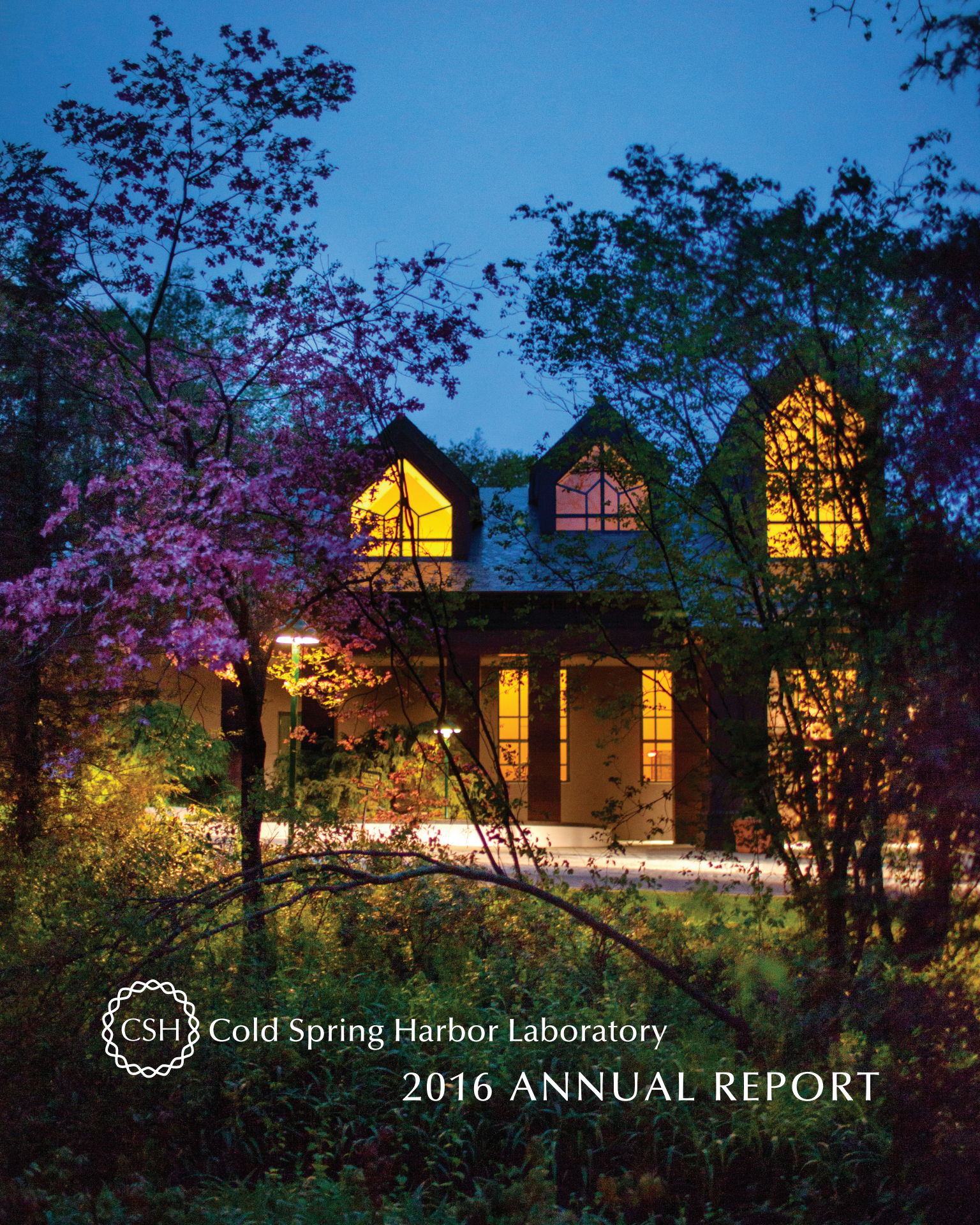
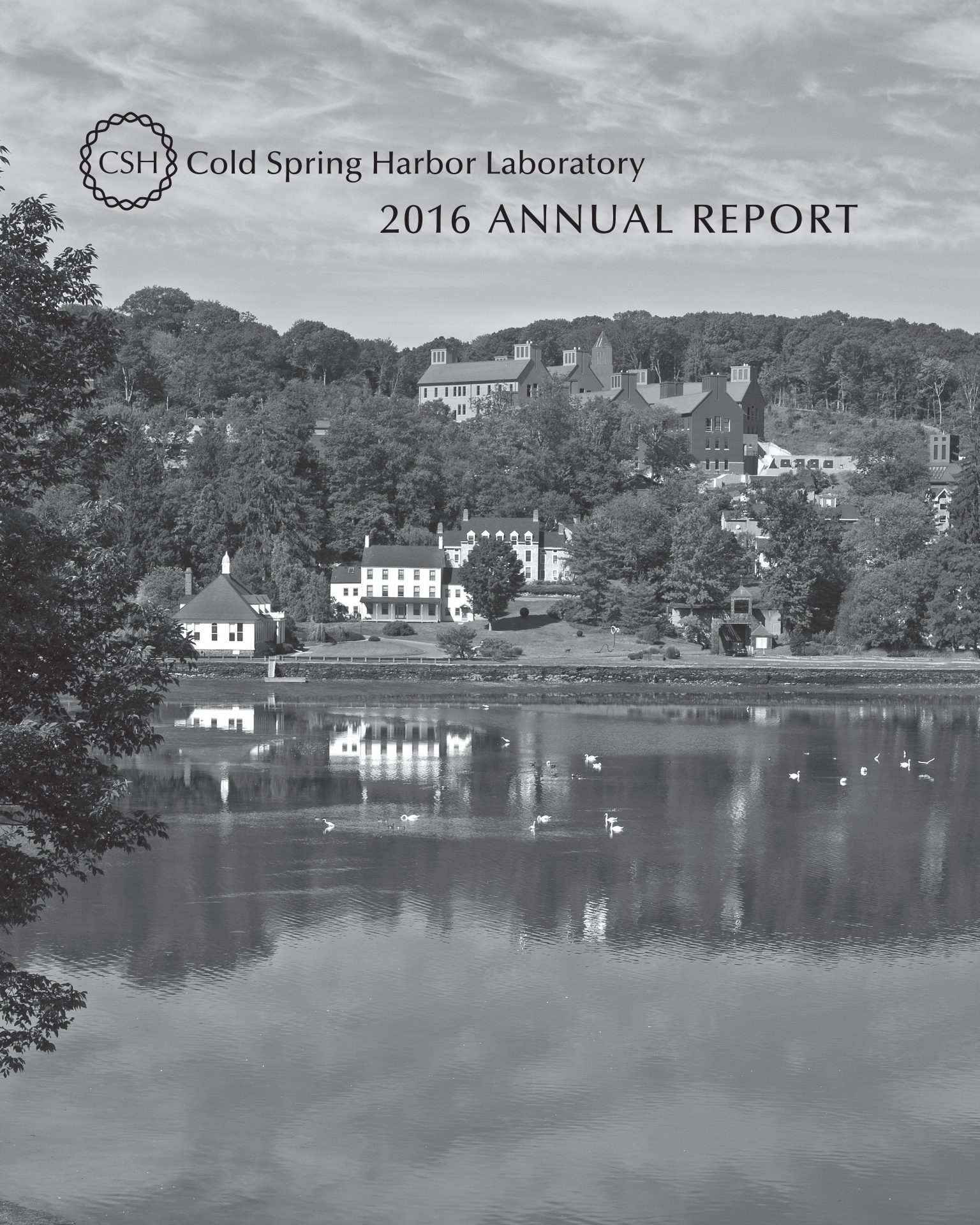


Cold Spring Harbor Laboratory

2016 ANNUAL REPORT


CSH
Cold Spring Harbor Laboratory
2016 ANNUAL REPORT

ANNUAL REPORT 2016

Cold Spring Harbor Laboratory
One Bungtown Road
Cold Spring Harbor, New York 11724

www.cshl.edu

Managing Editors Philip Renna, Dagnia Zeidlickis
Production Editor Kathleen Bubbeo
Science Writing Peter Tarr
Copy Editors Dorothy Brown, Joanne Dolce
Production Manager Denise Weiss
Nonscientific Photography Constance Brukin, Gina Motisi
Cover Designer Margot Bennett

Front cover: Photography by Constance Brukin

Contents

Officers of the Corporation

Jamie C. Nicholls, *Chairman*
Robert D. Lindsay, *Vice-Chairman*
Marilyn H. Simons, Ph.D., *Vice-Chairman*

Leo A. Guthart, *Treasurer*
Paul J. Taubman, *Secretary*

Bruce Stillman, Ph.D., *President and CEO*
W. Dillaway Ayres, Jr., *Chief Operating Officer*

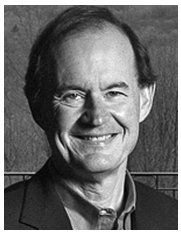
Board of Trustees

Individual Trustees

Lalit R. Bahl, Ph.D.
Senior Research Scientist, Renaissance Technologies Corp.

Tania A. Baker, Ph.D.
E.C. Whitehead Professor, Massachusetts Institute of Technology; Investigator, HHMI

Cornelia Isabella Bargmann
Investigator, HHMI; Torsten N. Wiesel Professor, The Rockefeller University, New York

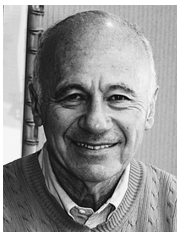


David Boies
Chairman, Boies, Schiller & Flexner LLP

Michael R. Botchan, Ph.D.
Richard and Rhoda Goldman Distinguished Professor of Biochemistry, University of California, Berkeley

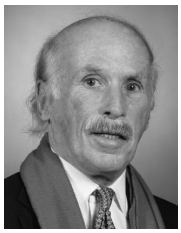
Charles I. Cogut
Senior M&A Counsel, Simpson Thacher & Bartlett LLP

Jacob Goldfield
New York, New York



Leo A. Guthart
Founder and CEO, Topspin Partners

Jeffrey E. Kelter
Senior Partner and CEO, KTR Capital Partners



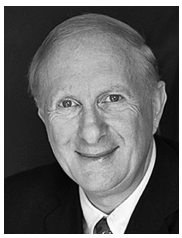
David M. Knott
Founder, Knott Partners

Laurie J. Landeau, V.M.D.
General Manager, Listowel, Inc.

Robert D. Lindsay
Co-Managing Partner, Lindsay Goldberg

Robert W. Lourie, Ph.D.
Head of Future Research, Renaissance Technologies Corp.

Elizabeth McCaul
Partner-in-Charge, New York Office, Promontory Financial Group, LLC



Howard L. Morgan, Ph.D.
Partner, First Round Capital; Director, Idealab

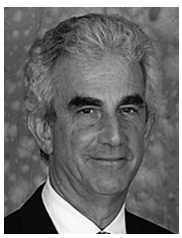
Jamie C. Nicholls
Chairman of the Board, New York, New York

Frank O'Keefe
Lloyd Harbor, New York

John C. Phelan
Managing Partner, MSD Capital, L.P.

Thomas C. Quick
Palm Beach, Florida

William S. Robertson
Robertson Foundation for Government



George Sard
Chairman and CEO, Sard Verbinnen & Co.

Thomas A. Saunders III
CEO, Ivor & Co., LLC

Charles L. Sawyers, M.D.
Investigator, HHMI; Chair, Human Oncology & Pathogenesis Program, MSKCC

Douglas Schloss
President and CEO, Rexford Management

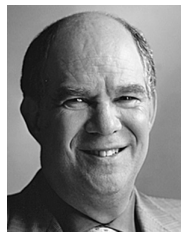
Individual Trustees *(continued)*

Marilyn H. Simons, Ph.D.
President, The Simons Foundation

Dinakar Singh
Founding Partner, TPG-Axon Capital

Samuel L. Stanley, Jr., M.D.
President, Stony Brook University

Bruce Stillman, Ph.D.
President, Cold Spring Harbor Laboratory

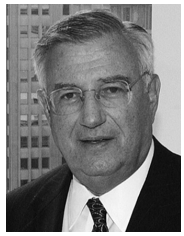


James M. Stone, Ph.D.
Chairman, The Plymouth Rock Company

Paul J. Taubman
PJT Partners, LLC

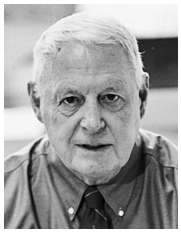
Edward Travaglianti
President, TD Bank Long Island

George Yancopoulos, M.D., Ph.D.
President, Research Laboratories, & CSO, Regeneron Pharmaceuticals, Inc.

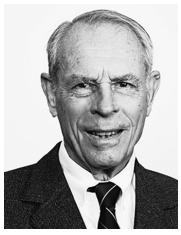


Roy J. Zuckerberg
Senior Director, Goldman Sachs & Co.

Honorary Trustees



Bayard Clarkson, M.D.
Memorial Sloan Kettering Cancer Center



Norris W. Darrell
Senior Counsel, Sullivan & Cromwell LLP

Charles F. Dolan
Cablevision Systems Corp.

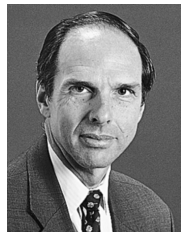
Helen Ann Dolan
Cablevision Systems Corp.

Lola N. Grace
Founder, Middle East Children's Institute

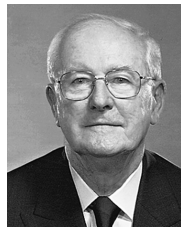
David H. Koch
Koch Industries, Inc.

Mary D. Lindsay
Laurel Hollow, New York

Nancy Abeles Marks
Nancy Marks Interiors and Carl Marks & Co., Inc.



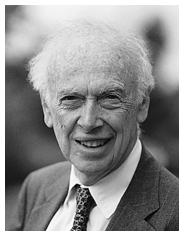
Eduardo G. Mestre
Vice Chairman, Evercore Partners



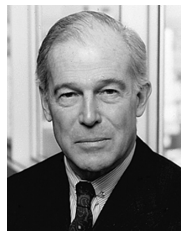
William R. Miller
New York, New York

Alan Seligson
NAK International

James Simons, Ph.D.
Chairman, The Simons Foundation



James D. Watson, Ph.D.
Chancellor Emeritus, Cold Spring Harbor Laboratory



Henry Wendt III
Friday Harbor, Washington

CSHL Scientific Advisory Council

The Scientific Advisory Council (SAC) is an external advisory group that advises the senior management of Cold Spring Harbor Laboratory (CSHL) on matters pertaining to science (both current and future), including the development of a research strategy to maintain CSHL as a world leader. The SAC is a nine-member Council, including a Chair of Council who is an individual known for scientific breadth and a detailed understanding of research management at the senior management level. The other eight members are world leaders in their respective fields and as such are able to provide advice on the different research areas of the Laboratory.

Frederick W. Alt, Ph.D. (Chair)
HHMI Investigator/Harvard University Professor and Co-Chief, Division of Molecular Medicine, Boston Children's Hospital

Steven Salzberg, Ph.D.
Johns Hopkins University School of Medicine, Professor of Medicine, Biostatistics and Computer Science

Kevan M. Shokat, Ph.D.
HHMI Investigator/University of California, San Francisco
University of California, Berkeley, Professor and Chair, Cellular and Molecular Pharmacology, Professor of Chemistry

Carol W. Greider, Ph.D.
Johns Hopkins School of Medicine, Professor and Director of Molecular Biology and Genetics

Angelika Amon, Ph.D.
HHMI Investigator/Massachusetts Institute of Technology
Kathleen and Curtis Marble Professor of Cancer Research, Koch Institute for Integrative Cancer Research

Eve Marder, Ph.D.
Brandeis University, Professor of Neuroscience, Biology Department and Volen Center

Markus Meister, Ph.D.
California Institute of Technology, Professor, Molecular and Cellular Biology

Joanne Chory, Ph.D.
HHMI Investigator/the Salk Institute for Biological Studies
Professor and Director, Plant Biology Laboratory

Max S. Wicha, M.D.
University of Michigan, Professor and Director, University of Michigan Comprehensive Cancer Center

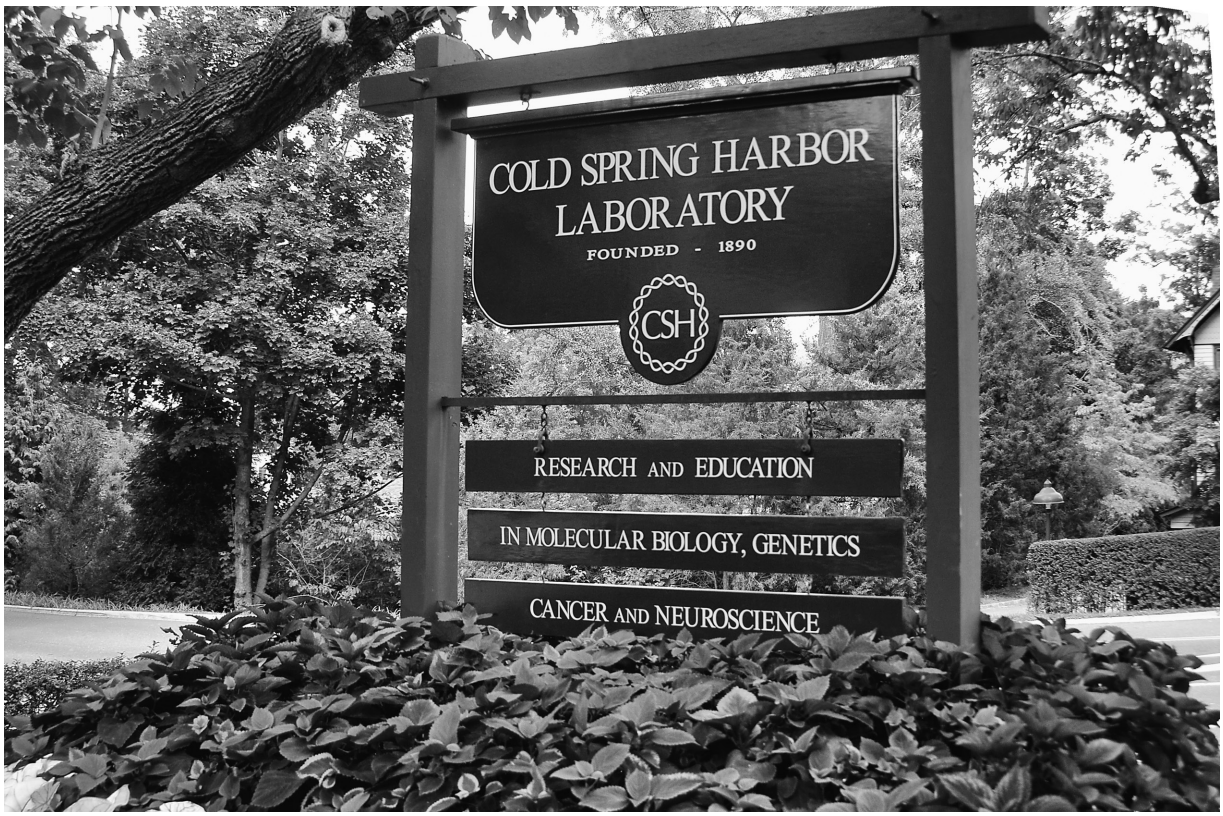
Governance

The Laboratory is governed by a Board of Trustees of up to 35 members that meets three or four times a year. Authority to act for the Board of Trustees between meetings is vested in the Executive Committee of the Board of Trustees. The Executive Committee is composed of the Officers of the Board and any other members who may be elected to the Executive Committee by the Board of Trustees. Additional standing and ad hoc committees are appointed by the Board of Trustees to provide guidance and advice in specific areas of the Laboratory's operations.

Representation on the Board of Trustees itself is divided between business and community leaders and scientists from major educational and research institutions.

The Laboratory is chartered as an educational and research institution by the Board of Regents of the Education Department of the State of New York. It is authorized to operate a graduate program under the name "Cold Spring Harbor Laboratory, Watson School of Biological Sciences" and thereat to confer the degrees of Doctor of Philosophy (Ph.D.), Master of Science (M.S.), and Doctor of Science (Sc.D.), Honorary.

It is designated as a "public charity" under Section 501(c)(3) of the Internal Revenue Code.



Committees of the Board

Academic Affairs

James M. Stone, Ph.D., Chair
W. Dillaway Ayres, Jr.
Lalit Bahl, Ph.D.
Tania A. Baker, Ph.D.
Michael R. Botchan, Ph.D.
Laurie J. Landeau, V.M.D.
Robert Lourie
Bruce Stillman, Ph.D.

Audit

Leo A. Guthart, Chair
Elizabeth McCaul
Edward Travaglianti

Commercial Relations

James M. Stone, Ph.D., Chair
W. Dillaway Ayres, Jr.
Robert Martienssen, Ph.D.
Bruce Stillman, Ph.D.

Development

Marilyn H. Simons, Ph.D., Chair
W. Dillaway Ayres, Jr.
David Boies
Charles I. Cogut, J.D.
Kristina Perkin Davison
Jeffrey Kelter
Laurie J. Landeau, V.M.D.
Howard Lee Morgan, Ph.D.
Jamie C. Nicholls
George Sard
Dinakar Singh
Bruce Stillman, Ph.D.
Paul J. Taubman
Roy J. Zuckerberg

DNA Learning Center

Laurie J. Landeau, V.M.D., Chair
W. Dillaway Ayres, Jr.
Charles I. Cogut, J.D.
Kristina Davison
Jeffrey Kelter
David A. Micklos
Jamie C. Nicholls
Ed Rover
Marilyn H. Simons, Ph.D.
Bruce Stillman, Ph.D.

Executive

Jamie C. Nicholls, Chair
W. Dillaway Ayres, Jr.
Charles I. Cogut, J.D.
Leo A. Guthart
Marilyn H. Simons, Ph.D.
Bruce Stillman, Ph.D.
James M. Stone, Ph.D.
Paul J. Taubman
Edward Travaglianti
Roy J. Zuckerberg

Finance

Leo A. Guthart, Chair
W. Dillaway Ayres, Jr.
Jamie C. Nicholls
William S. Robertson
Lari C. Russo
Marilyn H. Simons, Ph.D.
Edward Travaglianti

Investment

John C. Phelan, Chair
W. Dillaway Ayres, Jr.
Jamie C. Nicholls
Thomas Purcell
William S. Robertson
Thomas A. Saunders III
Dinakar Singh
Edward Travaglianti
Remy Trafelet

Nominating

Paul J. Taubman, Chair
W. Dillaway Ayres, Jr.
Charles I. Cogut
Leo A. Guthart
Robert D. Lindsay
Jamie C. Nicholls
Thomas A. Saunders III
Bruce Stillman, Ph.D.
Roy J. Zuckerberg

Robertson Research Fund

W. Dillaway Ayres, Jr.
Walter Goldschmidts
William Gridley
Leo A. Guthart
John H. Linnartz
Jamie C. Nicholls
William S. Robertson
Bruce Stillman, Ph.D.

Scientific Advisory Council

Fred Alt, Ph.D., Chair
Angelika Amon, Ph.D.
Cori Bargmann, Ph.D.
Joanne Chory, Ph.D.
Carol Greider, Ph.D.
Leonid Kruglyak, Ph.D.
Markus Meister, Ph.D.
Kevan Shokat, Ph.D.
Max Wicha, M.D.

PRESIDENT'S REPORT

Nurturing young scientific talent is a defining strength of Cold Spring Harbor Laboratory. The stress we place on investing in the rising generation explains unique aspects of our institution's culture. It likely also helps account for the outsized contribution that the Laboratory has made to modern biology.

The support we provide for young scientists, although both costly and risky, has consistently paid rich dividends. We had a spectacular reminder of this in 2016, a year that marks the 30th anniversary of the Cold Spring Harbor Fellows program.

My Own Experience

When I first arrived at CSHL in 1979, straight from graduate school in Australia, I was thrust into an environment foreign to the way most science is organized today. Then, senior scientists like CSHL Director Jim Watson and Assistant Director Joe Sambrook mentored many early-career scientists, and a select few were granted the freedom to pursue their own ideas and independently publish. Collaborations were encouraged, allowing individuals to achieve more than they might if working on their own.

Jim Watson offered me the opportunity to work on my own independent project, but I also was fortunate to have close mentorship from Mike Mathews, who, like me, was able to pursue his own interests unencumbered by the burdens of seeking research grants. Others, including Mike Botchan, Bob Tjian, David Lane, Ed Harlow, Mike Wigler, and Doug Hanahan, were granted similar independent opportunities at CSHL after completion of their graduate studies. Each was appointed to the Laboratory with a conviction that placing bright young people to work on thorny scientific problems in a highly collaborative atmosphere would have the highest impact.

The First Fellows

The CSHL Fellows program was established in 1986 to formalize our ability to attract very talented scientists at the earliest stage of their career. It was modeled after the prestigious Junior Fellows program at Harvard University, but, unlike that program, we provided space and substantial funds so that CSHL Fellows could pursue experimental research. In 1986, Richard Roberts, then a senior faculty member, and Jim Watson succeeded in recruiting a gifted young scientist from Harvard named Adrian Krainer. The very first CSHL Fellow, Adrian had just completed his doctorate under the mentorship of Harvard's Tom Maniatis, a former member of CSHL's faculty and then a member of the CSHL Board of Trustees. In 1986, Krainer could have pursued a conventional postdoctoral research position within a lab at any number of first-rank institutions. Instead, he came to CSHL because of the freedom that the Fellows program offered, backed up by the nurturing of Rich Roberts, a highly successful scientist who contributed amazing discoveries in the field Adrian was entering.

Forty years ago, Roberts, Louise Chow, and Phillip Sharp were the first to notice "split genes"—a phenomenon that we now refer to as RNA splicing. Although Roberts, Chow, and Sharp had come to Cold Spring Harbor at the beginning of their own independent careers in the early 1970s to study tumor viruses, Roberts, Chow, and colleagues at Cold Spring Harbor and separately Sharp, who moved to the Massachusetts Institute of Technology (MIT), discovered a phenomenon that transcended the cancer research in which they were engaged. Their research revealed that RNA messages copied from the genetic code in DNA are edited, or spliced, before the resulting

messenger RNAs (mRNAs) are translated into proteins. This discovery, recognized with a Nobel Prize in Physiology or Medicine conferred upon Roberts and Sharp in 1993, revolutionized our understanding of how genes are organized in our genome and how they evolved.

Adrian Krainer

As the first CSHL Fellow, Adrian worked on perfecting a method, called a cell-free system, that he had developed in his graduate work to investigate the cellular machinery involved in mRNA splicing. Here, Adrian was given a wide berth to follow a subject about which he was passionate. He seized the opportunity and began the painstaking process of isolating different components of the splicing machinery. He was not alone in this branch of research, but from the outset, it was clear to me that he would become a leader.

The Laboratory's investment in Adrian has been repaid in many ways over 30 years, but in 2016, both he and the Laboratory had the special satisfaction of seeing basic-science discoveries made in his lab come to fruition in the form of a drug that remedies a deadly neuromuscular disease of childhood called spinal muscular atrophy, or SMA. The day before Christmas of 2016, this drug, first identified by Krainer's team as ASO 10-27 and later developed and commercialized by Ionis Pharmaceuticals and Biogen as Spinraza®, was approved by the FDA for use in SMA patients.

Compensating for a genetic defect that reduces the amount of a particular protein in neurons that innervate muscles, the drug is already saving lives. It is a great example of why not only the Laboratory but the government of the United States make significant investments in basic research. Each CSHL Fellow is appointed for 3 years with full financial support from CSHL. Afterward, they can move up our academic ranks or transition to another institution. The CSHL Fellows program has launched many fine careers.

Carol Greider

Carol Greider, our second CSHL Fellow, began her independent career here in 1988, making discoveries about the structure of the enzyme that maintains telomeres, the ends of our chromosomes. These independent studies, along with her pioneering graduate studies at the University of California, Berkeley, were integral to the body of work recognized by her sharing the 2009 Nobel Prize in Physiology or Medicine.

A Youth Culture

There is an old expression that "youth is wasted on the young." Nothing could be further from the truth in science, which relies upon the young to move entire fields forward. It may sound paradoxical, but one of the strongest reasons for investing in young scientists has to do with what they do not know. I have come to appreciate this during my many years as president. Young people come into science with enthusiasm, fresh ideas, and the willingness to work hard. The very best of them are not burdened with current dogma.

An advantage of what I call the "naïveté of youth" is that they can pose important questions—some that may well have been asked by investigators of prior generations, but they do so unencumbered with accumulated bias that often slows the progress of more senior scientists. Because of the continual march of technology, young people can explore big questions with a fresh approach, using methods and machines that their predecessors did not possess. This story is repeated, over and over. Our young people are right now making discoveries and developing new technologies that will inform their science and empower the generation that follows them.

A start-up package today typically includes a guarantee by the Lab to cover up to $2.5 million in direct costs for the incoming faculty member, and with all factors included, it involves an expenditure of some $4 million over 5 years. That's a substantial commitment, but one we are happy to make because we want to see people succeed. During the initial 5-year period, we expect a junior faculty member to secure grants from government and private funders sufficient to support their independence. At the same time, we are deeply committed to supporting innovative aspects of their research, by connecting them with philanthropic supporters or funding from our endowment.

CSHL has a youth culture. Although we depend on senior faculty to mentor and guide the careers of our junior faculty, the ratio of senior to junior members is approximately 2 to 3. At peer institutions, the ratio is weighted heavily in the opposite direction: At Rockefeller University and the Salk Institute, for instance, there are seven or more senior faculty for every two or three junior faculty. Ours is a culture that bets on early-career scientists; our peers reward past achievement.

The same can be said of the grant culture at the National Institutes of Health (NIH). On average, a scientist receiving a first-time research (R01) grant does not receive it until age 42. The prerequisite is that the grantee is expected to have preliminary results; the NIH does not, presently, support people—rather, it supports projects. This is not the case at CSHL, where we give the brightest young people a chance to become successful. That we often succeed is reflected in the rapidity with which we promote junior faculty, compared with the much slower rate at universities.

A Community of Science

The Laboratory as a community of science is one of the strongest factors in our ability to perennially draw the best young talent. Our Meetings & Courses Program brings about 10,000 members of the biology community to our campus every year. This places everyone on our faculty and scientific staff right at the confluence of the many great rivers of information flowing into our field. My own exposure to the CSHL yeast course changed the direction of my career, and there are many other similar stories. Cold Spring Harbor is a place where you can branch out, change focus, and add new areas of expertise to the one for which you were initially recognized. This is part of what it means to have a youth culture.

Our institution also has remarkable state-of-the-art shared research resources, which enable faculty, students, and postdocs to use the most sophisticated apparatus and techniques. Such experimental resources not only promote collaboration but also allow small groups access to technologies that would be difficult to establish in their own lab. This is in part why a CSHL Fellow can be successful.

For all of its virtues, this is an expensive proposition that carries a fair amount of risk. When we invest in smart young investigators, we can never be certain of their success, and we know that they may leave us, to be replaced by another young person of talent and promise. This movement keeps us on the leading edge, but it means that we must keep investing in people, at costs that have soared over the last decade. Our Fellows program is supported only by our annual fund-raising, and to sustain it over the long run, we need substantial endowment funds to provide a constant level of support.

A Model for Others

Judging from our successful Fellows program and the similar program at the Whitehead Institute at MIT, investment in early career independence can lead to great science. I suggest that more institutions establish formal programs that allow for this opportunity. The NIH has introduced a funding mechanism to support early-career scientists straight out of graduate school, but it suffers

because these Fellows must first be appointed to a position at a university or research institute. Many universities do not have an organized environment to nurture a fellows program.

Mentoring is critical for an early-career independent researcher to succeed, as is access to shared research resources. But even before that, there needs to be a strong institutional commitment that encourages the best graduate students to pursue an independent career rather than continuing to train with a senior scientist. Although these positions are not for everyone, for the right person, immediate independence can be liberating and lead to great scientific achievements. Our nation should explore more opportunities for such talented people.

Bruce Stillman, Ph.D., F.R.S.
President and Chief Executive Officer

Highlights of the Year

Research

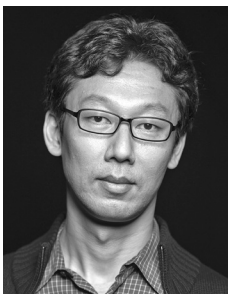
In 2016, the hundreds of scientists who work in Cold Spring Harbor Laboratory's 50-plus laboratory groups led or contributed to work disseminated in hundreds of peer-reviewed scientific papers. This research represents the full gamut of CSHL research activity, in Cancer, Neuroscience, Plant Biology, Quantitative Biology, and Genomics. Although it is impossible in a small space to adequately represent the scope of this work, its richness is suggested in the following sample of important findings.

Evidence of Earlier Neanderthal–Modern Human Mating


A. Siepel

We have been aware for some time that Neanderthals and humans mated at some point in their shared past. Initial analysis suggested such contacts occurred some 47,000–65,000 years ago. This year, an international team co-led by Adam Siepel reported an interbreeding event more than 100,000 years ago. This finding, the result of several kinds of advanced computer modeling algorithms comparing complete genomes of hundreds of contemporary humans with complete and partial genomes of four archaic humans, has implications for our knowledge of human migration patterns. The new data, interestingly, shows a signal of breeding in the opposite direction from that already known: bits of human DNA in the genome of a Neanderthal individual. This is in contrast to fragments of Neanderthal DNA in human genomes, which we have known about for years. The team's evidence of "gene flow" from descendants of modern humans into the Neanderthal genome applies to one specific Neanderthal. The modern human sequences in this "Altai Neanderthal" appear to derive from a group of modern human ancestors from Africa that separated early from other humans around the time present-day African populations diverged from one another, about 200,000 years ago.

Insights about NMDA Receptors in Action


H. Furukawa

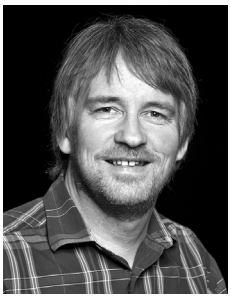
Hiro Furukawa and colleagues continue to make discoveries about the structure of NMDA (*N*-methyl-D-aspartate) receptors in the brain, whose dysfunction is linked to depression, schizophrenia, Alzheimer's, and other illnesses. In two published papers, they shed new light on how the various segments, or domains, of the receptor change shape when the receptor is activated, and how structural differences in variants of specific domains account for differences in how they interact with zinc, an important NMDA regulator. To do this, the team combined two molecular imaging techniques, X-ray crystallography and single-particle electron cryo-microscopy, and observed structures of the receptor in three specific configurations: the activated, nonactive, and inhibited states. Superimposing the crystal structure of the receptor in each of these states revealed which components move—typically by rotating slightly relative to one another—when the ion channel opens, a movement similar to the opening of a camera shutter. They also solved the mystery of why zinc binds much more readily to the "A" variant of the receptor as compared with the "B" form, which is structurally almost identical. Further, they explained why the important candidate drug ifenprodil binds at a site in the "B" form but not in the "A" form. These discoveries will help make receptor-modulating drugs more specific.

Statistical Assessments Underlie Feelings of Confidence

A. Kepecs

Adam Kepecs and his team suggest that the brain is constantly processing data to make statistical assessments that translate into the feeling we call confidence. This feeling of confidence, they assert, is central to decision-making, and, despite ample evidence of human fallibility, the subjective feeling relies on objective calculations, akin to the statistical computations a computer would make. If we did not have the ability to optimally assess confidence, we would routinely find ourselves in a state of indecision, or worse. In experiments with human subjects, Kepecs and colleagues tried to control for different factors that can vary from person to person. The aim was to establish what evidence contributed to each decision. In this way, they could compare people's reports of confidence with the optimal statistical answer. They created video games to compare human and computer performances. Participants rated confidence in each choice they were asked to make on a scale of 1 (a random guess) to 5 (high confidence). They found that human responses were similar to statistical calculations. The brain produces feelings of confidence that inform decisions the same way statistics pulls patterns out of noisy data. It is Kepecs' thesis that statistics—generated by the objective processing of sensory and other data—is the ultimate language of the brain.

New Stem Cell Pathway Points to Higher Staple Crop Yields



D. Jackson

A discovery by David Jackson's team explains how plants regulate the proliferation of their stem cells. This has implications for increasing the yield of maize and other staple crops, perhaps by as much as 50%. The newly discovered regulatory pathway is notable in that it channels signals emanating from a plant's extremities—emerging young leaves called primordia—to the stem cell niche, called the meristem, located at the plant's growing tip. Plant biologists have long known of another pathway, called the CLAVATA-WUSCHEL pathway, that regulates stem cell proliferation from within a portion of the meristem itself, called the organizing center (OC). WUSCHEL is a transcription factor that promotes stem cell proliferation. In the CLAVATA-WUSCHEL pathway, stem cells send back to the OC a negative signal, repressing the signal for proliferation. A similar feedback is established in the newly discovered pathway, although the signal begins in leaves. Having a signal coming from the leaves could act as a kind of environmental sensor, telling totipotent stem cells in the meristem to stop proliferating—a brake, applied from the older, more developed parts of the plant, for example, in response to environmental cues. In maize, the pathway encompassing FEA3—the receptor for the signal from the leaves—and its ligand, FCP1, which the team also discovered, is highly conserved across the plant kingdom. This fact points to the possibility of tweaking the components to achieve significant increases in yield in all the major staple crops.

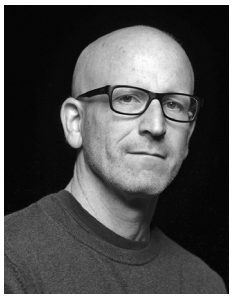
Killing Pancreatic Cancer Cells by Raising Antioxidant Levels

D. Tuveson

Most people have been taught to think that raising antioxidant levels in the body tends to keep cancer at bay. But David Tuveson's lab has demonstrated that in the specific context of pancreatic cells on the road to cancer or already in a malignant state, this is the last thing one wants to do. They find that reducing levels of antioxidants in pancreatic cancer cells can help kill them. More oxidants are being made in malignant cells, but more antioxidants are being made, too, countering the impact of rising oxidation. Without commensurately more antioxidants, malignant cells will die due to excessive oxidation. This suggests a new treatment strategy for the notoriously lethal illness, in which less than 9% of patients survive 5 years. The team focused on a protein called NRF2, a master gene-regulating protein one can tweak to disturb the redox balance in cancer cells. When NRF2 is active, cells synthesize glutathione, an important antioxidant. Can we reduce NRF2 activity or knock it out of action altogether? The team used pancreas organoids to show that when NRF2 is missing, the machinery that translates gene messages into proteins is

very sensitive to fluctuations in oxidant/antioxidant balance. Crucially, protein synthesis was not impacted in normal pancreas cells, indicating an opportunity to capitalize on synthetic lethality—circumstances in which a condition affecting all cells only kills unhealthy ones. The team is now testing various combinations of anticancer agents and glutathione inhibitors in organoid systems with the aim of advancing a novel form of treatment to the clinic.

A Revolutionary Brain-Mapping Method



A. Zador

Anthony Zador and colleagues have published a revolutionary way of mapping the brain at the resolution of individual neurons, which they successfully demonstrated in the mouse brain. Called MAPseq (multiplexed analysis of projections by sequencing), this approach makes it possible in a single experiment to trace the long-range projections of large numbers of individual neurons from a specific region or regions—using much less labor, time, and money than current methods require. MAPseq differs from "bulk tracing" methods now in common use, in which a marker is expressed by neurons and carried along their axons. These markers are good at determining all of the regions where neurons in the source region project to, but they cannot tell scientists that any two neurons in the source region project to the same region, to different regions, or to some of the same regions and some different ones. Zador's technology assigns unique barcode-like identifiers to large numbers of individual neurons via a single injection in any brain region of interest. Each injection consists of a deactivated virus that has been engineered to contain massive pools of individually unique RNA molecules, each of whose sequence—consisting of 30 nucleotides—is taken up by single neurons. The team published a proof-of-concept based on a test of MAPseq showing projections from the locus coeruleus (LC) in the mouse brain. The method will soon be used to map the brains of animals that model various neurodevelopmental and neuropsychiatric illnesses, to see how gene mutations strongly associated with causality alter the structure of brain circuits—and thus, presumably, brain function.

How the Brain Evaluates Results of Our Actions

B. Li

Much of what we do from day to day and even minute to minute is based on our evaluations—our ability to determine whether the consequences of our actions are better or worse than what we expected. But where in the human brain is evaluation performed? How do such evaluations inform our actions? Bo Li's team uncovered a neural circuit that processes evaluations and has succeeded in identifying its sources. They explain how choices are reinforced based on the results of our actions and how we assess those results. The team focused on substructures within the basal ganglia, "nuclei" in the forebrain that include the striatum and the globus pallidus. They discovered that a distinct grouping of neurons within the globus pallidus mediates the evaluation of outcomes. The area containing the set of neurons in the globus pallidus has been named the habenula-projecting globus pallidus, or GPh, which the team discovered is exclusively connected to a tiny structure nearby called the lateral habenula (LHb). In the GPh, information of opposing valence is integrated to determine whether the outcomes of a given action are better or worse than expected. Li's team will now look at the newly traced circuit in mouse models of depression, to see if the circuit as already traced in nondepressed mice is in any way altered.

Long Noncoding RNAs Play a Role in Cancer

Remarkably, only ~2% of the human genome encodes proteins. Nearly 80% of the rest of the genome is transcribed into RNA that does not code for proteins. David Spector and colleagues are learning how some of these RNAs play a role in cancer. They screened thousands of noncoding RNAs to find those expressed at high levels in two types of aggressive breast cancer. When they reduced the level of some of the most overexpressed of these RNAs from mammary tumor

D. Spector

samples, cellular features characteristic of cancer spread were significantly reduced. At first, the team found several hundred long noncoding RNAs (lncRNAs) that were expressed at higher than normal levels in both types of aggressive mouse tumors that they tested. Computational analysis enabled them to prioritize a subset of 30, dubbed mammary-tumor-associated RNAs, or MaTARs. With Ionis Pharmaceuticals, Spector's team designed a series of molecules that bind tightly to, and thereby destroy, specific RNA sequences. They used these antisense molecules to wipe out individual MaTARs in mammary-cancer-derived organoids. They found that individually eliminating 20 of the 30 MaTARs in these organoids diminished features associated with cancer, including cell proliferation, invasion, and migration. The team's next step is to administer antisense molecules to degrade specific MaTARs in mice, in the hope that this will decrease primary tumor mass and/or metastasis.

NETs Deployed by Immune Cells Are Hijacked by Cancers

M. Egeblad

A discovery by Mikala Egeblad and colleagues reveals how neutrophils, the most common type of white blood cell, can be "hijacked" by cancer cells. Using live-imaging technology, her team revealed that a remarkable weapon sometimes deployed by neutrophils against invaders like bacteria can aid metastasis. This astonishing weapon appropriated by cancer cells is a lattice of DNA, ejected from an activated neutrophil upon detection of a threat. These neutrophil extracellular traps or NETs form spider-web-like structures outside the neutrophil. The DNA that forms the backbone of the web is studded with tiny toxic enzymes that can degrade and digest invaders. The team found that cancer cells were able to induce nearby neutrophils to eject their NETs even when no infection or invader was present. Egeblad thinks NETs help aggressive cancer cells by literally eating through the proteins that form a tissue's scaffolding, thus opening up small holes and crevices that cancer cells can occupy. This can be a first step in forming a cancer colony at a site distant from the primary tumor. Working with Dr. Michael Goldberg at the Dana-Farber Cancer Institute, the team hitched the NET-degrading enzyme DNase to nanoparticles and directed these against triple-negative breast cancer in mice, markedly reducing, and for some mice even preventing, metastases to the lung. DNase is already approved for cystic fibrosis patients, so this approach has considerable translational potential.

CRISPR Yields Tomatoes That Flower and Ripen Weeks Earlier

Z. Lippman

Using a simple and powerful genetic method to tweak genes native to two popular varieties of tomato plants, a team led by Zachary Lippman has devised a rapid method to make them flower and produce ripe fruit more than 2 weeks faster than commercial breeders are currently able to do. This means more plantings per growing season and thus higher yield. In this case, it also means that the plant can be grown in latitudes more northerly than currently possible—an important attribute as the earth's climate warms. At the heart of the method are insights obtained by Lippman and colleagues about the evolution of the flowering process in many crops and their wild relatives as it relates to the length of the light period in a day. The hormone florigen and a counteracting "anti-florigen" hormone called SP (for SELF PRUNING) act together, in yin-yang fashion, to, respectively, promote or delay flowering. The team traced a loss of day-length sensitivity in domesticated tomatoes to mutations in a gene called *SP5G* (SELF PRUNING 5G). They discovered that although domesticated plants are insensitive to day length, there remains some residual expression of the antiflorigen *SP5G* gene. They used CRISPR to induce tiny mutations in the *SP5G* gene, inactivating it entirely. When this tweaked

gene was introduced to roma and cherry tomato varieties, the plants flowered earlier, and hence made fruits that ripened earlier. Tweaking another antiflorigen gene that makes tomato plants grow in a dense, compact, shrub-like manner made the early-flowering varieties even more compact and early-yielding. Lippman has thus demonstrated a means of "fast-forward breeding" that could enable growers to expand a plant's geographical range of cultivation, a valuable ability in a period of rapid climate change.

Research Faculty

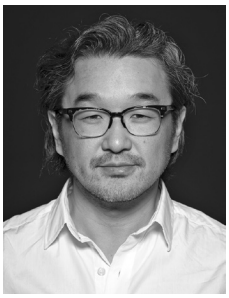
Awards

Associate Professor Bo Li and Assistant Professor Je Lee are members of international teams that won the 2016 Human Frontiers Science Program Research Project Grants. Bo's project is "Single-cell resolution imaging and optogenetics in the amygdala fear circuit in behaving animals." Je's project is "Complete cell lineage trees inferred by in situ genotyping of induced somatic mutations."

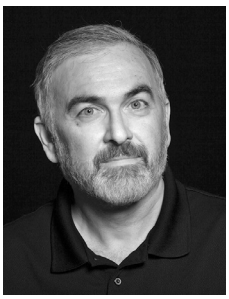
Professor Adrian Krainer was elected as one of 213 new members of the 236th class of the American Academy of Arts and Sciences. This organization's members include some of the world's most accomplished scholars, scientists, writers, and artists.

Associate Professor Chris Vakoc, M.D., Ph.D., was awarded the third annual Pershing Square Sohn Prize for Young Investigators in Cancer Research. A finalist in last year's prize, Chris' research employs a novel CRISPR technique that can reveal individual protein domains that sustain cancer cells.

Assistant Professor Camila dos Santos was selected as one of seven 2016 Rita Allen Foundation Scholars. She has found dramatic differences in the pace of breast development between first and second pregnancies, which appear to be mediated by epigenetic mechanisms—molecular changes that affect gene expression without altering DNA sequences. The award will allow the dos Santos lab to assess the relevance of these phenomena for breast cancer risk, which is 30% lower in women who have a full-term pregnancy before the age of 25.



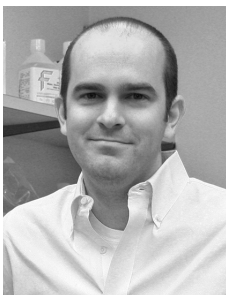
J. Lee



A. Krainer

Promotions and New Hires

Congratulations to Chris Hammell, who was promoted to Associate Professor. The Laboratory welcomed Ullas Pedmale, Assistant Professor, and Tsung Han Yeh, Research Assistant Professor. Rebecca Leshan, Ph.D., was recruited as the new Director of the Banbury Center. Michael Marchesiello is the new Vice President, Procurement.



C. Vakoc

CSHL–Northwell Affiliation 2015–2016 Progress

With the goal of bringing basic research insights into the clinic, the CSHL–Northwell Health affiliation links clinicians and scientists. In its second year, the affiliation supported more than 20 new projects at CSHL that have clinical implications. The first joint clinical trial in breast cancer was initiated from research in Nick Tonks' laboratory.

To facilitate future clinical developments, Dr. Robert Maki was named the first Director of the Center for New Cancer Therapies at the Northwell Health Cancer Institute. He holds a joint appointment as CSHL Professor and member of the CSHL National Cancer Institute (NCI)-designated Cancer Center.

The affiliation has launched new education initiatives, including a new research fellow track in clinical medical oncology and laboratory investigation and a new summer research program for Hofstra Northwell School of Medicine students. The first class of

C. dos Santos

J. Witkowski and R. Leshan

R. Maki

this program trained in the laboratories of David Tuveson, Michael Wigler, Camila dos Santos, and Mickey Atwal.

Half-day retreats held at the Banbury Center have seeded research collaborations between the two institutions. The topics of these meetings have included chronic lymphocytic leukemia and cancers of the prostate, pancreas, brain, breast, and lung.

CSHL Cancer Center

Professor David Tuveson, M.D., Ph.D., was named the new Director of the CSHL Cancer Center. Bruce Stillman held that role for 25 years, and the transition allows him to continue to focus on his long-held role as CSHL President and CEO.

D. Tuveson

Together with the Coalition Against Childhood Cancer (CAC2), CSHL hosted the first-ever "From Bench to Bedside and Beyond" conference convening a wide range of childhood cancer community stakeholders to advance research. Bruce Stillman opened the 2-day program, followed by presentations from Chris Vakoc and Mickey Atwal. Other speakers represented the National Cancer Institute, the nation's top basic and comprehensive cancer centers and the pharmaceutical industry. Other sponsors included Alex's Lemonade Stand Foundation, Bristol-Myers Squibb, United Therapeutics, and Amgen.

CAC2 PEDIATRIC CANCER RESEARCH CONFERENCE 2016

The CAC2 Conference logo

Business Development and Technology Transfer

The ability of CSHL to make game-changing discoveries and apply basic research in developing treatments for disease was demonstrated by FDA approval of the drug Spinraza™ at the end of 2016. This drug, based on the use of antisense oligonucleotide technology that utilizes and a short, RNA-based molecule that targets RNA splicing, is the first effective treatment for the lethal disease called spinal muscular atrophy (SMA). It took more than a decade for Dr. Adrian Krainer and a team of postdocs to understand the fundamental biological process of RNA splicing and convert that knowledge into a drug to counteract the genetic defect in children with SMA.

In 2008, Ionis Pharmaceuticals licensed technology from CSHL based on research conducted in the laboratory of Dr. Krainer, along with his postdoctoral fellow, Dr. Yimin Hua. Ionis

subsequently entered into an agreement with Biogen for development and commercialization of Spinraza™. In 2011, Ionis received FDA permission to begin clinical trials. On December 23, the FDA approved nusinersen, which will be sold by Biogen under the brand name Spinraza™.

The Business Development and Technology Transfer program supports Laboratory scientists interested in working with industry and investors who may need materials or transaction support to enable their research. Andrew Whiteley joined the team as the first CSHL Executive in Residence.

In 2016, the program concluded:

- 10 new license and option agreements
- 11 industry-sponsored research agreements totaling more than $4 million in new funds
- 10 new technology cases
- 18 new patent filings

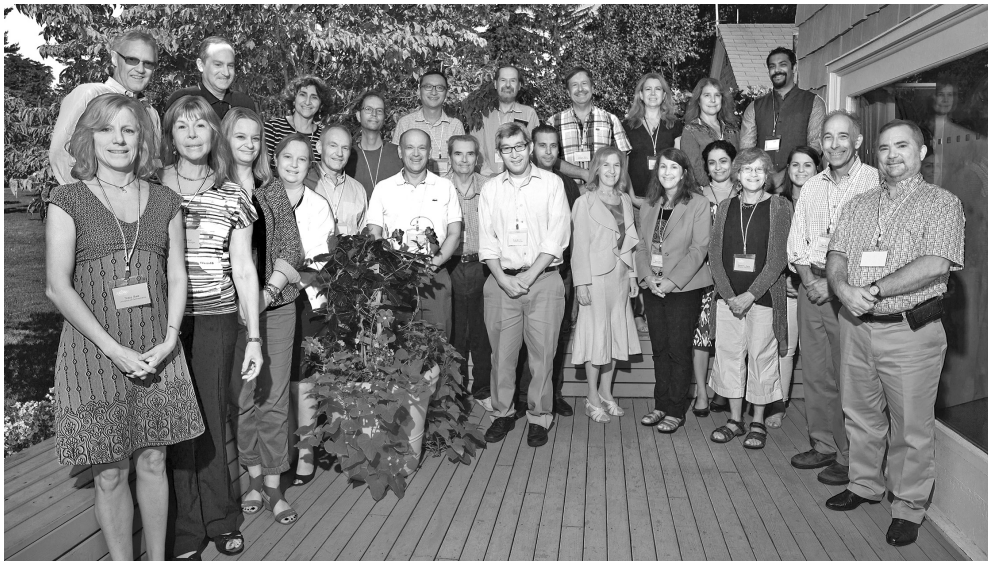
A. Krainer and Y. Hua

Education Programs

Banbury Conference Center

The Laboratory's science policy think tank, the Banbury Center, held 42 events, including 20 meetings drawing almost 600 participants. Several meetings were devoted to the topic of mental disorders. One problem that has continued to vex researchers—how to make accurate animal models of such disorders—was the topic of a meeting organized by Dr. Eric Nestler, a leading expert in the field. "Can We Make Animal Models of Human Mental Illness?" considered what behaviors in rodents might be taken as surrogates for behaviors in people.

Another highlight was a meeting devoted to Lyme disease, an example of an emerging infectious disease. Participants at "Diagnostic Tests for Lyme Disease: A Reassessment" made substantial progress in agreeing on next steps toward development of such tests, which are urgently needed.



Participants at Banbury meeting "Can We Make Animal Models of Human Mental Illness?," August 2016

Every year since 1998, a meeting on plant science has been convened at Banbury. This year's meeting focused on "Genomics-Based Accelerated Crop Breeding" and considered the impact of new gene-editing methods such as CRISPR on improving crop performance. Several meetings considered various aspects of cancer research, including possible methods of inhibiting transcription factor STAT-3 and analyzing ways of reducing the toxicity of oxidative chemotherapy.

Topices of focus at this year's meetings included the relationship between DNA science and archaeology and cellular diversity in the mammalian brain, among other subjects. Banbury meetings, which have historically convened diverse groups of experts from many disciplines around a challenge in biological and science policy, are made possible by generous support from individuals, foundations, and corporate sponsors.

Meetings & Courses Program

With roots in the famous Cold Spring Harbor Symposia series, dating to 1933, the scientific conference and advanced technology course program currently includes 55 meetings and 30 to 35 lab and lecture courses held over 2-year cycles. Since 2009, a parallel program of meetings has been held in Suzhou, China, under the aegis of the Cold Spring Harbor Asia (CSHA) program, a wholly owned subsidiary of the Laboratory.

Abstract book for Symposium 81: *Targeting Cancer*

In 2016, the CSHL Meetings attracted 7250 participants to the main campus. The 81st Cold Spring Harbor Symposium focused on Targeting Cancer, reflecting the enormous research progress achieved in recent years. It attracted almost 490 participants, including many of the world's leading cancer researchers. Several new meetings included Transposable Elements and the annual history of molecular biology meeting, which this year addressed HIV/AIDS Research. The program is supported by grants from the National Institutes of Health and the National Science Foundation, as well as the newly invigorated Corporate Sponsor Program.

In its seventh year of operation, the CSHA program held 17 scientific conferences in Suzhou and also arranged meetings on immunology and plant biology in Awaji, Japan, attracting more than 3500 scientists. This program is designed for scientists from the Asia/Pacific region, who make up more than 80% of attendance, and includes symposia, meetings, and Banbury-style discussion meetings. Support comes from a major sustaining grant from the Suzhou Industrial Park.

CSHL's Courses program covers a diverse array of topics in molecular biology, neurobiology, structural studies, and bioinformatics. More than 750 instructors, lecturers, and assistants come

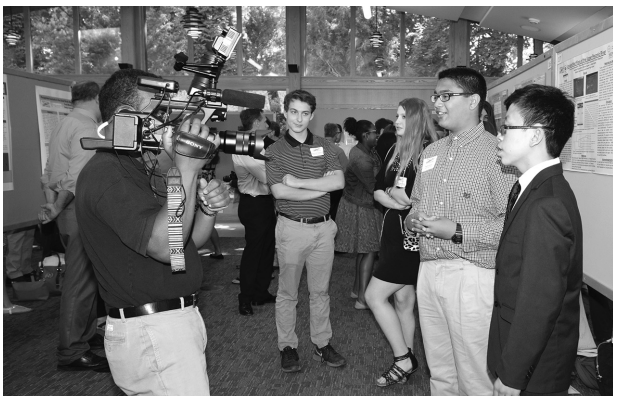
Suzhou meeting

to teach from universities, medical schools, research institutes, and companies around the world. In 2016, about 700 trainees, including advanced graduate students, postdocs, and faculty, attended courses lasting from 1 to 3 weeks.

Courses rely heavily on grants and foundation support, including major support from the Helmsley Charitable Trust, the Howard Hughes Medical Institute, National Institutes of Health, and the National Science Foundation (NSF). Specifically, the Helmsley Interdisciplinary Fellowship Fund provided major funding to 125 scientists to participate in courses outside their primary disciplines. Support from companies in the form of loaned equipment, reagents, and technical expertise is critical to offering participants training at the leading edge.

DNA Learning Center

The DNALC occupies the middle ground on a continuum of the science enterprise that spans from pure research to pure education. It adapts the latest methods and concepts from research to educational settings, empowering students and teachers to participate in real-time research experiences. In the last few years, most of the DNALC's projects funded by grants have focused on scaling research methods to reach larger numbers of students. This effort has expanded our reach from middle and high school students and teachers to undergraduate populations.



Barcode meeting

For the fifth year now, DNA barcoding projects have engaged students in biodiversity studies throughout the New York metropolitan area. Barcode Long Island involved 271 students from 31 schools in Nassau and Suffolk Counties, and the Urban Barcode project involved 214 students from 22 schools in the five boroughs of New York City. These students, mentored by DNALC-trained teachers with easy access to technology resources, investigated lichen biodiversity, ant biodiversity, mislabeling of herbal supplements, invertebrate bio-indicators of habitat health, microbes on smartphones, invasive plants along the Bronx River, and invasive beetles in a Long Island park.

CyVerse (Cyber Universe) is a $100 million NSF project that provides computer infrastructure to solve problems in modern biological research. As educational lead for the project, the DNALC brings students and teachers into the world of biological big data and high-performance computation. Leveraging the CyVerse infrastructure, the DNALC program in RNA sequence analysis allows undergraduate faculty to generate their own data sets of all of the genes active in an organism of their choice. Using bioinformatics tools at the "DNA Subway" created by the DNALC, faculty and students are working together to analyze nearly a trillion nucleotides of sequence data on the XSEDE national supercomputer system. The DNA Subway is the only graphical user interface for XSEDE, which is usually only accessed by experts via command-line programming.

DNA Subway logo

The DNALC is also participating in MaizeCODE, an NSF initiative to develop an encyclopedia of DNA elements that control gene action in corn. Building on CSHL's history of pioneering research on corn and our faculty expertise with the human ENCODE project, MaizeCODE will generate 150 new data sets of RNA and DNA sequence. The DNALC's task is to prepare undergraduate faculty and students to analyze these new data sets as they are released—at the same time and using the same tools as high-level researchers. Surprisingly, the vast majority of the corn genome sequence has not been carefully explored by researchers, so students will be trained to find

elements of gene structure and function that are missed by automated computer analysis. With this project, the DNALC will confirm that good research and education can be exactly the same thing.

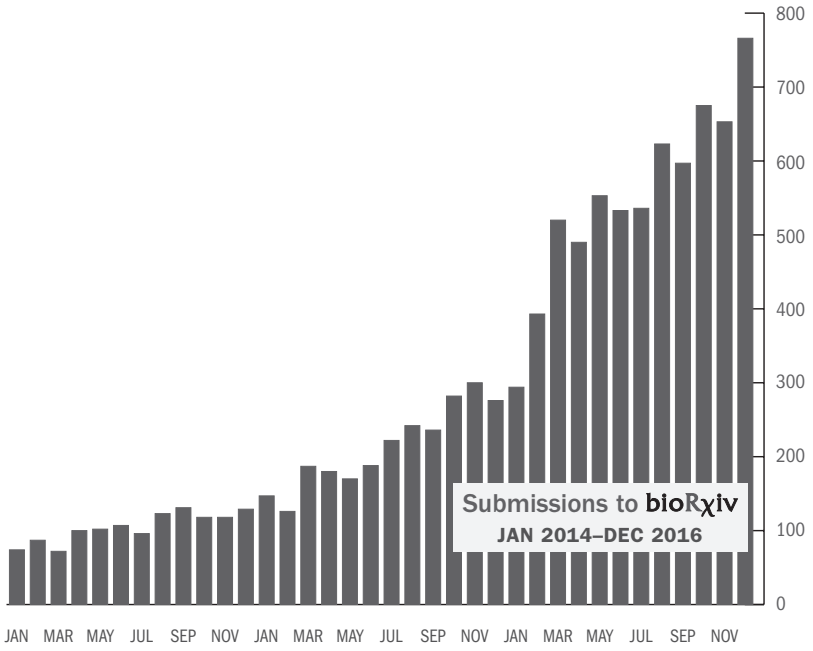
Underpinning these major efforts to bridge research and education are the hands-on laboratory experiences in genetics that the DNALC has historically provided to middle and high school teachers and students. In 2016, 20,884 students attended field trips at our teaching facilities in Cold Spring Harbor, Lake Success, and Harlem. An additional 9100 students received in-school lab instruction by DNALC staff, and 1349 students attended week-long summer camps at various locations across the metro New York area.

Five thousand visitors came to the Dolan DNALC facility for the Ötzi exhibit; 3600 students and members of the general public received guided tours about the 5000-year-old iceman mummy whose life-sized replica calls CSHL home. Monthly *Saturday DNA!* sessions for children and parents drew 171 students and family members. Half of these sessions involved art, with participants "drawing" with glowing microorganisms, creating biodiversity art, and making watercolor "portraits" of cells.

Cold Spring Harbor Laboratory Press

CSHL Press aims to provide scientists worldwide with authoritative, affordable, and appropriate information to further their research and career development. The Press carries forth the Laboratory's commitment to scientific communication.

A highlight this year was the remarkable growth of *bioRxiv*, the Laboratory's open preprint service. *bioRxiv* enables scientists to make draft manuscripts of papers immediately available to the research community and receive feedback before submission to peer-reviewed journals. Manuscript postings in 2016 were 2.5 times higher than in the previous year. With nearly 9000 papers from 40,000 authors in more than 80 countries, *bioRxiv* is now the world's largest source of life science preprints and is accessed more than a million times each month. Established in November 2013, *bioRxiv* has fundamentally changed the communication practices of scientists in biology and ignited similar movements in other sciences. This impact was made



Growth in submissions to *bioRxiv*

possible by the Laboratory's commitment to the founders' vision and by the generosity of Trustee Robert Lourie.

The long-established Press journals *Genome Research* and *Genes & Development* remain preeminent, with editorial teams adept at capturing the new ideas and technologies emerging in a broad range of disciplines. The *RNA* and *Learning & Memory* journals continue serving their specialized research communities in valuable ways. The newer review journals, *CSH Perspectives in Biology, CSH Perspectives in Medicine,* and *CSH Protocols*, continue to advance in stature and financial success. *CSH Molecular Case Studies* had a steady increase in submissions in its first full year and was accepted for indexing by the National Library of Medicine's PubMed service. The Press journals overall had a record yearly download of more than 17 million full-text articles.

The book-publishing program added 13 new titles to its list of 200, including a timely manual on the powerful CRISPR-based technologies. The bundling of print books with electronic editions for tablets, smartphones, and computers was warmly received by busy working scientists. The recent implementation of high-quality print-on-demand delivery for books now means that every title is always available to readers without the burden of costly inventory.

Watson School of Biological Sciences

This year, the Watson School welcomed its 18th incoming class and graduated its 13th. The achievements of the graduate program continued to grow. Students continued to graduate considerably faster than students in comparable Ph.D.-granting institutions and demonstrated an ability to secure excellent jobs. Twenty-two of our graduates have now secured tenure-track faculty positions and are receiving federal grants and publishing papers as independent researchers. Eight have been promoted to Associate Professor. Graduates have also moved into influential positions in administration, publishing, consulting, and industry.

At the 2016 graduation ceremony, seven Ph.D. degrees were awarded, bringing the total since the school's inception to 90. During the year, scientific papers published by students of the school appeared in major journals, bringing the cumulative total to more than 350. Current and former students won prestigious and highly competitive scholarships and fellowships.

2016 WSBS graduation ceremony

In August, the WSBS welcomed 11 new students. Members of the Class of 2016 were selected from 300 applicants.

This year, 1087 undergraduates from around the world submitted applications for 25 slots in the 10-week summer program to conduct advanced research in the laboratory of a CSHL faculty member. This Undergraduate Research Program, along with the equally innovative Partners for the Future Program that brings local high school students into our laboratories during their senior year, provides aspiring scientists with intellectual and social insights into life as a scientist.

125th Anniversary Capital Campaign

Led by Chairman Jamie C. Nicholls, the CSHL Board of Trustees set a goal: to match the institution's scientific preeminence with the financial resources to pursue pioneering research, regardless of the federal funding environment. The goal was achieved as the multi-year 125th Anniversary Capital Campaign closed in 2016, having raised $278 million—ahead of the $250 million goal. During the campaign period, an additional $180 million was donated to ongoing research and education programs—funds not counted toward the Campaign goals. Beginning with Charles Robertson's gift of the first CSHL endowment funds more than 50 years ago, philanthropy continues to be a key driver of discovery at the Laboratory.

The Capital Campaign, guided by the Board Development Committee and chaired by CSHL Trustee Marilyn Simons, focused on raising unrestricted endowment funds to support new research initiatives. A President's Fund of $136 million now provides flexibility to back basic discovery science and the development of new technologies that will have broad impact in many areas of science, agriculture, and health. The Campaign also supports CSHL's world-class Cancer Center, enabling renovation of the Demerec Laboratory and a new initiative linking development of cancer to nutrition and metabolism.

Campaign funds will support expansion of our world-renowned DNA Learning Center in New York City. An initial gift from trustee Laurie Landeau, V.M.D., generated donations of $25 million to bring hands-on lab experiences to public and private schools across the city. Thank you to all who supported the Campaign.

Demerec 3D rendering draft

Double Helix honorees Alan Alda and Roy Vagelos with CSHL President Bruce Stillman and Lesley Stahl

E. Witkin

Trustees

Douglas Schloss was elected to the Board of Trustees. Since 1994, Mr. Schloss has been CEO & Managing Member of Rexford Management. He previously managed arbitrage and investment activities at Marcus Schloss & Co. Mr. Schloss is a graduate of Princeton University and Harvard Business School.

Development

The 11th Double Helix Medals, hosted by Lesley Stahl, honored Alan Alda and Dr. P. Roy Vagelos and raised $4.3 million. The Women's Partnership for Science luncheon, honoring CSHL alumna and Lasker Prize winner Dr. Evelyn Witkin, raised $200,000. Led by the Cold Spring Harbor Laboratory Association, the Annual Fund contributed $7 million, with planned giving by the Helix Society playing a vital role.

Infrastructure

The major renovation of the Firehouse apartments marked the start of a long-term institutional project modernizing student, postdoc, and faculty housing units.

The Laboratory was confronted this year with the failure of the main seawall, originally constructed circa 1830. A temporary repair was effected this year, with its complete replacement planned for the near future.

CSHL seawall

A major chiller plant servicing four research buildings was replaced, increasing cooling capacity to these buildings as well as producing a net cost savings to the Laboratory of $1 million over the next decade. This was the first of a number of projects to modernize the Laboratory's physical plant infrastructure, providing greater reliability, lower operating costs, and the capacity to accommodate the future expansion of research facilities.

Public walking tour of main campus

Community Outreach

The Public Affairs Department works closely with faculty, students, and employees across the Lab to create opportunities for the public to engage with the institution. We offer the surrounding community opportunities to interact and experience CSHL in person through public lectures and talks, tours, concerts, and other special events on and off campus.

In 2016, 19 graduate students and postdocs who serve as expert guides for public walking tours helped acquaint more than 1300 visitors with the Laboratory's history and current pursuits. CSHL is also very well known by the local community's first graders who attend neighborhood public and private schools. More than 130 first graders participated in a hands-on science fair conceived, planned, and led by Watson School graduate students and DNA Learning Center instructors.

Twenty-four hours a day the Internet and social media channels allow CSHL to reach audiences outside of its immediate geography! Stories about the Laboratory's science, scientists, and educators come alive through multimedia products including video, interactive storytelling, and CSHL's own podcast, "Base Pairs." Please visit www.cshl.edu.

CSHL Public Lectures

March 24: David Micklos, Founder and Executive Director of CSHL's DNA Learning Center: *Asking the Wrong Questions about American Science Education*.

May 11: Chris Vakoc, M.D., Ph.D., Associate Professor, Cold Spring Harbor Laboratory: *Cocktails and Chromosomes*.

June 26: Jeremy Farrar, M.D., Infectious Disease Expert & Director of the Wellcome Trust: *Future of Global Health*. This was canceled because of Brexit.

July 6: Nicholas Tonks, Ph.D., F.R.S., Professor, Cold Spring Harbor Laboratory: *Drugging an Undruggable Target: A Scientific Journey from Discovery Research to a Clinical Trial.* This was cosponsored by CSHL, U.S. Trust–Bank of America, Northwell Health, and St. Johnland Nursing Center.

August 4: Zachary Lippman, Ph.D., Associate Professor, Cold Spring Harbor Laboratory: *Cocktails and Chromosomes*.

September 18: Raymond Dattwyler, M.D., Professor of Microbiology/Immunology and Medicine, School of Medicine of New York Medical College; **John Branda, M.D.**, Associate Director

Zador public lecture

of Clinical Microbiology Laboratories, Massachusetts General Hospital, Assistant Professor of Pathology, Harvard Medical School: *Update on Lyme Disease*.

October 16: Jon Cohen, Journalist; **Staffan Hildebrand**, Filmmaker; **Victoria Harden, Ph.D.**, Science historian: *HIV/AIDS Research: Its History and Future*. This was the public session of the 2016 CSHL/Genentech Center Conference on the History of Molecular Biology & Biotechnology.

October 17: Anthony Zador, M.D., Ph.D., Professor, Cold Spring Harbor Laboratory: *Can We Upload Our Mind to the Cloud?*

November 6: Philip R. Reilly, M.D., J.D., Author, geneticist, and former president of the American Society of Law, Medicine and Ethics: *EUGENICS: A Historical Perspective*. This was a 2016 Lorraine Grace lectureship on societal issues of biomedical research.

November 9: Molly Hammell, Ph.D., Assistant Professor, Cold Spring Harbor Laboratory: *Cocktails and Chromosomes*.

CSHL Public Concerts

March 18: Fei Fei Dong, piano
April 21: Stephen Waarts, violin
April 29: Claire Huangci, piano
May 20: Xun Wang, piano
August 26: Southampton Arts Festival Chamber Orchestra
September 9: Sang-Eun Lee, cello
September 16: Drew Petersen, piano
September 29: The Lysander Trio

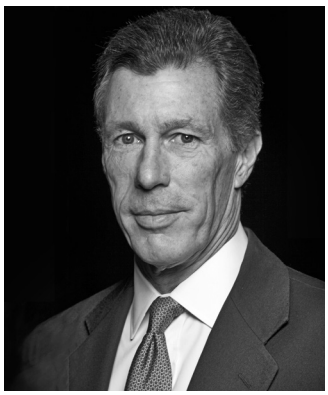
The Lysander Trio

Looking Forward

Thank you to the entire CSHL community for making 2016 such a successful year for our institution. I look forward to a future of CSHL breakthroughs in research and education that will undoubtedly change the world for the better.

Bruce Stillman, Ph.D., F.R.S
President and Chief Executive Officer

CHIEF OPERATING OFFICER'S REPORT



Cold Spring Harbor Laboratory enjoyed another positive year of operations. Revenues from federal and private grants, fund-raising, and internal operations were strong. Our endowment funds continued to grow and operating expenses were well controlled.

At year-end, we celebrated the success of our "125th Anniversary Capital Campaign," having exceeded our $250 million goal by nearly $30 million. Of the total funds raised, $186 million was earmarked for endowment. The campaign was effectively and energetically chaired by trustee Marilyn Simons, to whom we owe enormous gratitude, as we do to all of our trustees, who, as a group, contributed 65% of the total. At the same time, our 2016 Annual Fund added $7 million in unrestricted funds. This level of fund-raising is remarkable given the small size of our Development Office and the fact that the Laboratory's constituencies do not include undergraduate alumni or grateful patients.

The Laboratory's research investigators again achieved extraordinary success rates with federal grant awards leading to a 3% increase in federal funding year over year. The ongoing ability to secure a larger slice of a shrinking federal pie speaks to the excellence of the Laboratory's science. This, combined with substantial private support, allows the core basic research to move forward unencumbered. We are enthusiastic as well about our research/clinical collaboration with Northwell Health, which is supporting work that holds the promise of benefiting patients in the clinic.

Cold Spring Harbor's educational divisions are a critical component of the Laboratory's portfolio and major drivers of our international reputation and brand. They include the CSHL Meetings & Courses Program, the Banbury Center, the CSHL Press, the DNA Learning Center, and the Watson School of Biological Sciences. These activities are either self-sustaining or fully funded by endowment and/or philanthropic support. Most importantly, they reinforce the Laboratory's well-established reputation for scientific excellence around the world.

Particularly in an uncertain funding environment, endowment funds are the key to securing the Laboratory's future as a leading and enduring independent institution. It is for this reason that the recently completed fund-raising campaign placed such emphasis on endowment. Fortunately, our fund-raising success, in combination with steady investment returns, has resulted in healthy growth. Since the 2008 financial crisis, our total endowment fund has grown from $218 million to $470 million at year-end 2016. Interestingly, over the 8-year period, we realized $186 million in investment returns and received $188 million in endowment gifts for a total increase of $374 million. At the same time, we spent $122 million of the funds to support research and operations. Despite the growth, we recognize that capital market conditions make it increasingly difficult to expect investment returns that are substantially in excess of the current spending rate plus inflation. Consequently, the decision was made in 2016 to lower our annual spending rate on the endowment from 5% to 4.5%. Although this created an immediate $2 million reduction in available funds this year, the Laboratory was able to successfully absorb the shortfall. We are committed to holding to the 4.5% spending rate in 2017 as well.

All of this bodes well for the Laboratory's intellectual and financial strength going forward. However, we would be ill-advised to become complacent. The headwinds facing academic research in America have been well documented. Political polarization prevents the Congress from passing budgets—forcing the government to operate under "continuing resolutions." This has caused a 20% inflation-adjusted decline in the budget of the National Institutes of Health (NIH)—a primary source of funding for the country's outstanding research institutes, universities, and medical

schools. On top of this, the new administration has proposed a budget with an unprecedented 18% real-dollar cut to the NIH—a reduction that, if implemented, would be a game changer. As Nobel laureate Harold Varmus, a former Director of both the NIH and the National Cancer Institute, articulated in a recent *New York Times* Op-Ed piece, "A substantial NIH budget cut would undermine the fiscal stability of universities and medical schools, many of which depend on NIH funding; it would erode America's leadership in medical research; and it would diminish opportunities to discover new ways to prevent and treat diseases."

It would be a shame were the academic research community to face a challenge of this magnitude at a time of such great progress and promise.

W. Dillaway Ayres, Jr.
Chief Operating Officer

Long-Term Service

Back row (*left to right*): Patricia Wendel, Christopher Oravitz, David Spector, and Bruce Stillman; *center row* (*left to right*): Constance Hallaran, Michael Regulski, James Watson, and Frank Russo; *front row* (*left to right*): Susan Lauter, Christopher Hubert, Philip Renna, Maureen Morrow, and Lorraine McInerny.

The following employees celebrated milestone anniversaries in 2016:

40 years	Patricia Wendel
35 years	Terrance Chisum, Philip Renna
30 years	Christopher Hubert, Adrian Krainer, Susan Lauter, Vincent Meschan, Timothy Mulligan
25 years	Kathleen Cirone, Patricia McAdams, Christopher Oravitz, Michael Regulski, Frank Russo, Linda Van Aelst, Barbara Zane
20 years	William Carmona, Wendy Crowley, Constance Hallaran, Melissa Kramer, Oscar Lastra, Lorraine McInerny, Maureen Morrow, Marcie Siconolfi

Back row (*left to right*): Jerry Armstrong, David Spector, Dill Ayres and Bruce Stillman; *center row* (*left to right*): Andres Alarcon, Cesar Sisalima, Damian Desiderio, Erick Greene, Heather Cosel-Pieper, Louis Malfi, Joseph Carrieri, Bibiane Garite, and Stephanie Muller; *front row* (*left to right*): Hong Jie Shen, Amy Qiu Ji, Gail Sherman, Umamaheswari Ramu, Diane Fagiola, and Stephen Hearn.

15 years — Andres Alarcon, Jerry Armstrong, Edward Anderson, Joseph Carrieri, Maoyen Chi, Heather Cosel-Pieper, Damian Desiderio, Diane Errico, Diane Fagiola, Karen Filasky, Bibiane Garite, Erick Greene, Stephen Gregorovic, Stephen Hearn, Andriana Hincapie, Amy Qiu Ji, Louis Malfi, Alea Mills, Stephanie Muller, Umamaheswari Ramu, Hong Jie Shen, Gail Sherman, Cesar Sisalima, Doreen Ware

RESEARCH

CANCER: GENE REGULATION AND CELL PROLIFERATION

Camila dos Santos' laboratory studies the epigenetic regulation of normal and malignant mammary gland development, with an emphasis on the alterations brought by pregnancy. Significant changes mark the pre- and postpubescence mammary developmental stages, but those associated with pregnancy have the greatest effect on cellular function, tissue reorganization, and breast cancer susceptibility. Her group has recently found that mammary glands react differently to a second pregnancy than they do to the first one, with associated changes in DNA methylation. These findings suggested that pregnancy changes the state of mammary cells, and these changes may permanently alter how the cells react to the next pregnancy. In addition, the dos Santos lab is exploring how the pregnancy-induced epigenetic changes might influence cell transformation and the risk of breast cancer. This research uses genomic and computational approaches to define the pre- and postpregnancy mammary epigenome. An additional objective of the dos Santos' laboratory is to use functional genomics to discover novel transcriptional regulators that modulate mammary stem cell self-renewal, lineage specification, and cell transformation. The long-term objective of Camila's group is to improve the notion of the mammary epigenome during normal development and use this information to gain insight into new preventive and curative strategies to target breast cancer.

Human development requires the regulated activity of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome free of mutations. One type of mutation can arise from the activation of transposable elements (TEs). These virus-like parasites lay dormant within our genomes, but have the capacity to hop into new genomic locations, causing mutations as they break the surrounding DNA sequence. **Molly Hammell**'s lab explores the mounting evidence implicating transposon activity in a host of human diseases, with particular evidence for TE activation in neurodegenerative diseases: amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).

In **Leemor Joshua-Tor**'s lab, researchers study the molecular basis of nucleic acid regulatory processes using the tools of structural biology and biochemistry. One such regulatory process is RNA interference (RNAi), in which a small double-stranded RNA triggers gene silencing. Joshua-Tor and her team offered critical insight when they solved the crystal structure of the Argonaute protein and identified it as the long-sought Slicer. They then went on to explore the mechanism of the slicing event. The structure of human Argonaute 2 (hAgo2) bound to a microRNA (miRNA) guide allowed Joshua-Tor and her colleagues to understand how messenger RNA (mRNA) is cleaved during RNAi. This year, members of the Joshua-Tor lab explored the function of a very similar protein, Argonaute 1, that has no slicing ability, although it is almost identical in structure to the slicing hAgo2. Using biochemical methods and mutational analysis, they were able to identify key parts of the protein that are required for slicing activity. The lab also studies the generation of PIWI-interacting RNAs (piRNAs), which serve to protect the genome of germ cells. With colleagues in the Hannon lab, Joshua-Tor's team also determined the structure and function of Zucchini, a key nuclease in the initial generation of piRNAs in fruit flies. In other work, the lab is exploring the mechanisms of heterochromatin formation and gene silencing through the study of a protein complex called RNA-induced initiation of transcriptional gene silencing (RITS). Joshua-Tor is also well known for her work on the E1 helicase enzyme, which acts to unwind DNA strands during the DNA replication process.

Adrian Krainer's lab studies the mechanisms of RNA splicing, ways in which these mechanisms go awry in disease, and the means by which faulty splicing can be corrected. In particular, they study splicing in spinal muscular atrophy (SMA), a neuromuscular disease that is the leading genetic cause of death in infants. In SMA, a gene called *SMN2* is spliced incorrectly, making it only partially functional. The Krainer lab is able to correct this defect using a potentially powerful therapeutic approach. It is possible to stimulate protein production by altering mRNA splicing through the introduction of chemically modified pieces of RNA called antisense oligonucleotides (ASOs) into the spinal cords of mice. Previously, using ASOs in mice carrying a transgene of human SMN2, they developed a model for SMA using a technique they called TSUNAMI (shorthand for targeting splicing using negative ASOs to model illness). This year, they used the method to develop a mouse model for adult-onset SMA, and they are currently working to develop models for the study of other diseases caused by splicing defects, including familial dysautonomia. The Krainer lab has also worked to shed light on the role of splicing proteins in cancer. They have found that the splicing factor SRSF1 functions as an oncogene stimulating the proliferation of immortal cells. This year, they were surprised to find that SRSF1 can actually stop cell growth by stabilizing a powerful tumor suppressor protein called p53—suggesting that the cell is responding to the aberrant SRSF1 activity. This discovery offers insight into how tumors arise and the pathways that lead to transformation.

David L. Spector's laboratory studies the spatial organization and regulation of gene expression. Their recent studies showed an increase in random monoallelic gene expression on the differentiation of mouse embryonic stem cells (mESCs) to neural progenitor cells (NPCs). These data support a model in which stochastic gene regulation during differentiation results in monoallelic gene expression, and, for some genes, the cell is able to compensate transcriptionally to maintain the required transcriptional output of these genes. Therefore, random monoallelic gene expression exemplifies the stochastic and plastic nature of gene expression in single cells. In addition, the Spector lab is characterizing long nuclear retained noncoding RNAs (lncRNAs) that show altered levels of expression as mESCs transition from the pluripotent state to NPCs, and they are studying lncRNAs that are misregulated in cancer. Their efforts have focused on Malat1 lncRNA, which is one of the most abundant noncoding RNAs. The Spector lab previously identified a novel mechanism of 3′-end processing of this RNA. Current studies have revealed that altered levels of Malat1 lncRNA impact breast cancer initiation and progression. Studies are currently under way to elucidate the mechanism of action of this abundant nuclear-retained lncRNA.

Arne Stenlund and colleagues have obtained a detailed understanding of processes required for initiation of DNA replication from the papillomavirus, using this system to gain a general biochemical understanding applicable in other systems. Papillomaviruses are a large viral family that induces cell proliferation at the site of infection, usually giving rise to benign tumors. But certain types of human papillomaviruses (HPVs) generate tumors that progress toward malignancy. Among these are HPVs that cause most cervical cancers. Members of the Stenlund lab also pursue studies aimed at developing an effective small-molecule inhibitor of HPVs that might someday be used by women who do not receive the preventive anti-HPV vaccine now available or those already infected with HPV who would not be helped by the vaccine.

Bruce Stillman's lab studies the process by which DNA is copied within cells before they divide in two. Working with yeast and human cells, Stillman and colleagues have identified many of the cellular proteins that function at the DNA replication fork during the S phase, the portion of the cell-division cycle when DNA synthesis occurs. Among these proteins are those that facilitate the assembly of chromatin, the protein–DNA complexes that form the chromosomes. Current research focuses on the mechanism that initiates the entire process of DNA replication in eukaryotic cells. At the heart of this mechanism is a protein that binds to "start" sites on the chromosomes,

called the origin recognition complex (ORC). The Stillman lab is part of an ongoing collaboration that determined the cryo-electron microscopy (EM) structure of ORC proteins in complex with a group of proteins, called a helicase, that unwind DNA during replication. These images offer molecular insights into how the helicase is loaded onto DNA. The Stillman lab's research also focuses on the process by which duplicated chromosomes are segregated during mitosis. They found ORC at centrosomes and centromeres, structures that orchestrate chromosome separation during mitosis. At centromeres, ORC subunits monitor the attachment of duplicated chromosomes to the mitotic spindle that pulls the chromosomes apart when they are correctly aligned. Stillman's team has discovered that mutations in the Orc1 protein alter the ability of ORC to regulate both DNA replication and centrosome duplication. These mutations have been linked to Meier–Gorlin syndrome, a condition that results in people with extreme dwarfism and small brain size, but normal intelligence.

Cancer can be understood as a disease of dysfunctional gene expression control. Research in **Chris Vakoc**'s lab investigates how transcription factors and chromatin regulators cooperate to control gene expression and maintain the cancer cell state. This work makes extensive use of genetic screens to reveal cancer-specific functions for transcriptional regulators, as well as genomic and biochemical approaches to identify molecular mechanisms. One theme that has emerged from their efforts is that blood cancers are often vulnerable to targeting transcriptional coactivators, such as BRD4 and the SWI/SNF chromatin remodeling complex. Vakoc's team established that chemical inhibition of BRD4 shows therapeutic effects in mouse models of leukemia, a finding that has motivated ongoing clinical trials in human leukemia patients. The Vakoc lab has also developed a CRISPR-Cas9 screening approach that can reveal individual protein domains that sustain cancer cells. The lab is now deploying this technology in a diverse array of human cancers to reveal therapeutic opportunities and basic mechanisms of cancer gene control.

CHARACTERIZATION OF NEW MOUSE MODELS FOR THE INVESTIGATION OF PREGNANCY-INDUCED BREAST CANCER PROTECTION

C. dos Santos W.D. Frey R. Khatiwala

Epidemiological studies have revealed that women experiencing pregnancy at a young age have a significantly lower risk of developing breast cancer over their lifetime. Remarkably, the protective effects of pregnancy on breast cancer incidence have also been observed in rodents exposed to carcinogens. These observations suggest that pregnancy has an enduring effect on mammary gland cell biology that is well conserved. To develop a system to study the effects of pregnancy on mammary tumorigenesis, we developed a CAG-rtTA3-TRE-c-Myc mouse model, in which c-Myc overexpression is successfully achieved in doxycycline (DOX)-treated mammary epithelial cells (MECs). We used this model to test the effects of pregnancy on c-Myc-driven malignant transformation of MECs. Nulliparous and parous MECs were cultured with DOX for c-Myc induction. c-Myc overexpression reduced the proliferative activity of parous MECs, in marked contrast to the effects on nulliparous MECs, which were expanded for many passages within the culturing time. Altogether, these results lend support to our hypothesis that the output of oncogenic stimuli can be influenced by transitions through pregnancy in the cell of origin. To date, we have generated chromatin immunoprecipitation-sequencing (ChIP-seq) libraries for H3K27ac histone marks using transgenic nulliparous and parous MECs under DOX treatment for 40 days. Analysis of shared ChIP-seq peaks revealed enrichment for genes downstream from c-Myc signals, suggesting that nulliparous and parous MECs generally respond to the effects of c-Myc overexpression. Strikingly, H3K27ac peaks unique to parous MECs enriched for genes down-regulated in luminal breast cancer subtypes and in metastatic breast cancer, suggesting that a particular antitumor gene signature may be supported by the parous epigenome. We have also generated RNA-sequencing (RNA-seq) libraries to investigate the transcriptional output controlled by c-Myc in nulliparous and parous MECs. We found a significant enrichment for genes down-regulated in parous MECs that support tumor growth. We are currently developing the reagents to target such genes using CRISPR-Cas9 genomic editing to validate their anticancer properties. Ultimately, we will also merge ChIP-seq and RNA-seq data to link the most relevant modifications altering the predisposition of MECs to malignant transformation. The results of this study will reveal changes in the epigenetic landscape that confer a lasting resistance to breast cancer and will also supply a model to further test the impact of pregnancy-associated breast cancer protection.

Probing the Effects of Pharmacological Agents on the Prevention of Breast Cancer

C. dos Santos [in collaboration with P.F. Slepicka]

For the past decade, many groups have investigated the effects of pregnancy on reducing the risk of breast cancer, yet we still lack a clear understanding of the molecular mechanisms that prevent cancer development. The goal of this project is to identify pharmacological agents that can perturb the mammary gland epigenome in a manner that recapitulates the effect of pregnancy. Our previous findings indicate that Stat5, a transcription factor (TF) involved in tumorigenesis, is necessary for proper epigenetic modifications leading to lineage commitment in the mammary gland. We found that treatment with pimozide (PIM), a Food and Drug Administration (FDA)-approved drug and known modulator of Stat5 function, can mimic pregnancy-related signals. Interestingly, we observe that transitions through pregnancy are associated with exclusion of Stat5 from the nucleus of MECs and loss of genomic occupancy. This prompted us to consider the effects of PIM on pregnancy-dependent development. Our preliminary studies have revealed that PIM treatment enhanced several of the effects of pregnancy, including branching morphogenesis. Additionally, we have evaluated the effects of PIM treatment on the

development of mammary tumors. In these studies, we treated multiple mouse models of mammary tumorigenesis with PIM before the detection of tumors and monitored tumor development after treatment. Our preliminary studies provide evidence that PIM treatment delays mammary tumor onset and development. Altogether, these preliminary results lend support to our original hypothesis and provide new insights into the cancer-preventive effects of PIM treatment.

Characterization of the Molecular Basis of Pregnancy-Induced Breast Cancer Protection

R. Khatiwala [in collaboration with J. Rahman]

Our prior study suggests that pregnancy causes stable alterations of DNA methylation at a number of intergenic locations in MECs, which we hypothesize to represent enhancer DNA elements. In support of this view, many of these elements show enhancer activity when cloned into luciferase reporter constructs and transfected into parous MECs. These preliminary studies provide a rationale to investigate global enhancer activity in purified mouse MECs before and after pregnancy. To date, we have integrated ChIP-seq and RNA-seq data sets to search for regulatory regions and alterations in expression introduced by pregnancy. We are currently focused on several groups of regions/genes: (1) parous-specific regions enriched with active histone marks (H3K27ac and H3K4me1) indicative of increased expression in response to pregnancy signals, (2) parous-specific regions enriched with poised histone marks (H3K27ac and H3K27me3) that associate with genes that rapidly respond to pregnancy signals in parous MECs, and (3) parous-specific regions enriched with poised histone marks (H3K27ac and H3K27me3) that associate with genes unresponsive to the signals of a second pregnancy in parous MECs. This class of enhancer/promoter elements may be associated with genes that only activate in response to signals of a first pregnancy but become nonresponsive or delayed in response to the signals of consecutive pregnancies. We are currently focusing our analysis on the search for TFs that modulate the mammary epithelial epigenome, in addition to targeting candidate genes via CRISPR-based knockouts to determine whether the specific factors are functionally relevant to the preventive effects of pregnancy.

The Role of Bptf on Chromatin Accessibility at Enhancers and the Resultant Transplantation Capacity of Mammary Gland Stem Cells

W.D. Frey [in collaboration with P.F. Slepicka, A.M. Ouellette, and A. Chaudhry]

Lineage commitment and cell differentiation are processes driven by the reorganization of chromatin accessibility, which allows fate-specific TFs to control gene regulatory networks. We set out to investigate the role of epigenetic factors on chromatin remodeling and mammary stem cell (MaSC) maintenance. We found that Bptf, a bromodomain-containing factor and substituent of the nucleosome-remodeling factor (NURF) complex, was abundantly expressed in MaSCs and luminal progenitors, with a substantially lower expression in more differentiated MECs. Using a combination of tissue histology, flow cytometry, and epigenomics, we show that Bptf depletion in MaSCs impairs multiple stages of mammary gland development and leads to a substantial decrease in MaSC activity in transplantation-based experiments. Interestingly, Bptf depletion resulted in the accumulation of a novel MEC type, which we term Bsecs. We found that Bsecs have a unique transcriptional profile with similarities to luminal and myoepithelial cell types, and accumulation of this cell type may reflect a differentiation arrest driven by up-regulation of apoptotic pathways. Further investigation into chromatin accessibility suggests that depletion of Bptf leads to a considerable loss of open chromatin while increasing the accessibility of regions associated with apoptosis and cell cycle arrest—an observation that agrees with the global changes in gene expression of Bptf-depleted MECs. Collectively, our data support a role for Bptf in regulating the differentiation and survival of MECs through the control of chromatin remodeling at *cis*-regulatory elements and implicates Bptf as a novel epigenetic target in breast cancer.

Use of CRISPR Genomic Editing to Search for New Regulators of Mammary Stem Cell (MaSC) Self-Renewal

[In collaboration with A. Chaudhry and J. Ma]

The existence of MaSCs was first postulated from the observation that mammary glands can be regenerated by transplantation of epithelial fragments into

mice. The partial purification of mouse MaSCs using a combination of different cell surface markers has improved our understanding of their role in mouse mammary gland tumorigenesis. To address limitations in accurately purifying MaSCs, our lab developed a histone (H2b-GFP) label retention system in 2013 to improve MaSC isolation methods according to the expression of cell surface markers. Using this system, we found *CD1d*, a gene that encodes a cell surface marker, to be highly expressed by long-term label-retaining cells; thus, it represents a novel cell surface marker for MaSC purification. We found that CD1d-purified MaSCs displayed 247 short hypermethylated regions, termed CD1d-targeted regions. These regions are characterized by a stretch of hypermethylated CpGs that cover ~260 bp of DNA and are enriched with motifs recognized by a defined set of TFs. Interestingly, these regions are largely unmethylated in committed progenitors and differentiated cell types and, therefore, may represent regulatory regions involved in mammary epithelial cellular differentiation. In fact, transcriptome analysis showed that CD1d-targeted regions are often associated with genes that are preferentially down-regulated in CD1d MaSCs. Therefore, up-regulation of such genes may regulate mammary epithelial differentiation. To date, we have screened approximately 50 genes for their ability to increase self-renewal of the normal-like MEC-line Eph4. From this set of genes, we found five new regulators of MEC self-renewal. We are currently characterizing each of these genes in respect to their phenotypic and genomic alterations. Overall, this study will elucidate novel TFs involved in the epigenetic setting that are necessary for MaSC self-renewal.

INTEGRATING GENOMIC DATA SETS TO UNDERSTAND GENE REGULATION IN DEVELOPMENT AND DISEASE

M. Hammell Y. Hao, Y.-J. Ho, Y. Jin, W.-W. Liao, D. Molik, N. Rozhkov, R. Shaw

Human development requires the regulated expression of thousands of genes in hundreds of distinct cell types throughout life. One requirement for this process is that each cell must contain an intact, functional genome. Furthermore, there is a set of rules for how the genes in our genomes coordinate their activities, and these rules are encoded within gene regulatory networks. Many human diseases occur when these basic processes are altered—through either altering the genome itself (as in the mutations seen in cancers) or altering the way in which genes interact with each other. The focus of the Hammell lab is to understand how mutations in our genomes lead to both alterations in the function of the mutated gene itself and the repercussion of these alterations on the hundreds of other neighboring genes within the network. To this end, we use computational algorithms to integrate multiple types of genomic and transcriptomic sequencing data into models of cellular function. This includes an emphasis on developing novel tools for the statistical analysis of high-throughput data, developing novel algorithms for modeling the flow of signals through genetic pathways, and, importantly, testing these models using the tools of molecular genetics.

Endogenous Retroviral-Like Elements May Contribute to Neurodegeneration

Y. Jin, N. Rozhkov, R. Shaw

TDP-43 is an RNA-binding protein that is known to control proper processing of many RNA targets in neurons. Mutation of TDP-43 has been associated with a variety of neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), fronto-temporal lobar degeneration (FTLD), and Alzheimer's disease (AD). However, the normal function of TDP-43 in neuronal development and maintenance has not been fully characterized and few of its messenger RNA (mRNA) targets have been definitively associated with the neurodegenerative diseases that result from loss of TDP-43 function. In collaboration with the Dubnau lab at Stony Brook University, my group has explored the novel hypothesis that TDP-43 normally plays a large and hitherto uncharacterized role in regulating the expression of transposable elements (TEs). TEs are retroviral-like elements encoded within our genomes whose unregulated expression leads to genetic instability as well as cellular toxicity. Members of the Dubnau lab have shown that TE transcripts are elevated upon expression of mutant, aggregate-prone forms of human TDP-43 in the fly brain, and neurodegeneration results from expression of this hTDP-43 protein. Members of the Hammell lab have shown that TDP-43 binds widely to TE transcripts in mammals, and TDP-43 binding to TEs is lost in human patients diagnosed with FTLD, a disease characterized by TDP-43 proteinopathy. Although these studies support a role for TDP-43 in regulating TE expression, our future goals are centered on determining the role for misregulation of TEs in neurodegenerative disease. Ongoing research in the lab is focused on three main areas: (1) improving our ability to detect active transposons in ALS and FTLD patient samples, (2) examining the degree to which transposons and other retroviral-like elements are interfering with cellular function in ALS and FTLD patient tissues, and (3) understanding the basic biology of how TDP-43 interacts with the general transposon control machinery.

TEs have been historically difficult to study because of their highly repetitive nature. Nearly half of the human genome is composed of TE-derived sequences, with millions of copies of TEs scattered throughout the chromosomes. Although most of these copies are nonfunctional, thousands of TEs retain the ability to mobilize and create new copies of themselves elsewhere in the genome. The difficulty lies in differentiating these active TEs from the millions of other harmless

copies with nearly identical sequences. This presents both a technical challenge for experimentally isolating TE-derived sequences from the genomes of cells and a computational challenge for determining where each TE copy originates in the genome of a particular sample. Members of my lab have recently developed novel statistical inference methods to solve the computational challenge of analyzing TE expression in sequencing studies (Jin et al., *Bioinformatics 31:* 3593 [2015]; Jin and Hammell 2017). These statistical inference methods have been used to examine the basic mechanisms of controlling TE expression and activity in germline tissues in collaborative studies (Goh et al., *Genes Dev 29:* 1032 [2015]; Krug et al. 2017). Ongoing efforts will establish optimized analysis protocols for many different types of TE studies (genome resequencing studies, chromatin association studies, etc.). In addition, several members of the lab are developing optimized protocols for isolating and identifying novel TE insertion sites in the genomes of individual cells, as active TEs provide one source of genetic mutation that can occur somatically in adult cells. Together, these efforts will provide the tools with which to determine the extent of TE activity in patient samples.

SAKE: Analysis Software for Single-Cell RNA-Seq Data Sets

Y.-J. Ho, D. Molik

The Cancer Genome Atlas (TCGA) was a large-scale cancer profiling project that aimed to sequence the genetic mutations and expression profiles for hundreds of patients in dozens of different cancers. Rather than finding a few genetic mutations that explain most cancers, these TCGA studies uncovered thousands of mutations with surprisingly little overlap from patient to patient. This makes the task of designing targeted therapies to treat an individual patient's set of cancer-causing mutations a difficult one. Although the genetic mutations showed little pattern of recurrence among patients, the gene expression data did show clear expression patterns or cancer molecular subtypes. Moreover, expression subtypes are often predictive of patient survival rates, likelihood to metastasize, and response to targeted therapies. For the melanoma cancer samples that my lab generates, members of my lab have also been able to show that these molecular subtypes persist in two-dimensional (2D) cultured cell lines, in three-dimensional (3D) cultured organoid cells growing in Matrigel, in 3D tumors grown in a mouse xenograft model, and from tumor samples taken directly from patients. However, for all of these samples, the transcriptomes came from bulk tissue or cell populations, whose behavior and phenotypes are generally an average over a very large population of tumor cells. This ignores the heterogeneity that exists both among different cell types within a tissue and among different individual cells within a population. This problem is especially important in the context of human cancers, which are continually evolving to develop invasive properties as well as resistance to therapeutics. To address the problems of cellular heterogeneity, the field of biology needs better tools, both in terms of the technology for single-cell-sequencing assays and in terms of the statistical methods for analyzing data from single-cell assays.

Although our technological ability to generate single-cell data sets has improved considerably over the last few years, our ability to analyze these data sets with algorithms that can robustly detect expression patterns in the presence of high noise and sparse sampling is lagging behind. Members of my lab began by thoroughly testing dozens of statistical models for expression clustering to find a method that would be accurate, robust to noise, and computationally efficient for hundreds of libraries. Using thousands of iterative randomizations and a set of gold standard cell types, we have shown that an adapted version of nonnegative matrix factorization (NMF) shows the most robust performance and is computationally efficient enough to rapidly analyze data from hundreds to thousands of cells. Furthermore, this NMF-based clustering method provides quantitative confidence estimates on the number of clusters present, as well as the assignment of any given cell to a cluster, providing a method to assess the accuracy of our methods as compared with other analysis strategies. We have built a user-friendly web browser application to provide this adapted NMF analysis package to the scientific community, which we have entitled SAKE for single-cell analysis and klustering evaluation.

Mechanisms of Acquired Drug Resistance in Melanoma

Y.-J. Ho

The genetic basis of melanoma development is fairly well understood, with activating mutations in the

oncogene BRAF occurring in a majority of melanoma patient tumors—which also harbor hundreds of secondary mutations of unknown impact. Specific inhibitors that target activated BRAF, as well as the downstream MAPK/ERK signaling pathway, have been developed, which dramatically reduce the growth of melanoma cells in patients. However, the effects of these drugs typically extend patient life span for 6 months or less, as the tumors rapidly develop resistance to these targeted therapies. Although some tumors resistant to BRAF inhibitors acquire additional genetic lesions that elevate MAPK or AKT signaling, most therapy-resistant cell lines establish resistance without a clearly understood mechanism of resistance. Members of the Hammell lab are developing computational algorithms to understand what changes take place in melanoma cells that develop resistance to BRAF inhibitor therapy, and how these changes relate to the genotype and molecular subtype of different melanoma tumors.

One goal of the lab is to classify melanoma tumors by both their sets of cancer-specific mutations (their genomic subtype) and the gene transcripts they express (their transcriptional or molecular subtype). Large-scale-sequencing surveys of melanoma patients have suggested that hundreds of genomic subtypes exist, such that each patient has their own distinct mutational profile, with a few commonly mutated genes alongside thousands of patient-specific mutations. However, using novel algorithms to analyze the sets of gene transcripts expressed by these tumors, we have determined that cutaneous melanoma can be largely classified into a small number of transcriptional subtypes, and these transcriptional subtypes are highly predictive of overall patient survival rates. Moreover, tumors with each distinct subtype can be seen in human patient samples, mouse xenograft models of melanoma, and cell culture systems. This allows for the subtype-specific modeling of melanoma progression in animal and cell culture systems. Ongoing research is using cell culture models of each melanoma subtype to determine whether response to therapy differs for melanoma cells of different subtypes, and whether this can inform the likelihood of developing therapeutic resistance through different cellular pathways.

To this end, we have taken melanoma cells from two very different melanoma subtypes and sequenced hundreds of individual cells as they respond to MAPK inhibitors. This work, in collaboration with N. Anaparthy in Dr. Jim Hicks's lab at USC, has allowed us to explore how well these overall patterns we see in bulk cell extracts are recapitulated at the finer-grained single-cell level. We have used the SAKE single-cell analysis package described above to identify patterns in the expression profiles of these single cells and have identified several surprising findings. First, most cells do not alter their underlying molecular subtype when they develop resistance to targeted therapies. Second, cells within a given subtype have several options available for pathways to resistance (i.e., multiple separate clusters exist within each subset of therapy-resistant cells). Finally, a small minority of naïve cells that have never been exposed to cancer therapeutics are already expressing markers of therapy-resistant cells. This has led us to a model in which a small number of cells with some degree of innate resistance to targeted therapies provide a platform for adaptation on which the cells elaborate and evolve.

PUBLICATIONS

Chio II, Jafarnejad SM, Ponz-Sarvise M, Park Y, Rivera K, Palm W, Wilson J, Sangar V, Hao Y, Öhlund D, et al. 2016. NRF2 promotes tumor maintenance by modulating mRNA translation in pancreatic cancer. *Cell* **166:** 963–976.

In Press

Jin Y, Hammell M. 2017. Analysis of RNA-seq data using TE transcripts. *Methods Mol Biol* (in press).

Krug L, Chatterjee N, Borges-Monroy R, Hearn S, Liao WW, Morill K, Prazak L, Chang YH, Keegan R, Rozhkov N, et al. 2017. Retrotransposon activation contributes to neurodegeneration in a *Drosophila* TDP-43 Model of ALS. *PLoS Genet* **13:** e1006635.

STRUCTURAL BIOLOGY OF NUCLEIC ACID REGULATORY PROCESSES

L. Joshua-Tor E. Elkayam, A. Epstein, C. Faehnle, S. Goldsmith, J. Ipsaro, J. Jung, K. Meze, K. On, A. Tocilj

We study the molecular basis of nucleic acid regulatory processes—RNA interference (RNAi) and DNA replication, in particular. We use the tools of structural biology, biochemistry, and biophysics to study proteins and protein complexes associated with these processes to elucidate how they work. X-ray crystallography, cryo-electron microscopy (cryo-EM), and other structural techniques enable us to obtain the three-dimensional structures of these molecular machines. Biochemistry, biophysics, and molecular biology allow us to study properties that can be correlated to their function and biology.

Mechanisms of RNAi and Noncoding RNAs

RNAi has made an enormous impact on biology in a very short period of time. Not only are we still discovering new cellular pathways for the regulation of gene expression that use these pathways, but RNAi has become an extraordinarily useful and simple tool for gene silencing. Almost from its beginnings, people have used genetics, biochemistry, molecular biology, and bioinformatics to study the mechanism of RNAi and related pathways. To get a true mechanistic understanding of these pathways, however, we must understand how the components of the RNAi machinery work at a molecular level.

The let-7 Regulatory Network

C. Faehnle, K. Meze [in collaboration with J. Walleshauser, WSBS]

Although many studies focus on the identification of microRNA (miRNA) targets and the various downstream mechanisms of gene silencing, we are interested in the regulation of a particular miRNA, let-7. The pluripotency factor Lin28 inhibits the biogenesis of the let-7 family of mammalian miRNAs. Lin28 is highly expressed in embryonic stem cells and has a fundamental role in regulation of development, glucose metabolism, and tissue regeneration. Alternatively, Lin28 overexpression is correlated with the onset of numerous cancers, whereas let-7, a tumor suppressor, silences several human oncogenes. Lin28 binds to precursor let-7 (pre-let-7) hairpins, triggering the 3′ oligouridylation activity of TUT4/7. The oligoU tail added to pre-let-7 serves as a decay signal, as it is rapidly degraded by the exonuclease Dis3L2. Genetic disruption of *DIS3L2* is the primary cause of Perlman syndrome, a congenital disorder leading to fetal overgrowth and an increased susceptibility to Wilms' tumor development. Subsequent studies have shown that Wilms' tumors, a common pediatric kidney cancer, overexpress Lin28, underscoring the role of miRNA regulation in kidney tumorigenesis. In somatic cells, in the absence of Lin28, TUT4/7 promotes let-7 biogenesis by catalyzing single uridine addition to a subset of pre-let-7 miRNAs. We are studying the molecular basis and mechanism of Lin28-mediated recruitment of TUT4/7 to pre-let-7, and its effect on the uridylation activity of TUT4/7, switching it from a monouridylation activity to an oligouridylase, and the subsequent degradation of pre-let-7 by Dis3L2.

Binding of Modified Oligonucleotides to Human Argonaute-2

E. Elkayam [in collaboration with Alnylam Pharmaceuticals]

Efficient gene silencing by RNAi in vivo requires the recognition and binding of the 5′-phosphate of the guide strand of a small interfering RNA (siRNA) by the Argonaute protein. However, for exogenous siRNAs, this is limited by the rapid removal of the 5′-phosphate of

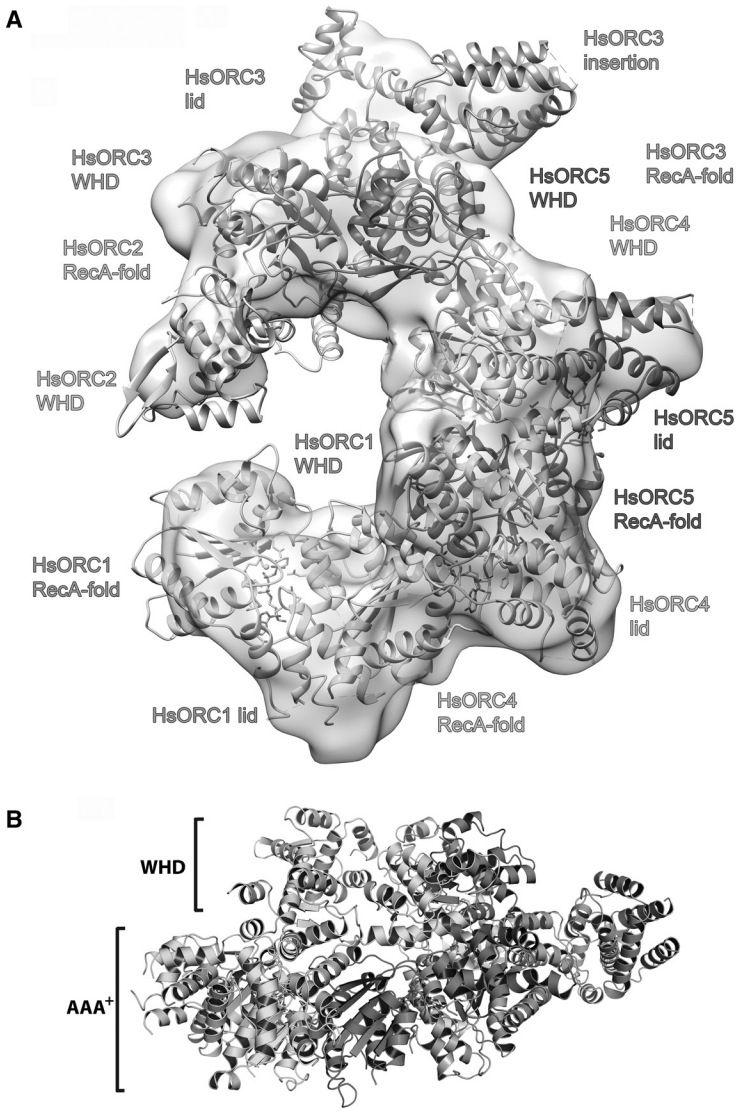


Figure 1. Structure of HsORC. (*A*) Ribbon diagram of HsORC modeled into cryo-EM density. The ORC motor module components are colored in green for ORC1, cyan for ORC4, and purple for ORC5. ORC2 is shown in wheat, and ORC3 is shown in salmon. ATP is shown in stick. The various domains for each subunit are labeled where the RecA-fold domain and the Lid domain together constitute the AAA+ domain. (*B*) Side view with the WHD layer on *top* and the AAA+ layer on the *bottom*. The ORC3 insertion can be seen extending from the particle on the *right*.

the guide strand by metabolic enzymes. We have been collaborating with Alnylam Pharmaceuticals, a leading biotech company with a focus on RNAi therapeutics, to understand the important RNA modifications they have been developing in the RNAi pathway, taking advantage of our ability to produce RNA-free human Argonaute-2. To this end, we have determined the crystal structure of human Argonaute-2 in complex with the metabolically stable 5′-(*E*)-vinylphosphonate (5′-*E*-VP) guide RNA at 2.5-Å resolution. The structure shows how the 5′ binding site in the Mid domain of human Argonaute-2 is able to adjust the key residues in the 5′-nucleotide binding pocket to compensate for the change introduced by the modified nucleotide. This observation also explains improved binding affinity of the 5′-*E*-VP-modified siRNA to human Argonaute-2 in vitro, as well as the enhanced silencing in the context of the trivalent *N*-acetylgalactosamine (GalNAc)-conjugated siRNA in mice relative to the unmodified siRNA.

DNA Replication

DNA replication is the most basic of life processes. Although the structure of Watson and Crick had immediate implications as to how the genetic material is copied, understanding the players involved, the intricate regulation, and even the simple mechanics of this process is a subject of intense investigation.

The Active Form of the Human Origin Recognition Complex

Binding of the origin recognition complex (ORC) to origins of replication marks the first step in the initiation of replication of the genome in all eukaryotic cells. We have been collaborating with Dr. Bruce Stillman here at Cold Spring Harbor Laboratory to understand the molecular mechanism of replication initiation. We determined the structure of the active form of human ORC (HsORC) by X-ray crystallography and cryo-electron microscopy (Fig. 1). The complex is composed of two lobes. The first, composed of the three nominally active ATP-hydrolyzing subunits, ORC1/4/5, which we named the motor module lobe, is organized in an architecture reminiscent of the DNA polymerase clamp loader complexes. A second lobe contains the ORC2/3 subunits. The complex is organized as a double-layered shallow corkscrew, with the AAA^+ and AAA^+-like domains forming one layer, and the winged-helix domains (WHDs) forming a top layer. CDC6 fits easily between ORC1 and ORC2, completing the ring and the DNA-binding channel, forming an additional ATP hydrolysis site. Analysis of the ATPase activity of the complex provides a basis for understanding ORC activity as well as molecular defects observed in Meier–Gorlin syndrome mutations.

In Press

Elkayam E, Parmar R, Brown CR, Willoughby JL, Theile CS, Manoharan M, Joshua-Tor L. 2017. siRNA carrying an (*E*)-vinylphosphonate moiety at the 5′ end of the guide strand augments gene silencing by enhanced binding to human Argonaute-2. *Nucl Acids Res* **45:** 3528–3536.

Ipsaro JJ, Shen C, Arai E, Xu Y, Kinney JB, Joshua-Tor L, Vakoc CR, Shi J. 2017. Rapid generation of drug resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS ONE* **12:** e0172177.

Tocilj A, On KF, Yuan Z, Sun J, Elkayam E, Li H, Stillman B, Joshua-Tor L. 2017. Structure of the active form of human Origin Recognition Complex and its ATPase motor module. *eLife* **6:** e20818.

RNA SPLICING

A.R. Krainer M. Akerman, O. Anczuków, J. Bassell, C.W. Hsu, H.-Y. Jeon, V. Kergourlay, Y.J. Kim, K.-T. Lin, Y.-H. Liu, W.K. Ma, L. Manche, T. Nomakuchi, J. Novatt, E. Panagot, M. Rahman, J. Scharner, L. Sheng, M.S. Wong, Q. Zhang

Mechanisms of Constitutive and Alternative Pre-mRNA Splicing

RNA splicing is required for expression of most eukaryotic protein-coding genes. The spliceosome selects authentic splice sites with very high fidelity, relying on limited sequence information present throughout introns and exons. In humans, >90% of genes are expressed via alternative splicing, giving rise to multiple protein isoforms. The choice of alternative splice sites is commonly regulated to alter gene expression, either tissue-specifically or in response to a developmental program or signaling pathways. The fact that multiple protein isoforms can be expressed from individual genes shows that the classical "one gene–one enzyme" paradigm is no longer valid, and provides an explanation for the unexpectedly small number of protein-coding genes uncovered by genome-sequencing projects.

Both constitutive and alternative splicing mechanisms involve numerous protein components, as well as five noncoding RNA components that are part of small nuclear ribonucleoprotein (snRNP) particles. The work in our lab focuses on the identification and molecular characterization of protein factors and sequence elements that are necessary for the catalysis and fidelity of splicing and/or for the regulation of alternative splice-site selection. We are interested in how the spliceosome correctly identifies the exons on pre-mRNA, and how certain point mutations in either exon or intron sequences cause aberrant splicing, leading to various human genetic diseases. Related areas of interest include the remodeling of messenger ribonucleoprotein (mRNP) architecture as a consequence of splicing, which influences downstream events such as nonsense-mediated mRNA decay (NMD); the various roles of alternative splicing misregulation in cancer; and the development of effective methods to correct defective splicing or modulate alternative splicing, especially in a disease context. A summary of some of our recently published studies is provided below.

Spliceosomal-Component Interactions

We collaborated with a former postdoc (Dr. Eric Allemand, Institut Pasteur, Paris) to complete a study of U2 snRNP interactions that he initiated here. In vivo–assembled U2 snRNP copurified with a subset of chromatin proteins, including histones and remodeling complexes such as SWI/SNF. An unbiased RNA interference (RNAi) screen revealed that alternative splicing is influenced by many chromatin factors, not all of which are physically associated with the spliceosome. These chromatin factors had context-specific effects on splicing, resulting in inclusion of certain alternative cassette exons and skipping of others. A direct assessment of the impact of chromatin on splicing using an in vitro cotranscriptional splicing assay with pre-mRNAs transcribed from a nucleosomal template showed that chromatin can influence the splicing competence of nascent pre-mRNPs.

Targeted Antisense Modulation of Alternative Splicing

Spinal muscular atrophy (SMA) is a common, autosomal-recessive motor-neuron degeneration disorder caused by homozygous deletion or mutation of the survival-of-motor-neuron gene, *SMN1*. A closely related *SMN1* paralog, *SMN2*, is present in all patients and differs from *SMN1* by a C to T transition in exon 7 that causes substantial skipping of this exon such that

SMN2 expresses only low levels of functional, full-length SMN protein. *SMN2* decreases the severity of SMA in a copy number–dependent manner. We previously developed an antisense oligonucleotide (ASO) that efficiently increases the extent of exon 7 inclusion during splicing of *SMN2* transcripts for therapeutic use in SMA. This translational research was performed in collaboration with Ionis Pharmaceuticals. After extensive preclinical testing, multicenter clinical trials (sponsored by Ionis and Biogen) with the ASO compound, nusinersen, were performed. The drug, which is administered to SMA infants and children by lumbar puncture, was well tolerated at all doses tested, and significant dose- and time-dependent improvements in survival, motor function, and achievement of developmental milestones were observed. Nusinersen (Spinraza™) was approved by the FDA on December 23 for broad use in all SMA types and all ages. This is the first and so far only approved treatment for SMA.

SMN protein is expressed ubiquitously and functions in RNA processing pathways that include trafficking of mRNA and assembly of snRNP complexes. Importantly, SMA severity is correlated with decreased snRNP assembly activity. In particular, the minor spliceosomal snRNPs are affected, and some U12-dependent introns have been reported to be aberrantly spliced in patient cells and animal models. It is likely that aberrant splicing of genes expressed in motor neurons is involved in SMA pathogenesis, but, as mentioned above, increasing evidence indicates that pathologies also exist in other tissues. We collaborated with Dr. Brage Andresen (University of Southern Denmark, Odense) to comprehensively analyze multiple tissues from an SMA mouse model by RNA sequencing. This study showed elevated U12-dependent-intron retention in all tissues examined, and such minor-intron retention was also observed on small interfering RNA (siRNA) knockdown of SMN in HeLa cells. Furthermore, retention of minor introns was reversed by ASO treatment to restore SMN levels in SMA mice, as were many downstream transcriptional changes. Finally, we proposed that missplicing of several minor intron-containing ion-channel genes may account for disrupted calcium homeostasis in SMA.

We also collaborated with Dr. John Staropoli (Biogen) to characterize gene-expression changes in an induced mouse model of SMA with or without ASO therapy. This study involved a technique we previously published, TSUNAMI, which uses an ASO to promote exon 7 skipping in *SMN2*-transgenic mice and thus phenocopy SMA in a dose-dependent manner, followed by rescue with a different ASO to restore exon 7 inclusion. As above, we found evidence of widespread intron retention, particularly of minor U12 introns, in the spinal cord of mice 30 days after SMA induction, which was then rescued by the therapeutic ASO. Intron retention was concomitant with a strong induction of the p53 pathway and DNA-damage response, manifesting as γ-H2A.X positivity in neurons of the spinal cord and brain. Widespread intron retention and markers of the DNA-damage response were also observed with SMN depletion in human SH-SY5Y neuroblastoma cells and human-induced pluripotent stem cell–derived motor neurons. We also found that retained introns, which tend to be high in GC content, served as substrates for the formation of transcriptional R-loops. We proposed that defects in intron removal in SMA promote DNA damage, in part through the formation of RNA:DNA hybrid structures, leading to motor neuron death.

The *SMN1* and *SMN2* genes are nearly identical, except for 10 single-nucleotide differences and a 5-nucleotide insertion in *SMN2* intron 6, exon 7, intron 7, and exon 8. SMA is subdivided into four main types with type I being the most severe. *SMN2* copy number is a key positive modifier of the disease, but it is not always inversely correlated with clinical severity. We previously reported the c.859G > C variant in *SMN2* exon 7 as a positive modifier in several SMA patients. In collaboration with Dr. Tom Prior (Ohio State University, Columbus) and former postdoc Dr. Yimin Hua (Soochow University, China), we recently identified A-44G in intron 6 as an additional positive disease modifier, present in a group of patients carrying three *SMN2* copies, but displaying milder clinical phenotypes than other patients with the same *SMN2* copy number. One of the three *SMN2* copies appears to have been converted from *SMN1*, and except for the C6T transition in exon 7, no other changes were detected. In the context of minigenes, *SMN1* C6T displayed an ~20% increase in exon 7 inclusion compared with *SMN2*. Through systematic mutagenesis, we found that the improvement in exon 7 splicing is mainly attributable to the A-44G transition in intron 6. Using RNA-affinity chromatography and mass spectrometry, we further uncovered binding of the RNA-binding protein HuR to the –44 region, where it acts as a splicing repressor. The A-44G change

markedly decreases the binding affinity of HuR, resulting in a moderate increase in exon 7 inclusion. This finding is consistent with our previous ASO screen in intron 6, which revealed the presence of a splicing silencer within this region.

Many pathogenic genetic variants have been shown to disrupt mRNA splicing. Besides mutations in the well-conserved splice sites, mutations in splicing regulatory elements (SREs) may deregulate splicing and cause disease. A promising therapeutic approach is to compensate for this deregulation by blocking other SREs with splice-switching ASOs. However, the location and sequence of most functional SREs are not well known. This study with Dr. Brage Andresen used single-nucleotide-resolution cross-linking and immunoprecipitation (iCLIP) to establish an in vivo–binding map for the key splicing regulatory factor hnRNP A1 and to deduce an hnRNP A1 consensus binding motif. We found that hnRNP A1 binding in proximal introns may be important for repressing exons. Inclusion of the alternative cassette exon 3 in *SKA2* was significantly increased by ASO-based treatment that targets an iCLIP-identified hnRNP A1 binding site immediately downstream of the 5′ splice site. Because pseudoexons are well suited as models for constitutive exons that have been inactivated by pathogenic mutations in SREs, we further used a pseudoexon in *MTRR* as a model and showed that an iCLIP-identified hnRNP A1 binding site downstream of the 5′ splice site can be blocked by ASOs to activate the exon. The hnRNP A1 binding map can be used to identify potential targets for ASO-based therapy. Moreover, together with the hnRNP A1 consensus binding motif, the binding map may be used to predict whether disease-associated mutations and SNPs affect hnRNP A1 binding and, consequently, mRNA splicing.

Targeted Antisense Modulation of NMD

NMD is another area in which we combined basic knowledge of RNA-processing mechanisms with antisense technology to develop a novel therapeutic approach. NMD is a cellular quality-control mechanism that is thought to exacerbate the phenotype of certain pathogenic nonsense mutations by preventing the expression of semifunctional proteins. NMD also limits the potential efficacy of therapies based on translational read-through compounds because it causes a reduction in the mRNA template available for translation. NMD depends in part on exon-junction complexes (EJCs), which are assembled upstream of exon–exon junctions on completion of splicing in the nucleus. The pioneer round of translation in the cytoplasm displaces bound EJCs, until the ribosome reaches the natural stop codon or a premature stop codon (PTC). In the latter case, the presence of one of more bound EJCs downstream from the PTC-bound ribosome constitutes a signal that initiates mRNA decay.

We developed a gene-specific method of NMD inhibition using ASOs to block EJC deposition downstream from a PTC. Using various target genes with disease-causing nonsense mutations (e.g., *HBB*, *MECP2*), we systematically screened sets of ASOs complementary to the region spanning the expected site of EJC deposition, looking for those that resulted in mRNA stabilization, allowing increased expression of truncated protein. We then combined this approach with an aminoglycoside read-through compound to effectively restore the expression of full-length protein. We verified the mechanism of action by showing that the effective ASOs blocked EJC binding in an in vitro splicing reaction, as well as by demonstrating the position-dependence of the ASO effect with respect to the downstream exon–exon junction. Finally, we showed that two ASOs targeting consecutive EJCs could be combined to stabilize an mRNA with a nonsense mutation in the antepenultimate exon. Our new method could potentially be used as a targeted antisense therapy in the context of nonsense mutations relatively near the 3′ end of genes—either alone when the truncated protein retains residual function or in conjunction with a translational read-through drug.

PUBLICATIONS

Allemand E, Myers MP, Garcia-Bernardo J, Harel-Bellan A, Krainer AR, Muchardt C. 2016. A broad set of chromatin factors influences splicing. *PLoS Genet* **12:** e1006318.

Anczuków O, Krainer AR. 2016. Splicing-factor alterations in cancers. *RNA* **22:** 1285–1301.

Arun G, Diermeier S, Akerman M, Chang KC, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, et al. 2016. Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. *Genes Dev* **30:** 34–51.

Bruun GH, Doktor TK, Borch-Jensen J, Masuda A, Krainer AR, Ohno K, Andresen BS. 2016. Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biol* **14:** 54.

Nomakuchi TT, Rigo F, Aznarez I, Krainer AR. 2016. Antisense oligonucleotide-directed inhibition of nonsense-mediated mRNA decay. *Nat Biotechnol* **34:** 164–166.

In Press

Doktor TK, Hua Y, Andersen HS, Brøner S, Liu YH, Wieckowska A, Dembic M, Bruun GH, Krainer AR, Andresen BS. 2017. RNA-sequencing of a mouse-model of spinal muscular atrophy reveals tissue-wide changes in splicing of U12-dependent introns. *Nucleic Acids Res* **45:** 395–416.

Jangi M, Fleet C, Cullen P, Gupta SV, Mekhoubad S, Chiao E, Allaire N, Bennett CF, Rigo F, Krainer AR, et al. 2017. SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. *Proc Natl Acad Sci* **114:** E2347–E2356.

Wu X, Wang SH, Sun J, Krainer AR, Hua Y, Prior TW. 2017. A-44G transition in SMN2 intron 6 protects patients with spinal muscular atrophy. *Hum Mol Genet* doi:10.1093/hmg/ddx166.

CELL BIOLOGY OF THE NUCLEUS

D.L. Spector G. Arun, B. Balasooriya, B. Benz, K.-C. Chang, S. Diermeier, N. El-amine, R. Hazra, S. Hearn, B. Liu, S. Russo, W. Xu, A. Yu

Most cellular processes can trace their beginnings to the nucleus where a gene is activated, resulting in the production of an RNA molecule, some of which code for proteins, whereas others are functional in the form of RNA. Although much biochemical information is available regarding many of the factors involved in gene expression, the spatial and temporal parameters that influence gene expression and the role of noncoding RNAs in regulating this multifaceted process are just beginning to be elucidated. Over the past year, our research has continued to focus on identifying and characterizing the role of long noncoding RNAs (lncRNAs) in cancer progression and/or differentiation. In addition, we have been examining the role of lineage commitment in establishing random autosomal monoallelic gene expression. Following is an overview of some of our accomplishments over the past year.

Identification of lncRNAs Involved in Breast Cancer Progression

G. Arun, B. Benz, K.-C. Chang, S. Diermeier, N. El-amine, S. Hearn, B. Liu, S. Russo, W. Xu [in collaboration with C.F. Bennett, R. MacLeod, and F. Rigo, Ionis Pharmaceuticals]

The mammalian genome encodes a huge repertoire of noncoding RNAs. *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1) is among the most abundant nuclear enriched lncRNAs. *MALAT1* is encoded at 11q13.1 in the human genome and its expression is up-regulated in many cancers, including breast, lung, prostate, and hepatocellular carcinoma (HCC). *MALAT1* is up-regulated in human mammary tumors, as well as different human mammary tumor cell lines, compared with normal tissue/cell samples. Up-regulation of *MALAT1* has been correlated with high metastatic progression and poor prognosis. In addition, *MALAT1* gene mutations have been shown to be recurrent in luminal type breast tumors. However, the mechanism by which *MALAT1* functions, and whether it is a driver in breast cancer progression, remains largely unclear. Our efforts are aimed at addressing the role of *MALAT1/Malat1* in metastatic breast cancer.

RNA-FISH studies on primary tumors and lung or brain metastases from patients with luminal type breast cancer revealed that *MALAT1* lncRNA is up-regulated three- to fourfold in metastatic nodules as compared with primary tumors. This is consistent with the possibility that *MALAT1* plays an important role in the metastatic progression of luminal breast cancers. We have used the *MMTV-PyMT* mouse model of luminal B breast cancer to characterize the role of *Malat1* in primary breast cancer and its subsequent metastasis. *Malat1* lncRNA was knocked down via subcutaneous administration of antisense oligonucleotides (ASOs) over a period of 7 wk, after which animals were killed, and primary tumors and lungs were removed for molecular and histological analyses. *Malat1* ASO treatment resulted in ~60% knockdown in the primary tumor concomitant with a significant differentiation of the primary tumor (Arun et al. 2016). Detailed histo-pathological analysis of ASO-treated tumors showed an increase in well-differentiated ductular tumors, whereas scrambled ASO-treated tumors progressed to solid undifferentiated carcinomas. Most interestingly, a marked decrease was observed in the incidence of lung metastases: ~70% fewer metastatic nodules in *Malat1* ASO-treated animals than scrambled ASO-treated animals. Further, *Malat1* ASO-treated ex vivo generated mammary tumor organoids from *MMTV-PyMT* mice resulted in an inhibition of branching morphogenesis, consistent with the in vivo reduction in metastasis. RNA-sequencing (RNA-seq) analysis of the primary tumors and tumor-derived organoids treated with *Malat1* ASO showed up-regulation of genes involved in differentiation and down-regulation of genes involved in migration and proliferation. Our data suggest that

Malat1 acting as a molecular scaffold is a repressor of tumor differentiation and its loss results in the differentiation of mammary tumors. Ongoing preclinical studies are aimed at moving these studies into a clinical trial for metastatic breast cancer.

In addition to our efforts in regard to *Malat1* lncRNA, we generated a comprehensive compendium of lncRNAs that show overexpression in mammary carcinomas compared with the normal mammary gland epithelium and, as such, represent additional therapeutic targets (Diermeier et al. 2016). We performed an RNA-seq screen on three physiologically relevant transgenic mouse models of luminal B (MMTV-PyMT) and HER2/neu-amplified (MMTV-Neu-NDL and MMTV-Cre;Flox-Neo-Neu-NT) subtypes of human breast cancer. We identified a total of 290 lncRNAs that are up-regulated at least twofold compared with normal mammary epithelial cells. From these up-regulated lncRNAs, we selected 30 previously uncharacterized transcripts for further evaluation. These transcripts were termed mammary tumor-associated RNA 1–30 (MaTAR1–30). Interestingly, 20 MaTARs show either exclusive expression or strong up-regulation in mammary tumors compared with any other adult organ. Based on our in vitro results, we chose the five most promising MaTARs and evaluated their therapeutic potential in vivo by subcutaneous injections of specific ASOs into female MMTV-PyMT mice. Tumor growth rates were monitored for up to 9 wk, and effects on tumor volume and morphology were investigated with respect to nonsilenced control mice injected with a scrambled ASO. Interestingly, independent knockdown of three MaTARs significantly reduced tumor growth (up to 40%). Primary tumors were isolated and analyzed regarding their morphological and histological properties using hematoxylin and eosin (H&E) staining of tissue sections. We observed moderate to severe tumor necrosis on treatment with ASOs independently targeting two MaTARs, and increased differentiation in tumors treated with ASOs targeting a third MaTAR. In the case of the MMTV-PyMT mouse model, lungs are of specific interest as they represent the main site of metastasis. Therefore, the lungs of ASO-treated mice were examined regarding the number and size of metastatic nodules compared with control animals. Notably, treatment with ASOs targeting one MaTAR resulted in an 80% decrease of metastatic nodules. Ongoing studies are pursuing the molecular function of these and other MaTARs and their potential as therapeutic targets in metastatic breast cancer and differentiation.

The Role of a lncRNA, *Platr4*, in Lineage Commitment

R. Hazra

We have previously published a high-depth RNA-seq analysis of poly(A)$^+$ RNA from mouse embryonic stem cells (ESCs) and neural progenitor cells (NPCs) to determine the expression of lncRNAs in a biologically robust manner in the Cast/Bl6 and AB2.2 backgrounds. This analysis focused on "intergenic" lncRNAs with the goal of identifying novel examples of how these RNA molecules may regulate ESC or NPC states. A total of 958 lncRNAs (biotypes "lncRNA" and "processed transcripts") were identified that were expressed in ESCs and NPCs (in both genetic backgrounds and excluding lncRNAs overlapping annotated small RNAs). After filtering RNA-seq reads, considering only lncRNAs with ≥5 FPKM, lncRNAs were manually curated to exclude any antisense lncRNAs and comparatively well-studied lncRNAs such as *Malat1*, *Neat1*, and *Firre*. From the 958 lncRNAs identified, weighted gene correlation network analysis was performed and identified 30 lncRNAs, which clustered tightly among the known pluripotency factors in the top quartiles of importance, suggesting their functional integration into the ESC gene expression program. We referred to these lncRNAs as *Platr1–30* (pluripotency-associated transcript; ranked by relative module importance). From these 30 presently uncharacterized *Platr* lncRNAs, we have focused efforts on *Platr4* and have identified a role of *Platr4* in lineage commitment. *Platr4* is a 986-nucleotide poly(A)$^+$ transcript comprising two exons, consistent with our RNA-seq and northern analysis in ESCs. Cellular fractionation of mouse ESCs indicated that *Platr4* is mainly present in the nuclear fraction and associated with chromatin. Using the CRISPR-Cas9 genome-editing system, we have generated multiple mouse ESC lines (AB2.2 and V6.5) with deletion of the *Platr4* promoter resulting in a significant depletion (homozygous deletion, *Platr4*-knockout) of the *Platr4* transcript (up to 99%) as measured by qRT-PCR and single-molecule RNA-FISH analysis. Deletion of *Platr4* showed abnormalities in the spontaneous contraction of embryoid bodies (EBs) derived from *Platr4*-knockout ESCs, compared with

control cells. Consistent with these data, the expression of cardiac troponin T (*cTnT*), an integral component of the contraction machinery, shows decreased levels in *Platr4*-knockout EBs. Further, morphological abnormalities of EBs were observed with smaller size and a darker cavity in knockout cells. Moreover, the relative expression levels of transcripts expressed in endoderm (*Sox17*, *Foxa2*) and mesoderm (*Tbx5*, *Gata4*) were markedly reduced in *Platr4*-depleted cells compared with control. In addition, deletion of *Platr4* in mouse ESCs resulted in significantly reduced expression of the cardiovascular gene network during targeted differentiation of cardiomyocytes (CMs). These data suggest that *Platr4* plays a critical role in lineage commitment and we are currently pursuing its molecular role in this process.

Probing the Role of a Highly Expressed lncRNA in HCC

A. Yu [in collaboration with Carmen Berasain, University of Navarra, Pamplona, Spain]

HCC is the most common type of liver malignancy and is one of the most lethal forms of cancer. HCC is the fifth-most frequently diagnosed cancer and the third-leading cause of cancer-related deaths worldwide. Interestingly, one lncRNA that came out of our ESC screen, lnc05, is also highly up-regulated in HCC cells compared with normal mouse hepatocytes, as well as in human HCC and cirrhotic liver tissue samples. Northern blot analysis revealed that lnc05 is 714 nt long. Knockout of lnc05 in HepA1-6 cells resulted in a 23% increased doubling time, or a 47% reduction in proliferation as assessed by expression levels of Ki-67 and colony formation assays. Flow cytometry analysis to evaluate the fraction of cells in each stage of the cell cycle found a 26% increase of cells in S phase and a 41% decrease in G_2. Based on these data, we are pursuing a potential role of lnc05 in HCC cellular proliferation.

Random Autosomal Monoallelic Gene Expression and Differentiation

B. Balasooriya

Monoallelic gene expression describes the transcription from only one of two homologous alleles of a particular gene in a diploid cell. We previously performed an unbiased RNA-seq screen to identify random monoallelically expressed genes taking advantage of a hybrid ESC line, which is an F1 cross between C57Bl/6 and CAST/Ei strains, so that the expressed single-nucleotide polymorphisms (SNPs) would reflect which allele the transcript is derived from. We identified a 5.6-fold increase from just 67 to 376 genes, showing random autosomal monoallelic expression during differentiation of mouse ESCs to NPCs, indicating that monoallelic expression is acquired on lineage commitment. These 376 genes represented ~3% of expressed genes in NPCs.

Over the past year, we initiated studies to determine the impact of various differentiation paradigms (ectoderm, mesoderm, endoderm) on the establishment of stochastic autosomal monoallelic gene expression at single-cell resolution. We have successfully established protocols to derive (1) CMs and cardiac smooth muscle cells (mesoderm), (2) neurons and glial cells (ectoderm), and (3) proximal lung epithelial cells (basal, secretory, and ciliated cells—endoderm). Successful differentiation into the three germ layers, and the presence of ESC-derived lineage-specific cells, were confirmed by qPCR for enrichment of lineage-specific markers.

We first focused on the mesoderm lineage. We used three different male F1 hybrid mESC lines established from three different blastocysts (biological triplicates). Before differentiation to CMs and cardiac smooth muscle cells, we performed karyotyping to confirm that all three mESC lines were diploid. We differentiated these ESCs toward cardiac progenitor cells (CPCs) and then allowed them to mature in culture to CMs (beating) and cardiac smooth muscle cells. mESCs, CPCs, CMs, and cardiac smooth muscle cells were independently harvested from different in vitro cultures at different time points, and prepared for single-cell RNA-seq using the Fluidigm C1 Single-Cell Auto Prep System. After single cells were primed into compartments, cell numbers (single or multiple cells) and the quality (debris or healthy) of the cells were evaluated by light microscope. Total RNA in each mammalian cell ranged from 10 to 30 pg. cDNAs from poly(A)$^+$ RNA were synthesized using the Clontech SMART-Seq v4 Ultra Low Input RNA Kit. We used 0.3 ng of dsDNA per cell, from high-quality cells for library preparation. Indexed libraries were generated using Illumina Nextera XT

DNA library preparation index kit, which is capable of multiplexing (up to 96 available indexes). Indexed libraries generated from each sample from one experiment were pooled together. We obtained eight pooled libraries, and library quality and quantities were assessed using an Agilent 2100 bioanalyzer using high-sensitivity DNA chips. In total, we obtained 215 libraries for ESCs, 175 libraries for CPCs, and 232 libraries for CMs and cardiac smooth muscle cells. We performed 125-nt paired-end sequencing for these eight pooled libraries (one pooled library per lane) using an Illumina HiSeq2500–V4 flow cell platform at the Cold Spring Harbor Laboratory Genome Center. This sequencing platform is capable of producing reads with 99% accuracy and, on average, 2.5 million reads per library. The high accuracy is essential in downstream analysis, as we will be relying on SNPs to assess allele-specific gene expression.

PUBLICATIONS

Arun G, Diermeier S, Akerman M, Chang K-C, Wilkinson JE, Hearn S, Kim Y, MacLeod AR, Krainer AR, Norton L, et al. 2016. Differentiation of mammary tumors and reduction in metastasis upon *Malat1* lncRNA loss. *Genes Dev* **30:** 34–51. [Commented on in Mendell JT. 2016. Targeting a long noncoding RNA in breast cancer. *N Engl J Med* **374:** 2287–2289.]

Diermeier SD, Chang K-C, Freier SM, Song J, Demerdash OE, Krasnitz A, Rigo F, Bennett CF, Spector DL. 2016. Mammary tumor associated RNAs impact tumor cell proliferation, invasion and migration. *Cell Rep* **17:** 261–274.

MOLECULAR BIOLOGY OF PAPILLOMAVIRUSES

A. Stenlund M. Kornaj

The papillomaviruses are a group of viruses that infect and transform the basal epithelium, inducing proliferation of the cells at the site of infection. The resulting tumors (warts) are in most cases benign and will usually regress after some time, but certain types of human papillomaviruses (HPVs) give rise to tumors that are prone to progress toward malignancy—especially and frequently, cervical carcinoma. Indeed, HPV infection appears to be a necessary cause of invasive cervical carcinoma and thus represents one of the few firmly established links between viral infections and the development of cancer.

An impediment to the study of papillomaviruses has been the inability to define simple in vitro cell culture systems for analysis of the viral life cycle. These viruses normally require specialized differentiating cells that only with difficulty can be generated in cell culture. However, for a bovine papillomavirus (BPV-1), a convenient cell culture system exists in which viral gene expression, oncogenic transformation, and viral DNA replication can be studied. Thus, BPV has become a useful model for these aspects of the viral life cycle. The DNA replication properties of the papillomaviruses show some unique and interesting characteristics. As part of their normal life cycle, these viruses can exist in a state of latency, which is characterized by maintenance of the viral DNA as a multicopy plasmid in infected cells. The copy number of the viral DNA is tightly controlled and the viral DNA is stably inherited under these conditions. Papillomaviruses, therefore, provide a unique opportunity to study plasmid replication in mammalian cells. In addition, the viral DNA replication machinery represents one of the most promising targets for antiviral therapy.

In previous years, we have reported the characterization of the papillomavirus replicon and the identification of the viral components that are required for viral DNA replication. In recent years, we have directed our attention toward the biochemical events that are associated with initiation of DNA replication. We are studying the biochemical properties of the viral E1 and E2 proteins and how these two proteins interact with the viral origin of DNA replication and with the cellular replication machinery to generate initiation complexes. Our studies show that the E1 protein has all the characteristics of an initiator protein, including origin of replication (ori) recognition, DNA-dependent ATPase activity, and DNA helicase activity. The transcription factor E2, whose precise function has remained more elusive, appears to serve largely as a loading factor for E1. Through direct physical interactions with both E1 and the ori, E2 provides sequence specificity for the formation of the initiation complex.

We are currently attempting to elucidate how the E1 and E2 proteins orchestrate the precise biochemical events that precede initiation of DNA replication at the viral ori. These events include binding of the initiator to the ori, the initial opening of the DNA duplex (melting), as well as the assembly and loading of the E1 replicative helicase at the replication fork. Our studies so far indicate that these activities are generated in an ordered process that involves the sequential assembly of E1 molecules on the ori. This sequential assembly generates different complexes with different properties that, in turn, recognize ori, destabilize the double helix, and function as the replicative DNA helicase.

The Amino-Terminal Domain in E1 Controls the Formation of an E1 Double Trimer, and the Transition to the Double Hexameric E1 Helicase through Multiple Mechanisms

The papillomavirus E1 protein is a representative of a subgroup of hexameric helicases. However, E1 is not a dedicated helicase but also serves several additional functions in the viral life cycle. E1 takes part in initiation of viral DNA replication in several different ways: The E1 DNA-binding domain (DBD) provides sequence-specific binding of E1 to the viral ori. A particular form of E1, a double trimer (DT), can then generate local melting of the viral ori. Once the DNA is melted locally, an E1 double hexameric (DH) is formed on the melted DNA and unwinds the DNA in front of the replication fork.

Obviously, the use of a single polypeptide to provide multiple functions presents many challenges. In addition to accommodating multiple different activities in one polypeptide, mechanisms have to exist that allow switching between the different activities. Furthermore, mechanisms have to exist that prevent the different activities from interfering with each other. For example, in the DT to DH transition, E1 switches from site-specific DNA binding and complex formation dictated by E1 binding sites in the ori to formation of hexamers dictated by the oligomerization domain and binding to single-stranded DNA (ssDNA). How this feat is accomplished is not understood.

We have previously shown that the E1 protein contains two different DNA-binding activities. One of these is the E1 DBD, which binds to specific sites in the origin of DNA replication and directs binding of E1 as a head-to-head dimer. The second DNA-binding activity consists of a β-hairpin structure in the helicase domain, which directs non-sequence-specific DNA binding. These two elements are required to form the DT and DH complexes. In addition to the DT and DH complexes that form on the origin of DNA replication, E1 can form two sequence-independent complexes on double-stranded DNA. These complexes, a trimer in the presence of ADP, and a hexamer that depends on ATP hydrolysis, provide information about the intrinsic ability of E1 to oligomerize and are thought to relate to the formation of the DT and DH, respectively.

We wanted to determine whether the formation of the trimer and hexamer also relies on the same two DNA-binding elements that are important for formation of the DT and DH. We therefore generated mutations in the DBD and β-hairpin in the helicase domain, residues involved in DNA contacts, and tested these for trimer and hexamer formation. Strikingly, although the β-hairpin was required for trimer formation, the DBD was not. However, the DBD was critically important for hexamer formation—showing that the formation of these complexes relies on different types of DNA-binding activities.

Because the β-hairpin by itself is not sufficient for trimer formation, we wanted to examine E1 for additional DNA-binding activities. Although E1 is well studied, functions have not been assigned to all parts of the protein, and especially the amino-terminal's approximately 150 residues remain mysterious. A part of this sequence is involved in nuclear import and export, but these sequences account for only a fraction on the amino-terminal domain. Given the parsimony that usually characterizes viral proteins, it is likely that additional functions are present in the amino-terminal domain.

To identify a DNA-binding activity present in the amino-terminal domain, we generated fragments from this domain and expressed these as GST fusion proteins in *Escherichia coli*. In this manner, we could identify a nonspecific DNA-binding activity, located between residues 70 and 120. Point mutations that disrupted the DNA-binding activity of the 70,120 peptide resulted in a failure to form the DT complex, consistent with a role for this DNA-binding activity in DT formation.

A peculiar aspect of this additional DNA-binding activity is that DNA binding by this peptide is not detectable in the context of the intact amino-terminal domain. However, successive deletions from the amino terminus gradually increased the DNA-binding activity of this fragment. These results show that the DNA-binding activity present in $E1_{70\text{-}120}$ is hidden or sequestered in the presence of the intact amino terminus. This sequestration can also be observed for other activities present in the amino-terminal domain. The β subunit of the kinase CK2 binds to the same peptide that mediates DNA binding and directs phosphorylation of an adjacent site. In the intact amino-terminal domain, phosphorylation of this site cannot be observed, but deletion of the amino terminus results in phosphorylation. We believe that the sequestration is reflective of a mechanism in which the DNA-binding activity is either exposed or sequestered to allow different kinds of E1 complexes to form with DNA. Consistent with this idea, the amino-terminal DNA-binding activity, which clearly is required for DT formation, when prematurely exposed inhibits DH formation. This new understanding of E1 complex formation on DNA now makes it possible to recapitulate in vitro the changes in E1 DNA binding that accompany the transition from DT to DH.

DNA REPLICATION AND CHROMATIN INHERITANCE

B. Stillman A. Amendolara, K. Bhalla, H.-C. Chou, M. Chugh, M. Hossain, Y. Hu, S. Nandi, Y.-J. Sheu, P. Wendel

Our studies continue to focus on the process of how our genome is duplicated before cells enter into mitosis and segregate the identical sets of chromosomes into the daughter cells. This process starts with the assembly of a prereplicative complex (pre-RC) at each origin of DNA replication as cells enter and progress through the G_1 phase of the cell division cycle. Pre-RC assembly first requires the origin recognition complex (ORC) and its partner protein CDC6 to bind to DNA at sites that are destined to become origins of DNA replication. In the yeast *Saccharomyces cerevisiae*, origins are specified in a semi-sequence-specific manner, whereas, as far as we know, origin specification in human cells is not DNA-sequence-directed. Nevertheless, similar processes are thought to occur to assemble pre-RCs all over the genome in G_1 phase, thereby marking these sites for subsequent initiation of DNA replication as the cell enters into S phase. The pre-RCs are activated by a complex series of events that are orchestrated by the two protein kinases, the S-phase cyclin-dependent kinase (S-CDK) and the Dbf4-dependent Cdc7 protein kinase (DDK), activities we have studied in previous years. In the past year, we have continued to study how DNA replication is regulated and have extended our finding that the ORC and CDC6 proteins that are themselves regulated by S-CDKs, in turn, control the timing of the synthesis of the S-CDKs and also their activity, providing a regulatory feedback mechanism for control of entry into cell division.

Once ORC and CDC6 bind to DNA and mark the location of future origins of DNA replication, the assembly of a pre-RC at these sites involves the recruitment of two hexamers of the Mcm2-7 proteins that are each bound to their chaperone, Cdt1. The process of assembly of pre-RCs was reconstituted in vitro with purified proteins a number of years ago, and careful biochemical studies in a number of laboratories have characterized pre-RC assembly intermediates, including ORC-DNA complexes, ORC-Cdc6-DNA complexes, ORC-Cdc6 attached to a Cdt1-bound Mcm2-7 hexamer on DNA, and the Mcm2-7 double hexamer that encircles the double-stranded DNA in the final pre-RC structure. Over the years, in collaboration with Huilin Li (now at the Van Andel Institute in Michigan) and former postdoctoral fellow Christian Speck (Imperial College London), we have characterized the assembly of pre-RCs in the yeast system using cryo-electron microscopy (cryo-EM). These studies continue and a high-resolution structure of a large replication intermediate was recently reported. In addition, with our Cold Spring Harbor colleague Leemor Joshua-Tor and her colleagues, notably Ante Tocilj, we have determined the structure of the human ORC protein in its active ATPase form.

ORC and CDC6 Control Cyclin-Dependent Protein Kinase

The Cyclin E- and A-dependent protein kinase activities are required in human cells to control both the initiation of DNA replication and centriole duplication in centrosomes. In previous studies, we had shown that ORC1, the largest subunit of the human ORC, contained a domain that functions as an inhibitor of both Cyclin E– and Cyclin A–dependent protein kinases. The two CDK inhibitory activities were genetically separable because mutations that block the ability of ORC1 to inhibit Cyclin A-CDK activity still allow ORC1 to inhibit Cyclin E-CDK activity. Moreover, mutations in ORC1 in patients with the microcephalic dwarfism Meier–Gorlin syndrome eliminate the ORC1 Cyclin E-CDK inhibitory activity but leave the Cyclin A-CSDK inhibitory activity intact. Thus, the separate ORC1 CDK inhibitory activities modulate the ability of both Cyclin E- and A-CDKs to control the DNA replication and centriole duplication.

We have now found that in human cells ORC1 also controls transcription of the gene encoding Cyclin E (*CCNE1*), a gene that requires the E2F1-DP1 transcription factors. ORC1 binds both the retinoblastoma (RB) tumor suppressor protein and the histone methyltransferase SUV39H1, which trimethylates the repressive histone mark on histone H3 at lysine 9 (H3K9me3). ORC1, RB, and SUV39H1 all bound to the *CCNE1* promoter near the E2F1 binding sites in early G_1 phase and repressed transcription of the gene. The repression prevents expression of Cyclin E, a protein that is normally involved in the commitment of a cell to cell division. Thus, ORC1 is involved in transcriptional gene silencing in human cells. Remarkably, when we discovered ORC in *S. cerevisiae* in 1992, we soon showed that ORC in yeast is required for transcriptional gene silencing of the mating-type genes. Although the mechanism is different, the conserved nature of the role of ORC1 in transcriptional gene silencing may be more widespread than previously known.

We also showed that the CDC6 protein, an ATPase that is highly related to ORC1 and forms part of the ORC-CDC6 DNA recognition complex for the initiation of DNA replication, binds to Cyclin E and reverses the ORC1 repression of *CCNE1* transcription. Both the genes encoding Cyclin E and CDC6 are regulated by the E2F1-DP1 transcription factors and activated as a cell commits to cell division. Cyclin E and CDC6 directly bind to each other and to the ORC1 protein at the *CCNE1* transcription promoter, and they cooperate to relieve the ORC-RB-SUV39H1 repression of *CCNE1*. Thus, a small amount of expression of CDC6 and Cyclin E that is induced by the activation of the Cyclin D-CDK4 during mid G_1 phase promotes a CDC6-Cyclin-E CDK feedback loop in which CDC6 and Cyclin E cooperate to relieve the repression imposed on the Cyclin E gene, thereby amplifying the amount of Cyclin E and committing cells to establish pre-RCs and enter into S phase (Fig. 1).

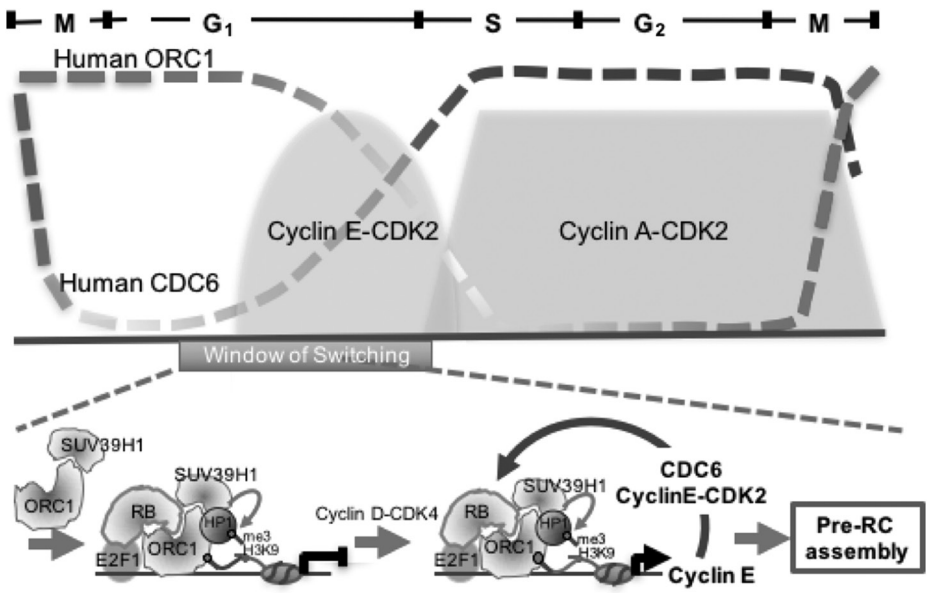


Figure 1. Cell-cycle control of transcription of the gene encoding Cyclin E. (*Top*) Unlike in yeast cells, ORC1 in human cells is expressed in a cell division cycle-dependent manner, being degraded as cells enter into S phase. On the other hand, CDC6 is degraded at the M phase to G_1 transition, and thus ORC1 and CDC6 are only present in the cell at the same time during mid G_1 phase. (*Bottom*) ORC1 binds SUV39H1 and RB and represses the *CCNE1* gene. After activation of Cyclin D-CDK4, which antagonizes RB, a small amount of CDC6 and Cyclin E are made from their E2F1 regulated gene, and CDC6 and Cyclin E bind to each other and feedback inhibits ORC1-mediated repression of CCNE1.

Structures of ORC-CDC6 and Other DNA Replication Intermediates

In previous studies with Huilin Li and Christian Speck and their colleagues, we had determined the low-resolution structure of the ORC-Cdc6-Cdt1-Mcm2-7 complex, called OCCM, which is a pre-RC assembly intermediate. With the vast improvement of cryo-EM microscopes, detectors, and data analysis software, it is now possible to gain much higher resolution structures. With this in mind, the structure of the *S. cerevisiae* OCCM bound to *ARS1* origin DNA was determined with an average of 3.9-Å resolution. The structure shows that ORC and Cdc6 form a shallow spiral-shaped ring that surrounds the DNA (Fig. 2, left panel). The carboxy-terminal winged-helix (WH) domains of ORC-Cdc6 form a collar that interacts with the WH domains of the Mcm2-7 subunits, forming a ring-shaped Mcm2-7 hexamer surrounding the DNA. Because ATP hydrolysis was prevented in forming the OCCM, the Mcm2-7 chaperone Cdt1 remained associated with the OCCM and formed a three-domain structure that embraced one-half of the Mcm2-7 hexamer. The Mcm2-7 hexamer ring was partially closed, with the amino-terminal domains of the Mcm2 and Mcm5 subunits locked together, whereas a gap was still present at the carboxy-terminal domains of the Mcm2 and Mcm5 subunits. The structure provides valuable insight into how the pre-RC is assembled, but also provides a view of how ORC and Cdc6 cooperate to recognize origin DNA in a semi-sequence-specific manner in budding yeasts, the focus of current studies.

In collaboration with Leemor Joshua-Tor and her colleagues, the structure of the human ORC (Orc1-5) complex was determined by separately determining the structure of ORC1-ORC4-ORC5 bound to ATP and ORC2-ORC3 and then fitting the two high-resolution structures into a cryo-EM map of human ORC 1-5 subunits determined in collaboration with Huilin Li. Using molecular modeling, we were able to place the CDC6 protein in the structure as well as in the model in double-stranded DNA. Like the yeast ORC-CDC6, this structure also formed a shallow, right-handed spiral that is expected to encircle the DNA and a ring of ATP binding proteins (Fig. 2, right panel). The structure of both the yeast and human ORC were very similar. More interestingly, they both resembled the structure of the DNA polymerase clamp loader replication factor C (RFC) that we had discovered many years ago and whose structure was determined by John Kuriyan and Michael O'Donnell and their colleagues. The ATPase-driven clamp loaders load a ring-shaped clamp onto DNA, which then interacts with DNA polymerases to tether them to the template DNA during DNA synthesis. Thus, there has been a remarkable structural similarity between the ATPase-driven machines ORC-CDC6 and RFC that load ring-shaped proteins onto double-stranded DNA at different stages of DNA replication.

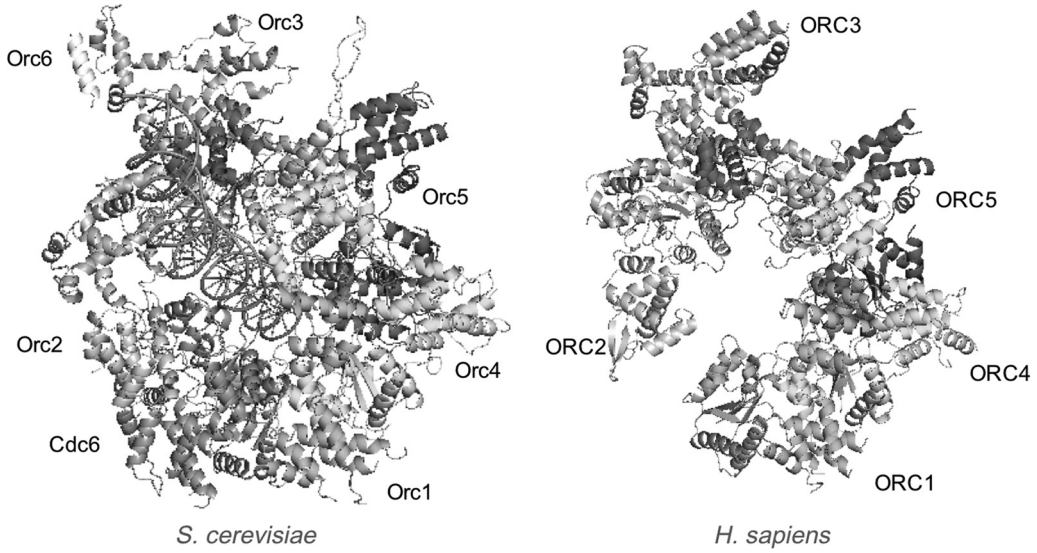


Figure 2. Comparison of the structure of *Saccharomyces cerevisiae* ORC-Cdc6 on ARS1 origin DNA with the human ORC1-5 complex. The structures are shown in the same relative orientation.

The structure of the human ORC complex in its active ATPase form also yielded new information about how the ATPase activity of ORC is controlled, and provided insight into the nature of two mutations in Meier–Gorlin syndrome that cause microcephalic dwarfism. Mutations in both ORC1 and ORC4 near the ATP binding site suggest roles for these amino acid residues in controlling the hydrolysis of ATP and also suggest that the ATPase activity of ORC is critical for normal organ development in humans.

PUBLICATIONS

Hossain M, Stillman B. 2016. Opposing roles for DNA replication initiator proteins ORC1 and CDC6 in control of Cyclin E gene transcription. *eLife* pii: e12785.

Hossain M, Stillman B. 2016. Meier–Gorlin syndrome. In *The initiation of DNA replication in eukaryotes* (ed. Kaplan DL), pp. 503–521, Springer International, Cham, Switzerland.

MacDiarmid JA, Langova V, Bailey D, Pattison ST, Pattison SL, Christensen N, Armstrong LR, Brahmbhatt VN, Smolarczyk K, Harrison MT, et al. 2016. Targeted doxorubicin delivery to brain tumors via minicells: Proof of principle using dogs with spontaneously occurring tumors as a model. *PLoS ONE* **1:** e0151832.

Sheu Y-J, Kinney J, Stillman B. 2016. Concerted activities of Mcm4, Sld3 and Dbf4 in control of origin activation and DNA replication fork progression. *Genome Res* **26:** 315–330.

Sun J, Yuan Z, Stillman B, Speck C, Li H. 2016. Structure and function studies of replication initiation factors. In *The initiation of DNA replication in eukaryotes* (ed. Kaplan DL), pp. 427–441. Springer International, Cham, Switzerland.

In Press

Tocilj A, Yuan Z, On KF, Sun J, Elkayam E, Li H, Stillman B, Joshua-Tor L. 2017. Structure of the active form of human origin recognition complex and its ATPase motor module. *eLife* 10.7554/eLife.20818.

Yuan Z, Riera A, Bai· L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B. 2017. Structural basis of MCM2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol* **24:** 316–324.

TRANSCRIPTION AND CHROMATIN DEREGULATION IN CANCER

C.R. Vakoc A. Bhagwat, A. Hohmann, Y. Huang, B. Lanning, B. Lu, J. Milazzo, S. Polyanskaya, J. Roe, C. Shen, J. Shi, T. Somerville, Y. Tarumoto, Y. Wei, X. Wu, Y. Xu, Z. Yang

Broad alterations of gene expression are necessary, and often sufficient, for cells to undergo malignant transformation. As a consequence, cancer cells are vulnerable to perturbations of individual transcription regulators, including DNA-binding transcription factors and chromatin regulatory machineries. For many years our laboratory has taken a genetic screening approach to identify essential transcriptional regulators in cancer cells in an effort to expose basic regulatory mechanisms and opportunities for therapeutic intervention. Our initial efforts used short hairpin RNA (shRNA) screening but, more recently, we have embraced the use of clustered regularly interspersed short palindromic repeat (CRISPR)–CRISPR-associated protein 9 (Cas9) genome editing as a tool for annotating essential regulators in cancer. The overarching goal of our research is to gain a mechanistic understanding of how individual regulators become essential in specific malignant contexts. Our efforts have historically focused on identifying essential chromatin regulators in the mixed-lineage leukemia (MLL)-fusion subtype of acute myeloid leukemia, but our work now includes a broad investigation of "epigenetic phenomena" in diverse cancers, as described below.

Therapeutic Targeting of Transcriptional Coactivators BRD4, SWI/SNF, and TFIID in Acute Myeloid Leukemia

A. Bhagwat, A. Hohmann, C. Shen, Y. Xu

Original studies from our laboratory in 2011 identified the coactivator protein BRD4 as a leukemia dependency and drug target in acute myeloid leukemia. In 2013, BRD4 inhibitors entered Phase 1 clinical trials in leukemia patients, with the first evidence of antileukemia activity in human beings reported in the past year. A major focus of our current work has been to define the disease-relevant molecular mechanism of BRD4 function in leukemia. We have pursued this question using a host of genetic and biochemical approaches and have made substantial progress in defining a "BRD4 pathway" comprising critical *cis*- and *trans*-acting components that fuel uncontrolled cell growth in leukemia.

In the past year, we have published a study that defines a key mechanism underlying the antileukemia activity of BRD4 inhibitors. We have found that BRD4 uses the multisubunit Mediator complex as an effector at a subset of its occupied sites, which includes enhancers at the Myc locus. Remarkably, a short treatment with BET inhibitors causes a large-scale displacement of several Mediator subunits from the Myc enhancers and from various other regulatory sites in the genome. In addition, we found that the genetic perturbation of several Mediator subunits leads to leukemia cell differentiation, which mimics the effects of targeting BRD4. Collectively, these studies suggest that BRD4 facilitates Mediator recruitment to promote transcriptional activation of certain target genes. This represents progress in defining downstream components of BRD4 function that are functionally suppressed by BET bromodomain inhibition.

Another essential coactivator in leukemia that was identified via shRNA screening is the SWI/SNF chromatin remodeling complex. In 2016, we published a study showing that acute myeloid leukemia (AML) cells require the BRD9 subunit of the SWI/SNF chromatin remodeling complex to sustain *MYC* transcription, rapid cell proliferation, and a block in differentiation. Based on these observations, we derived small-molecule inhibitors of the BRD9 bromodomain, which selectively suppressed the proliferation of mouse and human AML cell lines. To establish these effects as on-target, we engineered a bromodomain-swap allele of *BRD9*, which retains functionality despite a radically altered bromodomain pocket. Expression of this allele in AML cells conferred resistance to the antiproliferative effects of our compound series, thus establishing BRD9 as the relevant cellular

target. Furthermore, we used an analogous domain-swap strategy to generate an inhibitor-resistant allele of *EZH2*. Our study provided the first evidence for a role of BRD9 in cancer and reveals a simple genetic strategy for constructing resistance alleles to show on-target activity of chemical probes in cells.

A major goal of our current research on coactivators in AML is to reveal discrete perturbations of coactivators that may augment the antileukemia activity while sparing the toxicity to normal tissues. A genetic screening strategy performed in our lab has revealed that TAF12, a small subunit of the TFIID coactivator complex, will suppress leukemia cell growth while having no effect on nontransformed myeloid cells. On a mechanistic level, we have found that TAF12 provides a surface on the TFIID complex that interacts with the transcription factor Myb. Hence, targeting TAF12 will suppress Myb function to eradicate leukemia in mice. We are currently pursuing short peptides that can disrupt this interaction to target leukemia cells in vivo.

Enhancer Reprogramming in Pancreatic Ductal Adenocarcinoma

J. Roe [in collaboration with T. Somerville, CSHL]

Unlike earlier stages of tumor formation, the metastasis process is not generally associated with recurrent genetic alterations and, hence, is likely driven by nongenetic mechanisms. Here we have evaluated the hypothesis that global changes in enhancer activity drive the transition from primary tumor to distant metastasis. To this end, we derived a collection of epithelial "organoid" cultures from normal murine pancreatic ducts or from premalignant, primary tumor, or metastatic pancreatic ductal adenocarcinoma mouse models. These samples were cultured under identical conditions and subjected to genome-wide evaluations of histone H3K27 acetylation, histone H3K4 monomethylation, and overall chromatin accessibility. This analysis revealed, unexpectedly, that metastasis is associated with thousands of enhancer alterations (both gains and losses) when compared with earlier stages of tumor formation. A motif analysis of metastasis-specific enhancers implicated the FOXA1 and GATA5 transcription factors as promoting these regulatory alterations. FOXA1 and GATA5 expression is massively up-regulated in murine and human pancreas cancer metastasis, which we show experimentally is sufficient to drive enhancer activation and the acquisition of several metastasis-promoting capabilities, as shown using a variety of in vitro and in vivo assays. The pro-metastatic effects of FOXA1 and GATA5 occur independently of any epithelial-to-mesenchymal transition but instead are associated with a reengagement of endoderm development gene expression programs. Collectively, our findings suggest that enhancer reprogramming by developmental transcription factors is a key mechanism underlying the metastatic transition in pancreatic cancer.

In addition to the epigenomic evaluation of pancreatic cancer organoids described above, we have also used CRISPR–Cas9-based genetic screens of transcription factors and chromatin regulators in pancreas cancer cell lines. These efforts have identified novel candidate regulators of cell identity in this disease, and our ongoing efforts seek to understand the biological consequences of these factors.

Cross Talk between Signal Transduction Cascades and the Cancer Cell Epigenome

B. Lanning, S. Polyanskaya, Y. Tarumoto, Y. Wei, Z. Yang

A major new effort in the laboratory initiated in the past year has been to conduct CRISPR–Cas9-based genetic screens to probe the essentiality of signaling proteins in diverse malignant contexts with the goal of understanding how such factors regulate the epigenome of cancer cells. Our screening effort is being directed at kinases, phosphatases, ubiquitylation proteins, glycosyltransferases, and various cell surface molecules. The long-term goal of these efforts will be to identify essential signaling molecules in specific contexts of hematological, sarcoma, and carcinoma biology and thereafter expose the underlying molecular pathway. These efforts may expose therapeutic opportunities and basic mechanisms of gene expression control in cancer.

We have recently discovered that acute myeloid leukemias are uniquely addicted to the kinases LKB1 and SIK3. This finding was unexpected, as LKB1 is widely assumed to be a tumor suppressor in other cancer contexts. In AML cells, we have shown that LKB1 functions upstream of SIK3 to promote cell expansion, whereas these factors are dispensable for the growth of nonleukemia cancer cell lines. Ongoing efforts are seeking to determine how LKB1/SIK3 regulates the

metabolic and epigenetic profile of leukemia cells to support this disease process.

Essential Epigenetic Regulators in Small Cell Lung Cancer

Y. Huang, X. Wu

A new area of interest in the laboratory has been to study how epigenetic processes become dysregulated in the pathogenesis of small cell lung cancer. A recently completed CRISPR–Cas9 screen has identified the transcription factor POU2F3 as a master regulator of a novel subtype of small cell lung cancer, which represents between 15% and 30% of cases of this disease. In this patient subset, POU2F3 is expressed at extremely high levels and lung cancer cells are addicted to this high level of POU2F3. We have found that POU2F3 is causally linked to upregulation of c-MYC. Moreover, we have performed a focused RNA-seq-based screen of cancer drugs to reveal candidate small molecules that are capable of suppressing POU2F3 functional output. Ongoing efforts will determine whether personalized therapeutic regimens can be constructed that will target POU2F3$^+$ lung cancer. In addition, we are exploring how histone modifying enzymes support the pathogenesis of this disease and the essential function of POU2F3.

PUBLICATIONS

Bhagwat AS, Roe JS, Mok BY, Hohmann AF, Shi J, Vakoc CR. 2016. BET bromodomain inhibition releases the mediator complex from select *cis*-regulatory elements. *Cell Reports* **15:** 519–530.

Hohmann AF, Martin LJ, Minder JL, Roe JS, Shi J, Steurer S, Bader G, McConnell D, Pearson M, Gerstberger T, et al. 2016. Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat Chem Biol* **12:** 672–679.

Huang YH, Vakoc CR. 2016. A biomarker harvest from one thousand cancer cell lines. *Cell* **166:** 536–537.

Martin LJ, Koegl M, Bader G, Cockcroft XL, Fedorov O, Fiegen D, Gerstberger T, Hofmann MH, Hohmann AF, Kessler D, et al. 2016. Structure-based design of an in vivo active selective BRD9 inhibitor. *J Med Chem* **59:** 4462–4475.

Somerville TD, Vakoc CR. 2016. Modeling the epigenetic chain reaction downstream of DNMT3A(R882H). *Cancer Cell* **30:** 9–10.

Tasdemir N, Banito A, Roe JS, Alonso-Curbelo D, Camiolo M, Tschaharganeh DF, Huang CH, Aksoy O, Bolden JE, Chen CC, et al. 2016. BRD4 connects enhancer remodeling to senescence immune surveillance. *Cancer Discov* **6:** 612–629.

Zhang Q, Zeng L, Shen C, Ju Y, Konuma T, Zhao C, Vakoc CR, Zhou MM. 2016. Structural mechanism of transcriptional regulator NSD3 recognition by the ET domain of BRD4. *Structure* **24:** 1201–1208.

In Press

Oran AR, Adams CM, Zhang XY, Gennaro VJ, Pfeiffer HK, Mellert HS, Seidel HE, Mascioli K, Kaplan J, Gaballa MR, et al. 2016. Multi-focal control of mitochondrial gene expression by oncogenic MYC provides potential therapeutic targets in cancer. *Oncotarget* (in press).

Christopher Hammell's lab is interested in understanding gene regulatory processes that give rise to robust phenotypes associated with normal development in animals (specifically, how the timing of developmental processes is controlled), as well as the alterations in these pathways that give rise to diseases such as cancer (as in the alterations in mitogenic pathways in melanoma). Hammell and colleagues approach this elemental problem by using a variety of model organisms and patient-derived cancer cell lines. To directly identify the components that function in controlling normal developmental timing, they use the small nematode *Caenorhabditis elegans*, applying forward and reverse genetic approaches. In contrast to the extreme robustness of cell-fate lineage in *C. elegans*, in which specification of developmental programs is hardwired, mutations that alter conserved signaling pathways in melanoma create relatively plastic developmental landscapes that allow these lesions to become aggressive tumors. Notably, the gene regulatory architecture of melanoma cells allows them to acquire resistance to therapeutic agents. Hammell's team is interested in epigenetic mechanisms that contribute to resistance, specifically dramatic changes in gene expression patterns and intracellular signaling pathways. They are performing high-throughput screens to identify cellular factors that allow these rewiring events to occur, with the idea that these components would make ideal therapeutic targets to complement existing clinical strategies.

Je H. Lee's lab studies how cells interact with their microenvironment to regulate gene expression during development. Single-cell heterogeneity in gene expression can result from spatial differences in cell–cell and cell–extracellular matrix interactions. Such differences contribute to stochastic evolution of tumor cells as well as morphogenesis during normal development. However, the spatial control of gene expression in complex tissues, embryos, or tumors remains poorly understood, because most genome-wide studies sample bulk tissues or dissociated single cells.

Lee and colleagues have recently developed a method to sequence RNA molecules directly within single cells and tissues using subcellular resolution imaging, and they showed subtle differences in cell–cell/extracellular matrix (ECM) signaling and gene expression genome-wide in situ. By clustering transcripts into functionally or morphologically discrete regions, they have found many unique spatial markers and signaling pathways. The Lee lab focuses on the role of noncoding RNA in chromatin remodeling and tumor progression using mouse and organoid models of human cancer. They use in situ sequencing, cell lineage tracing, and single-cell profiling to understand how noncoding RNA affects tumor cell evolution in their native context. Their long-term goal is to develop better tumor classification tools and anticancer therapeutics using our understanding of the tumor microenvironment.

Alea Mills is studying genetic pathways important in cancer, aging, and autism, identifying the genetic players and determining how aberrations in their functions culminate in human disease. Through innovative use of a technique called "chromosome engineering," the Mills group discovered that one of the most common genetic alterations in autism—deletion of a 27-gene cluster on chromosome 16—causes autism-like features in mice. These autism-like movement impairments can be identified just days after birth, suggesting that these features could be used to diagnose autism. Mills has also used chromosome engineering to identify a tumor suppressor gene that had eluded investigators for three decades. The gene, called *Chd5*, was shown by Mills to regulate an extensive cancer-preventing network. This year, the Mills lab uncovered how *Chd5* acts as a tumor suppressor: It binds to a protein found within chromatin to turn specific genes on or off, halting cancer progression. The epigenetic role of *Chd5* in development, cancer, and stem-cell maintenance is currently being investigated. The Mills lab is also studying p63 proteins, which

regulate development, tumorigenesis, cellular senescence, and aging in vivo. They succeeded in halting the growth of malignant tumors by turning on production of one of the proteins encoded by the *p63* gene, called TAp63. TAp63 also exerts other protective effects. This year, the Mills lab generated a mouse model that allowed them to find that TAp63 is required to prevent a genetic disorder known as EEC (ectrodactyly-ectodermal dysplasia cleft lip/palate syndrome), which is characterized by a cleft palate and major deformities of the skin and limbs in infants. In addition, they recently discovered that a different version of *p63*, called ΔNp63, reprograms stem cells of the skin to cause carcinoma development—the most prevalent form of human cancer. Modulation of these proteins may offer new ways to treat human malignancies in the future.

Scott Powers' work focuses on gene alterations that cause cancer and factors that influence responses to specific anticancer drugs. His lab uses technologies that probe the entire genome to identify candidate cancer genes and evaluate their functional role in cell transformation and tumor biology. They also use whole-genome technologies to guide development of novel cancer diagnostics and therapeutics. Using DNA copy number analysis, the Powers group pinpoints novel amplified oncogenes and then applies functional studies to address the mechanisms by which they are implicated in oncogenesis. They have successfully applied this approach in breast, liver, colon, and lung cancers. Powers has also had an important role in the development of a distinctive CSHL approach to functional study of cancer genes. Called integrative oncogenomics, it is a rapid, large-scale screen for genes that are deleted or amplified in human cancers and suspected of being tumor suppressors, in the case of deletions, or oncogenes, in the case of amplifications.

Michael Wigler's work provides a new paradigm for understanding and exploring human disease. The Wigler lab studies human cancer and the contribution of new mutations to genetic disorders. The cancer effort (with James Hicks, Alex Krasnitz, and Lloyd Trotman) focuses on breast and prostate cancers. It involves collaborative clinical studies to discover mutational patterns predicting treatment response and outcome and the development of diagnostics to detect cancer cells in bodily fluids such as blood and urine. The major tools are single-cell DNA and RNA analysis. The single-cell methods, which are in development, are also being applied to problems in neurobiology (with Josh Huang and Pavel Osten) to characterize neuronal subtypes, somatic mutation, and monoallelic expression. The Wigler lab's genetic efforts are in collaboration with Ivan Iossifov and Dan Levy, and this team focuses on determining the role of new mutations in pediatric disorders. In a large-scale population sequencing project with W. Richard McCombie and the Genome Sequencing Center at Washington University in St. Louis, and supported by the Simons Foundation, the team has proven the contribution of this mechanism to autism. The work further suggests a relationship between the mutational targets in autism and the process of neuroplasticity that lies at the heart of learning. Smaller-scale population studies of congenital heart disease and pediatric cancer (collaborations with scientists at Columbia University and Memorial Sloan Kettering Cancer Center, respectively) also point to new mutation as a causal factor in these disorders.

THE REGULATION OF DEVELOPMENTAL TRANSITIONS DURING ANIMAL DEVELOPMENT

C.M. Hammell C. Aquirre-Chen K. Doerfel J. Wang

In normal development, the identity, behavior, and biological function of individual cells are coordinated with their neighbors to form distinct tissues. The defined structure and specification of tissues endows them with unique functional and physiological properties that exceed the capabilities of individual cells. Setting up these architectures is an incredibly complex process that requires aspects of spatial and temporal gene regulation to be coordinated to control facets of cell proliferation and cell fate determination. Although the last few decades have illuminated a variety of conserved regulatory mechanisms that mediate spatial gene regulation, our understanding of how temporal patterning is set up lags behind. From what we do know, the genes and mechanisms that control developmental timing are very well conserved, and disruptions in this regulation result in a variety of human diseases, including cancer. In particular, the expression and activity of particular microRNAs play key roles in regulating developmental transitions, and our lab is focused on determining the mechanisms by which the expression and activity of these regulatory RNAs are established.

Oscillatory Gene Expression and Developmental Patterning

K. Doerfel, C.M. Hammell, S. Persuad, A. Tao, and J. Wang

Previous work from our laboratory has determined that a large fraction of the *Caenorhabditis elegans* transcriptome is expressed in an oscillatory fashion (Fig. 1B,C). Animals show periodic patterns of gene expression throughout larval development, and these patterns are tied to the essential molting cycle. Importantly, misregulation of this patterned expression leads to cell fate specification defects. Previous genetic screens had also identified the *Caenorhabditis elegans* ortholog of the Period protein as being involved in controlling the dynamics of these expression patterns. Using phenotypes associated with *lin-42(lf)* mutants, we identified suppressors. Several of these suppressors encode transcription factors (BLMP-1, ELT-3, NHR-23, and NHR-25) that likely function upstream of LIN-42 in controlling oscillatory transcription. In the last year, we have sought to understand how these transcription factors control the various features of cyclical patterns of gene expression. In essence, there are three main components or features of oscillatory gene expression that could be encoded genetically. These are amplitude, phase, and duration of pulse.

Through CRISPR-Cas9-mediated genome editing, we GFP-tagged each of these transcription factors to determine their dynamic expression patterns. Surprisingly, only two of these components (NHR-23 and NHR-25) showed pulsatile expression, with the remaining two (BLMP-1 and ELT-3) showing constitutive expression. Using this expression data, ChIP-sequencing, as well as high-resolution transcriptome data, we sought to build a model for how the functions of these components generate cyclical patterns of gene expression. By determining the global transcriptional target sets of each of these transcription factors, we showed that these four components target similar gene sets, and the relationship between these factors is statistically significant when compared with other transcription factors that also drive expression in the skin of animals (Fig. 1A). Analysis of phenotypes associated with *blmp-1, elt-3, nhr-23*, and *nhr-25* mutants indicated that BLMP-1 and ELT-3 function to amplify most oscillatory gene expression, and mutations in *nhr-23* or *nhr-25* alter a distinct fraction of this class of mRNAs. Next, through the analysis of high-resolution mRNA-seq data, we determined that there are two main pulses of transcription during each larval stage (Fig. 1C). These two main pulses of transcription can be explained by the synergistic activities of three complexes: (1) NHR-23 controls the early phases of transcription, whereas (2) NHR-25 functions to control later transcriptional pulses and (3) BLMP-1/ELT-3 functions to amplify these patterns of expression (Fig. 1D). Furthermore, we determined

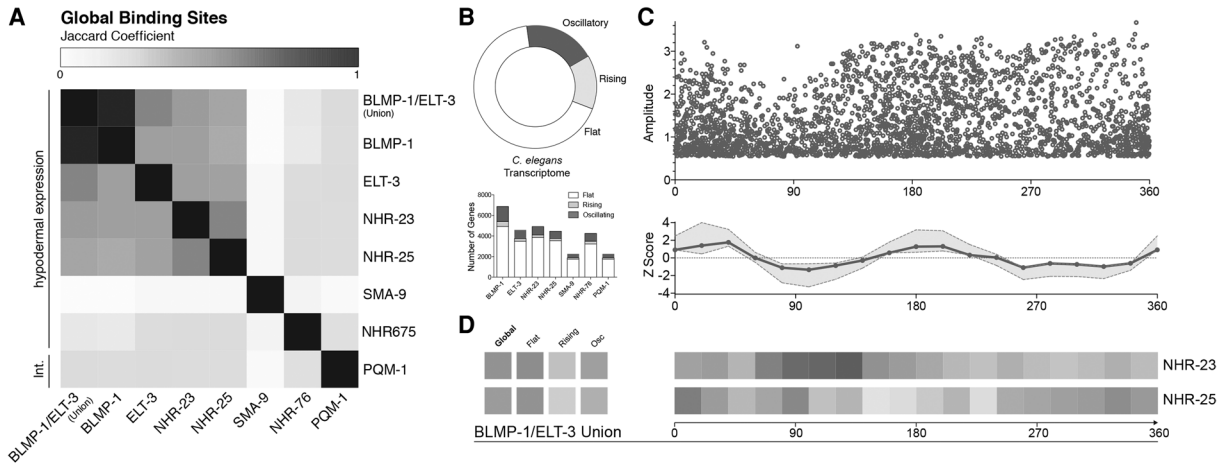


Figure 1. BLMP-1, ELT-3, NHR-23, and NHR-25 bind similar target genes and function together to generate phased transcriptional patterns. (*A*) Jaccard coefficient calculations of a variety of transcription factors that are expressed in the hypodermis or intestine. (*B*) The proportions of the *Caenorhabditis elegans* larval transcriptome parsed into expression dynamics categories and a distribution of various transcription factor targets for several transcription factors in these same categories. (*C*) mRNAs that show oscillatory expression patterns can be compartmentalized into two phases of expression. (*D*) BLMP-1/ELT-3 and NHR-23 cotarget gene sets that are expressed in the early phases of each larval stage. In contrast, BLMP-1/ELT-3 and NHR-25 cotarget gene sets that are expressed in the late phases of each larval stage.

that LIN-42 directly interacts with NHR-23 to potentially form the core of a functional developmental clock.

PQN-59 Encodes a Novel "Prion-Like" Protein that Regulates the Activity of the Core MicroRNA Machinery

C.M. Hammell and J. Wang

In addition to identifying genes that regulate the transcription of microRNA genes, we identified several candidates that potentially regulate the biochemical function of microRNAs while residing in the microRNA-induced silencing complex (miRISC). Among this candidate set of genes is a *C. elegans* "prion-like" gene, *pqn-59*. Pqn-59 encodes a highly conserved protein that harbors an ubiquitin-associated domain (UBA), as well as several stretches of primary AA sequence that have a propensity to form amyloid structures (Fig. 2A). Loss of this gene, via RNA interference (RNAi) or mutation, suppresses developmental phenotypes associated with several hypomorphic alleles of heterochronic microRNAs without dramatically altering microRNA expression.

To explore how PQN-59 may function in microRNA-mediated gene regulation, we generated a CRISPR-Cas9-mediated GFP allele of pqn-59 to study its expression and localization during development. We found that, in contrast to many of the other regulatory components that mediate changes in miRNA levels, PQN-59::GFP was expressed throughout development in most cells of developing embryos and larva. Furthermore, PQN-59::GFP is localized to the cytoplasm of these cells and excluded from the nucleus (Fig. 2C–E). Because PQN-59 harbors domains that may form amyloids, we investigated whether PQN-59::GFP formed insoluble network throughout the cytoplasm by using fluorescence recovery after photobleaching (FRAP) assays. By comparing the diffusibility of PQN-59::GFP to soluble GFP in these assays, we determined that PQN-59::GFP may form a cytoplasmic matrix (Fig. 2F,G). Consistent with the prion-like domains functioning in this process, removal of these domains increases the solubility of these fusion proteins and these domains are sufficient to form amyloid-like structures in yeast cells (data not shown). We next used the FRAP assay to determine that many of the core miRISC components, as well as several of the mRNA degradation machinery components that miRISC uses to regulate gene expression, also show low diffusibility in the cytoplasm of *C. elegans* cells (data not shown). This suggested to us that PQN-59 may form

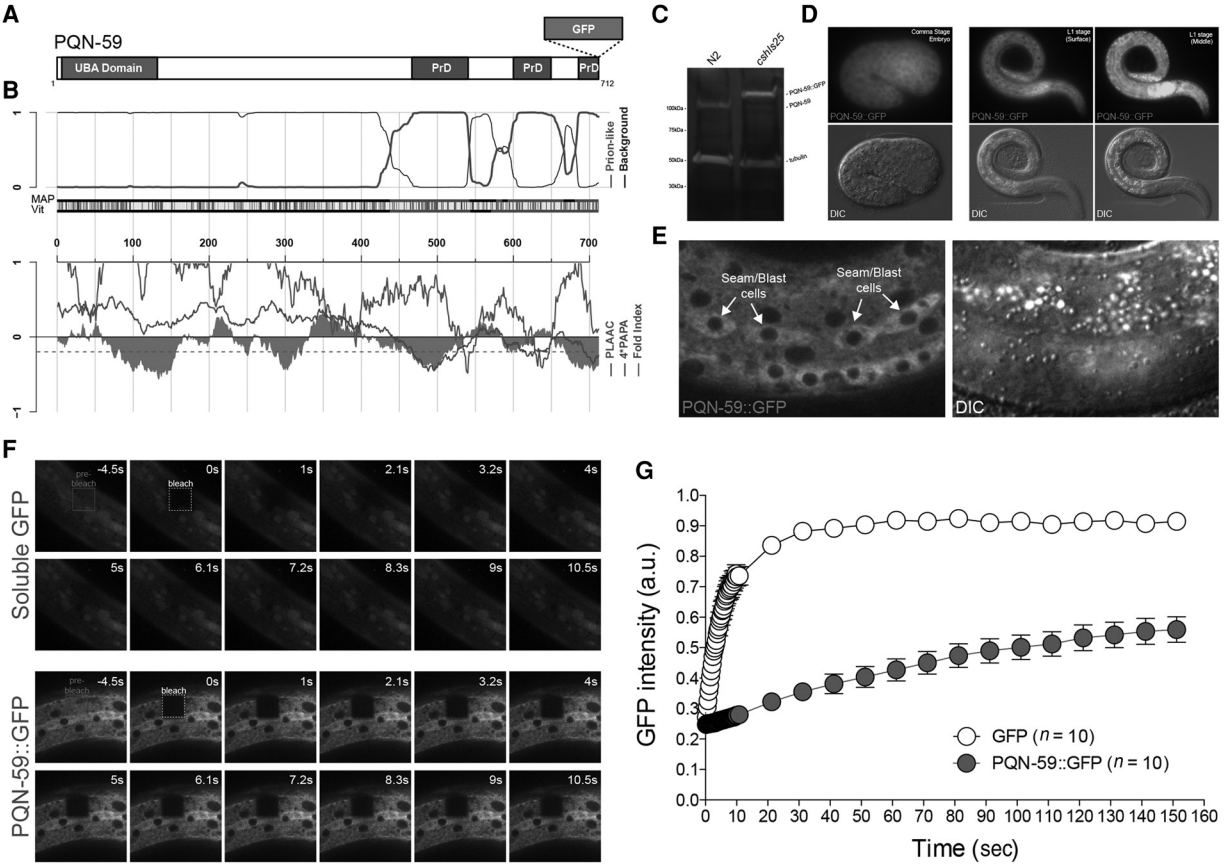


Figure 2. *pqn-59* encodes a ubiquitously expressed protein that forms an insoluble cytoplasmic matrix. (*A*,*B*) PQN-59 is a predicted prion-like protein that harbors a UBA domain and several stretches of amino acids that are predicted to form an amyloid. (*C–E*) A CRISPR-Cas9 GFP-tagged allele is expressed in most *Caenorhabditis elegans* tissues and is localized exclusively to the cytoplasm. (*F*,*G*) FRAP experiments indicate that PQN-59::GFP has limited infusibility in hypodermal tissues.

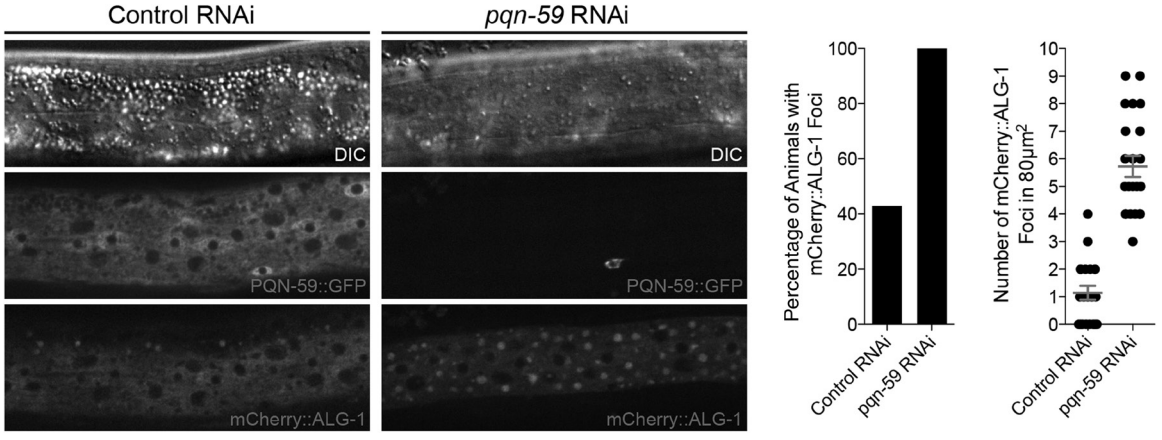


Figure 3. Depletion of *pqn-59* results in the mislocalization of the microRNA-specific argonaute protein. Under normal conditions, both mCherry::ALG-1, a core miRISC component, and PQN-59::GFP are localized throughout the cytoplasm of hypodermal cells. Depletion of *pqn-59* by RNAi causes ALG-1 to mislocalize to dense, cytoplasmic aggregates. *pqn-59* RNAi results in the percentage of animals that have any mCherry::ALG-1 foci in hypodermal cells, as well as the concentration of mCherry::ALG-1 foci in these tissues.

a cytoplasmic matrix that functions to keep miRISC and associated proteins from diffusing. To test this hypothesis, we examined the localization of miRISC machinery under conditions in which *pqn-59* was depleted via RNAi. Surprisingly, depletion of PQN-59 causes the core miRISC argonaute protein (ALG-1) to form large aggregates in the cytoplasm (Fig. 3). This suggests that PQN-59 dampens miRNA activity by regulating the local concentration of miRISC. Future experiments will determine whether PQN-59 functions to prevent other cellular components from inappropriately accumulating in the cytoplasmic compartment.

PUBLICATION

Aguirre-Chen C, Kim N, Mendivil Ramos O, Kramer M, McCombie WR, Hammell CM. 2016. *C. elegans* PVD neurons: A platform for functionally validating and characterizing neuropsychiatric risk genes. *bioRxiv* doi: 10.1101/053900.

SPATIOTEMPORAL RECONSTRUCTION OF THE TRANSCRIPTOME IN DEVELOPMENT AND CANCER

J. Lee O. Chaudhary D. Ghosh E. Rozhkova
X. Cheng M. McCartan

Our lab has been developing new sequencing chemistries and computational methods to enable the visualization of gene expression patterns genome-wide across whole embryos or human tissues in their native context. Current approaches use single-cell RNA sequencing combined with single-molecule fluorescent in situ hybridization (FISH) for mapping the cellular heterogeneity in space; however, these methods do not allow for a rapid and efficient de novo mapping of the transcriptome, requiring practically unscalable efforts to generate a reference map or validate individual gene clusters experimentally.

We are now developing two approaches to randomly sample the whole transcriptome using three-dimensional (3D) imaging, in which the transcripts are detected in a stochastic manner using in situ RNA sequencing (fluorescent in situ sequencing, FISSEQ) or genome-wide FISH. Here, the low sensitivity of FISSEQ yields a sparse set of highly distinct transcript signatures immobilized in the intact tissue environment across thousands of cells or embryos. Because this process is random, cells and embryos with similar topology, orientation, and gene expression signatures can be combined to generate a high-resolution atlas of gene expression using a computational algorithm similar to that of cryoEM.

To enable highly sensitive, yet targeted, reconstruction of gene expression in space, we have developed a new type of sequencing chemistry in collaboration with several industry partners. Here, sequencing by ligation is performed directly on the RNA template, and one can choose a large number of specific genes to be amplified for visualization. We are now developing ways to stochastically turn on the fluorescence from individual amplicons across thousands of embryos in parallel. This could result in highly sensitive FISH-like detection sensitivity and next-generation sequencing (NGS)-like sequence specificity, while enabling a cryoEM-like reconstruction method using stochastic and random transcript detection.

Inferring Cell Lineage by In Situ Sequencing of Induced Somatic Mutations

O. Chaudhary, M. McCartan

We are in the first year of a Human Scientific Frontier Grant (M. Averof, M. Telford, and J. Lee) whose aim is to spatially map the cell fate phylogeny in 3D with single-cell resolution during early *Drosophila* development. Here, a single reporter bearing 32 Cas9-targeted repeats is combined with tuned guide RNAs (gRNAs), which generate somatic mutations with a low frequency in the reporter over time. The sequential accumulation of random deletions in the 32 targeted loci is then sequenced in situ for phylogenetic reconstruction and 3D visualization.

We have generated *Drosophila* strains with the appropriate mutational frequencies and performed phylogenetic reconstruction simulations to determine the optimal number of repeats and sequence read length required for at least 20% accuracy in cell fate reconstruction. We are now developing sequencing chemistries that can discriminate the transcripts from these repeats with high specificity. Our approach enables one to scale the number of reporters to various subcellular compartments for simultaneous imaging so that more than 200 independent mutation-bearing loci can be sequenced in situ, enabling cell fate reconstruction in later developmental time points with high accuracy.

Although other groups have used similar somatic mutation methods for phylogenetic reconstruction, they lack sufficient spatial resolution or scalability for barcode detection in vivo. Our goal in 2017–2018 is to show the specificity and the sensitivity of our approach for at least one or two repeat elements and perform cell fate mapping experiments during *Drosophila* blastoderm development, in combination with the whole-embryo transcriptome mapping described above.

The Genetic Basis of Glioma Histopathology and Clonal Evolution

D. Ghosh, X. Cheng, E. Rozhkova

In this STARR Cancer Consortium–funded collaboration with Dr. Timothy Chan at Memorial Sloan Kettering Cancer Center (MSKCC) and Dr. Molly Hammell (CSHL), we are developing experimental protocols and computational tools necessary to implement FISSEQ for paraffin-embedded clinical brain tissue samples. Because low-grade glioma (LGG) has a well-characterized histological appearance, genetic signatures, and molecular markers (i.e., IDH, ATRX) and the clinical history, we are focusing on Stage II–III astrocytoma for our approach. Here, the key challenge is to generate a moderate density of highly fluorescent amplicons that can be sequenced in situ using FISSEQ. By combining many similar cell, region, or tissue features of interest, our goal is to create a spatially defined map of LGG stage- or histology-specific gene expression.

So far, we have increased our FISSEQ imaging speed by 1000-fold and developed new sequencing reagents with the help of ThermoFisher. We expect to solve the problem of generating brighter amplicons for imaging relatively soon. Finally, we hope to incorporate our cell lineaging tool from above in mice glioma models later to study the effect of tissue architecture, microenvironment, and metabolism on clonal evolution of brain glioma precursors.

The Genetic Basis of Tissue Patterning in Space

E. Rozhkova, X. Cheng

This work is funded by the National Institute of General Medical Sciences (NIGMS) R35 to develop tools required to quantify the forces regulating gene expression during tissue patterning in development. We now have a clear methodological and computational framework to tackle this problem using 4D tomography based on FISSEQ or stochastic FISH described in the introduction. Our goal in 2017 and 2018 includes imaging up to 5000 *Drosophila* embryos and approximately 10,000 polarized epithelial stem cells in parallel using FISSEQ for a detailed map of the transcriptome in space. We will use genetic mutants and well-established pair-rule genes to estimate the sensitivity and the specificity of our approach.

The main biological aim of our grant is to understand how cells use a gradient of mechanotransduction (i.e., FAT) to regulate the morphogen interpretation and signal thresholds for error-free cell positioning during development. Although we are establishing our technology, we are also using an in vitro model of embryonic stem cell patterning, single-cell analysis, and FISSEQ to show that local mechanotransduction signaling leads to single-cell heterogeneity and tissue patterning in culture.

Translating Advanced Sequencing Technologies into General Tissue Stains

O. Chaudhary, M. McCartan

Although modern advances such as NGS, single-cell analysis, and single-molecule imaging have been transformative in biomedical research, clinical pathologists still can far outperform when it comes to tumor diagnosis and staging in terms of timeliness, scalability, cost, and ease of implementation. These considerations impact patient management in real time in the setting of limited dollars and aging populations in health care.

We are inspired by how hematoxylin and eosin (H&E) and other general tissue stains have remained the bedrock tools in medicine despite their broad specificity. Rather than taking the standard bioinformatics approach to NGS, our lab is developing NGS chemistries into general tissue stains that can paint multiple mutations, transcriptional signatures, and metabolic status. The key challenge here is to have a pool of probes that make up each "stain" be single-nucleotide-specific and to eliminate nearly all off-target sequence staining in situ. We have now met structural and sequence requirements for RNA detection with >90% specificity and sensitivity in situ, and we are exploring ways to remove false-positive binders.

Our goal over the next several years is to work with industry partners to achieve robust signal generation and RNA signal detection, so that any RNA nucleotide sequences can be sequentially stained using a standard set of universal or target-specific reagents. With proper encoding, we expect that a large number of transcriptional patterns can be appreciated using a small number of successive staining, potentially providing a means of assessing genetic determinants of tissue diagnosis in a clinical setting.

In Press

Lee J. 2017. Quantitative approaches for investigating the spatial context of gene expression. *WIREs Syst Biol Med* **9:** e1369.

CANCER GENETICS

A. Mills L. Banks D.-W. Hwang A. Macias C. Stahlhut-Espinosa
Y. Chang A. Jaganathan P. Shrestha S. Sun

Our laboratory is focused on determining the genetic/epigenetic basis of cancer and neurodevelopmental syndromes. We have discovered new genes impacting these conditions, revealed new insight into how the encoded proteins work in normal cells, and determined how deregulation in these processes contributes to disease. These findings have had a major impact in the scientific community and have affected how clinicians analyze and treat patients with these syndromes.

Major discoveries include (1) identifying *p63* as a gene affecting development, aging, and cancer; (2) defining the genetic basis of autism; (3) discovering *CHD5* as a gene that prevents cancer; and (4) determining that Chd5 loss causes male infertility.

p63 in Development, Aging, and Cancer

We discovered *p63*, a gene related to *p53*—a gene encoding a tumor suppressor defective in more than half of all human cancers. Although similar to p53, it was not at all clear how *p63* worked. We found that a lack of p63 leads to aging, as p63 is needed for stem cell renewal, and lack of p63 causes curvature of the spine, hair loss, and severe skin lesions. Yet, there is a fine balance, as an excess of one version of p63 (ΔNp63α) causes carcinoma—the most prevalent type of human cancer. In contrast, we found that a different version of p63 (TAp63) prevents cancer. Our work showing that TAp63 inhibits tumor growth, even when p53 is absent, was surprising: It had always been believed that p53 was absolutely essential for inhibiting cancer. Instead, we found that TAp63 can do the job alone.

We discovered that p63 was needed for development: Its loss causes malformations of the limbs, skin, and palate. This finding led others to interrogate *p63* and to reveal that its mutation causes seven different human syndromes involving birth defects affecting the limbs, skin, and palate. By generating mouse models for one of these syndromes (ectrodactyly, ectodermal dysplasia, clefting [EEC] syndrome), we revealed new insight into why some children with EEC syndrome have symptoms that are very severe and even life-threatening, whereas other children with EEC—even those in the same family with the same *p63* mutation—have symptoms that are barely noticeable. During the past year, we published a collaborative study implicating p63 in cartilage development. We are currently working to understand how p63 regulates stem cell biology and how its perturbation leads to cancer.

CHD5, A New Cancer-Preventing Gene

We discovered *CHD5* as a tumor suppressor mapping to human 1p36—a region of our genomes frequently deleted in cancer. 1p36 deletions occur in many different types of human tumors, including those of the epithelia, brain, and blood. Although this suggested that a cancer-suppressing gene resided in this region, its identity remained a mystery. By generating mice with deletions and duplications of the genomic region corresponding to 1p36 using chromosome engineering technology—a strategy with which we can generate precise chromosome rearrangements in the mouse—we pinpointed a region of the genome with potent tumor-suppressive activity. Using genetic and molecular approaches, we identified *CHD5* as the tumor suppressor gene in the region and found that its product worked as a "circuit breaker" that empowers a network of tumor suppressors. In addition, we discovered that *CHD5* was frequently deleted in human glioma. Chromosome engineering proved so powerful for discovering the 1p36-encoded gene that had challenged the cancer community for over three decades that we also used it for studying neurodevelopmental syndromes, including schizophrenia and autism. This technology was essential for several collaborative studies focused on copy number variations responsible for autism. We also used the chromosome engineering strategy to identify cancer genes in a collaborative project reported this past year.

We continue to focus on defining the role of CHD5 in chromatin dynamics and deciphering how

dysregulation of CHD5 and the pathways it regulates leads to disease. We found that Chd5 uses its plant homeodomains to bind histone 3 and that this is essential for tumor suppression. Our work paved the way for further discoveries, and *CHD5* is now known to be mutated in human cancers of the breast, ovary, and prostate, as well as in melanoma, glioma, and neuroblastoma. Furthermore, recent reports indicate that patients with high levels of *CHD5* have much better overall survival than those with low levels. We found that Chd5 is essential for packaging DNA, and that loss of Chd5 causes improperly packaged DNA that is prone to DNA damage. Intriguingly, Chd5's absence is particularly important during the process of sperm maturation—an event in which the DNA is first unpackaged and then repackaged using an elaborate series of steps—and that deficiency of Chd5 causes male infertility. We discovered that Chd5 is expressed highly in neurons, and within the last year we found that Chd5 plays a pivotal role in the brain, suggesting that inappropriate DNA packaging contributes to neurodevelopmental syndromes such as autism. We are currently delving deeper into the mechanisms whereby Chd5-mediated regulation of chromatin affects gene expression cascades that regulate neuronal stem cells and how deregulation of this process sets the stage for neurodevelopmental syndromes and cancer.

PUBLICATIONS

Liu Y, Chen C, Xu Z, Scuoppo C, Rillahan CD, Gao J, Spitzer B, Bosbach B, Kastenhuber ER, Baslan T, et al. 2016. Deletions linked to *TP53* loss drive cancer through p53-independent mechanisms. *Nature* **531:** 471–475.

In Press

Mills AA. 2017. The chromodomain helicase DNA-binding chromatin remodelers: Family traits that protect from and promote cancer. *Cold Spring Harb Perspect Med* doi: 10.1101/cshperspect.a026450.

Taniguchi Y, Kawata M, Ho Chang S, Mori D, Okada K, Kobayashi H, Sugita S, Hosaka Y, Inui H, Taketomi S, et al. 2017. Regulation of chondroitin survival in mouse articular cartilage by p63. *Arthritis Rheumatol*. **69:** 598–609.

CANCER GENOMICS

S. Powers F. Celeste A.P. Delgado J. Li A. Mofunanya M. Rao X. Zhao

Functional Genomic Screening for Oncogenic Drivers and Therapeutic Targets

A. Mofunanya, J. Li

Much of our work is funded by the National Cancer Institute (NCI) to work individually and collaboratively (CTD²; https://ocg.cancer.gov/programs/ctd2) to use cancer genome data for the discovery and development of human cancer therapeutics. We have continued to follow up on a pooled open reading frame (ORF) oncogenic functional screening of a set of overexpressed genes in human breast cancer. We have found that *CEACAM5*, a well-known tumor marker thought to be a passive passenger, plays an active role in tumorigenesis, as does *SCUBE3*, a secreted protein that we are pursuing as a therapeutic target.

Accurately Modeling Multigenic and Large Copy Number Alterations in Human Breast Cancer

M. Rao

Multigenic and large copy number alterations are a dominant feature of human breast cancer, but there is still a lack of experimentation to determine specific effects. We are using CRISPR-Cas9 genomic engineering to introduce an amplifiable dihydrofolate reductase (*DHFR*) gene into multiple genomic sites of normal mammary epithelial cells to examine the effects of localized amplification of different regions on tumorigenic properties. As predicted from prior work, amplification only occurs when *TP53* is inactivated.

Bad Luck Hypothesis

This work was done in collaboration with Y. Hannun.

There is a strong correlation between tissue-specific cancer risk and the lifetime number of tissue-specific stem-cell divisions. Whether such correlation implies a high unavoidable intrinsic cancer risk has become a key public health debate with the dissemination of the "bad luck" hypothesis. Contrary to others, we found that the correlation between stem-cell division and cancer risk does not distinguish between the effects of intrinsic and extrinsic factors. We found that intrinsic risk is better estimated by the lower bound risk controlling for total stem-cell divisions, and the rates of endogenous mutation accumulation by intrinsic processes are not sufficient to account for the observed cancer risks.

Would Inducing Reversion to Normal Cellular Phenotypes Be Better than Killing Cancer Cells?

This work was done in collaboration with R. Pollack.

How can we stop cancer progression? Targeted therapies aim to overcome alterations in tumor suppressor genes and oncogenes. Secondary mutational escape from those treatments has become a major impediment because it leads to resistance. However, treatments based on selection and analysis of stable revertants could create more durable remissions by reducing the selective pressure that leads to rapid drug resistance.

PUBLICATIONS

Cancer Target Discovery and Development Consortium. 2016. Transforming big data into cancer-relevant insight: An initial, multi-tier approach to assess reproducibility and relevance. *Mol Cancer Res* **14:** 675–682.

Lee HJ, Li CF, Ruan D, Powers S, Thompson PA, Frohman MA, Chan CH. 2016. The DNA damage transducer RNF8 facilitates cancer chemoresistance and progression through twist activation. *Mol Cell* **63:** 1021–1033.

Li J, Sordella R, Powers S. 2016. Effectors and potential targets selectively upregulated in human KRAS-mutant lung adenocarcinomas. *Sci Rep* **6**: 27891.

Powers S. Pollack RE. 2016. Inducing stable reversion to achieve cancer control. *Nat Rev Cancer* **16:** 266–270.

Scott KL, Powers S. 2016. Functionally assessing candidate drivers advances precision cancer medicine. *Cancer Cell* **30:** 187–189.

Wu S, Powers S, Zhu W, Hannun YA. 2016. Substantial contribution of extrinsic risk factors to cancer development. *Nature* **529:** 43–47.

In Press

Merkin RD, Vanner EA, Romeiser JL, Shroyer AL, Escobar-Hoyos LF, Li J, Powers RS, Burke S, Shroyer KR. 2017. Keratin 17 is overexpressed and predicts poor survival in estrogen receptor–negative/human epidermal growth factor-2–negative breast cancer. *Hum Pathol* **62:** 23–32.

CANCER AND HUMAN GENETICS

M. Wigler J. Alexander, P. Andrews, T. Forcier, I. Hakker, J. Kendall, Y.-H. Lee, S. Li, B. Ma, S. Marks, J. McIndoo, L. Rodgers, M. Ronemus, J. Rosenbaum, A. Stepansky, Z. Wang, M. Wigler, B. Yamrom, C. Yoon

Our lab works in three areas: genomics, cancer, and human genetics, the latter with a strong emphasis on autism. We do this in close collaboration with three other CSHL laboratories headed by Dan Levy, Alex Krasnitz, and Ivan Iossifov. Our autism sequence data comes from collaboration with the New York Genome Center, where a good portion of our computation also occurs. I now describe our efforts in these three areas.

Genomics

Most of our efforts in this area surround issues of structural variation (SV) and genome assembly.

Copy number. We have a long-term interest in detecting large-scale copy number variation in both normal and cancer genomes, and its role in genetic disorders. This interest has led to methods for assessing cancer outcome, detecting cancer cells, identifying oncogenes and tumor suppressors, as well as exploring the involvement of copy number variation in human genetic disorders. Those applications are all discussed later, but recently we published a method called SMASH (Wang et al. 2016) that detects copy number variation as cheaply as is theoretically possible by DNA sequencing. We break subject DNA into pieces just large enough for mapping to the genome, ligate them into read-length fragments, sequence those fragments, map the pieces of each read to the genome, and create copy number profiles using specially designed algorithms based on map density. Cost reduction comes from having multiple independent mappings per read, up to six per read fragment. We have used SMASH for screening cancer specimens and cases of pediatric disorders. This effort was partly supported by the Simons Foundation and was a collaboration with Jude Kendall, Mike Ronemus, Dan Levy, and Zihua Wang.

SV. Most of our exploration of human genetics derives from analysis of whole-genome sequence data from short-read platforms. SV, including indels, translocations, inversions, microsatellite instability, pseudogene formation, and transpositions, is a rich source of genetic anomalies. No available genome analysis tool addressed all these things, so we have developed our own. We call it MUMdex, and the underlying approach is to use exact matches (MUMs or maximal unique matches) to a reference genome. When the distance or orientation between two MUMs in a read differs from their expected distance or orientation in the reference genome, then we have detected a candidate structural variant. With sufficient examples of these events, the candidate grows stronger. Gathering the data over populations, we can then determine if a structural variant is rare or common. Gathering the data over a family, we collect if a structural variant is de novo in a child. Comparing a cancer to the host genome, we find the somatic structural variants of the cancer. Unlike other methods, MUMdex is indifferent to the type of structural variant or its size. A preprint is published in the BioRxiv (Andrews et al. 2016). The tool compares well with the various other more standard tools that are pieced together to do various aspects of the same thing. MUMdex has more general utility than SV discovery: it can also find single-nucleotide variants (SNVs), if they are not too dense in a given region of the genome, and with some tweaking we should be able to extend it to discover SV such as transpositions involving repetitive elements. This effort received partial support from the Simons Foundation, and was a collaboration with Peter Andrews and Dan Levy.

Large-scale haplophasing. A persistent problem in genomics is haplophasing: how to assign variants in an individual to the proper parent in the absence of the parental DNA. We have devised a solution to the large-scale problem, which we call "HaHa" or haplotyping by halving. Nuclei from the individual are cleaved by cryotome and subnuclear fragments are

sorted into single compartments by flow sorting. We then use single nuclear sequencing, a methodology pioneered by this laboratory. With algorithms designed by Dan Levy, using techniques akin to hidden Markov processing, variants in the same subnuclear fraction are aggregated into parental haplophases. The method has been validated in cases in which family DNAs were available and for genomes that were partially phased by long-range sequencers. This work was partly supported by a grant from Calico, the Simons Foundation, and a subaward from a grant from the National Human Genome Research Institute (NHGRI) to the New York Genome Center. This work was a collaboration with Dan Levy and Partha Mitra.

Assembly by template mutagenesis. Another persistent problem in genomics is the genome assembly problem, which arises when the structures to be assembled have a duplicated structure and/or sufficient variation between haplotypes to allow phased assembly with a given sequence platform. One solution to this problem has been the development of long-read sequence platforms such as PacBio and Nanopore. We have approached it by conversion of short-read sequencers, which are vastly cheaper and more accurate than the long-read sequencers. We published the theory for this in 2014 (Levy and Wigler, *Proc Natl Acad Sci 111:* E4632 [2014]). The principle is to introduce random mutations into the template structures, thereby introducing enough variation into each template to make them unique, which in turn permits the long-range assembly of each template from short-read data. We have now reduced the method to practice using partial bisulfite mutagenesis, as published in *BioRxiv* (Kumar et al. 2017). Further methods are in development to make protocols more robust and encode efficient algorithms for sequence data processing. This work was supported by the Simons Foundation and a subaward from a grant from the NHGRI to the New York Genome Center. This work was a collaboration with Vijay Kumar and Dan Levy.

Cancer

Most of our effort in the cancer area focuses on risk assessment, response to therapy, and early detection. Although the economic incentives for improvements in clinical care for these advances are poor, the benefits to the patient are obvious.

Single-cell analysis of prostate cancer biopsy. Many more patients are biopsied for prostate cancer than are treated, and the guides for treatment are quite complex and not standard, but are based on morphological Gleason score of core prostate biopsies. We seek to bring genomic methods to this problem. Sparse DNA sequencing of single-cell nuclei from prostate core biopsies is a rich source of quantitative parameters for evaluating neoplastic growth and aggressiveness. These include the presence of clonal populations, the phylogenetic structure of those populations, the degree of the complexity of copy number changes in those populations, and measures of the proportion of cells with clonal copy number signatures. The parameters all show good correlation to the Gleason score, derived from individual prostate biopsy tissue cores and radical prostatectomy surgical specimens. Indeed, these genomic parameters correlate better with the Gleason score of the radical prostatectomy than does the Gleason score of the core biopsies. This is highly relevant because primary treatment decisions are dependent on the biopsy and not the surgical specimen. Thus single-cell analysis has the potential to augment traditional core histopathology, improving both the standardization and accuracy of risk assessment, and hence treatment decisions. Future emphasis will be on making single-cell analysis cheaper so that our methods can become affordable clinical tools. This work was a collaboration between Alex Krasnitz, Joan Alexander, Jude Kendall, and two clinicians, Ashutosh Tewari (Mt. Sinai School of Medicine) and Herbert Lepor (NYU Medical School).

Measuring minimal residual disease (MRD). The concept of MRD has its origins in the treatment of leukemia and lymphoma, in which the persistence of neoplastic cells can be followed by their natural occurrence in the blood. The notion of using molecular techniques to assess MRD also has its roots in leukemia and lymphoma and, in particular, following chronic myelogenous leukemia (CML) through the presence of the Philadelphia chromosome translocation. In principle, measuring MRD allows the clinician to determine if therapy is being successful, and to make a determination whether to continue, switch the therapeutic regimen, or take a radically different approach. In fact, for leukemias and lymphomas, this is common practice. It is now apparent that even solid cancers shed cells and

DNA into the blood, and this opens up the possibility of following such cancers in a similar manner. After performing the sequence analysis of the presenting neoplasm, we determine the cancer-specific nucleotide mutations. We then apply techniques we developed for error-free sequencing. These allow the detection of single-nucleotide cancer variants at less than one part per million. The basis for this method is template tagging with randomly generated oligonucleotides (which we call "varietal tags") (Hicks et al., U.S. patent application publication U.S. 2014 and U.S. patent publication 0065609 [March 6, 2014]). By integrating information over each unique template, sequencing errors are drastically reduced. We are applying the method to measuring MRD from leukemias and breast cancer in a clinical collaboration with Northwell Health hospital system. This work is a collaboration with Zihua Wang, Dan Levy, Alex Krasnitz, Joan Alexander, Jude Kendall, and Peter Andrews from CSHL, and Steve Allen and Dan Budman, both clinicians from Northwell Health. We received research support from the Breast Cancer Research Foundation (BCRF), Northwell–CSHL Alliance, and the Simons Foundation.

Early detection of cancer incidence. Cancer kills by spreading to distant sites. At the time of the first clinical presentation, metastasis has typically already occurred. Were it otherwise, most cancers would be curable by surgery. It follows that there may be a window of time when detection of cancer and its timely extirpation will result in a cure. Some, perhaps most, cancer spreads to distant sites through blood. We analyze the feasibility of the early detection of cancer in blood using genomic analysis. As any early detection method must not cause false alarms and must be verifiable and actionable, we further propose isolating the suspect cells for further analysis. The method rests on detecting the presence of recurrent patterns of copy number profile in some of the candidate cells using sparse single-cell sequencing. We show feasibility by simulation, using all copy number profiles from nearly 3900 cases described in the Cancer Genome Atlas, and generating single-cell data in silico. For the latter, we devise procedures for sampling from actual diploid single-cell data to produce data for a cell with a given profile. The algorithm for detection requires searching for connected components of highly correlated cell profiles. We recently published our theoretical analysis. This work was a collaboration with Alex Krasnitz and Dan Levy. We received research support from the Simons Foundation.

Autism

Autism spectrum disorder (ASD), a collection of developmental delay syndromes characterized by deficient social skills and communication, receives a strong contribution from genetics. Whereas in the past our focus was genetic contribution from de novo mutation in candidate genes, we now look more broadly into the spectrum of causation. This work is supported by the Simons Foundation.

Shared ancestral variation. Given the proportion of autism that is multiplex, we do not expect that all autism is explainable by de novo mutation, and that transmission of risk variants plays a role. There has been persistent evidence of shared variation from case-control studies. These approaches universally use a liability threshold model developed by Yang et al. (*Nat Genet 42*: 565–569 [2010]), applied initially to estimate the genetic contribution to quantitative traits such as height on a simple random sample from the population. We have not been satisfied that this method has been correctly applied, and so investigated the question with our own data and methods. We developed a method of analysis (A2DS) that tests, generally, whether shared genomic variants contribute to a disorder. Using a standard measure of genetic relation, test individuals are compared with a cohort of discordant sib-pairs (CDS) to derive a comparative similarity score. We ask whether a test individual is more similar to an unrelated affected than to the unrelated unaffected sibling from the CDS and then sum over such individuals and pairs. Statistical significance is judged by randomly permuting the affected status in the CDS. In the analysis of published genotype data from the SSC (Simons Simplex Collection) and the AGRE (Autism Genetic Research Exchange) cohorts of children with ASD, we find strong statistical significance that the affected are more similar to the affected than to the unaffected of the CDS (p-val ~ 0.00001). Fathers in multiplex families have marginally greater similarity (p-val = 0.02) to unrelated affecters. These results do not depend on ethnic matching or gender. This was a collaboration with Kenny Ye from Albert Einstein College of Medicine, and others, and has recently been published (Ye et al. 2017).

De novo genetic damage and loss of motor skills. In individuals with ASD, de novo mutations have previously been shown to be significantly correlated with

lower IQ, but not with the core characteristics of ASD: deficits in social communication and interaction, restricted interests, and repetitive patterns of behavior. We extend these findings by showing in the SSC that damaging de novo mutations in ASD individuals are also significantly and convincingly correlated with measures of impaired motor skills. This correlation is not explained by a correlation between IQ and motor skills. We find that IQ and motor skills are distinctly associated with damaging mutations and, in particular, that motor skills are a more sensitive indicator of mutational severity, as judged by the type and its gene target. We use this finding to propose a combined classification of phenotypic severity: mild (little impairment of both), moderate (impairment mainly to motor skills), and severe (impairment of both). A preprint of this work has been recently published (Buja et al. 2017), a collaboration with Ivan Iossifov of CSHL and Andreas Buja of the University of Pennsylvania.

Contribution from de novo indels in the introns of candidate target genes. Copy number profiling and whole-exome sequencing have allowed us to make remarkable progress in our understanding of the genetics of autism over the past 10 years, but there are major aspects of the genetics that are unresolved. Through whole-genome sequencing, additional types of genetic variants can be observed. These variants are highly abundant, and knowing which are functional is challenging. We have analyzed whole-genome sequencing data from 510 of the SSC's quad families and focused our attention on intronic variants. Within the introns of 546 high-quality autism target genes, we identified 63 de novo indels in the affected and only 37 in the unaffected siblings. The difference of 26 events is significantly larger than expected (p-val = 0.01), and using reasonable extrapolation shows that de novo intronic indels can contribute to at least 10% of simplex autism. The significance increases if we restrict to the half of the autism targets that are intolerant to damaging variants in the normal human population—the half we expect to be even more enriched for autism genes. For these 273 targets, we observe 43 and 20 events in affected and unaffected siblings, respectively (p-val of 0.005), and a drop from 26 to 23. There was no significant signal in the number of de novo intronic indels in any of the control sets of genes analyzed. We see no signal from de novo substitutions in the introns of target genes. A preprint of this work has been recently published (Munoz et al. 2017) and is a collaboration with Ivan Iossifov.

PUBLICATIONS

Andrews PA, Iossifov I, Kendall J, Marks S, Muthuswamy L, Wang Z, Levy D, Wigler M. 2016. MUMdex: MUM-based structural variation detection. *bioRxiv* doi: 10.1101/078261.

Wang Z, Andrews P, Kendall J, Ma B, Hakker I, Rodgers L, Ronemus M, Wigler M, Levy D. 2016. SMASH, a fragmentation and sequencing method for genomic copy number analysis. *Genome Res* **26:** 844–851.

In Press

Buja A, Volfovsky N, Krieger A, Lord C, Lash AE, Wigler M, Iossifov I. 2017. Damaging mutations are associated with diminished motor skills and IQ in children on the autism spectrum. *Neuron* (submitted).

Kumar V, Rosenbaum J, Zihua W, Forcier T, Ronemus M, Wigler M, Levy D. 2017. Mutational sequencing for accurate counter and long-range assembly. *bioRxiv* doi: 10.1101/149740.

Munoz A, Yamrom B, Lee Y-A, Marks S, Lin K-T, Wang Z, Krainer AR, Darnell RB, Wigler M, Iossifov I. 2017. De novo indels within introns contribute to ASD incidence. *bioRxiv* doi: 10.1101/137471.

Ye K, Iossifov I, Levy D, Buja A, Krieger A, Wigler M. 2017. Deducing the contribution of shared variants from discordant siblings: Applications to autism. *Proc Natl Acad Sci* **114:** 7073–7076.

CANCER: SIGNAL TRANSDUCTION

Mikala Egeblad and colleagues study tumors and, in particular, the contributions of the microenvironment in which the cancer cells arise and live. Solid tumors are abnormally organized tissues that contain not only cancer cells but also various stromal cell types and the extracellular matrix, and these latter components constitute the microenvironment. Communications between the different components of the tumor influence its growth, response to therapy, and ability to metastasize. Among the tumor-associated stromal cells, the lab's main focus is on myeloid-derived immune cells, a diverse group of cells that can enhance angiogenesis and metastasis and suppress the cytotoxic immune response against tumors. Egeblad is interested in how different types of myeloid cells are recruited to tumors and how their behaviors—for example, their physical interactions with cancer cells and other immune cells—influence cancer progression, including metastasis. The Egeblad lab studies the importance of the myeloid cells using mouse models of breast and pancreatic cancer and real-time imaging of cells in tumors in live mice. This enables them to follow the behaviors of and the interactions between cancer and myeloid cells in tumors during progression or treatment. This technique was instrumental when the lab recently showed that cancer drug therapy can be boosted by altering components of the tumor microenvironments, specifically reducing either matrix metalloproteinases (enzymes secreted by myeloid cells) or chemokine receptors (signal receptors on myeloid cells). This year, the Egeblad lab collaborated with Scott Powers' group to understand how normal cells surrounding a tumor promote cancer growth. They found that normal cells signal to tumors through multiple pathways, and blocking these signals together has the greatest effect on inhibiting tumor growth—offering a new strategy to fight cancer.

Douglas Fearon's laboratory studies the interaction between cancer and the immune system. Their underlying premise is that the tumor microenvironment is immune suppressive because cancer cells elicit responses characteristic of wound healing and tissue regeneration. This approach has led to the finding that activated fibroblasts in the tumor stroma mediate immune suppression in several mouse models of cancer, including the autochthonous model of pancreatic ductal adenocarcinoma of the Tuveson lab. Their understanding of the basis of immune suppression is evolving, but they know that it involves the production of the chemokine, CXCL12, by the fibroblastic stromal cells, binding of this CXCL12 by pancreatic cancer cells and exclusion of T cells from the vicinity of the cancer cells. T-cell exclusion, which also occurs in several types of human adenocarcinomas, causes antagonists of T-cell checkpoints to be ineffective, despite the presence of cancer-specific $CD8^+$ T cells. This immune suppression is interrupted by administering AMD3100, an inhibitor of CXCR4, the receptor for CXCL12, which leads to the rapid accumulation of T cells among cancer cells, thereby uncovering the efficacy of anti-PD-L1 and eliminating cancer cells. Because human pancreatic cancer has certain immunological characteristics of the mouse model, a phase 1 clinical trial of AMD3100 in patients with pancreatic cancer will be initiated. Some of the next steps are to determine the biological process that causes cancer cells to express nonmutated, shared antigens, and the means by which dormant metastases escape immune elimination.

Darryl Pappin's lab develops chemical and computational methods for analysis of proteins and peptides. These are fundamental tools for proteomics, and they are vital in many fields of biological investigation. Proteins and peptides are typically analyzed via mass spectrometry, a method that involves fragmenting samples by colliding them with gas atoms in a vacuum. Masses of the resulting fragments are measured, and computer algorithms match the results with known or predicted molecules whose amino acid sequences are either known or inferred. Pappin has developed search engines for mass spectrometry data that enable investigators to sift hundreds of thousands

of experimental spectra at a time for database matches. He also seeks to reduce sample complexity via an approach he calls chemical sorting. This includes the use of chelation to enrich phosphopeptides from the total peptide pool and the use of specific affinity-tagged small-molecule inhibitors to segregate classes of kinases or phosphatases for more specific mass spectroscopic analysis.

Despite their large variety of genetic abnormalities, cancer cells have been found to be extremely sensitive to the reversal of certain mutations. **Raffaella Sordella** and colleagues study why cells in certain cancers are responsive to the inhibition of one particular gene or gene product. Why, for instance, do non-small-cell lung cancer (NSCLC) cells that have a particular mutation in the epidermal growth factor receptor (EGFR) respond dramatically to its inhibition by the drug Tarceva? This occurs in 15%–20% of patients, the great majority of whom, within 1–3 years, develop resistance. Various mutations have been implicated in about one-half of resistant patients. Sordella and colleagues have discovered a new resistance mechanism in a subpopulation of NSCLC cells that are intrinsically resistant to Tarceva. These tumor cells were observed to secrete elevated amounts of a growth factor called transforming growth factor-β (TGF-β), which in turn increases secretion of interleukin-6 (IL-6), an immune signaling molecule. Significantly, these effects were independent of the EGFR pathway. The team therefore hypothesizes that inflammation is one of the factors that can render a tumor cell resistant to treatment with Tarceva. In other work, Sordella collaborates with the Krainer lab to study whether alternative splicing has a role in the failure of p53-mediated senescence to halt oncogenesis in certain lung cancers.

Nicholas Tonks and colleagues study a family of enzymes called protein tyrosine phosphatases, or PTPs, which remove phosphate groups from proteins and other signaling molecules, such as lipids, in cells. Disruption of PTP function is a cause of major human diseases, and several of the PTPs are potential therapeutic targets for such diseases. Tonks' group seeks to fully characterize the PTP family, understanding how PTP activity is controlled and how PTPs modify signaling pathways. In addition, they are working to determine how those pathways are abrogated in serious illnesses, including cancer, diabetes, and Parkinson's disease. The overall goal is to identify new targets and strategies for therapeutic intervention in human disease. Tonks and colleagues have defined new roles for PTPs in regulating signaling events in breast cancer, identifying three PTPs as novel potential tumor suppressors. They have characterized the regulation of PTP1B by reversible oxidation, showing that it is regulated by covalent modification of the active site by hydrogen sulfide (H_2S) under conditions of endoplasmic reticulum (ER) stress that are linked to protein-folding-related pathologies such as Parkinson's and Alzheimer's. In addition, they have generated recombinant antibodies that selectively recognize the oxidized conformation of PTP1B; these antibodies display the ability to promote insulin signaling in cells and suggest novel approaches to therapy for diabetes. Finally, they have also discovered a novel mechanism for allosteric regulation of PTP1B activity, offering the possibility of developing small-molecule drugs that could inhibit the phosphatase and thereby modulate signaling by insulin and the oncoprotein tyrosine kinase HER2, potentially offering new ways to treat insulin resistance in type 2 diabetes and breast cancer.

Lloyd Trotman's recent research path begins at his discovery some years ago that the loss of a single copy of a master tumor-suppressing gene called PTEN is sufficient to permit tumors to develop in animal models of prostate cancer. His team later found that complete loss of PTEN paradoxically triggers senescence, an arrested state that delays or blocks cancer development in affected cells. These findings explained why many patients only display partial loss of this tumor suppressor when diagnosed with prostate cancer. Now, the team is researching ways to restore the PTEN protein levels in these patients. This therapeutic approach could slow disease progression and, thus, greatly reduce the need for surgical removal of the prostate or similar drastic interventions that carry the risks of incontinence and impotence. Their second approach to combat

prostate cancer is to model the lethal metastatic disease in genetically engineered mice. They are developing a novel approach that allows for quick generation and visualization of metastatic disease. The efficacy of existing and novel late-stage therapies, such as antihormonal therapy, can then be tested and optimized in these animals. At the same time, the Trotman lab is exploring the genome alterations associated with metastatic disease and resistance to therapy. To this end, they use single-cell and multicell genome sequencing techniques developed at CSHL by Dr. Wigler and Dr. Hicks.

David Tuveson's lab uses mouse and human tissue models of neoplasia to explore the fundamental biology of these diseases and thereby identify new diagnostic and treatment strategies. His team's main focus is pancreatic cancer, a lethal malignancy that has eluded clinical solutions despite intensive study. The lab's approaches at CSHL run the gamut from designing new model systems of disease to inventing new therapeutic and diagnostic platforms for rapid evaluation in preclinical and clinical settings. For example, they have adopted a new method of culturing tissue fragments indefinitely in cell culture, enabling deep analysis with genetic and pharmacological probes. In addition, therapeutic experiments in mouse models have revealed an important role of redox metabolism and stromal interactions on influencing therapeutic response. This year, they used the mouse model system to identify the mechanism of a promising drug treatment for pancreatic cancer. The lab found that using the drug in combination with more standard chemotherapeutic drugs stopped the tumor growth and lengthened life span for the mice, suggesting that the combination therapy may help overcome the drug resistance that is so commonly found in cancers. Tuveson's lab also has a strong link to clinical trials locally and internationally, with confirmation in early-phase trials the ultimate goal. Collectively, their strategy in the preclinical and clinical arena is codified as the "Cancer Therapeutics Initiative," and this initiative will provide these same approaches to the entire CSHL cancer community.

Dr. Tuveson serves as Director of the Lustgarten Foundation Pancreatic Research Laboratory at CSHL and as Director of Research for the Lustgarten Foundation.

Linda Van Aelst's lab studies how aberrations in intracellular signaling involving enzymes, called small GTPases, can result in disease. They are particularly interested in Ras and Rho GTPases, which help control cellular growth, differentiation, and morphogenesis. Alterations affecting Ras and Rho functions are involved in cancer and various neurodevelopmental disorders. Van Aelst's team has extended its prior study of mutations in a Rho-linked gene called *oligophrenin-1* (*OPHN1*), part of an effort to connect the genetic abnormalities associated with mental retardation to biological processes that establish and modify the function of neuronal circuits. In addition to a role for *OPHN1* in activity-driven glutamatergic synapse development, lab members have obtained evidence that *OPHN1* has a critical role in mediating mGluR-LTD (long-term depression), a form of long-term synaptic plasticity, in CA1 hippocampal neurons. Their findings provide novel insight not only into the mechanism and function of mGluR-dependent LTD but also into the cellular basis by which mutations in *OPHN1* could contribute to the cognitive deficits observed in patients. Defects in cortical neurogenesis have been associated with cerebral malformations and disorders of cortical organization. The Van Aelst team discovered that interfering with the function of the Rho activator DOCK7 in neuronal progenitors in embryonic cerebral cortices results in an increase in the number of proliferating neuronal progenitors and defects in the genesis of neurons. In an extension of these studies, the Van Aelst team this year showed that DOCK7 has a central regulatory role in the process that determines how and when a radial glial cell progenitor "decides" to either proliferate (i.e., make more progenitor cells like itself) or give rise to cells that will mature, or "differentiate," into pyramidal neurons. These lines of research provide novel insight into mechanisms that coordinate the maintenance of the neural progenitor pool and neurogenesis.

Hongwu Zheng's lab aims to define the complex biology of malignant glioma pathogenesis, with the ultimate goal of translating the developed knowledge into patient benefits. Although eerily similar in terms of their self-renewal capacity and distinct phenotypic plasticity, malignant glioma cells conspicuously lack the terminal differentiation traits possessed by their normal counterparts—neural progenitors. With the use of multiple approaches combining human cancer genomics, animal modeling, and stem-cell biology, Zheng has unraveled the causal relationship between aberrant differentiation and ensuing gliomagenesis. Perhaps more importantly, his team has shown that forced restoration of differentiation capacity within glioma cells can drastically attenuate their tumorigenic potential. This finding fits well with the team's overall strategy, which is to target differentiation control pathways as a novel avenue for malignant glioma treatment. To this end, they have sought to (1) develop various animal models to recapitulate human glioma pathogenesis and use them to trace and investigate in vivo tumor initiation/progression and (2) identify key pathways/players controlling normal and neoplastic neural progenitor cell renewal and fate determination.

THE INFLUENCE OF THE TUMOR MICROENVIRONMENT ON CANCER PROGRESSION

M. Egeblad J. Albrengues E. Bružas J. Cappellani M. Fein X. He V. Küttner L. Maiorino J. Park L. Puckett M. Shields R. Wysocki

Solid tumors are aberrant tissues. Like organs, solid tumors are composed of functional cells and stroma. The stroma is the supportive framework of the organs and includes the extracellular matrix (ECM) and fibroblasts, adipocytes, cells of the vascular system, and immune cells. Interactions between epithelium and stroma are essential for normal organ development as well as for tumor progression. In solid tumors, the stroma is also known as the tumor microenvironment.

We study how the tumor microenvironment influences cancer cells in the context of tumor initiation, growth, drug resistance, and metastasis. We use mouse models of breast and pancreatic cancer together with real-time spinning disk confocal microscopy in living mice (intravital imaging). This allows us to study how cancer cell proliferation, migration, and survival are influenced by stromal components in real time.

Neutrophil Extracellular Traps Promote Breast Cancer Metastasis

J. Albrengues, E. Bružas, M. Fein, V. Küttner, L. Maiorino, J. Park, R. Wysocki

More than a century ago, Dr. Stephen Paget observed that metastases develop preferentially in certain organs, suggesting that factors external to cancer cells influence metastasis and that targeting such factors might reduce the ability of cancer cells to colonize distant organs. It is now recognized that communications between cancer cells and host cells, occurring through, for example, growth factors and cytokines, do indeed play an important role in the formation of metastasis.

We compared the microenvironment of tumors formed from the metastatic 4T1 and the nonmetastatic 4T07 cell lines, originally isolated from the same spontaneous mouse breast tumor. We identified significant differences in the amount of cytokines that were secreted by the cancer cells and in the nature of the inflammatory immune cell infiltrate between the metastatic and nonmetastatic tumors. The most profound difference was in the secretion of cytokines known to mediate neutrophil recruitment, which was highly upregulated in the metastatic breast tumors compared with the nonmetastatic. Correspondingly, infiltration of neutrophils was also fivefold higher in the metastatic tumors. Neutrophils—the most abundant leukocyte in human blood—have previously been shown to promote metastasis, but they can also kill disseminated cancer cells under certain conditions.

Neutrophils and their precursors are sensitive to many chemotherapy regimens, resulting in dangerously low neutrophil numbers (neutropenia) during the course of treatment. Neutropenia carries a risk of life-threatening infections, and the American Society of Clinical Oncology recommends prophylactic treatment with neutrophil-stimulating factors, including granulocyte colony stimulating factor (G-CSF), for certain chemotherapeutic regimens. It is therefore important to determine the conditions under which neutrophils promote metastatic spread.

Neutrophils' normal function is to kill harmful microorganisms, and they can do so through three different means: (1) phagocytosis, a process whereby bacteria or fungi are engulfed and digested; (2) degranulation of cytotoxic enzymes into the extracellular space; and (3) formation of neutrophil extracellular traps (NETs). These are pathogen-trapping structures generated by expulsion of the neutrophil's DNA with associated proteolytic enzymes. NETs can be formed in tissues, as well as intravascularly. When they form inside the vasculature, they can damage the vascular cells. Recently, it was shown that NETs induced by systemic bacterial infection and formed within the vasculature aided in the metastatic seeding of lung cancer cells in the liver. NETs have been documented in human pancreatic, liver, and gastric cancers, but whether they play a role in the progression of these cancers remains unclear.

Traditionally, studies on metastasis have relied primarily on measurements made at the end point of the process, the establishment of micro- or macrometastases. However, the metastatic process is dynamic and characterized by the ability of cancer cells to move from one part of the body to another: Cells exit the primary tumor, invade the local tissue, enter blood or lymphatic vessels, and are transported to a distant site where they exit the vessels and move into the tissue. Thus, a different level of understanding of metastasis might be achieved using technologies that can follow these dynamic processes in vivo. To observe how circulating cancer cells and neutrophils interact on arrival in the lungs—a major site of metastatic colonization in breast cancer—we developed confocal intravital lung imaging (CILI), a modification of a lung-imaging approach used with two-photon microscopy. Using CILI, we observed NET-like structures around metastatic 4T1 cancer cells that had reached the lungs of mice. NETs were also found in the aggressive triple-negative human breast cancer subtype. In vitro, we found that metastatic, triple-negative breast cancer cells can induce metastasis-supporting NETs in the absence of infection, through secretion of G-CSF. The NETs, in turn, stimulated the invasion and migration of breast cancer cells. Inhibiting NET formation or digesting NETs with DNase I blocked these processes in vitro. Excitingly, treatment with NET-digesting, DNase I–coated nanoparticles significantly reduced lung metastases in mice. Our data suggest that induction of NETs by cancer cells is a previously unidentified metastasis-promoting tumor–host interaction and a new therapeutic target.

Cancer Cells Orchestrate Escape from the Adaptive Immune Response

J. Cappellani, M. Fein, X. He, L. Puckett

The immune system is very efficient at eliminating pathogens that can cause harm to the organism. The immune system also has the potential to eliminate neoplastic cells. The concept of "immune surveillance" was first described more than 50 years ago, and refers to the ability of immune cells to detect tumor cells and destroy them. T cells, part of the adaptive immune system, are critical for tumor immune surveillance. Immune surveillance may lead to a period in which cancer cells are kept in check by the immune system, and the tumor neither expands nor regresses. Eventually, tumors develop means to escape immune control. Mechanisms of escape include decreasing the surface expression of antigen-presenting major histocompatibility complex (MHC) I molecules. This makes the cells effectively invisible to the T cells. Another mechanism is increased expression of programmed cell death ligand 1 (PD-L1) on the cancer cells, which can bind the PD-1 receptor on activated T cells, resulting in protection against T-cell-mediated killing. Current immunotherapies that target this

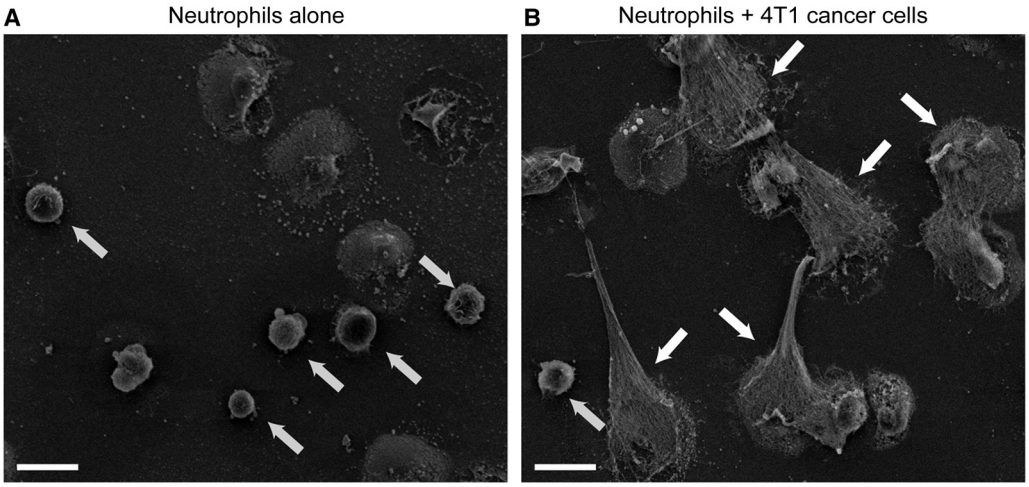


Figure 1. Breast cancer cells induce neutrophil extracellular traps (NETs). (*A*) Scanning electron microscopy of neutrophils three hours after plating and (*B*) coculturing with 4T1 breast cancer cells. Light gray arrows point to intact neutrophils. White arrows point to NETs. Scale bars, 10 µm.

so-called immune checkpoint have led to long-lasting regression in several cancers.

Chemokines, or chemotactic cytokines, have critical roles in mediating recruitment of immune cells to sites of inflammation and tumors. For example, the C-C chemokine ligand 2 (CCL2) recruits CC chemokine receptor (CCR2)-expressing immune cells to tumors. The chief role of CCR2 in cancer has therefore been considered to be the regulation of immune cell infiltration, such as the recruitment of monocytes to tumors after treatment with chemotherapy.

Breast cancer cells also express CCR2, but the effects of CCR2 expression and signaling in these cells are largely unknown—in part, because they were thought to be minor compared with the roles of CCR2 in host cells. We have now found that *Ccr2* in MMTV-PyMT breast cancer cells has a surprising role in inhibiting tumor immune surveillance. This effect is driven by multiple changes in the immune environment, including increased infiltration and activation of cytotoxic T cells and cross-presenting $CD103^+$ dendritic cells, as well as higher expression of MHC-I and lower expression of PD-L1 on cancer cells. These results establish a novel role for cancer cell CCR2 in suppressing an adaptive immune response, thereby promoting immune escape.

Our data suggest that CCR2, a chemokine receptor normally considered the regulator of monocyte infiltration into tumors, has an unexpected and critical role when expressed in cancer cells—by regulating an immune escape mechanism not previously reported. Our data, together with previous findings regarding the role of CCR2 in host cells in protecting tumors against chemotherapy, make CCR2 an attractive target in combination with both chemotherapy and immunotherapy.

Collagen Architecture in Pancreatic Cancer Progression

M. Shields

Tumors are often discovered as nodules that are harder than the surrounding tissue. This reflects changes in ECM stiffness and architecture. The interstitial ECM consists of macromolecules such as fibrillar collagens, fibronectin, and proteoglycans. Type I collagen is the major fibrillar collagen in many tissues and forms a scaffold that provides stability. Type I collagen also has signaling functions mediated by, for example, integrins and discoidin domain receptors. The synthesis and proteolytic remodeling of the fibrillar type I collagen increases in many tumors, most notably in pancreatic tumors.

Collagen cross-linking in mouse models of mammary carcinoma delays tumor onset and slows tumor progression. Furthermore, collagen architectural structure becomes abnormal with progression of breast cancer: Whereas collagen fibers are curly and oriented in parallel to normal or hyperplastic epithelium, there is a progressive change in the fibers so that they are straighter and mostly perpendicular to the tumor border in the late stages. This changed architecture promotes cell invasion by enabling cells to migrate along the collagen fibers or enhancing integrin signaling. Enzymes of the lysyl oxidase (LOX) family are important for collagen cross-linking. Cross-linking results in straighter fibers, and LOXs have been shown to promote cancer progression and metastasis of breast cancer.

Pancreatic cancer has higher levels of type I collagen than breast cancer, and we therefore hypothesized that collagen cross-linking would have a significant effect on reducing tumor initiation and metastatic spread. However, using both pharmacological inhibitors and genetic approaches, we instead found an increase in the growth of pancreatic tumors and metastatic spread. Intravital imaging in mice showed that pancreatic cancer cells also migrate along collagen fibers, and collagen architecture is altered by cross-linking inhibitors in pancreatic cancer. In addition, we have determined that intracellular signaling pathways, downstream from integrins, are up-regulated after inhibition of LOX in pancreatic cancer cells. This suggests that collagen architecture or the cancer cells' response to collagen is fundamentally different between breast and pancreatic cancer.

PUBLICATIONS

Arun G, Diermeier S, Akerman M, Chang K-C, Wilkinson JE, Hearn S, Kim Y, MacLeod, AR, Krainer A, Norton L, et al. 2016. Differentiation of mammary tumors and reduction in metastasis upon Malat1 LncRNA loss. *Genes Dev* **30:** 34–51.

Park J, Wysocki RW, Amoozgar Z, Maiorino L, Fein MR, Jorns J, Schott AF, Kinugasa-Katayama Y, Lee Y, Won NH, et al. 2016. Cancer cells trigger neutrophils to induce metastasis-supporting extracellular DNA traps. *Sci Transl Med* **8:** 361ra138.

RESOLVING THE CHEMOKINE/CHEMOKINE RECEPTOR CHALLENGE OF POORLY IMMUNOGENIC TUMORS

D.T. Fearon J. Li A. Pommier
M. Ludwig Z. Wang
N. Memos

The Fearon lab is conducting preclinical experiments that will lead to an understanding of why cancer-specific T cells do not control the growth of pancreatic ductal adenocarcinoma (PDAC). We are considering not only the factors that are restricted to the microenvironment of the tumor but also systemic factors that may alter immune reactions within the tumor. Our research makes only one assumption, which is that the immune systems of patients with PDAC have responded to the cancer. This premise is based on our finding that genetically modified mice with spontaneously arising PDAC, which is considered to mimic closely the human disease, have $CD8^+$ T cells that are specific for antigens expressed by cancer cells, but not normal cells, of the pancreas. These $CD8^+$ T cells are capable of controlling PDAC growth in these mice, at least temporarily, when systemic and intratumoral immune-suppressive reactions are blocked. Therefore, the essential immune reaction has occurred, which is the recognition by the host immune system of the tumor. Our aim is to discover the two general sets of signals that explain why immune recognition does not result in immune killing of the cancer cells: an absence of immune-activating signals that characterize infected tissue and the presence of immune-suppressing signals that characterize regenerating tissue.

Cachexia and Systemic Immune Suppression

A wasting condition termed cachexia affects a majority of patients with PDAC and is the proximate cause of death in at least 20% of patients. Clinically, patients lose their appetite, and the resulting decrease in caloric intake results in extreme loss of adipose tissue and skeletal muscle. Although cachexia has not been considered to be related to tumor immunology, we considered the possibility of a relationship because cachexia often occurs also with chronic infections, such as HIV and tuberculosis. In addition, although there is no "cachexia" in patients with acute infections, there is often a temporary loss of appetite that is coupled with elevated core body temperature, which induces a hypermetabolic state that aggravates the relative caloric insufficiency. Therefore, alterations in systemic metabolism may have been selected during evolution for their effect on immune reactions.

We studied two models in mice of cancer-induced cachexia: a transplanted colorectal tumor model and an autochthonous, spontaneously occurring PDAC. In both, we discovered that tumor-induced alterations of the host metabolic response to caloric deficiency caused intratumoral immune suppression. We found that the cytokine IL-6, which is produced by stromal cells in the tumors, reduced the hepatic ketogenic potential through suppression of PPARα, the transcriptional master regulator of this metabolic response. Ketogenesis is an essential metabolic compensatory response that provides energy to the brain in the form of ketone bodies that are generated from free fatty acids released from adipose tissues. Therefore, when these tumor-bearing mice were challenged with caloric deficiency, the resulting relative hypoketonemia triggered a marked rise in glucocorticoid levels, which caused further breakdown of skeletal muscle for the purpose of releasing amino acids that may be converted into glucose by the liver. We found by analysis of the transcriptomes of the tumors that multiple intratumoral immune pathways were suppressed by this hormonal stress response, but most striking was the suppressed expression of CXCL9, CXCL10, and CXCL11, three chemokines that attract the immune cells that mediate the killing of cancer cells. The potential clinical relevance of these findings was shown by the finding that even more modest elevations of corticosterone levels blocked the immunotherapeutic

response of mouse PDAC. Therefore, tumor-induced IL-6 impairs the ketogenic response to reduced caloric intake, resulting in a systemic metabolic stress response that blocks anticancer immunotherapy in the mouse. We are now collaborating with clinical investigators to assess the effects of cachexia on the outcome of immunotherapy in humans.

The Regulation of Chemokines and Chemokine Receptors in the Tumor Microenvironment

The immune system evolved to control or eliminate infections. Of the various types of infections, tumors most closely resemble viral infections in that the antigens being recognized by the adaptive immune system on cancer cells and virally infected cells are complexes comprised of major histocompatibility complex (MHC) class I membrane proteins and peptides that are derived from intracellular proteins. If the peptides are "foreign," the T-cell receptors of $CD8^+$ T cells will bind to MHC class I/peptide complexes and stimulate the T cells to kill the target cells.

To accomplish an effective immune reaction, the immune system has the unique challenge of transitioning from a dispersed group of cells that are present in secondary lymphoid tissues and the blood to an organized aggregation of immune cells at tissue sites of infection that is capable of orchestrating the killing of cells presenting foreign antigens. This transition is mediated by a system of chemokines that are produced at sites of infection and chemokine receptors on immune cells that mediate the migration and accumulation of these cells in the infected tissues. Effector $CD8^+$ T cells and other immune cells that have roles in host defense against viral infections, such as $CD4^+$ TH1 cells, natural killer (NK) cells, and dendritic cells (DCs), express several chemokine receptors, but CXCR3, which responds to CXCL9, CXCL10, and CXCL11, has a nonredundant role in attracting these cells to virally infected tissues. Perhaps not surprisingly, this chemokine/chemokine receptor system also mediates attraction of immune cells to tissues in which $CD8^+$ T-cell-mediated killing is the predominant immune reaction (such as allografts undergoing immune rejection), the autoimmune syndrome of type I diabetes, and epithelial tumors (the last example being the focus of the Fearon lab).

We have recently discovered that in an immunogenic colorectal cancer mouse model, there are high messenger RNA (mRNA) levels for CXCL9, CXCL10, and CXCL11, but relatively low expression of CXCR3. This observation has suggested to us that the immune cells expressing CXCR3 had not responded to the intratumoral presence of these chemokines. This possibility was supported by the relatively low frequency of $CD8^+$ T cells, $Batf3^+$ DCs, and NK cells in the tumor as compared with that of monocytes and macrophages, which use other chemokine receptors. Thus, we propose that one means by which immunogenic tumors may escape immune control is by impairing the function of CXCR3. Although we have also found that treatment of mice with drugs that inhibit CXCR4, another chemokine receptor on immune cells, leads both to enhanced intratumoral accumulation of $CD8^+$ T cells and immune control of tumor growth, we suggest that inhibition of CXCR3 is mediated by the interaction of CXCR4 with its chemokine ligand CXCL12. We are conducting experiments to assess this possibility, which could provide the mechanistic explanation for our earlier results showing that treating PDAC-bearing tumors with AMD3100, a small-molecular inhibitor of CXCR4, allowed the intratumoral accumulation of T cells and uncovered the sensitivity of the tumor to a T-cell checkpoint antagonist, antibody to PD-L1.

Finally, if the identification of CXCR3 as the chemotactic receptor that mediates the intratumoral accumulation of the immune cells that orchestrate tumor killing is correct, then we must find a means to induce the expression of the chemokines for this receptor. Although the levels of CXCL9, CXCL10, and CXCL11 may be sufficient in immunogenic tumors, these chemokines are not well expressed in poorly immunogenic tumors—so even if inhibition of CXCR3 is relieved by an inhibitor of CXCR4, the intratumoral immune reaction will possibly be ineffective. Therefore, the Fearon lab is determining the signals in immunogenic tumors that stimulate the production of these chemokines.

PUBLICATION

Flint TR, Janowitz T, Connell CM, Roberts EW, Denton AE, Coll AP, Jodrell DI, Fearon DT. 2016. Tumor-induced IL-6 reprograms host metabolism to suppress anti-tumor immunity. *Cell Metab* **24:** 672–684.

In Press

Fearon DT. 2017. Explaining the paucity of intratumoral T cells: A construction out of known entities. *Cold Spring Harb Symp Quant Biol* **81:** 219–226.

Flint TR, Fearon DT, Janowitz T. 2017. Connecting the metabolic and immune responses to cancer. *Trends Mol Med* **23:** 451–464.

Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, et al. 2017. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214:** 579–596.

MASS SPECTROMETRY LABORATORY

D. Pappin A. Makarenko N.S. Turna
K. Rivera J. Wilson
J. Snider

NRF2 Promotes Tumor Maintenance by Modulating Messenger RNA Translation in Pancreatic Cancer

K. Rivera, J. Wilson, D. Pappin [in collaboration with C. Chio, Y. Park, Y. Hao, E. Lee, B. Da Silva, C. Schoepfer, H. Tiriac, M. Hammell, D. Tuveson, CSHL; S.M. Jafarnejad, M. Sonenberg, McGill University; W. Palm, Memorial Sloan Kettering; V. Sangar, Institute of Systems Biology, Seattle, WA; J.E. Wilkinson, H.C. Crawford, University of Michigan; J.M. Buscaglia, Stony Brook University; G.M. DeNicola, Weill Cornell; E.E. Schmidt, Montana State University]

Pancreatic cancer is a deadly malignancy that lacks effective therapeutics. The CSHL Tuveson laboratory had previously reported that oncogenic Kras induced the redox master regulator Nfe2l2/Nrf2 to stimulate pancreatic and lung cancer initiation. In collaboration with the Pappin laboratory, they showed that NRF2 is necessary to maintain pancreatic cancer proliferation by regulating messenger RNA (mRNA) translation. Specifically, loss of NRF2 led to defects in autocrine epidermal growth factor receptor (EGFR) signaling and oxidation of particular translational regulatory proteins, resulting in impaired cap-dependent and -independent mRNA translation in pancreatic cancer cells. Combined targeting of the EGFR effector AKT and the glutathione antioxidant pathway mimicked Nrf2 ablation to potently inhibit pancreatic cancer ex vivo and in vivo, representing a promising synthetic lethal strategy for treating the disease.

Vitamin C Selectively Kills KRAS and BRAF Mutant Colorectal Cancer Cells by Targeting GAPDH

K. Rivera, D. Pappin [in collaboration with C. Chio, CSHL; J. Yun, E. Mullarky, C. Lu, K.N. Bosch, A. Kavalier, E.G. Giannopolou, O. Elemento, Z. Chen, L.E. Dow, S.S. Gross, L. Cantley, Weill Cornell; J. Roper, Tufts University; N. Papadopoulos, Howard Hughes, Johns Hopkins University; J.M. Asara, Harvard Medical School]

More than half of human colorectal cancers (CRCs) carry either KRAS or BRAF mutations and are often refractory to approved targeted therapies. In collaboration with the Cantley laboratory, we found that cultured human CRC cells harboring KRAS or BRAF mutations are selectively killed when exposed to high levels of vitamin C. This effect is a result of increased uptake of the oxidized form of vitamin C, dehydroascorbate (DHA), via the GLUT1 glucose transporter. Increased DHA uptake causes oxidative stress as intracellular DHA is reduced to vitamin C, depleting glutathione. Reactive oxygen species then accumulate and inactivate glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Inhibition of GAPDH in highly glycolytic KRAS or BRAF mutant cells leads to an energetic crisis and cell death not seen in KRAS and BRAF wild-type cells. High-dose vitamin C impairs tumor growth in *Apc/Kras*^{G12D} mutant mice. These results provide a mechanistic rationale for the therapeutic use of vitamin C for CRCs with KRAS or BRAF mutations.

Use of a Lys/Arg-N Protease for Mass Spectrometry Applications

J. Wilson, K. Rivera, N.S. Turna, D. Pappin

Directed in vacuo peptide fragmentation using mass spectrometry (MS) can generate specific ion fragment series that simplify MS/MS collision spectra. This is most readily achieved by controlling where charge is located on the peptide, either enzymatically or by chemical modification. To produce predominantly amino-terminal ions, charge must be placed at or near the amino termini of peptides. This can be elegantly achieved using a proteolytic enzyme with amino-terminal cleavage specificity at basic amino acids, and we have discovered and characterized a novel thermostable protease with amino-terminal arginine and lysine specificity. The enzyme can be used as a substitute for trypsin, generates predominantly amino-terminal ions, and completes digestions in 1–2 h at 60°C.

The enzyme shows a temperature optimum of ~60°C and broad pH specificity centered at pH 8.5, with at least 50% activity remaining at pH 5 and pH 11. The protease is most active in low ionic strength MS compatible acetate buffers and requires both calcium and zinc. Specificity for amino-terminal cleavage at arginine and lysine was ~95%, with the remaining ~5% amino-terminal to larger aliphatic residues. In whole-cell lysates, the numbers of identified peptides and proteins were similar to trypsin, yet generated in ~20× shorter digestion times. Eighty percent of the enzyme activity was retained in 0.5% SDS and specificity was equal for both lysine and arginine residues. As expected, b-ion series in MS/MS spectra were significantly more intense than y-ions, and relative b/y ion intensities were almost exactly reversed compared with trypsin. In complex mixtures, a shift to lower peptide charge states was also observed. This active metalloprotease is suitable for extremely rapid (1–2 h) digestion of samples for high-throughput MS analysis. The peptides produced by this enzyme always have major basic centers at the peptide amino termini, producing more easily interpretable MS/MS fragmentation spectra by generating a dominant b-ion series. The Pappin MS lab continues to use the enzyme routinely for characterization of posttranslational modifications such as phosphorylation, in which the dominant b-ion fragment series have enabled a significantly higher rate of detection.

Efficient Reassignment of a Frequent Serine Codon in Wild-Type *Escherichia coli*

K. Rivera, D. Pappin [in collaboration with J.M. Ho, M. Connolly, G.M. Church, Harvard Medical School; N.M. Reynolds, L.T. Guo, J. Ling, D. Soll, Yale University]

Expansion of the genetic code through engineering the translation machinery has greatly increased the chemical repertoire of the proteome. This has been accomplished mainly by read-through of UAG or UGA stop codons by the noncanonical aminoacyl-transfer RNA (tRNA) of choice. Although stop codon read-through involves competition with the translation release factors, sense codon reassignment entails competition with a large pool of endogenous tRNAs. In collaboration with the Pappin laboratory at CSHL, the Soll lab (Yale) used an engineered pyrrolysyl-tRNA synthetase to incorporate 3-iodo-l-phenylalanine (3-I-Phe) at a number of different serine and leucine codons in wild-type *E. coli*. Quantitative liquid chromatography (LC)-MS/MS measurements of amino acid incorporation performed in selected reaction monitoring experiments in the Pappin laboratory revealed that the 3-I-Phe abundance at the Ser208AGU codon in superfolder green fluorescent protein (GFP) could be as high as 65% ± 17%. This method also allowed quantification of other amino acids (serine, 33% ± 17%; phenylalanine, 1% ± 1%; threonine, 1% ± 1%) that compete with 3-I-Phe at both the aminoacylation and decoding steps of translation for incorporation at the same codon position. Reassignments of different serine (AGU, AGC, UCG) and leucine (CUG) codons with the matching tRNA(Pyl) anticodon variants were met with varying success, and the findings provide a guideline for the choice of sense codons to be reassigned. The results indicated that the 3-iodo-l-phenylalanyl-tRNA synthetase (IFRS)/tRNA(Pyl) pair could efficiently outcompete the cellular machinery to reassign select sense codons in wild-type *E. coli*.

PUBLICATIONS

Chio II, Jafarnejad SM, Ponz-Sarvise M, Park Y, Rivera K, Palm W, Wilson J, Sangar V, Hao Y, Öhlund D, et al. 2016. NRF2 promotes tumor maintenance by modulating mRNA translation in pancreatic cancer. *Cell* **166:** 963–976.

Herrera VL, Steffen M, Moran AM, Tan GA, Pasion KA, Rivera K, Pappin DJ, Ruiz-Opazo N. 2016. Confirmation of translatability and functionality certifies the dual endothelin1/VEGFsp receptor (DEspR) protein. *BMC Mol Biol* **17:** 15.

Ho JM, Reynolds NM, Rivera K, Connolly M, Guo LT, Ling J, Pappin DJ, Church GM, Söll D. 2016. Efficient reassignment of a frequent serine codon in wild-type *Escherichia coli*. *ACS Synth Biol* **5:** 163–171.

ROLE OF CROSS TALK BETWEEN INTERLEUKIN-6 AND TRANSFORMING GROWTH FACTOR β1 IN CHEMORESISTANCE

R. Sordella A. Gray M. Miller N. Pisterzi P. Safaric Tepes N. Shirole
N. Korimerla D. Pal N. Sachan S. Sentuk Z. Yao

Once considered as a uniform population of cells, it has become evident that a single tumor can be made up of thousands of cancer cells that look and behave differently from one another. This observed diversity arises from changes in the DNA sequence of particular genes, changes in the activity of genes, and the ability of cells to transition between different "states." As a result, individual cells within a tumor may react differently to certain anticancer drugs. Whereas most of the cells may be sensitive to the treatment and die, some might be resistant and survive. Hence, to design better cancer therapies we have to identify the driving forces behind tumor intrinsic heterogeneity. In the past year, we have identified three novel nongenetic molecular mechanisms that can contribute to intratumor diversity.

An Epigenetic Switch Regulates the Ontogeny of Axl +ve/EGFR-TKI-Resistant Cells by Modulating miR-335 Expression

Multiple mechanisms of acquired resistance to tyrosine kinase inhibitors (TKIs) have been described and validated in both experimental systems and patients' samples. These mechanisms include secondary/gatekeeper mutations in *EGFR* (T790M), c-Met amplifications, *PI3K* mutations, and the acquisition of mesenchymal and small-cell lung cancer features. More recently, the expression of Axl has also been reported as a novel mechanism of acquired resistance.

Axl is a *trans*-membrane receptor tyrosine kinase whose overexpression has been described in several human cancers. Axl possesses an extracellular domain with two amino-terminal immunoglobulin (Ig)-like domains and two fibronectin type III (FNIII) repeats that bind to the arrest-specific 6 (Gas6) ligand. The intracellular domain of Axl consists of a prototypical tyrosine kinase domain that modulates downstream effector signaling pathways in a selective manner. The binding of Axl to Gas6 upon its paracrine or autocrine secretion enables the *trans*-autophosphorylation of Axl intracellular tyrosine kinase domain and, consequently, the activation of multiple downstream signaling cascades. The Axl–Gas6 axis has been shown to promote processes implicated in tumorigenesis, including cell proliferation, cell survival, angiogenesis, invasion, and metastasis, as well as acquired resistance to multiple drug treatments.

In the context of non-small-cell lung cancer (NSCLC), higher levels of Axl and Gas6 have been observed in tumors that developed resistance to erlotinib. In these tumors, the targeting of Axl by chemical or genetic means restored erlotinib sensitivity. Alternatively, forced expression of an active Axl kinase in erlotinib-sensitive tumor cells was sufficient to induce erlotinib resistance.

Despite these documented findings, the molecular mechanisms leading to the ontogeny of Axl-positive cells remain poorly understood. Unlike other RTKs, no mutations or amplifications of Axl locus have been described in erlotinib-resistant cells.

We were able to show that Axl +ve cells are already present in erlotinib-naïve cells, and the expression of Axl is regulated through an epigenetic mechanism centered on the methylation of a specific CpG island present in the promoter of *MEST*, a gene contained in the second intron of the micro RNA miR-335. We found that *MEST* and miR-335 were consistently down-regulated in Axl +ve cells in both tumor-derived cell lines and primary human tumor samples. We observed that the forced down-regulation of miR-335 in Axl −ve cells was sufficient to induce phenotypic and molecular features that are characteristic of Axl +ve cells (e.g., epithelial to mesenchymal transition, decreased sensitivity to erlotinib).

Interestingly, miR-335 has already been shown to serve a critical role in the suppression of cancer reinitiation, progression, and metastasis, as well as in the regulation of drug resistance in breast, ovarian, and gastric cancer.

Consistent with a model of epigenetic/stochastic generation of Axl +ve cells, our data also suggest that the transition between Axl +ve and Axl –ve cells is highly plastic. This process is reminiscent of other cell-state transitions in which a stochastic/epigenetic mechanism has been proposed to explain the origin of different cell states.

In summary, our studies have defined a novel mechanism that couples epigenetic/stochastic inheritance to the ontogeny of the Axl +ve cell state. In principle, our proposed mechanism could explain the molecular basis of tumor cells' adaptability to erlotinib treatment.

Intracellular α-Ketoglutarate Is Required for p53 Separation-of-Functions Isoforms to Reprogram the Cells toward the Acquisition of Prometastatic Features

Close to half of all human cancers have mutations in the tumor suppressor TP53. Among the mutations characterizing TP53, truncating mutations are common in human tumors and, in general, are thought to give rise to p53-null alleles. Among the latter, *TP53* exon-6 truncating mutations are an exception as they have been shown to produce proteins that functionally and molecularly resemble the naturally occurring p53-psi isoform. p53-psi is generated by an alternative-splicing event that gives rise to a truncated p53 protein devoid of domains required for DNA binding, oligomerization, and nuclear localization. Collectively, p53-psi and the p53-psi-like exon-6 truncating mutations have been referred to as TP53 separation-of-function isoforms. Although lacking transcriptional activity and canonical TP53 tumor suppressor activity, p53-psi and TP53 exon-6 truncating mutations are capable of reprograming cancer cells toward the acquisition of prometastatic features. Consistently, *TP53* exon-6 truncating mutations occur at higher than expected frequency in tumors and are associated with poor prognosis and higher tumor metastatic spread. Differently from other TP53 alleles, TP53 separation-of-function isoforms rely on mitochondria localization and interaction with the mitochondria matrix protein cyclophilin D to elicit EMT, increased motility, extracellular matrix (ECM) invasion and metastatic colonization.

We have asked whether p53-psi and *TP53* exon-6 truncating mutations could modify the mitochondrial metabolism, and whether changes in certain metabolites could contribute to the cell-state reprograming induced by p53-psi-like isoforms. In fact, although mitochondria were initially thought to be exclusively involved in metabolite production and energy regulation, an accumulating body of evidence indicates that these cellular organelles are central to many cell-signaling networks. In addition to supplying much of the ATP, metabolites, and reducing molecules (e.g., NADH, FADH) essential for DNA replication and the maintenance of cellular activities, mitochondria produce metabolites such as succinic acid (SA) and α-ketoglutaric acid (aKG) that can modulate the activity of enzymes, signal transduction pathways, transcription factors, and epigenetic regulators.

We found that in cancer cells expressing p53 separation-of-function isoforms, high levels of intracellular aKG were required for the maintenance of epithelial-to-mesenchymal transition and the increased motility and invasion that characterized these cells. Consistent with a described role of αKG in the maintenance of cell identity via epigenetic modifications, we observed that the expression of p53 separation-of-functions isoforms regulates specific chromatin modifications such as decreasing histone H3K27me3 levels and ten-eleven translocation (Tet)-dependent DNA demethylation that could be reversed by treatment with the aKG inhibitor 2-HG. Consummately, our findings unearth a novel function of p53 separation-of-function isoforms in the regulation of mitochondrial metabolism and reveal that intracellular aKG levels can contribute to the reprogramming of cells toward the acquisition of prometastatic features.

Signal Integration of PDGFR and IL-6 Networks Boost the Metastatic Potential of Cancer Cells

Tumor-associated inflammation has been recently recognized as one of the hallmarks of cancers. In particular, inflammatory cytokines secreted in the tumor microenvironment have been shown to promote cancer pathogenesis and progression. Among them, interleukin 6 (IL-6) is increasingly expressed in many tumors and associated with poor prognosis. IL-6 can be produced and secreted by the same tumor cells or a variety of cells present in the tumor-microenvironment (TME)-like macrophages, neutrophils,

fibroblasts, and endothelial cells. On binding to its receptor, IL-6 triggers diverse signaling pathways that could result, depending on the biological context, in different outcomes such as cell proliferation, increased motility, resistance to proapoptotic stimuli, etc. The pleiotropic and highly contextual nature of IL-6 biological activity is not clearly understood. Intriguingly, we observed that platelet-derived growth factor receptor (PDGFR), different than other receptor tyrosine kinases such as epidermal growth factor receptor (EGFR), can phosphorylate the suppressor of cytokine signaling 3 (SOCS3), a key negative regulator of the IL-6 signaling pathway, at a specific tyrosine residue (Y165). As a consequence, SOCS3 is destabilized and degraded. Notably, we observed cells exposed to IL-6 and PDGF, in combination but not as single factors, or depletion of SOCS3 undergo dramatic changes in their morphology and behaviors and become highly metastatic. Altogether these findings suggest that PDGFR, by phosphorylating SOCS3, could modify IL-6 signaling and its biological effects. In principle, this could explain some of the contextual activity of IL-6. Importantly, our studies could potentially lead to the development of novel anticancer treatment strategies for the many tumors overexpressing PDGFR.

PROTEIN TYROSINE PHOSPHATASES AND THE CONTROL OF SIGNAL TRANSDUCTION

N.K. Tonks C. Bonham, G. Fan, C. Felice, N. Krishnan, K. Nicholas, I. Rus, O. Shrestha, D. Song, S. Zhang, W. Zheng

The reversible addition and removal of phosphate to proteins, which is termed protein phosphorylation, is the central feature of the mechanism of signal transduction—the process by which cells respond to stimuli in their environment. The activities of the enzymes that catalyze the addition (kinases) and removal (phosphatases) of phosphate groups are coordinated in signal transduction pathways to mediate the cellular response to environmental stimuli, and the functions of these enzymes are frequently disrupted in human diseases. The ability to modulate such signal transduction pathways selectively with drugs holds enormous therapeutic potential. In the area of tyrosine phosphorylation–dependent signal transduction, drug discovery efforts to date have emphasized the protein tyrosine kinases (PTKs); however, considering the reversibility of protein tyrosine phosphorylation, there is the potential to manipulate signal transduction pathways at the level of both PTKs and protein tyrosine phosphatases (PTPs). Although the PTPs have been garnering attention as potential therapeutic targets, they remain largely an untapped resource. The long-term objectives of the work of the Tonks lab are to characterize the structure, modes of regulation, and physiological function of members of the PTP family of enzymes. Through the application of basic research to understand the mechanism of PTP action, the Tonks lab is trying to devise creative new approaches to exploit these enzymes as targets for therapeutic intervention in major human diseases, including diabetes, obesity, and cancer.

During the last year, Ioana Rus completed her Ph.D. and has gone back to medical school at Stony Brook University to complete her M.D./Ph.D. qualification. Wei Zheng left the lab to pursue postdoctoral studies in Melbourne, Australia.

Defining the Signaling Function of PTP1B in HER2-Positive Cancer to Exploit Further Its Potential as a Therapeutic Target

PTP1B plays a positive role in promoting signaling events downstream from the oncoprotein tyrosine kinase HER2; consequently, small-molecule inhibitors of PTP1B may represent novel therapeutics for treatment of tumorigenesis and malignancy associated with elevated HER2. We showed that trodusquemine/MSI-1436 is a novel allosteric inhibitor of PTP1B, which attenuates HER2-dependent tumorigenesis and abrogates metastasis in the NDL2 mouse model of breast cancer. Currently, MSI-1436 is being tested in a Phase 1 clinical trial in HER2-positive breast cancer patients (ClinicalTrials.gov: NCT02524951), in collaboration with the Northwell Health Monter Cancer Center. Our current focus is on examining the effect of MSI-1436 on tyrosine phosphorylation in HER2-positive breast cancer cells, such as BT474, to define its mechanism of action. Within the field, PTPs are still frequently viewed as broad-specificity housekeeping enzymes; however, contrary to such a view, the effects of MSI-1436 on tyrosine phosphorylation are subtle. Various structural analyses conducted previously in the lab provided critical insights that defined the catalytic mechanism of PTPs, which led to our development of PTP "substrate trapping" mutants capable of forming stable complexes with target substrates in vivo. This technology, which may be applied to all PTPs, allows for the examination of PTP substrate specificity in various cellular contexts. We are now integrating this substrate-trapping technology with recently described proximity-dependent biotin identification (BioID) labeling strategies. Proximity-based labeling offers the ability to apply a stable "tag" to proteins in live cells, which are in the

vicinity of a defined protein of interest. Bait proteins are expressed as chimeric fusions with a promiscuous biotin ligase enzyme ($BirA^{R118G}$), which catalyzes the "activation" of exogenous biotin, leading to covalent biotinylation of nearby interacting proteins and thereby avoiding the need to ensure these transient and low-abundance complexes remain intact during cellular lysis and enrichment. By coupling with quantitative, high-throughput mass spectrometry–based proteomics and phosphoproteomics, we have established and validated a single workflow capable of identifying direct substrates of PTP1B and their respective sites of tyrosine dephosphorylation. As HER2-positive patients display either de novo or acquired resistance to Herceptin (trastuzumab), identification of alternative or combinatorial targets for therapeutic intervention is desperately needed. In our first initiative, we are directly implementing this "chemical-omics" platform into disease-relevant models of breast cancer to identify signaling axes influenced by PTP1B activity. We have developed functional PTP1B knockout breast cancer cell lines of acquired and de novo Herceptin resistance using CRISPR-CAS9 gene editing. By reintroducing CRISPR-CAS9-resistant PTP1B-BioID expression constructs, we are generating stable cell lines for examination in tissue culture and animal xenograft models to uncover mechanistic targets of PTP1B offering additional therapeutic avenues in a clinically relevant state of Herceptin resistance.

In an exciting development that epitomizes the advantages of an academic setting, with its freedom to think differently and pursue unique opportunities, we have identified an orally bioavailable analog of MSI-1436 that has taken us in an unanticipated direction. Our mechanistic studies have now revealed that this molecule binds copper with high affinity ($k_D = 5$ nM) and unique specificity. We have tested this compound in an animal model of Wilson's disease, which is caused by mutations in the copper transporter ATP7B and is associated with accumulation of copper in the liver, brain, and other tissues. We have shown that this compound chelated copper in liver and brain when delivered orally to the Wilson's disease mouse model, without any apparent detrimental effects on the animals. We are excited also about the potential cancer applications of this discovery. Recently, there have been reports concerning the application of copper chelation as a cancer therapy. In particular, the observation that copper exerts a stimulatory effect on methyl ethyl ketone (MEK) and is required for oncogenic BRAF signaling in melanoma has stimulated considerable interest in clinical trials with copper chelators such as tetrathiomolybdate. Our compound offers several advantages over tetrathiomolybdate, and we plan to test it broadly in cell and animal cancer models.

A New Role for PTP1B as a Target for Therapeutic Intervention in Rett Syndrome

We have continued to investigate the role of PTP1B as a regulator of tropomyosin receptor kinase B (TRKB) function and as a therapeutic target for Rett syndrome (RTT). This is an X-linked neurological disorder presenting with autistic features that are caused primarily by mutations in a transcriptional regulator, methyl CpG binding protein 2 (MECP2). Current treatment options for RTT are limited and focus on alleviating some of the neurological symptoms. Previously, we showed that the *PTPN1* gene, which encodes PTP1B, is a target of MECP2, and disruption of MECP2 function was associated with increased levels of PTP1B in RTT models. Pharmacological inhibition of PTP1B, focusing primarily on an active site-directed inhibitor CPT-157633, was shown to ameliorate the effects of MECP2 disruption in mouse models of RTT, including improved survival in young male ($Mecp2^{-/y}$) mice and improved behavior in female ($Mecp2^{-/+}$) heterozygous mice. Furthermore, we showed that PTP1B was a negative regulator of tyrosine phosphorylation of the tyrosine kinase TRKB, the receptor for brain-derived neurotrophic factor (BDNF). Consequently, the elevated levels of PTP1B that are generated by disruption of MECP2 function in RTT would be expected to represent a barrier to BDNF signaling. In fact, inhibition of PTP1B led to increased tyrosine phosphorylation of TRKB in the brain, which would augment BDNF signaling. Taken together this work presents PTP1B as a mechanism-based therapeutic target for the treatment of RTT, validating a novel strategy for treating the disease by modifying signal transduction pathways with small-molecule drugs. We have extended these analyses to include trodusquemine/MSI-1436, our novel allosteric inhibitor of PTP1B that is being tested in a Phase 1 clinical trial in HER2-positive breast cancer patients. Like CPT-157633, this structurally and mechanistically distinct inhibitor of

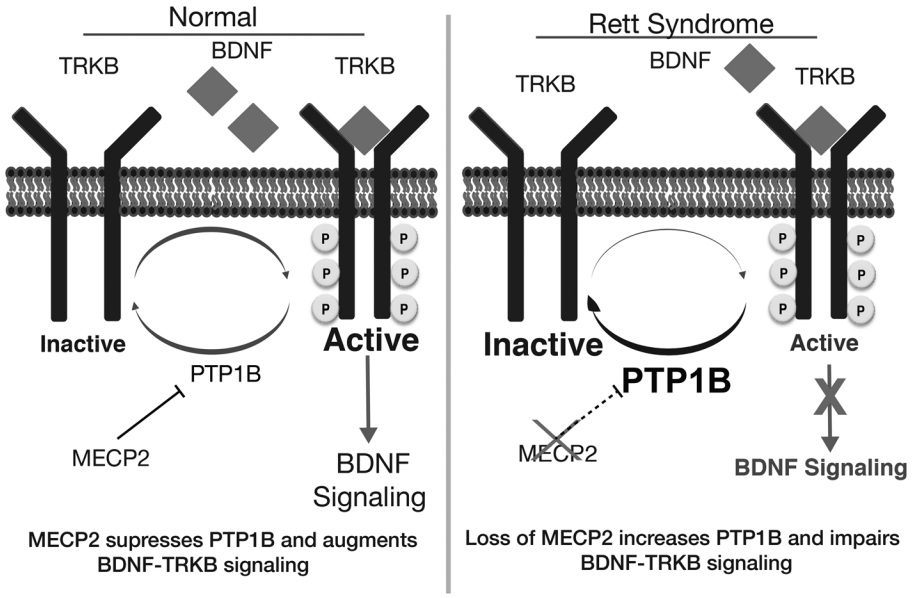


Figure 1. Model to illustrate methyl CpG binding protein 2 (MECP2)-mediated regulation of brain-derived neurotrophic factor (BDNF)-tropomyosin receptor kinase B (TRKB) signaling. MECP2 suppresses expression of PTP1B, which augments BDNF-induced signaling through the TRKB protein tyrosine kinase (PTK). Functional loss of MECP2 in Rett syndrome (RTT) results in an increase in levels of PTP1B, which attenuates BDNF-TRKB signaling. (Taken from Krishnan et al. 2015. *J Clin Invest* **125:** 3163.)

PTP1B ameliorates the effects of MECP2 loss in RTT mouse models. Now, we plan to look more broadly at which aspects of the RTT phenotype are affected by PTP1B, with the goal of defining end points that would be used in a clinical trial in RTT patients.

Redox Regulation of PTP Function

PTP1B is a major regulator of the signaling pathways initiated by insulin, which controls glucose uptake and metabolism, and leptin, which controls appetite. Gene-targeting studies showed that PTP1B-null mice are healthy, display enhanced insulin sensitivity, do not develop type 2 diabetes, and are resistant to obesity when fed with a high-fat diet. Furthermore, depletion of PTP1B expression with antisense oligonucleotides elicits antidiabetic and antiobesity effects in rodents, as well as human subjects. Another major project in the lab is to validate a novel approach to the development of inhibitors of PTP1B as drug candidates for the treatment of diabetes and obesity, which represent perhaps the greatest healthcare challenges facing the nation. Major programs in industry have focused on developing small-molecule inhibitors of PTP1B to promote insulin and leptin signaling in resistant states. Nevertheless, these efforts have been frustrated by technical challenges arising from the chemical properties of the PTP active site. In particular, although it is possible to generate potent, selective, and reversible active site-directed inhibitors, the tendency for such molecules to be highly charged, such as pTyr-substrate mimetics, presents problems with respect to their oral bioavailability and limits their drug development potential. As a result, industry views PTP1B and other PTPs as challenging. Consequently, innovative strategies are required to generate inhibitors of this highly validated target that may be readily exploited for drug development.

In this project, the goal is to harness a physiological mechanism for redox regulation of PTP1B function that reflects a new tier of control of tyrosine phosphorylation–dependent signaling. Previously, we observed that the activity of PTP1B is attenuated by reversible oxidation of an essential cysteinyl residue at the active site of the enzyme. The architecture of the PTP active site is such that this essential cysteinyl residue displays unique properties that favor its role as a nucleophile in catalysis but also render it prone to oxidation. Insulin stimulation of mammalian cells leads

to enhanced production of intracellular H_2O_2, which causes reversible oxidation of PTP1B and inhibition of its enzymatic activity, which, in turn, concomitantly promotes the signaling response to insulin. We have shown that mild oxidation of PTP1B, such as occurs in response to insulin, results in profound conformational changes in the active site of the enzyme that transiently inhibit substrate binding and catalysis. These structural changes, however, are reversible and the enzyme can be reduced back to its active state. Therefore, reversible oxidation of PTP1B in response to insulin provides a mechanism for fine-tuning the signaling response to the hormone. Previously, we used phage display to identify conformation-sensor antibodies that recognize the reversibly oxidized form of PTP1B (PTP1B-OX) selectively and stabilize this inactive state, inhibiting its reactivation by reducing agent and thereby inhibiting phosphatase activity. We showed that expression of these antibodies in cells enhanced insulin-induced signal transduction. These data provide proof of concept that stabilization of the oxidized, inactive form of PTP1B with appropriate therapeutic molecules may offer a novel paradigm for phosphatase drug development.

Over the past year, we completed studies to define the molecular basis for recognition of PTP1B-OX by the conformation sensor antibody scFv45. A mass spectrometry analysis of the ability of scFv45 to protect PTP1B-OX from proteases highlighted the importance of basic residues in a loop composed of residues 36–46. Similarly, investigation of the specificity of scFv45 for PTP1B-OX, compared with its closest relative TCPTP, showed that whereas scFv45 did not bind TCPTP-OX, it did bind TCPTP-mut2 (TC-mut2 F39L/E41K/R43K), which contains three residues substituted from PTP1B. Conversely, binding between scFv45 and PTP1B-mut2 (1B-mut2 L37F/K39E/K41R), which contains just three residues from TCPTP, was abrogated. Finally, we generated a crystal structure of the conformation-sensor antibody scFv45. Although we were able to purify complexes of scFv45 with PTP1B-OX, it was not possible to generate crystals, despite extensive efforts. Nevertheless, modeling of the complex revealed an important role for acidic residues in the complementarity-determining regions of scFv45 and these same basic residues in PTP1B at the binding interface.

To test the hypothesis that the effects of the PTP1B-OX-directed conformation-sensor antibodies could be mimicked by small molecules, we developed an assay that allowed us to conduct a pilot screen of the LOPAC library of compounds (Sigma-Aldrich). This revealed two such hits, one of which was sanguinarine. We screened additional analogs of sanguinarine, ultimately focusing on chelerythrine. As with scFv45, chelerythrine bound directly to PTP1B-OX, but not to PTP1B-mut2, and inhibited the reduction and reactivation of PTP1B-OX, but did not affect TCPTP. We have now shown that treatment of HEK cells with chelerythrine led to enhanced insulin signaling, and this effect was abrogated by expression of catalase to promote decomposition of H_2O_2. Furthermore, we have shown that treatment of high-fat diet (HFD)-fed C57Bl6/J mice with chelerythrine resulted in weight loss, whereas there was no effect of treatment with saline or protopine, an inactive analog of chelerythrine. We observed a 3% decrease in body weight in HFD-fed mice treated with chelerythrine, which plateaued after 14 days; however, there was no effect in mice fed a normal chow diet. Treatment with chelerythrine, but not saline or protopine, also improved glucose tolerance and insulin sensitivity, coincident with enhanced phosphorylation of the insulin receptor β-subunit and AKT in liver. In addition, it led to enhanced leptin signaling in the hypothalamus. Overall, these data are consistent with a mechanism in which chelerythrine, like scFv45, stabilizes PTP1B-OX to enhance insulin and leptin signaling. These data are currently under review for publication.

Efforts to Define How Loss of Expression of Specific PTPs That Regulate Cell Migration and Invasion In Vitro Contribute to Tumorigenesis In Vivo

The *PTPN23* gene is located on chromosome 3p21, a region that is spontaneously lost in 8%–10% of breast cancers, and low expression of PTPN23 coincides with poor survival. Our previous work identified PTPN23 as an important regulator of mammary epithelial cell migration and invasion. Now, in an orthotopic transplantation model, we observed that when PTPN23-deficient Comma 1Dβ cells were transplanted in the cleared mammary fat pad, suppression of PTPN23 induced severe primary tumor development in 52 weeks (18 out of 24) and a few cases (two out of 24) of lung metastasis. Immunohistochemistry analysis detected

extensive phosphorylation of Tyr 142 in β-catenin in tumor samples, but not in mammary glands repopulated with control cells or in normal mammary glands. We have shown that the primary effects of PTPN23 in this context are exerted through FYN, an SRC family kinase. Under normal conditions, PTPN23 dephosphorylates the autophosphorylation site in FYN, thereby suppressing its activity. Following suppression of PTPN23, FYN activity is elevated, leading to enhanced phosphorylation of Tyr 142 in β-catenin. More recently, we turned to tumor xenograft models because we could generate data in 8–10 weeks, compared with ~52 weeks in the transplantation model. We used RNAi to suppress PTPN23 in BT474 cells and tested the effects of AZD0530 (saracatinib), a small-molecule inhibitor of SRC family kinases, on tumor xenografts in SCID/Beige mice. Whereas suppression of PTPN23 accelerated tumor formation from BT474 xenografts, treatment with AZD0530 reversed the effect, coincident with inhibition of FYN activation. In a complementary study, we tested the effects of using CRISPR-CAS9-based gene targeting to suppress PTPN23 alone or together with FYN in CAL51 breast cancer cells. Using a xenograft model in SCID/Beige mice, we observed that suppression of FYN rescued tumor outgrowth because of loss of PTPN23 in CAL51 cells, consistent with the effects of AZD0530. These data provide mechanistic insights into the tumor suppressor function of PTPN23 and suggest that FYN may be a therapeutic target for tumors defined by LOH of PTPN23. A paper presenting this study is currently in preparation.

Use of CRISPR-CAS9-Based Gene Targeting to Define PTP Function in Acute Myeloid Leukemia

In a recent reorganization of the CSHL Cancer Center, a new Shared Resource focusing on "functional genomics" was created, under the direction of Chris Vakoc. One area of focus is CRISPR-CAS9-based gene targeting. Considering our previous systematic analysis of PTP function using RNAi, we decided to exploit this opportunity. Working in collaboration with the Shared Resource, we generated a pooled subgenomic RNA (sgRNA) library that targets 101 mouse PTPs and allows us to examine the function of members of the PTP gene family systematically in various models. Aberrant tyrosyl phosphorylation–dependent signaling is associated with cell proliferation, antiapoptosis, and drug resistance in leukemia progression. In this project, we decided to identify and characterize acute myeloid leukemia (AML)-dependent PTPs. Using the CRISPR-CAS9 gene-editing tool, we performed a negative screen by infecting an AML cell line, RN2, with a pooled sgRNA library that targets 101 mouse PTPs. By comparing the sgRNA abundance at the beginning and end of culture by MiSeq sequencing, our screen revealed a small set of PTPs, which are required for RN2 cell proliferation and survival, including PTPN23, DUSP12, and PTPMT1. In light of our existing interest in PTPN23, we have pursued further characterization of this enzyme.

Previous studies showed that PTPN23 is an ESCRT (endosomal sorting complexes required for transport)-associated protein. ESCRTs are multimeric protein complexes mediating a number of important physiological processes, including multivesicular body (MVB) formation, cytokinetic abscission, autophagy, membrane repair, and retroviral budding. The selectivity in control of various cellular events is regulated by assembly of different subsets of ESCRTs and specialized adaptor BRO1 domain-containing proteins—BRO1 in yeast and PTPN23 and ALIX in animals. PTPN23 and ALIX share a similar domain organization, except that PTPN23 has an additional carboxy-terminal catalytic domain. Our loss-of-function CRISPR-CAS9 screen revealed that PTPN23 was required for mouse AML cell proliferation and survival but dispensable for normal mouse bone marrow cells. Validation in three additional human AML cells showed similar dependency on PTPN23. We constructed a series of PTPN23 truncations, and found that the minimal segment of human PTPN23 that can rescue the mouse sgRNA knockout is a fragment comprising residues 1-872. The boundaries of this construct correspond to the full-length form of ALIX. Therefore, the current focus is to examine whether PTPN23 functions through ESCRT pathways in this context and to define the specificity between the effects of PTPN23 and ALIX. Each functional domain in PTPN23(1-872) is being substituted by the ALIX counterpart, and vice versa, and we are examining whether domain-swapped constructs can rescue the phenotype. In addition, we are testing known binding partners of PTPN23 for their contribution to its function in supporting AML cell proliferation and survival.

A New Role for the Non–Receptor Tyrosine Kinase FER in Potentiating Metastasis in Ovarian Cancer

We have been investigating metastasis of ovarian cancer cells, which disseminate readily within the peritoneal cavity, promoting metastasis, and are often resistant to chemotherapy. Ovarian cancer patients tend to present with advanced disease, which also limits treatment options; consequently, new therapies are required. Previously, we showed that in ovarian cancer cell lines protein content did not correlate with mRNA expression or epigenetic, DNA methylation data; however, quantitative proteomic analysis allowed cell line segregation based on carboplatin sensitivity, suggesting potential markers for therapy response and treatment outcome. In developing some of the ideas from this study, we investigated the oncoprotein tyrosine kinase MET, which is the receptor for hepatocyte growth factor (HGF), has been implicated in ovarian tumorigenesis and has been the subject of extensive drug development efforts. We showed a novel ligand- and autophosphorylation-independent activation of MET through the non–receptor tyrosine kinase FER. We showed that the levels of FER were elevated in ovarian cancer cell lines relative to those in immortalized normal surface epithelial cells, and that suppression of FER attenuated the motility and invasive properties of these cancer cells. Furthermore, loss of FER impaired the metastasis of ovarian cancer cells in vivo. Mechanistically, we showed that FER phosphorylated a signaling site in MET, Tyr 1349. This enhanced activation of RAC1/PAK1 and promoted a kinase-independent scaffolding function that led to recruitment and phosphorylation of GAB1 and the specific activation of the SHP2-ERK-signaling pathway. Overall, this analysis provides new insights into signaling events that underlie metastasis in ovarian cancer cells, consistent with a prometastatic role of FER and highlighting its potential as a novel therapeutic target for metastatic ovarian cancer.

PUBLICATIONS

Fan G, Zhang S, Gao Y, Greer A, Tonks NK. 2016. HGF-independent regulation of MET and GAB1 by non–receptor tyrosine kinase FER potentiates metastasis in ovarian cancer. *Genes Dev* **30:** 1542–1557.

Neel BG, Tonks NK, eds. 2016. *Protein tyrosine phosphatases in cancer*. Springer, New York.

UNDERSTANDING PTEN AND PROSTATE CANCER METASTASIS

L. Trotman A. Ambrico, I. Casanova-Salas, H. Cox, M. Lee, G. Matthew, D. Nowak, J. Vilaro, K. Watrud

Resistance to therapy of metastatic prostate cancer (CRPC) is responsible for the annual deaths of some 250,000 U.S. men. Although there is considerable progress in development of improved antihormone therapy for treatment of metastatic disease, this standard of care approach will invariably fail at some point.

Our focus is to understand the mechanisms driving human prostate cancer in its most lethal form: metastatic disease. We have studied the human genetics behind the transition from indolent to lethal CRPC and combined it with viral transgene delivery into prostate. With this approach, we have succeeded in generating a unique, fast, and faithful mouse model for CRPC. We have termed this system RapidCaP, as it allows us to generate any genetically mutant mouse prostate cancer with a much accelerated time frame compared with breeding-based approaches. Now, we use RapidCaP for analysis and therapy of metastatic disease.

At the same time, we aim to better understand how the phosphatase and tensin homolog (PTEN) tumor suppressor works. This has given us new insights into how cancer successfully attacks PTEN protein in early-stage prostate cancer before the *PTEN* gene is mutated.

The Nuclear Transport Receptor Importin-11 Is a Tumor Suppressor that Maintains PTEN Protein

D.G. Nowak, M. Chen, K. Watrud, A. Ambrico, T.M. Herzka, M.E. Zeeman, M. Minderer, W. Zheng, S.H. Ebbesen, C. Stahlhut, V.M. Wang, L. Wills, L.C. Trotman [in collaboration with J. Wilkinson, University of Michigan; B. Robinson, N. Navrula, A. Nasar, N.K. Altorki, V. Mittal, B.M. Stiles, Weill Cornell Medical College; M. Castillo-Martin, C. Cordon-Cardo, Mount Sinai School of Medicine; K.S. Plafker, S.M. Plafker, Oklahoma Medical Research Foundation; S. Powers, R. Sordella, CSHL]

Tumors evolve when cells gain function of proto-oncogenes and/or lose the function of tumor suppressor genes. PTEN is one of the most frequently deleted or down-regulated tumor suppressors that we know of. However, cancers do not always deactivate the two gene copies coding for the PTEN protein, instead the levels of PTEN protein in cancer can still be kept low as actual protein levels are critical for tumor suppression. The search for a recurrent cancer-associated gene alteration that causes PTEN protein degradation has, however, remained fruitless.

In our study, we showed that Importin-11 protein, encoded by the *IPO11* gene, traffics the tumor suppressor PTEN into the nucleus and helps to protect it from cytoplasmic proteins that cause PTEN degradation. PTEN levels are under tight control to execute growth restriction but still allow for survival. The ubiquitin (Ub) system has emerged as the prime regulator of PTEN levels. They need to be rapidly lowered upon insults that threaten cell survival, like ischemia or wounding. We recently found that the E2 Ub-conjugating enzyme UBE2E1, which is an IPO11 cargo, is a limiting factor for PTEN degradation. Using in vitro and in vivo gene-targeting methods, we show that IPO11 loss results in degradation of PTEN, lung adenocarcinoma, and neoplasia in mouse prostate with aberrantly high levels of UBE2E1 in the cytoplasm. These findings explain the correlation between loss of IPO11 and PTEN protein in human lung tumors. In addition to maintaining PTEN levels, IPO11 could be equally important in supporting nuclear-specific PTEN functions that were previously reported or remain to be discovered. It is not clear at present how ubiquitinated PTEN precisely interacts with Importin-11. The K13 and K289 sites, although residing in two different PTEN domains, are part of unstructured loops that could provide flexible, independently ubiquitinated interaction sites for the IPO11 transport receptor. This would be similar to the efficient interaction of importin-α with the classical NLS-peptide sequence, which often works equally well regardless of amino- or carboxy-terminal fusion to a transgene and independently of the transgene sequence.

Our work explains the nuclear accumulation of PTEN observed in many healthy tissues and, because IPO11 mutant mice develop prostate or lung tumors, also implicates Importin-11 as a novel tumor suppressor. Furthermore, we find that IPO11 status predicts disease recurrence and progression to metastasis in patients choosing radical prostatectomy. Thus, our data introduce the IPO11 gene as a tumor suppressor locus that is of special importance in cancers that still retain at least one intact PTEN allele. We conclude that Importin-11 function is crucial to PTEN stability. Thus, our work introduces a novel locus that could be an important target of the chromosome 5q deletions seen in many solid tumors beyond prostate and lung cancer.

Selective Killing of PTEN-Deficient Cancer Cells

G. Mathew, A. Naguib, K. Watrud, A. Ambrico, T. Herzka, W. Zheng [in collaboration with C.R. Reczek, N.S. Chandel, Northwestern Medical School]

Resistance to therapy of CRPC is responsible for the annual deaths of some 250,000 U.S. men. Although there is progress in development of improved drugs for treatment of metastatic disease, the vast majority of patients are in one type of treatment approach, antihormone therapy. This approach goes back to discoveries made in the 1940s and has improved the condition of most men with the disease. However, it is not a cure: Emergence of androgen resistance is invariably observed. We are interested in searching for alternatives that specifically target PTEN-deficient cancer cells.

The genes for p53 and *PTEN* genes are frequently inactivated in lethal prostate cancer. Their diminished activity provides a growth, proliferative, and antiapoptotic advantage. In mouse prostate cancer models, loss of p53 does not result in malignancy, yet codeletion of *Pten* and *Trp53* genes causes prostatic adenocarcinoma, illustrating the need to combat the cooperative power of these combined genetic lesions. To identify drugs specific to p53/Pten-deficient cells, we used an array platform to determine the most narrowly targeted approach that could selectively kill Pten/p53-deficient, but not Pten-normal, cells.

We generated *Pten* wild-type (wt) and *Pten* null cell lines from $Trp53^{fl/fl}$ or $Trp53^{l/fl}$, $Pten^{fl/fl}$ mouse embryonic fibroblasts (MEFs), and virally delivered Cre. The two cell lines were treated with 92 different chemotherapy drugs (phenotype microarray plates from Biolog). Deguelin and rotenone showed selective effects on the *Pten*-deficient, but not the wt, cells. We showed that these drugs kill by inhibiting mitochondrial complex I (MCP1) function, and we are now defining the mechanism by which the killing effect is specific to just the Pten null cells even if MCP1 is equally well inhibited in both cell types.

PUBLICATION

Labbé DP, Uetani N, Vinette V, Lessard L, Aubry I, Migon E, Sirois J, Haigh JJ, Bégin LR, Trotman LC, et al. 2016. PTP1B deficiency enables the ability of a high-fat diet to drive the invasive character of PTEN-deficient prostate cancers. *Cancer Res* **76**: 3130–3135.

CANCER MEDICINE LABORATORY

D. Tuveson

K. Aibel
B. Alagesan
L. Baker
L. Bertoncello
G. Biffi
R. Burkhart
D. Cheng
I.I. Chio
B. Da Silva
E. Elyada
D. Engle
M. Feigin
L. Guerriero
A. Handly-Santana
C.-I. Hwang
E. Lee
K. Miyabayashi
S. Moran
T. Oni
Y. Park
C. Schoepfer
P. Sridevi
H. Tiriac
C. Tonelli
S. Xia
M. Yao
G. Yordanov
C.M. Young
K. Yu

In 2016, pancreatic cancer overtook breast cancer to become the third leading cause of cancer-related deaths in the United States. Our laboratory studies the fundamental biology of pancreatic ductal adenocarcinoma (PDA) with the objective of designing novel therapeutic strategies to treat this malignancy. To enable these studies, in addition to genetically engineered PDA mouse models that we have generated over the past 20 years, we are using organoid models of human and mouse PDA that we published in 2015. These novel organoid model systems have unraveled previously uncharacterized facets of this deadly disease. For example, organoids were used to investigate the fundamental mechanisms used by PDA cells to detoxify reactive oxygen species (ROS), revealing new therapeutic opportunities. Additionally, studies focused on the tumor microenvironment by developing a new organoid-stromal coculture have identified distinct subpopulations of cancer-associated fibroblasts (CAFs). We are optimizing culture of human PDA organoids for pharmacotyping studies so that we can compare organoid drug response to patient therapeutic response in the clinic. In collaboration with Dr. David Spector (CSHL), Dr. Hans Clevers (Hubrecht Institute, the Netherlands), and Dr. Aldo Scarpa and Dr. Vincenzo Corbo (ARC-Net, Italy), we were recently awarded a National Institutes of Health (NIH) contract to develop hundreds of organoids from various human cancer tissues, as part of the Human Cancer Models Initiative (HCMI) consortium. Finally, the Preclinical Experimental Therapeutics-X Facility (PET^{x}) has started operation and will enhance the numerous preclinical studies that we have initiated.

Protein Translation as a Redox Dependency in Pancreatic Cancer

This work was done in collaboration with D. Pappin and M. Hammell, CSHL; N. Sonenberg, McGill University; Craig Thompson, Memorial Sloan Kettering Cancer Center; H. Crawford, University of Michigan, Ann Arbor; J. Buscaglia, Stony Brook University; and E. Schmidt, Montana State University.

ROS activate the Nrf2/Nfe212 transcription factor, which directs expression of the oxidative stress response genes. We previously reported that in pancreatic cancer, oncogenic Kras stimulates Nrf2 expression and thereby promotes pancreatic cancer initiation. By establishing a sensitive redox proteomics method with our organoid models, we identified translational regulatory proteins to be specifically oxidized when Nrf2 was deleted from pancreatic cancer cells. Both cap-dependent and -independent mRNA translations are impaired in Nrf2-deficient pancreatic cancer cells as a result of both the direct oxidation of the protein translation machinery and impairment of mitogenic signaling, and both antioxidants and exogenous mitogens can mitigate this defect. Nrf2 loss led to defects in autocrine epidermal growth factor receptor (EGFR) signaling. Using organoids and in vivo studies in mice, we showed that combined targeting of the EGFR effector pathways, such as AKT and the glutathione antioxidant pathway, mimicked Nrf2 depletion and potently inhibited pancreatic cancer, representing a promising synthetic lethal strategy for treating the disease. Expanding on this work, we are currently identifying tractable pharmacological methods that recapitulate Nrf2 loss in pancreatic cancer and will choose the most efficacious and least toxic approaches

in preclinical models that can be translated for an investigational clinical trial. This includes the PI3K inhibitor BLY719, PIM1/2 inhibitor AZD1208, Mnk1/2 inhibitor eFT508, and also an inhibitor of eIF4E/4G interaction (4EGI-1). Preliminary data showed heightened sensitivity of Nrf2-deficient tumor organoids to BLY719. Small molecule inhibitors with heightened activity in the Nrf2-deficient setting will be tested in combination with prooxidants for synergy. Prooxidants that are being tested include buthionine sulfoximine, auranofin, and high-dose vitamin C. The most active combinations of prooxidants and targeted agents will then be used in vivo for preclinical evaluation. Besides implicating mRNA translation as an important facet of PDA biology, it broadly implicates regulation of this process in other tissues when redox bursts occur, potentially by inducing a state of cell pausing such that bioenergetics can be applied to repair the damage.

Fibroblast Heterogeneity in Pancreatic Cancer

This work was done in collaboration with M. Egeblad, D. Fearon, and J. Crawford, Hofstra Northwell School of Medicine; H. Clevers, Hubrecht Institute; R. Hynes, Massachusetts Institute of Technology; and P. Robson, Jackson Laboratory.

PDA is histopathologically distinct from other cancers because of its high content of nonneoplastic tissue. The dense desmoplastic stroma of PDA contains acellular extracellular matrix (ECM) proteins and glycosaminoglycans and a large number of nonneoplastic cell types, including immune cells and fibroblasts. CAFs in PDA may be derived from several sources, and the resident pancreatic stellate cells (PSCs) are one major source. PSCs store vitamin A–containing lipid droplets in normal pancreatic tissues and respond to inflammation and neoplasia by proliferating and differentiating into myofibroblasts that synthesize ECM components. To investigate whether there are different types of CAFs in pancreatic cancer, we evaluated the spatial distribution of α-smooth muscle actin (α-SMA), a hallmark of myofibroblasts, in human pancreatic tumors and in KPC (*Kras*$^{LSL\text{-}G12D/+}$; *Trp53*$^{LSL–R172H/+}$; *Pdx-1-Cre*) mouse tumors. We identified a distinct CAF subpopulation of myofibroblasts, which we have termed "myCAFs," because of their marked expression of α-SMA. myCAFs are located immediately adjacent to neoplastic cells in both mouse and human PDA tissue. Using a novel coculture system involving CAFs and organoids embedded in Matrigel, we showed that organoid-activated CAFs produce a desmoplastic stroma, recapitulating the nature of PDA tumors. These cocultures importantly revealed cooperative interactions between the PSCs and organoids, with each cell type supporting the growth of the other. Surprisingly, we also identified another subpopulation of CAFs that were located more distantly from neoplastic cells. This other subpopulation of CAFs lacked elevated α-SMA expression and, instead, secreted IL6 and additional inflammatory mediators, and therefore we have termed them inflammatory CAFs or "iCAFs." Our in vivo and organoid-CAF cocultures show a spatial separation between iCAFs and myCAFs, suggesting that direct juxtacrine interactions with cancer cells are required for myCAFs but preclude the development of iCAFs. Using RNA sequencing analysis, we compared the transcriptomes of quiescent PSCs (PSCs embedded alone in Matrigel), iCAFs (SMA^{low}; IL6^{high} PSCs), and myCAFs (SMA^{high}; IL6^{low} PSCs), and found clusters of genes uniquely up-regulated in either myCAFs or iCAFs. Based on these encouraging results, we are further analyzing CAF heterogeneity in vivo by single-cell RNA sequencing of CAFs present in both KPC and human PDAC samples in a collaborative project with Dr. Paul Robson at Jackson Laboratories. Our aim is to find populations of CAFs that are characterized by unique gene expression patterns and identify their roles within the tumor microenvironment. We are also investigating the functional differences and the mechanisms of formation of myCAFs and iCAFs by exploiting our in vitro coculture platform and performing in vivo orthotopic coculture transplantations in mice. Finally, we are continuing our studies characterizing the nature of the PDA stroma in collaboration with Richard Hynes at MIT. Our cross-species comparison has identified previously known matricellular proteins such as collagens and fibronectin and has also revealed novel proteins that play roles in PDA genesis.

Human Organoid Platform for Development of PDA Therapeutics

This work was done in collaboration with M. Wigler, A. Krasnitz, and J. Crawford, Hofstra Northwell School of Medicine.

In collaboration with several hospitals, including Northwell Health, Thomas Jefferson University,

Figure 1. Heatmap of therapeutic response using normal, early, and late disease human organoids screened against the NCI oncological small molecule library. Organoids are dosed 24 h postplating with 2 μM of each compound. Hierarchical clustering of organoids and compounds is based on Euclidean distance. Percent viability values are the average of two replicates.

MSKCC, Stony Brook University, Johns Hopkins Medical Institute, and Winthrop Hospital, we have obtained more than 100 human pancreatic tumor tissues and have successfully established more than 70 organoids from pancreatic tumor, metastases, and biopsy samples. Our laboratory recently developed a robust therapeutic assay to characterize drug responses of human organoids. Using this platform, for each organoid, we will initially establish the proliferation rate and dose response to standard-of-care chemotherapeutic agents. The sensitivity profiles observed in our initial studies are robust and can be reproduced at various organoid passages. In a preliminary study of standard-of-care chemotherapeutics for PDAC, we observed patient-specific sensitivity profiles. Importantly, both single agents and combinations of two or more compounds can be tested in this platform, enabling the ex vivo modeling of current PDAC therapeutic regimens. In addition to standard-of-care compounds, we have established an organoid-based high-throughput drug screening methodology using the National Cancer Institute (NCI) library of oncological small molecules (129 chemotherapeutic and targeted agents) (Fig. 1). For this project, we are currently expanding the size of our drug library to include promising investigational compounds, as well as pathway-specific inhibitors, which we will test in at least 100 different human organoid lines. An important objective of this study is to identify therapeutic responses in organoids that are predictive of a patient's sensitivity or resistance to therapy. Successful completion of our initial retrospective trial will establish the organoids as a platform that provides predictive therapeutic and molecular information to a physician within a clinically relevant time frame.

Therapeutic Development for Pancreatic Cancer

At the PETx facility at Woodbury, we have started rederiving mice to populate the animal rooms of this facility with genetically engineered mouse models (GEMMs). We are continuing our in vivo therapeutic studies using both the KPC mice and newly developed organoid-based orthotopic models of PDA. In one such study, we investigated the combination of a MEK inhibitor (AZD6244) with an ERBB inhibitor (neratinib, an irreversible ERBB receptor tyrosine kinase inhibitor). Although no significant activity was observed on treatment with single agents, combined treatment with these drugs resulted in decreased tumor volumes over a short time course. Accordingly, combinatorial drug approaches using available and investigational therapies will be applied in pilot studies to different genetic models of PDA in efforts to identify new therapeutic approaches for this disease.

PUBLICATIONS

Chio II, Jafarnejad SM, Ponz-Sarvise M, Park Y, Rivera K, Palm W, Wilson J, Sangar V, Hao Y, Öhlund D, et al. 2016. NRF2 promotes tumor maintenance by modulating mRNA translation in pancreatic cancer. *Cell* **166:** 963–976.

In Press

Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, et al. 2017. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med* **214:** 579–596.

Ras AND Rho GTPases AND THE CONTROL OF SIGNAL TRANSDUCTION

L. Van Aelst C. Fitzpatrick S. Nakamuta
N.B. Gallo Y. Tai
A. Mirko M. Wang

Research in our laboratory is focused on the role of Ras and Rho GTPase family members in signal transduction. Ras and Rho family members play key roles in cellular activities controlling cell growth, differentiation, and morphogenesis. Alterations in Ras and Rho functions have been causally linked to both cancer and brain/mental disorders, the latter including intellectual disability (ID), schizophrenia, epilepsy, and mood disorders. Our interests lie in understanding how defects in Ras- and Rho-linked proteins contribute to the development of these diseases. Toward this end, our lab has continued to define the functions of selected GTPases, their regulators, and effectors, using animal models of cancer and neurodevelopmental/neurological disorders. Below, our key projects are highlighted.

Multifunctional Role of the X-Linked Intellectual Disability Protein Oligophrenin-1 in Cognition and Behavior

Oligophrenin-1 (*OPHN1*), which encodes a Rho-GTPase-activating protein, was the first identified Rho-linked ID gene. It was initially identified by the analysis of a balanced translocation t(X;12) observed in a female patient with mild ID. Subsequent studies revealed the presence of *OPHN1* mutations in families with a syndromic form of ID, with affected individuals commonly showing mild to severe ID and behavioral problems. Most of the *OPHN1* mutations identified to date were shown, or predicted, to result in loss of OPHN1 function; however, the pathophysiological role of *OPHN1* has remained poorly defined. To gain insight into how mutations in *OPHN1* could lead to the cognitive deficits, we initially focused on defining OPHN1's role in hippocampal plasticity associated with learning and memory. Our studies unveiled multiple roles for OPHN1 at hippocampal CA1 synapses. In addition to an essential role in controlling activity-driven glutamatergic synapse development, we found that temporal regulation of OPHN1 translation plays a critical role in mGluR-dependent long-term depression (LTD), a form of plasticity linked to drug addiction and cognitive disorders that is dependent on rapid new protein synthesis. We subsequently showed that OPHN1's involvement in mGluR-LTD and its ability to control glutamatergic synapse development are independent of each other. Whereas the latter requires OPHN1's Rho-GAP activity and association with Homer 1b/c proteins, the former is dependent on OPHN1's interaction with Endophilin A2/3. Together, our findings provided first insights into how mutations in *OPHN1* could contribute to the cognitive deficits in individuals with *OPHN1* mutations.

Apart from the above-described role for OPHN1 in hippocampal plasticity and learning, we recently discovered that it also plays a critical role in moderating stress-induced depressive-like behaviors in the learned helplessness (LH) model of depression. In this model, animals are exposed to an unpredictable and uncontrollable stressor and, subsequently, evaluated for their coping capabilities (i.e., resilient vs. helpless/depression-like behavior) in a testing session that consists of an aversive but escapable stressor. We found that *Ophn1*-deficient mice ($Ophn1^{flox//Y}$;β-$actin^{Cre-/+}$) subjected to the LH procedure show a marked increase in helpless behavior compared with control littermates. Importantly, *Ophn1*-deficient mice did not display altered motor activity and were capable of learning and performing the avoidance task. Upon further examination of the neuronal cell types involved, we found that *Ophn1* deficiency in parvalbumin (PV), but not somatostatin (SOM), interneurons promotes LH. Moreover, and importantly, we found that *Ophn1* depletion in PV interneurons exclusively in the prelimbic area of the medial prefrontal cortex (mPFC) is sufficient to induce helpless behavior. Together, these

findings uncovered the importance of the X-linked intellectual disability (XLID) protein OPHN1 in the establishment of resilience to stress and revealed its involvement in mPFC PV interneuron function.

DOCK7 Drives Tangential Migration of Interneuron Precursors in the Postnatal Forebrain

Throughout life, new interneurons are added to the olfactory bulb (OB), enabling OB neural circuits to continuously adapt. This process critically relies on the proper migration of ventricular-subventricular zone (V-SVZ)-derived neuroblast precursors, which migrate along a long path, known as the rostral migratory stream (RMS), to the OB. Although progress has been made toward identifying extracellular factors and matrix components that guide the migration of these cells, still little is known about the intracellular mechanisms that govern the dynamic reshaping of the neuroblasts' morphology required for their migration along the RMS. In particular, how tangentially migrating V-SVZ-derived neuroblasts control and coordinate leading process extension and nucleokinesis to accomplish efficient migration in the RMS remains largely unknown.

We showed that DOCK7, a member of the DOCK180 family of atypical Rho-GEFs, is essential for the tangential migration of neuroblasts in the postnatal mouse forebrain. Interestingly, using a molecular replacement strategy combined with live cell imaging, we uncovered that DOCK7 regulates the migration of these cells by controlling both leading process extension and somal translocation. Moreover, we found that DOCK7 controls these two processes by acting on distinct signaling pathways. Namely, DOCK7 governs leading process stability/growth via a Rac-dependent pathway, likely by modulating microtubule networks. In parallel, DOCK7 also regulates F-actin remodeling at the cell rear to promote somal translocation via a previously unrecognized myosin phosphatase-RhoA-interacting protein-dependent pathway. Thus, our study identifies DOCK7 as a key effector that regulates and coordinates two distinct phases/steps of neuroblast migration. Thereby, it not only offers new insight into DOCK7 function, but it also enhances our understanding of the mechanisms that govern the tangential migration of V-SVZ neuroblasts in the postnatal forebrain.

TGF-β/Smad Signaling through DOCK4 Facilitates Lung Adenocarcinoma Metastasis

The cytokine TGF-β plays an important, and complex, role in epithelial tumorigenesis. During early stages of tumorigenesis, TGF-β typically functions as a tumor suppressor. At later stages, however, it can act as a potent promoter of multiple events driving the metastatic process, which comprises local motility/invasion, entry of cancer cells into the blood stream (intravasation), exit from the blood vessels (extravasation), and colonization of distant organs. The relevance of TGF-β signaling for disease progression has been particularly recognized in tumors in which cancer cells retain the core TGF-β signaling components, as is frequently the case in breast and lung cancers. However, a major remaining challenge is the identification of TGF-β target genes that drive specific events during metastasis, especially because TGF-β modulates gene expression in a highly cell- and context-specific manner. Although some progress has been made in the context of breast cancer metastasis, the genes and mechanisms that mediate the prometastatic effects of TGF-β in lung adenocarcinoma (ADC) remain largely unknown.

To identify molecular mechanisms that mediate the prometastatic effects of TGF-β in lung ADC, we took a candidate gene approach and started by scrutinizing members of the DOCK180-related protein superfamily, which, as mentioned above, emerged as a distinct class of Rac and/or Cdc42 GEFs. We found that in lung ADC cells, expression of DOCK4, but not other DOCK180-family members, is rapidly and robustly induced by TGF-β in a Smad-dependent manner. Subsequently, we showed that DOCK4 is a direct TGF-β/Smad target gene and, importantly, high DOCK4 expression correlates with activated TGF-β signaling and poor prognosis in human lung ADC. These findings prompted us to assess the role of DOCK4 in mediating the prometastatic effects of TGF-β in lung ADC in vivo in a mouse model. Remarkably, we found that DOCK4 induction is essential for TGF-β-driven lung ADC metastasis. Specifically, we observed that blockade of TGF-β-mediated DOCK4 induction attenuates the ability of lung

ADC cells to extravasate into distant organ sites, resulting in a marked reduction in metastatic burden. At a cellular level, our evidence supports a model in which TGF-β-induced DOCK4 facilitates extravasation by stimulating lung ADC cell protrusive activity, motility, and invasion, without promoting epithelial-to-mesenchymal transition (EMT), and, intriguingly, that it does so by driving Rac1 activation. So far, Rac1 has been only linked to TGF-β via a noncanonical pathway. Thus, our findings identified the atypical Rac1 activator DOCK4 as a novel key component of the TGF-β/Smad pathway that promotes lung ADC cell extravasation and metastasis.

Recently, we expanded this line of research toward identifying genes that mediate lung ADC organ-specific metastases, with a particular focus on genes that mediate colonization of specific organs. To this end, we established a multiple organ metastasis model system using mice injected intracardially with lung ADC cell populations derived from primary tumors arising in $Kras^{G12D}/p53^{-/-}$ mice. With this model system, we initiated an in vivo RNA interference (RNAi) screen designed to assess the possible involvement of a select set of genes in modifying the potential of lung ADC cells to metastasize to specific host organs. These studies were performed in collaboration with Kenneth Chang. Interestingly, we found that silencing of this gene set resulted in a significant decrease in occurrence of brain metastases and an increase in the occurrence of bone metastases. These data suggest that one or more genes in this set modulate(s) lung ADC metastatic cell homing to and/or colonization of the brain and bone. We are currently investigating which of these genes, individually or in combination, modulate organ-specific homing and/or colonization of lung ADC cells in the brain and bone. These studies have the potential to identify novel mediator(s) and suppressor(s) of lung ADC metastases to brain and bone.

In Press

Nakamuta S, Yang Y-T, Wang C-L, Gallo NB, Yu J-R, Tai Y, Van Aelst L. 2017. Dual role for DOCK7 in driving tangential migration of interneuron precursors in the postnatal forebrain. *J Cell Biol* (in press).

Van Aelst L, Downward J, McCormick F, eds. 2017. *Ras and cancer in the 21st century*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (in press).

UNDERSTANDING MALIGNANT GLIOMA PATHOGENESIS

H. Zheng C. Wu L. Xu Q. Zhang

Malignant glioma is the most common and lethal type of brain tumor and is resistant to all currently available therapies. In its most aggressive form, glioblastoma (GBM) has a median survival of 12–15 months after initial diagnosis. Our group applies an integrated approach combining in vivo animal model systems with in vitro biochemical and cellular assays to investigate the genetic and epigenetic programs governing pathogenesis of malignant glioma. Our long-term goal is to uncover the molecular and developmental programs behind the complex biology of glioma initiation and progression so as to enable the medical community to develop mechanism-based therapeutic strategies and targets against this lethal disease. During the past year, we have continued to focus our research around two major areas: (1) developing genetically engineered animal model systems to recapitulate the genetic and pathological aspects of various human glioma subtypes and using these glioma animal models as tools to dissect in vivo tumor propagation and their response to experimental therapeutic treatments and (2) identifying the genetic and epigenetic pathways governing normal and neoplastic neural stem and glioma cell fate determination. Below we highlight some of our progress.

Characterizing ATRX Functions during Neural Development and Glioma Pathogenesis

Alpha thalassemia/mental retardation syndrome X-linked (ATRX) is a chromatin remodeling factor of the Snf2 family, and its germline mutation causes an X-linked developmental disorder featuring mental retardation, developmental delay, and α-thalassemia. Frequent loss-of-function mutations of *ATRX* are identified in a variety of cancer types, including pediatric and adult gliomas, pancreatic neuro-endocrine tumor, neuroblastoma, and osteosarcoma. More recently, mutations of *ATRX* in human tumors have been found to be mutually exclusive with *TERT* promoter mutations, but correlate with activation of a telomerase-independent telomere maintenance mechanism termed alternative lengthening of telomeres (ALT). ATRX is known to form a histone chaperone complex with its binding partner DAXX and facilitate replication-independent deposition of histone variant H3.3 at heterochromatic regions. But how ATRX exerts its tumor-suppressing and telomere maintenance functions remains unclear.

In collaborations with the Lowe and Vakoc groups, we previously initiated a project to probe chromatin remodelers controlling neural stem cell (NSC) and/or glioma cell self-renewal and differentiation. By conducting a screen of a short hairpin RNA (shRNA) library targeting epigenetic regulators, we identified ATRX, whose depletion promotes mouse glioma cell growth specifically under differentiation induction conditions. More recently, using CRISPR-Cas9 or shRNA-mediated approaches, we further revealed that depletion of ATRX expression in IDH1^{R132H}-transduced *p53*$^{-/-}$ or *p53*$^{-/-}$ *Pten*$^{-/-}$ mouse NSCs strongly promotes in vivo glioma formation, showing that ATRX is a bona fide glioma tumor suppressor. Mechanistically, our analysis of NSC lineage differentiation and RNA-Seq data indicates that ATRX plays a crucial role in controlling neural development. Depletion of ATRX in mouse NSCs strongly blocks neuronal lineage maturation but has only limited impact on NSC proliferation or differentiation of other neural cell lineages such as astrocytes and oligodendrocytes. This is consistent with our observation that Atrx is highly expressed in neuronal cells compared with other cell lineages in adult mouse brains. The findings that ATRX expression level increases along the neuronal differentiation path and with strongest expression confined to postmitotic neuron cells support the idea that ATRX is required for proper induction and/or maintenance of the neuronal differentiation state during central nervous system (CNS) development.

Dynamic nucleosome deposition, eviction, and remodeling are essential for epigenome and high-order chromatin structure maintenance. As canonical H3.1 and H3.2 histones are only synthesized during the

S phase of the cell cycle, the mammalian cells use the histone variant H3.3 for replication-independent nucleosome assembly during the differentiation process. As a result, adult tissues are highly enriched for histone H3.3. Because histone H3 is the major substrate of chromatin modification, we reason that ATRX-mediated H3.3 deposition may play an essential role to ensure proper neuronal lineage progression and maturation during differentiation-mediated chromatin reorganization. We are currently conducting chromatin analysis to probe epigenetic functions of ATRX during neuronal differentiation.

ATRX/DAXX Loss Induces Activation of Alternative Lengthening of Telomeres Independent of Telomere Length

Telomere maintenance is essential for cell immortalization and tumorigenesis. Although the majority of human malignancies rely on reactivation of telomerase, 10%–15% of them use recombination-directed ALT to protect chromosomal ends. Despite accumulated evidence that has linked ALT with loss of histone variant H3.3 chaperone components ATRX and DAXX, their roles in telomere maintenance and tumor progression remain unclear.

In collaboration with the Dr. Lieberman's group at Wistar Institute, we recently showed that depletion of ATRX or DAXX in human NSCs and glioma cells, but not in mouse NSCs and glioma cells, disrupts normal telomeric DNA replication and induces ALT, irrespective of preexisting telomerase activity and telomere length. We further revealed that the ALT activation is contingent on dysfunctional telomeric H3.3 loading. Reconstituted wild-type *DAXX*, but not mutants defective of either H3.3 or ATRX interaction, abolishes ALT in *DAXX*-deleted cells, linking ALT activation to defective assembly of telomeric nucleosomes displaced outside of S phase. Along with ALT, we show that *ATRX* or *DAXX* deletion also elicits persistent telomeric DNA damage response (DDR) that negatively impacts cell growth. Mitigation of telomeric DDR by ectopic *hTERT* overexpression restores proliferative capacity of *ATRX*- or *DAXX*-deleted cells without abolishing ALT induction. Our studies therefore establish ALT as a telomere repair mechanism activated in response to disrupted telomere replication.

These findings indicate that the ATRX/DAXX chaperone complex plays an essential role in ensuring proper S-phase telomeric DNA replication by assisting replication-independent nucleosome assembly. In the absence of ATRX/DAXX-mediated H3.3 loading, evicted nucleosomes outside of S phase cannot be reassembled. The resulted unpackaged telomeric sequences, because of their GC-rich and repetitive natures, form higher-order secondary structures that impede fork progression during telomere replication. The accumulation of telomere replicative stress then further elicits recombination-based ALT to repair the disrupted telomere replication. Meanwhile, the severely stalled or collapsed replication forks induce DDR that consequentially suppresses cell growth. Our model suggests that persistent telomeric DDR caused by *ATRX* or *DAXX* loss would act as a checkpoint control limiting the progression of early-stage *ATRX*-mutant tumors. Indeed, analysis of human patient specimens indicated that *ATRX*-mutant low-grade gliomas constitute a significantly higher percentage of cells positive for DNA damage marker γH2AX as compared with *ATRX*-mutant high-grade or wild-type gliomas. Moreover, the observation that the ATRX loss-induced DDR is markedly lowered in high-grade *ATRX* mutant tumors suggests that escaping from the checkpoint control is essential for *ATRX* mutant tumor progression. Identification of the mechanism(s) behind the damage response checkpoint control therefore may have important clinical implication for future therapeutic development targeting ALT tumors.

In Press

Singh DK, Kollipara RK, Vemireddy V, Yang XL, Sun Y, Regmi N, Klingler S, Hatanpaa KJ, Raisanen J, Cho SK, et al. 2017. Oncogenes activate an autonomous transcriptional regulatory circuit that drives glioblastoma. *Cell Rep* **18:** 961–976.

NEUROSCIENCE

How does the brain encode stimuli from the outside world, within and across sensory modalities, to generate specific perceptions that trigger complex behaviors? How is the brain shaped by sensory experience, and what modifications occur in neuronal circuits that allow us to learn and remember? These are the questions guiding the work of **Dinu Florin Albeanu**, who is using the olfactory bulb and olfactory cortex of mice as the subject of his current studies. Airborne chemicals, translated into neuronal signals by specific receptors in the nose, are sent directly to the olfactory bulb. Advances in optical imaging and optogenetics combined with electrophysiological recordings enable Albeanu and colleagues to monitor and/or alter patterns of activity at unprecedented synaptic and millisecond resolution, in real time, as animals are engaged in various behaviors. For survival, rodents need to identify the smells of objects of interest such as food, mates, and predators, across their recurring appearances in the surroundings, despite apparent variations in their features. Furthermore, animals aptly extract relevant information about their environment across different sensory modalities, combining olfactory, visual, or auditory cues. By recording neuronal activity in the input and output layers of the olfactory bulb, as well as feedback from olfactory cortical areas and neuromodulatory signals, Albeanu and his team aim to understand computations the bulb performs and how this information is decoded deeper in the brain. They have recently published evidence suggesting that the mouse olfactory bulb is not merely a relay station between the nose and cortex, as many have supposed. Using optogenetic tools and a novel patterned illumination technique, they discovered that there are many more information output channels leaving the olfactory bulb for the cortex than there are inputs received from the nose. They are currently investigating how this diversity of bulb output is generated, as well as how downstream areas, such as the pyriform and parietal cortex, make use of such information during behaviors.

The study of decision-making provides a window into the family of brain functions that constitute cognition. It intervenes between perception and action and can link one to the other. Although much is known about sensory processing and motor control, much less is known about the circuitry connecting them. Some of the most interesting circuits are those that make it possible to deliberate among different interpretations of sensory information before making a choice about what to do. **Anne Churchland**'s lab investigates the neural machinery underlying decision-making. Lab members use carefully designed paradigms that encourage experimental subjects to deliberate over incoming sensory evidence before making a decision. Recent results show that rats and humans have a statistically similar decision-making ability. To connect this behavior to its underlying neural circuitry, the researchers measure electrophysiological responses of cortical neurons in rodents as they perform designated tasks. The lab's current focus is on parietal cortex, which appears to be at the midpoint between sensory processing and motor planning. Churchland and colleagues also use theoretical models of varying complexity to further constrain how observed neural responses might drive behavior. This approach generates insights into sensory processing, motor planning, and complex cognitive function.

Hiro Furukawa's lab studies receptor molecules involved in neurotransmission. Its members mainly focus on the structure and function of NMDA (*N*-methyl-D-aspartate) receptors—ion channels that mediate excitatory transmission. Dysfunctional NMDA receptors cause neurological disorders and diseases, including Alzheimer's disease, Parkinson's disease, schizophrenia, depression, and stroke-related ischemic injuries. The Furukawa lab is working to solve the three-dimensional

structure of the very large NMDA receptor by dividing it into several domains. They seek to understand the pharmacological specificity of neurotransmitter ligands and allosteric modulators in different subtypes of NMDA receptors at the molecular level. Toward this end, they use cutting-edge techniques in X-ray crystallography to obtain crystal structures of the NMDA receptor domains and validate structure-based functional hypotheses by a combination of biophysical techniques, including electrophysiology, fluorescence analysis, isothermal titration calorimetry, and analytical centrifugation. Crystal structures of NMDA receptors serve as a blueprint for creating and improving the design of therapeutic compounds with minimal side effects for treating neurological disorders and diseases. During the last several years, the team discovered and mapped several regulatory sites in specific classes of NMDA receptors, progress that now opens the way to the development of a new potential class of drugs to modulate receptor activity.

Josh Huang and colleagues study the assembly and function of neural circuits in the neocortex of the mouse. The neocortex consists of a constellation of functional areas that form a representational map of the external (sensory, social) and internal (visceral, emotional) world. These areas are strategically interconnected into elaborate information processing networks that guide behavior. The group's overarching hypothesis is that, at the cellular level, cortical processing streams and output channels are mediated by a large set of distinct glutamatergic pyramidal neuron types, and functional neural ensembles are regulated by a diverse set of GABAergic interneuron types. Understanding cortical circuit organization requires comprehensive knowledge of these basic cellular components. The Huang lab uses state-of-the-art genetic approaches to systematically target cell types and facilitate the application of a full set of modern techniques for exploring cortical circuits. Among GABAergic interneurons, the chandelier cell is one of the most distinctive cell types that control pyramidal neuron firing at the axon initial segment. Huang and colleagues are studying the developmental specification, activity-dependent circuit integration, and functional connectivity of chandelier cells, an entry point toward understanding a local circuit module. Regarding pyramidal neurons, they are systematically characterizing the developmental origin, axon projection pattern, and input connectivity of multiple classes of pyramidal neuron types, focusing on the forelimb motor cortex. They combine a range of approaches that include genetic and viral engineering, genetic fate mapping, gene expression profiling, cellular imaging, electrophysiology, and behavior analysis. Recently, they began to integrate their studies in the context of the motor cortex control of forelimb movements.

Adam Kepecs and colleagues are interested in identifying the neurobiological principles underlying cognition and decision-making. They use a reductionist approach, distilling behavioral questions to quantitative behavioral tasks for rats and mice that enable the monitoring and manipulation of neural circuits supporting behavior. Using state-of-the-art electrophysiological techniques, they first seek to establish the neural correlates of behavior and then use molecular and optogenetic manipulations to systematically dissect the underlying neural circuits. Given the complexity of animal behavior and the dynamics of neural networks that produce it, their studies require quantitative analysis and make regular use of computational models. The team also has begun to incorporate human psychophysics to validate its behavioral observations in rodents by linking them with analogous behaviors in human subjects. Currently, the team's research encompasses study of (1) the neural basis of decision confidence, (2) the division of labor among cell types in prefrontal cortex, (3) how the cholinergic system supports learning and attention, and (4) the social decisions that rely on stereotyped circuits. A unifying theme is the use of precisely timed cell-type- and pathway-specific perturbations to effect gain and loss of function for specific behavioral abilities. This year, the Kepecs lab was able to link foraging decisions—the choice between staying or going—to a neural circuit and specific cell types in the prefrontal cortex. In other work, they

identified a class of inhibitory neurons that specializes in inhibiting other inhibitory neurons in the cerebral cortex and conveys information about reward and punishment. Through manipulations of genetically and anatomically defined neuronal elements, the team hopes to identify fundamental principles of neural circuit function that will be useful for developing therapies for diseases such as schizophrenia, Alzheimer's disease, and autism spectrum disorder.

Alexei Koulakov and colleagues are trying to determine the mathematical rules by which the brain assembles itself, with particular focus on the formation of sensory circuits such as those involved in visual perception and olfaction. The visual system of the mouse was chosen for study, in part, because its components, in neuroanatomical terms, are well understood. What are not known is how projections are generated that lead from the eye through the thalamus and into the visual cortex, how an individual's experience influences the configuration of the network, and what parameters for the process are set by genetic factors. Even less is known about the assembly of the neural net within the mouse olfactory system, which, in the end, enables the individual to distinguish one smell from another with astonishing specificity and to remember such distinctions over time. These are among the challenges that engage Koulakov and his team.

Understanding the link between neural circuits and behavior has been the focus of research in **Bo Li**'s lab. The team is particularly interested in studying the synaptic and circuit mechanisms underlying reward processing, attention, and learning and memory, as well as synaptic and circuit dysfunctions responsible for maladaptive behaviors that are related to major mental disorders. They integrate in vitro and in vivo electrophysiology, imaging, molecular, genetic, optogenetic, and chemogenetic techniques to probe and manipulate the function of specific neural circuits—with a focus on the fear and reward circuits—in the rodent brain, and to determine how these circuits participate in adaptive or maladaptive behavioral responses in various tasks.

Partha Mitra seeks to develop an integrative picture of brain function, incorporating theory, informatics, and experimental work. In the ongoing Mouse Brain Architecture Project, Mitra and colleagues are generating a brain-wide connectivity map for the mouse using a shotgun approach in which neuronal tracer substances are injected systematically on a grid in the brain. Currently, approximately 500 tracer-injected mouse brains may be viewed through a virtual online digital microscope on the project portal (http://mouse.brainarchitecture.org). The project requires a petabyte of data, posing big-data computational challenges that the lab is finding novel ways of meeting. In another application of whole-brain digital neuroanatomy, Mitra is collaborating with Josh Huang to characterize the distribution of the cell bodies and processes of subtypes of GABAergic neurons in mouse brains to understand the differences between a normal mouse and mouse models of autism spectrum disorders. In parallel, Mitra is undertaking theoretical work at the interface between physics, engineering, and biology by bringing methods from statistical physics to bear on problems in network control theory and multivariable statistics. Biological networks involve large numbers of variables, and it is expected that insights and analytical methods derived from this work will apply to biological networks such as the whole-brain network being determined in the Mouse Brain Architecture Project.

Pavel Osten's lab works on identification and analysis of brain regions, neural circuits, and connectivity pathways that are disrupted in genetic mouse models of autism and schizophrenia. Osten hypothesizes that (1) systematic comparison of multiple genetic mouse models will allow determination of overlaps in pathology—neural circuit endophenotypes—responsible for the manifestation of neuropsychiatric disorders and (2) neural circuit–based classification of autism and schizophrenia will provide key circuit targets for detailed mechanistic studies and therapeutic development. Osten and colleagues have developed the first systematic approach to the study

of neural circuits in mouse models of psychiatric diseases, based on a pipeline of anatomical and functional methods for analysis of mouse-brain circuits. An important part of this pipeline is high-throughput microscopy for whole-mouse-brain imaging, called serial two-photon (STP) tomography. This year, they used this method to describe the first whole-brain activation map representing social behavior in normal mice. They are currently focusing on using this approach to study brain activation changes in two mouse models of autism: the 16p11.2 df/+ mouse model, which shows an increased propensity to seizures and hyperactivity, and the CNTNAP2 knockout mouse model, which shows abnormal social behavior.

Stephen Shea's lab studies the neural circuitry underlying social communication and decisions. He uses natural social communication behavior in mice as a model to understand circuits and processes that are evolutionarily conserved and therefore shared broadly across species, likely contributing to disorders such as autism. Shea and colleagues have examined how emotion and arousal enable mice, via their olfactory systems, to store memories of other individuals and related social signals. The team has exploited the intimate relationship between memory and emotion to effectively create memories in anesthetized mice, allowing them unprecedented access to neurobiological processes that typically only occur during behavior. The lab has been making a detailed analysis of the changes in neural connections that underlie odor memory. The team is particularly focused on an enigmatic cell type (granule cells or GCs) that has long been hypothesized to be crucial for memories, but has resisted direct study. They have developed methods for recording, giving them the first glimpse of the dynamics of these cells while the animal is learning an odor. The results show unexpectedly complex population dynamics among the GCs that were independently predicted by a model of odor learning developed in Alexei Koulakov's lab. The two labs are collaborating to discern how GC population activity gets integrated by olfactory bulb output neurons and pinpoint the synaptic circuit that underlies this form of learning. In parallel, another member of the lab is using imaging techniques to determine how memories are stored among broad neuronal ensembles, at a different level of the system. Recently, the lab made a key breakthrough, developing the ability to record from GCs in awake animals and discovering that their activity is dramatically modulated by state of consciousness. Finally, the Shea lab completed a series of studies of a different form of social recognition: auditory recognition of pup vocalizations by their mothers. Through this research, they have shown that a mouse model of Rett syndrome shows deficits in communication and learning not unlike those in human patients. Grants from the Simons and Whitehall Foundations are allowing the lab to extend this work by directly linking these deficits to the action of the gene *MeCP2* in the auditory cortex.

The **Jessica Tollkuhn** lab seeks to understand how transient events during brain development exert lasting effects on gene expression, circuit function, and, ultimately, behavior. They study how sex-specific neural circuits in rodents are established and modulated by the gonadal hormones estrogen and testosterone. The cognate receptors for these hormones are nuclear receptor transcription factors, which orchestrate modification of local chromatin environment and thus exert long-term effects on gene expression. However, the genes regulated by these receptors, as well as the specific mechanisms they use, remain poorly understood in the brain. This is, in part, because the extraordinary cellular heterogeneity of the brain complicates analysis of the small subpopulations of neurons that mediate sex-specific behaviors.

Having recently identified sex differences in both gene expression and chromatin in brain regions known to regulate sex-specific behaviors, the lab is now working to understand how hormones generate these molecular sex differences during development, through the use of biochemical, genomic, and behavioral analyses. They have developed a method that permits genome-wide analysis of histone modifications or DNA methylation in genetically defined populations of neurons.

They hypothesize that these epigenetic data, combined with gene expression profiling, define the molecular signature of the critical period for sexual differentiation of the brain. Their goal is to provide a mechanistic link between the transcriptional effects of hormone signaling during development and the consequent social behaviors displayed in adulthood.

What is a memory? When we learn an association, information from two different sensory streams somehow becomes linked together. What is this link in terms of neural activity? For example, after a few bad experiences, we learn that the "green" smell of an unripe banana predicts its starchy taste. How has the neural response to that green smell changed so it becomes linked to that taste? What are the underlying mechanisms—what synapses change strength, what ion channel properties change? These are the questions that drive research in **Glenn Turner**'s laboratory. His team addresses these questions by tracking neural activity using a combination of different techniques. Using electrophysiological methods, they can examine individual neurons with very high resolution, monitoring synaptic strength and spiking output. They have also developed functional imaging techniques to monitor the activity of the entire set of cells in the learning and memory center of the fly brain. This comprehensive view of neural activity patterns enables them to actually predict the accuracy of memory formation in separate behavioral experiments. This year, the Turner lab was able to map the activity of a particular region of the brain that is associated with learning and memory. They found that a remarkably small number of neurons are required for flies to distinguish between odors. The Turner lab also studied the role of a specific type of cells, known as Kenyon cells, that receive input via several large claw-like protrusions. These neurons use their claws to recognize multiple individual chemicals in combination to remember a single scent. By examining the effects of learning-related genes on these processes, they can, in the future, connect their network-level view of memory formation to the underlying molecular mechanisms that govern the basic cellular and synaptic changes that drive learning.

Anthony Zador and colleagues study how brain circuitry gives rise to complex behavior. Work in the lab is focused on two main areas. First, they ask how the cortex processes sound, how that processing is modulated by attention, and how it is disrupted in neuropsychiatric disorders such as autism. Recently, the lab found that when a rat makes a decision about a sound, the information needed to make the decision is passed to a particular subset of neurons in the auditory cortex whose axons project to a structure called the striatum. In the second major line of work in the Zador lab, they are developing new methods for determining the complete wiring instructions of the mouse brain at single-neuron resolution, which they term the "Connectome." In contrast to previous methods, which make use of microscopy, these methods exploit high-throughput DNA sequencing. Because the costs of DNA sequencing are plummeting so rapidly, these methods have the potential to yield the complete wiring diagram of an entire brain for just thousands of dollars.

Yi Zhong's lab studies the neural basis of learning and memory. The team works with fruit fly models to study genes involved in human cognitive disorders, including neurofibromatosis, Noonan syndrome (NS), and Alzheimer's disease. Mutations leading to a lack of function of the *neurofibromatosis 1* (*NF1*) gene cause noncancerous tumors of the peripheral nervous system, as well as learning defects. The lab's analyses of *Drosophila NF1* mutants have revealed how expression of the mutant gene affects a pathway crucial for learning and memory formation. The *NF1* gene and a gene called *corkscrew*, implicated in NS, share a biochemical pathway. Recently, the lab succeeded in linking changes in this pathway because of specific genetic defects in NS with long-term memory deficiencies. In fly models, they discovered the molecular underpinnings of the "spacing effect"—the fact that memory is improved when learning sessions are spaced out between rest intervals. Zhong's team also has succeeded in reversing memory deficits in mutant flies, work suggesting longer resting intervals for NS patients might reverse their memory deficits. They also

identified a means of reversing memory loss in fruit flies while suppressing brain plaques similar to those implicated in Alzheimer's disease by blocking epidermal growth factor receptor (EGFR) signaling, a pathway commonly targeted in cancer. Separately, having discovered that memory decay is an active process, regulated by the Rac protein, the team has proposed that Rac's role in erasing memory is related to its influence on downstream cytoskeleton remodeling agents. This year, the Zhong lab explored how neurons control our response to different scents, offering insight into how the brain distinguishes between food odors that are attractive and repulsive. They discovered that neurons expressing a particular peptide were only activated by food odors, and the amount of activation predicted how much a fly was attracted to a particular odor.

UNDERSTANDING NEURONAL CIRCUITS IN THE MAMMALIAN OLFACTORY BULB

F. Albeanu A. Banerjee, W. Bast, H.G. Chae, Y. Chen, M. Davis, M. Dussauze, P. Gupta, M. Koh, M. Modi, G. Otazu, G. Vargova, C. Velasquez

The focus of our research group is on understanding how neuronal circuits encode and interpret inputs from the environment across different contexts, senses, and brain states to generate meaningful behaviors. Toward this end, we use optogenetic methods (such as fast multiphoton laser scanning imaging of genetically encoded neuronal activity reporters or patterned illumination of light-gated neuronal activity switches) coupled with electrophysiological measurements (extracellular and intracellular recordings). We use the rodent olfactory system as a model and monitor neuronal inputs, outputs, and feedback loops across different layers of the circuit.

We want to understand (1) how inputs get processed at different synapses of the underlying neuronal circuits; (2) what roles top-down feedback signals play in decoding specific stimulus features as a function of prior experience, and (3) how neuronal representations change across brain states and contexts to support cognitive processes such as selective attention. The broad scope of this effort is observing how perceptions arise and relate to action during specific behaviors.

We modulate activity of select neuronal populations in a reversible manner in awake animals to understand how fundamental sensory encoding problems are solved by the olfactory bulb (OB) and olfactory (piriform) cortex networks.

Investigating Selective Attention and Motor Control across the Brain in a Novel Self-Controlled Odor Localization Paradigm

Sensory environments are complex and rich in information, and animals actively navigate through them in a closed loop. Animals use different mechanisms to actively control their sensory environment through iterative sensorimotor transformations and selectively process information only from the source of interest and ignore irrelevant information (selective attention). Yet, in lab settings, most stimuli are delivered in isolation to passive subjects that have no control over the incoming stimuli. The rodent olfactory system offers a suitable model for understanding how sensory information is transformed into motor action—rodents display excellent behavioral acuity for tracking down intermittent odor plumes, despite the presence of several distractor odor sources in the environment. We capitalize on this behavior and aim to understand what processes in the brain allow an animal to selectively localize a sensory object of interest, in closed-loop versus open-loop situations. Briefly, in this paradigm, head-fixed mice are trained to move a lever, whose movement is coupled to the real-time location of an odor source, to park a target odor in front of their snout, which triggers availability of reward. As the lever is moved back and forth, the location of the odor source changes from left to right, following transfer functions that change across trials, such that animals need to rely on sensory feedback from the odorant (Fig. 1). To mimic natural complexity, distractor odor sources are also introduced in the environment, independent of the lever movement. Within this framework, animals not only locate where the odor is coming from, but also selectively respond only to the odor stream that remains coupled to their own movement. Coupled with activity recordings and manipulations, the self-controlled sensory stimulus delivery in head-fixed animals enables a powerful experimental platform for investigating processes ranging from stimulus localization and identification to sensorimotor integration in open- versus closed-loop fashion and selective attention.

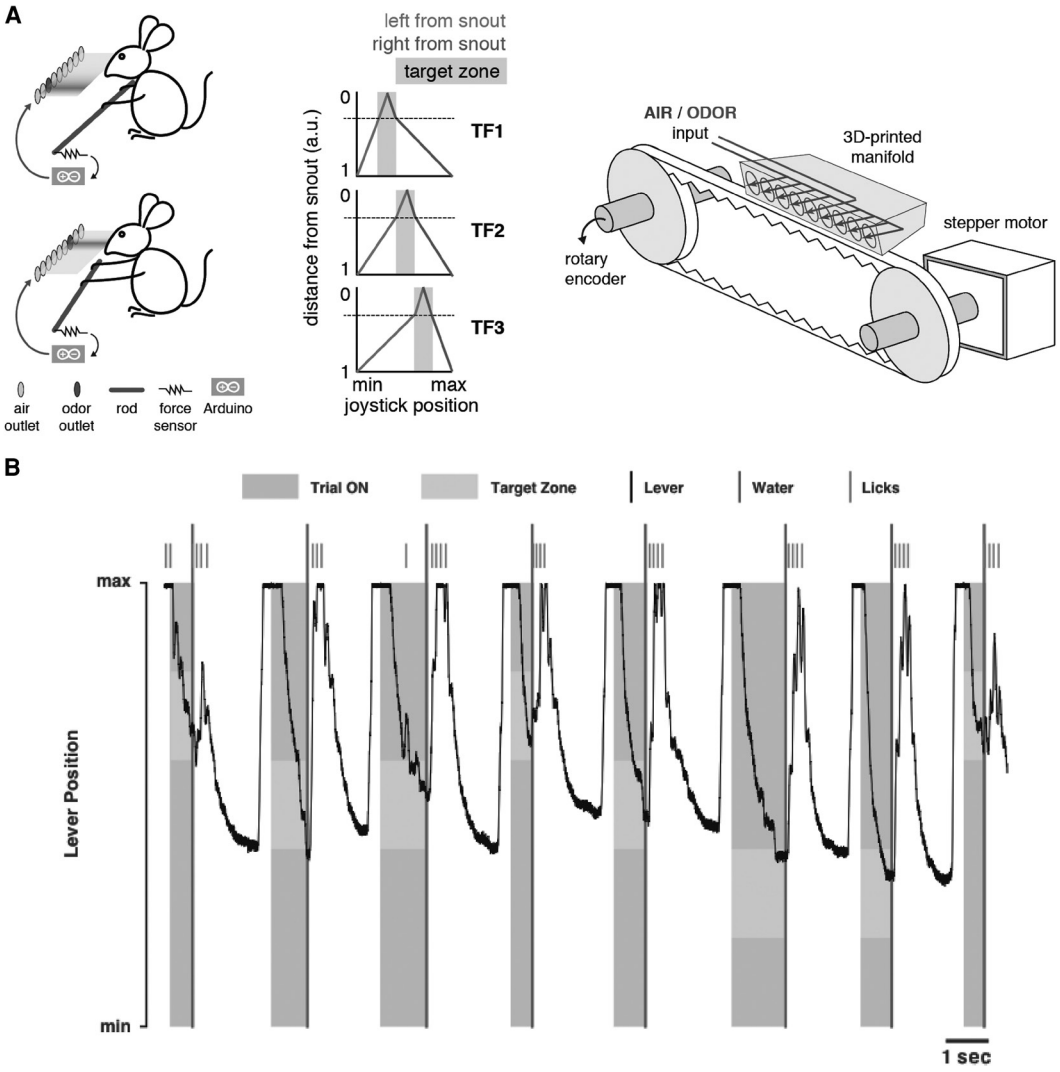


Figure 1.

High-Throughput Mapping of OB Projections across the Brain

Progress in understanding the logic of olfactory coding has been hindered by the inability to map the information flow from glomeruli through downstream brain circuits. Current methods, based on sparse fluorescent labeling, do not allow large-scale characterization of individual neuron projection statistics nor the mapping of their functional convergence to distinct target brain areas. The OB relays sensory information represented by olfactory sensory neurons through its output neurons, the mitral and tufted (M/T) cells, to higher brain areas. The major OB target areas, the piriform cortex (PC), anterior olfactory nucleus (AON), olfactory tubercle, cortical amygdala, lateral entorhinal cortex, and hippocampus, have been proposed to perform distinct computations ranging from odor detection and localization, guiding spatial navigation to odor identification to innate, or learned, stimulus value assignment. It remains largely unknown whether these functional differences emerge locally or are (in part) dictated by differential inputs from the OB. Using a novel high-throughput sequencing technology (multiplexed analysis of projections by sequencing, MAPseq), we are investigating the projection patterns of individual M/T cells across the brain. In particular, we label a large set of individual M/T

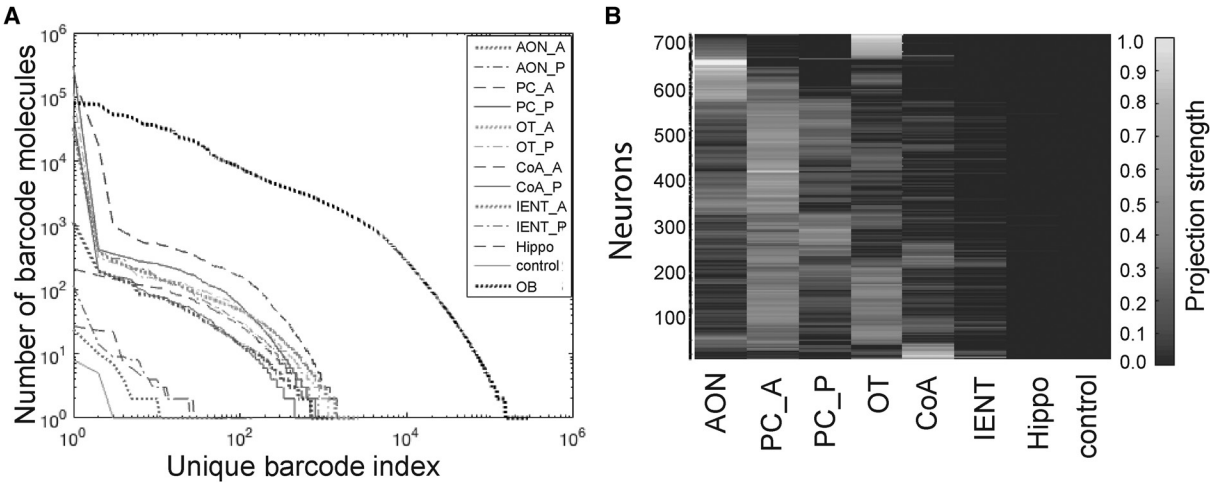


Figure 2. Multiplexed analysis of projections by sequencing (MAPseq) experiments reveal biases in olfactory bulb (OB) projection patterns. (*A*) Distribution of abundance of barcodes in the injection site (OB) and target areas in one example experiment. (*B*) A heat map of all approximately 700 projection patterns from two animals. Abundance of each barcode is normalized to one across the target areas and color-coded.

cells (~500/experiment) with unique barcoded RNAs using a viral strategy and trace their axonal projections to different target brain areas by sequencing the barcoded RNA. Our approach enables systematic labeling of M/T cells across multiple aspects of the bulb that sample inputs from different sets of glomeruli. We cut 200-μm-thick coronal sections along the anteroposterior (AP) axis of fresh frozen brains and capture regions of interest (ROIs) pertaining to the major target areas via laser microdissection. By examining the projection patterns of thousands of M/T cells, we investigate whether M/T cells broadcast information indiscriminately or have preferential projections to specific brain areas (Fig. 2).

Understanding the Relationship between Olfactory Perceptual Discriminability and Glomerular Response Features

For rodents, the ability to recognize and discriminate particular combinations of volatile compounds is essential for their survival. Mice can easily report the difference between weak, similar odors in rich sensory scenes, even when stronger odorants fluctuate in the background. To date, the neural mechanisms underlying such behavior remain unknown.

To understand the neural basis of odor discrimination, we measured and manipulated the activity of the input nodes of the olfactory system, the glomeruli. By using wide-field optical imaging in conjunction with odor stimulation, we tracked the position of glomeruli and quantified their odor response properties; this allowed us to define different sets of affine and nonaffine glomeruli with a variable number of components. We aim to determine the relationship between the discriminability of olfactory stimuli and the similarity of glomerular odor response profiles. We additionally quantified the discriminability of the stimuli with the degree of overlap between different sets of glomeruli, as well as the physical separation of glomeruli on the bulb surface.

Toward this end, and to assess the specificity of photostimulation, we express red-activatable channelrhodopsin1 (ReaChR) in all mature olfactory sensory neurons and GCaMP6f in the OB output neurons. We use digital micromirror device (DMD)-based patterned illumination to selectively stimulate combinations of glomeruli on the dorsal surface of the bulb with subglomerular resolution (~10 μm) and high temporal precision (3 msec) in awake, head-fixed mice. Before optogenetic stimulation, using a large odor panel (up to approximately 100 stimuli), we identified the exact locations of glomeruli, revealing their shapes and response tuning to the odors sampled. We further create glomerular light patterns of known odor response similarity (within the range of our panel) and project specific glomerular inputs. In a two-alternative forced-choice discrimination task, we systematically relate the similarity of these light

patterns to the perceived difference between them. Further, using a novel strategy to decouple patterned photostimulation and two-photon imaging across different axial planes, we are monitoring the responses of M/T cells in the deeper layers of the bulb.

We are further implementing strategies that will enable noninvasive, functional dissection of neuronal networks with cellular resolution in behaving animals. This will be brought about via a closed-loop strategy involving real-time control of activity of select neurons with simultaneous monitoring of the concomitant effects of these manipulations on neuronal outputs within the circuit and elsewhere in the brain. Briefly, we are using digital holography methods via spatial light modulators (SLMs) to optogenetically control neurons of interest at single-cell level and DMD-based methods to control cell-type-specific populations across large brain regions (Fig. 3). This allows us to both replicate and systematically manipulate stimulus-evoked activity patterns in a circuit. We are simultaneously using two-photon calcium imaging and electrophysiology within the same and different brain regions (OB vs. olfactory cortex) to dissect how the alteration of select circuit elements, or their specific properties, affects the output of the network. This closed-loop approach will make it possible to determine the spatial-temporal integration rules within the bulb and the olfactory cortex, investigate the relevance of spike-time codes, and reveal underlying decoding schemes.

Task-Related Representations in the Cortical-Bulbar Feedback

Sensory systems format information about objects in the environment into neural representations used by the brain depending on context and prior experience. Sensory representations emerge from the interplay between feedforward inputs, as well as local and feedback signals across brain areas. Given that the output neurons of the main OB—the first processing station of olfactory input—are richly modulated by task contingencies, we asked whether the feedback from PC provides an underlying substrate. To this end, we engage mice in a serial reversal learning task in which

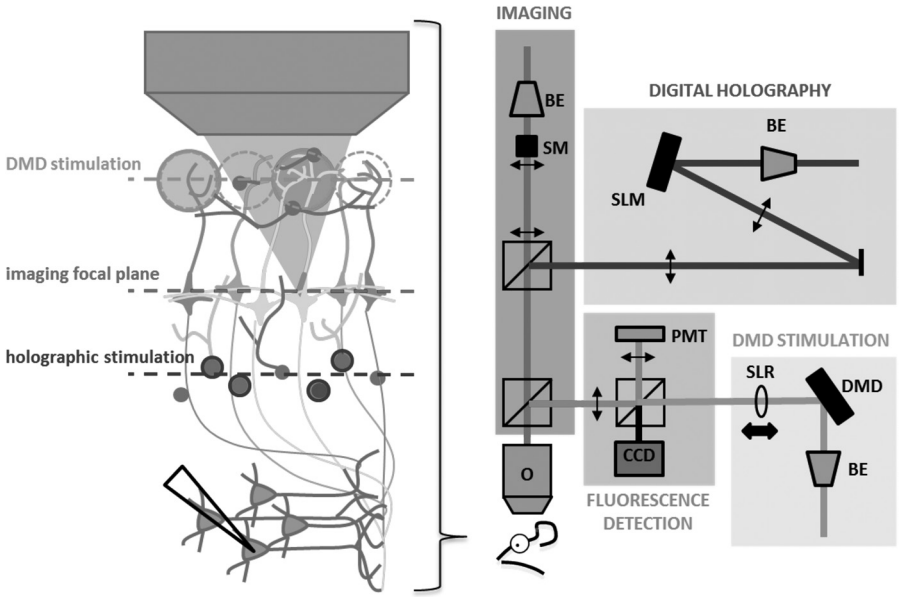


Figure 3. Combined imaging and photostimulation setup. (*Left*) Experimental configuration, combining scanning two-photon imaging (red), DMD photostimulation (blue), and holographic photostimulation (brown). (*Right*) Schematic of the experimental microscope. DMD stimulation is used to create spatiotemporal light patterns on the surface (<100 µm). Digital holography is used to photostimulate deeper (<500 µm) in the brain with cellular resolution. Calcium activity is monitored in an independent optical plane by two-photon imaging and by electrodes in downstream brain regions. BE, beam expander; SM, scan mirrors; O, objective; PMT, photomultiplier; SLR, camera lens; DMD, digital micromirror device; SLM, spatial light modulator; CCD, charge-coupled device.

a stimulus (odorant or sound) carries distinct motivational value according to different contingencies. In parallel, we monitored the activity of GCaMP5-labeled cortical-bulbar feedback axons. We report the existence of two largely independent populations of axon terminals, responsive in sound versus odor trials, and spanning diverse temporal profiles across both response polarities. Correlation and dimensionality reduction population analysis identified sets of choice-selective and -independent boutons active in both odor and sound trials (Fig. 4). Sound responses occurred mainly during reward (hit) and error (false alarm) trials, further suggesting that the cortical-bulbar feedback carries information on action outcome. Optogenetic inactivation of cortical feedback axons locally in the bulb substantially impaired the task performance compared with controls for both odor and sound trials. These effects were dominated by increased error rates (false alarms) and persisted beyond the light-on trials. In ongoing experiments, we are monitoring the cortical feedback and OB outputs while varying the size of reward, interval between sensory cues and time when reward is available, and identity of odor-sound pairs.

Taken together, our results suggest that cortical feedback conveys information related to task contingencies, such as learning rules and reward, and is well positioned to flexibly control the bulb output during olfactory behaviors.

Two Parallel Feedforward and Feedback Pathways for Olfactory Information Processing

Areas at the sensory periphery send feedforward signals to the cortex and, in turn, receive massive top-down cortical feedback. To date, the function and degree of specificity of such feedback loops remain poorly understood. In the mammalian olfactory system, the OB output neurons—the mitral cells (MCs) and tufted cells (TCs)—differ in their intrinsic properties, local connectivity, and projection targets. Two important cortical targets of OB outputs—the AON and the PC—receive feedforward input predominantly from TCs and MCs, respectively. In turn, both AON and PC send massive glutamatergic projections back to local inhibitory OB interneurons (mainly granule cells, GC). We asked whether and how top-down feedback inhibition received by MCs and TCs is functionally segregated based on their projection targets. We monitored the odor responses of MCs and TCs separately via multiphoton calcium imaging in awake, head-fixed mice, while selectively silencing neuronal activity in the PC or the AON. We found that cortical feedback indeed segregated—silencing PC selectively increased the response amplitude and odor correlations of MCs, but not TCs (Fig. 5; Otazu et al., *Neuron 86*: 1461 [2015]), whereas silencing AON exerted a substantially stronger effect on the odor responses of TCs (increased response amplitude and odor correlations) compared with the MCs. We further investigated whether MCs and TCs process sensory input differentially, and measured their responses across two stimulus dimensions: odorant identity and concentration. MCs showed greater response heterogeneity across odors, as well as concentrations, whereas TCs showed higher population correlations across odorants and nearly monotonic scaling with increasing stimulus concentration. Additionally, the cross-validated linear decoding performance for concentration-invariant odor identity was found to be significantly better from the TC population than from the MC population. Therefore, we propose that sensory information in these two parallel loops is formatted differently. Within this scenario, we are currently testing the hypothesis that cortical targets of TCs (AON, olfactory tubercle) are poised to infer concentration-invariant odor identity using linear decoding schemes, whereas the cortical targets of MCs such as the PC are more suited for nonlinear decoding strategies. In summary, we describe two parallel feedforward and feedback streams of odor processing in the mammalian brain with differential inhibitory organization and neural representation for odor identity and concentration, further providing a testable framework during olfactory behaviors.

Mapping Odor Space onto Neuronal Representations in the Olfactory System

This work was done in collaboration with the Koulakov lab.

Unlike many other sensory systems, low-dimensional metrics for characterizing stimuli have remained elusive for olfaction, and it is unclear what features of chemical stimuli are represented by neurons.

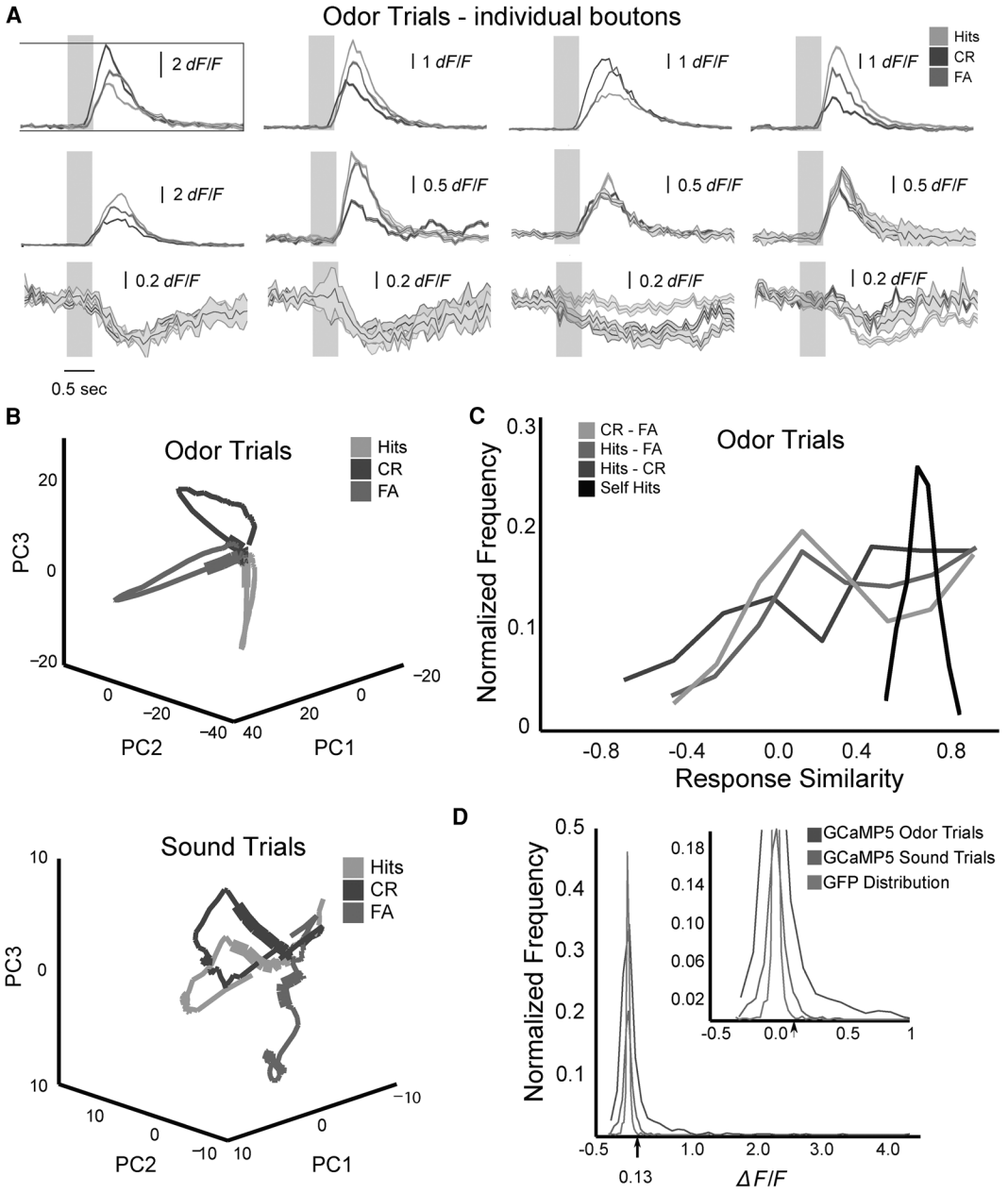


Figure 4. Cortical feedback responses to odor stimuli cluster differently across different task conditions. (*A*) Excitatory and inhibitory trial-averaged responses from individual feedback boutons exemplifying differential modulation across task conditions (thick line, average response; shaded area ± SEM; gray bar, stimulus [350 msec]). (*B*) Ensemble responses of boutons across trial outcomes projected onto the first three principal components (64% of variance). Each trajectory represents the bouton population activity over time for odor (*top*) and sound trials (*bottom*). Origin marks start of trial and thick line marks the stimulus period. (*C*) Histograms of pairwise response similarity for odor trials across conditions. Gray lines indicate uncentered correlation between responses of single boutons for different trial outcomes; black line shows distribution of self-correlations between two halves of hit trials resampled by bootstrapping 1000 times. (*D*) Histograms of average Δ*F*/*F* for GCaMP responses versus green fluorescent protein (GFP). Each point is average fluorescence of a 600-msec window starting at the stimulus onset. The arrow denotes the 99th percentile of the GFP distribution, used as significance threshold for GCaMP responses for further analysis.

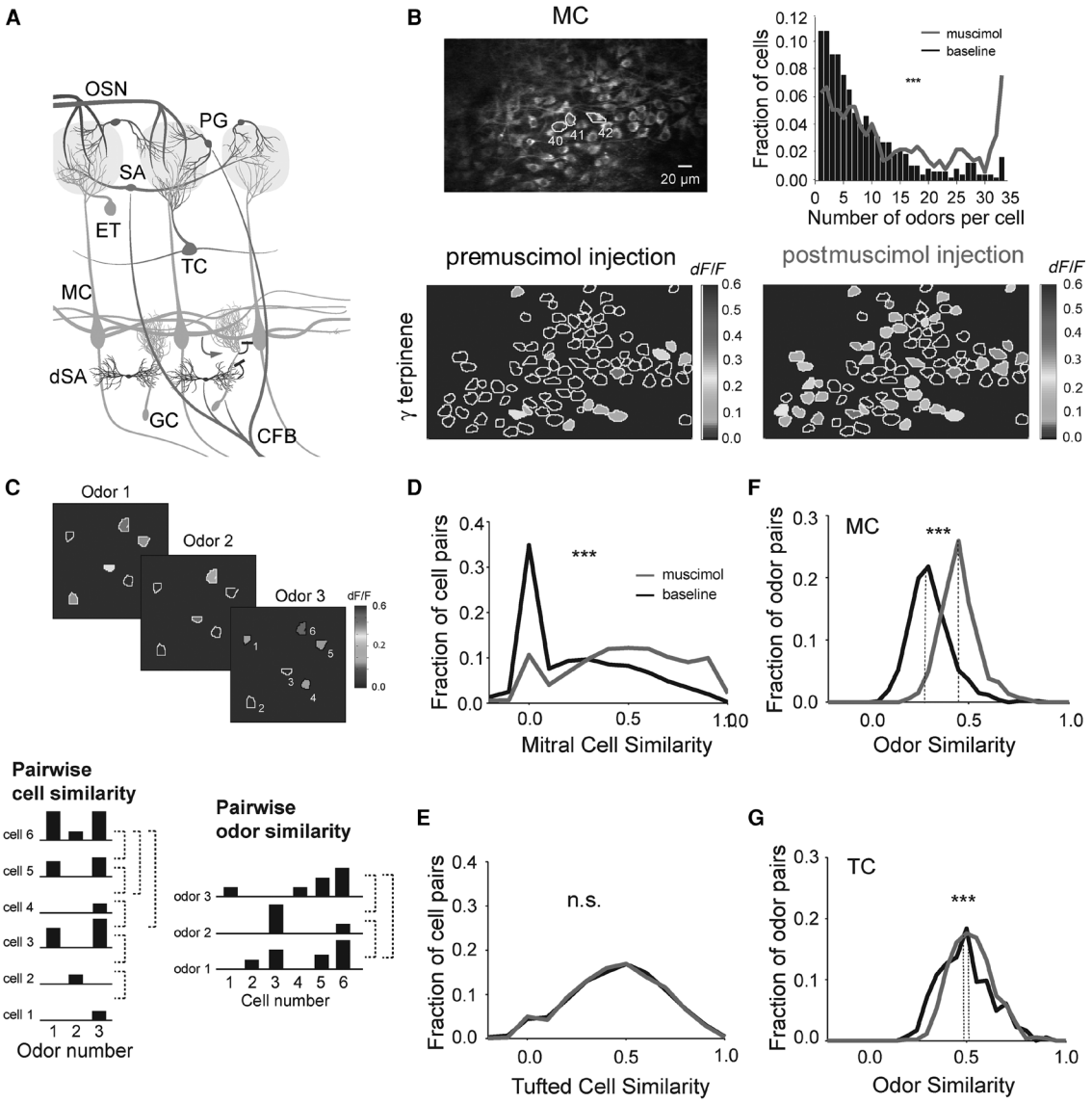


Figure 5. Cortical-bulbar feedback decorrelates the OB output. (*A*) Olfactory bulb (OB) circuit and neuronal types; OSN, olfactory sensory neurons; PG, periglomerular cells; SA, superficial short axon cells; ET, external tufted cells; TC, tufted cells; MC, mitral cells; dSA, deep short axon cells; GC, granule cells; CFB, cortical feedback fibers synapsing onto inhibitory interneurons across the OB. (*B*) (*Top left*) average resting fluorescence of an example field of view in the MC layer (~220 µm from surface). (*Bottom*) Color map showing average fluorescence change in response to γ terpinene in the field of view before (*left*) and after (*right*) muscimol injection (injected in anterior piriform cortex [PC] to suppress cortical activity and thus disable cortical feedback). (*Top right*) Histogram of the number of odors individual MCs responded to before (black bars) and after (red trace) muscimol injection. *** Indicate significance level ($p < 0.001$, Wilcoxon signed rank test). (*C*) Schematic exemplifying pairwise cell similarity and pairwise odor similarity calculations for a given field of view; (*left*) cartoon showing responses of six identified regions of interest (ROIs; yellow outlines) within a given field of view across three odors; color indicates the average response amplitude ($\Delta F/F$) for each ROI; (*center*) an odor response spectrum (ORS) is calculated for each ROI (cell) as the vector containing the average $\Delta F/F$ for each odor; pairwise cell similarity is calculated as the uncentered correlation between the ORS vectors for each pair of cells (indicated by dotted lines); (*right*) a cell response spectrum (CRS) is calculated for each odor as the vector containing the average $\Delta F/F$ for each cell on presentation of the given odor; pairwise odor similarity is calculated as the uncentered correlation between the CRS vectors for each pair of odors. (*D–G*) Histograms of pairwise cell (*left*) and odor (*right*) similarity of MC and TC before (black, baseline) and after (red) muscimol injection. *** Indicate significance level ($p < 0.001$, Wilcoxon signed rank test); dotted lines indicate the median.

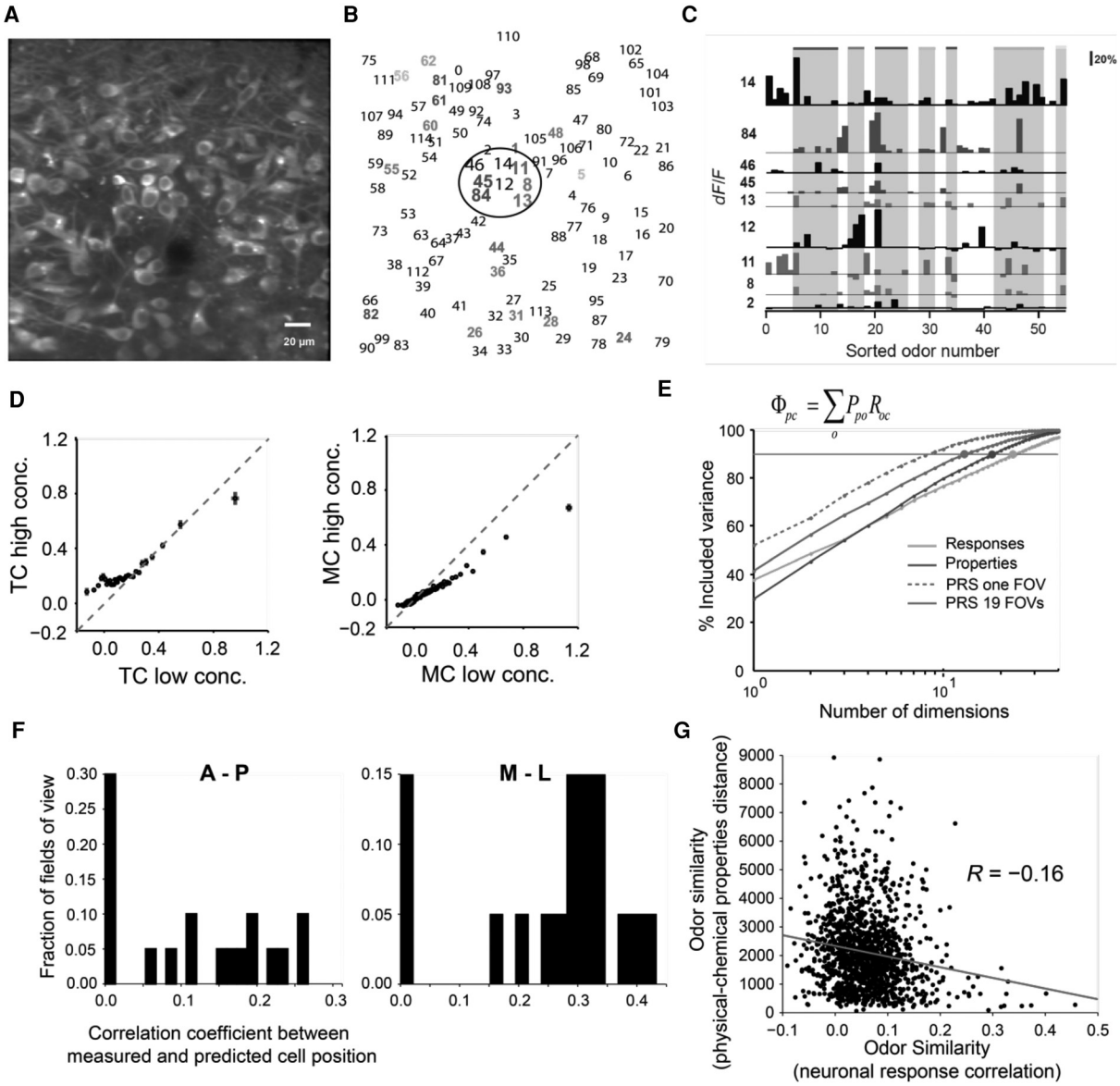


Figure 6. Mapping odor space on mitral cell (MC) activity patterns. (*A*) Average resting fluorescence multiphoton image in the MC layer. (*B*) Numbers indicate the relative positions of 112 MC bodies in the imaged field of view. (*C*) Example odor response spectra of MC bodies within an arbitrarily picked 75-µm-diameter region in *A* sorted according to functional chemical groups (tiglates, thiazoles, ethyl esters, aldehydes, ketones, alcohols, and acids). (*D*) Scatterplots showing odor-induced change in tufted (*left*) and mitral (*right*) cells with increasing odor concentration. (*E*) Dimensionality of odorant properties, mitral and tufted (M/T) neuronal responses, and M/T receptive fields (properties response strength, PRS, Φ). (*F*) A LASSO/jackknife predictor yields a set of anteroposterior (A-P) and mediolateral (M-L) cell positions that are weakly correlated with observed values. (*G*) Odor pairwise similarity across a set of 1660 physical-chemical properties versus neuronal responses.

We would like to relate neural activity in the early olfactory system of mice to the physicochemical properties of odorants. We imaged odor-evoked responses in identified TCs and MCs in awake mice using multiphoton microscopy. Similarity of odors computed through a comprehensive set of 1664 physicochemical features was a poor predictor of similarity of the corresponding neuronal representation by MCs or TCs. Dimension reduction revealed that fewer dimensions (~14) in the combined neural-chemical space than

in neural space alone (~25) were required to explain 90% of variance in neural responses across the population, suggesting that factors other than the physicochemical properties we considered are required to fully explain the neural responses (Fig. 6). Responsive MCs and TCs were spatially dispersed, and cells within a local region were functionally heterogeneous, with only limited and variable dependence of M/T cell position on odorant characteristics. Our data indicate that novel descriptors are needed to link chemical space to neuronal representations, and that odor information leaves the bulb in a mosaic pattern with substantial local diversity.

Other Collaborative Projects with CSHL Groups

Huang: DLP-based patterned stimulation to functionally map the motor cortex.

Koulakov, Lee, Zador: Sequencing the OB—bridging the gap between glomerular odor responses and odor receptor sequences by identifying the molecular identity of glomeruli.

Li: Fiber optic–based approach to monitor neuronal activity in punishment and reward neuronal circuits during behavior. We are focusing on monitoring and manipulating activity in the lateral division of central amygdala and the insula cortex.

Osten: Developing light-sheet-based approaches for fast optical reconstruction of neuronal circuits.

Zador: Optical monitoring and manipulation of neuronal activity in genetically and anatomically defined cortical circuits in animal models of cognition.

PUBLICATIONS

Kebschull JM, Garcia da Silva P, Reid AP, Peikon ID, Albeanu DF, Zador, A. 2016. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA (MAPseq). *Neuron* **91:** 1–13.

Yu K, Garcia da Silva P, Albeanu DF, Li B. 2016. Central amygdala somatostatin neurons gate passive and active defense behaviors. *J Neurosci* **36**: 6488–6496.

In Press

Anselmi F, Banerjee A, Albeanu DF. 2015. Patterned stimulation in the brain. In *New techniques in system neuroscience* (ed. Douglass AD), pp. 235–270. Springer International Publishing, Cham.

INVESTIGATING NEURAL CIRCUITS FOR SENSORY INTEGRATION AND DECISION-MAKING

A. Churchland A. Brown, B. Cascone, L. Chartarifsky, J. Eichhorn, S. Gluf, A. Juavinett, M. Kaufman, A. Licata, E. Lu, S. Musall, F. Najafi, H. Nguyen, D. Nunes Raposo, O. Odoemene, S. Pisupati, H. Read, M. Ryan

Making use of sensory information requires more than simply relaying incoming signals from the sensory organs. It requires interpreting information, classifying it, drawing inferences, and, ultimately, using the context of behavioral goals to make a decision about its meaning. A decision is a commitment to a particular choice or proposition at the expense of competing alternatives. In some situations, decisions involve integration of evidence—that is, they make use of multiple pieces of information from the environment or memory. These decisions can provide a framework in which to investigate complex cognitive processes and open a window into higher brain function in general.

Although previous experiments have begun to reveal how neural systems combine evidence to make decisions, they have left a critical gap in our understanding. Specifically, very little is currently known about the neural mechanisms that make it possible to combine information from multiple sensory modalities for decisions. The gap is apparent, although it is clear from behavioral observations that neural systems can combine multisensory information: When parsing speech in a crowded room, for example, the listener makes use of both auditory information (the speaker's vocal sounds) and visual information (the speaker's lip movements). Understanding the neural mechanisms of multisensory decisions is critical for two reasons. First, it is essential for a complete understanding of sensory perception because real-world stimuli rarely activate a single sense in isolation. Therefore, understanding how the brain interprets incoming information requires understanding how the brain merges information from different senses. Second, it is likely of clinical importance as several developmental abnormalities appear to be related to difficulties in integrating sensory information. For example, abnormalities in multisensory processing are a hallmark of subjects with autism spectrum disorder. Impairments in multisensory processing are also observed in subjects with a collection of sensory abnormalities known together as sensory processing disorder and may also be evident in patients with Rett syndrome and dyslexia. Understanding the neural mechanisms of multisensory integration could inform treatment of those conditions. Our long-term goal is to understand how the brain can make decisions that integrate inputs from our multiple senses, stored memories, and innate impulses. Our projects in previous years constituted the very first steps toward achieving this goal: We developed a multisensory behavior that could be precisely quantified in both humans and rodents (Raposo et al., *J Neurosci 32:* 3726 [2012]; Sheppard et al., *J Vision 13:* 1 [2013]). In addition, we built on the foundation by measuring the responses of neurons in rodents engaged in the behavior, affording insight into the neural circuits underlying multisensory decisions (Raposo et al., *Nat Neurosci 17:* 1784 [2014]).

Microcircuits within Posterior Parietal Cortex That Support Decision-Making

M. Kaufman, L. Chartarifsky, S. Pisupati, E. Lu

There are two projects within this category. The goal of the first project is to gain a deeper understanding of the neural circuits that allow auditory, visual, and multisensory decision-making. This year, we finalized a project that used optogenetics to manipulate neural activity and thus to understand the causal role of a candidate decision-making structure, the posterior parietal cortex (PPC) (Licata et al. 2017). This paper reveals that the PPC plays a causal role in visual decisions, but not in auditory decisions or multisensory integration.

To identify the key region that supports multisensory integration, Lital Chartarifsky and Sashank Pisupati are performing brainwide manipulations of candidate areas. Their approach is to disrupt neural activity in a candidate area and look at the effects on auditory, visual, and multisensory decisions. They disrupt activity using both optogenetic and chemogenetic approaches and are using a model-based system to evaluate and interpret the incoming data. This model-based system is a major step forward because connecting decision-making to its underlying behavior is a notoriously difficult problem (Churchland and Kiani 2016).

Population Dynamics of Neurons during Decision-Making

M. Kaufman, F. Najafi [in collaboration with D. Barabasi, Notre Dame]

The goal of this project is to understand how population activity changes from one neural structure to the next to support behavior. To tackle this question, Matt Kaufman is measuring the responses of 300 to 400 neurons simultaneously using two-photon imaging. This approach will revolutionize the kinds of questions about decision-making that we can address, especially when used in conjunction with emerging mathematical techniques for analysis (Churchland and Abbott 2016). This year, Matt was supported by a grant from the Simons Collaboration on the Global Brain. Matt will present this work at the Society for Neuroscience Annual Meeting (San Diego, 2017) and also was selected to present a poster at the Computational and Systems Neuroscience Conference (Cosyne; Salt Lake City, UT). Simon Musall was a coauthor on both posters; Daniel Barabasi, a summer student and part of the URP at CSHL, was also an author on the Cosyne poster. Farzaneh Najafi is, likewise, interested in population dynamics. She uses two-photon microscopy to image neural responses during decision-making in mice. She will focus on how excitatory and inhibitory neurons together contribute to evolving population responses. Her approach will include a consideration of the animal's strategy, such as the degree to which it uses the outcome of previous trials to guide current decisions. Farzaneh will present her ongoing findings from this work at the Society for Neuroscience Annual Meeting (San Diego, CA) and also was selected to present a poster at the Computational and Systems Neuroscience Conference (Salt Lake City, UT).

Both projects have benefitted from technical support from the Albeanu lab, our neighbors in the Marks building, and valued collaborators both at Cold Spring Harbor and Columbia University.

Brainwide Macrocircuits that Support Decision-Making

S. Musall, S. Gluf, O. Odoemene, H. Nguyen

The goal of this project is to understand how multiple neural areas make up brainwide macrocircuits, which are circuits that support decision-making. The project has three key components.

The first component is led by Onyekachi Odoemene and relies on inactivation methods to determine the role of multiple visual areas in transforming sensory information into signals that guide action. Onyekachi uses imaging methods to identify primary and secondary visual areas and then separately inactivates these and examines the effect on behavior. Research technician Hien Nguyen has provided support for this project, especially the behavioral component.

The second component is led by Simon Musall. Simon uses a combination of wide-field and two-photon imaging to understand which neural areas are active during different moments of a perceptual decision. To do so, he has been developing new behaviors in mice. Research technician Steven Gluf has provided support for this project, especially the behavioral component.

PUBLICATIONS

Churchland AK, Abbott LF. 2016. Conceptual and technical advances define a key moment for theoretical neuroscience. *Nature Neurosci* **19:** 348–349.

Churchland AK, Kiani R. 2016. Three challenges for connecting model to mechanism in decision-making. *Curr Opin Behav Sci* **11:** 74–80.

In Press

Licata AM, Kaufman MT, Raposo D, Ryan MB, Sheppard JP, Churchland AK. Posterior parietal cortex guides visual decisions in rats. *J Neurosci* **37:** 4954–4966.

MOLECULAR ANALYSIS OF NEURONAL RECEPTORS AND ION CHANNELS

H. Furukawa M. Regan N. Tajima
A. Romero-Hernandez J.X. Wang
N. Simorowski

The research in the Furukawa lab aims to understand molecular events that initiate cellular signal transductions involved in neurotransmission in the mammalian brain with scope to develop therapeutic compounds for treatment of neurological diseases and disorders, including schizophrenia, depression, stroke, and Alzheimer's disease. To achieve our goals, we conduct structural and functional studies on ion channels that control intracellular calcium signaling on stimulation by voltage and/or neurotransmitters. Those ion channels regulate strength of neurotransmission, the fundamental process for neuronal communication. Dysfunction of the ion channels studied in our group is highly implicated in the neurological disorders and diseases described above. The abnormal activation of the ion channels is caused by a number of factors, including excessive transmission of neurotransmitters and point mutations in the ion channels, which alter their functional properties. To understand functions of normal and abnormal ion channels, we use structural biology techniques, including X-ray crystallography and single-particle cryo-electron microscopy (cryo-EM), to determine three-dimensional (3D) atomic structures of target ion channels and test structure-based mechanistic hypotheses by site-directed mutagenesis in combination with biochemical and biophysical techniques such as electrophysiology. In 2016, we achieved a major breakthrough in understanding how *N*-methyl-D-aspartate receptor (NMDAR) "moves" itself to open its transmembrane ion channel.

X-Ray Crystallography and Electrophysiology of the Amino-Terminal Domain of NMDARs

The only available structures for the heterodimeric NMDAR amino-terminal domains (ATDs) to date are those bound to allosteric inhibitors, ifenprodil, and Ro25-6981, representing the allosterically inhibited state. For that reason, we conducted structural studies without allosteric inhibitors to capture a potentially active ATD conformation. The crystal structure of GluN1-GluN2B ATDs in the absence of an allosteric inhibitor (apo-GluN1b-GluN2B ATD) was obtained at 2.9-Å resolution (Fig. 1). This was achieved by crystallizing the purified GluN1b-GluN2B ATD proteins complexed to a Fab fragment made from mouse monoclonal IgGs (made by the CSHL antibody facility) to improve the quality of the crystals. The crystallographic analysis shows heterodimeric GluN1-GluN2B ATDs that have a bilobed architecture composed of the regions previously called R1 and R2 in the structure of GluN1b-GluN2B ATD bound to an allosteric inhibitor, ifenprodil (Fig. 1). There are a number of differences between the structures of the apo-GluN1b-GluN2B ATD and the ifenprodil-GluN1b-GluN2B ATD. The most apparent difference is the separation of GluN1b-R1 and GluN2B-R2 in the apo-GluN1b-GluN2B ATD due to the ~20° rigid-body opening of the GluN2B ATD bilobed structure in the apo-GluN1b-GluN2B ATD compared with that in ifenprodil-GluN1b-GluN2B ATD (Fig. 1D). Another major difference is the rearrangement of GluN1b and GluN2B subunits, which involves an ~15° rotation relative to one another (Fig. 1E). The above difference between the apo-GluN1b-GluN2B ATD and the ifenprodil-GluN1b-GluN2B ATD brings the lower lobes (R2) of GluN1-GluN2B significantly closer together in the apo-GluN1b-GluN2B ATD than in ifenprodil-GluN1b-GluN2B ATD (Fig. 1E). For example, the distance between Cαs of GluN1b Lys178 and GluN2B Asn184 in apo-GluN1b-GluN2B ATD is 4.4 Å closer than in the ifenprodil-GluN1b-GluN2B ATD (Fig. 1E).

Because subunit arrangement in the apo-GluN1b-GluN2B ATD is different from that previously observed in the ifenprodil-GluN1b-GluN2B ATD, due to the 15° rotation between GluN1b and GluN2B and

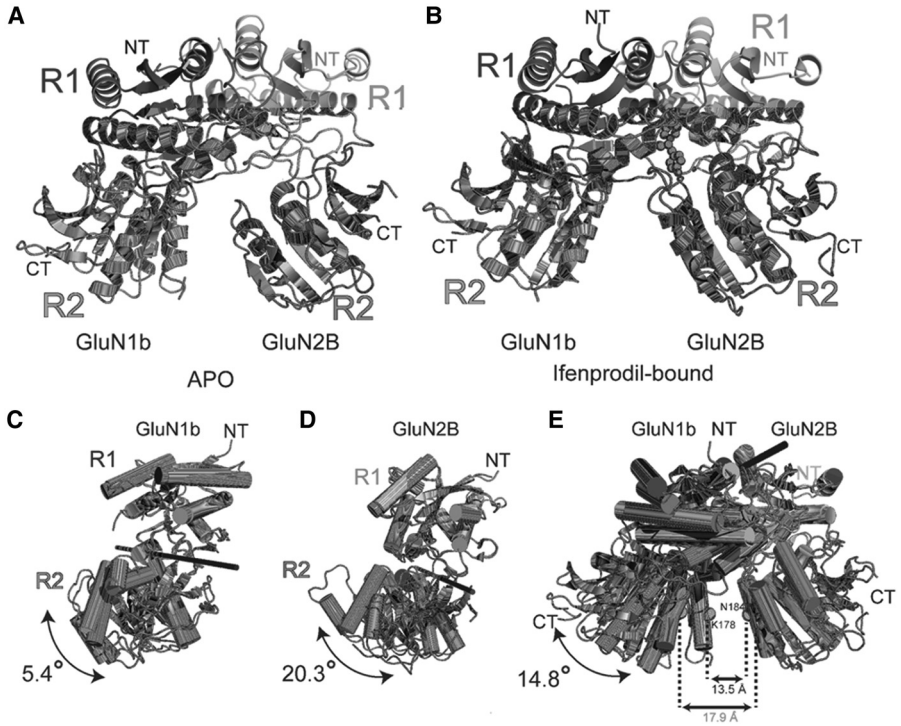


Figure 1. Structures of GluN1b-GluN2B ATD heterodimers. (*A,B*) Crystal structure of the GluN1b-GluN2B amino-terminal domain (ATD) heterodimer in the apo state solved at 2.9 Å (*A*) in comparison with the ifenprodil-bound structure (Protein Data Bank [PDB] ID:3QEL) (*B*). Both structures are aligned so that the GluN2B R1 lobe is viewed from the same angle. R1 and R2 lobes are colored in magenta and light pink for GluN1b ATD and cyan and yellow for GluN2B ATD. Ifenprodil is represented as green spheres. (*C,D*) Superposition of R1 lobes of GluN1b (*C*) and GluN2B (*D*) illustrates the relative "opening" between R1 and R2 lobes. Color-coding for the apo structure is the same as in *A* and ifenprodil-bound structure is in blue. (*E*) Structural comparison of the GluN1b-GluN2B ATD heterodimers in apo- and ifenprodil-bound forms by superposing the GluN2B R1 lobes reveals ~15° rotation of GluN1b ATD relative to GluN2B ATD. Black rods indicate the axis of rotation between two states. Color-coding is the same as in *C* and *D*.

the opening of the GluN2B ATD bilobed structure as described above, we next sought to validate the physiological relevance of this novel GluN1-GluN2B subunit arrangement. Toward this end, we tested whether an intersubunit disulfide bond can form at the subunit interface observed in the crystal structure of the apo-GluN1b-GluN2B ATD but not in the ifenprodil-GluN1b-GluN2B ATD in the context of the intact GluN1/GluN2B NMDAR by mutating GluN1 and GluN2B residues that are proximal to each other. We assumed that a spontaneous disulfide bond should form between the mutated cysteines in the intact GluN1/GluN2B NMDAR if the subunit interface observed in the crystal structure is physiological. Here, we engineered cysteine residues at GluN1b Phe113/GluN2B Ala107 and GluN1b Gly331/GluN2B Glu75, expressed and purified the mutant GluN1b/GluN2B NMDAR in the context of the intact ion channel, and assessed for a band shift in western blot conducted under nonreducing conditions (Fig. 2). In the two selected positions, the disulfide bonds are formed only when the cysteine mutant of GluN1 and that of GluN2B are coexpressed and detected by an anti-GluN1 and an anti-GluN2B western blot in the absence of β-mercaptoethanol (Fig. 2). When the cysteine mutants of one subunit are coexpressed with the wild-type (WT) of the other subunit no disulfide bonds are formed, indicating that the disulfide bonds are specifically formed by the engineered cysteines. Taken together, the above experiments show that the GluN1-GluN2B subunit arrangement observed in the apo-GluN1b-GluN2B ATD crystal structure exists in the context of the intact GluN1b/GluN2B NMDAR.

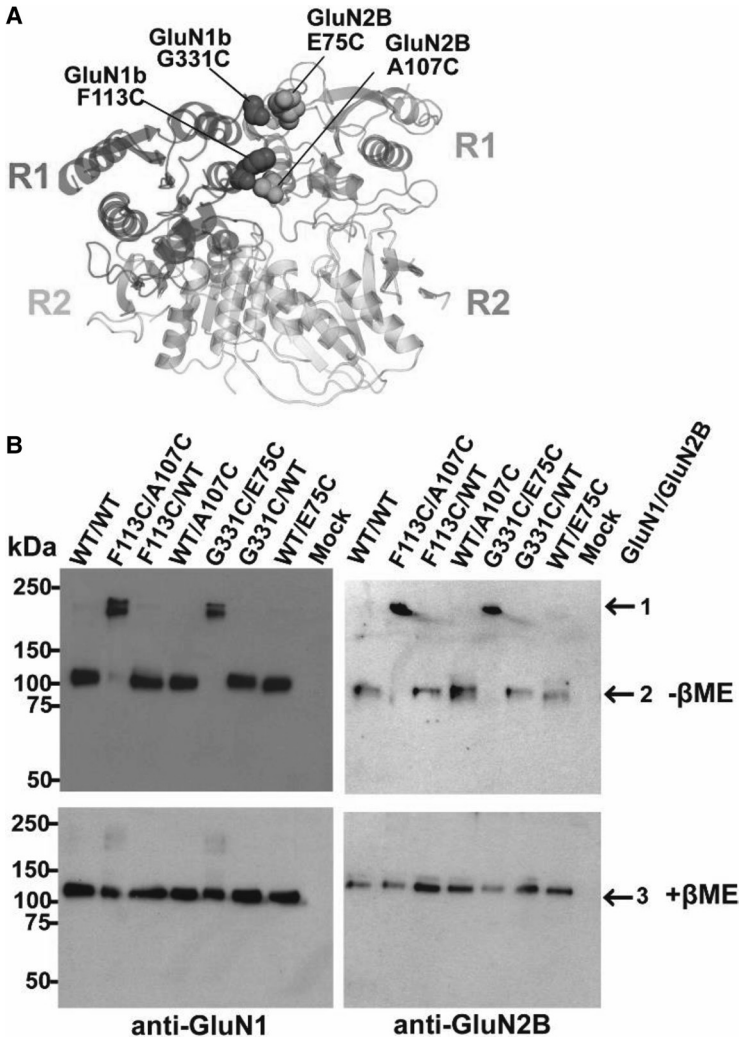


Figure 2. Validation of the crystal structure by disulfide cross-linking. (*A*) Crystal structure of the apo-GluN1b-GluN2B ATD showing locations of the mutated residues, GluN1b Phe113, GluN1b Gly331, GluN2B Ala107, and GluN2B Glu75 in spheres. (*B*) Western blots using anti-GluN1 (*left*) and anti-GluN2B (*right*) antibodies on purified intact GluN1b/GluN2B NMDA receptor that lacks the carboxy-terminal domain (CTD). *Upper* and *lower* panels are blots run in the absence and presence, respectively, of β-mercaptoethanol (βME). Bands indicated by arrow 1 are consistent with the molecular weight of GluN1-GluN2B heterodimers, whereas those indicated by arrows 2 and 3 are consistent with the molecular weights of monomers of GluN1-4b and GluN2B.

To understand what functional state the crystal structure of the apo-GluN1b-GluN2B ATD may represent, we next attempted to stabilize the conformation observed in the crystal structure and assessed the ion channel activity. Toward this end, we engineered cysteines at the positions in the lower lobes (R2) of the GluN1b and GluN2B ATDs (GluN1b Ala175Cys/GluN2B Gln180Cys and GluN1b Lys178Cys/GluN2B Asn184Cys), which are facing each other, and "trapped" the conformation by tethering the engineered cysteines with bifunctional methanethiosulfonate (bi-MTS) reagents (Fig. 3). The distances between the mutated residues are closer in the apo-GluN1b-GluN2B ATD than in ifenprodil-GluN1b-GluN2B ATD as mentioned above (Fig. 1E). When bi-MTS binds to the lower lobes of the GluN1b-GluN2B heterodimers, we reasoned that the conformation observed in the apo-GluN1b-GluN2B ATD with the open GluN2B bilobed architecture and the rearranged GluN1-GluN2B subunit orientation should be trapped. Experimentally, we coexpressed the cysteine mutants of GluN1b and GluN2B in *Xenopus* oocytes and probed

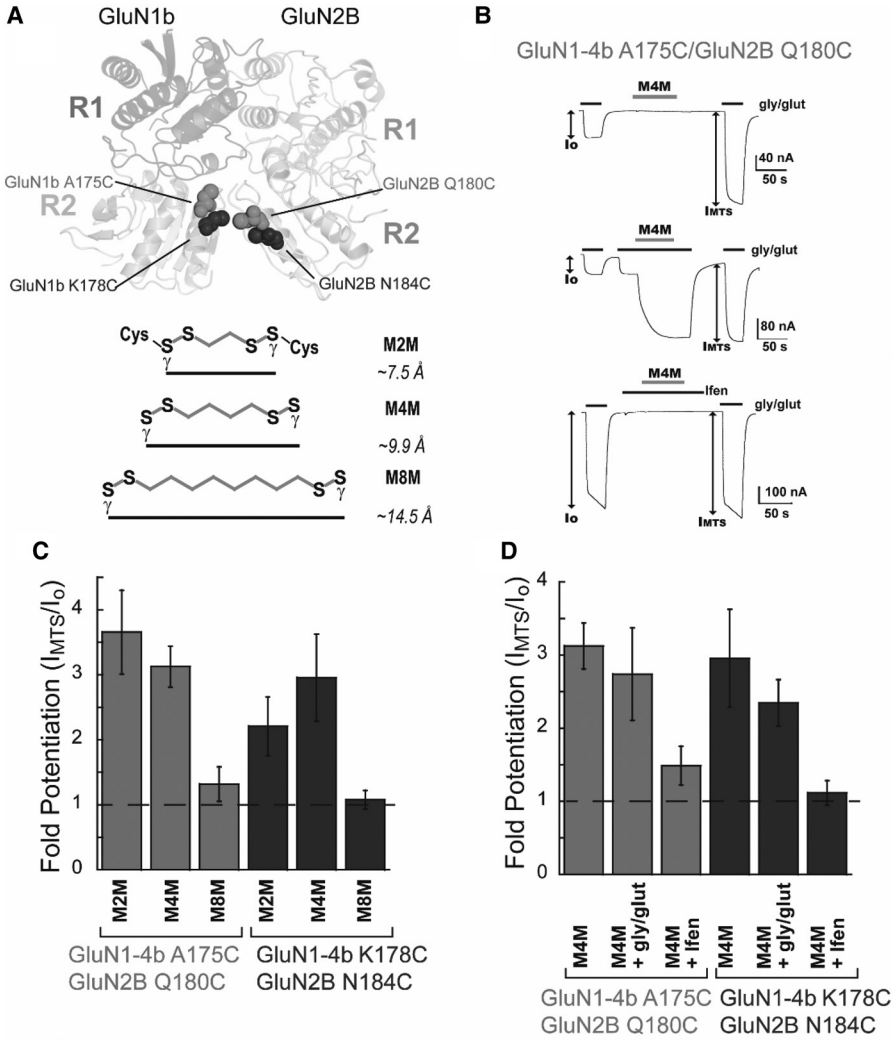


Figure 3. Conformational trap shows the apo-GluN1b-GluN2B ATD structure to represent "active" form. (*A*) Location of engineered cysteines in the crystal structure of the apo-GluN1b-GluN2B ATD. The cysteine mutant pairs, GluN1-4b Ala175Cys/GluN2B Gln180Cys (in green spheres) and GluN1-4b Lys178Cys/GluN2B Asn184Cys (in blue spheres), are coexpressed in *Xenopus* oocytes and cross-linked by bifunctional methanethiosulfonate (bi-MTS) with different linker lengths (M2M, M4M, and M8M). (*B*) Application of 200 μM M4M in the presence or absence of 100 μM agonists (glycine [gly]/glutamate [glut]) potentiates the macroscopic current measured at the holding potential of −60 mV by two-electrode voltage clamp (TEVC). No potentiation was observed when M4M was applied in the presence of ifenprodil (Ifen). Shown here are the representative recording profiles for the GluN1-4b Ala175Cys/GluN2B Gln180Cys pair. (*C*,*D*) "Fold Potentiation" is presented as I_{MTS}/I_o as measured in *B* for bi-MTS with different linker lengths (*C*) and M4M applied in different functional states (*D*) for both of the cysteine mutant pairs tested here. Error bars represent ±SD for data obtained from at least five different oocytes per experiment.

the effect of the bi-MTS reagents on the macroscopic current of NMDAR by two-electrode voltage clamp (TEVC). We initialized this experiment by testing bi-MTS with the four-carbon linker (M4M in Fig. 3A) because the estimated distances between the γ-sulfur atom of the mutated cysteines in GluN1b Ala-175Cys/GluN2B Gln180Cys and GluN1b Lys178Cys/GluN2B Asn184Cys in apo-GluN1b-GluN2B ATD are ~10 Å and ~9 Å, respectively, roughly matching the length of M4M. The application of M4M to the GluN1b Ala175Cys/GluN2B Gln180Cys and GluN1b Lys178Cys/GluN2B Asn184Cys mutants in the presence or absence of glycine and glutamate potentiates the NMDAR currents by approximately three- to

fourfold (Fig. 3B,C). No such effect is observed when the cysteine mutants of one subunit are coexpressed with the WT of the other subunit, indicating that the observed functional effect is specific to the engineered cysteines. We suggest that this potentiating effect by the bi-MTS conformational trap favored the NMDAR ion channel to reside in the "active" form. The potentiation effect was also observed when M2M was applied to both of the above mutant pairs, indicating that the GluN1b-GluN2B distance in R2 may move even closer than that observed in the crystal structure. In contrast, when adding M8M, a bi-MTS agent that is 4–5 Å longer than the intercysteine distances observed in the apo-GluN1b-GluN2B ATD, no potentiating effect was observed, supporting the view that the distance between the R2 lobes of GluN1b-GluN2B must be reduced during activation (Fig. 3C). Finally, when M4M was applied in the presence of ifenprodil, we observe little or no potentiating effect, indicating that it traps the active conformation of GluN1b-GluN2B ATDs but not the inhibited conformation as represented by the crystal structure of the ifenprodil-GluN1b-GluN2B ATD (Fig. 1B,D). Taken together, these experiments indicate that the protein conformation observed in the crystal structure of the apo-GluN1b-GluN2B ATD likely represents the active conformation that facilitates ion channel opening.

Cryo-EM Study of NMDARs to Probe Conformational Alterations

The crystal structure we reported in 2014 represents the GluN1a/GluN2B NMDARs in an allosterically inhibited state in which agonists glycine and glutamate occupy the ligand-binding pockets in the ligand-binding domain (LBD) and an allosteric modulator resides at the ATD. The ion channel at the transmembrane domain (TMD) is closed. Our effort to obtain the structure representing the active conformation has not gone well with X-ray crystallography because of the low quality of diffraction data. Thus, we turned to cryo-EM to analyze the sample in the presence of glycine and glutamate but no ifenprodil to capture the active state of the NMDAR. Most importantly, we sought to capture the protein conformation represented in the apo-GluN1b-GluN2B ATD in the context of the intact GluN1b-GluN2B NMDARs. We obtained cryo-EM structures of the intact heterotetrameric rat GluN1b-GluN2B NMDAR ion channel in the presence of glycine and L-glutamate and in the absence of ifenprodil, the allosteric inhibitor, in collaboration with Nikolaus Grigorieff at Janelia Research Campus/HHMI. The cryo-EM data were collected with a Titan/KRIOS 300-keV microscope coupled with the K2 direct electron detector. The images were corrected for electron-induced movement, and single-particle analysis was conducted to reconstruct the 3D structure. The size of NMDARs (380 kDa) is large enough to conduct reliable 3D reconstruction, thus indicating that cryo-EM may be a superior method to obtain structures in the medium range resolution (5–10 Å).

The cryo-EM structures were reconstructed at resolutions higher than 7 Å, thus permitting us to reliably recognize secondary structure elements (Fig. 4). The cryo-EM structures show conservation of general features observed in the recent full-length NMDAR crystal structures we reported last year, including a dimer of GluN1-GluN2B heterodimers arrangement at the ATD and LBD layers, the domain swap between ATD and LBD, and pseudo-fourfold symmetrical subunit arrangement at the TMD. Importantly, 3D classification of the cryo-EM data revealed patterns of conformational variations (Fig. 4). Overall, there are roughly three distinct conformations, which we define as "nonactive1," "nonactive2," and "active." When compared with the crystal structure of the ifenprodil, glycine, and L-glutamate complexes, all of the 3D classes contain a GluN2B ATD open bilobed architecture with an ~14°–21° opening similar to the crystal structure of the apo-GluN1b-GluN2B ATD. This opening of the GluN2B ATD results in a separation of the GluN1-GluN1 ATDs by as much as ~29 Å in the intact NMDAR compared with the ifenprodil-bound form (Fig. 4). Thus, relative to the LBD and TMD, it is the R1 lobe moving to close the bilobed architecture of the GluN2B ATD upon ifenprodil binding, and that closes the gap between the two GluN1 ATDs to inhibit receptor activity.

The two conformations, nonactive 1 and nonactive 2, do not display the ~15° rotation of the GluN1b and GluN2B subunits relative to one another at the ATD as observed in the crystal structure of the apo-GluN1b-GluN2B ATD, which represents the active state. The arrangements of the dimer of the

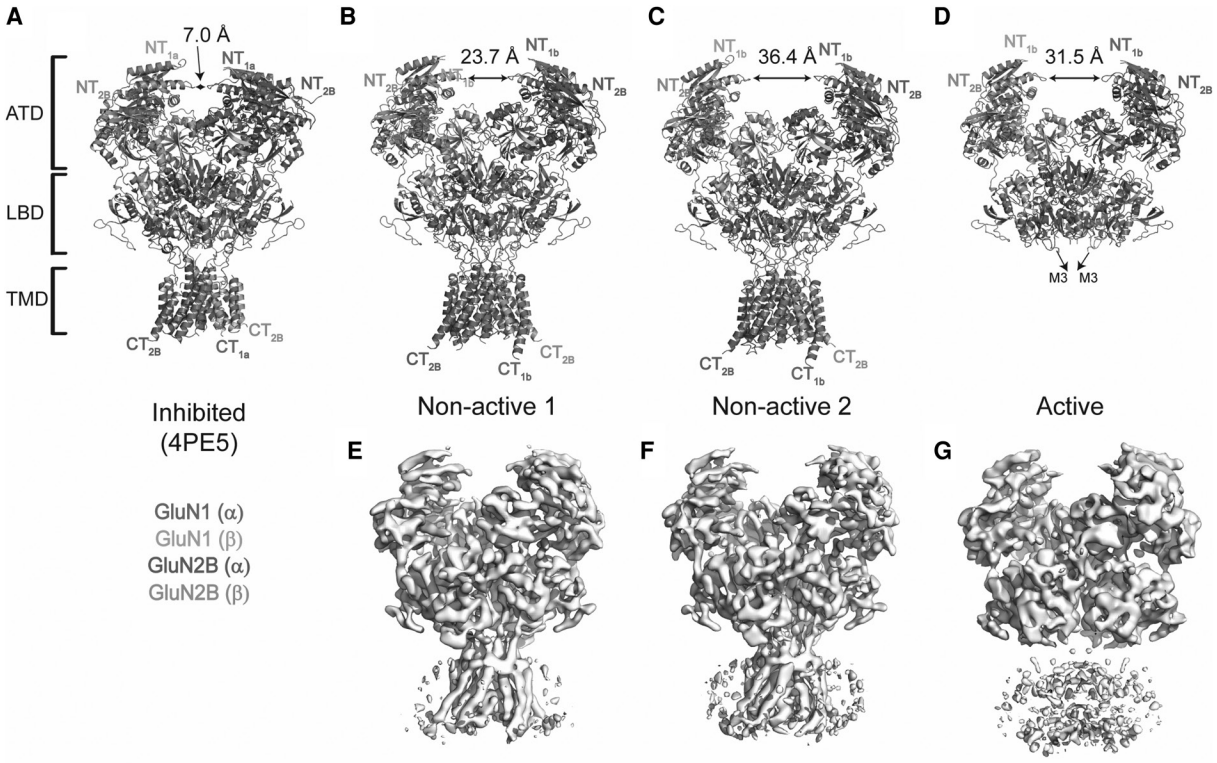


Figure 4. Cryo-electron microscopy (cryo-EM) structures of GluN1b/GluN2B NMDA receptors in the presence of glycine, glutamate, and ifenprodil. The resolution of the structures is sufficiently high to identify secondary structures (*E–G*). The crystal structure solved in the presence of glycine, glutamate, and ifenprodil we reported in 2014 (*A*) is compared with three-dimensional (3D) classes of cryo-EM structures solved in the presence of glycine and glutamate (*B–D*).

GluN1b-GluN2B ATD dimers (Fig. 5), as well as the dimer of GluN1b-GluN2B LBD dimers, remains similar to those observed in the crystal structure ("inhibited" conformation). Consequently, the ion channel pores at the TMD remain closed, confirming that both cryo-EM classes likely represent nongating or nonactive conformations. The difference in nonactive 1 and nonactive 2 is the extent of bilobe opening in the GluN2B ATD, where nonactive 2 has a 7° more open conformation resulting in ~13 Å larger separation between the GluN1 ATDs (Fig. 6). Although we tentatively call these two conformations nonactive, it remains uncertain whether they represent functional states equivalent to the "preactivated" state observed in non-NMDARs or a "desensitized" state.

One of the cryo-EM classes shows an open cleft of the bilobed GluN2B ATD architecture by ~21° and a GluN1b-GluN2B heterodimeric subunit rotated by ~11°, strikingly similar to the apo-GluN1b-GluN2B ATD crystal structure representing the "active ATD conformation" (Fig. 4). In the heterotetrameric NMDAR, the GluN1b-GluN2B heterodimer pairs rotate by 10.6° in opposite directions (Fig. 5). Importantly, this 3D class containing the active conformation of the ATD also shows large differences in the subunit arrangement of LBDs compared with the other 3D classes that contain a nonactive ATD, and is also different from the recent crystal structures of the glycine, L-glutamate, and ifenprodil complexes. Specifically, when transitioning from the nonactive 2 to active conformation, the two pairs of GluN1b-GluN2B LBD heterodimers rotate by ~13° (Fig. 6). These subunit movements in the LBD cause movement of the residues GluN1b Asp744 and GluN2B Gly724 at the LBD-TMD linkers (Fig. 5). For example, compared with the nonactive 2 class, the ring formed by GluN1b Asp744 and GluN2B Gly724 in the active class dilates by 7.6 Å and 5 Å, respectively (Fig. 5), which is likely the movement resulting in gating of the ion channel. Furthermore,

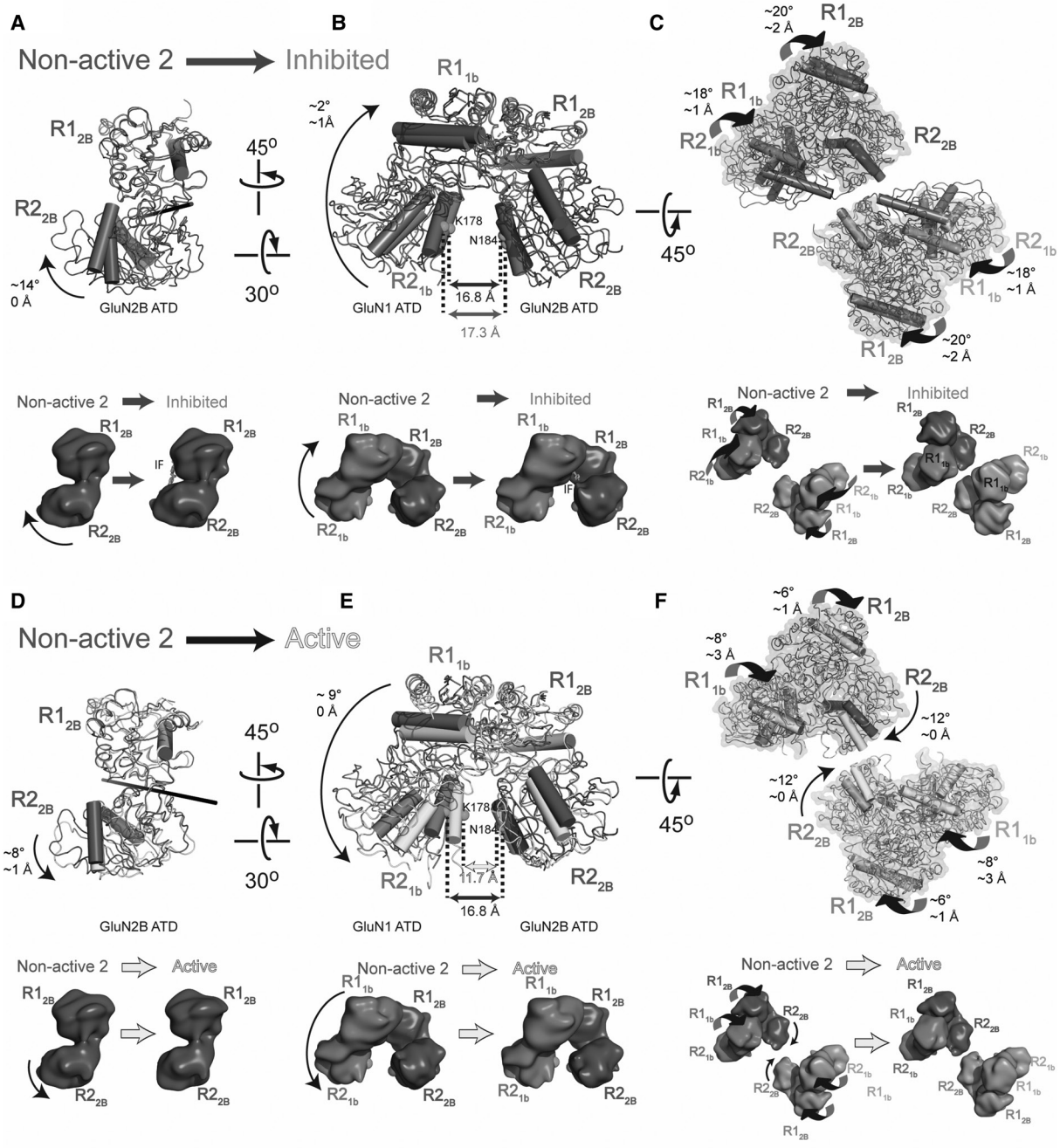


Figure 5. Changes in conformation and heterotetrameric subunit arrangement in ATD during ifenprodil inhibition and receptor activation. Comparison of 3D classes determined by cryo-EM shows movements of the bilobed architecture in GluN2B ATD (*A*,*D*), and rearrangement of the GluN1b-GluN2B ATD heterodimer (*B*,*E*) and heterotetramer (*C*,*F*) during transition from the agonist-bound "nonactive" conformation (GluN1 (α), GluN1 (β), GluN2B (α), and GluN2B (β) in red, orange, blue, and cyan, respectively) to the ifenprodil-bound "inhibited" conformation (gray) (*A*–C) or to the "active" conformation (yellow) (*D*–*F*). (*A*,*D*) Superposition of GluN2B ATD R1 lobes shows relative movement of R2 that rotates around the axis (black rods) in different functional states. The indicated arcs show the direction of the rotation. The rotational and translational movements are measured using PyMOL. (*B*,*C*,*E*,*F*) GluN1b-GluN2B from different 3D classes are superposed onto one another at the R1 lobe of GluN2B ATD and probed for rearrangement in the pattern of subunit arrangement. The distance of the R2 lobes in the GluN1b-GluN2B heterodimers is measured between GluN1b Lys178 and GluN2B Asn184 (*B*,*E*; green spheres).

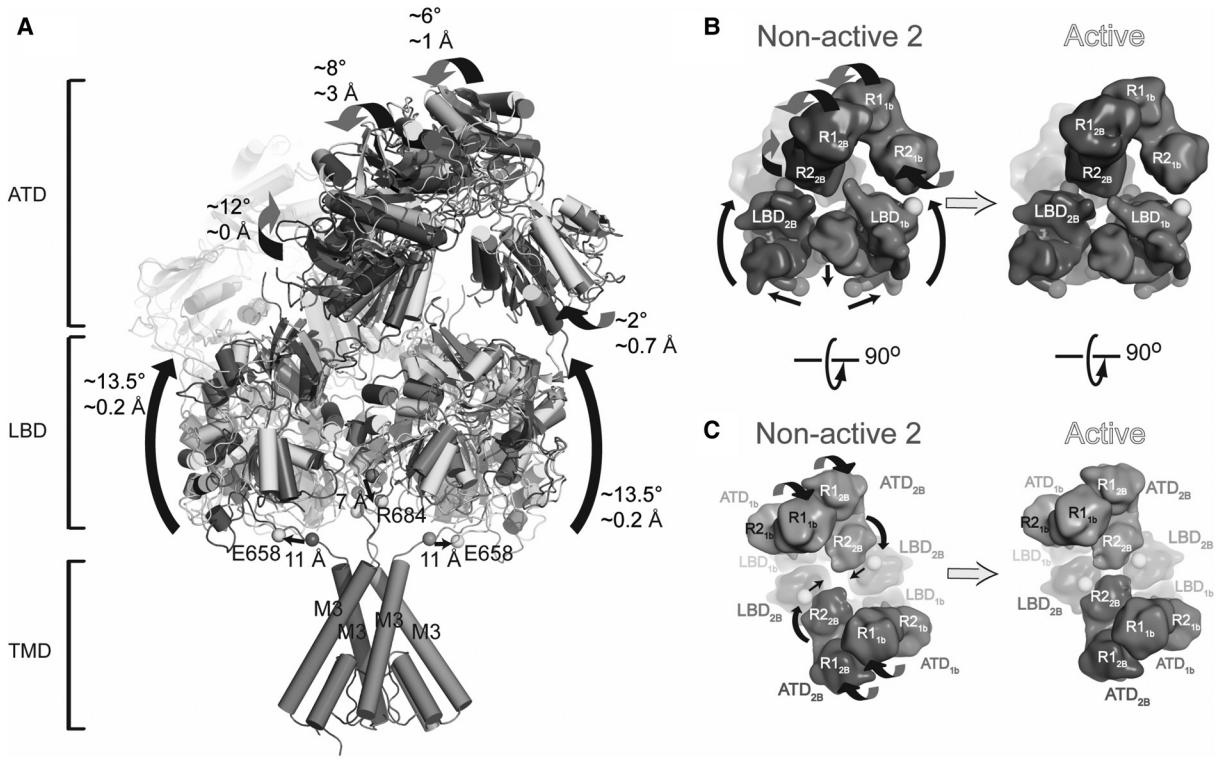


Figure 6. Concerted movement of the ATD and ligand-binding domain (LBD) opens the gate. (*A*) Structural comparison of NMDA receptors in the "nonactive 2" conformation (GluN1 (α), GluN1 (β), GluN2B (α), and GluN2B (β) in red, orange, blue, and cyan, respectively) and the "active" conformation (yellow) is performed by aligning the centers of mass (COMs) of the ATD heterotetramer, LBD heterotetramer, and individual LBDs. Here, the structure is rotated ~45° clockwise along the *y*-axis relative to those in Fig. 3 and viewed from the side. The M1 and M4 helices from the transmembrane domain (TMD) of the nonactive 2 conformation are omitted for clarity. The indicated arcs indicate the direction of the rotation of the LBDs and R1 and R2 lobes of the ATDs for both GluN1b and GluN2B when changing from the nonactive 2 conformation to the active conformation. The first-ordered residues on the linker between the M3 helices on TMD and the LBD in the active state structure (GluN1b Arg 684 and GluN2B Glu 658) are shown as spheres on both structures, and the direction of the transition of these residues from the nonactive 2 to active conformations is indicated by arrows. (*B*,*C*) Schematic diagram highlighting the conformational changes at the extracellular domains during the transition from the nonactive 2 conformation to the active conformation, viewed from side of the intact tetramer (*B*) and top of the ATD (*C*). GluN1b Arg 684 and GluN2B Glu 658 are shown as green spheres and the residues at the ATD-LBD linker (GluN1b Ser 416 and GluN2B His 405) are shown as yellow spheres.

when focusing on the residues located right above the pore formed by the M3 TMD helices, the concerted movement between the ATD and LBD going from nonactive 2 to active described above results in the vertical movement of GluN1b Arg684 and the lateral separation of GluN2B Glu658 by 7 Å and 11 Å, respectively, to dilate the channel gate (Fig. 6). Thus, this cryo-EM class is structurally and functionally consistent with an active conformation for GluN1/GluN2B NMDARs. Perhaps caused by the intrinsic disorder, the density for the TMD for the active conformation is not resolved in sufficient detail to directly observe opening of the ion channel. Nevertheless, the structure clearly shows the movement of the GluN1b-GluN2B heterotetramer, which is likely coupled to the ion channel opening. Finally, the comparison of the cryo-EM classes with GluA2 AMPA receptor in the "pre-open" state shows that there is a greater difference between the active and pre-open states than between the nonactive and pre-open states, consistent with the observation that the TMD ion channel is closed in the nonactive state as in the pre-open state in GluA2 AMPA receptors.

PUBLICATIONS

Malinauskas T, Furukawa H. 2016. Production of heteromeric transmembrane receptors with defined subunit stoichiometry. *Structure* **24:** 653–655.

Regan MC, Furukawa H. 2016. Deeper insights into the allosteric modulation of ionotropic glutamate receptors. *Neuron* **91:** 1187–1189.

Romero-Hernandez A, Simorowski N, Karakas E, Furukawa H. 2016. Molecular basis for subtype specificity and high-affinity zinc inhibition in the GluN1-GluN2A NMDA receptor amino-terminal domain. *Neuron* **92:** 1324–1336.

Tajima N, Karakas E, Grant T, Diaz-Avalo, R, Grigorieff N, Furukawa H. 2016. Activation of NMDA receptors and the mechanism of inhibition by ifenprodil. *Nature* **534:** 63–68.

ASSEMBLY AND FUNCTION OF NEURAL CIRCUITS IN THE NEOCORTEX

Z.J. Huang X. An, S.W. Galbavy, D. Huilgol, S. Kelly, K. Krishnan, J. Levine, J. Lu, K. Matho, M.A. Paul, R. Raudales, P. Wu

We study the developmental assembly and functional organization of neural circuits in the neocortex. The neocortex consists of a constellation of functional areas that form a representational map of the external and internal world. These areas are interconnected into elaborate information processing networks that guide behavior. Cortical processing streams and output channels are mediated by a large set of glutamatergic pyramidal neuron (PyN) types, and functional neural ensembles are regulated by a diverse set of GABAergic interneuron types. We use genetic approaches in the mouse to systematically target cell types and facilitate the application of modern techniques for exploring cortical circuits. Among GABAergic interneurons, the chandelier cell (ChC) is one of the most distinctive cell types that controls PyN firing at the axon initial segment. We are studying the developmental specification, activity-dependent circuit integration, and functional connectivity of ChCs, an entry point toward understanding a local circuit module. Regarding PyNs, we are systematically characterizing the developmental origin, axon projection pattern, and input connectivity of multiple classes of PyN types, focusing on the forelimb motor cortex. We combine a range of approaches that include genetic and viral engineering, genetic fate mapping, gene expression profiling, cellular imaging, electrophysiology, and behavior analysis. Recently, we begin to integrate these studies in the context of the motor cortex control of forelimb movements.

Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex

Systematic genetic access to GABAergic cell types will facilitate studying the function and development of inhibitory circuitry. However, single gene-driven recombinase lines mark relatively broad and heterogeneous cell populations. Although intersectional approaches improve precision, it remains unclear whether they can capture cell types defined by multiple features. Here, we show that combinatorial genetic and viral approaches target restricted GABAergic subpopulations and cell types characterized by distinct laminar location, morphology, axonal projection, and electrophysiological properties. Intersectional embryonic transcription factor drivers allow finer fate mapping of progenitor pools that give rise to distinct GABAergic populations, including laminar cohorts. Conversion of progenitor fate restriction signals to constitutive recombinase expression enables viral targeting of cell types based on their lineage and birth time. Properly designed intersection, subtraction, conversion, and multicolor reporters enhance the precision and versatility of drivers and viral vectors. These strategies and tools will facilitate studying GABAergic neurons throughout the mouse brain.

Transcriptional Architecture of Synaptic Communication Delineates Cortical GABAergic Neuron Identity

Understanding the organizational logic of neural circuits requires deciphering the biological basis of neuron type diversity and identity, but there is no consensus on defining a neuron type. We analyzed single-cell transcriptomes of anatomically and physiologically characterized cortical ground truth populations and conducted a computational genomic screen for transcription profiles that distinguish them. We discovered that cardinal GABAergic neuron types are delineated by a transcriptional architecture that encodes their synaptic communication patterns. This architecture comprises six categories of approximately 40 gene families, including cell adhesion molecules, transmitter-modulator receptors, ion channels, signaling proteins, neuropeptides

and vesicular release components, and transcription factors. Combinatorial expression of select members across families shapes a multilayered molecular scaffold along the cell membrane that may customize synaptic connectivity patterns and input–output signaling properties. This molecular genetic framework of neuronal identity integrates cell phenotypes along multiple axes and provides a foundation for discovering and classifying neuron types.

Selective Inhibitory Control of PyN Ensembles and Cortical Subnetworks by ChCs

The neocortex comprises multiple information processing streams mediated by subsets of glutamatergic pyramidal cells (PCs) that receive diverse inputs and project to distinct targets. How GABAergic interneurons regulate the segregation and communication among intermingled PC subsets that contribute to separate brain networks remains unclear. Here, we show that a subset of GABAergic ChCs in the prelimbic cortex (PL), which innervates PCs at spike initiation site, selectively control PCs projecting to the basolateral amygdala (BLAPC) compared with those projecting to contralateral cortex (CCPC). These ChCs, in turn, receive preferential input from local and contralateral CCPCs as opposed to BLAPCs and BLA neurons (the PL-BLA network). Accordingly, optogenetic activation of ChCs rapidly suppresses BLAPCs and BLA activity in freely behaving mice. Thus, the exquisite connectivity of ChCs not only mediates directional inhibition between local PC ensembles but may also shape communication hierarchies between global networks.

Genetic Targeting of PyN Subtypes in Mouse Neocortex

A major obstacle to studying the development and organization of neural circuits in the cerebral cortex is the diversity of neuron types. Glutamatergic PyNs constitute ~80% of cortical neurons; are endowed with large capacity for information coding, storage, and plasticity; and carry the output of cortical computation. PyNs consist of diverse subtypes based on their specific laminar locations, axonal projection patterns, and gene expression profiles. Subsets of PyNs form multiple and hierarchical subnetworks of information processing, with distinct output channels to cortical and subcortical targets that subserve sensory, motor, cognitive, and emotional functions. Importantly, PyN subtypes are differentially affected in various neuropsychiatric and neurodegenerative disorders. We have begun to build a comprehensive genetic tool set for major PyN subtypes in the mouse. We are using intersection, subtraction, and inducible strategies to target PyN subtypes. We have already generated 12 knock-in driver lines that target restricted populations of corticothalamic, corticofugal, corticostriatal, and corticocortical PyNs. Combined with anterograde and retrograde viral vectors, we begin to achieve increasingly specific targeting of PyN subpopulations.

Distinct Progenitor Programs Underlie Striatal Compartments and Circuit Organization

The functional circuitry of the striatum is characterized by two organizational plans: one dividing the striatum into the striosome and matrix compartments, which are thought to mediate evaluation and action, and the second dividing it into direct and indirect circuits, which are thought to promote or suppress behavior. The relationship between these two organizations is unknown, leaving a conceptual gap in understanding the corticobasal ganglia system. We used genetic fate mapping to fill this gap. We show that striosome-matrix compartmentalization arises from a division among lateral ganglionic eminence progenitors giving rise to striatal projection neurons (SPNs). An early set of radial glia progenitors (RGSs) produces striosomal SPNs mainly through direct neurogenesis, whereas later-activated progenitors (RGMs) generate matrix SPNs through amplification of intermediate progenitors. Remarkably, direct and indirect pathway SPNs arise within both RGS and RGM pools, suggesting that striosome-matrix architecture is the fundamental plan from which integrated basal ganglia circuitry emerges.

Activity-Dependent Regulation of ChC Circuit Integration

Proper distribution of cortical inhibitory interneurons is crucial in regulating the balance of local circuitry and routing of cortical information flow. Although progress has been made in understanding the specification and long-range migration of cortical interneurons,

the mechanisms that regulate their density and cortical integration remain unclear. The ChCs innervate PyNs at the axon initial segment—the spike initiation site—thus, the proper distribution of ChCs may play a major role in shaping functional PyN ensembles. After their specification in the late embryonic medial ganglionic eminence, ChCs migrate with stereotyped route and schedule to arrive at designated cortical layers by the end of the first postnatal week. Surprisingly, we found that young ChCs then undergo massive apoptosis throughout the developing cortex between P7 and P14. In particular, we observed an additional 50% reduction of ChCs at the border region between primary visual cortex (V1) and lateral secondary visual cortices (V2L) compared with the surrounding areas. The V1-V2L border receives input from callosal neuron (CN) axons of the contralateral visual cortex. By blocking CN growth or activity using Kir2.1 or inhibitory DREADD, we found that contralateral CN activity between P7 and P14 regulates the survival and density of ChCs at the border region. Further, ectopic CN axon projection induced by monocular enucleation at birth resulted in corresponding ectopic ChC elimination. Importantly, blocking retinal activity by monocular TTX injection between P7 and P14 had no effect on CN axon projection, but reduced ChC elimination at ipsilateral V1-V2L border. Together, these results suggest that the density of ChCs at V1-V2L border is regulated by contralateral CN axons and retinal inputs. CNs receive inputs from the temporal retina that represents the central visual field and contribute to the seamless fusion of the left and right visual field. We hypothesize that activity-dependent elimination of ChCs at V1-V2L border may contribute to the development of a fast bilateral signaling pathway between the CNs that correlates their discharges when stimulated by the same orientation stimulus. Current effort is directed toward understanding the role of ChCs in the interhemispheric callosal projection pathway that integrates the cortical representation of the central visual field.

Genetic Dissection of Motor Cortex Circuits That Control Forelimb Movement

The motor cortex has been implicated in the volitional control of forelimb movements, a rich set of behavioral skills that allow rodents to manipulate their environment according to sensory inputs, internal goals, and experience. However, the underlying circuitry mechanisms remain largely unexplored. In particular, different types of glutamatergic PyNs (e.g., defined by their axon projection patterns) likely form multiple interareal processing streams and output channels that mediate sensorimotor integration and descending motor commands, but the cellular basis of these neural networks and output pathways is poorly understood. We have initiated a systematic genetic and viral targeting of PyNs based on their lineage, birth order, gene expression, and axon projection. Genetic access to these PyNs provides unprecedented opportunities to dissect motor cortex circuitry and output pathways. We are using multiple driver lines combined with anterograde and retrograde viral methods to examine the projection patterns and connectivity of several infragranular layer subpopulations. In addition, we are using optogenetic manipulation of these PyNs to explore whether they mediate specific components of forelimb movements in head-fixed awake mice.

MECP2 Regulates Cortical Plasticity Underlying a Learned Behavior in Adult Female Mice

Neurodevelopmental disorders are marked by inappropriate synaptic connectivity early in life, but how disruption of experience-dependent plasticity contributes to cognitive and behavioral decline in adulthood is unclear. Here, we show that pup-gathering behavior and associated auditory cortical plasticity are impaired in female $Mecp2^{het}$ mice, a model of Rett syndrome. In response to learned maternal experience, $Mecp2^{het}$ females showed transient changes to cortical inhibitory networks typically associated with limited plasticity. Averting these changes in $Mecp2^{het}$ through genetic or pharmacological manipulations targeting the GABAergic network restored gathering behavior. We propose that pup-gathering learning triggers a transient epoch of inhibitory plasticity in auditory cortex that is dysregulated in $Mecp2^{het}$. In this window of heightened sensitivity to sensory and social cues, Mecp2 mutations suppress adult plasticity independently of their effects on early development.

PUBLICATIONS

Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. 2016. Exploiting single-cell expression to characterize co-expression replicability. *Genome Biol* **17:** 101.

He M, Tucciarone J, Lee S, Nigro MJ, Kim Y, Levine JM, Kelly SM, Krugikov I, Wu P, Chen Y, et al. 2016. Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. *Neuron* **91:** 1228–1243.

Wall NR, De La Parra M, Sorokin JM, Taniguchi H, Huang ZJ, Callaway EM. 2016. Brain-wide maps of synaptic input to cortical interneurons. *J Neurosci* **36:** 4000–4009.

In Press

Krishnan K, Lau BY, Ewall G, Huang ZJ, Shea SD. 2017. MECP2 regulates cortical plasticity underlying a learned behaviour in adult female mice. *Nat Commun* **8:** 14077.

Lu J, Tucciarone J, Padilla-Coreano N, He M, Gordon JA, Huang ZJ. 2017. Selective inhibitory control of pyramidal neuron ensembles and cortical subnetworks by chandelier cells. *Nature Neurosci* (in press).

NEUROBIOLOGY OF COGNITION AND DECISION-MAKING

A. Kepecs N. Bobrowski-Khoury, Q. Chevy, E. Demir, E. Gibson, T.S. Gouvea, Q.A. Li, S.J. Li, P. Masset, T. Ott, H.J. Pi, T. Pinkhasov, S. Starosta, J.F. Sturgill, C. Toschi

Understanding the mysteries of cognition is an age-old aspiration of humankind. Neuroscience has made great strides toward understanding systems for sensory processing and motor output. Yet, studying cognition at the level of neurobiological substrate seemed hopeless until recently. The goal of our laboratory is to bridge the chasm between cognitive and circuit neuroscience by applying a unique, multifaceted skill set and approach. We translate psychological questions into the language of neuroscience by developing quantifiable, well-controlled behavioral tasks for rodents. We then couple these tasks with targeted, high-resolution, and high-throughput monitoring and manipulation of the neural circuits mediating cognitive behaviors. Given the complexity of animal behavior and the dynamics of the neural networks that produce it, our studies rely on computational models to guide and sharpen the neurobiological questions. Finally, we use human psychophysics to validate our behavioral observations in rodents by linking them with analogous behaviors in humans. Behavioral links to humans can also serve to identify behavioral dimensions that are predictive of disordered mental states, and our goal is to bridge our studies in animals to psychiatric disorders. Using this integrated approach, our long-term goal is to "reverse engineer" the computational and neurobiological principles underlying cognition and decision-making.

In terms of topics, our approach is multifaceted: We study (i) the roles of uncertainty and confidence in decision-making, (ii) foraging decisions about whether to stay or to switch, (iii) the division of labor between different cell types in prefrontal cortex, (iv) how the cholinergic system supports learning, and (v) social decisions that rely on shallow, stereotyped circuits. A unifying theme is the use of cell-type- and pathway-specific perturbations to effect gain and loss of function for particular behavioral abilities. We hope that by identifying the neural processes underlying behavior in "humanized" mouse models of cognition, we can understand what goes awry in the brain during mental illness. Ultimately, we hope these insights will enable us to develop novel therapeutic strategies for psychiatric disorders such as addiction, major depression, schizophrenia, and autism spectrum disorder.

Confidence in Rats, Humans, Brains, and Statistics

T.S. Gouvea, P. Masset, T. Ott

We experience confidence as a feeling, but it is a ubiquitous process for predicting outcomes in the face of the pervasive uncertainty in the world. Knowing one's degree of confidence confers benefits for a broad range of activities from the sophisticated to the mundane: managing a stock portfolio or deciding whether to carry an umbrella. Indeed, appropriate estimates of confidence can drive information-seeking behaviors, learning, and attention by reducing the level of uncertainty. Conversely, the pathological misevaluation of confidence contributes to a wide range of neuropsychiatric conditions, including anxiety, obsessive-compulsive disorder, and addiction. In psychology, confidence judgments have been long studied as a central example of a metacognitive process unique to humans. More recently, we and others attempted to show that nonhuman animals are also capable of confidence judgments. We have developed a set of behavioral tasks and a theoretical framework that rigorously translates the psychological concept of confidence into a formally defined decision variable. Using this approach we identified orbitofrontal cortex (OFC) neurons that encode the confidence associated with a perceptual decision. Further, we have derived a mathematical framework for decision confidence from first principles of statistics (Hangya et al. 2016).

We showed that key properties of statistical decision confidence match human self-reported confidence (Sanders et al. 2016), providing a deep link between objective and subjective notions of confidence. We also showed that reversible lesions of OFC specifically disrupt confidence reporting (Lak et al., *Neuron 84:* 190 [2014]).

Based on these results, we are now ready to pursue the neural circuit basis of confidence. To pursue this, we designed a rodent auditory task, inspired by the well-known random dots task used to probe visual decisions (Sanders et al. 2012). Here, rats must determine which of two Poisson click trains has the higher underlying rate and report their confidence with the amount of time they are willing to invest in each decision. Using this task, we seek to ask two fundamental questions: How are representations of confidence in OFC generated from sensory inputs? How and where do OFC output neurons route this information downstream to guide decision-making? We also wondered whether OFC computes a metacognitive signal, irrespective of the sensory modality used to make the decision. In a dual modality task with interleaved auditory and olfactory decisions, our data show that many OFC neurons encode decision confidence across modalities, revealing an abstract confidence signal.

Beyond our rodent work, we are now in a position to use these quantitative measures of decision confidence in humans, through collaborations, with neuroimaging and genetic approaches. Moreover, by showing that a single confidence measure is applicable to humans and rodents, our results strengthen the case for using the rat as a model system for translational studies in cognition and psychiatry.

Categorical Representations of Decision Variables in OFC

J. Hirokawa, Doshisha University; A. Vaughan

OFC, like other cortical regions, creates internal representations of the external world in the form of neural activity, which is structured to support adaptive behavior. In many cortical regions, individual neurons respond to specific features that are matched to the function of each region and statistics of the world. In OFC, like elsewhere in frontal cortex, neurons display baffling complexity, responding to a mixture of sensory, motor, and other variables. Here, we use an integrated approach to understand the architecture of higher-order cortical representations and show that discrete groups of OFC neurons encode distinct decision variables. Using rats engaged in a complex task, combining perceptual and value-guided decisions, we found that OFC neurons can be grouped into distinct, categorical response types. These categorical representations map directly onto decision variables, such as reward size, decision confidence, and integrated value, in a choice model explaining our task. We propose that, like sensory neurons, frontal neurons form a sparse and overcomplete population representation aligned to the natural statistics of the world—in this case, spanning the space of decision variables required for optimal behavior.

Neural Circuit Logic of OFC during a Postdecision Confidence Task

J. Hirokawa, Doshisha University; P. Masset, T. Ott
[in collaboration with M. Lagler and T. Klausberger, Medizinische Universität Wien]

Neurons in OFC represent an array of decision variables from confidence to reward size and chosen value. However, whether OFC neurons with particular projection targets specialize in transmitting distinct decision variables, and thereby route signals in target-selective manners, remains unknown.

We are taking two complementary approaches to answer this question. First, we target specific projection neurons using retrograde viruses and then use optogenetic stimulation to identify these in electrophysiological recordings. Using this technique, we have found that OFC projections to ventral striatum show a characteristic response in which negative value signals are sustained throughout the intertrial interval to the beginning of the subsequent trial. Second, in collaboration with the Klausberger laboratory, we use juxtacellular labeling to target neurons based on their functional response profiles. Once labeled, neurons are subjected to detailed ex vivo analysis of the axonal projection patterns. Using these techniques, we have begun to record and identify OFC neurons that specifically signal the confidence-dependent waiting time. We expect that the combination of juxtacellular and optogenetically identified extracellular recordings will enable us to determine the cell-type-specific circuit logic of OFC.

Midbrain Dopamine Neurons Signal Confidence-Dependent Prediction Errors during Perceptual Decisions

This work was done in collaboration with A. Lak (University of Cambridge) and M. Sakagami (Tamagawa University).

In the struggle of life, animals survive by following a simple dictum: win big and win often. Finding bigger (e.g., large food reward) and more likely wins is particularly challenging when these are not available in their nearby environment. Inspired by the study of animal behavior, a machine-learning approach called "reinforcement learning" provides a rigorous framework for how to select winning behaviors. The key to reinforcement learning is to adjust the expected reward values associated with each behavior based on the outcomes of one's actions. These adjustments to reward values are based on the discrepancy between the received and predicted value, or the "reward prediction error," signaled by midbrain dopamine neurons. This computation requires a prediction about the value of upcoming outcomes. In value-guided behavioral tasks, this prediction is based on the reinforcement history.

Many types of decisions are not only based on previous outcomes but also ambiguous percepts. For such decisions, predicting outcome values requires resolving the perceptual ambiguity and computing an instantaneous estimate of the probability of choice accuracy that reflects trial-by-trial fluctuations in perception. However, it has been unclear whether dopamine neurons are able to signal prediction errors incorporating such estimates. We used a computational model to show that dopamine neurons in monkeys performing a perceptual decision task comply with the predictions of a reinforcement learning model, extended with a belief state about the perceptual stimulus. As a feature of the belief state computation, this model generated trial-by-trial predictions of the probability of decision correctness, termed "decision confidence." Similarly, we found that dopamine responses did not simply reflect the average value of the perceptually ambiguous stimuli, but were predictive of the trial-to-trial fluctuations in decision confidence. These confidence-dependent dopamine responses emerged before the monkeys' choice initiation and, thus, could potentially impact impending actions. Finally, by manipulating the reward size, we showed that dopamine neurons reflect both the magnitude of the upcoming reward and the confidence in achieving it. Thus, dopamine responses convey the teaching signals that are also appropriate for learning perceptual decisions.

Neural Circuits and Quantitative Measures of Impulsive Choice in Mice

H.J. Pi, T. Pinkhasov, S. Starosta

Impulsivity, a failure of inhibition in deliberative decision-making, is present in many psychiatric disorders and significantly increases the risk of suicide, violence, and criminal behavior. Yet, the underlying neuronal mechanisms remain elusive. Whether a particular decision made too early is impulsive is difficult to determine because misevaluation of expected outcome or misestimation of time could lead to similar consequences. Therefore, as an initial step toward understanding the underlying neural circuits, we sought to develop a task that isolates the contribution of impulsivity to individual choices and separates it from reward valuation. We focus on a specific subtype of impulsivity, impulsive actions, defined as the failure to withhold a response to a stimulus despite negative consequences. We are examining three brain regions implicated in different facets of impulsivity because of their involvement in cognitive and/or motoric inhibition: ventral tegmental area, a major dopaminergic center; the dorsal raphe nuclei, a serotonergic nucleus; and anterior cingulate cortex. We are using optogenetic activation to disentangle the circuit-specific contributions to impulsivity from these brain regions. Ultimately, we hope our circuit-based understanding of impulsivity will contribute to the design of circuit-specific treatments for impulsivity disorders.

Circuit and Behavioral Functions of Cortical Chandelier Neurons

Q.A. Li [in collaboration with Z.J. Huang and J. Tucciarone, CSHL]

Chandelier cells (ChCs) are perhaps the most unique GABAergic interneurons in cortex, and have fascinated neuroscientists for a long time. GABAergic interneurons are widely accepted to be inhibitory cells. ChCs exclusively innervate the axon initial segment of excitatory pyramidal neurons, the site for action potential generation. Consequently, they have been

thought to function as "veto cells" by suppressing spike initiation in pyramidal neurons. This assumption has been called into question by in vitro studies showing that ChCs can, paradoxically, be excitatory. Until recently, it was not possible to study these neurons in the intact brain, and therefore even elementary questions about function, such as whether they are excitatory or inhibitory, remain unanswered. Josh Huang's lab's (CSHL) successful development of a novel genetic mouse line has enabled the reliable targeting of ChCs, and been a game changer for the field. The mouse line enables us, for the first time, to probe both the circuit impact and behavioral function of ChCs in behaving mice. First, we will label ChCs in mPFC with ChR2 to make them activable. Then, we will combine electrophysiological and optogenetic techniques to identify and manipulate the intracortical circuit between ChCs and postsynaptic pyramidal neurons in vivo. We aim to provide direct evidence of ChCs' impact on their local circuit when are they recruited during behavior.

Cortical VIP Interneurons, Disinhibitory Control, and Reinforcement Learning

Q. Chevy, H.J. Pi [in collaboration with Z.J. Huang, CSHL; Z. Szadai and B. Rozsa, Hungarian Academy of Sciences]

We have recently identified a disinhibitory cortical circuit motif that appears to be a conduit for fast neuromodulatory action in cortex. This circuit is controlled by a class of inhibitory interneurons that express vasoactive intestinal polypeptide (VIP) and inhibit other interneurons, thereby disinhibiting a subpopulation of principal neurons. Functionally, we showed that VIP interneurons in the auditory cortex are recruited in response to specific reinforcement signals such as reward and punishment. Based on these observations we wondered whether the VIP-controlled cortical circuit has a general function across cortex, and what these neurons might signal beyond reinforcers.

To explore the generality of these observations across cortex regions, we are collaborating with B. Rozsa (KOKI, Hungarian Academy of Sciences), who developed a state-of-the-art three-dimensional (3D) random-access AOD two-photon imaging system that allows recordings of the sparse VIP population across large volumes (~$600 \times 600 \times 800$ μm). In preliminary data, we observed that most VIP neurons are activated by reward and punishment across multiple cortical regions, suggesting that their behavioral recruitment has a cortex-wide function in reinforcement learning.

We have also begun to record VIP neurons in a cued-outcome task to determine what aspect(s) of reinforcement they signal. Do they respond simply to the delivery of primary reinforcers, reinforcement prediction errors, or reinforcement surprise? Our preliminary data suggest that cue responses emerge with learning and are proportional to outcome value, suggesting a more sophisticated role in reinforcement learning. By establishing the detailed circumstances under which VIP interneurons are recruited and identifying their generality and circuit mechanisms, we expect that these studies will reveal fundamental principles about cortical microcircuits that are applicable across cortical regions.

Neural Representation of Social Decisions and Rewards

N. Bobrowski-Khoury, E. Demir [in collaboration with R. Axel, Columbia University]

Social behavior is integral to animals' survival and reproduction; social deficits are at the heart of cognitive disorders such as autism spectrum disorder, which have proven profoundly difficult to study in model organisms. Mice, like humans, are social animals. To interact, cooperate, and compete with others, mice have to collect information about each other's identity, fertility, and likely intent. We would like to understand how social information is represented, computed, and used by mice. In rodents, a main source of information for social decision-making and reward valuation is the chemosensory system. The neural circuits supporting these tend to be shallow, from sensory input to motor actions, and highly stereotyped, enabling the systematic dissection of this system.

Male mice deposit scents in the environment that elicit attraction behavior in females. One component of these scents is darcin, a mouse urinary protein that is sufficient to induce innate attraction in sexually receptive female mice and also acts as an unconditioned stimulus in associative learning paradigms. We have identified a neural circuit that extends from the vomeronasal organ to the medial amygdala and mediates the innate response to darcin. Genetic silencing of either the accessory olfactory bulb or the medial amygdala eliminates the innate attraction to darcin.

Optical reactivation of darcin-activated medial amygdala neurons elicits attraction behavior.

We are also interested in understanding basic rules that mice use to choose partners. For this purpose, we have developed a psychophysical social behavior task, the "social carousel," inspired by perceptual psychophysics and game theoretic traditions that have been instrumental in studying other facets of cognition. Our task enables reliable, quantitative, and high-throughput analysis of social interactions in mice. This task contrasts appetitive rewards with opportunities for social interaction. Water-deprived animals are trained to alternate between two platforms: One platform provides a small water reward, and the other allows for social interaction with caged mice drawn at random from a revolving magazine. Subjects can choose to engage in extended social interactions with the caged mice at the expense of delaying the water reward. This task allows us to infer the "social value" of a mouse based on the trade-off between social interactions and appetitive rewards. In addition, this task is compatible with our electrophysiological studies because it is devised for precise stimulus delivery and reproducible behavioral contingencies.

Cholinergic Basal Forebrain in Attention and Learning

S. Li, J.F. Sturgill [in collaboration with B. Hangya, Hungarian Academy of Sciences]

Basal forebrain (BF) cholinergic neurons constitute a major neuromodulatory system implicated in normal cognition functions, including learning, memory, and attention. Cognitive deficits in Alzheimer's disease, Parkinson's dementia, age-related dementias, and normal aging are correlated with the extent of deterioration of BF cholinergic neurons. Cholinergic projections densely innervate neocortex and release acetylcholine, which is thought to regulate arousal, attention, and learning. However, precise behavioral function of these projections is poorly understood because identified cholinergic neurons have not been recorded during behavior. Therefore, we recorded cholinergic neurons using optogenetic identification in mice performing an auditory detection task requiring sustained attention. We observed a noncholinergic BF population—but not cholinergic neurons—that were correlated with trial-to-trial measures of attention. Instead, cholinergic neurons responded to reward and punishment with unusual speed and precision across two BF nuclei that innervate distinct cortical areas. This reinforcement response invites comparison to dopamine neurons for which a key conceptual advance was that they compute reward prediction error: the difference between reward expectation (as informed by a predictive cue) and the reward received. Therefore, we adopted an analogous behavioral and computational approach to understand the principles governing cholinergic neuron activation. We used fiber photometry to measure bulk GCaMP signal in cholinergic neurons in a cued probabilistic outcome task. After the mice learned the task, cholinergic neurons responded to reward-predicting cues, and the neurons' reward responses were diminished by cued expectations, similar to dopamine neurons. These results reveal that the cholinergic system broadcasts a rapid and precisely timed reinforcement signal that could support fast cortical activation and plasticity.

A Receptor Complementation Strategy to Efficient, Tropism-Free Retrograde Targeting of Neurons

S. Li

Retrogradely transported neurotropic viruses enable researchers to target neurons based on their long-range projections and have thus become indispensable tools for linking neural connectivity with function. However, a major limitation of viral techniques is that they rely on cell-type-specific molecules for uptake and transport. Known as "viral tropism," viruses can only infect cells that express surface receptors compatible with the virus' capsule. As a result, viruses fail to infect variable subsets of neurons, depending on the complement of surface receptors expressed by the neurons. We developed a strategy to overcome this problem by enhancing neurons' expression of receptors compatible with the retrograde virus of interest, in this case, canine adenovirus type 2 (CAV-2). We designed adeno-associated virus (AAV) vectors to cause cells in the source region to express the coxsackievirus and adenovirus receptor (CAR) throughout candidate projection neurons' axonal arbors. Enhancement of CAR expression greatly increased retrograde labeling rates with CAV-2, which we showed in several long-range projections in both rats and mice. Our results

show a strategy to abrogate endogenous viral tropism, and thereby facilitate efficient retrograde viral targeting for functional analysis of neural circuits.

Nanophotonic Silicon Probes for Multisite Optical Stimulation

Q.A. Li [in collaboration with A. Mohanty, M.A. Tadayon, and M. Lipson, Columbia University]

Optogenetic studies involve activating genetically defined neuronal populations with light. Deep-brain optical excitation at a high spatiotemporal resolution would enable activation of specific neural populations and in-depth study of neural circuits. Conventionally, a single fiber is used to flood light into a large area of the brain, limiting spatial resolution. The scalability of silicon photonics enables neural excitation over large areas with single-cell resolution similar to that achieved with electrical probes. However, active control of these optical circuits has not been shown for optogenetics. We worked with the Lipson group to design and test the first active, integrated optical switch for neural excitation in the visible regime (473 nm). This switch should enable control of multiple light beams for deep-brain neural stimulation. Using a silicon nitride waveguide platform, the Lipson group developed a cascaded Mach–Zehnder interferometer (MZI) network located outside the brain to direct light to eight different grating emitters located at the tip of the neural probe. Using integrated platinum microheaters to control the switch output, we achieved on–off extinction ratios of >8 dB and a switching speed of 20 μsec. These devices were tested in vivo in visual cortex for single-neuron optical activation. Directly light-activated neurons showed robust spike firing activities with low first-spike latency and small jitter, indicating successful light delivery. This nanophotonic platform enables high-multiplexed, high-resolution optical stimulation in deep-brain regions. In the future, this technology can be combined with recording probes to enable simultaneous optical stimulation in arbitrary spatiotemporal patterns and electrophysiological recordings, both on a massively parallel scale.

PUBLICATIONS

Hangya B, Sanders JI, Kepecs A. 2016. A mathematical framework for statistical decision confidence. *Neural Comput* **28:** 1–20.

Kobak D, Brendel W, Constantinidis C, Feierstein CE, Kepecs A, Mainen ZF, Romo R, Qi XL, Uchida N, Machens CK. 2016. Demixed principal component analysis of neural population data. *Elife* **5:** e10989.

Neuro Cloud Consortium. 2016. To the cloud! A grassroots proposal to accelerate brain science discovery. *Neuron* **92:** 622–627.

Pouget A, Drugowitsch J, Kepecs A. 2016. Confidence and certainty: Distinct probabilistic quantities for different goals. *Nat Neurosci* **19:** 366–374.

Sanders JI, Hangya B, Kepecs A. 2016. Signatures of a statistical computation in the human sense of confidence. *Neuron* **90:** 499–506.

In Press

Hirokawa J, Vaughan A, Kepecs A. 2017. Categorical representations of decision-variables in orbitofrontal cortex. *bioRxiv* doi: 10.1101/135707.

Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A. 2017. Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Curr Biol* **27:** 821–832.

Mohanty A, Li Q, Tadayon MA, Bhatt GR, Cardenas J, Miller SA, Kepecs A, Lipson M. 2017. Reconfigurable visible nanophotonic switch for optogenetic applications. *Proc SPIE* doi: 10.1117/12.2251663.

PRINCIPLES OF NEURAL COMPUTATION

A. Koulakov B. Baserdem T. Ngoc
H. Giaffar S. Shuvaev
D. Kepple

Our laboratory works on theories of neural computation. Our overall strategy is to use methods developed in mathematics, physics, machine learning, computer science, and statistics to build experimentally testable models of neural networks and their functions. In most cases, we base our theories on what is known about particular biological systems; however, given that the principles of brain function remain unclear, in many cases, we resort to building machine-learning theories. This means that we formulate the problems solved by the brain in a mathematically rigorous fashion and hypothesize how an engineer would solve the problem. We then use these solutions to form experimentally testable predictions. Testing these predictions in collaboration with our experimental colleagues helps us refine our models. For example, we are interested in understanding how connectivity is established in the brain. We have proposed several theories that may determine the rules of making connections between neurons based on a limited set of instructions contained in the genome. These theories address several levels of organization, including computational, biological, engineering, and evolutionary. Our theories may explain the differences between connectivities in normal and abnormal brain circuits. We are also interested in understanding the principles of perceptual invariance—that is, how can sensory systems represent objects in the environment despite substantial variations in intensity and background. Visual percepts, for example, retain basic features, such as perceived shape and color composition, despite variable luminance, spectral composition, scale, and position of the stimuli. Although we study the question of perceptual invariance in application to well-defined problems, we believe that the principles that we will uncover may generalize across sensory modalities. Finally, we are pursuing the question of how modern theories of machine learning and artificial intelligence can apply to brain function. Although reinforcement learning, deep learning, long short-term memory networks, etc. are successful in solving a variety of artificial intelligence problems, their mapping onto brain circuits remains unclear. We attempt to bring these systems closer to satisfying the constraints imposed by biology. We hope that the convergence of machine-learning theories and biology will help us learn more about brain function.

Neural "Relativity" and Duality in Sensory Coding

This work was done in collaboration with D. Rinberg (New York University).

Sensory systems are constantly facing the problem of computing the stimulus identity, which is invariant with respect to several features. In the olfactory system, for example, odorant percepts have to retain their identity despite substantial variations in concentration, timing, and background. This computation is necessary for us to be able to navigate in chemical gradients or within variable odorant plumes. How can the olfactory system robustly represent odorant identity despite variable stimulus intensity? We propose a novel strategy for the encoding of intensity-invariant stimulus identity that is based on representing relative rather than absolute values of the stimulus features. We therefore call this strategy a neural "relativity" principle. We propose that, once stimulus features are extracted at the lowest levels of the sensory system, the stimulus identity is inferred on the basis of their relative amplitudes. Because, in this scheme, stimulus identity depends on relative amplitudes of features, identity becomes invariant with respect to variations in intensity and nonlinearities of neuronal responses. For example, in the olfactory system, odorant quality can be represented by the identities of the strongest responding odorant receptor types. In this project, we study how this information can be used to reconstruct information about an odorant. We show that this information is sufficient to ensure the robust recovery of a

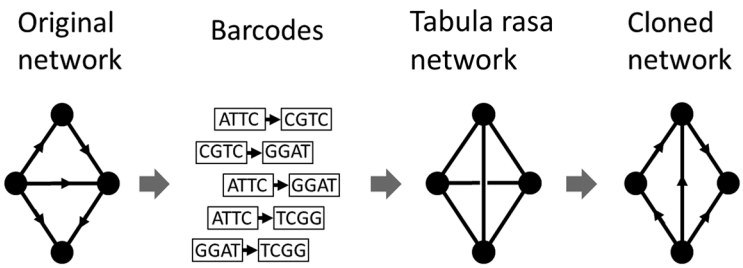


Figure 1. Network cloning as a way to copy connectivity from one network to another. The original network is read out into a set of barcodes carrying information about connections. Each half of the barcode represents one of the cells that are connected, whereas the link represents the direction of the connections. These barcodes are then introduced into a tabula rasa network that has no structure. Barcodes are capable of shaping the tabula rasa network to match the target connectivity.

sparse stimulus (odorant) via minimization of a cost function representing parsimony of the stimulus. Such a minimization has to be performed under the constraints imposed by the relationships between stimulus features. We relate this problem to a dual problem conventionally solved in applied mathematics. We study how the dual problem can be solved by the neural networks in the piriform cortex and olfactory bulb. Our theory yields testable predictions for the structure of olfactory connectivity.

Evolution of Cortical Parcellation and Connectivity

In this project, we study the distribution of brain and cortical area sizes (parcellation units [PUs]) obtained for three species: mouse, macaque, and human. We find that the distribution of PU sizes is close to log-normal. We propose the mathematical model of evolution of brain parcellation based on iterative fragmentation and specialization. In this model, each existing PU has a probability to be split that is dependent on PU size only. This model suggests that the same evolutionary process may have led to brain parcellation in these three species. Interestingly, a similar model had been proposed in the 1940s by Kolmogorov to describe the distribution of mineral pieces. Kolmogorov's model has been successful in explaining features of distributions of explosive shell fragments and pieces produced by rock grinding and crushing. We both evaluate the statistical distribution of brain region sizes and propose an evolutionary model that is somewhat distinct from Kolmogorov's theory. We also apply Hebbian learning theories to the distribution of the strengths of connections between brain regions. Within our model, region-to-region (macro)connectivity is given by the outer product form. We show that most experimental data on nonzero macaque cortex macroscopic-level connections can be explained by the outer product power-law form suggested by our model (62% for area V1). We propose a multiplicative Hebbian learning rule for the macroconnectome that could yield the correct scaling of connection strengths between areas. We, thus, propose an evolutionary model that may have contributed to both brain parcellation and mesoscopic level connectivity in mammals.

Theory of Network Cloning Using DNA Barcodes

This work was done in collaboration with A. Zador (CSHL).

The connections between neurons determine the computations performed by both artificial and biological neural networks. Recently, we have proposed SYNseq, a method for converting the connectivity of a biological network into a form that can exploit the tremendous efficiencies of high-throughput DNA sequencing. In SYNseq, each neuron is tagged with a random sequence of DNA—a "barcode"—and synapses are represented as barcode pairs. SYNseq addresses the analysis problem, reducing a network into a suspension of barcode pairs. Here, we formulate a novel and complementary synthesis problem: How can the suspension of barcode pairs be used to "clone" or copy the network back into an uninitialized tabula rasa network? Although this synthesis problem might be expected to be computationally

intractable, we find that, surprisingly, this problem can be solved efficiently using only neuron-local information. We present the "one barcode one cell" (OBOC) algorithm, which forces all barcodes of a given sequence to coalesce into the same neuron, and show that it converges in a number of steps that is a power law of the network size. Rapid and reliable network cloning with single-synapse precision is thus theoretically possible.

PUBLICATION

Ferrante DD, Wei Y, Koulakov AA. 2016. Mathematical model of evolution of brain parcellation. *Front Neural Circuits* **10:** 43.

THE FUNCTION AND PLASTICITY OF CENTRAL SYNAPSES IN ADAPTIVE AND MALADAPTIVE BEHAVIORS RELATED TO PSYCHIATRIC DISORDERS

B. Li S. Ahrens, A. Bouhuis, C. Bravo-Rivera, H. Deng, C. Fernandes Henriques, A. Furlan, J. Giovanniello, G. Hwang, L. Mejia, M. Stephenson-Jones, X. Xiao, T. Yang, K. Yu, X. Zhang

Understanding the relationship between synapse, circuit, and behavior has been the focus of research in our lab. We are particularly interested in understanding the synaptic and circuit mechanisms underlying cognitive functions, as well as synaptic and circuit dysfunction that may underlie mental disorders, including anxiety, depression, schizophrenia, and autism. To address these questions we use in vitro and in vivo electrophysiology and imaging, molecular, genetic, optogenetic, and chemogenetic methodologies to probe and manipulate the function of specific neural circuits in the rodent brain and determine their role in behaviors such as fear regulation, anxiety, reward- and motivation-related behaviors, and autism-related behaviors. We are currently undertaking the following major lines of research.

The Role of the Amygdala Circuitry in Fear Regulation and Anxiety

Our previous studies show that the central amygdala (CeA) has a key role in learning and expression of defensive responses to threats. In particular, our studies indicate that somatostatin-expressing (SOM^+) neurons in the lateral division of the central amygdala (CeL) are essential for the acquisition and recall of conditioned freezing behavior, which has been used as an index of defensive response in laboratory animals, during Pavlovian fear conditioning. We also show that SOM^+ CeL neurons are activated by threat-predicting sensory cues following fear conditioning and that activation of these neurons suppresses ongoing actions and converts an active defensive behavior to a passive response. Furthermore, inhibition of these neurons using optogenetic or molecular methods promotes active defensive behaviors. Our results provide the first in vivo evidence that SOM^+ neurons represent a CeL population that acquires learning-dependent sensory responsiveness during fear conditioning and furthermore reveal an important role of these neurons in gating passive versus active defensive behaviors in animals confronted with threat.

More recently, we show that another major class of CeL neurons, the protein kinase C-δ-expressing (PKC-δ^+) neurons, is essential for the synaptic plasticity underlying learning in the lateral amygdala, as it is required for lateral amygdala neurons to respond to unconditioned stimulus (US) and furthermore carries information about the US to instruct learning. Our results indicate that PKC-δ^+ CeL neurons constitute a key node in a pathway that imparts information about US to the lateral amygdala (LA) during fear conditioning (FC), hence revealing a previously unknown amygdala functional organization in which the CeL is upstream of the LA in processing aversive US during learning. Our findings also revise a prevalent model for the functional organization of amygdala circuits, which posits that PKC-δ^+ CeL neurons are "fear-off" neurons—a CeL population that shows inhibitory conditioned stimulus (CS) responses following fear conditioning—that act to suppress fear responses through inhibition of amygdala output. In fact, we show that a substantial population of PKC-δ^+ CeL neurons are essentially "fear-on" neurons and function in the opposite manner by conveying aversive US signals.

Brain Circuits Involved in Susceptibility and Resilience to Stress

Some individuals are resilient, whereas others succumb to despair in repeated stressful situations. The

neurobiological mechanisms underlying such divergent behavioral responses remain unclear. In a recent study, we used an automated method for mapping neuronal activity in search of signatures of stress responses in the entire mouse brain. We used serial two-photon tomography to detect expression of c-fosGFP, a marker of neuronal activation, in *c-fosGFP* transgenic mice subjected to the learned helplessness (LH) procedure, a widely used model of stress-induced depression-like phenotype in laboratory animals. We found that mice showing "helpless" behavior had an overall brain-wide reduction in the level of neuronal activation compared with mice showing "resilient" behavior, with the exception of a few brain areas, including the locus coeruleus, that were more activated in the helpless mice. In addition, the helpless mice showed a strong trend of having higher similarity in whole-brain activity profile among individuals, suggesting that helplessness is represented by a more stereotypic brain-wide activation pattern. This latter effect was confirmed in rats subjected to the LH procedure, using 2-deoxy-2[18F] fluoro-D-glucose positron emission tomography to assess neural activity. Our findings reveal distinct brain activity markers that correlate with adaptive and maladaptive behavioral responses to stress and provide a framework for further studies investigating the contribution of specific brain regions to maladaptive stress responses.

A Basal Ganglia Circuit for Evaluating Action Outcomes

The basal ganglia, a group of subcortical nuclei, play a crucial role in decision-making by selecting actions and evaluating their outcomes. Although much is known about the function of the basal ganglia circuitry in selection, how these nuclei contribute to outcome evaluation is less clear. Here, we show that neurons in the habenula-projecting globus pallidus (GPh) are essential for evaluating action outcomes and are regulated by a specific set of inputs from the basal ganglia. We found in a classical conditioning task that individual mouse GPh neurons bidirectionally encode whether an outcome is better or worse than expected. Mimicking these evaluation signals with optogenetic inhibition or excitation is sufficient to reinforce or discourage actions in a decision-making task. Moreover, cell type–specific synaptic manipulations revealed that the inhibitory and excitatory inputs to the GPh are necessary for mice to appropriately evaluate positive and negative feedback, respectively. Finally, using rabies virus–assisted monosynaptic tracing, we discovered that the GPh is embedded in a basal ganglia circuit wherein it receives inhibitory input from both striosomal and matrix compartments of the striatum and excitatory input from the "limbic" regions of the subthalamic nucleus (STN). Our results provide the first direct evidence

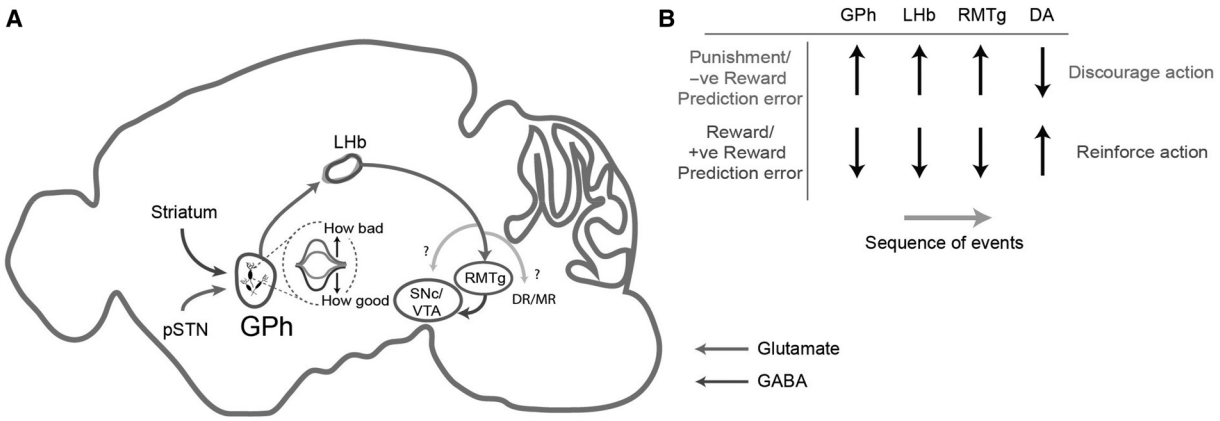


Figure 1. The proposed function of the basal ganglia and midbrain evaluation circuits. (*A*) Schematic showing the activity of GPh neurons and the downstream circuitry controlling the midbrain dopaminergic system. (*B*) Proposed sequence of events by which GPh activity may influence the firing rate in downstream structures. Upward arrows indicate an increase in firing; downward arrows indicate a decrease in firing. RMTg, Rostromedial tegmental nucleus; SNc, substantia nigra pars compacta; VTA, ventral tegmental area; DA, dopamine; DR, dorsal raphe; MR, median raphe. "?" indicates that alternative circuits downstream from the LHb, including the serotonergic raphe nuclei, may constitute other key pathways that also process the GPh-LHb prediction error signals that we show in this study. (*A*, Modified from the Allen Mouse Brain Atlas, Allen Institute for Brain Science; available from http://mouse.brain-map.org.)

that information about the selection and evaluation of actions is channeled through distinct sets of basal ganglia circuits, with the GPh representing a key locus in which information of opposing valence is integrated to determine whether action outcomes are better or worse than expected (Fig. 1).

PUBLICATIONS

Kim Y, Perova Z, Mirrione MM, Pradhan K, Henn FA, Shea S, Osten P, Li B. 2016. Whole-brain mapping of neuronal activity in the learned helplessness model of depression. *Front. Neural Circuits* **10:** 3.

Stephenson-Jones M, Yu K, Ahrens S, Tucciarone JM, van Huijstee AN, Mejia LA, Penzo MA, Tai L-H, Wilbrecht L, Li B. 2016. A basal ganglia circuit for evaluating action outcomes. *Nature* **539:** 289–293.

Yu K, Garcia da Silva P, Albeanu DF, Li B. 2016. Central amygdala somatostatin neurons gate passive and active defensive behaviors. *J Neurosci* **36:** 6488–6496.

In Press

Yu K, Ahrens S, Zhang X, Schiff H, Ramakrishnan C, Fenno L, Deisseroth K, Zhou P, Paninski L, Li B. 2017. The central amygdala controls learning in the lateral amygdala. *bioRxiv* doi: https://dx.doi.org/10.1101/126649.

INTEGRATIVE SYSTEMS NEUROBIOLOGY

P.P. Mitra L. Cartagine, D. Ferrante, A. Field-Pollatou, B. Huo, Z. Lodato, I. Malladri, J. Mulhern, S. Savoia, A. Tolpygo

We study complex neurobiological systems using a combination of experimental and computational approaches. The primary area of experimental work in our laboratory continues to be the Mouse Brain Architecture (MBA) project, with a goal to generate a mesoscale connectivity map. The project has recently reached a major landmark, completing the acquisition of a full data set (636 whole mouse brains, 1037 injection sites) with one anterograde tracer (AAV). The MBA data set and associated metadata are freely viewable with the aid of a three-dimensional (3D) injection browser and a high-resolution image viewer at the project website http://mouse.brainarchitecture.org. Our neuroinformatics research involves the development of analytical tools and informatics infrastructures to process, analyze, and integrate large volumes of neurobiological data in various brain connectivity projects. In collaborative studies, we apply the methods developed for the MBA project to other mouse strains, zebra finch, marmoset, macaque monkey, and human. We highlight ongoing efforts and development below.

We completed a joint study with Josh Huang at CSHL (supported by the Simons Foundation) to obtain a brainwide census of GABAergic interneurons in the mouse brain during development and in genetic models of autism and released all 140 whole-brain image data to the public on a dedicated portal (http://mouse.brainarchitecture.org/hua/). The goal of the cell counts project is to generate brainwide maps of inhibitory neuron subpopulations and long-range projects in these autism spectrum disorder models, and in wild-type controls, to determine the neuroanatomy of dysfunctional inhibition relevant to autism.

We are in the third year of collaboration with Pavel Osten at CSHL (funded by the National Institutes of Health [NIH] through the BRAIN Initiative [BRAINI] program) to obtain an accurate whole-brain census of neuron classes in cell type–specific reporter mice and develop neuroinformatics infrastructure to analyze and integrate such data. In the first release, cell type–specific connectivity using reporter mice is used (http://mouse.brainarchitecture.org/ost2/). A cell type–based atlas with initial broad focus on GABAergic inhibitory neurons is available at http://mouse.brainarchitecture.org/ost/. The revamped portal includes cell counts (CSV) and overlays displaying the centroids of automatically detected cells. We have optimized a custom version of our Java-based Djatoka image server and added OpenLayers, the same technology used for displaying annotations on Google Maps. This allows us to add the annotations to each image and control the placement of box/polygon selections and other markup tools. Upgrades to the display, markup, and counting tools are ongoing.

On a separate BRAINI grant, we have joined a team effort led by Sacha Nelson (Brandeis University) to a cross-species study of neuronal cell types in transgenic strains of rats and mice. Data collection is in progress, and released data sets are available at the dedicated portal http://mouse.brainarchitecture.org/listing/nel/.

We also have major ongoing collaborations with Marcello Rosa (Monash University, Australia) and Hideyuki Okano (RIKEN, Japan) on a project in which methods developed for the MBA pipeline are adapted to the marmoset brain to develop its mesoscale circuit map and its first digital atlas. We have set up a high-throughput histology pipeline at the RIKEN Brain Science Institute with the goal of producing 25 marmosets per year, mapping four tracer injections in each brain. The project will test our ability and computational toolbox for "big image data" processing, as the images generated are 5–8× larger than those for our baseline mouse images. Data on a single marmoset brain constitute approximately 1400 images with 25 k × 30 k pixels. Currently, we automatically identify strongly labeled regions and cell bodies (fluorescent: green fluorescent protein, tdTomato, Fast Blue) and we can predict cell centers based on close-to-round/centroid parameters. As we

iterate the algorithms and improve detection, we plan to combine cell density and tracer location data with segmented anatomical regions, producing a detailed map for cortically injected regions.

Dr. Mitra is a distinguished (visiting) professor at the Indian Institute of Technology (IIT), Madras, where he initiated a Center for Computational Brain Research. At the Center, he lectures and collaborates with a team at the Department of Computer Science and Engineering at IIT Madras. Last summer, the Mitra lab hosted two IIT Madras computer science doctoral candidates for a rotation to further this collaborative effort, including presentations at the Society for Neuroscience.

MBA Project

A. Tolpygo, S. Savoia, D. Ferrante, J. Mulhern, L. Cartagine, Z. Lodato

The project aims to construct a comprehensive mesoscale wiring map of the adult mouse brain. We use classical neuroanatomical approaches scaled to a high-throughput data acquisition pipeline. Using stereotaxic targeting, we inject neurotracer substances in one of the predetermined sites systematically mapped on a 3D grid that covers an entire hemispheric brain volume. The sample brains are sectioned, histologically processed, and digitally scanned. To the public, we release the registered high-resolution section images and metadata through our web portal (http://mouse.brainarchitecture.org/). The total number of brains on the portal now stands at more than 1350, representing more than 1500 injections. We have finalized the anterograde (AAV) data set and have made progress in acquiring the retrograde counterpart. Our primary target for the next phase is to complete the retrograde injections using cholera toxin subunit B (CTB) (importantly, these were not performed by the companion project for projection mapping by the Allen Institute of Brain Research) and perform the critical brain-to-brain analysis. In the next cycle, there will be a continuous rollout of the analysis tools that will be used to analyze both data sets and associated projects. Table 1 shows the relative advantages of the MBA data set. To efficiently analyze such large image sets, it is necessary to create a computational pipeline with automatic processing steps with solutions to segmentation tasks at different levels of anatomical detail. This includes, among others, tissue/background separation, cell/tissue segregation, extraction of specific anatomical compartments, detection of injection site, and other cellular signals. A short summary of our growing toolbox and collaborations is presented below.

Atlas-Based Segmentation Using Nissl Data Sets

This work was done in collaboration with M. Miller and B. Lee (Johns Hopkins University).

The unique aspect of the MBA data set is the imaging of Nissl-stained sections for every alternating signal section (fluorescent or immunohistochemical). Effectively, two 3D volumes can be elicited from each mouse brain, with the Nissl stack as an anatomical reference for each brain. This collaboration looks into morphing the Allen reference atlas (ABAv2 2011) annotations to map the brain regions

Table 1. AAV data set comparison with the Allen Brain Institute Phase-I (doi: 10.1038/nature13186) and Mouse Connectome Project (doi: 10.1016/j.cell.2014.02.023)

Parameter	MBA	AIBS phase-I	MCP neocortex
Anterograde tracing	AAV, BDA	AAV	PHAL, BDA
Retrograde tracing	CTB, RV	N/A	CTB, FG
Nissl	Alternating	N/A (CCF)	NeuroTrace
Total injections	1037 (AAV)	469 (AAV)	402 (all tracers)
Bolus diameter (μm)	250–600	400–1000	250–500
Z-sampling (μm)	20/40	100	200
Number of sections per brain	550–600	140	34–72
Sectioning plane	Coronal	Coronal	Coronal
Image resolution (μm/pixel)	0.46	0.35	0.35

The Mouse Brain Architecture (MBA) data set has finer section spacing (2.5× and 5× more sections) while providing direct measurements of cytoarchitecture via the alternate Nissl series. The data set features a higher frequency of injections in the sample volume, with smaller mean distances and similar injection volumes.

onto each Nissl slice. The deformable mapping and computed diffeomorphisms are applied to the atlas segmentation, producing a segmented target that can be viewed with a regional overlay. Using this process, we were able to map the fluorescent injection and Nissl-stained histology slices into a single coordinate space to give the correspondence between structural information provided by the Nissl and connective information provided by the injection. To address this, we applied the multichannel large deformation diffeomorphic metric mapping (LDDMM) algorithm, a deformable registration framework. In our case, the algorithm is used to register Nissl target images to the Nissl atlas (annotated with regions), and the computed deformation is applied to the atlas to bring it into the target space for each individual brain. Rigid registration is successfully used to overlay the fluorescent images on the Nissl as a result of the tape-transfer alternating sections. Figure 1 shows a sample AAV-injected brain showing a Nissl and fluorescent section with the anatomical boundaries overlay. The MBA portal enables seamless change from a structural Nissl section to an injection region by means of a transparency slider. The final goal of this collaboration is to produce a Nissl atlas based on the MBA project's Nissl-stained images.

Detection of Labeled GABAergic Interneurons via Classical Image Processing Techniques

This work was done in collaboration with S. Das (IIT Madras).

In this study, using CRE lines of C57/BL6 mice and gene deletion mouse models of autism (generated earlier with J. Huang at CSHL), the target consists of specific classes of interneurons that express fluorescent label. Green fluorescent probes (GFPs) are used widely to study the structural and functional anatomy of the brain, and as hundreds to thousands of cells may appear in a single image, it is necessary to develop automated methods for detection and counting of the objects within the image. We have developed an iterative image processing algorithm based on distance transformation and ridge processing techniques with a high degree of precision (0.94) and recall (0.95) when compared with manual detection of the cell nuclei. Because the automated detection is fast, consistent, and high-yield, it will significantly reduce the labor required to generate whole-brain data. Currently, the system above is deployed as a web interface applied to estimate the spatial distribution of GFP-tagged GABAergic neuronal types in the wild-type and autism spectrum model

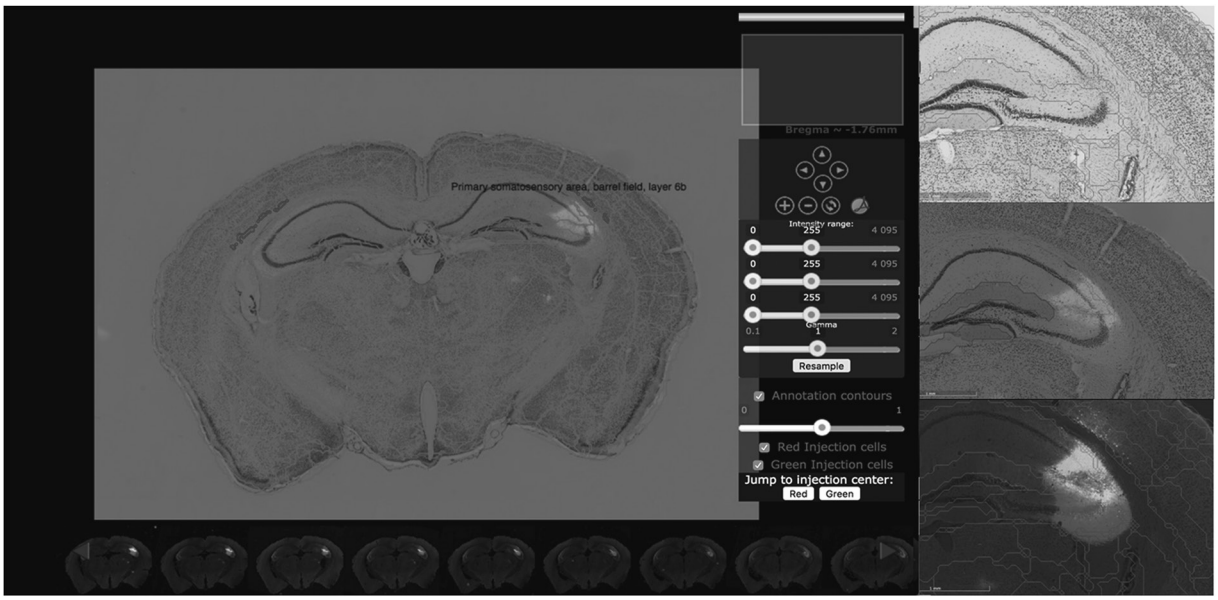


Figure 1. AAV-injected brain processed through the new registration, segmentation, and annotation pipeline. Deployed on the web: Selectable, registered, anatomical boundaries are displayed over each individual section. Fluorescent sections are registered to the Nissl and allow for instant comparison and localization.

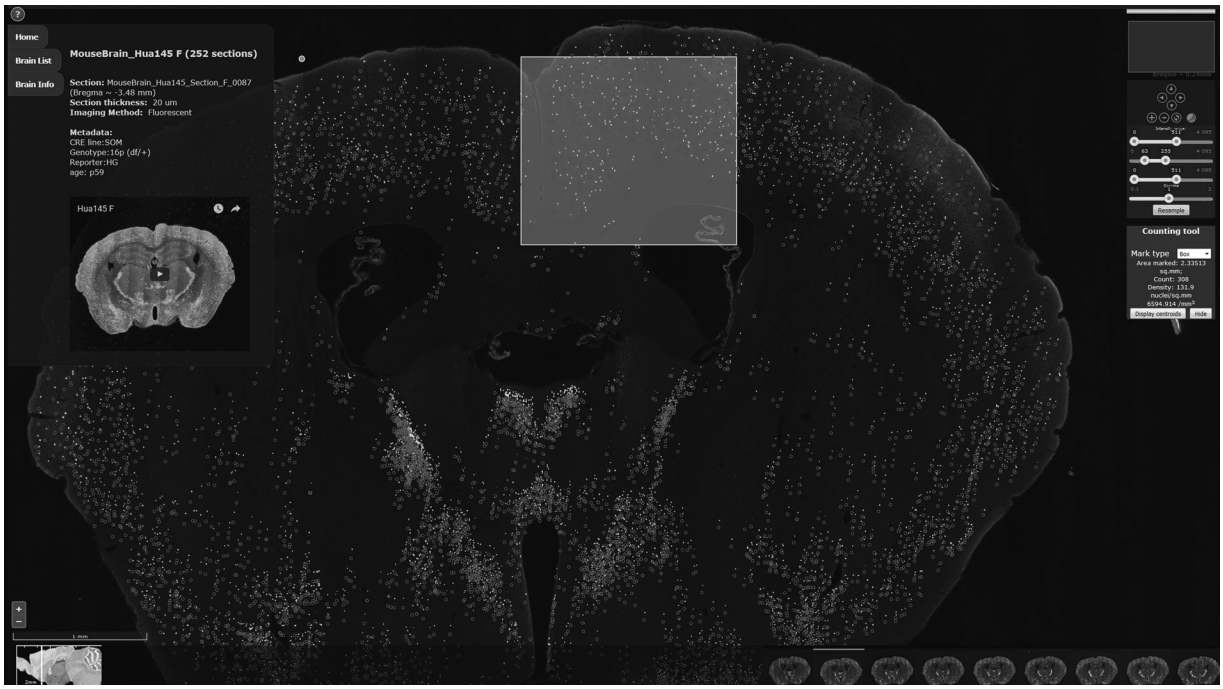


Figure 2. Web interface implementation of the cell detection algorithm designed for the counting of GFP nuclei with test data set Somatostatin-expressing GABAergic neurons (SOM+) as ground truth. The interface is designed with zoom, ROI specification, automatic evaluation of cell counts, and the detection of cell centers.

(16pdf/+). The algorithm may also be used as a standalone function, with the input being the entire image.

Neuronal Tree Extraction Based on Computational Geometry and Topology Methods

This work was done in collaboration with Y. Wang and S. Wang (Ohio State University).

We have been working with the team of Prof. Yusu Wang to analyze the mesoscale circuit structure of the AAV connectivity data set using methods from computational topology and geometry (Morse theory and persistent homology). This approach to segmenting and tracing anterograde axonal projections is robust to noise and gaps in data and retains global shape information, which is lost in the connectivity matrix derived from the same data in conventional graph-theoretical analysis. As we continue to skeletonize the collection of neuronal trees emanating from an injection site and develop metrics to compare tree morphologies, our goal is to align trees while finding consensus trees using the persistent homology-based methods. Further, by mapping these trees onto consensus Nissl and compartmentalized structures, we can begin to create objective representations of key components of neuronal pathways, from injection to projection sites. By combining this geometrical information, overlaying branching and tree maps, we can enhance the regional connectivity-based summaries (threshold, density) currently used in traditional matrices.

Advances toward 3D Reconstruction of Traumatic Vascular Injury

This work was done in collaboration with L. Latour (NINDS, CRNM).

Processing of neuroanatomical tissue is subject to many forms of distortion. With growth in whole-slide imaging and development of high-throughput approaches to histology based on our tape-transfer method, we investigate tissue morphology of posttraumatic brain injury samples as relating to traumatic microbleeds. We obtain high-quality histological material and coregister to magnetic resonance imaging (MRI) data sets for pathology and subsequent analysis. Development of whole-brain processing techniques is ongoing and will permit unbiased study of tissues without constraint to

small region blocking, and offers the unique opportunity to translate new MRI techniques and high-resolution 7T+ scanning to specific 3D histopathology work.

PUBLICATIONS

Ashika, Chakraborti S, Deodhare D, Kumar PS, Mitra PP, Sharma A. 2016. Relation extraction from *Neuroscience* articles in PubMed.

Das S, Gopal VV, Pahariya G, Ferrante DD, Mitra PP. 2016. Automated detection of GFP labeled nuclei in whole-brain light-microscopic datasets for mouse with high precision and recall. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Ferrante DD, Wang S, Li Y, Wang Y, Mitra PP. 2016. Computational topology algorithms for comparing neuronal tree shapes using persistent homology. In *Abstracts presented at the Brain Initiative Meeting*, Seattle, WA.

Ferrante DD, Lee B, Miller M, Mitra PP. 2016. Mouse Brain Architecture LDDMM-based multi-channel registration pipeline. In *Abstracts presented at the Brain Initiative Meeting*, Seattle, WA.

Ferrante DD, Wang S, Li Y, Wang Y, Mitra PP. 2016. Computational topology algorithms for skeletonizing whole mouse brain tracer-injection data using discrete Morse Theory. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Jeong M, Kim Y, Kim J, Ferrante DD, Mitra PP, Osten P, Kim D. 2016. Comparative three-dimensional connectome map of motor cortical projections in the mouse brain. *Sci Rep* **6:** 20072.

Ksendzovsky A, Tolpygo AS, Walbridge S, Diamond J, Ferrante DD, Cummins A, Heiss J, Kapur J, Mitra PP, Zaghloul KA. 2016. Ex-vivo retrograde transport in rat cortex; a model for human cortical connectivity. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Li Y, Ascoli G, Mitra PP, Wang Y. 2016. Metrics for comparing neuronal tree shapes based on persistent homology. *bioRxiv* doi.org/10.1101/087551.

Lin M, Takahashi YS, Weber K, Hossain K, Huo B, Tolpygo AS, Ferrante DD, Bai S, Rosa MG, Okano H, Mitra PP. 2016. Computational infrastructure to enable whole-brain mesoscale circuit mapping for marmoset. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Majka P, Chaplin TA, Yu HH, Tolpygo AS, Mitra PP, Wójcik DK, Rosa MGP. 2016. Towards a comprehensive atlas of cortical connections in a primate brain: Mapping tracer injection studies of the common marmoset into a reference digital template. *J Comp Neurol* **524:** 2161–2181.

Pahariya G, Vangala VG, Das S, Mitra PP, Ferrante DD. 2016. Analysis of Cre:H2B-GFP labeled GABAergic interneurons data from the Mouse Brain Architecture project. In *Abstracts presented at the Annual Workshop for the Center of Computational Brain Research*, Chennai, India.

Sharma A, Sharma A, Deodhare D, Chakraborti S, Sreenivasa P, Mitra PP. 2016. Case representation and retrieval techniques for neuroanatomical connectivity extraction from PubMed. In *Case-based reasoning research and development*, pp. 370–386. LNCS, Vol. 9969.

Singh A, Parag T, Ferrante DD, Mitra PP. 2016. Automated cell segmentation of Nissl stained mouse brain images. In *Abstracts presented at the Annual Workshop for the Center for Computational Brain Research*. Chennai, India.

Singh A, Parag T, Ferrante DD, Mitra PP. 2016. Automated segmentation of Nissl-stained somata from whole-brain histological image data. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Takahashi YS, Weber K, Lin M, Hossain K, Huo B, Tolpygo AS, Ferrante DD, Hata J, Chan J, Mizukami H, et al. 2016. A high-throughput neurohistological pipeline for whole-brain mesoscale circuit mapping for marmoset. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Tolpygo AS, Ferrante DD, Ram K, Savoia S, Lee BC, Mechler F, Franciotti N, Miller MI, Mitra PP. 2016. A comprehensive data set for whole-brain mesoscale connectivity mapping in mouse using injections of a viral anterograde tracer (AAV) on a brain-wide grid. In *Abstracts presented at the Society for Neuroscience 2016*, San Diego, CA.

Vogelstein JT, Amunts K, Andreou A, Angelaki D, Ascoli G, Bargmann C, Burns R, Clai C, Chance F, Chun M, et al. 2016. Grand challenges for global brain sciences. *arXiv* arXiv:1608.06548.

IDENTIFICATION OF DISRUPTED BRAIN CIRCUITS IN MOUSE MODELS OF AUTISM AND SCHIZOPHRENIA

P. Osten G. Fitzgerald, R. Munoz Castaneda, A. Narasimhan, R. Palaniswamy, E. Szelenyi, J. Taranda, K.U. Venkataraju

Quantitative Understanding of Brain Structure and Its Relationship to Behavioral Variability and Disease

What are the neural circuits that drive innate behaviors, such as social behaviors? How can we distinguish normal individual-to-individual variation from abnormal brain features that underlie the risk for neurodevelopmental and psychiatric disorders?

Long-standing barriers to understanding these fundamental questions have been the difficulty in surveying the entire brain at cellular or subcellular resolution, and the lack of tools to quantitatively compare cell distributions in the brains of different animals. A major goal of my laboratory has been to overcome this barrier by developing increasingly high-throughput and high-resolution methods for standardized, automated, and quantitative analyses of whole-brain anatomy and function. Although we initially developed these methods in the context of mapping activated neurons in the brain using immediate early genes, we have now sufficiently improved the throughput, scalability, and sensitivity of whole-brain mapping to enable us to broaden our areas of enquiry and directly address previously unanswerable questions related to brain structural variation, neural circuit function, and the nature of brain aberrations related to human neurologic diseases. Our ongoing research is divided into four programs.

Program 1: Cell-Type-Based Understanding of the Mouse Brain

Brain cell-type anatomy informs all topics in the neurosciences: The distribution and ratios of cell types and their wiring into neuronal circuits underlie the vast diversity of mammalian behaviors. To enable a better understanding of brain circuits and function, we developed the first pipeline of imaging and computational methods for unbiased cell-type atlasing in the mouse brain. We have successfully applied these methods to generate the first quantitative brain-wide maps for seven inhibitory and modulatory cell types proposed to play key and cell-type-specific roles in essentially all forms of brain processing and cognitive functions. One notable finding is that we uncovered a previously unknown hierarchical interneuron organization in the isocortex: The ratios of two major inhibitory cell types, the parvalbumin-positive (PV^+) and somatostatin-positive (SST^+) interneurons, vary significantly in different cortical regions. Computational models using these distributions predicted that they should differentially impact the local circuit processing features (the modeling is a collaboration with Xiao-Jing Wang's lab at New York University [NYU]).

Gender-based brain structural differences are of increasing biomedical interest because of unexplained skewing of genetically promoted neuropsychiatric disorders and male and female differences in drug responses. We directly compared interneuron densities in male and female brains and, surprisingly, identified nine subcortical areas that differ in cell-type composition, eight of which contained more cells (PV^+ or VIP^+) in females. This contrasts with previous studies that identify increased total neurons in males in sex dimorphic areas, highlighting the power of unbiased anatomic studies to discover unexpected circuit features.

We have made these data available as a resource via the BrainArchitecture portal of the Partha Mitra lab at CSHL. Based on the success of this work, we plan to expand this effort by building a mouse brain atlas of cell-type distributions and morphologies for more

than 100 neuronal and glial cell types. The long-term impact of these studies will be to establish a ground truth data set of cell-type distribution and variability in the mammalian brain. Because many current studies of brain function are aimed at dissecting the function of previously identified and well-studied brain regions, we anticipate that this unbiased whole-brain resource will inspire and enable a wide variety of new discoveries.

Program 2: The Structure and Function of Neural Circuits Involved in Social Behaviors

A key question in systems neuroscience is to understand how brain-wide patterns of neuronal activity represent sensory encoding, cognitive processes, and, ultimately, behaviors. Traditionally, the circuit underlying a behavior has been identified by a piecemeal approach: A specific hypothesis about the involvement of a particular brain region is tested by correlating its neural activity with behavior, and the causality of the relationship is tested by loss- or gain-of-function experiments. Although it is immensely powerful, iterating this approach over the set of pathways that are potentially active during any given behavior or that may be altered in genetic mouse models would be prohibitively time- and cost-consuming.

To alleviate this difficulty, we developed an unbiased approach for identifying complete sets of activated brain regions for any behavior based on whole-brain mapping of activity-driven IEG (e.g., *c-fos*) expression (Ragan et al., *Nat Methods 9:* 255 [2012]; Kim et al., *Cell Rep 10:* 292 [2015]; Renier et al., 2016). In the first whole-brain c-fos mapping study, we identified differentially activated brain regions in the male brain after exposure to males, females, or a novel object (Kim et al., *Cell Rep 10:* 292 [2015]). This study revealed comprehensive brain activation maps discriminating between the three behaviors and, furthermore, it showed that our assay can be used to correlate the level of activation (the number of c-fos ± cells) per brain region and the time the individual animal spent in specific behaviors, such as c-fos$^+$ counts in bed nuclei of stria terminals and amygdalar nuclei during sniffing/exposure to female pheromones, or in frontal cortical areas, midline thalamus, and ventral striatum during motivational behavior such as close following. This type of analysis may be particularly relevant for dissecting out subtle and variable changes in brain activation in genetic mouse models for neuropsychiatric disorders (see below).

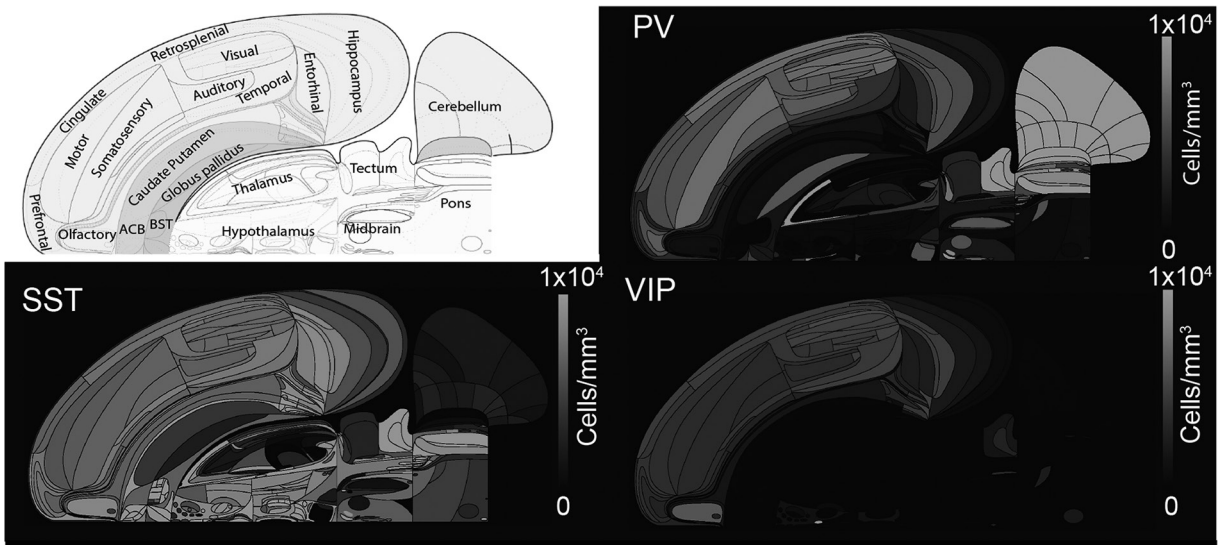


Figure 1. Whole-brain maps of three major neuronal cell types in the mouse brain—the parvalbumin-positive (PV$^+$), somatostatin-positive (SST$^+$), and vasoactive intestinal peptide–positive (VIP$^+$) neurons. Cell densities of each of the cell types are represented per mm^3.

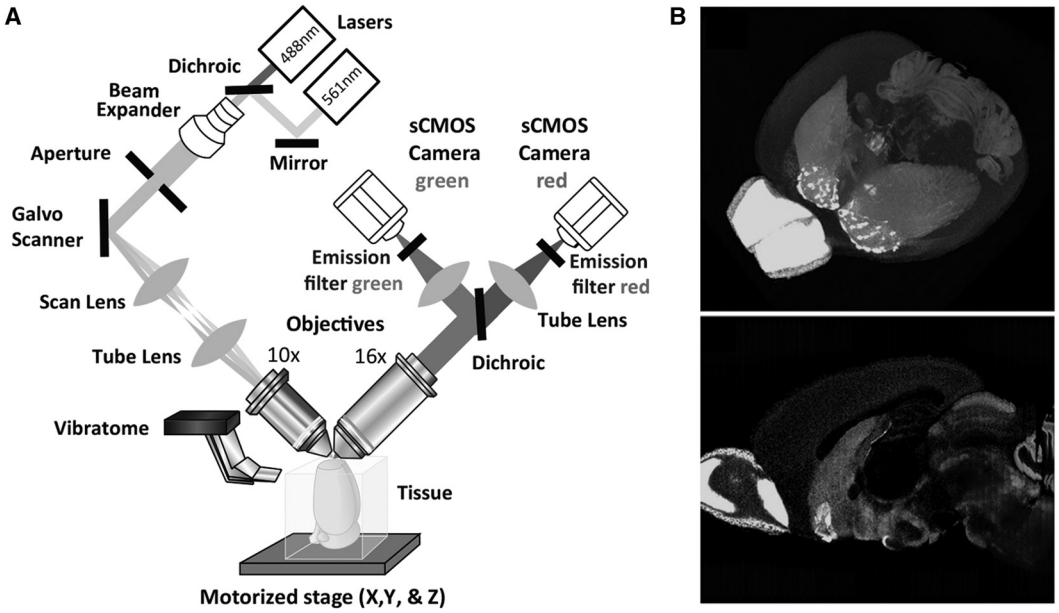


Figure 2. Oblique light sheet tomography (OLST). (*A*) Schema of the OLST instrument with embedded clear mouse brain. (*B*) Whole-brain distribution of all interneurons in the mouse brain imaged by OLST.

Program 3: Deciphering Mouse Models of Human Neurodevelopmental and Psychiatric Disorders

Human genetic studies are uncovering genes and loci linked to pervasive neurologic diseases such as autism and schizophrenia. Yet, we know very little about which neuronal cell types, circuits, or functions are impacted by these genes and mutations, impeding a search for treatments.

Identifying vulnerable circuits in mouse genetic models of neurodevelopmental and psychiatric disorders has been a major motivation for the development of whole-brain mapping and analysis methods in my laboratory. To accomplish this goal, we began by studying brain-wide structural and circuit changes in two genetic mouse models. First, we have identified an incomplete penetrance phenotype linking altered excitation/inhibition balance, increased seizure susceptibility, disrupted non-REM sleep, and increased repetitive behaviors in the mouse model of the 16p11.2 deletion syndrome associated with autism and other neurodevelopmental deficits. This finding recapitulates both the partial penetrance of genetic lesions in autism, including in monozygotic twins, and the high prevalence of electroencephalography (EEG) abnormalities and seizures in autism. Importantly, we have replicated the finding of incomplete phenotypic penetrance in a second isogenic background, establishing the 16p11.2 del mice as a unique model for studying the developmental processes that can influence the gene-to-phenotype relationship and contribute to the well-documented incomplete and variable penetrance in human neurodevelopmental disorders. The application of the same methods across multiple genetic risk models holds the promise of identifying common circuit deficits underlying convergent psychiatric manifestations.

Program 4: Technology Development

Our collaboration with the lab of Marc Tessier-Lavigne led to the development of a new method pipeline for analysis of brains processed by iDISCO+ and imaged by light-sheet fluorescence microscopy (LSFM; Renier et al. 2016). More recently, we have developed new microscopy for cleared brains, named oblique light sheet tomography (OLST), which combines light-sheet imaging and vibratome sectioning to allow optimal image resolution throughout the sample (Narasimhan et al., in press; this work is in collaboration with Florin Albeanu, CSHL), and we have developed a third

iteration of our method pipeline for the OLST data sets. Finally, this year, we have developed a novel version of OLST for superresolution imaging (OLST^{SR}) of cleared brains, with an unprecedented capability of handling cubic millimeters of tissue.

Summary

Motivated by the gap in knowledge presented by difficulties in quantifying cellular phenotypes across the entire mammalian brain, my lab has pioneered a series of high-throughput and high-resolution methods to enable new discoveries relevant to brain circuit functions, gender dimorphism, social behavior, brain evolution, and genetic risk for human neurologic disease.

PUBLICATIONS

He M, Tucciarone J, Lee S, Nigro M, Kim Y, Levine J, Kelly J, Krugikov I, Wu P, Chen Y, et al. 2016. Strategies and tools for the combinatorial targeting of GABAergic neurons in mouse neocortex. *Neuron* **91:** 1228–1243.

Hou XL, Huyn M, Huang KW, Atwater E, Croney D, Taranda J, Guo L, Mathai J, Zeidel ML, Osten P, et al. 2016. Central control circuit for context-dependent micturition. *Cell* **167:** 73–86.

Jeong M, Kim Y, Kim J, Ferrante DD, Mitra PP, Osten P, Kim D. 2016. Comparative three-dimensional connectome map of motor cortical projections in the mouse brain. *Sci Rep* **6:** 20072.

Kim Y, Perova Z, Mirrione MM, Pradhan K, Henn FA, Shea S, Osten P, Li B. 2016. Whole-brain mapping of neuronal activity in the learned helplessness model of depression. *Front Neural Circuits* **10:** 1–11.

Niedworok CJ, Brown AP, Jorge Cardoso M, Osten P, Ourselin S, Modat M, Margrie TW. 2016. aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data. *Nat Commun* **7:** 11879.

Renier N, Adams EL, Kirst C, Wu Z, Azevedo R, Kohl J, Autry AE, Kadiri L, Umadevi Venkataraju K, et al. 2016. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* **165:** 1789–1802.

In Press

Narasimhan A, Umadevi Venkataraju K, Mizrachi J, Albeanu DF, Osten P. 2017. A high resolution whole brain imaging using oblique light sheet tomography. *bioRxiv* doi: 10.1101/132423.

NEURAL CIRCUITRY FOR SOCIAL COMMUNICATION

S. Shea B. Cazakoff B. Lau
R. Dvorkin A. Nowlan
A. Kyalwazi

The broad goal of my laboratory's research is to understand how the brain detects and interprets sensory stimuli to guide flexible behavior. We are particularly interested in how neural activity and plasticity in olfactory and auditory brain circuits facilitate communication and social behavior. We are revealing neural mechanisms that allow organisms to detect and recognize familiar individuals, gather information about their identity and social status, and select appropriate behaviors. Mice can acquire a surprisingly detailed profile of a social partner from the smells and sounds it emits during their encounter. For example, they can discern the other mouse's sex, genetic identity, reproductive state, levels of distress or sexual interest, or even recently consumed foods. As you might imagine, proper interpretation of these social signals is indispensable for survival and mating success.

It is quite reasonable to ask, "Why do you care about the social life of small rodents?" There are really two reasons. First, we want to identify the fundamental principles that govern how the brain adaptively controls complex behavior. Natural social behaviors are well-suited for this purpose, because the brain is exquisitely adapted through evolution to resolve and integrate social cues and link them to powerful behavioral responses. Second, we hope to pinpoint and repair neural circuitry defects that impair appropriate use of social information. Difficulty with social perception and cognition are core features of the autism spectrum disorders (ASDs). For example, patients may have trouble perceiving and interpreting communication gestures such as speech, facial expressions, and "body language." This broad feature is also evident in many mice that carry genetic variants identified in human ASD populations. Therefore, if we can ascertain the neural circuit substrates of social behavior in normal mice, we can make and test predictions for how the circuitry is affected in the mouse models. The results are likely to tell us more about the synaptic modifications that occur in human autism.

In recent years, we have focused on understanding the neural activity and circuitry in primary sensory brain areas that support adaptive behavior. We are now moving deeper into the brain to ascertain where the sensory data from those regions are collected and integrated into hormonal and electrical signals that promote appropriate behavioral choices. We are particularly interested in how all stages of neural processing are flexible according to experience and behavioral state.

Dynamics of Neuronal Inhibition during Olfactory Learning

B. Cazakoff

One of the most fundamental questions in neuroscience is "How does ongoing behavior in a dynamic environment affect the neural processing and perception of sensory stimuli?" For example, behavioral state and previous experience can dramatically change sensory responses. Rather than a one-way street, sensory perception is probably better viewed as an ongoing and dynamic interaction between the brain and its environment. Understanding this interaction in greater detail is a crucial goal for understanding brain and behavior. Flexibility in sensory processing is absolutely necessary for optimal decisions regarding stimuli, and, if impaired, that could lead to negative behaviors such as perseveration or impulsivity. To approach this goal, we are tracking activity in the main olfactory bulb (MOB) of mice as they learn and update associations of specific stimuli with specific outcomes.

Arguably, the most critical and finely tuned instrument providing a mouse with information about its world is its sense of smell. Detection and interpretation of olfactory stimuli in the brain begin at the MOB. Sensory receptor neurons in the nose relay information about odors to projection neurons, which carry the output of the MOB to deeper brain structures such as the olfactory cortex. Despite the rather

peripheral location of the MOB, one synapse away from the primary sensory neurons, the bulb is heavily targeted by signals that originate deeper in the brain. These "feedback" inputs are well-positioned to modulate MOB sensory responses according to internal factors such as behavioral state and experience. Feedback is obviously important for maintaining and updating stimulus associations in a changing environment.

Among the neurons in the MOB, the inhibitory granule cells (GCs) are certainly the most numerous, arguably the most important, and yet somehow the most mysterious. GCs are the primary target of feedback connections, making them a conduit for the olfactory bulb to integrate odor information with signals originating deep in the brain. Furthermore, functional studies suggest that they are necessary for olfactory discrimination and learning. Thus, GCs are a crucial link between the projection neurons that carry odor information out of the MOB and central feedback carrying signals about positive and negative outcomes associated with specific stimuli. Nonetheless, because they do not yield to conventional recording techniques, GC activity patterns in vivo are poorly understood, and their electrophysiological properties in awake animals are completely unknown.

Our group developed methods to overcome this technical roadblock, and subsequently we were the first to report results from recordings of individual GCs in awake mice that were passively presented odors. We found that the activity of these cells was dramatically different from the activity seen in anesthetized mice, indicating that they are heavily modulated by behavioral state. Nevertheless, that study did not address how GCs may be further affected by the animal's participation in a task requiring it to learn and update associations of specific odors with specific outcomes. We developed such a task in head-fixed mice that allows us to monitor the neuronal activity of GCs as the mouse rapidly adjusts its behavior in response to changes in odor meaning. We are specifically interested in how the mouse's ability to solve this task involves the GCs.

This year, Brittany Cazakoff successfully extended her recordings of GCs into mice that are not only awake, but actively learning new associations of odor stimuli with reward (water) and punishment (bitter taste). She has developed a behavioral paradigm in which the mouse learns that two new odors each signal either impending reward or the bitter tastant, respectively. They are also asked to learn that the odor associations have switched. The mice show that they have learned by licking when reward is signaled and withholding licks when the odor portends bitter taste. This learning develops in a time frame that allows us to monitor activity in individual GCs throughout. These experiments reveal several very intriguing preliminary findings. First, some of these mice successfully learned the task, while some did not. Interestingly, we see stronger responses to odors in GCs from mice that had successfully been trained as compared with mice that did not learn the task (Fig. 1A). Second, many cells change their response to an odor when it begins to predict a new outcome (Fig. 1B). Third, we found that responses often fluctuate over time in the trials following this switch. These fluctuations were not random, but rather were related to the animal's fluctuations in task accuracy (Fig. 1C).

Auditory Plasticity Sharpens Vocal Perception during Maternal Learning

K. Krishnan, B. Lau, A. Nowlan

Far beyond the limits of our hearing, in the ultrasound range, mice are continuously "speaking" (or vocalizing) to one another in a "language" that we have only just begun to understand. Many distinct vocalization types are produced by males, females, juveniles, and adults in a variety of behavioral situations. We would like to better understand the perceptual significance of these calls to the mice and how they are used to guide behavioral choices. One form of vocalization that is reasonably well understood is the ultrasonic distress vocalization (USV). Young mice, before vision and full mobility, will occasionally become separated from the nest. This is stressful for them and they will therefore call out to their mother with a very high frequency peep. New mothers develop sensitivity to these cries and respond by moving toward their source to retrieve the pup. Moreover, females who have never given birth do not innately show approach responses to pups or their calls; however, they can learn to perform the behavior with experience. We refer to this group as "surrogates." Accurate maternal learning by mothers and surrogates is suspected to involve rewiring (or "plasticity") in the auditory cortex.

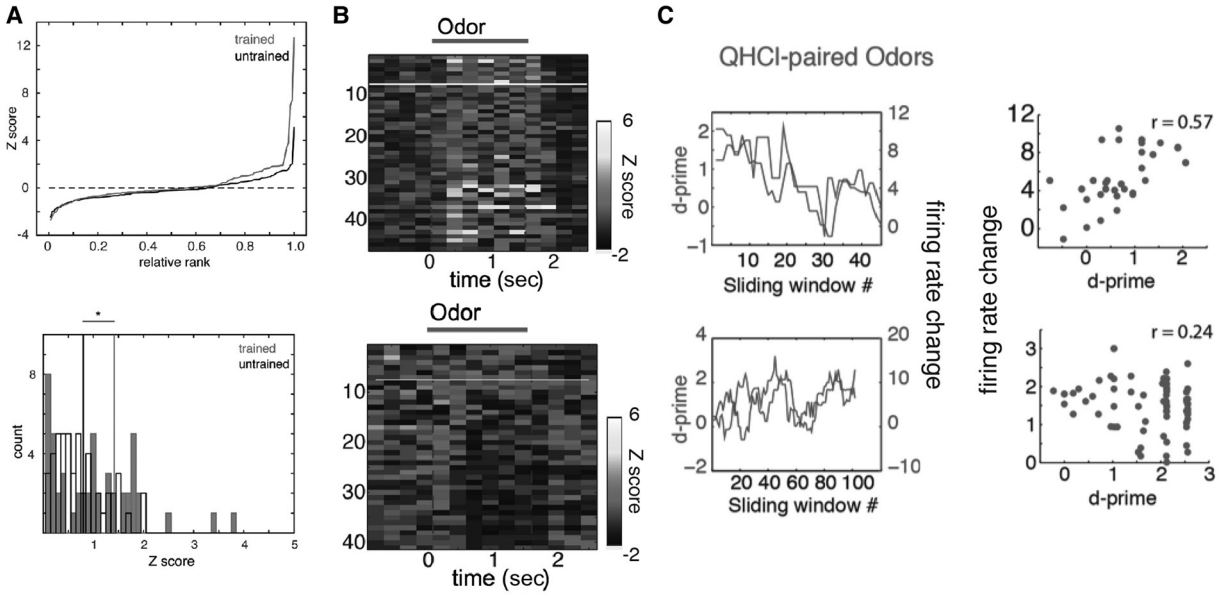


Figure 1. Recordings from granule cells (GCs) in the main olfactory bulb (MOB) of actively learning mice. (*A*) GC odor responses are stronger in trained mice than in mice that failed to learn the task. (*Upper*) Plot of response strength comparing trained and untrained mice. (*Lower*) Histogram of excitatory response strength comparing trained and untrained mice. (*$p < 0.05$, Mann-Whitney U test). (*B*) GC responses are dynamic during learning and track behavioral performance. (*Upper*) Two-dimensional peristimulus time histogram depicting the changes in a GC's response to an odor over approximately 50 trials. The white horizontal line indicates the time of the switch from the odor signaling water to the odor signaling quinine (QHCl). The blue dotted lines indicate odor onset and odor offset. (*Lower*) Another example of a different cell showing changes in its response to the odor following the transition to QHCl. (*C*) Example data from two cell-odor pairs for the correlation of behavioral performance with normalized GC activity. In each example, the *left* panel shows *d'* (accuracy) data (computed over a sliding 10-trial window) and response strength (computed as the average change from baseline firing during the odor) plotted over trials. The *right* panel is a scatterplot of *d'* data versus response strength.

Alexandra Nowlan, a new graduate student in the lab, is hoping to visualize auditory cortical plasticity day by day in a mouse learning to care for pups by literally peering through a window into the brain. Together with technician Clancy Kelahan, she has begun implanting glass windows that allow her to use neural imaging techniques to take daily snapshots of the brain's response to pup calls. Now that she has perfected the method, Ally will begin tracking changes in these patterns as a female mouse is growing more proficient in care of the pups. Another aspect of her project is her identification of a novel input to the auditory cortex from a region of the amygdala that processes social odors. Preliminary data Alexandra has gathered suggests that this pathway actually modulates sound processing according to these odors, and the pathway may be important for maternal learning. Our working model is that this pathway combines olfactory stimuli, such as the smell of pups, with auditory stimuli, such as the vocalizations of pups.

Postdoctoral fellows Billy Lau and Keerthi Krishnan led a related collaboration with CSHL professor Dr. Josh Huang. They examined how vocal perception of pup calls is affected in mice that are missing one copy of a gene called *MeCP2*. Impairments in the function of this gene are understood to cause the ASD Rett syndrome. Indeed, we find that females that possess only a single copy of *MeCP2* are not able to develop proficiency at gathering pups. In a paper published online at the end of 2016, we showed that this likely happens because MECP2 (the protein product of the gene) plays a critical role in maintenance and plasticity of the auditory cortex by acting on inhibitory interneurons. In a particularly exciting set of experiments, we were able to repair inhibitory function with genetic and pharmacological manipulations, thereby restoring maternal gathering behavior. This finding suggests that there may be a way to achieve cognitive improvement in humans, even after brain development.

Before Billy and Keerthi left CSHL at the end of 2016, they completed experiments for a second study focused on the specific activity changes in different types of auditory cortical neurons during maternal learning. Based on our recently published findings, we predicted that one of the central participants in auditory cortical plasticity is a network of inhibitory neurons that express a protein called parvalbumin (PV). Notably, deletion of *MeCP2* only in this minority of neurons is sufficient to disrupt pup care. Therefore, Billy made neuronal recordings in awake behaving females of both genotypes that differed in their maternal experience. The data show that when a normal adult female mouse is first exposed to pups, her auditory cortex becomes "disinhibited" (i.e., there is suppression of the inhibitory network). This seems to be attributable, specifically, to suppression of the PV neurons. In contrast, we observed no disinhibition in *MeCP2*-deficient mice. The results of this study are consistent with our model that MECP2 regulates plasticity in adults and juveniles through its effects on PV neurons.

Neural Activity Signaling Reward during Social Encounters

R. Dvorkin

The ability to detect and act on rewards is an indispensable component of adaptive behavior. Rewards often take the form of food, drugs of abuse, or, in the case of humans, currency. However, a variety of sensory social cues can also be rewarding. In fact, outside of the laboratory, social rewards likely guide a substantial fraction of natural motivated behaviors. Improper reward signaling in response to social cues may contribute to social behavior deficits and disinterest in social interaction in, for example, ASDs. To understand the pathological basis of this symptom, it is critical to identify the neural circuits that compute the value of social cues.

Social interactions and stimuli carry well-established reward value in humans and nonhuman primates. Extensive evidence establishes that engaging in social interactions, such as maternal behavior or sexual behavior, is powerfully rewarding to rodents, including mice. The desire for access to these activities can be harnessed to condition changes in behavior. Remarkably, early postpartum interactions with pups can even be a more powerful reward than cocaine. The rewarding properties of social interaction are at least partly driven by the rewarding nature of specific sensory cues experienced during a social encounter. Olfactory stimuli from conspecifics show robust reward characteristics in mice. For example, male-soiled bedding, adult male courtship vocalizations, and the ultrasonic cries of their pups are typically strongly attractive to female mice.

Former postdoctoral fellow Dennis Eckmeier successfully developed reliable methods for making stable recordings of individual neurons during actual social encounters between mice. This work is now being continued by postdoctoral fellow Roman Dvorkin. Their specific goal is to record from neurons in deep brain neuromodulatory centers during these encounters. Neurons that release neuromodulators, such as noradrenaline and dopamine, are likely responsive to rewarding social signals and may modulate encoding of sensory data. Understanding the context-dependent activity patterns of these neurotransmitters is therefore critical to developing models for how they affect behavior. They have recorded from a small brainstem structure called the "locus coeruleus" in both males and females during courtship. We anticipate that the activity will reflect both reward processing and motivational components.

In Press

Krishnan K, Lau BYB, Ewall GE, Huang ZJ, Shea SD. 2017. MeCP2 regulates cortical plasticity underlying a learned behavior in adult female mice. *Nat Commun* **18:** 14077.

SEX DIFFERENCES IN THE BRAIN: FROM GENES TO BEHAVIOR

J. Tollkuhn R. Bronstein M. Wu

Our lab seeks to understand the mechanisms that shape and regulate sex differences in the brain. Females and males differ in many behaviors and are differentially affected by mental health disorders, but the distinct developmental trajectories that give rise to these sex differences remain poorly understood. We couple genomic and molecular biology approaches with behavioral analyses to understand how gene regulatory events in early life have lasting effects on brain function and behavior. Males and females display different behavioral responses to the same social stimulus, such as the presence of a novel adult male or the cries of pups. These behaviors are mediated by sexually dimorphic neural circuitry, which develops under the control of the gonadal hormones estrogen and testosterone during a neonatal critical period. Manipulating hormone signaling during this critical period produces lasting effects on gene expression, brain wiring, and adult behavior.

Human males experience developmental testosterone surges, the intensity of which correlates with increased male-typical social behaviors and interests during childhood. Male mice similarly undergo a surge of testosterone at birth that rapidly subsides within the first few days of life. Our studies and those of others have found that in the brains of mice, it is the female sex hormone estrogen that is the primary mediator of the effects of neonatal testosterone. Estrogen is produced from testosterone by aromatase, which is expressed in discrete locations within the brain. Neonatal estrogen then acts to permanently establish the sex-specific neural circuitry that mediates behavior in the adult. Females given estrogen at birth display male-typical fighting behavior as adults with no additional hormone supplementation. Furthermore, neonatal estrogen is necessary and sufficient to produce sex-typical expression patterns of both its own receptor (ERα) and its androgen receptor (AR), the cognate receptor of testosterone. Both of these receptors are required for correct displays of sex-specific behaviors such as mating, aggression, and maternal care. As with such behaviors, the actions of neonatal estrogen on gene expression persist into adult life; switching hormonal profiles to that of the opposite sex does not produce corresponding alterations in AR or ERα expression. Thus, the actions of neonatal estrogen determine the capacity of both genes and circuits to respond to hormones in adulthood. These long-lasting effects suggest that neonatal estrogen organizes the developing brain through an epigenetic mechanism whereby the transient signal at birth irreversibly modifies the chromatin state of genes. The nature of this mechanism, as well as the identity of the genes regulated by gonadal hormones in the brain, is the focus of our current research program.

Genetic Dissection of a Sexually Dimorphic Neural Circuit

In males, genetic deletion of ERα abolishes fertility, alters hypothalamic-gonadal-pituitary (HPG) axis signaling, and impairs sexual and territorial behaviors. These innate behaviors are primarily regulated by hypothalamic regions that receive pheromonal chemosensory information from the accessory olfactory pathway. ERα is expressed in many of these areas, including the medial amygdala (MeA), the principal nucleus of the bed nucleus of the stria terminalis (BNST), the medial preoptic area (MPOA), and the ventrolateral region of the ventromedial hypothalamus (VMHvl). ERα-expressing neurons of the VMHvl and posterior ventral medial amygdala (MePV) are largely glutamatergic, whereas ERα cells in the MPOA, BNST, and posterior dorsal medial amygdala (MePD) are predominantly GABAergic. These regions are organized into a sexually dimorphic circuit whereby chemosensory information from conspecifics is sent from the accessory olfactory bulb to the MeA and BNST, which are interconnected and project to hypothalamic regions such as the POA and VMHvl, and additional output areas. This circuit mediates the display of both male and female innate social behaviors, but the role of ERα in orchestrating sex differences in this hierarchy remains unexplored.

As a first step in understanding the specificity of ERα signaling in the male brain, we have deleted this receptor in either excitatory glutamatergic neurons or inhibitory GABAergic neurons and assessed alterations in behavior and gene expression in mutant males. Surprisingly, we find that males lacking ERα in excitatory neurons display wild-type levels of male mating, aggression, and territory marking, indicating that although ERα is abundantly expressed in excitatory neurons, including those in the MePV and VMHvl, it is not required in these neurons for the masculinization of behavior. In contrast, deletion of ERα in GABAergic inhibitory neurons alters mating and aggression behavior and abolishes male-typical territorial marking. Twenty-five percent of mutant males attack females, a behavior never displayed in wild-type animals, suggesting that these mutants have difficulty discriminating between the sexes. Furthermore, males lacking ERα in inhibitory, but not excitatory, neurons have dysmasculinized gene expression of two key sexually dimorphic genes. AR is expressed more highly in males compared with females, whereas estrogen receptor beta (ERβ/Esr2) is more abundant in females. Mutant males show decreased AR and increased ERβ (Fig. 1). Together, these results suggest that ERα masculinizes the brain by organizing inhibitory inputs from BNST and MeA onto glutamatergic neurons that drive behavioral output.

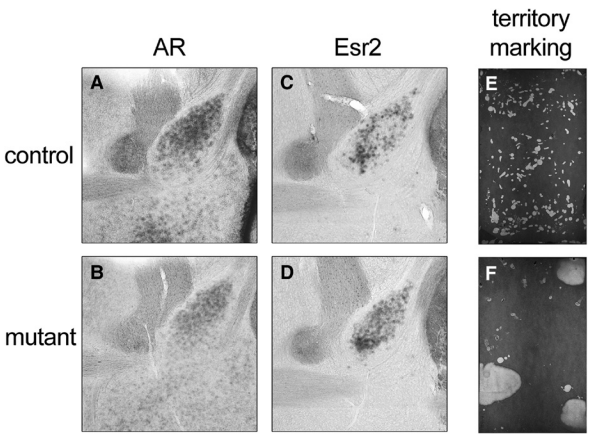


Figure 1. Loss of ERα in GABAergic neurons feminizes gene expression and behavior. AR (*A*) expression in the BNST is reduced in males lacking ERα in inhibitory neurons (*B*), whereas ERβ (*Esr2*) (*C*) expression is increased in mutant males (*D*). Male mice typically territory-mark a novel environment by dispersing urine spots over the area (*E*), whereas females pool their urine in corners. Mutant males pool urine in a female-typical pattern (*F*).

Identification of Sex-Specific Gene Programs in the Brain

To understand how neonatal estrogen directs sexual differentiation of the BNST and MeA, we have performed transcriptomic profiling of these regions in male and female pups. To circumvent the heterogeneity of the mammalian brain, we are performing our RNA-sequencing (RNA-seq) experiments specifically in ERα$^+$ neurons. This specificity is achieved through use of the RiboTag mouse allele that permits Cre-defined tagging of ribosomes, followed by purification of the affiliated translating mRNA. We have identified novel genes with sex-biased expression in these regions. We are particularly interested in Nfix, a member of the nuclear factor one (NFI) family of transcription factors. Nfix is expressed more highly in male BNST, and this expression pattern can be induced in females via a single dose of estradiol at birth. NFI factors are known to both regulate chromatin accessibility and interact with steroid receptors in nonneural tissues, but the role of Nfix in sexual differentiation of the brain has not been explored. We have initiated collaboration with Dr. Richard Gronostajski at State University of New York (SUNY) Buffalo to investigate the requirement of Nfix in brain masculinization. We plan to delete Nfix in ERα neurons and assess alterations in gene regulation and male behaviors.

Sex Differences in the Epigenome

Our lab is particularly interested in identifying the *cis*-regulatory elements that impart sex specificity to gene expression. To gain access to chromatin from our limited populations of ERα neurons, we are using a protocol termed isolation of nuclei tagged in specific cell types (INTACT). This approach uses a Cre-inducible nuclear envelope tag to permit purification of chromatin from genetically defined cell types. The chromatin accessibility landscape of a cell both reflects its developmental history and defines its current transcriptomic capacity. We are using the assay for transposase accessible chromatin (ATAC-seq) to identify accessible chromatin regions in ERα$^+$ neurons from BNST and MeA. ATAC-seq uses a modified Tn5 transposase to simultaneously cut DNA into small segments and insert adapters for high-throughput sequencing. As transposase preferentially cuts in nucleosome-depleted,

open chromatin areas, the regions sequenced represent functional genomic regulatory elements. Our initial experiments have identified approximately 72,000 accessible genomic regions in ERα neurons from BNST and approximately 63,000 from MeA. Only a few hundred peaks show a sex bias, and we find that these sex-biased peaks are enriched for specific transcription factor binding sites, including Ets, Rfx, and Forkhead family members. We are currently performing transcription-factor footprinting to characterize how these factors are directly bound to their motifs. We are also beginning our analyses in adult animals to determine how adult hormone profiles alter gene expression and chromatin accessibility. Collectively, these experiments identify promoters and enhancers that direct sex differences in gene expression, and implicate specific transcription factors in their regulation.

Many serious mental health disorders display a sex bias in incidence, age of onset, or presentation: Males are more likely to develop a neurodevelopmental disorder such as autism, schizophrenia, or language delay, whereas adolescent-onset mood disorders such as anxiety and depression are more common in females. There is a clear need for a mechanistic understanding of the molecular events that underlie such sex biases. Our goal is to identify genes and genomic regions that impart sex specificity to neuronal function and to reveal general principles that underlie hormonal regulation of differential gene expression between the sexes. Understanding the molecular processes underlying sexual differentiation of the brain will, therefore, reveal new insights into the developmental trajectories of psychiatric disorders, ultimately leading to novel targets and treatments.

NEURAL CODING AND MEMORY FORMATION IN THE *DROSOPHILA* OLFACTORY SYSTEM

G. Turner M. Brill M. Modi
T. Hige Y. Shuai

The brain has a tremendous capacity to form different memories. These memories are, in normal circumstances, highly accurate for particular stimuli and situations. How do patterns of neural activity in the brain convey this specificity? We address this question—the accuracy of memory—by studying the formation of olfactory memories in *Drosophila*.

We focus on a brain area known as the mushroom body (MB). This area is essential for learning Pavlovian associations between odors and reward or punishment. Unlike earlier layers of the olfactory circuit, MB neurons show highly odor-specific responses, and activity patterns in the MB are relatively sparse. This specificity is thought to underlie the accuracy of memory, as modifying the synapses of highly odor-specific neurons would lead to relatively precise memories. Sparse representations by highly stimulus-specific neurons are a general feature of brain areas involved in learning and memory, including the hippocampus and cerebellum.

Using the simplicity and genetic manipulability of *Drosophila*, our goal is to understand several fundamental properties of sparse representations. What mechanisms give rise to the stimulus specificity? What exactly is meaningful about the activity patterns? Does the precise timing of activity matter, or is it simply which cells respond? How are sparse activity patterns modified by learning?

Functional Connectomics in the Mushroom Body

T. Hige [in collaboration with S. Takemura, L. Scheffer, H. Hess, S. Plaza, P. Rivlin (the connectomics group), Janelia Research Campus]

We previously showed that the output synapses of the MB undergo plasticity during learning (Hige et al., *Neuron 88:* 985 [2015]). Using focused ion beam scanning electron microscopy (FIB-SEM), the connectomics group has reconstructed the complete set of synaptic connections within one of the output regions of the MB. This is the site where information about odor, carried by the MB neurons, converges with information about punishment and reward, carried by dopaminergic neurons. Dopaminergic input is essential to trigger plasticity of the synapses between Kenyon cells in the MB and the downstream MB output neurons. However, it is not clear which side of the synapse the dopamine acts on, the pre- or postsynaptic partner. The FIB-SEM results suggested that dopamine synapses on both Kenyon cells (KCs) and mushroom body output neurons (MBONs). We evaluated whether these anatomically identified synaptic connections are functional by optogenetically stimulating dopaminergic cells and monitoring for a response in the MBON with whole-cell recordings. We were able to show that this connection, not predicted by any of the previous work in the field, is indeed functional. This suggests that plasticity in this network may involve modifications of both pre- and postsynaptic elements, a hypothesis we are currently testing with our electrophysiological approach.

Visual Learning and the Mushroom Body

T. Hige

Previous work by R. Tanimoto's lab established that *Drosophila* are capable of forming Pavlovian associations between colored sectors of a behavioral arena and a reinforcing punishment. Lesion studies suggested that mushroom bodies were required for this type of learning, in addition to its long-established role in olfactory learning. However, no anatomical connection had been observed between visual areas of the *Drosophila* brain and the MB. Our collaborators in R. Tanimoto's lab identified the first putative visual inputs to the MB. These cells targeted an anatomically distinct subregion of MB. We identified genetic labels for this subset of MB neurons and examined

their stimulus response properties with intracellular recording techniques. We showed that this subpopulation of neurons indeed responds selectively to visual stimuli with different colors. Surprisingly, they showed no response to any olfactory stimuli we tested. These results, thus, established that visual and olfactory modalities both arrive in the MB, but in segregated populations of cells. These modalities are subsequently integrated downstream, in MB output neurons, which collect information across both visual and olfactory classes of cells.

Dopamine Controls the Signal-to-Noise Ratio of Signal Transmission through the Mushroom Body

M. Modi

Neuromodulators, such as dopamine and octopamine, are essential for learning, but how they act and what their precise role is during the learning process has not been established. We examined the effects of these neuromodulators on activity in the MB. Specifically, we tested the effects of increasing steady-state levels of dopamine by directly applying it to the brain. Surprisingly, we found that dopamine affects the signal-to-noise ratio of odor representations in the MB. Preliminary results suggest that dopamine increases the reliability of MB responses. Reliability is difficult to achieve in sparse representations because neurons typically respond with small numbers of spikes. Thus, it seems particularly useful for the circuit to use neuromodulators to increase signal-to-noise depending on the behavioral context the animal is in—essentially, dopamine could control the salience of a stimulus.

Additionally, we showed that changing baseline levels of dopamine diminishes odor responses in the neurons that are downstream of the MB. Together, these results suggest a model in which the dopamine acts to make the circuit a selective filter, increasing the signal-to-noise ratio of responses in the MB while decreasing signal transmission through the MB. These results may be an excellent entry point for studying the fundamental cognitive process of attention, which is essentially a process of filtering out distracting information to focus on pertinent inputs. We are now testing this possibility and investigating the precise mechanisms of the dopamine effects.

Is the Timing of Spikes Important for the Neural Code in the Olfactory System?

M. Brill

Neuronal signals propagate from one cell to the next via electrical impulses, or spikes, that evoke neurotransmitter release. These trains of spikes can be considered like the ones and zeros of the binary code of a computer. Spike trains have many different features that could potentially carry information—the rate of spikes, the timing of the onset, the relative timing of firing across neurons. Although spike rate is widely recognized as important, other features of those spike trains more related to spike timing could also be important signaling factors. Understanding what features of these spike trains convey information is a fundamental aspect of understanding the brain. It is a little like finding the key to a computer's binary code.

We have begun investigating this fundamental question by constructing artificial spike trains with these different features (rate vs. latency) present in different degrees. Using optogenetic techniques, we can control the activity patterns at an early layer of the olfactory system. By then monitoring downstream layers of the circuit using in vivo imaging techniques, we can assess what features of those artificial spike trains are effectively transmitted to those downstream neurons (i.e., what features of the code are relevant). The project is at an early stage, and we are currently focusing on the importance of the onset of spiking activity. If inhibition is slow and takes time to develop, early-onset spikes could potentially be more effective at driving downstream circuitry than those arriving late. Combining this optogenetic control of spike trains with measures of the impact on behavior could reveal which spike train features are important for the perception of the animal, getting to the core features of the neural code in this system.

PUBLICATION

Vogt K, Aso Y, Hige T, Knapek S, Ichinose T, Friedrich AB, Turner GC, Rubin GM, Tanimoto H. 2016. Direct neural pathways convey distinct visual information to *Drosophila* mushroom bodies. *Elife* **15:** 5.

In Press

Takemura S-Y, Aso Y, Hige T, Wong A, Lu Z, Xu CS, Rivlin P, Hess H, Zhao T, Parag T, et al. 2017. A connectome of a learning and memory center in the adult *Drosophila* brain. *Elife* (in press).

CIRCUITRY UNDERLYING CORTICAL PROCESSING AND DECISION-MAKING

A. Zador F. Burbach, F. Carnevale, X. Chen, G. Henry, L. Huang, J. Kebschull, F. Marbach, A. Reid, C. Stoneking, Y. Sun, V. Vagin, A. Vaughan, H. Zhan, A. Zhang

The Zador laboratory is interested in how neural circuits underlie cognition, with a focus on the cortical circuits underlying auditory processing and decision-making. To address these questions, we use a combination of molecular, electrophysiological, imaging, and behavioral approaches. In addition, we are developing a method that will allow us to use high-throughput DNA sequencing to determine the wiring diagram of the brain at single-neuron resolution—rapidly and efficiently, at low cost.

MAPseq of Locus Coeruleus

J. Kebschull, A. Reid [in collaboration with P. Silva, I. Peikon, and D. Albeanu, CSHL]

Neurons transmit information to distant brain regions via long-range axonal projections. In the mouse, area-to-area connections have only been systematically mapped using bulk labeling techniques, which obscure the diverse projections of intermingled single neurons. We have developed MAPseq (multiplexed analysis of projections by sequencing), a technique that can map the projections of thousands or even millions of single neurons by labeling large sets of neurons with random RNA sequences ("barcodes"). Axons are filled with barcode mRNA, each putative projection area is dissected, and the barcode mRNA is extracted and sequenced. Applying MAPseq to the locus coeruleus (LC), we find that individual LC neurons have preferred cortical targets. By recasting neuroanatomy, which is traditionally viewed as a problem of microscopy, as a problem of sequencing, MAPseq harnesses advances in sequencing technology to permit high-throughput interrogation of brain circuits.

MAPseq of Visual Cortex

J. Kebschull [in collaboration with T. Mrsic-Flogel, University of Basel, and others]

The wiring diagram of the neocortex determines how information is processed across dozens of cortical areas. Each area communicates with multiple others via extensive long-range axonal projections, but the logic of interareal information transfer remains unresolved. In sensory neocortex, previous work suggests that neurons typically innervate single cortical areas, implying that information is distributed via ensembles of dedicated pathways. Alternatively, single neurons could broadcast information to multiple cortical targets. Distinguishing between these models has been challenging because the projection patterns of only a few individual neurons have been reconstructed. Here, we map the projection patterns of axonal arbors from 595 individual neurons in mouse primary visual cortex (V1) using two complementary methods: whole-brain fluorescence-based axonal tracing and high-throughput DNA sequencing of genetically barcoded neurons (MAPseq). Although our results confirm the existence of dedicated projections to certain cortical areas, we find that these are the exception, and that the majority of V1 neurons broadcast information to multiple cortical targets. Furthermore, broadcasting cells do not project to all targets randomly, but rather they comprise subpopulations that either avoid or preferentially innervate specific subsets of cortical areas. Our data argue against a model of dedicated lines of interareal information transfer via "one neuron–one target area" mapping. Instead, long-range communication between a sensory cortical area and its targets is based on a principle whereby individual neurons copy information to, and potentially coordinate activity across, specific subsets of cortical areas.

Mapping Neuronal Projections Using In Situ Barcode Sequencing

X. Chen, G. Henry, L. Huang, J. Kebschull, Y. Sun, H. Zhan

We have previously developed MAPseq to map neuronal projections at cellular resolution. In MAPseq, individual neurons are labeled with unique RNA barcodes, and the projection patterns of neurons are read out by sequencing the barcodes present in a particular brain area. Although MAPseq preserves cellular resolution, the locations of the neurons are lost during barcode sequencing. To retain the locations of the neurons, we performed MAPseq by sequencing the barcodes in the neuronal cell bodies in situ, thus preserving the locations of the barcoded neurons. Using in situ MAPseq, we mapped neuronal projections from the mouse auditory cortex. We found all major classes of projection neurons known in the auditory cortex, as well as interactions among projection targets within a class. Such interactions would have been challenging to find using conventional bulk tracing techniques. We are now extending in situ sequencing to also allow sequencing of barcodes at the axonal terminals and combining with transcriptomic identifiers.

High-Throughput Interrogation of Cortical Activities and Projections at Cellular Resolution

X. Chen, F. Marbach, C. Stoneking, A. Vaughan

Corticocortical projection neurons in the mouse V1 showed activities tailored to their projection targets. It is not known, however, if such target-specific functionality can be generalized to other types of cortical projection neurons or cortical neurons in other cortical areas. To answer this question, we are combining in vivo two-photon calcium imaging with in situ MAPseq to examine both the activities and the projection targets of cortical neurons. Compared with stand-alone in situ MAPseq, combining with in vivo imaging requires tissue sectioning with minimal distortion and tissue loss. Tape-transfer cryo-sectioning minimizes tissue distortion, but current methods do not stabilize the tissue enough and are incompatible with Illumina sequencing chemistry. We have adapted a tape-transfer system to be compatible with Illumina sequencing, and it is also better at stabilizing tissue during sequencing. Our preliminary results suggest that tissue sections produced with our system are able to maintain good sequencing signal over multiple sequencing cycles with minimal tissue loss. We are optimizing our tape-transfer system for sequencing efficiency and reproducibility.

SynSeq—High-Throughput Barcode Sequencing to Map Brain Connectivity

J. Kebschull, I. Peikon, D. Ravens, Y. Sun, V. Vagin

The mouse brain consists of hundreds of millions of neurons connected by billions of synapses. The precise patterns of these synaptic connections determine both normal and pathological behavior. However, there are at present no rapid and cost-effective methods for determining circuit wiring with single-synapse precision. We have been developing a novel approach, SynSeq, to map neural connectivity. SynSeq exploits high-throughput next-generation sequencing. In SynSeq, each neuron is labeled with a unique random RNA barcode. These barcodes are then dragged to both the pre- and postsynaptic terminals by modified synaptic proteins. The proteins are then joined covalently at the terminals, immunoprecipitated, and the pre- and post-barcodes are joined into a single piece of DNA for sequencing. The barcode pairs can then be easily read out as synaptic connections. A paper describing this approach was recently published.

Optogenetic Manipulation of Auditory Striatal Neurons during a Frequency Discrimination Task

F. Carnevale

Making decisions based on sensory evidence is a fundamental operation of the brain. However, the neural circuits underlying this key cognitive function remain elusive. We have previously shown that optogenetic activation and inactivation of corticostriatal neurons can bias behavioral choices during an auditory discrimination task (Znamenskiy and Zador, *Nature 497:* 482 [2013]). Others have shown that optogenetic manipulation of direct and indirect striatal neurons during other behaviors facilitates and inhibits movement (Tai et al., *Nat Neurosci 15:* 1281 [2012]). We are therefore testing whether manipulating neuronal activity in auditory striatum, a region of striatum receiving information from auditory cortex, can bias behavioral choices during auditory discrimination. By optogenetically activating and inactivating direct and indirect striatal

subpopulations selectively, we will test the general hypothesis that the corticostriatal pathway provides a mechanism for the formation of arbitrary sensorimotor transformations. This will help us constrain network-level models of the neural computations that take place during perceptual decision-making.

Fluorescent In Situ Sequencing of Brain Slices Embedded in Expanded Hydrogels

G. Henry

In the mouse, distinct regions of the brain are connected to each other by long-range axonal connections. Understanding the architecture of these connections is a critical problem in neuroscience because many behaviors require the coordinated action of multiple neuronal circuits, which are often located in distinct regions of the brain. Although long-range projections can be mapped by a variety of methods, throughput and cost often limit progress. A solution to this problem involves the use of high-throughput sequencing technologies. For example, a recent publication from the Zador laboratory describes MAPseq (Kebschull et al. 2016b). In this method, a population of neurons is infected with a viral library that encodes a vast array of RNA-based barcodes. The axonal projections of these neurons are then mapped by sequencing cDNA derived from both the source and various target regions. Although this method enables rapid projection mapping, we would also like to know the precise spatial distribution of barcodes. Toward this end, we hope to use the recently described fluorescence in situ sequencing of RNA (FISseq) method (Lee et al., *Nat Protoc 10:* 442 [2015]) to map the source region and MAPseq to map target regions. If FISseq mapping is sufficiently robust, we will also attempt to map both the source and target regions using this method alone. The hypothalamus and the regions of the brain connected to it will be the focus of this work.

Mapping Brain-Wide Corticocortical Projections at Single-Cell Resolution by Barcoded RNA Sequencing

L. Huang, J. Kebschull

Neurons transmit information to distant brain regions via long-range axonal projections. Elucidating neuronal projection patterns between various brain areas at single-cell resolution is crucial in understanding neural circuit functions. Conventional neuronal tracing methods (e.g., bulk tracing with fluorescence) are able to reveal mesoscale connectome between brain areas, but with obscure single-cell information. To date, an efficient method for mapping individual neuron projection in high throughput is lacking. To overcome this, we (Kebschull et al. 2016b) recently introduced MAPseq, a method that allowed us to determine the projections of hundreds of neurons in a single animal. MAPseq uses random sequences of RNA as barcodes to recast neuroanatomy into a form that allows us to exploit the tremendous gains in high-throughput DNA sequencing. We are scaling up MAPseq by two orders of magnitude, and applying it to determine the all-to-all projections of more than 50,000 cortical neurons in a single mouse, at single-neuron resolution. The brain-wide MAPseq approach has the potential to uncover the entire corticocortical projection structure at single-cell resolution, and will provide insight into principles that underlie the organization of cortical circuits.

Role of Primary Auditory versus Visual Cortices in Decision-Making

A. Zhang

Sensory information of different modalities is processed and extracted through distinct but analogous series of hierarchically organized brain areas. Visual information travels from the retina to the lateral geniculate nucleus of the thalamus and then on to V1. Auditory information travels similarly from the cochlea, through the medial geniculate nucleus of the thalamus, to primary auditory cortex (A1). Feature-encoding principles are fairly analogous at these early stages of processing, and perceptual decision-making work has begun to elucidate the mechanisms by which such features can then be associated with appropriate motor outputs. However, because of varied stimulus complexities, task demands, and animal models, it has been impossible to directly compare the mechanisms that underlie visual-based associations to those that underlie auditory-based associations.

To overcome these challenges, we have developed a novel visual discrimination task that uses spatial location as the task-relevant variable. We believe that this task is directly analogous, and thus permits comparison, to the existing auditory discrimination task currently used in the lab. Animals are able to acquire this

task quickly and to high levels of performance. We are now investigating the neuronal representations in V1 in performing animals at various time points during the task. We predict that V1 neurons will respond in this task in both similar and distinct ways compared with A1 neurons in the auditory task; where these similarities and differences lie will help us to understand and compare the contributions of primary sensory areas to a perceptual decision-making task.

Role of Auditory Cortex for Belief Updating

A. Funamizu, F. Marbach

In perceptual decision-making, optimal behavior requires integration of sensory evidence and prior knowledge about stimulus probability. Although a series of studies has shown how prior knowledge affects decisions, little is known about how humans and animals update this before adapting to changing and uncertain environments.

Recent studies hypothesize that the prior belief is represented in cerebral cortex. As the auditory cortex-to-striatum pathway is essential for auditory perceptual decisions, auditory cortex is a candidate region for representing the prior belief for auditory stimuli. Here, we train mice in an auditory discrimination task in which the probability of categories A and B stimulus presentations changes in blocks, and monitor population activity in the auditory cortex with two-photon microscopy.

Our preliminary results showed that the stimulus probability biased mice toward choices associated to high-probability stimuli. A state-prediction behavior model, which recursively updates stimulus probability, fits the observed bias in the behavior. We predict that updating of the prior belief happens in the auditory cortex and that the cortical column hierarchically estimates the posterior belief of the stimulus by combining sensory evidence and the prior belief in a Bayesian manner.

Simple and Fast Brain-Wide Screening for Plasticity between Spatially or Genetically Defined Cell Populations with Single-Synapse Resolution

S. Ghosh, H. Zhan

There is wide support for the hypothesis that changes in synaptic strength mediated by long-term potentiation (LTP) underlie many forms of learning and memory. However, there is currently no high-throughput method for discovering which synapses in which pathways are modified by learning. We are developing SYNPLA (SYNaptic Proximity Ligation Assay), a sensitive, specific and high-throughput method for detecting the synaptic plasticity among candidate neuronal populations. SYNPLA exploits the proximity ligation assay to detect synaptic colocalization of presynaptic neurexin with postsynaptic GluA1-containing AMPA receptors. GluA1-containing receptors rapidly traffic into the synapse after LTP. Synaptic colocalization of neurexin and GluA1, and hence the SYNPLA signal, is therefore increased at recently potentiated synapses. SYNPLA results in greatly amplified, punctate fluorescent signal, allowing the easy and high-throughput identification of labeled synapse using light microscopy. We have demonstrated the function of SYNPLA in cultures of dissociated hippocampal neurons, and are now working to increase signal-to-noise and throughput of SYNPLA in slice cultures and ex vivo tissue.

Corticostriatal Plasticity Underlying Learning of Stimulus–Motor Association in an Auditory Discrimination Task

S. Ghosh

Animals use complex sensory cues from their environment to make a variety of decisions in their lives. Such behavior requires coming together of sensory discrimination, decision-making, and appropriate motor actions. An auditory discrimination task called the "tonecloud" task was established in the lab to understand the brain circuits involved in such decision-making and how these circuits evolve during learning of sensorimotor associations. Previous studies from the lab have shown that the connections between auditory cortex and the auditory striatum are instrumental for an animal to perform this task (Znamenskiy and Zador, *Nature 497:* 482 [2013]). Moreover, learning of this discriminatory task results in a specific pattern of plasticity in this pathway, which is determined by the frequency tuning of the neurons and the corresponding motor movement (Xiong et al., *Nature 521:* 348 [2015]). Nevertheless, the striatal substrates of this plasticity and its effects on striatal circuitry remain largely unexplored. The striatum is an inhibitory nucleus in

the basal ganglia composed primarily of inhibitory projection neurons called the medium spiny neurons (MSNs). These neurons can express either D1- or D2-type dopamine receptors in which the D1-MSNs constitute the "direct pathway" and the D2-MSNs form the "indirect pathway." The balance between these two supposedly antagonistic pathways is considered critical in controlling movement (Kreitzer and Malenka, *Neuron 60:* 543 [2008]). We hypothesize that upon learning the tonecloud task, changes in strength of cortical synapses onto a given MSN occur depending on whether that neuron belongs to direct or indirect pathway and whether it is tuned to low- or high-frequency sounds. To test our hypothesis, we are training D1/D2-CRExAi14 mice injected with Channelrhodopsin in the auditory cortex on the tonecloud task. After learning, we will obtain acute brain slices from these animals and use patch-clamp recording to test the strength of cortical inputs on D1/D2 cells tuned to high- and low-frequency sounds.

PUBLICATIONS

Kebschull JM, Garcia da Silva P, Zador AM. 2016a. A new defective helper RNA to produce recombinant Sindbis virus that infects neurons but does not propagate. *Front Neuroanat* **10:** 56.

Kebschull JM, Garcia da Silva P, Reid AP, Peikon ID, Albeanu DF, Zador AM. 2016b. High-throughput mapping of single neuron projections by sequencing of barcoded RNA. *Neuron* **91:** 975–978.

In Press

Han Y, Kebschull J M, Campbell RAA, Cowan D, Imhof F, Zador AM, Mrsic-Flogel TD. 2017. A single-cell anatomical blueprint for intracortical information transfer from primary visual cortex. *bioRxiv* doi: 10.1101/148031.

Peikon I, Kebschull J, Vagin V, Ravens D, Sun YC, Brouzes E, Corrêa Jr I, Bressan D, Zador A. 2017. Using high-throughput barcode sequencing to efficiently map connectomes. *Nucleic Acids Res* doi: 10.1093/nar/gkx292.

NEURAL BASIS OF LEARNING AND MEMORY IN *DROSOPHILA*

Y. Zhong J. Beshel

Because my lab has moved to Tsinghua University at Beijing, my remaining National Institutes of Health (NIH) grant was transferred to another relevant lab at CSHL, and the remaining postdoc Dr. Jennifer Beshel moved to the Lab to continue her research. Therefore, the following research will be completed in Josh Dubnau's lab, although the major findings were achieved before the transfer, which resulted in a publication listed below. During 2016, I worked with Dr. Beshel to get the manuscript published.

Conserved Genetic and Neural Circuit to Modulate Obesity-Linked Behaviors

J. Beshel

Leptin, a typically adipose-derived "satiety hormone," has a well-established role in weight regulation. Here, we describe a functionally conserved model of genetically induced obesity in *Drosophila* by manipulating the fly leptin analog Unpaired 1 (upd1). Unexpectedly, cell-type-specific knockdown reveals upd1 in the brain, not the adipose tissue, and mediates obesity-related traits. Disrupting brain-derived upd1 in flies leads to all the hallmarks of mammalian obesity: increased attraction to food cues, increased food intake, and increased weight. These effects are mediated by domeless receptors on neurons expressing *Drosophila* Neuropeptide F, the orexigenic mammalian Neuropeptide Y homolog. In vivo two-photon imaging reveals upd1 and domeless receptors inhibit this hedonic signal in fed animals. Manipulations along this central circuit also create hypersensitivity to obesogenic conditions, emphasizing the critical interplay between biological predisposition and environment in overweight and obesity prevalence. We propose that adipose- and brain-derived upd/leptin may mediate different aspects of weight or body size regulation through distinct circuits.

In Press

Beshel J, Dubnau J, Zhong Y. 2017. A leptin analog locally produced in the brain acts via a conserved neural circuit to modulate obesity-linked behaviors in *Drosophila*. *Cell Metab* **25:** 208–217.

PLANT BIOLOGY

David Jackson and colleagues study genes and signals that regulate plant growth and architecture. They are investigating a unique way in which plant cells communicate, by transporting regulatory proteins via small channels called plasmodesmata. These channels, which direct the flow of nutrients and signals through growing tissues, are regulated during development. The team discovered a gene encoding a chaperonin, *CCT8*, that controls the transport of a transcription factor SHOOTMERISTEMLESS (STM) between cells in the plant stem cell niche, or meristem. STM is critical for stem cell maintenance, and studies of the *CCT8* gene indicate that movement of STM between cells is required for this function. The lab also continues to identify other genes that control plant architecture through effects on stem cell maintenance and identity, and their work has implications for crop yields. Recent examples include discovery of a subunit of a heterotrimeric G protein that is conserved throughout animals and plants, and their studies indicate that this gene controls stem cell proliferation. They have found that, in plants, the G protein interacts with a completely different class of receptors than in animals. Their discovery helps to explain how signaling from diverse receptors is achieved in plants. This year, they also showed that weak mutations in one of the receptor proteins can enhance seed production in maize, which could lead to yield increases. Separately, the lab has characterized system-wide networks of gene expression by means of "next-generation" profiling and chromatin immunoprecipitation methods that have revealed many new hypotheses in developmental networks controlling inflorescence development. They are also developing a collection of maize lines that can drive expression of any reporter or experimental gene in any tissue type—tools of great interest to maize researchers that are being made available to the broader scientific community, enabling experiments never before possible in crop plants.

Zachary Lippman's research focuses on the process of flowering and flower production, which is a major contributor to plant reproductive success and agricultural yield. By identifying genes that control how tomato plants produce their flowers in their characteristic repeated zigzag arrangement (e.g., tomatoes on a vine), Lippman's lab is addressing when and how flowering branches, known as inflorescences, develop on plants, particularly fruit-bearing plants. Of particular interest is how these "reproductive phase transitions" have contributed to the evolution of diverse inflorescence branching patterns in tomato's larger Solanaceae family, which includes plants that make just one flower, such as pepper and petunia, in each inflorescence, to plants whose inflorescences produce dozens of branches and hundreds of flowers, such as many wild species of tomato. Using a combination of genetic, genomic, and molecular approaches, Lippman is dissecting the gene networks that are responsible for the variation in inflorescence branching found in nature. He hopes to leverage these discoveries to improve crop yields. Already, his work on genes that are responsible for the production and activity of a universal flowering hormone, known as florigen, has resulted in novel approaches to fine-tune plant architecture and flower production, boosting yield beyond leading commercial varieties. To continue hunting for new genes, Lippman has adopted a systems-biology approach and next-generation sequencing technology to capture those genes that are active as stem cells mature from a vegetative to a reproductive state. Nearly 4000 genes were found to reflect the existence of a "maturation clock," and one of the clock genes, known as *Terminating Flower*, acts as a key regulator to maintain a progressive pace to flowering—which in turn dictates how many flowers are produced on each tomato inflorescence. Finally, the Lippman lab determined the genome sequence of the "currant tomato," the wild ancestor of larger-fruited cultivated tomatoes, to better understand how flower and fruit production changed during the process of crop domestication.

Epigenetic mechanisms of gene regulation—chemical and conformational changes to DNA and the chromatin that bundles it—have had an important impact on genome organization and inheritance and cell fate. These mechanisms are conserved in eukaryotes and provide an additional layer of information superimposed on the genetic code. **Robert Martienssen**, a pioneer in the study of epigenetics, investigates mechanisms involved in gene regulation and stem cell fate in yeast and model plants including *Arabidopsis* and maize. He and his colleagues have shed light on a phenomenon called position-effect variegation, caused by inactivation of a gene positioned near densely packed chromosomal material called heterochromatin. They have discovered that small RNA molecules arising from repeating genetic sequences program that heterochromatin. Martienssen and colleagues have described a remarkable process by which "companion cells" to sperm in plant pollen grains provide them with instructions that protect sperm DNA from transposon damage. They found that some of these instructions, or epigenetic marks, could be inherited in the next generation. With collaborators in Mexico, Martienssen has also coaxed *Arabidopsis*, a flowering plant, to produce egg cells without meiosis, an important step toward a long-time goal of plant breeding: generating clonal offspring to perpetuate hybrid vigor. The lab has also shown that when RNA polymerase II has transcribed a stretch of DNA, the RNA interference mechanism causes the enzyme to release its hold on the DNA and fall away. This allows the replication fork to progress smoothly and the DNA strands to be copied; histone-modifying proteins, which follow right along, establish heterochromatin. Martienssen's group also continues to work on problems related to the creation of plant-based biofuels. This year, as part of a collaborative project to generate a high-quality full genome map of the oil palm plant, Martienssen and his colleagues identified a single gene that controls the yield of oil palm trees. This discovery will increase yields and should lessen the environmental burden of oil palm production, which often threatens already endangered rainforest lands.

Plants and animals interact with their environment. Because plants are unable to move around, they are sensitive to their surrounding environment and modify their development according to external signals. Plants face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Yet, plants respond to these biotic and abiotic factors and survive substantial fluctuations in their environment. Plants also must balance the range of potential threats and benefits confronting them and should make appropriate decisions on resource allocation. Such adaptability is essential given the sessile nature of the plants. The mechanisms that underlie this adaptability likely involve complex signaling to generate the appropriate response. In some adaptive responses, for example, when the plants have to cope with climate change and increased competition for light, there is a decrease in productivity (yield, biomass) as the plant relocates resources to better adapt.

Vincent Pedmale lab's research seeks to determine the mechanisms behind how a plant perceives and successfully adapts to its environment. Pedmale and colleagues also aim to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield.

The growing tips of plants, called meristems, contain a population of stem cells that serve as a persistent source of daughter cells from which new organs, such as leaves, arise. **Marja Timmermans** and colleagues are studying the genetic networks that regulate plant stem cell activity. Using genomic approaches, they have defined gene expression signatures that distinguish indeterminate stem cells from their differentiating derivatives. They have also worked out the mechanism that suppresses stem cell fate to allow cells to differentiate and have shown that this process requires a highly conserved epigenetic gene-silencing mechanism. In particular, Timmermans' group has

shown that specific DNA-binding proteins mediate the recruitment of Polycomb repressive complexes to stem cell factors, an action that stably represses their expression in differentiating organs. This work addresses a major unresolved question in the field of epigenetics: How do Polycomb proteins, which do not bind DNA themselves, recognize defined targets? Plant stem cells also produce signals important for the patterning of lateral organs. The lab has discovered that small RNAs can traffic from cell to cell and are among the stem-cell-derived signals. They have found that polarity in leaves is established via opposing gradients of mobile small RNAs that act as morphogen-like signals. Their most recent findings identified a third small RNA gradient involved in maintenance of organ polarity. These findings illustrate the complexity with which small RNAs generate developmental patterns. Currently, they are investigating parameters of small RNA mobility and the unique patterning properties of resulting small RNA gradients. Mathematical modeling predicts that such gradients might serve to generate robustness during development.

DEVELOPMENTAL BIOLOGY—STEM CELLS, SIGNALING, AND CROP PLANT ARCHITECTURE

D. Jackson D. Chatterjee, H. Claeys, E. Demesa-Arevalo, T. Eck, P. Huang, B.I. Je, D.Z. Li, L. Liu, T. Skopelitis, A. Volz, J. Wang, Q. Wu, F. Xu, X. Xu, M. Yuan, S. Zafar

Our research asks how the growth of plants is controlled, with the ultimate goal of improving crop plant yield. We identify the genes, signals, and pathways that regulate plant architecture and development. Living organisms develop by carefully controlling the movement of information molecules between cells during their growth. We are interested in discovering the signals that carry this information, finding out how the signals are transmitted, and learning how they function.

A major focus of our lab has been identification of genes that control stem cells in plants. In the past year, we reported the discovery of a new pathway in stem cell control (using feedback signals from differentiating cells) that has a significant impact on crop yields. We also discovered the biological relevance of cell-to-cell transport of a nuclear protein, SHOOT MERISTEMLESS, and several new roles for G proteins in stem cells. We greatly expanded our use of CRISPR genome editing, which, as in many systems, is revolutionizing the way we do genetics. We also continue to generate tools for the maize research community by creating a collection of lines that can drive expression of any reporter or experimental gene in any tissue type. Such tools are of great interest to researchers and are being made available to the broader scientific community, enabling experiments never before possible in crop plants.

The Regulation of Meristem Size in Maize

B.I. Je, Q. Wu, F. Xu [in collaboration with R. Meeley, M. Komatsu, H. Sakai, DuPont Crop Genetics]

All plant organs derive from populations of stem cells called meristems. These stem cells have two purposes: to divide to replace themselves and to give rise to daughter cells, which will differentiate into lateral organs. Consequently, meristems must precisely control the size of the stem cell niche via a network of positive and negative feedback signals. A loss of function in a negative regulator of stem cell fate can result in an enlarged or fasciated meristem phenotype and a dramatic alteration in the morphology of the maize ear and tassel.

Maize is an excellent genetic model system because of a large collection of developmental mutants and a sequenced reference genome. Our lab has undertaken a forward genetic approach to identify key regulators of stem cell homeostasis and meristem size. Two previously cloned mutants, *fasciated ear2* and *thick-tassel dwarf1*, encode orthologs of the *Arabidopsis thaliana* genes *CLAVATA1* and *CLAVATA2*, indicating that the well-known *CLAVATA-WUSCHEL* regulatory feedback loop is conserved from dicots to monocots. However, little else is known about the control of this important developmental process in maize. Here, we describe our progress in identifying additional genes contributing to stem cell niche homeostasis.

In previous work, we found that the maize *COMPACT PLANT2* (*CT2*) gene, which encodes the predicted α subunit of a heterotrimeric GTPase (Gα), functions in the CLAVATA pathway to control meristem size through its interaction with FEA2. We performed in vitro enzyme activity assays to show that CT2 has GTP-binding and GTPase activity. To understand the significance of CT2 GTPase activity in meristem development, we generated a constitutively active (GTPase-dead) version of CT2 (CA-CT2) by introducing the Q223L mutation in the context of a native construct. Using yeast-3-hybrid, we showed that the Q223L mutation abolished the interaction between CT2 and the G$\beta\gamma$ dimer, suggesting that $CT2^{Q223L}$ is indeed constitutively active. Expression of *CA-CT2* in a *ct2* mutant background resulted in interesting phenotypes, including higher spikelet density and kernel row number (KRN), larger shoot apical and ear inflorescence meristems, and smaller leaf angles compared with normal sibs. These results

suggest that manipulation of the GTPase activity of maize Ga subunit is a promising approach to improve agronomic traits of maize.

In addition to CT2, the sole canonical Gα of maize, there are three "extra large G proteins" (XLGs), which have a domain showing homology with Gα, as well as additional domains. Maize also contains one *G*β and six *G*γ genes. To understand the functions of these G protein subunits, we used CRISPR-Cas9 to knock out the *G*β subunit gene, *ZmGB1*, and the three maize *XLGs*. Surprisingly, we found that both *Zmgb1* single and *xlg1/2/3* triple mutants were lethal at the seedling stage (Fig. 1). The *Zmgb1* mutants showed overaccumulation of H_2O_2, constitutive activation of MAP kinases, and up-regulation of *PR1* (*pathogenesis-related 1*) and *PR5*, two immune marker genes. These results suggest that *ZmGB1* mutation caused autoimmune symptoms. The mechanism underlying the lethality of *xlg1/2/3* triple mutants remains to be unraveled; however, our preliminary observation showed that they have defects in root development. Moreover, knocking out *XLG*s in a *ct2* mutant background enhanced its dwarf phenotype. Further studies will focus on understanding the roles of Gβ and XLGs in maize development and immune responses. Another fasciated ear mutant that we have cloned is *fasciated ear 3* (*fea3*). *FEA3* encodes a predicted leucine-rich repeat receptor-like protein, related to *FEA2*. In situ hybridization and imaging of red fluorescent protein (RFP)-tagged transgenic plants show that FEA3 is expressed in the organizing center (OC) of the shoot apical meristem (SAM) and leaf primordia, as well as in the root apical meristem. Remarkably, expression of maize WUSCHEL, a marker for the stem cell niche-organizing cells, spreads downward in *fea3* mutants, which is strikingly different from its response in the known *CLAVATA* stem cell mutants.

To determine whether FEA3 responds to a CLAVATA3 (CLV3)-related (CLE) peptide, we tested its sensitivity to different peptides. *fea3* mutants showed reduced peptide sensitivity, but interestingly responded to a different CLE peptide, ZmFCP1 (FLORAL ORGAN NUMBER2-like CLE protein 1), compared with FEA2. ZmFCP1 is expressed in leaf primordia, suggesting that it serves as a signal from differentiating cells to repress meristem growth via the FEA3 receptor. Our results suggest that the FEA3 receptor functions in a new pathway for stem cell control that is spatially distinct from the known CLV receptors and uses a different peptide signal. We also found that *FEA3* and *FCP1* homologs function in meristem control in *Arabidopsis*, suggesting this new pathway is universal in flowering plants. To better understand the functional mechanism of FEA3, we are using an *fea3* ethylmethanesulfonate (EMS) suppressor screen to identify potential downstream signaling components. Meanwhile, IP-MS (immunoprecipitation-mass spectrometry) with FEA3-tagged transgenic plants will be used

Figure 1. *Zmgb1* mutants are lethal at the seedling stage.

to find FEA3-interacting proteins. Plants carrying FEA3 fused with amino- or carboxy-terminal tandem affinity purification (TAP) tags (yellow fluorescent protein [YFP]-STREP tag) have been constructed and will be used for TAP followed by mass spectrometry (MS) to isolate FEA3 interactors. As a backup, FEA3-tagged lines were also crossed to a *branched silkless; Tunicate* (*bd-Tu*) double mutant, which makes a mass of proliferating meristems and results in strong enrichment of developmental protein complexes. In addition, CRISPR technology is being used to knock out FEA3 homologs to study the function of FEA3-like receptor-like proteins in maize. We already obtained different CRISPR alleles for four FEA3 homologs and made crosses with *fea3.* Characterization and mutant analysis will be performed to determine their potential function in meristem development regulation.

Weak alleles of fasciated ear mutants can improve maize yield traits, such as KRN, by increasing meristem size and number of primordia while maintaining structural integrity of the meristem. We found that *fea3* weak allele hybrids also enhance yield traits as well as overall yield. These results are particularly exciting because in our previous studies of weak *fea2* alleles we found an increase in KRN but no overall increase in ear weight because of a compensatory reduction in kernel size. Therefore, the newly identified *FEA3* signaling pathway could be used to develop new alleles for crop improvement.

Having found multiple receptors and CLE peptide ligands, we were intrigued by how specificity is obtained. Using peptide assays, we found that FEA2 also participates in ZmFCP1 signaling, but CT2 does not. CORYNE (CRN, a membrane-localized kinase) is a CLV2-interacting kinase first identified in *Arabidopsis*. We identified *Zmcrn* mutants and found that they also have a fasciated ear phenotype (Fig. 2). *Zmcrn* mutants were resistant to ZmFCP1 peptide treatment, and we also confirmed the molecular interaction between FEA2 and ZmCRN. In contrast, ZmCRN and CT2 did not interact physically, and our genetic evidence indicates ZmCRN acts in parallel to CT2. These data suggest that FEA2 is involved in perception of diverse peptides and transmits signals from different ligands to different downstream components.

Aberrant phyllotaxy 2 (*Abphyl2*) is a gain-of-function mutant in maize with enlarged shoot meristems and a decussate phyllotactic pattern. Previous studies in our lab showed that this mutation is caused by transposition of a glutaredoxin gene, *male sterile converted anther1* (*MSCA1*), resulting in altered expression pattern of *MSCA1* in *Abphyl2*. Consistently, *msca1* loss-of-function mutants have reduced meristem size. The growth defect of meristems in the *Abphyl2* and *msca1* mutants suggests a potential novel function of glutaredoxins in meristem development. Two MSCA1 close homologs, including

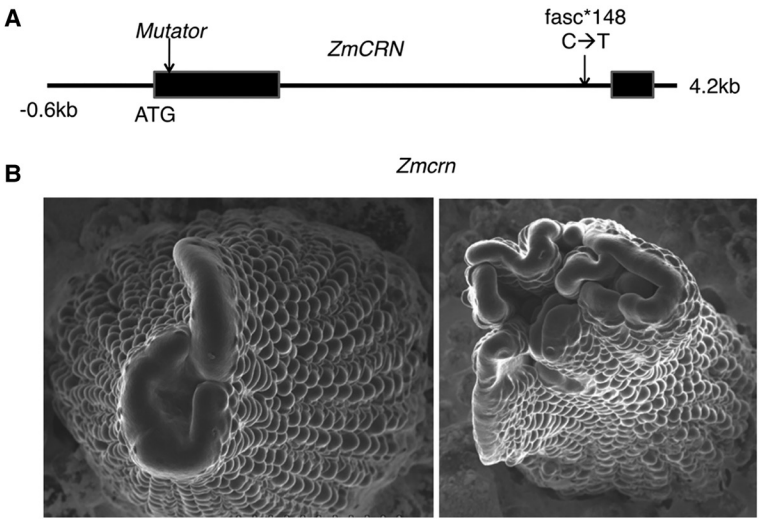


Figure 2. *Zmcrn* mutant phenotypes. (*A*) Schematic of the *Zmcrn* mutant allele. The arrows indicate two independent mutant alleles. (*B*) *Zmcrn* mutants show fasciated ear primordia.

GRX2 and GRX5, are also expressed in meristem. We used CRISPR to knock out both genes, and *grx2 grx5* double mutants showed deceased SAM size compared with wild type (WT), and *grx2 grx5 msca1* triple mutants showed reduced KRN and spikelet intensity, suggesting that GRX2 and GRX5 are also involved in meristem development. Interestingly, MSCA1 interacts with a TGA transcription factor (TF), *FASCIATED EAR4*, and these proteins overlap in expression in the nucleus, some cells of the SAM, and inflorescence meristem. Genetic analysis showed that *fea4* is epistatic to *msca1* with respect to SAM size, supporting that they interact in a pathway. To investigate how the activity of FEA4 is regulated by MSCA1 and its homologs, we crossed FEA4-YFP and FEA4-RFP transgenic plants with *msca1 grx2 grx5* triple mutants. Experiments are in process to investigate how MSCA1 regulates the function of FEA4 and other putative targets through posttranslational modification.

Beyond these advances, we are mapping additional weak fasciated ear mutants from EMS mutagenesis screens. We have determined candidate map positions for several of these mutants and are proceeding with fine mapping and molecular cloning. For one of the mutant *fea148*, we mapped the mutation to a region between 153 and 158 Mbp on chromosome 3. We found a C-to-T mutation, which led to an early stop codon in *ZmCRN1*. *ZmCRN1* encodes a membrane-localized kinase homologous to *Arabidopsis* CORYNE, which is essential in stem cell division and differentiation of the SAM. The identification of a new mutant allele of *Zmcrn1* in maize confirmed its role in regulating meristem development. In addition to the usual positional cloning approach, we are trying to identify additional *fea* mutants by whole genome sequencing.

Genetic Redundancy in Circuits Controlling Meristem Development in Plants

L. Liu, E. Demesa-Arevalo, B.I. Je, F. Xu, T. Skopelitis, [in collaboration with Z. Lippman, CSHL; M. Bartlett, University of Massachusetts, Amherst; Z. Nimchuck, University of North Carolina at Chapel Hill; B. Yang, Iowa State University]

The CLE (CLAVATA3/endosperm surrounding region–related) peptides are fundamental players in meristem maintenance in plants, as they are mobile signals that establish the feedback circuit signaling between the stem cell differentiation and division. Disruption of this pathway generates overproliferation in meristems or fasciation. In our lab, we have described different fasciated mutants encoding leucine-rich repeat receptor-like kinases or receptor-like proteins (LRR-RLK or RLP); however, the signals perceived by many of these receptors remain elusive. Genetic evidence in maize suggests a divergence in signaling pathways controlling meristem size. The putative orthologs in maize for CLAVATA1 (THICK TASSEL DWARF1, TD1) and CLAVATA2 (FASCIATED EAR2, FEA2) and FASCIATED EAR3 (FEA3), have synergistic effects, suggesting additional pathways control meristem size. Recently, 49 CLE peptide genes were identified in maize, suggesting either specialization or redundancy in these signaling pathways.

To analyze the role of CLE peptides in maize meristem regulation and their involvement in redundant circuits, we are generating CRISPR knockouts. We first analyzed expression patterns from publicly available datasets combined with our transcriptional profiles from meristematic tissues, and identified 27 candidates to be likely expressed in SAM and inflorescence meristems. So far, we have made CRISPR alleles for approximately 10 of them. Mutant alleles for ZmCLE7 showed ear fasciation and increased number of silks, phenotypes already described for *fea2* and *td1*, including overproliferation of unfertilized ovules characteristic of *glf-td1* (Fig. 3). *Zmfcp1* mutants showed fasciation in tassels and male sterility as a result of anther abortion (Fig. 3). To identify *Zm*CLE peptides that regulate meristem size in a redundant way, we are crossing our mutant alleles to make higher order mutants, including with mutants in receptors. Analyzing genetic interactions between CLE ligands and receptors will help us to identify the ligands for these putative receptors.

In *Arabidopsis*, the CLV3-CLAVATA1 (CLV1) ligand-receptor kinase pair negatively regulates meristem cell proliferation, and the *clv1* mutant phenotypes are enhanced by mutations in the related receptor kinases *BARELY ANY MERISTEM 1*, *2*, and *3* (*BAM1*, *BAM2*, and *BAM3*). In maize, we identified 7 *CLAVATA1* or *BAM*-like genes, including the known gene *TD1*. To characterize the function of BAM-like genes in maize, we are using CRISPR-Cas9 to create mutations of all of them. The genetic interactions between different BAM-like genes and CLEs will allow us to dissect the signaling pathways from CLEs to BAMs for meristem maintenance in maize.

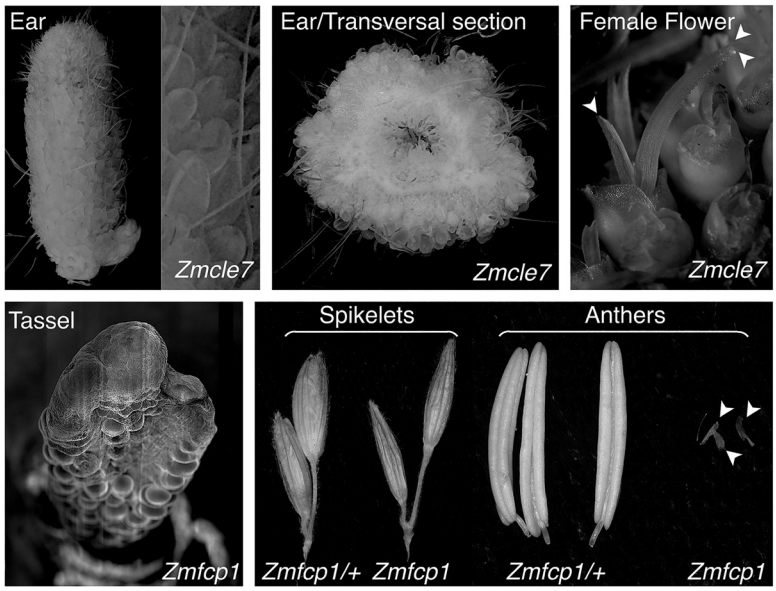


Figure 3. Phenotypes obtained by CRISPR knockouts of ZmCLE7 and ZmFCP1. (*Upper* panel) *Zmcle7* ears are strongly fasciated and show a "glassfingers" phenotype (close-up); they also develop an internal cavity because of meristem overproliferation (transverse section); female flowers of *Zmcle7* also had three silks instead of two (arrowheads). (*Lower* panel) *Zmfcp1* tassels are mildly fasciated and later showed male sterility as a result of anther abortion (arrowheads).

Control of Shoot Branching and Determinacy

H. Claeys, E. Demesa-Arevalo, X. Xu, M. Yuan, T. Skopelitis [in collaboration with M. Komatsu, H. Sakai, DuPont Crop Genetics; B. Yang, Iowa State University]

The *RAMOSA* (*RA*) genes in maize function to impose determinacy on axillary meristem growth; consequently, *ra* loss-of-function mutants (*ra1*, *ra2*, and *ra3*) have more highly branched inflorescences. *RA3* encodes a trehalose phosphate phosphatase, which catalyzes the conversion of trehalose-6-phosphate (T6P) to trehalose. T6P is an important regulatory metabolite that connects sucrose levels, and thus the sugar status, to plant growth and development, but its mode of action is still unclear. *RA3* is expressed at the base of axillary inflorescence meristems, and localizes to distinct puncta in both nuclear and cytoplasmic compartments, suggesting that its effect on development may not be simply metabolic. These data support the hypothesis that *RA* genes may serve as mediators of signals, maybe a sugar signal, originated at the boundary domain and regulating determinacy. *RA3* itself may have a transcriptional regulatory function, because it affects the expression of specific genes.

We are taking genetic approaches to identify factors that act in the same pathway with *RA3*, by screening for enhancers of the *ra3* phenotype. Typically, *ra3* mutants have three to eight branches only at the base of the ear. We mutagenized *ra3* mutants and looked for plants that have more branches and/or have branches in the upper part of the ear. So far, 20 mutants have been isolated, and mapping is under way. So far, four independent mutant alleles of *REL3* (*RAMOSA ENHANCER LOCUS 3*), a *RA3* paralog, were identified, and we confirmed that *REL3* is the causative gene using additional CRISPR-Cas9-generated alleles. All EMS-induced alleles contain single amino acid substitutions that affect their enzymatic activity, explaining the phenotype. We are currently investigating the role of other RA3 paralogs using CRISPR-Cas9. Another mutant was mapped to a gene encoding an RNA-binding protein that is known to play a role in inflorescence development, and we are also focusing on understanding this genetic interaction. Other putatively identified mutants are being confirmed.

In a parallel approach to understand how RA3 functions, we are screening for interacting proteins. To this end, a yeast 2 hybrid screen was performed,

and a number of predominantly nuclear proteins were found to interact with RA3, which fits with its partially nuclear localization. In parallel, ear primordia from plants expressing a FLAG-HA-tagged RA3 protein were used for IP-MS, revealing a number of potential in vivo interactors. To overcome limited tissue availability, we crossed the FLAG-HA-tagged RA3 maize lines with a double mutant, *bd*;*Tu,* that transforms the ear into a "cauliflower" with massively overproliferating meristems. The biological roles of a number of the interactors are currently being studied using insertional mutants and mutant alleles generated using CRISPR-Cas9. We are analyzing CRISPR alleles in two families of RA3 interactors: the RNA recognition motif (RRM) *Zea mays* SCAFFOLD ATTACHMENT FACTOR B (*Zm*SAFB) and ZmSAFB-like and the vascular plant one-zinc-finger (*Zm*VOZ) TFs, to analyze their genetic interactions with *ra3.* Preliminary data showed *Zmsafb ra3* mutants have two times more branches in ear and tassels compared with the single *ra3* mutants. Additionally, we generated mRFP1-*Zm*SAFB and ZmVOZ1-RFP transgenic lines to characterize localization patterns in ear development of these candidate interactors. Using immunolabeling, we analyzed mRFP1-ZmSAFB subnuclear localization in detail, and found it is excluded from heterochromatin, and enriched in euchromatin (in which it is partially colocalized with the active form of RNA polymerase II), suggesting *Zm*SAFB functions in transcriptional activation (Fig. 4).

Finally, we are combining a number of approaches to directly understand the contribution of the enzymatic function of RA3 to its biological function. We are complementing the *ra3* mutant with a catalytically dead mutant version of RA3 generated by BAC recombineering. We are also using CRISPR-Cas9 to generate mutants in TREHALOSE PHOSPHATE SYNTHASES (TPSs), which catalyze the preceding step to RA3 in the trehalose metabolic pathway. If accumulation of T6P is causing the *ra3* phenotype, then combining *ra3* with *tps* mutants should alleviate the phenotype, giving important insights into its mechanism.

Natural Variation in Inflorescence Architecture

H. Claeys, Q. Wu

Maize inflorescence architecture has been a target for extensive selection by breeders since domestication; hence, different maize inbreds vary greatly in these traits. The genetic basis underlying this diversity is largely unknown, but is of great interest for both fundamental and applied science. To identify natural variation relevant to inflorescence traits, we looked for inbred backgrounds that can enhance or suppress the phenotypes of different mutants. We focused on the 25 NAM (nested association mapping) founder inbreds because they were selected to capture the diversity of maize germplasm, and because of the genetic tools available for these lines. We have crossed these 25 inbreds to our collection of mutants (in a B73 background) and screened the F2 for plants with suppressed or enhanced phenotypes compared with the mutant in the B73 background.

fea2-0 is strongly enhanced in the NC350 background (Fig. 5). We mapped this enhancement to a single major-effect locus on chromosome 5 using both bulked segregant analysis and crosses to NC350/B73 recombinant inbred lines (RILs). The region we identified contains *THICK TASSEL DWARF1* (*TD1*), a gene that is known to be involved in meristem size regulation, and we found that the NC350 allele of *TD1* is expressed at a lower level than the B73 allele, strongly suggesting that natural variation in this gene causes the phenotype. Interestingly, we have also been able to show using B73-NC350 heterogeneous inbred families (HIFs) that the NC350 allele of *TD1* positively

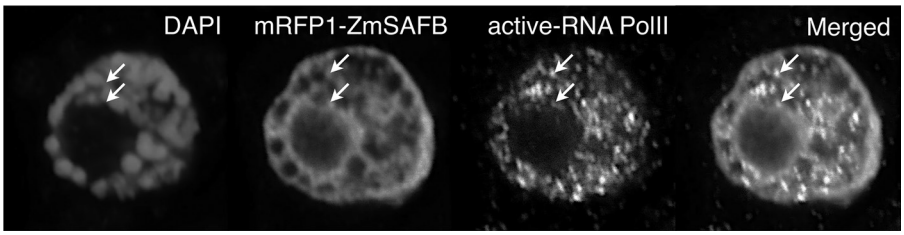


Figure 4. Immunofluorescence in mRFP-ZmSAFB transgenic lines. mRFP1-ZmSAFB is colocalized with the active form of RNA POL II (arrows) in the nuclei.

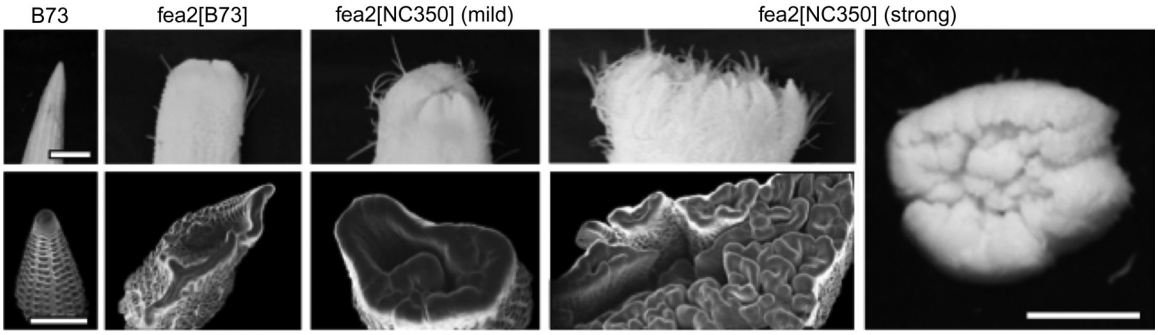


Figure 5. Enhancement of *fea2-0* in the NC350 background. *fea2-0* ear tips are enlarged and flattened in B73 compared with the wild type. However, in the NC350 background, we see a range of much more severe phenotypes. (*Upper* panels) Pictures of unfertilized ears. Scale bar, 1 cm. (*Lower* panels) Scanning electron microscopy of immature ears. Scale bar, 1 mm.

affects KRN, demonstrating its usefulness in plant breeding. This confirms our lab's recent findings that hypomorphic alleles of genes involved in meristem size regulation lead to increased KRN, linking fasciation and KRN. We are currently investigating how NC350 ensures normal meristem development by looking at how other genes in this network have adapted to compensate for the weak allele of *td1*. Additionally, we found that Ki11 enhances the *ra3* phenotype, but this enhancement is more complex, with several underlying loci. Nonetheless, we obtained rough map locations for two enhancers and one suppressor from Ki11, and are generating mapping populations for further fine-mapping.

We also screened for natural modifiers of *ct2*, and found it was dramatically enhanced in NC350 and HP301 backgrounds and suppressed in CML69. In the enhanced plants, the ear inflorescence meristems were extremely fasciated and showed multiple branches. By using scanning electron microscopy, the enhanced fasciation phenotype was detectable at very early stages of ear development (~2 mm ear length). The *ct2* suppressors in the CML69 background only suppressed the fasciated ear phenotype, whereas other *ct2* phenotypes, such as dwarf, wide and erect leaves, and dense spikelets were not affected, indicating the suppressor may function specifically in ear inflorescence meristem regulation. The bulk segregant analysis (BSA) results suggest that there are multiple loci that contribute to the modification of *ct2* phenotypes. We are backcrossing the modified plants with *ct2* in B73 background to dissect the loci that are contributing to the suppressed or enhanced phenotypes of *ct2*.

The Effects of Drought on Early Inflorescence Development

H. Claeys [in collaboration with H. Cline, B. Meeley, DuPont Crop Genetics]

Drought stress is one of the major environmental factors limiting maize yield. Some progress has been made in studying how drought affects grain filling during the later stages of reproductive development, but almost nothing is known about how drought affects early inflorescence development, in which the number of grains is being determined. It has been found that yield is significantly affected when maize is subjected to drought at this stage. Therefore, we set out to address this question in collaboration with DuPont-Pioneer and the Eveland lab at the Danforth Center, the Vollbrecht lab at Iowa State University (ISU), the Dinneny lab at the Carnegie Institution, and the Hake lab at University of California (UC)-Berkeley.

Last year, we completed a large pilot experiment in which we profiled ear tips from plants grown in automated greenhouses at a Pioneer facility in Iowa. Water was withheld from half of the plants during early ear development. This revealed a large transcriptional reprogramming of inflorescence meristems in response to drought, affecting many known developmental regulators. We identified the gibberellin and T6P pathways as potential mediators of this response, and preliminary follow-up experiments have shown that ears of *ra3* mutants, which are affected in T6P metabolism, are more affected by drought. We are currently performing a second large transcriptome experiment to identify early signaling events after drought with greater developmental resolution. We

have also identified inbreds with different sensitivity of ear growth to drought, which we plan to profile in the second phase of our drought experiments.

Generation of a pOp-LhG4 Transactivation System for High-Throughput Protein Localization and Functional Studies in Maize

E. Demesa-Arevalo, Q. Wu, B.I. Je, T. Skopelitis
[in collaboration with A. Chan, J. Craig Venter Institute; A. Sylvester, University of Wyoming]

Despite the abundance of resources available for the study of various model organisms, the classical approach of characterizing single gene mutants still provides invaluable information regarding gene function. Visual information, in regard to time and location of gene expression, as well as the ability to study gene products at the protein level, complement the genetic approaches to study the biological role of a gene. We previously generated transgenic lines tagging various genes with fluorescent markers using their endogenous promoters. Using our experience in promoter analysis, we are establishing tools for complex analysis at cellular, tissue, or organ level. We developed a pOp-LhG4 transactivation system in maize that allows us to express proteins in specific tissues and/or transient stages during development. The pOp-LhG4 system includes (1) a transcription activator, LhG4, which is a fusion between a high-affinity DNA-binding mutant of lac repressor, *LacI*^{His17}, and transcription-activation-domain-II of GAL4 from *Saccharomyces cerevisiae*; and (2) a chimeric promoter, pOp, that consists of *lac* operators cloned upstream of a minimal CaMV promoter, not activated in the reporter lines until crossed with the LhG4 activator line. We have already generated 35 LhG4 drivers, as well as eight pOp responder lines. We confirmed the reliability of our system by crossing pOp::ZCN8 plants with a constitutive promoter line, pEF1A::LhG4, phenocopying the early flowering phenotype of ZCN8 overexpression. This system has been used to test new hypotheses in meristem regulation by differentiated cells; expressing pOp::*Zm*FCP1 using a leaf primordia-specific driver line, pYABBY14::LhG4, strongly inhibited meristem growth, confirming this feedback regulation. We also continue to develop fluorescent lines targeting different meristem domains, and new markers include: translational fusions for ZmWUSCHEL1 (ZmWUS1-mRFP1) and TERMINAL EAR1 (TE1-YFP) expressed in specific domains. Additionally, we are collaborating to standardize fluorescence assisted cell sorting (FACS) protocols using our root tissue-specific marker lines p*Zm*WOX5:4XNLS-tagRFPt, p*Zm*SHR:4XNLS-tagRFPt, and p*Zm*SCR:4XNLS-tagRFPt. Currently, we have close to 200 transgenic lines, including translational fusions, promoter fusions, promoter activator lines (LhG4), and responder lines (pOp). Data on the characterization of our FP and LhG4 transgenic lines, including confocal micrographs, movies, and recent publications, can be found on our website: http://maize.jcvi.org/cellgenomics; seed are also available through the Maize Genetic Cooperation Stock Center.

Mechanism of Active Transport of TF through Plasmodesmata

M. Kitagawa, D. Kumar, T. Skopelitis, J. Wang, R. Balkunde
[in collaboration with S. Harrington, C. Faulkner, John Innes Centre, United Kingdom]

In plants, certain TFs are actively and selectively transported between cells to target cells to specify their fates. These TFs are transported through plasmodesmata (PD), membrane-lined channels traversing the cell wall. To this date, however, the mechanism underlying the active and selective transport of TFs through PD has been largely unknown. Previously, we established a system for evaluating the capacity of the active transport of TFs in *Arabidopsis* seedlings using a mobile homeodomain TF, KNOTTED1 (KN1). Using this system, we isolated mutants in which the active transport of TFs may be defective. One such mutant encodes an aspartyl-transfer RNA synthetase (asp-RS), an enzyme that attaches L-aspartate onto its transfer RNA (tRNA) to function in translation. Interestingly, we found enlarged SAM and fasciation of stems in this mutant (Fig. 6), suggesting this asp-RS regulates SAM development. Next, we observed the expression domain of *WUSCHEL* (*WUS*), which is a marker gene of the OC of the SAM, and found that the OC was expanded or split in the enlarged mutant SAMs (Fig. 6), suggesting that this asp-RS regulates SAM organization. We next examined the subcellular localization of this asp-RS by fluorescent protein (FP) fusions and found a punctate pattern, implying that this asp-RS is located in specific organelles (Fig. 6).

In the past year, we isolated two new mutants in which the active transport of TFs could be defective,

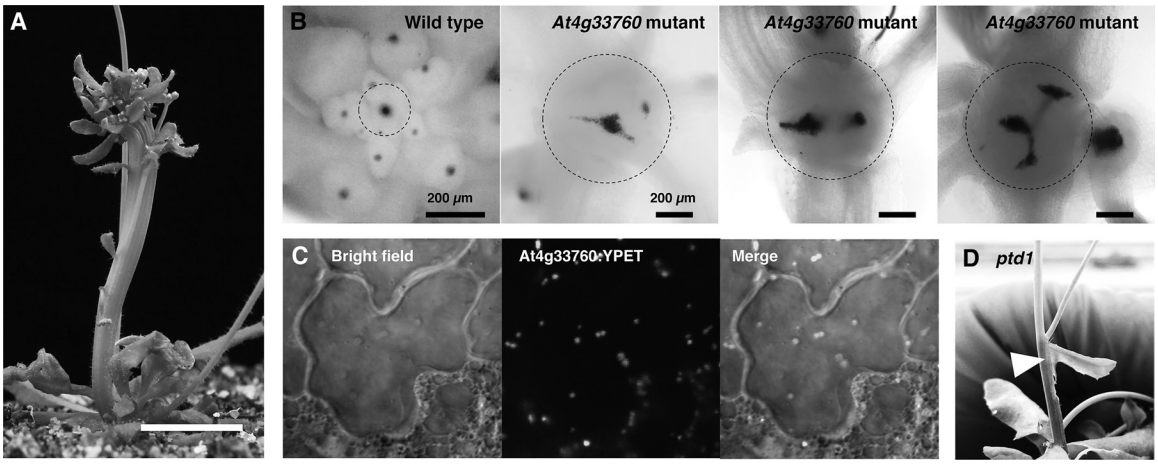


Figure 6. Functional analysis of *asp-RS* and *ptd1* phenotype. (*A*) Fasciated stem of *asp-RS* mutant. (*B*) Expression domain of *WUS* (*WUS* promoter::β-*glucuronidase* [*GUS*]) in the SAM of wild-type and *asp-RS* mutants. (*C*) Subcellular localization of *asp-RS*-YFP in a leaf epidermal cell. (*D*) Fusion between a leaf and a stem in *ptd1*.

plasmodesmata targeted-transport defective (*ptd*) *1* and *ptd2*, by the screen described above. In these mutants, the capacity of KN1 transport between cells is decreased, and fusions between leaf and stem can be found in *ptd1* (Fig. 6). We previously found that blocking active transport of the *Arabidopsis* KN1 homolog SHOOTMERISTEMLESS (STM) shows similar phenotypes, implying that the active transport of particular TFs, including STM, may be defective in *ptd1*.

We are also attempting to reveal the mechanism underlying regulation of PD-mediated signaling by a reverse genetics approach. We are focusing on PD-localized receptor-like proteins of the PD LOCATED PROTEIN (PDLP) family. *Arabidopsis* has eight PDLPs, and one of them, PDLP5, facilitates β-1,3-glucan (callose) deposition at PD to regulate PD-mediated signaling by salicylic acid signaling during plant-pathogen interactions. In contrast, PDLP function in plant development is unknown. Previously, we made a triple mutant *pdlp1-pdlp2-pdlp3*, but it had no obvious growth or developmental phenotype, suggesting that PDLPs may act redundantly. Therefore, we are aiming to knock out all *PDLPs* via CRISPR-Cas9, in collaboration with Sophie Harrington and Christine Faulkner at the John Innes Centre, United Kingdom. We transformed a CRISPR-Cas9 construct containing single guide RNAs for all *PDLP*s (*PDLP1–8*) into *Arabidopsis* and confirmed that the gene editing has occurred at least in *PDLP5*, *-6*, and *-8*.

Comprehensive Annotation of the Maize Genome Function by FACS and Chromatin Immunoprecipitation–Sequencing

X. Xu, E. Demesa-Arevalo, T. Skopelitis

To fully develop the functional annotation of maize genome sequence as a resource for the scientific community, we are conducting the profiling of B73 cell types by FACS and genome-wide TF-binding analysis by chromatin immunoprecipitation sequencing (ChIP-seq). These are important goals of the MaizeCODE project, which is an initial analysis of functional elements in the maize genome.

FACS has provided valuable cell-type specific messenger RNA (mRNA) profiling for the *Arabidopsis* and rice community. In maize, we are now using the FP promoter lines developed by our lab to isolate specific cell types or domains for a subset of profiling experiments, including PolII and histone modification. The current FP promoter lines cover many different cellular domains such as central zone (pZmWUSCHEL1-mRFP), boundary zone (pZmRAMOSA3-RFP), and endodermal zone (pZmSCARECROW4-TagRFP). To generate enough tissue for protoplasting and sorting, we are currently crossing these FP lines into the *bd*;*Tu* mutant that can produce large amounts of meristem tissue. Once tissues are harvested, we will extract RNA from the sorted protoplasts for RNA-seq and chromatin modification and small RNAs analysis.

To provide genome-wide TF-binding site analysis, we are also conducting ChIP-seq for diverse TFs that

function in different aspects of maize growth and development. Currently, we are working on approximately 12 TFs that belong to several different families. For example, we have the MADS box TF, TUNICATE1A (TU1A), the ZINC FINGER (ZF) TF, RAMOSA1 (RA1), the bZIP TF, FASCIATED EAR 4 (FEA4), and the GATA TF, TASSELSHEATH1 (TSH1). We have generated transgenic lines for these TFs with different FP tags such as YFP or RFP. To overcome the limited tissue availability, we are also crossing these lines into *bd*;*Tu* mutants. We are currently working on the ChIP using green fluorescent protein (GFP) or RFP antibodies to pull down the TF-DNA complex. Future work will use next-generation sequencing and peak calling to identify the specific binding sites of these TFs.

PUBLICATIONS

Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson D, Kausch AP, et al. 2016. Advancing crop transformation in the era of genome editing. *Plant Cell* **28:** 1510–1520.

Baker RF, Leach KA, Boyer NR, Swyers MJ, Benitez-Alfonso Y, Skopelitis T, Luo A, Sylvester A, Jackson D, Braun DM. 2016. Sucrose transporter ZmSut1 expression and localization uncover new insights into sucrose phloem loading. *Plant Physiol* **172:** 1876–1898.

Je BI, Gruel J, Lee YK, Bommert P, Arevalo ED, Eveland AL, Wu Q, Goldshmidt A, Meeley R, Bartlett M, et al. 2016. Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nat Genet* **48:** 785–791.

Somssich M, Je BI, Simon R, Jackson D. 2016. CLAVATA-WUSCHEL signaling in the shoot meristem. *Development* **143:** 3238–3248.

Urano D, Maruta N, Trusov Y, Stoian R, Wu Q, Liang Y, Jaiswal DK, Thung L, Jackson D, Botella JR, Jones AM. 2016. Saltational evolution of the heterotrimeric G protein signaling mechanisms in the plant kingdom. *Sci Signal* **9:** ra93.

Urano D, Miura K, Wu Q, Iwasaki Y, Jackson D, Jones AM. 2016. Plant morphology of heterotrimeric G protein mutants. *Plant Cell Physiol* **57:** 437–445.

In Press

Balkunde R, Kitagawa M, Xu XM, Wang J, Jackson D. 2017. SHOOT MERISTEMLESS trafficking controls axillary meristem formation, meristem size and organ boundaries in *Arabidopsis*. *Plant J* **90:** 435–446.

Jackson D. 2017. No sex please, we're (in)breeding. *EMBO J News Views* (in press).

PLANT DEVELOPMENTAL BIOLOGY, STEM CELLS, AND FLORAL BRANCHING SYSTEMS

Z.B. Lippman J. Dalyrymple C.-T. Kwon S. Soyk
A. Hendelman Z. Lemmon C. Xu
A. Krainer D. Rodriquez-Leal S. Zhou

Our lab seeks to reveal, understand, and manipulate the genes that control when, where, and how flowers are produced on plants. Flowers form on inflorescences, the reproductive branches that originate from stem cells located in growing tips called meristems. The number of inflorescences produced on a plant, as well as how many branches and flowers form on each inflorescence, can vary dramatically both within and between species. At the center of this diversity lie two critical processes of stem cell regulation that are the focus of research in the lab: (i) stem cell maturation, during which stem cells transition from a vegetative to a reproductive growth program, and (ii) stem cell proliferation, which controls stem cell population size. We are taking advantage of extensive natural and mutant variation in inflorescence production and architecture in tomato and related nightshades to explore how differences in stem cell maturation and population size explain the remarkable diversity of inflorescence production, complexity, flower production, and reproductive success observed in nature and agriculture.

Variation in the Flowering Gene *SP5G* Promotes Day Neutrality and Early Yield in Tomato

S. Soyk, A. Krainer

Over the last two decades, hundreds of studies on crop domestication have shown that human actions have selected for diverse, often species-specific changes in plant growth and development. Prominent examples include reduced branching in maize, loss of seed shattering in rice, and increased fruit size in tomato, to name but a few. Yet, one common target that allowed the expansion of crops beyond their centers of origin was modifying when and how plants produce flowers in response to the environment. Specifically, the wild ancestors of most crops flower according to seasonal changes in day length (day sensitivity), which often limits cultivation to a narrow geographical range. The basic mechanisms underlying day sensitivity have been studied extensively in the model *Arabidopsis*. However, for many major crops and their wild ancestors, the genes and mechanisms that were selected to allow worldwide food production are unknown.

By uniting quantitative genetics with CRISPR gene-editing technology, we have discovered a major gene that controls day sensitivity in wild tomatoes, and whose activity was modified at the transcriptional level to facilitate the introduction of cultivated tomato in the 16th century to Europe from the equatorial region of South America where the wild ancestor of tomato was domesticated. Surprisingly, we found this critical adaptation arose before domestication, from a regulatory change in a flowering repressor gene, *SELF PRUNING 5G* (*SP5G*), that is closely related to the universal flowering hormone florigen. By dissecting the mechanism of how *SP5G* inhibits flowering, we revealed a strategy to engineer more compact tomato varieties that flower rapidly regardless of day length and bear fruits faster that most cultivars. Significantly, we then translated these findings to agriculture by demonstrating that this novel "double-determinate" growth habit provides a highly desirable trait of early yield for field tomatoes. Our work illustrates how discoveries in basic mechanisms of plant growth and development can be united with new gene-editing technology to allow rapid and immediate improvement of elite germplasm in plant breeding.

The Evolution of Inflorescence Diversity in the Nightshades and Heterochrony during Meristem Maturation

Z. Lemmon

The amazing diversity of shoot architecture found in flowering plants is an impressive example of morphological evolution. Over the past several decades, a

combination of forward and reverse genetics has revealed key genes and pathways that control both vegetative and inflorescence shoot architecture. More recently, genome-wide methods have been used to reveal and begin dissecting the gene regulatory networks underlying this diversity. In all plants, inflorescences arise from vegetative meristems that mature gradually to reproductive states in response to environmental and endogenous signals. The morphology of the shoot meristem maturation process is conserved across distantly related plants, raising the question of how diverse inflorescence architectures arise from seemingly common maturation programs.

In tomato and related *Solanaceae* species, inflorescence complexity is reflected in a continuum from solitary flowers to highly branched structures bearing hundreds of flowers. Because reproductive barriers between even closely related *Solanaceae* have precluded a genetic dissection, we captured and compared meristem maturation transcriptomes from five domesticated and wild species reflecting the evolutionary range of *Solanaceae* inflorescence architecture. We found that these divergent species share hundreds of dynamically expressed genes, highly enriched for transcription factors that define a series of vegetative, transition, and floral molecular maturation states. Quantitative differences in transcriptome variation among the species pointed to changes in maturation schedules during evolution, explaining differences in inflorescence architecture and flower production. Notably, maturation schedules were marked by a high level of transcriptome expression divergence during the reproductive transition, primarily driven by heterochronic shifts of transcriptional regulators with known roles in flowering. Heterochronic expression of dynamic genes during this critical transitional window of shoot development modulates the rate at which shoots acquire floral fate. Thus, evolutionary

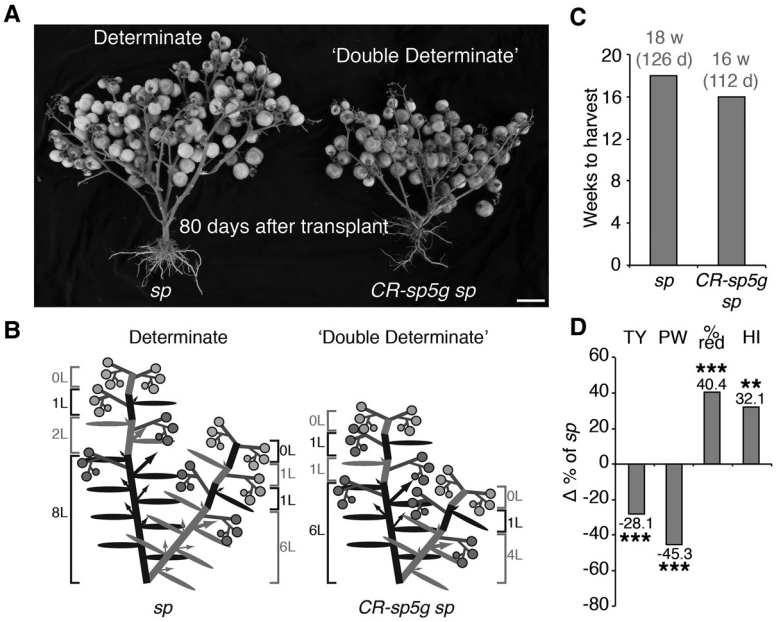


Figure 1. Introducing *sp5g* mutations into the *sp* determinate background results in highly compact "double determinate" plants with early yields. (*A*) Representative *sp* and *CR-sp5g sp* plants 80 days after transplanting, showing early fruit ripening on *CR-sp5g sp* double determinate plants. Leaves were removed to show fruit set. (*B*) Diagrams depicting the shoot architectures of *sp* and *CR-sp5g sp* plants. Bars and ovals represent shoots and leaves. Gray and black bars represent successive sympodial shoots. Arrows represent axillary shoots, and green lines depict inflorescences, with colored circles represent maturing fruits. The number of leaves on primary, axillary, and sympodial shoots is indicated. (*C*) Number of weeks to final harvest (90% red fruits) in the *sp* and *CR-sp5g sp* genotypes. *CR-sp5g sp* plants reach final harvest 2 weeks faster than *sp* plants. (*D*) Total yield (TY), plant weight (PW), percentage of red fruits (%red), and harvest index (HI) from *CR-sp5g sp* plants as percentage difference (Δ%) to *sp* controls 78 days after transplanting. Mean values (±s.e.m.) were compared with *sp* using *t* tests (**$p < 0.01$, ***$p < 0.001$). Scale bar, 10 cm.

diversity for inflorescence architecture in the nightshades is driven by subtle alterations in transcriptome dynamics during a critical transitional window of meristem maturation. More broadly, our findings parallel the recent description of a transcriptome inverse hourglass model for animal embryogenesis, suggesting both plant and animal morphological variation is guided by a mid-development period of transcriptome divergence.

Exploring the Role of Meristem Size in the Diversity of Pepper Plant Form

Z. Lemmon

We are continuing to study shoot and inflorescence architecture variation in peppers to broaden our knowledge and perspective on agriculturally important traits in this important solanaceous crop. We are integrating quantitative genetic, genomic, and reverse genetics tools to investigate the molecular networks directing *Capsicum* meristem development and resulting shoot and flower production. Genetic diversity within and between pepper species is being assayed with high-throughput genotyping by sequencing technology on more than 400 accessions. We have also developed multiple biparental mapping populations and shown that several architectural traits segregate in a manner suggesting involvement of only a few loci. By integrating genome-wide association studies (GWASs) with classical quantitative trait locus (QTL) mapping, we hope to identify the genes and polymorphisms underlying pepper shoot and inflorescence architecture variation. This work will enable comparative developmental studies with tomato and related single-flower edible *Solanaceae*, such as *Physalis peruviana*

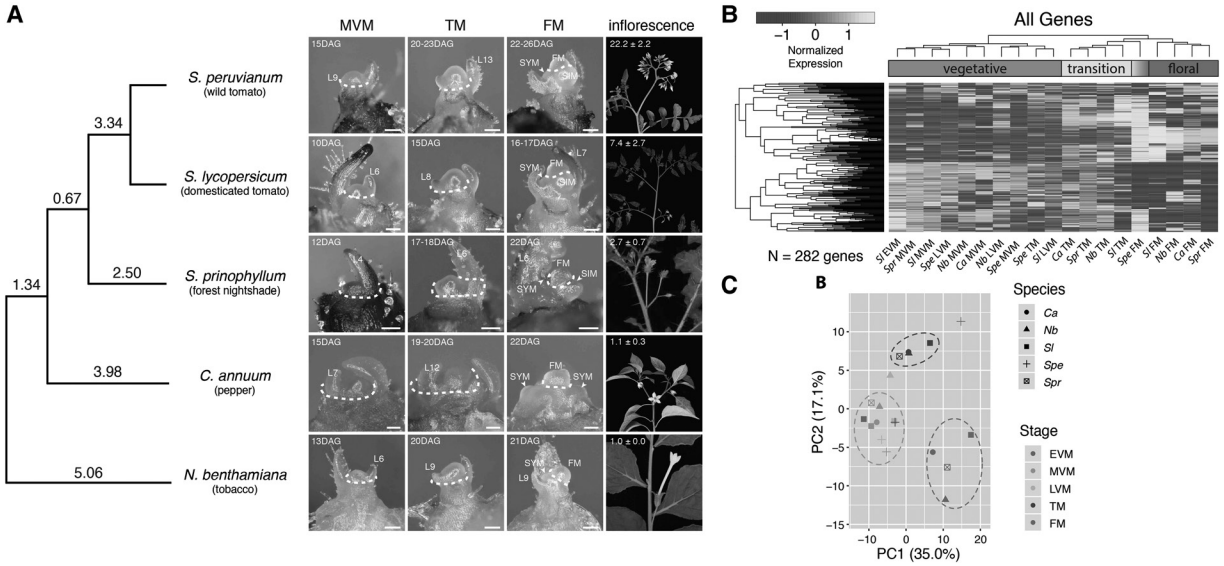


Figure 2. The SlBOP transcriptional regulators interact with TMF to control meristem maturation and inflorescence. Phylogeny, meristem ontogeny, and variation in maturation schedules in five *Solanaceae* species representing a quantitative range of inflorescence complexity. (*A*) Phylogeny of *Solanum peruvianum* (*Spe*, wild tomato species), *Solanum lycopersicum* (*Sl*, domesticated tomato), *Solanum prinophyllum* (*Spr*, forest nightshade), *Capsicum annuum* (*Ca*, pepper), and *Nicotiana benthamiana* (*Nb,* tobacco). Shown at right are stereoscope images of the middle vegetative meristem (MVM), transition meristem (TM), floral meristem (FM), and inflorescence from the five species arranged from most complex (top) to least complex (bottom) inflorescence architecture. Average number of flowers produced (± standard deviation, SD) by the mature inflorescence is shown in the inflorescence frame. (*B*) Dynamically expressed genes were *z*-score normalized within each species, and hierarchical clustering was visualized by heatmap. (*C*) Principal component analysis (PCA) of normalized expression. The first two principal components (PCs) account for ~52% of the overall variation, with PC1 primarily accounting for differences between vegetative and floral meristem stages. PC2 primarily separates the transition stage from vegetative and floral. Regions in which the majority of vegetative, transition, and floral samples cluster together are indicated by dashed ovals. Three notable outliers from the main vegetative, transition, and floral groups include the *Spe* TM and FM, and the *Nb* LVM, reflecting heterochronic shifts in expression that imply a delay (*Spe*) and acceleration (*Nb*) in meristem maturation.

(Cape gooseberry), which we are developing into a new model system through genomic sequencing and CRISPR with collaborators at other institutions.

Control of Inflorescence Architecture in Tomato by BTB/POZ Transcriptional Regulators

C. Xu

Several years ago we discovered the *TERMINATING FLOWER* (*TMF*) gene, encoding a transcriptional regulator that acts in the earliest stages of meristem maturation to prevent precocious flower formation and ensure the development of multiflowered inflorescences. *TMF* defined a new flowering pathway, but the molecular mechanism was unclear. We discovered that TMF protein functions in transcriptional complexes with a family of cofactors known as the BLADE-ON-PETIOLE (BOP) proteins, defined by the classical BTB/POZ interaction domain that has widespread importance in both plant and animal development. Although there have been dozens of studies on the *BOP* family in diverse species, little is known about the mechanisms of BOP protein function—and surprisingly, prominent roles in flowering or inflorescence complexity have not been found. By using CRISPR-Cas9 to create a complete series of single- and higher-order tomato *BOP* mutants, we found that, along with shared functions with other species, all three tomato *BOP* genes are major regulators of flowering and inflorescence complexity. Most notably, eliminating *BOP* activity entirely resulted in the transformation of multiflowered inflorescences into single flowers, revealing a critical role for the *BOP* family in modulating meristem maturation (Fig. 3). Our findings provide the first insights into a new molecular mechanism underlying one of the many developmental roles of the deeply conserved BOP family. More broadly, our findings address a wide range of

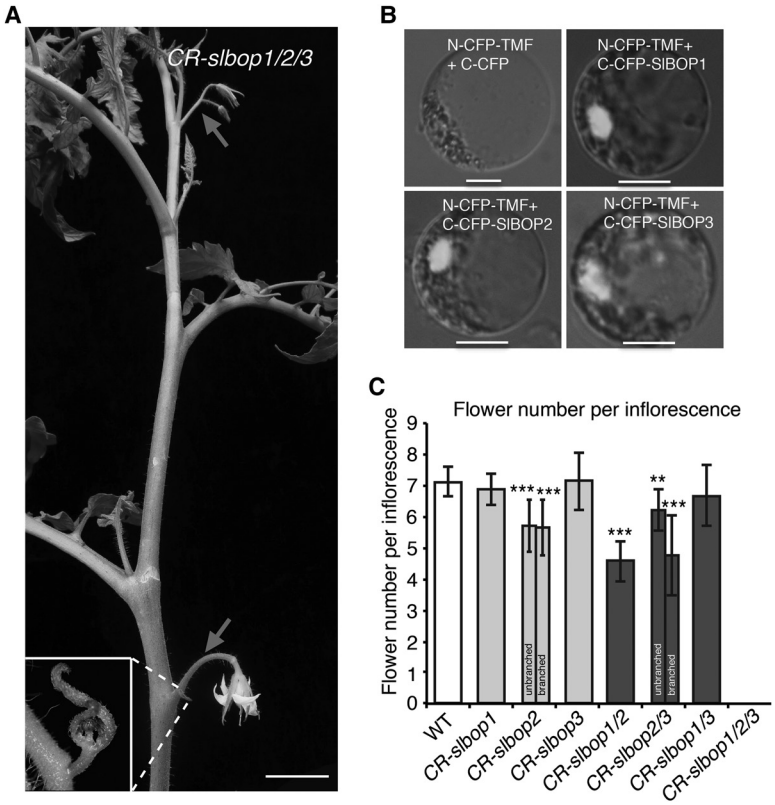


Figure 3. The SlBOP transcriptional regulators interact with TMF to control meristem maturation and inflorescence complexity. (*A*) Triple-knockout mutant of *SlBOP1/2/3* using CRISPR-Cas9 results in plants that produce inflorescences with 1–2 flowers. (*B*) BiFC in tomato protoplasts showing all three SlBOP proteins interact with TMF. (*C*) Quantification of flowers per inflorescence in all single, double, and triple *slbop* mutants generated by CRISPR.

questions on the links among the reproductive transition, stem cell fate, and diversity of inflorescence architecture and flower production in plants. Future identification of the targets of these transcriptional complexes will facilitate uncovering gene networks that control inflorescence complexity.

Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Locus

S. Soyk, Z. Lemmon

A recurring target during crop domestication and improvement was modification of inflorescence complexity. In major cereal crops like rice, wheat, corn, and barley, humans selected for inflorescences with greater complexity to increase flower and grain production. Yet, for many crops—particularly fruit-bearing species like tomato—inflorescence architecture has remained largely unchanged from wild ancestors. Improving tomato inflorescence architecture has remained challenging despite a century of intensive breeding and a rich resource of valuable variation in distantly related wild tomato relatives. Specifically, genetic incompatibilities and the challenge of transferring complex polygenic traits have precluded exploiting wild species to improve the tomato inflorescence.

As our work has shown, quantitative diversity in *Solanaceae* inflorescence architecture depends on the rate at which meristems mature from a vegetative to a reproductive state. Our working hypothesis is that weak delays in the maturation process could lead to slightly more complex inflorescences. Thus, modification of this developmental program could potentially be exploited to create beneficial changes in tomato inflorescence architecture. By screening a large tomato core collection of more than 6000 accessions, we identified a class of moderately branched inflorescence variants that carry mutations in two related MADS-box transcription factor genes. One mutation arose early in domestication and alone enlarged the leaf-like organs on fruits and was selected as fruit size increased. The other mutation arose within the last century and eliminated the flower abscission zone on the fruit stem, providing the classical "jointless" trait that reduced fruit drop and facilitated mechanical harvesting in large-scale tomato production. Each mutation was selected independently for improved flower and fruit traits, but redundancy between these two transcription factors caused undesirable branching when breeding for both traits. Surprisingly, we found that many modern elite breeding inbreds and cultivated hybrids carry both mutations, showing that breeders were able to overcome undesirable branching by selecting for suppressors. However, these suppressors imposed a barrier to improving inflorescence architecture and flower production. By dissecting the genetic interaction between a collection of natural and engineered mutations in these MADS-box genes and a family member, we created a range of inflorescence complexity, which facilitated breeding of higher-yielding hybrids. This work suggests that resolving similar cases of negative epistasis in both plant and animal breeding could help improve productivity in agriculture.

PUBLICATIONS

Lemmon Z, Park SJ, Jiang K, Van Eck J, Schatz M, Lippman ZB. 2016. The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res* **26:** 1676–1686.

Xu C, Park S, Van Eck J, Lippman ZB. 2016. Control of inflorescence architecture in tomato by BTB/POZ transcriptional regulators. *Genes Dev* **30:** 2048–2061.

In Press

Soyk S, Muller N, Park SJ, Schmalenbach I, Jiang K, Hayama R, Zhang L, Van Eck J, Jimenez-Gomez JM, Lippman ZB. 2017. Variation in the flowering gene *SELF PRUNING 5G* promotes day-neutrality and early yield in tomato. *Nature Genet* **49:** 162–168.

EPIGENETIC INHERITANCE IN PLANTS AND FISSION YEAST

R. Martienssen
B. Berube
S. Bhattacharjee
F. de Sousa Borges
E. Ernst
D. Grimanelli
M. Gutbrod
J.-J. Han
R. Herridge
H.S. Kim
S.-C. Lee
A. Molla-Morales
J.-S. Parent
M. Regulski
J. Ren
B. Roche
A. Schorn
A. Shimada
J. Simorowski
R. Umamaheswari
C.J. Underwood

Plants and fission yeast provide excellent models for epigenetic mechanisms of transposon regulation, heterochromatic silencing, and gene imprinting, important for both plant breeding and human health. We are investigating the role of RNA interference (RNAi) in heterochromatic silencing in the fission yeast *Schizosaccharomyces pombe* and the model plant *Arabidopsis thaliana*, as well as continuing our work on *Zea mays*. In fission yeast, we have found that RNAi promotes DNA replication and repair, as well as histone modification, in part by releasing RNA polymerase II in mitotic cells. In quiescence, RNAi becomes essential because it is required for release of RNA polymerase I from ribosomal DNA repeats. Genetic screens have revealed dependency on transcription, heterochromatin, and chromosome segregation, and provide insight into dicer mutations in cancer. We have found that long terminal repeat (LTR) retrotransposons in the mouse are controlled by transfer RNA (tRNA) fragments that match the highly conserved primer binding site (PBS), and plant retrotransposons are similarly targeted by a microRNA (miRNA) in pollen—a remarkable degree of conservation in transposon control. We continue to develop duckweeds for biofuel by sequencing the genome and developing an efficient transformation system in the clonally propagated aquatic macrophyte *Lemna minor*. This year we said goodbye to postdocs Jong Jin Han and Almudena Molla-Morales, who left for positions in patent law and biotechnology in Washington, D.C., and Madrid, respectively. Charlie Underwood graduated from the Watson School after completing a secondment to Cambridge University in the laboratory of Ian Henderson and moved on to a postdoc position at Keygene in the Netherlands. Dr. Daniel Grimanelli returned to IRD-Montpellier in France after a very productive and enjoyable sabbatical, and Jura Magalhaes returned to EMBRAPA in Brazil after a joint sabbatical with Doreen Ware. We welcomed Watson School student Ben Berube.

RNAi Is Essential for Cellular Quiescence

B. Roche, R.A. Martienssen [in collaboration with B. Arcangioli, Institut Pasteur, Paris]

Most cells in nature are not actively dividing, yet are able to return to the cell cycle given the appropriate environmental signals. There is now ample evidence that quiescent ("G_0") cells are not shut down but still metabolically and transcriptionally active. Quiescent cells must maintain a basal transcriptional capacity to maintain transcripts and proteins necessary for survival. This implies a tight control over RNA polymerases: RNA Pol II for mRNA transcription during G_0, but especially RNA Pol I and Pol III to

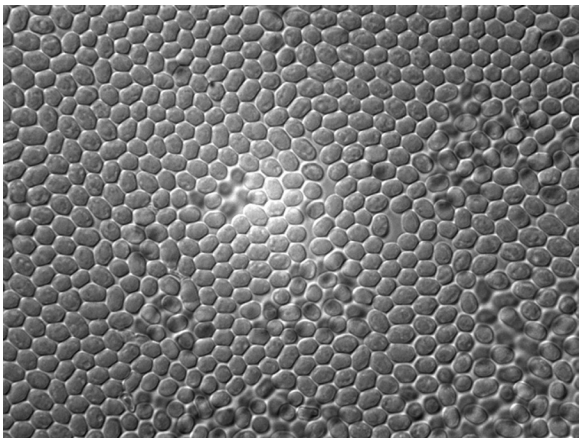


Figure 1. Single layer of quiescent (G_0) fission yeast cells, imaged by light microscopy.

maintain an appropriate level of structural RNAs, raising the possibility that specific transcriptional control mechanisms evolved in quiescent cells. We have found that RNAi is a major requirement for quiescence in the fission yeast *S. pombe*. RNAi mutants lose viability at G_0 entry and are unable to maintain long-term quiescence. We identified suppressors of G_0 defects in cells lacking Dicer (*dcr1*Δ), which mapped to genes involved in chromosome segregation, RNA polymerase-associated factors, and heterochromatin formation. We propose a model in which RNAi promotes the release of RNA polymerase in cycling and quiescent cells: (i) RNA polymerase II release mediates heterochromatin formation at centromeres, allowing proper chromosome segregation during mitotic growth and G_0 entry, and (ii) RNA polymerase I release prevents heterochromatin formation at ribosomal DNA during quiescence maintenance. Our model may account for the codependency of RNAi and histone H3 lysine 9 methylation throughout eukaryotic evolution, as well as the requirement for RNAi in mammalian stem cells. Quiescence is a key process in tumor progression and drug resistance, and our results may shed light on the predominance of dicer mutations in cancer.

The Conserved RNA-Binding Cyclophilin, Rct1, Regulates Small RNA Biogenesis and Splicing Independent of Heterochromatin Assembly

A.Y. Chang, S.E. Castel, E. Ernst, H.S. Kim, R.A. Martienssen

RNAi factors and their catalytic activities are essential for heterochromatin assembly in *S. pombe*. This has led to the idea that small interfering RNAs (siRNAs) can promote H3K9 methylation by recruiting the cryptic loci regulator complex (CLRC), also known as recombination in K complex (RIKC), to the nucleation site. The conserved RNA-binding protein Rct1 (AtCyp59/SIG-7) interacts with splicing factors and RNA polymerase II. Here, we show that Rct1 promotes processing of pericentromeric transcripts into siRNAs via the RNA recognition motif. Surprisingly, loss of siRNA in *rct1* mutants has no effect on H3K9 di- or trimethylation, resembling other splicing mutants—suggesting that posttranscriptional gene silencing per se is not required to maintain heterochromatin. Splicing of the *Argonaute* gene is also defective in *rct1* mutants and contributes to loss of silencing, but not to loss of siRNA. Our results suggest that Rct1 guides transcripts to the RNAi machinery by promoting splicing of elongating noncoding transcripts.

The *A. thaliana* Mobilome and Its Impact at the Species Level

C. LeBlanc, R.A. Martienssen [in collaboration with P. Fransz, T. Gerats, I. Schubert, K. Schneeburger, J.A. Jeddeloh, V. Colot]

Transposable elements (TEs) are powerful motors of genome evolution—yet a comprehensive assessment of recent transposition activity at the species level is lacking for most organisms. Using genome sequencing data for 211 *A. thaliana* accessions taken from across the globe, we found thousands of recent transposition events involving half of the 326 TE families annotated in this plant species. The composition and activity of the "mobilome" varies extensively among accessions in relation to climate and genetic factors. Moreover, TEs insert equally throughout the genome, but are rapidly purged by natural selection from gene-rich regions because they frequently affect gene expression. Remarkably, loci controlling adaptive responses to the environment are the

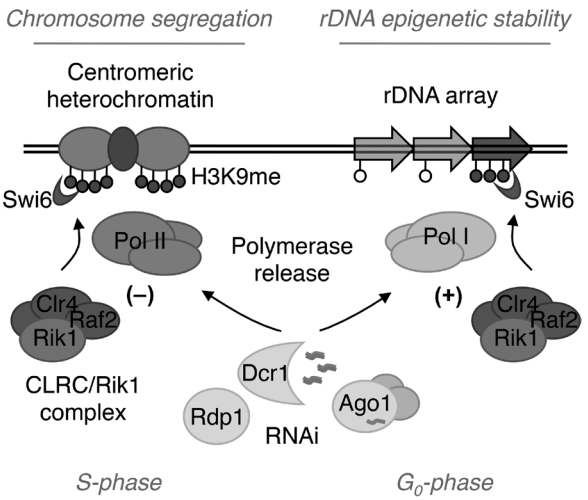


Figure 2. RNA interference (RNAi) is essential for quiescence. Shown here is a model for the dual function of RNAi in RNA polymerase release in cycling and quiescent cells. RNAi helps to remove RNA polymerase II from centromeric repeats in dividing cells, affecting chromosome segregation but not viability. In quiescent cells, RNAi is also required to remove RNA polymerase I from ribosomal DNA, uncovering an essential function.

most frequent transposition targets observed. These findings show the importance of transposition as a recurrent generator of large-effect alleles. Chromosomal inversions can provide windows onto the cytogenetic, molecular, evolutionary, and demographic histories of a species. A paracentric 1.17-Mb inversion on chromosome 4 was created by Vandal transposon activity, splitting an F-box and relocating a pericentric heterochromatin segment in juxtaposition with euchromatin without affecting the epigenetic landscape. Examination of the RegMap panel and the 1001 *Arabidopsis* genomes revealed more than 170 inversion accessions in Europe and North America, showing a robust association between the inversion and fecundity under drought. We also find linkage disequilibrium between the inverted region and the early flowering Col-FRIGIDA allele. Single-nucleotide polymorphism (SNP) analysis suggests the origin of the inversion to be southeastern Europe ~5000 years ago and that of the FRI-Col allele to be northwestern Europe, and reveals the spreading of a single haplotype to North America during the 17th–19th centuries. The "American haplotype" was identified in several European localities, potentially because of return migration.

Natural Variation and Dosage of the HEI10 Meiotic E3 Ligase Control *Arabidopsis* Crossover Recombination

C.J. Underwood, R. Martienssen [in collaboration with I. Henderson, Cambridge University, UK]

During meiosis, homologous chromosomes undergo crossover recombination, which creates genetic diversity and balances homolog segregation. Despite these critical functions, crossover frequency varies extensively within and between species. Although natural crossover recombination modifier loci have been detected in plants, causal genes have remained elusive. Using natural *A. thaliana* accessions, we identified two major recombination quantitative trait loci (*rQTL*) that explain 56.9% of crossover variation in Col × Ler F_2 populations. We mapped *rQTL1* to semidominant polymorphisms in *HEI10*, which encodes a conserved ubiquitin E3 ligase that regulates crossovers. Null *hei10* mutants are haploinsufficient and, using genome-wide mapping and immunocytology, we show that transformation of additional *HEI10* copies is sufficient to more than double euchromatic crossovers. However, heterochromatic centromeres remained recombination-suppressed. The strongest *HEI10*-mediated crossover increases occur in subtelomeric euchromatin, which is reminiscent of sex differences in *Arabidopsis* recombination. Our work reveals that HEI10 naturally limits *Arabidopsis* crossovers and has the potential to influence the response to selection. Remarkably, natural variation in the human ortholog of Hei10 also underlies variation in recombination frequency in human populations.

Loss of Histone H3K9 Methylation and Non-CG DNA Methylation Unleashes Heterochromatic Recombination in *Arabidopsis*

C.J. Underwood, R. Martienssen [in collaboration with I. Henderson, Cambridge University, UK]

Methylation of lysine-9 in histone H3 (H3K9) is a characteristic mark of constitutive heterochromatin that is highly conserved from unicellular eukaryotes to mammals and plants. Fission yeast, mouse, and *Arabidopsis* mutants in histone H3K9 methyltransferases show increased transcription of TEs and repetitive sequences due to increased RNA polymerase II accessibility. We sought to test how the histone H3K9 methylation pathway restricts access of the meiotic recombination machinery to heterochromatic sequences in the flowering plant, *A. thaliana*. Using a pollen fluorescent reporter system with markers flanking the pericentromeric heterochromatin on *Arabidopsis* chromosome 3 (CEN3), we observed increased recombination in histone H3K9 methyltransferase mutants. In mutants of the histone H3K9 methyltransferases, *kryptonite suvh5 suvh6* triple mutants, in which all H3K9 methylation is removed, we observed large increases in pericentromeric recombination. As the histone H3K9 methylation pathway in flowering plants is maintained through a positive feedback loop with the chromomethylases, we have also determined CEN3 recombination rate in the respective mutants and found consistent effects in CEN3 recombination. Mapping of double-strand breaks and crossovers genome-wide indicates transposons may be involved in the increased recombination rate. Our results could account for the curious observation that Ds transposons from maize are preferentially recovered in the *CHROMOMETHYLASE3* gene when launched from proximal locations in the *Arabidopsis* genome.

Live-Cell Analysis of DNA Methylation during Sexual Reproduction in *Arabidopsis* Reveals Context- and Sex-Specific Dynamics Controlled by Noncanonical RNA-Directed DNA Methylation

A.J. Schorn, R.A. Martienssen, D. Grimanelli [in collaboration with M. Ingouff, IRD-Montpellier; F. Berger, GMI, Vienna]

Cytosine methylation is a key epigenetic mark in many organisms, important for both transcriptional control and genome integrity. Although relatively stable during somatic growth, DNA methylation is reprogrammed genome-wide during mammalian reproduction. Reprogramming is essential for zygotic totipotency and to prevent transgenerational inheritance of epimutations. However, the extent of DNA methylation reprogramming in plants remains unclear. Here, we developed sensors reporting with single-cell resolution CG and non-CG methylation in *Arabidopsis*. Live imaging during reproduction revealed distinct and sex-specific dynamics for both contexts. We found that CHH methylation in the egg cell depends on DOMAINS REARRANGED METHYLASE 2 (DRM2) and RNA POLYMERASE V (Pol V), two main factors of RNA-directed DNA methylation (RdDM), but does not depend on Pol IV. Our sensors provide insight into global DNA methylation dynamics at the single-cell level with high temporal resolution and offer a powerful tool to track CG and non-CG methylation both during development and in response to environmental cues in all organisms with methylated DNA, as we illustrate in mouse embryonic stem cells.

LTR Retrotransposon Control by tRNA-Derived Small RNAs

A.J. Schorn, M.J. Gutbrod, C. LeBlanc, R. Martienssen

Transposon reactivation is an inherent danger in cells that lose epigenetic silencing during developmental reprogramming. In the mouse, LTR retrotransposons, or endogenous retroviruses (ERVs), account for most novel insertions and are expressed in the absence of histone H3 lysine 9 trimethylation in preimplantation stem cells. We found abundant 18-nt tRNA-derived small RNA (tRF) in these cells and ubiquitously expressed 22-nt tRFs that include the 3′ terminal CCA of mature tRNAs and target the tRNA PBS to be

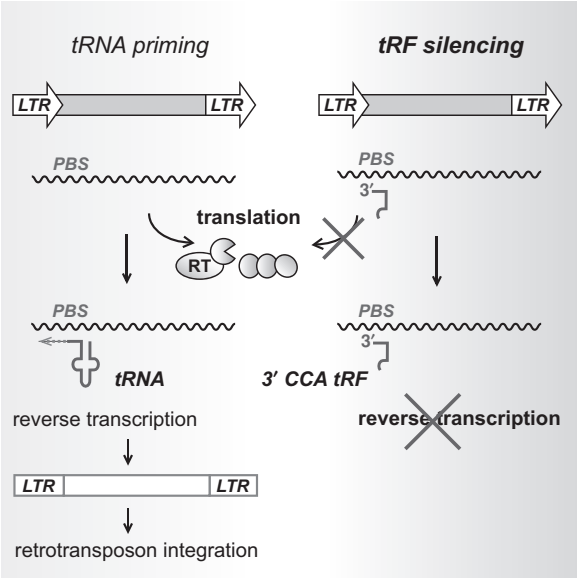


Figure 3.

essential for ERV reverse transcription. We found that the two most active ERV families, IAP and MusD/ETn, are major targets and strongly inhibited by tRFs in retrotransposition assays. Coding-competent ERVs are posttranscriptionally silenced by 22-nt tRFs, whereas 18-nt tRFs specifically interfere with reverse transcription and retrotransposon mobility (Fig. 3). The PBS offers a unique target to specifically inhibit LTR retrotransposons, and tRF targeting is a potentially highly conserved mechanism of small RNA-mediated transposon control.

microRNA-Directed Regulation of Transposon Small RNA Prevents Interploidy Hybridizations

F. de Sousa Borges, J.-S. Parent, F. Van Ex, R. Martienssen [in collaboration with C. Köhler, Uppsala University, Uppsala, Sweden]

TEs are epigenetically regulated to control their activity and preserve genome stability. In plants, miRNAs are involved in this process by triggering biogenesis of siRNA from target transcripts. We have found that the pollen-expressed miRNA845b targets retrotransposons and triggers the production of siRNAs that prevent interploidy hybridizations in *A. thaliana*. Strikingly, miR845b activity has a profound and dosage-dependent effect on the accumulation of most

21- and 22-nt TE-derived siRNAs in mature pollen, which seems to occur at the onset of meiosis and gametogenesis. We have further shown that natural variation at the *MIR845* locus results in miR845b depletion and reduced levels of 21- and 22-nt siRNA in the *Arabidopsis* accession Ler-0, which allows the formation of viable triploid seeds (a form of dosage compensation). Ectopic expression of a functional *MIR845b* transgene in Ler-0 largely restores TE siRNA biogenesis in pollen. This dose-dependent miRNA-directed boost of TE siRNAs in male meiocytes is able to regulate TE silencing and genomic imprinting in the seed, preventing spontaneous polyploidization events from spreading within populations.

High-Throughput Identification of Pseudouridine in Small RNA

R.P. Herridge, R. Martienssen

Pseudouridine is an isomer of uridine and is highly abundant in the RNA of all organisms. Until recently, pseudouridine was predominantly associated with structural RNAs such as rRNA, tRNA, and small nuclear RNA (snRNA). Using genome-wide next-generation sequencing techniques, pseudouridine has been identified in mRNA, small nucleolar RNA (snoRNA), and noncoding RNAs of yeast and human cells. One important role of pseudouridine is nuclear export of tRNAs via the export factor Los1p, a function that is impaired upon mutation of pseudouridine synthase 1 in yeast. *PAUSED* is the *Arabidopsis* homolog of Los1p and has an effect on the abundance and nuclear export of some miRNAs, suggesting that similar mechanisms might play a role in miRNA/siRNA export. We set out to develop techniques with which to analyze and confirm the presence of pseudouridine in small RNA in an effort to determine its prevalence and function in this class of RNA. CMCT is a chemical that specifically binds to pseudouridine and blocks reverse transcriptase (RT); CMCT treatment was used to block RT in the preparation of small RNA libraries, leading to a depletion of pseudouridine-containing reads (pseudouridine-si-Sequencing). This procedure was capable of detecting known pseudouridine-containing sites in rRNA, as well as tRNAs and snoRNAs. Novel small RNAs, including some miRNAs, were predicted to contain pseudouridine. Comparing pseudouridine-si-Seq data with other publicly available data sets, we were able to find ARGONAUTE proteins that preferentially bind pseudouridine-containing siRNA. We have also developed techniques to confirm pseudouridine presence in small RNA via northern blot and showed the ability of this method to discriminate a synthetic small RNA with and without pseudouridine.

Maternal Silencing of *EVADÉ* LTR Retrotransposon Is Mediated by AGO9-DDM1-RDR6-Dependent Epigenetic Pathways in *Arabidopsis*

S-C. Lee, J.S. Parent, E. Ernst, R. Martienssen

Retrotransposons are transcriptionally activated by stress conditions or through developmental epigenetic modifications. In *Arabidopsis*, 21-nt epigenetically activated small interfering RNAs (easiRNAs) are produced from retrotransposon transcripts to trigger transcriptional gene silencing or posttranscriptional gene silencing. In the SWI2/SNF2 chromatin remodeling mutant *ddm1*, overaccumulation of most retrotransposon transcripts occurs, but only results in retrotransposition for a handful of retrotransposons. To examine the possibility of translational control via 21-nt easiRNAs produced by RNA-dependent RNA polymerase 6 (RDR6), we obtained total and microsome-polysomal RNA-seq data from *ddm1* and *ddm1rdr6*. RNA transcripts of the majority of retrotransposons were associated with polysomes, indicating potentially active translation and very limited translational regulation by RDR6. After translation, LTR-retrotransposon products form virus-like particles in which reverse transcription makes complementary DNA, which is eventually integrated into plant genomes. We purified virus-like particle DNA (VLP-DNA) from *ddm1* and *ddm1rdr6* and identified LTR retrotransposons using next-generation sequencing. Among gypsy retrotransposons, *ATHILA* elements were more enriched in VLP-DNA of *ddm1rdr6* compared with *ddm1*, but VLP-DNA of *EVADÉ*, a Copia LTR retrotransposon known to escape many genome surveillance mechanisms, was the most abundant in both *ddm1* and *ddm1rdr6*. Our data showed that transcript levels and copy number of *EVADÉ* were significantly increased in *ddm1rdr6* as compared with *ddm1*. It has been reported that *EVADÉ* is maternally silenced, but the exact mechanisms are unclear. Our genetic analyses showed that AGO9 and DDM1 impact maternal silencing in

female gametophytes, whereas RDR6 has sporophytic effects on the female germ cell. DNA methylation levels of *EVADÉ* elements remained low in mutants defective in maternal silencing, whereas wild-type plants showed fully recovered DNA methylation in hybrid progeny. Interestingly, *DRM2* did not contribute to recovery of DNA methylation, which suggests that noncanonical RNA-directed DNA methylation pathways are involved in maternal silencing.

PUBLICATIONS

Fransz P, Linc G, Lee CR, Aflitos SA, Lasky JR, Toomajian C, Hoda A, Peters J, van Dam P, Ji X, et al. 2016. Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. *Plant J* **88:** 159–178.

Pichugina T, Sugawara T, Kaykov A, Schierding W, Masada K, Uewaki J, Grand RS, Allison JR, Martienssen RA, Nurse P, et al. 2016. A diffusion model for the coordination of DNA replication in *Schizosaccharomyces pombe*. *Sci Rep* **6:** 18757.

Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V. 2016. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5:** e15716.

Roche B, Arcangioli B, Martienssen RA. 2016. RNA interference is essential for cellular quiescence. *Science* **354:** aah5651.

In Press

Chang AY, Caste, SE, Ernst E, Kim HS, Martienssen RA. 2017. The conserved RNA binding cyclophilin, Rct1, regulates small RNA biogenesis and splicing independent of heterochromatin assembly. *Cell Rep* (in press).

Ingouff M, Selles B, Michaud C, Vu TM, Berger F, Schorn AJ, Autran D, Van Durme M, Nowack MK, Martienssen RA, et al. 2017. Live-cell analysis of DNA methylation during sexual reproduction in *Arabidopsis* reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev* (in press).

Roche B, Arcangioli B, Martienssen RA. 2017. Transcriptional reprogramming in cellular quiescence. *RNA Biol* (in press).

Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. 2017. LTR-retrotransposon control by tRNA-derived small RNAs. *Cell* (in press).

Underwood CJ, Henderson IR, Martienssen RA. 2017. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr Opin Plant Biol* (in press).

Ziolkowski PA, Underwood CJ, Lambing C, Martinez-Garcia M, Lawrence EJ, Ziolkowska L, Griffin C, Choi K, Franklin FC, Martienssen RA, Henderson IR 2017. Natural variation and dosage of the HEI10 meiotic E3 ligase control *Arabidopsis* crossover recombination. *Genes Dev* (in press).

MOLECULAR SIGNALING EVENTS UNDERLYING ENVIRONMENTAL CONTROL OF PLANT GROWTH

U. Pedmale L.N. Lindbäck K. Schwartz
K. O'Neill O. Spassibojko

A fundamental question in biology, which remains unanswered, is how the environment of the organism regulates its growth and development. Both plants and animals interact with their environment; however, plants grow postembryonically as they are incapable of moving around. Unlike animals, plants do not have specific organs that see or hear various stimuli; yet, plants are sensitive to their surrounding environment and modify their growth according to various external and internal signals. Plants regularly face variability in growth conditions—temperature, light quality and quantity, herbivores, pathogens, water availability, etc. Plants respond to these biotic and abiotic factors and survive substantial fluctuations in their environment. Plants also must balance the range of potential threats and benefits confronting them and should make appropriate decisions on resource allocation. Remarkably lacking a brain, plants can successfully integrate various cues and make appropriate decisions about growth. Such adaptability is essential to the sessile nature of the plants. In some adaptive responses, for example, when the plants have to cope with climate change and increased competition for light, there is a decrease in productivity (yield, biomass) as the plant reallocates resources to adapt better.

The goal of our laboratory is to determine the mechanisms behind how a plant perceives and successfully adapts to its environment. We also aim to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield. Our lab primarily studies the effect of light environment on plant growth and development. Light is among the most relevant environmental signals because it not only drives photosynthesis, but also provides critical information about the local growth environment, as well as seasonal time. Light is perceived by a complex array of photoreceptors, which include phytochromes (PHYA-E), cryptochromes (CRY1-2), phototropins (PHOT1-2), zietlupe family, and UVR8. Plants have developed various adaptive responses to interpret and use light directionality, quantity, and quality. In vegetational shading, when plants are under the shade of another plant, they perceive a decrease in the ratio of red to far-red light (R:FR) due to absorption of red light by chlorophyll and reflection of R:FR by the neighboring foliage. Simultaneously, there is also decrease in blue light and the available photosynthetically active radiation (PAR).

We focus on blue-light-absorbing CRYs; apart from being an excellent genetic and molecular tool to tease out the complexities of growth and adaptation, there are still many open questions about the molecular function of CRYs in plants. Understanding the role of CRYs is not only appealing for agriculture, but it also has an impact on human health, which could make this field appealing to diverse funding agencies. CRYs regulate growth and development and provide circadian entrainment to both plants and animals. In metazoans, disruption of CRY activity is linked to cancer, altered behavior, magnetoreception, and metabolism. Therefore, understanding CRY function in plants is important not only because of its crucial role in plant growth, but also because it may impact human health.

Mechanism of Light Control of Root Development

During shading (Fig. 1), many aerial organs elongate rapidly, whereas the root growth is reduced with the delay in the emergence of lateral roots. Roots not only serve as a mechanical anchor, but play a vital role in the well-being of the entire plant. Therefore, a robust and well-developed root system is required for healthy plant growth. As one can imagine, there is a negative

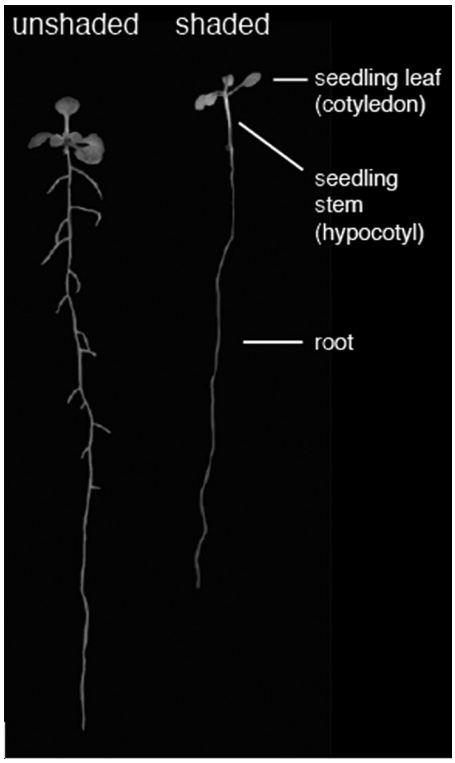


Figure 1. Light quality and quantity control of plant growth. *Arabidopsis* seedlings of the same age and genotype but grown in separate light conditions, shaded and unshaded, and have different body plans. Under shaded conditions, the seedling rapidly elongates its stem and grows taller to outcompete its neighbor to maximize light capture at the cost of root and leaf growth.

cycle occurring during shading; shoot-perceived shade leads to reduced root growth, which in turn is unable to support the shoot, leading to unproductive plants. However, this phenomenon is an excellent model to understand growth at a systems level because of the different growth phenotypes observed in the various organs of the same plant. Additionally, the ability to explore the nature of the interorgan and long-distance communication is used to signal when a distant organ is exposed to an adverse environment. Unfortunately, and surprisingly, not much is known about the mechanisms that underlie reduced root growth seen during shading.

During shading, newly synthesized auxin hormone in the cotyledons is required for the hypocotyl growth. First, we measured free auxin levels in the dissected root and shoot of seedlings exposed to low R:FR. As expected, we saw an increase in free auxin in the shoot and, interestingly, a sixfold accumulation in the root. However, we failed to see the activity of various auxin reporters in the root compared with the hypocotyl. High auxin levels are known to stimulate lateral root growth, but, intriguingly, we did not detect growth of lateral roots—but lateral root primordia were present. These results indicate that although auxin accumulation occurred, auxin signaling or perception was blocked in the root during shading. We also performed broad-time-course transcriptomic analysis from excised cotyledons, hypocotyl, and roots from *Arabidopsis* seedlings exposed to shade and nonshading conditions. Our transcriptomic analysis detected induction of auxin-responsive genes in the hypocotyl, but not in the root. Interestingly, genes induced during abiotic and biotic stress responses were significantly up-regulated in the root. To further test whether the roots in shaded plants activated stress responses to slow down their growth, we also used various reporters that are activated during stress or defense against pathogens. Therefore, our results suggest that stress responses in the root are likely inhibiting its growth. Furthermore, it is known that the plant can either grow or defend, and it is probable that in shaded plants, by turning on the defense or stress response in the roots, resources are diverted to the shoot organs, which then compete for light. However, the molecular determinant that switches a plant between these two states of defense/stress and growth is unknown and is of immense interest to biologists. We are taking several approaches to identify this molecular switch.

Molecular Determinants of CRY2 Protein Signaling and Stability

CRYs were first identified in plants and then discovered in animals. However, the signaling events from the photoactivated CRYs to growth and development programs are not known. In animals and plants, CRY protein levels and activity are tightly modulated to influence signaling outcome. Therefore, to elucidate the CRY-signaling pathway, our lab has purified CRY2-containing protein complexes from *Arabidopsis* seedlings exposed to low intensity of blue light, which is typically encountered under shading. We identified CRY2-associated proteins by tandem affinity purification and MudPIT. Interestingly, the orthologs of some of the CRY2-associated proteins were also present in CRY protein complexes purified

from human cells. This indicates that there could be a similar signaling mechanism in these two different evolutionary lineages.

Two of the highly enriched proteins in the CRY2-associated complex were two homologs of a deubiquitinase protein. We found out that CRY2 and these deubiquitinating proteins can interact directly in the nucleus of the cell. In seedlings treated with a deubiquitinase inhibitor, NEM, we did not detect CRY2 protein. We hypothesize that CRY2 is ubiquitinated continuously, but in certain situations, it recruits deubiquitinases to protect itself from proteasomal degradation such that downstream signaling can proceed. Next, we will test CRY2 levels in the deubiquitinase single and double mutants and plants where they are overexpressed. We are also in the process of identifying the cell type in which CRY2 interacts with these deubiquitinases, using cell-type-specific tools our lab has developed. Efforts are under way to identify ubiquitylated residues in the CRY2 protein and evaluate the effect of catalytically dead deubiquitinases in plants. Unlike animals, substrates for the large number (approximately 50) of plant deubiquitinases remain unidentified, except for histones. Therefore, we are presented with a unique opportunity to study CRY2 as one of the novel substrates of plant deubiquitinases and its role in plant growth and development. Conversely, to identify the E3 ubiquitin ligase responsible for targeting plant CRY2 for degradation, we are undertaking a forward genetic screen.

The Role of CRY2 in RNA Metabolism

We also identified many RNA-binding proteins that copurified with CRY2. Two recent papers also reported copurification of RNA-binding proteins with human CRY1/2. However, we do not know the significance of CRYs associating with RNA-binding proteins. Interestingly, the CRY2 nuclear speckles (Fig. 2) resemble speckles formed by pre-mRNA splicing factors, SR proteins, and other RNA-binding proteins in plants and animals, indicating that CRYs likely have a role in RNA metabolism. This observation may provide mechanistic insights into posttranscriptional control, known to be essential for animal and plant circadian biology, and control of alternative splicing in plants by light. We are currently focusing on two unknown proteins, which are hypothesized to bind to modified RNA—and their mutants resemble *cry2* mutant plants, indicating an epistatic relationship between them. Furthermore, similar to our experiments in plants, we are also studying their orthologs in animals and whether they can also interact directly with mammalian CRY2. RNA modifications are emerging as important regulators of various cellular processes not limited to protein translation, mRNA degradation, alternative splicing, and nuclear export. Also, we are collaborating with Dr. Chuan He at the Howard Hughes Medical Institute and the University of Chicago to map various RNA modifications in different genotypes at an organ- and tissue-specific level. We are generating loss-of-function mutants of these RNA-binding proteins to understand their impact on plant growth, circadian rhythms, alternative splicing, and other physiological responses.

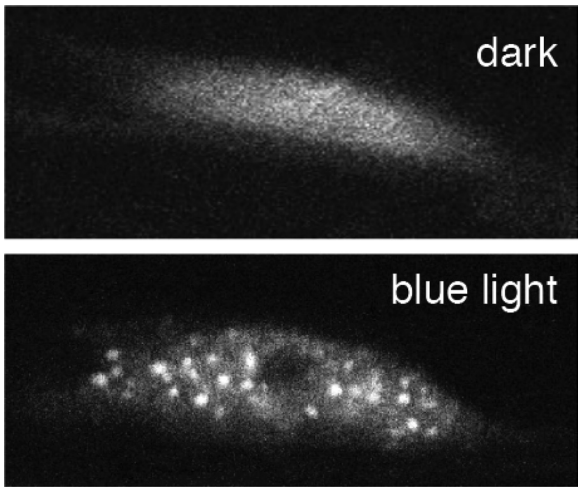


Figure 2. *Arabidopsis* CRY2 is a nuclear protein and forms nuclear speckles after exposure to blue light. The image is from *Arabidopsis* seedlings expressing mCitrine-CRY2.

PLANT DEVELOPMENTAL GENETICS

M. Timmermans A. Husbands K. Petsch
S. Knauer D. Skopelitis

Small RNA Gradients Create Stable Developmental Boundaries

D. Skopelitis

Specification of adaxial–abaxial (upper/lower) polarity in the developing leaf drives the flattened outgrowth of the leaf blade and directs the differentiation of distinct cell types within the leaf's adaxial/top and abaxial/bottom domains. Both are important innovations in the evolution of land plants that maximize photosynthesis while minimizing water loss to the environment. In addition to being a key developmental process, adaxial–abaxial patterning is of particular interest from a mechanistic point of view and has proven to be an excellent model to study small RNA-mediated gene regulation. We previously showed that the specification of adaxial–abaxial polarity involves a unique patterning mechanism in which the tasiR-ARF and miR166 small RNAs function as mobile positional signals. Movement of these small RNAs from their defined source of biogenesis in the top and bottom epidermis, respectively, creates opposing small RNA gradients across the leaf, which we showed are interpreted to yield sharply defined expression domains of their targets and polarity in developing primordia. This work provided the first direct evidence that small RNAs are mobile and function as instructive signals in development.

We are currently studying the mechanism through which mobile small RNAs pattern their targets. Mathematical modeling of the tasiR-ARF–ARF3 interaction predicts that small RNA gradients resulting from mobility are uniquely suited to generate sharply defined boundaries of target gene expression. To test this experimentally, we took advantage of the *pARF3:ARF3-GUS* reporter generated previously. In an *rdr6* mutant background, which blocks tasiR-ARF production, this reporter is expressed throughout the developing leaf. In this background, we express an artificial microRNA (miRNA) targeting *ARF3* (miRARF) from different leaf-specific promoters to modulate the position, direction, and steepness of this small RNA gradient. The outcomes of these experiments show that boundary formation is an inherent property of the small RNA gradient itself. The threshold-based readout of a small RNA gradient is highly sensitive to small RNA levels at the source, allowing plasticity in the positioning of a target gene expression boundary. In addition to generating sharp expression domains of their immediate targets, the readouts of opposing small RNA gradients were found to underlie formation of stable and uniformly positioned developmental boundaries. These patterning properties of small RNAs are reminiscent of those of morphogens in animal systems. However, their unique mode of action and the fully intrinsic nature of their gradients distinguish mobile small RNAs from classical morphogens and present a novel direct mechanism through which to relay positional information. Mobile small RNAs and their targets thus emerge as highly portable and evolutionarily tractable regulatory modules through which to create pattern in development and beyond.

Dissecting Small RNA Mobility in Plants

D. Skopelitis [in collaboration with K. Hill and S. Klesen, University of Tuebingen]

Given the scope of miRNA-regulated gene networks, the cell-to-cell movement of small RNAs has important implications with respect to their potential as instructive signals in development or response to physiological and stress stimuli. We are using artificial miRNAs targeting easy-to-score reporter genes to study parameters of miRNA movement such as tissue specificity, directionality, dose dependence, and the kinetics of movement. To investigate miRNA mobility in multiple distinct developmental contexts, we are expressing an artificial miRNA targeting a cell-autonomous green fluorescent protein (GFP) reporter (miR-GFP) from a number of promoters with distinct spatiotemporal

patterns of expression. Our findings support the idea that miRNAs traffic via plasmodesmata, small channels that connect most plant cells, and indicate that the number of cells across which a small RNA moves is determined in part by its abundance. Moreover, our findings indicate that the intercellular movement of small RNAs is indeed a developmentally regulated process. Depending on the tissue context, small RNAs can show directional movement or act strictly cell-autonomously. Interestingly, production of miR-GFP in young leaf primordia surrounding the shoot apical meristem (SAM) resulted in silencing of GFP inside the SAM, showing that miRNAs traffic from determinate organs into the shoot stem cell niche. In addition, miR-GFP was able to move from the vasculature below the SAM into the niche. These findings suggest that miRNAs may function as mobile signals between differentiated tissues and stem cells, possibly integrating environmental/physiological cues and plant development.

A key outcome of this work is that the developmental regulation of small RNA mobility follows rules that are distinct from those that govern protein trafficking. Using the insights gained from the above experiments, we are performing forward genetic screens to identify factors underlying regulated miRNA trafficking. The first screen has yielded several mutants that either increase or decrease the range of miRNA movement from the epidermis, which we are currently analyzing at the molecular level.

The START Domain Regulates HD-ZIPIII Activity and Organ Polarity

A. Husbands

Our previous observations on the expression and function of miR166 and tasiR-ARF reveal that adaxial–abaxial patterning involves a cascade of positional signals. The mobile signals that pattern the newly formed leaf are distinct from those that maintain polarity during subsequent development. This project aims to identify additional signals in adaxial–abaxial patterning. One such signal may involve the adaxial-promoting HD-ZIPIII transcription factors (TFs), which contain a putative lipid-binding START domain. Modeling of the START domain of the HD-ZIPIII member PHABULOSA (PHB) suggests structural similarity to human PC-TP, a phosphatidylcholine-binding START domain. This model, and sequence conservation with other START domains, was used to predict amino acids in PHB critical for START domain ligand binding. Analysis of *Arabidopsis* lines that express PHB-YFP fusion proteins with mutations in key residues lining the START ligand-binding pocket indicates that the START domain is required for proper HD-ZIPIII function. Such mutations block PHB-mediated target gene expression, indicating positive regulation of HD-ZIPIII activity by an unknown ligand.

Experiments to determine the mechanism through which the START domain controls HD-ZIPIII function show that subcellular localization is not impacted by ligand binding. Likewise, ligand binding does not impact the suite of PHB-interacting partners, as determined by quantitative mass spectrometry performed in collaboration with the CSHL proteomics facility. However, ligand binding does affect PHB's ability to bind DNA, possibly by blocking dimerization. To assess this, we established single-molecule pull-down (SiMPull) in plants, which uses fluorescence-based detection to interrogate protein–protein interactions and can quantitatively determine the heterogeneity and stoichiometry of protein complexes. Using SiMPull, we showed that mutating or deleting the START domain diminishes PHB's capacity to form dimers. Our data further indicate an additional, distinct role for the START domain in augmenting the transcriptional potency of PHB. To independently validate these findings, we fused the START domain to a heterologous TF, and found that the behavior of this START fusion protein neatly mirrored that of PHB. Taken together, we propose a model in which the presence of a START domain turns HD-ZIPIII proteins into potent TFs, but only in those cells in which they complex with their ligand. Currently, we are using a number of biochemical approaches to identify the endogenous ligand bound by HD-ZIPIII proteins. The results from immunoprecipitation-mass spectrometry experiments are very encouraging and identify an interesting ligand candidate.

Novel DICER-LIKE1 siRNAs Bypass the Requirement for DICER-LIKE4 in Development

K. Petsch [in collaboration with O. Tam and M. Hammell, CSHL; R. Meeley, DuPont-Pioneer; P. Manzotti and G. Consonni, University of Milan]

Dicer enzymes function at the core of RNA silencing to defend against exogenous RNA or as an endogenous

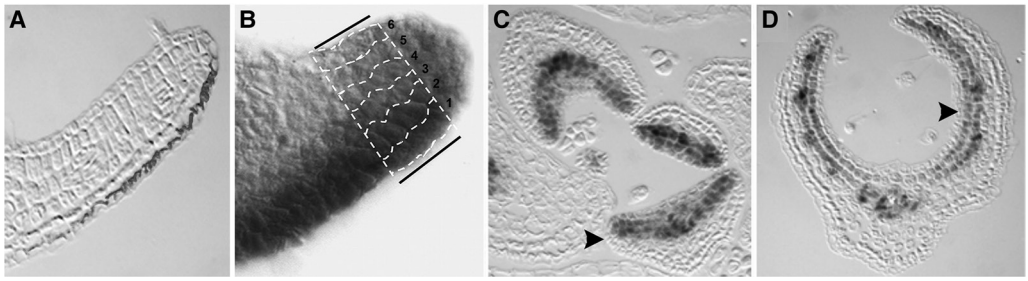


Figure 1. Gradients of mobile small RNAs have morphogen-like patterning activities. Mobility of miRNAs from their site of biogenesis in the bottom epidermis (*A*) yields a miRNA gradient (*B*) that through a threshold-based readout establishes an on–off pattern of target gene expression (*C*,*D*).

mechanism of gene regulation. Plant DICER-LIKE4 (DCL4) performs dual functions, acting in antiviral defense, as well as in development via the biogenesis of tasiR-ARFs. These small RNAs play an essential role in the grasses and act to spatially define the expression domain of AUXIN RESPONSE FACTOR3 (ARF3) TFs. However, contrary to tasiR-ARFs' essential function in development, DCL4 proteins show strong evidence of recurrent adaptation typical of host factors involved in antiviral immunity. Here, we address how DCL4 balances its role in development with pressures to diversify in response to viral attack. We show that, in contrast to other tasiR-ARF biogenesis mutants, *dcl4* null alleles condition an uncharacteristically mild phenotype, correlated with normal expression of select *arf3* targets. Loss of DCL4 activity yields a class of 22-nt tasiR-ARF variants associated with the processing of *arf3* transcripts into 22-nt secondary small interfering RNAs (siRNAs) by DCL1. Our findings uncover the presence of a novel DCL1-dependent siRNA pathway that bypasses the otherwise adverse developmental effects of DCL4 mutations (Fig. 1). This novel pathway is predicted to have important implications for DCL4's role in antiviral defense by reducing the selective constraints on DCL4 and allowing it to diversify in response to viral suppressors.

A High-Resolution Gene Expression Atlas for the Maize Shoot Apex

S. Knauer [in collaboration with M. Javelle and M. Scanlon, Cornell University; G. Muehlbauer, University of Minnesota; J. Yu, Kansas State University; P. Schnable, Iowa State University]

The SAM, a specialized stem cell niche at the growing shoot tip, integrates developmental and environmental signals to direct the initiation and patterning of new organs such as leaves. Its activity throughout the plant's lifetime is tightly controlled. To gain insight into gene regulatory networks behind stem cell maintenance and organogenesis, we generated a high-resolution gene expression atlas of 10 distinct domains and cell types within the vegetative maize shoot apex using laser microdissection and RNA deep sequencing. We found that ~10% of all transcribed genes are differentially expressed across these tissue types, including a valuable collection of cell type–specific genes. Interestingly, very few functional categories are enriched among the differentially expressed genes, which we show reflects prominent subfunctionalization within gene families. Strong enrichments were, however, seen for TF families, and principal component analysis identified unique TF signatures predictive of meristematic and vascular fate. Analysis of TF binding sites within promoter regions of stem cell–specific genes predicts a hierarchical network in which the combinatorial actions of diverse TF families underlie their spatially restrictive pattern of expression. Natural variation present at these TFs is associated with key plant architectural traits in genome-wide association studies (GWASs), providing functional support for these findings. Moreover, through cluster analysis we identified genes whose expression specifically marks the functional zones of the maize SAM: that is, the stem cell–harboring central zone, the organogenic peripheral zone, and the organizing center. Genes defining these zones are, in part, conserved between maize and *Arabidopsis*, but they also reveal remarkable differences and novel gene functions associated with these domains in maize that are validated by mutational analysis. In summary, our findings identify unique TF signatures as master regulators of

cell identity within the SAM that balance stem cell maintenance and organogenesis.

PUBLICATIONS

Husbands A, Aggarwal V, Ha T, Timmermans MCP. 2016. In planta Single-Molecule Pull-down (SiMPull) reveals tetrameric stoichiometry of HD-ZIPIII:LITTLE ZIPPER complexes. *Plant Cell* **28:** 1783–1794.

Kuhlemeier C, Timmermans MCP. 2016. The Sussex signal: Insights into leaf dorsiventrality. *Development* **143:** 3230–3237.

Plavskin Y, Nagashima A, Perroud P-F, Hasebe M, Quatrano RS, Atwal GS, Timmermans MCP. 2016. Ancient *trans*-acting siRNAs confer robustness and sensitivity onto the auxin response. *Dev Cell* **36:** 276–289.

GENOMICS

There has been a growing appreciation in recent years that gene function is frequently context-dependent, with a large part of that context provided by the activities of other genes. But trying to understand how genes interact to produce function is a hugely complicated problem and one that appears likely to become more so as genomic information becomes more detailed. **Jesse Gillis** and colleagues are computational biologists who are presently challenging an oft-taken approach to the problem in which the functions of genes are interpreted in the context of networks derived from gene association data. Such networks consist of millions of interactions across thousands of genes, derived from protein-binding assays, RNA coexpression analysis, and other sources. Historically, many attempts to understand gene function through networks have leveraged a biological principle known as "guilt by association." It suggests that genes with related functions tend to share properties (e.g., physical interactions). In the past decade, this approach has been scaled up for application to large gene networks, becoming a favored way to grapple with the complex interdependencies of gene functions in the face of floods of genomics and proteomics data. Gillis' work centers on identifying the limits of the approach and making fundamental improvements to its operation, as well as applying those improvements to neuropsychiatric gene network data.

Thomas Gingeras and colleagues study where and how functional information is stored in genomes. These efforts help explain the biological and clinical effects of disease-causing gene mutations in humans and other organisms. Gingeras is a leader of the ENCODE (ENCyclopedia of DNA Elements) and the mouseENCODE and modENCODE (model genome ENCODE) projects of the National Institutes of Health. His research has altered our understanding of the traditional boundaries of genes, revealing that almost the entire lengths of genomes in organisms ranging from bacteria to humans can be transcribed into RNA (pervasive transcription), and most RNA products made by a cell are not destined to be translated into proteins (noncoding, or ncRNAs). In fact, ncRNAs are proving to be involved in a variety of other important biological functions. Some have been shown to be critical components in the pre- and posttranscriptional and translational processes, as scaffolds on which large protein complexes are assembled and as extracellular signals. The initial studies that led to these observations have been extended to cover the entire human genome.

Gholson Lyon's lab focuses on analyzing human genetic variation and its role in severe neuropsychiatric disorders and rare diseases, including Tourette syndrome, attention-deficit hyperactivity disorder (ADHD), obsessive–compulsive disorder (OCD), intellectual disability, autism, and schizophrenia. By recruiting large groups of related individuals living in the same geographic location (e.g., Utah), Lyon's lab can study the breadth and depth of genetic variants in a similar environmental background. Using the exome—the parts of the genome that code for protein—and whole-genome sequencing, his lab looks for mutations that segregate with syndromes in the various populations. A second focus of the Lyon lab is to study the mechanistic basis of a new rare disease that they described in 2011. This is the first human disease involving a defect in the amino-terminal acetylation of proteins, a common modification of eukaryotic proteins performed by amino-terminal acetyltransferases (NATs). The team has been using several different cellular model systems to better understand the disease pathophysiology and the basic process of amino-terminal acetylation. This year, Lyon collaborated with a team of researchers from other universities and companies to use precision medicine to successfully treat a patient with severe OCD. His symptoms were treated with deep brain stimulation, and the team used whole-genome sequencing to try to understand the molecular basis of his disease. The patient experienced significant relief

from his symptoms and his quality of life returned, suggesting that similar methods may hold tremendous promise in the future.

The insights of **W. Richard McCombie** and colleagues have led to the introduction and optimization of novel methods of high-throughput genome sequencing. His team has made it possible to catalog variation among individual organisms in a way that would have been unthinkable 10 years ago. They have brought online a new generation of Illumina sequencers and optimized their function to a level at which 8–10 trillion DNA bases can be sequenced in a month. McCombie's team has been involved in international efforts culminating in genome sequences for maize, rice, and bread wheat—three of the world's most important food crops. They have also had an important role in projects to sequence the flowering plant *Arabidopsis thaliana* (the first plant genome sequence) and the fission yeast *Schizosaccharomyces pombe,* as well as the human genome and other important genomes. McCombie's group is currently involved in several important projects to resequence genes in patient samples that are of special interest to human health, including *DISC1* (a strong candidate gene for schizophrenia), looking for genetic variants implicated in bipolar illness and major recurrent depression. They are also looking for genes that contribute to cancer progression using whole-genome sequencing or a method called exome sequencing, which they developed with Greg Hannon to look at mutations in the regions of the genome that code for proteins.

Using multidisciplinary approaches that combine computational analysis, modeling, and prediction with experimental verification, **Doreen Ware**'s lab seeks a deeper understanding of the evolution of genome sequences in plants and their implications for agricultural improvement. By looking comparatively across the genomes of plants in the same lineage, they seek answers to the following questions: How are genes conserved and lost over time? What are the fates of duplicated genes? What is the impact of structural variation on phenotypic variation? Ware's team also studies gene regulation in plants, focusing on gene regulatory networks, targeting transcription factors and microRNA genes with the objective of understanding how these parts of the plant genome work together in determining spatial and temporal expression of genes. The lab had an important role in the project to produce a haplotype map reference genome of maize, spearheading the most comprehensive analysis of the crop yet. This has provided important information on the variation of the reference genome, as well as comparative data showing changes in the genome acquired through domestication and breeding. They have devoted special attention to examining diversity within maize, grape, and tomato, aiming to accelerate the development of strategies to introduce new germplasm that is needed to meet demands of increasing population and a changing environment. The lab also has brought fully sequenced genomes into an integrated data framework to enhance the power of their comparative studies. This past year, Ware was named as its principal investigator for the National Science Foundation–funded Gramene project, a comparative genomics resource for agriculturally important crops and models to support sustainable food and fuel production. Ware, as principal investigator for plants, has also helped lead an effort funded by the Department of Energy to create—out of many separate streams of biological information—a single, integrated cyber-"knowledgebase" for plants and microbial life.

GENE NETWORK ANALYSIS

J. Gillis S. Ballouz M. Crow

Research in the Gillis lab involves two interwoven elements: improving the interpretability of network analysis and characterizing transcriptional data in the brain. These topics form a naturally complementary unit because the complexity of the brain as a system means that it is essential that the methods for analyzing it yield clear and precise signals. A dominant interest within computational biology is the analysis of gene networks to provide insight into diverse levels of functional activity, typically starting with regulatory interactions and moving up to more diffuse associations important for understanding systemic dynamics. Gene associations (of various sorts) are believed to encode functional interaction, and this interaction is frequently shown to be able to substantially predict gene function. This approach, commonly called "guilt by association," is embedded in everything from prioritization of de novo variants to uncovering novel regulatory interactions or mechanisms of disease. Although black-box-style network analyses are common, explaining the basis of how and why methods work is more rarely attempted. In the Gillis lab, we are developing network-based methods and software that improve both the sophistication and breadth of data available for determining how genes interact to produce function, particularly focusing on how genes interact to cause disease or cell phenotypes. Broadly, our research can be divided into methods development and our own research applications, often performed in collaboration with other labs to test computational predictions experimentally. In addition to Jesse Gillis, the members of the Gillis lab are postdocs Sara Ballouz and Maggie Crow.

Functional Outliers in Rare Disease

Assessing gene expression from individual patients in rare disease cohorts is a statistical problem; we have little power to detect changes in a robust fashion. One typical way to overcome this $n = 1$ problem is to gain more power by looking for a joint signal among differentially expressed genes through their shared functions. These functional signals are usually molecular pathways to which the genes belong or gene ontology (GO) terms they share. An alternative to this approach is to look at the genes that should be frequently co-regulated but now display an unusual expression signature and, thus, potentially harbor a disease signal that is both disruptive and unique. In this project, we examine the possibility that the candidate disease genes display "outlier" or unexpected expression rather than reflecting a collective pattern of dysregulation.

In collaboration with the Lyon lab, we collected and sequenced one quad and five trios from the rare *TAF1* syndrome cohort. The probands from each family have a unique variant in their *TAF1* transcription factor and share many phenotypic attributes. As we are looking for unusual differential expression in a disease context, we were able to design a novel family-specific differential expression analysis that exploits expected overlaps and differences in the transcriptomic profiles of the parents relative to their affected child. Then, by tallying coexpression patterns from more than 3000 expression samples in 75 experiments, we generated a frequency of common expression value for all gene pairs across most of the genome. Genes with jointly common expression values were then filtered away, leaving us with a small number of what we called "outlier" genes, characteristic of each proband. We first note that filtering removed all GO functional enrichment. Importantly, we found a single outlier gene, the calcium channel subunit *CACNA1I*, that recurs in five of the six pedigrees. Notably, this gene is recurrently implicated in other neurological diseases such as schizophrenia and autism, making it a very plausible candidate. The sole family in which no signal was present was a copy number variant (CNV) carrier, whereas the other probands had different single nucleotide variants, implying a potentially different underlying molecular mechanism. Our outlier-based analysis revealed otherwise difficult to observe signals in the case of *TAF1* syndrome cohort, and we anticipate that future transcriptomic studies of rare disorders would benefit from this type of analysis.

Single-Cell Coexpression Networks

In a close collaboration with the Huang lab at CSHL, we have been exploring the basis of cellular identity, and pathways underlying it, using single-cell expression data. Single-cell RNA sequencing (RNA-seq) has emerged as a powerful technique to explore heterogeneity among cells, and it has now been applied to many human and mouse tissue types at multiple stages of development, including the lung, spleen, brain, retina, and lumbar dorsal root ganglia, among others. Because these studies are seeking to discover novel cell identities, they have tended to use unsupervised clustering of cells as defined by their expression profiles. Once novel groupings are determined, the researchers can assess them in more traditional ways (i.e., with differential expression), often with the intent of finding novel biomarkers for the subtypes identified. Although it is straightforward to test enrichment of existing pathways based on these gene sets, it is far less common to search for novel pathways in these newly defined cells. Coexpression analysis is particularly well-suited for this task, and in recent work we have developed a host of meta-analytic methods specifically targeted toward single-cell expression data. Single-cell RNA-seq is known to have specific and distinctive technical artifacts, and so uncovering novel biology within it requires approaches tailored both to avoid those confounds and to exploit its greater transcriptional specificity. Broadly, our methods are distinguished from others both by focusing on coexpression signatures, rather than individual genes, and also by exploiting preexisting knowledge as to the function of genes, a form of machine learning.

Learning Cell Identity

The exceptional diversity of neurons has been appreciated from the time of Ramón y Cajal. His depictions of cells that varied wildly in size, shape, and connectivity provided foundational guidance in the field of neuroscience. The precise extent of this diversity has been debated since that time, with each new technology indicating novel facets of neuronal identity: from morphology, to electrical activity, and now single-cell transcriptomics. When faced with a transformative technology, like single-cell RNA-seq, how are we to make sense of the varied and distinct results of papers attempting to reshape the landscape of molecular neuroscience? One approach is to build on a firm bedrock of our prior knowledge; comparison across studies to detect replicability is another approach. Our ongoing work with the Huang lab combines these two ideas in a naturally complementary way: By assessing clusters of cells for their overlap across studies with respect to the known functional properties of the genes that describe them, we characterize not just the overlap in transcriptional signal, but also its functional implications. These innovations, and our method of cross-laboratory study design, are critical means of defining the overall state of neuroscientific knowledge as to cell identity.

We first measured the replicability of neuronal identity by comparing more than 13,000 individual small cytoplasmic RNA-sequencing (scRNA-seq) transcriptomes and then assessing cross-data set evidence for novel pyramidal neuron and cortical interneuron subtypes identified by scRNA-seq. We found that 24/45 cortical interneuron subtypes and 10/48 pyramidal neuron subtypes have evidence of replication in at least one other study and provide lists of candidate marker genes. Across tasks we found that any large set of variably expressed genes can identify equivalent cell types across data sets with high accuracy, indicating many of the transcriptional changes defining cell identity are pervasive and easily detected.

Functional Convergence in Autism

Disagreements over genetic signatures associated with disease have been particularly prominent in the field of psychiatric genetics, creating a sharp divide between disease burdens attributed to common and rare variation, with study designs independently targeting each. Meta-analysis, including using underlying data, within each of these study designs is routine but no method for combining the results across study designs exists. Traditional meta-analytic approaches are not calibrated to detect functional convergence within this data, because significant disease gene associations from the different study designs arise in quite different ways with quite different null hypotheses. In this work, we develop a general solution that integrates the disparate genetic contributions constrained by their observed effect sizes to determine functional convergence in the underlying architecture of complex diseases, which we illustrate on autism spectrum disorder (ASD) data.

Our approach looks not only for similarities in the functional conclusions drawn from each study type individually but also those that are consistent with the known effect sizes across these studies. We name this the "functional effect size trend," and it can be understood as a generalization of a classic meta-analytic method, the funnel plot test. Because this allows us to combine many different studies, and to do so in a structured way, it has the potential to find overlapping functional signals that would otherwise be impossible to see. We took candidate disease gene data from multiple ASD studies across thousands of individuals (approximately 7000) and study designs, including whole-exome sequencing and genome-wide association studies. We split the candidate genes by variant class (common and rare) and effect size (low to high) into 14 gene sets, in the range of 20–100 associated genes per set, and performed functional analyses, controlling for set size. We detected remarkably significant trends in aggregate with 20 individually significant properties (FDR < 0.01), many in areas researchers have targeted based on different reasoning, such as fragile-X mental retardation protein (FMRP) interactor enrichment (false discovery rate [FDR] ~ 0.006). We are also able to detect novel technical effects and we see that network enrichment from protein–protein interaction data is heavily confounded with study design, arising readily in control data. We perform a number of downstream analyses, including tests for robustness and genome-wide rankings of disease association. Our focus is on ASD and gene-level associations, but the method can be extended to other diseases and other types of tests as data permits (e.g., phenotype screening, grouping of disorders, variants within studies).

Machine Learning Functional Properties of Genes

Increasingly, biologists have turned to computational methods to sift through the vast array of preexisting genomics data for validation that a candidate variant has a molecular role in the phenotype of interest or to prioritize a candidate as disease causal. These computational methods usually fit under the rubric of "machine learning," using network data that represent the interaction of genes or their products. Using our burgeoning genomics data to characterize gene function has been one of the central research objectives of computational biology in the postgenomics era. Despite this intense focus, the abundance of public data, and methods development in machine learning and "big data" analysis, progress has been surprisingly uncertain. Attaching novel functions to genes based on prior data remains difficult, and a large number of genes still have comparatively little information attached to them. This is a critically important problem in evaluating disease candidate variants, many of which involve otherwise unstudied genes. To better understand the forces driving algorithmic and data performance, developers have focused on comparative assessment. Despite some important progress enabled by these assessments, the underlying mechanisms driving methods performance are poorly understood. Each developer may use separate data resources, different algorithms, and quite specific means for parsing the data in combination, making it quite hard to explore factors affecting performance. Field-wide progress is hampered by our inability to know why a specific algorithm worked on a particular data set for a potentially narrow task.

To overcome this lack of detail in assessment, we developed a computational infrastructure that allows us to implement representative samples of cutting-edge machine learning algorithms. These algorithms can then be customized to specific biological problems and, when benchmarked against prior data, obtain very high levels of performance. Moreover, we learn more by combining them through meta-analysis than we can from their individual results, as is often the case in machine learning. Because we now have in-house versions of multiple algorithms, we can explore variance and data dependencies in a way not typically possible, leading to several useful heuristics for meta-analysis we have recently published.

Epigenetic Meta-Analysis

Chromatin accessibility provides an important window into the regulation of gene expression. Recently, the assay of transposase accessible chromatin with sequencing (ATAC-seq) was developed to profile genome-wide chromatin accessibility. Although pipelines for implementation have been developed on an ad hoc basis for the analysis of individual data sets, there has been little comparative or aggregate evaluation.

This is critical both to determine appropriate methodologies, controls, and efficacy, as well as to determine the global biological landscape of chromatin accessibility across diverse conditions. One major technical problem to address is that the counts of ATAC-seq reads underlying each peak vary substantially within a single sample and also between samples. Such variation makes the comparison to determine presence and absence of peaks (i.e., the open and closed state of chromatin) more difficult and less statistically well grounded. In this work, we analyzed 197 ATAC-seq data samples from 13 studies to evaluate the robustness of results, as well as their specificity across studies. We find that peaks are promiscuously identified, with approximately 34,000 peaks per sample on average. These can be substantially overlapped with transcription start sites (TSSs), with 11,000 genes on average overlapping with the called peaks. Among those genes, 447 genes have peaks at their TSSs from all 197 samples we analyzed. We evaluate the properties of these genes in detail, including mean expression across a diverse corpus of data, coexpression between this set and other genes, and functional enrichment. Finally, we suggest a novel approach to evaluate the robustness of peak signals and sensitivity by bootstrapping reads and recalling peaks for each resampling, whose calls are then aggregated. This yields peak calls that are highly robust to variation in noise as a source of peaks within data itself. We applied this novel data to our own ATAC-seq conditional experiment with a reduction of likely spurious peaks to improve specificity.

Heuristics for the Interpretation of Gene Sets

One branch of our research tackles an almost philosophical problem of broad impact in bioinformatics: How can we evaluate methods in which no gold standard exists? This is a particular problem when the bioinformatics methods are targeted to functional interpretation, in which not only is a gold standard challenging, but knowing whether method performance will generalize is a major impediment. The most striking example of this problem occurs in function enrichment methods, which are ubiquitous, diverse, and very hard to assess.

We propose a solution to this problem by looking to assess not whether enrichment results are correct, but whether they meet basic criteria for coherence. In particular, we show that assessing algorithm outputs for uniqueness and robustness is highly useful across a variety of problems. A substantial part of the work involves formalizing a test for uniqueness or robustness that can be applied to any enrichment method, but ultimately the principles map to surprisingly intuitive biological properties in the form of an assessment of the role of multifunctional genes within study results.

Gene set analysis to translate gene lists into enriched functions is among the most common bioinformatic methods, yet few would advocate taking observed *p*-values at face value. Not only is there no agreement on the algorithms themselves, there is no agreement on how to benchmark them. We evaluate the robustness and uniqueness of enrichment results. We show that both properties turn out to be closely linked to multifunctionality, defined as the number of annotated functions a gene has. We show that multifunctional genes are more likely to appear in genomics study results and drive the generation of biologically nonspecific enrichment results because of their overlaps in annotations. Additionally, individual multifunctional genes easily generate apparent, yet highly fragile, significance. By providing a means of benchmarking when enrichment analyses report nonspecific and nonrobust findings, we are able to assess where we can be confident in their use. We find that gene set enrichment methods that correct for biases are far likelier to return robust and unique results. We provide a software implementation for multifunctionality assessment in ErmineJ, but our general approach can be adapted to any enrichment method.

PUBLICATIONS

Ballouz S, Gillis J. 2016. AuPairWise: A method to estimate RNA-seq replicability through co-expression. *PLoS Comput Biol* **12:** e1004868.

Crow M, Paul A, Ballouz S, Huang ZJ, Gillis J. 2016. Exploiting single-cell expression to characterize co-expression replicability. *Genome Biol* **17:** 101.

O'Meara MJ, Ballouz S, Shoichet BK, Gillis J. 2016. Ligand similarity complements sequence, physical interaction, and co-expression for gene function prediction. *PLoS ONE* **11:** e0160098.

Verleyen W, Ballouz S, Gillis J. 2016. Positive and negative forms of replicability in gene network analysis. *Bioinformatics* **32:** 1065–1073.

In Press

Ballouz S, Pavlidis P, Gillis J. 2017. Using predictive specificity to determine when gene set analysis is biologically meaningful. *Nucleic Acids Res* (in press).

Ballouz S, Weber M, Pavlidis P, Gillis J. 2017. EGAD: Ultra-fast functional analysis of gene networks. *Bioinformatics* **33:** 612-614.

GENOME ORGANIZATION AND REGULATION AND FUNCTIONAL ROLES OF NONCODING RNAs

T.R. Gingeras P. Batut, S. Chakrabortty, C. Danyko, C. Davis, A. Dobin, J. Drenkow, G. Nechooshtan, S. Patel, A. Prakash, A. Scavelli, L.-H. See, D. Yunusov, C. Zaleski

Human-ENCODE Project Transition to EN-TEx Project

C. Danyko, C. Davis, A. Dobin, J. Drenkow, A. Scavelli, L.-H. See, C. Zaleski

Phase 3 of ENCODE (3) was completed in mid-year and all data generated for the transcriptome portion of the ENCODE3 project are available at https://www.encodeproject.org/search/?type=experiment&lab.title=Thomas%20Gingeras,%20CSHL. An extension to the ENCODE3 project was initiated at this time entitled the ENCODE-GTEX (EN-TEx) subproject. The aim of this project is the construction of personal genomes from four deceased individuals (two males and two females) whose families have donated their bodies for the construction of these personalized genomes. Approximately 25 tissues from different organs have been donated from each individual. Approximately 20 assays have been used to analyze transcriptional, genetic, and epigenetic states for each tissue sample from each individual. These data types started with an accurate representation of the donor's phased diploid genome complete with single-nucleotide and structural variations that is used to provide detailed information concerning the gene annotations of their individual transcriptomes and regulatory elements. In collaboration with Dr. Michael Schatz's (Johns Hopkins University), Dr. Bing Ren's (University of California at San Diego), and Dr. Roderic Guigo's (CRG, Barcelona) laboratories, the genome sequences of each of the four donors have recently been assembled into phased diploid genomes using ~60× coverage of Illumina paired-end short reads, ~35× coverage of 10× Genomics-linked read technology, and 100× coverage of long-range Hi-C data. For one genome, 55× coverage of PacBio sequencing was generated. These multiple technological approaches permit additional information concerning the detection of the structural variations present in each individual. Additionally, we have generated approximately 500 long (>200 nt), short (<200 nt), and RAMPAGE profiles and ChIP-seq data sets for the four individuals.

The results of these analyses will address a number of important questions relevant to the scientific community concerning the value and construction of personalized genomes.

1. *What are the most effective biotechnologies for establishing personalized genomes?*
   This analysis will compare and contrast the relative merits of Illumina, 10× Genomics, Hi-C, and PacBio sequencing for discovering variants and phasing them into complete personalized genomes. Special attention will be placed on developing the most cost-effective strategies by combining different technologies appropriately so that our approach can be applied to large numbers of samples in the future. Related work to be performed will focus on finalizing computational methodologies using phased diploid genomes as the framework for personalized allele-specific analysis by developing optimized read mapping and quantification pipelines. Similarly, the comparison of the GTEx and ENTEx protocols for RNA sequencing will serve to evaluate the gain of using stranded versus nonstranded approaches.
2. *What do we learn from a personalized genome instead of the reference?*
   This analysis will focus on studying gene expression and regulation empowered by the personalized genomes established in Aim 1. Here, we will be able to address important questions such as how many novel or disrupted genes and transcripts are there in a typical person and how many novel or

disrupted regulatory regions are there. This is crucially important for estimating how reliable it is to transfer functional annotation from one person to another. One further investigation is to determine the impact of noncoding mutations, both single-nucleotide and structural variations, in candidate regulatory regions on gene expression. We will answer this question by comparing the gene expression of the ENTEx samples to normalized expression values from larger collections of samples such as in GTEx and identifying variations in the putative regulatory regions for these genes.

3. *Can we use the genomic variants as natural perturbations of the ENCODE (Encyclopedia of DNA Elements) encyclopedia?*
   This analysis will measure the impact of genomic variants identified in Aim 1 on the Phase 3 ENCODE encyclopedia. For example, this will allow us to address whether variants overlapping annotated regulatory elements affect binding and expression. The availability of the phased genomes will allow us for the first time to investigate whether there is linkage between the perturbations and the perturbed elements (i.e., whether or in which circumstances the effect of a mutation in a regulatory element affects regulated elements in the same chromosome or in the two chromosomes). Although not every annotated feature will have a variant in one of these individuals, this will, nevertheless, allow us to consider the role of a large number of these features, much larger than would be practical to study using CRISPR-Cas9 or other directed gene-editing techniques. This aim will also allow us to test whether previously identified elements not interrupted by variants are confirmed in primary tissues.

Database Management and Software Development

C. Davis, A. Dobin, C. Zaleski

Database Management: With the end of phase 3 of the ENCODE project, data submissions of all accumulated data have been completed and cleanup of missing metadata or incomplete/retired submissions has been accomplished. This included the continued execution of the ENCODE sequencing production pipeline on the final ENCODE 3 data sets. RNA sequence mapping, multiple quality control steps, and file transfer to the ENCODE Data Coordination Center (DCC) have also been completed. Migration and upgrade of our laboratory sequencing and metadata database system (TDM) to a more robust server have been completed. This included bug fixes, feature updates, and better network routing. We are now prepared to take on new and larger data sets that are and will be generated for other still ongoing projects (see below).

Participation in the recently initiated MaizeCODE and Northwell projects has required the integration of these data sets into our in-house data management schema. Very early tasks have included creation of a data model to store sample/sequencing metadata using JOSN Schema. Setup and administration of the MaizeCODE and Northwell wikis have been completed and now serve as a primary mode of communication for all members of these projects. In collaboration with the CSHL information technology group, a specific storage "share" utilizing Gingeras lab SONAS space to be accessible among all MaizeCODE and Northwell participants has been created.

Software Development: Personal genomics is envisaged to become an essential component of precision medicine, holding the promise for identifying genetic predispositions for common and complex diseases, diagnosis of hereditary disorders, individual treatment of cancer, and genotype-guided drug research. Millions of personal genomes will be sequenced in the next few years; however, the tools are lacking for personalized processing of the functional data types such as RNA-seq and ChIP-seq, which at present are routinely mapped to the haploid reference genome. Here, we present an extension of popular RNA-seq aligner STAR, the STAR-Diploid software that was developed to map RNA-seq and ChIP-seq reads to the fully phased diploid personal genomes.

In the first step, STAR-Diploid uses the personal variants, including single-nucleotide variants, short indels, and large structural variants, to build the personal diploid genome sequence from the reference assembly. The reference annotations are arithmetically lifted over to each of the haplotypes. Next, the reads are mapped to both haplotypes simultaneously to produce diploid genomic alignments. Mapping to the personal diploid sequence virtually eliminates the reference bias that plagues the alignment to the haploid reference genome. The diploid alignments are then converted to the reference coordinates while preserving the haplotype information. The final output of

the pipeline consists of haplotype-specific alignments and signal (wiggle) tracks in the reference coordinates, which can be visualized in the standard genomic browsers, as well as allele-specific counting of reads per gene. Furthermore, STAR-Diploid converts diploid genomic alignment into diploid transcriptomic alignments that are input into RSEM for allele-specific quantification of transcripts and gene expression.

To show the effectiveness of the STAR-Diploid algorithm, we have used it to process a large collection of long and small RNA-seq, RAMPAGE, and ChIP-seq data from the ENCODE-GTEx (EN-TEx) collaboration (approximately 20 tissues for four donors). The personal diploid genomes were constructed with the variants phased by means of 10× Genomics Chromium sequencing and Hi-C short reads for chromosome-span phasing, with RNA-seq data used to supplement and resolve phasing conflicts. We have compared allelic imbalance across the multiple tissues obtained from the same donor, as well as across multiple individuals for the same tissue, for the purpose of understanding the genotypic and cell-type contributions to allele-specific expression (ASE). Using long-range variant phasing information from the 10× Genomics and Hi-C data, we have identified potential causative mutations in regulatory regions responsible for the observed ASE.

Conserved Noncoding Transcription and Core Promoter Regulatory Code in Early *Drosophila* Development

P. Batut

Recently, we completed a high-throughput transcriptional start site (TSS) mapping in tightly resolved time series to establish genome-wide promoter activity profiles throughout embryonic development in five *Drosophila* species spanning 25–50 million years (MY) of evolution. Combining TSS identification at single-nucleotide resolution with quantitative measurements of developmental expression patterns, we uncovered unique features of expression timing and core promoter structure to generate novel insights into transcriptional regulation. We observed that distinct types of core promoters are selectively active in three broad phases of embryonic development: specific combinations of core motifs mediate transcription during early, intermediate, and late embryogenesis. Each individual class of core promoters is functionally associated with distinct sets of transcription factors. As a result, a two-tier model of transcriptional control is proposed in which core promoters and enhancers mediate, respectively, coarse-grained and fine-grained developmental regulation. It also has been shown that noncoding transcription is far more widespread than anticipated in *Drosophila*, with 3973 promoters driving the expression of long non-coding RNAs (lncRNAs) during embryogenesis. Through the analysis of the fine structure and sequence conservation of their core promoters, we show that evolutionarily conserved lncRNAs are under strong purifying selection at the levels of primary sequence and expression specificity. As an exemplar we functionally characterize the schnurri-like RNA (SLR) locus, which expresses a hitherto uncharacterized lncRNA. This lncRNA is conserved in a spatiotemporal pattern suggestive of a role in early dorsoventral patterning.

In summary, these results have uncovered different classes of core promoters and distinct functional roles for these core promoters in regulating transcription by defining windows of opportunity for activation by enhancer sequences. These studies also reveal a vastly underappreciated aspect of developmental transcriptomes by showing that noncoding transcription is extremely prevalent, tightly regulated in a manner similar to coding genes, and, crucially, deeply conserved.

Extracellular Vesicles (EVs) Are Centers for the Sorting and Processing of Various Types of RNAs

S. Chakrabortty, G. Nechooshtan, S. Patel, A. Prakash

We have previously observed (Chakrabortty et al., *RNA 21:* 1 [2015]) that the K562 chronic myelogenous leukemia cell line produces EVs that induce a cell death phenotype in primary cells of different types. This is illustrated using BJ foreskin fibroblasts as recipient nontumor cells (Fig. 1). A comprehensive characterization of the nature and diversity of RNAs packaged into EVs derived from nine different primary and transformed cell types has been performed, revealing a reproducible, nonrandom, and cell-type-specific sorting of the different types of RNAs into EVs by the cell origin. This sorting of specific RNAs into EVs is a dynamic process that varies based on the physiological conditions of the source cells. As seen

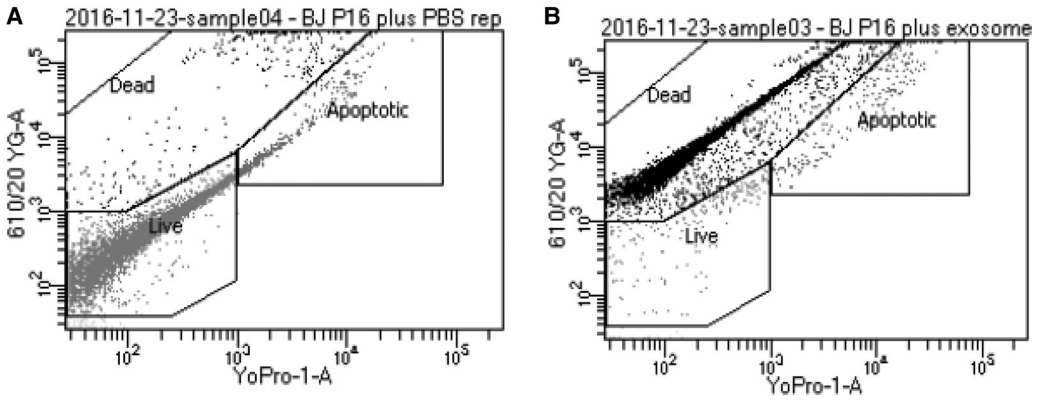


Figure 1. Cell-sorted profiles of BJ primary cells before and after treatment with EVs from K562 cancer cells. (*A*) 10^5 BJ cells suspended in same solution as used in EV-treated cells. (*B*) 10^5 BJ cells treated with EVs derived from 1.35×10^8 K562 cancer cells. BJ cells are stained with Yo-Pro (apoptosis) and propidium iodide (postapoptosis dead cells). See Chakrabortty et al., *RNA 21:* 1 [2015] for results of treatment of BJ cells with EVs from other primary cells and empty lipofection vessels.

previously, EV RNAs are capable of eliciting a transcriptional response by the recipient cells, as well as a corresponding striking cell death phenotype, specifically on primary but not transformed cells. We have shown that a 29–31-nt RNA fragment of the RNY5 RNA is processed in the EVs. Synthetic copies of this fragment are capable of recapitulating the cell death phenotype induced by the EVs from transformed cells. Several questions surround these observations concerning the processing agent that forms the 31–29-nt RNA fragments in the EVs: whether there are other RNAs found in the EVs that are the result of processing of longer precursors and whether this in vitro cell death phenotype of primary cells is observed in vivo.

RNY5 Processing Agent

Formation of the 31–29-nt processed species of the full-length RNY5 (83 nt) can be recapitulated in vitro using EV protein extracts, but not cellular protein extracts. During 2016, work has focused on trying to determine the agent(s) responsible for the processing of RNY5 transcript. In collaboration with Jonathan Ipsaro from the Joshua-Tor lab, we applied chromatographic separation to EV protein extract. Although proteomic analysis of fractions active in RNY5 RNA processing is ongoing, we also noticed that processing of RNY5 by EV protein extract is sensitive to RNase inhibitor. Because RNase inhibitor is known to be specific to pancreatic RNase family members, we are currently generating a CRISPR-Cas9 knockout of RNase A1, the most abundant and widely expressed family member. In addition, because different classes of nucleases leave different ends on processed fragments (with/without phosphate on 5′ or 3′ ends), we are interrogating the 5′ and 3′ ends of fragments processed in vitro and in vivo to corroborate the involvement of this class of nucleases. Pancreatic RNases are secreted enzymes, whereas RNase inhibitor is an intracellular protein. We hypothesized that this might explain why RNY5 is processed in the extracellular milieu and not in the cell. Supporting this line of research, we found that when RNase inhibitor is inactivated by N-ethylmaleimide, cellular protein extracts are able to process RNY5 RNA, as seen when using EV extracts.

Other Processed EV RNAs

Recently, efforts in the laboratory have focused on transfer RNA (tRNA) fragments that are found within EVs to determine whether these RNA biotypes are also subject to selective processing within EVs. As background for this question, our latest results indicate that a selective enrichment of processed fragments whose parent tRNA is transcribed from specific tRNA genes are differentially distributed between a cell of origin and its vesicles. This enrichment appears to be independent of the anticodon marked by the particular gene and, instead, is based on sequences present within the 5′ and 3′ halves of the molecules. Further studies of the relative distributions of 3′ halves, 5′ halves, and full-length tRNA within cells and vesicles

indicate that for some tRNA transcripts, the 5′ halves of that tRNA are exclusively detected in vesicles and not observed in their parent cells. These observations suggest the possibility that these 5′ fragments were processed from full-length tRNA precursor molecules within EVs for export outside of the parent cell.

A full-scale analysis of tRNA fragments was undertaken to test this hypothesis. Our results suggest that there is a significant enrichment specifically of 5′ halves of three tRNA genes—tRNA23550, tRNA23608, and tRNA7861. These results indicate that specific tRNA genes are enriched in EVs of cells with a great degree of reproducibility. Additionally, when 5′-, 3′-halves and full-length tRNA transcripts were individually compared with the EVs and source cells, we observed that the full-length transcripts of the tRNA are enriched within cells, whereas for the same genes their 5′ tRNA halves are enriched in EVs. However, for tRNAs that are more abundant in cells and not seen to be enriched in EVs, we see that they predominantly exist as full-length transcripts. This finding leads to the hypothesis that full-length mature tRNA transcripts of certain tRNA genes are specifically sorted into EVs, in which they may undergo processing, which leads to the formation of the 5′ tRNA halves and a systematic depletion/degradation of the corresponding 3′ tRNA halves. Further, in vitro processing assays performed with EV extracts have suggested that when synthetic full-length mature tRNA23550 is exposed to EV extract, the resulting fragments of tRNA match the length of the 5′ tRNA half. These findings lead us to hypothesize that for the tRNA transcripts sorted into EVs, there might be a specific processing step that occurs within EVs that produces an enrichment of the 5′ tRNA halves of these genes. Other interesting molecules, which were seen to be enriched within EVs of several cell lines, were microRNAs: miRNA-17-92a, miRNA-93, and miRNA-103b.

PUBLICATIONS

Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis C, Gingeras TR, Guigó R. 2016. Gene-specific patterns of expression across organs and species. *Genome Biol* **17:** 151.

Djebali S, Rodríguez Martin B, Palumbo E, Pervouchine DD, Breschi A, Davis C, Dobin A, Alonso G, Rastrojo A, Aguado B, et al. 2016. Recurrent chimeric transcripts in human and mouse. *J Anim Sci* **94** (suppl 5): 200–201.

Dobin A, Gingeras TR. 2016. Optimizing RNA-seq mapping with STAR. In *Data mining techniques for the life sciences*, 2nd ed (ed. Carujo O, Eisenhaber F), pp. 245–262. Humana, New York.

In Press

Breschi A, Gingeras TR, Guigó R. 2017. Comparative transcriptomics in human and mouse. *Nat Rev Genet* **18:** 425–440.

Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Davis C, Gingeras TR, Frankish A, Harrow J, Guigó R, Johnson R. 2017. High-throughput annotation of full-length long noncoding RNAs with Capture Long-Read Sequencing (CLS). *bioRxiv* 105064.

HUMAN GENETICS AND GENOMIC MEDICINE

G. Lyon J. Crain R. Kleyner A. Solowinska
M. Doerfel J. Malcolmson D. Tegay
L. Kimoto J. O'Rawe Y. Wu

The Amino-Terminal Acetylation of Proteins

M. Doerfel, Y. Wu, J. Crain [in collaboration with S. Lyons, CSHL]

More than 85% of human proteins are acetylated at their amino-terminal amino group. Hence, amino-terminal acetylation (NTA) is one of the most abundant modifications of eukaryotic proteins. Despite its discovery more than 30 years ago, very little is known about the cellular effects/functions of this modification. In humans, six distinct amino-terminal amino-acetyltransferases (NATs) catalyze the transfer of an acetyl group from acetyl-CoA to the amino-terminal amino group of their specific target proteins. The major human acetyltransferase, NatA, consists of an auxiliary subunit, Naa15, and a catalytic active subunit, Naa10. We have previously described two families with a lethal X-linked disorder of infancy called Ogden syndrome. This disorder comprises a distinct combination of an aged appearance, craniofacial anomalies, hypotonia, global developmental delays, cryptorchidism, and cardiac arrhythmias. Using X chromosome exon sequencing, we identified a c.109T > C (p.Ser37Pro) variant in Naa10 as contributing to this disease. Biochemical analysis and immunoprecipitation assays in combination with liquid chromatography–mass spectrometry (LC-MS) showed a reduced catalytic capacity and revealed an impaired binding of the S37P mutant toward specific interaction partners, including Naa15 and Naa50. Analysis of the amino-terminal acetylome of patient cells revealed a decreased acetylation of a subset of NatA substrates, indicating that a reduced binding capability and an affected enzymatic activity of the Naa10 S37P mutation is a prominent feature in Ogden syndrome. We are studying NTA of proteins in yeast, mice, and patient-derived human cell lines, as it is very likely that the complexity of NTA increased substantially during evolution of higher organisms. The long-term goal is to gain an understanding of this understudied protein modification in yeast (*Saccharomyces cerevisiae*), mice, and humans. We have already conducted extensive analyses of the pathway in yeast, primarily as an easier system in which to understand some of the known phenotypes, including the mating defect in yeast. Characterization of NAA10/NAA15 knockout yeast strains revealed various phenotypes, including growth defects at elevated temperatures and altered sensitivity toward cytotoxic stresses. These effects could be rescued by overexpressing human wild-type (wt) Naa15/Naa10 from plasmids; however, overexpressing mutant Naa15/Naa10 S37P only partially rescues these effects. Interestingly, introduction of both human Naa15/Naa10 wt and S37P mutant into the endogenous locus of the corresponding yeast genes failed to reverse the effects. We also continued our efforts toward establishing induced pluripotent stem cells (iPSCs) from skin fibroblasts from one of the boys with Ogden syndrome.

The role of NTA, as performed by any of the NATs, is poorly understood on a whole-organism level, particularly in mammals. Therefore, we have been working with several mouse strains, either lacking Naa10 completely or having severe knockdown of Naa10 expression, which display increased early postnatal lethality and various abnormalities, including congenital heart defects, peripheral nervous system (PNS) defects, a C7 cervical vertebrosternal rib, piebaldism, and genital and renal abnormalities. These mice also develop hydrocephaly during embryogenesis, which in some cases gets worse in the first few weeks of life and becomes associated with abnormal gait and neurologic abnormalities, ultimately leading to death. In addition, the majority of the homozygous female mice are infertile, along with some infertility even in the heterozygous female mice. These data show a relatively specific set of defects that emerge in late embryogenesis and/or early neonatal life, thus arguing that Naa10 is important

for normal development—but also hinting that there must be some unknown biological redundancy in this pathway, at least in mice, given that these mice do not all die during embryogenesis. We have been using much of the new equipment at Woodbury, including the CT scanner and ultrasound machine, in active collaboration with Scott Lyons.

Outlier Gene Expression Reveals Recurrent Dysregulation in Rare Disease Pedigrees

S. Ballouz, M. Dörfel, J. Crain, M. Crow, J. Gillis

We have been generating and analyzing RNA sequencing from blood-derived RNA from one quad and five trios from the rare TAF1 syndrome cohort. We have resubmitted a manuscript in which we describe a new "outlier-based analysis" that reveals otherwise difficult-to-observe signals, including an outlier gene involving the calcium channel subunit CACNA1I, which recurs in five of the six pedigrees. The sole family in which no signal was present was a copy number variant (CNV) carrier, whereas the other probands had different single nucleotide variants, implying a potentially different underlying molecular mechanism. Notably, this gene is recurrently implicated in other neurological diseases, such as schizophrenia and autism—making it a very plausible candidate for further functional studies.

Whole-Genome Sequencing of One Complex Pedigree Illustrates Challenges of Genomic Medicine

H. Fang, Y. Wu, H. Yang, M. Yoon, K. Wang, and others

Human phenotype ontology (HPO) has risen as a useful tool for precision medicine by providing a standardized vocabulary of phenotypic abnormalities to describe presentations of human pathologies; however, there have been relatively few reports combining whole-genome sequencing (WGS) and HPO, especially in the context of structural variants. We illustrate an integrative analysis of WGS and HPO using an extended pedigree, which involves Prader–Willi syndrome (PWS), hereditary hemochromatosis (HH), and dysautonomia-like symptoms. A comprehensive WGS pipeline was used to ensure reliable detection of genomic variants. Beyond variant filtering, we pursued phenotypic prioritization of candidate genes using Phenolyzer. Regarding PWS, WGS confirmed a 5.5-Mb de novo deletion of the parental allele at 15q11.2 to 15q13.1. Phenolyzer successfully returned the diagnosis of PWS and pinpointed clinically relevant genes in the deletion. Further, Phenolyzer revealed how each of the genes is linked with the phenotypes represented by HPO terms. For HH, WGS identified a known disease variant (p.C282Y) in HFE of an affected female. Analysis of HPO terms alone fails to provide a correct diagnosis, but Phenolyzer successfully revealed the phenotype–genotype relationship using a disease-centric approach. Finally, Phenolyzer also revealed the complexity behind dysautonomia-like symptoms, and seven variants that might be associated with the phenotypes were identified by manual filtering based on a dominant inheritance model. The integration of WGS and HPO can inform comprehensive molecular diagnosis for patients, eliminate false positives, and reveal novel insights into undiagnosed diseases. Because of extreme heterogeneity and insufficient knowledge of human diseases, it is also important that phenotypic and genomic data are standardized and shared simultaneously.

KBG Syndrome Involving a Single Base Insertion in *ANKRD11*

J. Malcolmson, R. Kleyner, D. Tegay [in collaboration with K. Ward, J. Coppinger, A. Maughan, G. Maughan, L. Nelson, Utah; K. Wang, California; R. Robison, Utah]

KBG syndrome is a rare autosomal dominant genetic condition characterized by neurological involvement, macrodontia, and distinct facial, hand, and skeletal features. More than 100 cases have been reported; however, it is likely that KBG syndrome is underdiagnosed because of lack of comprehensive characterization of the heterogeneous phenotypic features. We described the clinical manifestations in a male referred at 11 years of age, who showed symptoms, including epilepsy, developmental delay, distinct facial features, and hand anomalies, without positive genetic diagnosis. Subsequent exome sequencing identified a novel de novo heterozygous single base pair insertion (c.6015dupA) in *ANKRD11*, which was validated by Sanger sequencing. We predicted that this insertion leads to a premature stop codon and loss of function in *ANKRD11,* thereby implicating it as contributing to the proband's symptoms and yielding a molecular diagnosis of KBG syndrome for the case.

SCN8A Mutation in Child Presenting with Seizures and Developmental Delays

J. Malcolmson, R. Kleyner, D. Tegay [in collaboration with W. Adams, K. Ward, J. Coppinger, L. Nelson, Utah; K. Wang, California; R. Robison, Utah]

The *SCN8A* gene encodes the Nav1.6 neuronal voltage-gated sodium channel α subunit. Mutations in this gene have been associated with early infantile epileptic encephalopathy type 13. With the use of whole-exome sequencing, a missense mutation was identified in a 4-year-old female who initially showed symptoms at the age of 5 months, after she received routine vaccinations. Determining the molecular etiology of this proband's epileptic encephalopathy improved her management.

Expanding Collection and Sequencing of Other Rare Genetic Syndromes

Y. Wu [in collaboration with H. Fang, CSHL; R. Robison, Utah; K. Wang, California; A. Rope, Oregon, and others]

We continue to meet and collect many families in Utah and elsewhere with very rare, idiopathic genetic syndromes. The total number of DNA samples collected to date is approaching 2000, and this includes detailed phenotyping information. We have been making extensive use of HPO terms, and the primary investigator (PI) was an author on a review concerning the current progress with the development and integration of HPO in various research settings. We also participated in writing a review concerning the indel-calling algorithm, Scalpel. The PI also participated in the long-read sequencing of a Chinese genome, along with reporting a novel mutation in a case of Ehlers–Danlos syndrome.

Collaborating on Genetics of Tourette Syndrome

This work was done In collaboration with The Tourette Syndrome Association International Consortium for Genetics.

The PI continues to collaborate on this international effort to understand the genetics of Tourette syndrome (TS). Psychiatric comorbidity is common in TS; when present, these conditions typically cause more distress and impairment than do tics. High rates of attention-deficit/hyperactivity disorder (ADHD) and obsessive-compulsive disorder (OCD) are well documented and thought to be core components of the TS phenotype; however, few studies have fully characterized other comorbidities. We, therefore, continue to characterize the prevalence and impact of psychiatric comorbidity in a large sample of individuals with TS and their family members.

PUBLICATIONS

Darrow S, Hirschtritt ME, Davis LK, Illmann C, Osiecki L, Grados M, Sandor P, Dion Y, King R, Pauls D, et al., and the Tourette Syndrome Association International Consortium for Genetics. 2016. Identification of two heritable cross-disorder endophenotypes for Tourette syndrome, *Am J Psychiatr* **174:** 387–396.

Desai A, Connolly JJ, March M, Hou C, Chiavacci R, Kim C, Lyon GJ, Hadley D, Hakonarson H. 2016. Systematic data-querying of large pediatric biorepository identifies novel Ehlers–Danlos syndrome variant. *BMC Musculoskelet Disord* **17:** 80.

Dörfel MJ, Fang H, Crain J, Klingener M, Weiser J, Lyon GJ. 2016. Proteomic and genomic characterization of a yeast model for Ogden syndrome. *Yeast* **34:** 19–37.

Fang H, Bergmann EA, Arora K, Vacic V, Zody MC, Iossifov I, O'Rawe JA, Wu Y, Jimenez Barron LT, Rosenbaum J, et al. 2016. Indel variant analysis of short-read sequencing data with Scalpel. *Nat Protoc* **11:** 2529–2548.

Fang H, Wu Y, Jiménez-Barrón LT, O'Rawe JA, Highnam G, Mittelman G, Lyon GJ. 2016. Whole genome sequencing of one complex pedigree illustrates challenges with genomic medicine. *BMC Med Genomics* **10:** 10.

Hirschtritt ME, Lee PC, Pauls DL, Dion Y, Grados MA, Illmann C, King RA, Sandor P, McMahon WM, Lyon GJ, et al.; for the Tourette Syndrome Association International Consortium for Genetics. 2016. Social disinhibition is a heritable subphenotype of tics in Tourette syndrome. *Neurology* **87:** 497–504.

Kleyner R, Malcolmson J, Tegay D, Ward K, Maughan G, Nelson L, Wang K, Robison R, Lyon GJ. 2016. KBG syndrome involving a single-nucleotide duplication in *ANKRD11*. *Cold Spring Harb Mol Case Stud* **2:** a001131.

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. 2016. The Human Phenotype Ontology in 2017. *Nucleic Acids Res* **45:** D865–D876.

Malcolmson J, Kleyner R, Tegay D, Adams W, Ward K, Coppinger J, Nelson L, Meisler MH, Wang K, Robison R, et al. 2016. SCN8A mutation in a child presenting with seizures and developmental delays. *Cold Spring Harb Mol Case Stud* **2:** a001073.

Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, Fu A, Li Q, Li N, Gong S, et al. 2016. Long-read sequencing and de novo assembly of a Chinese genome. *Nat Commun* **7:** 12065.

DEVELOPMENT OF NEXT-GENERATION SEQUENCING TECHNOLOGY

W.R. McCombie E. Antoniou, Y. Berstein, E. Ghiban, S. Goodwin, S. Iyer, M. Kramer, J. Lihm, S. Mavruk Eskipehlivan, O. Mendivil Ramos, S. Muller, R. Wappel

In 2016, we continued to progress on several fronts. First among these was our work in the optimization and utilization of long-read-sequencing technology. We acquired the second generation of the Pacific Biosciences instrument in early 2016 and began optimizing the instrument. In 2016, we sequenced a maize genome with it, which we will continue assembling (in collaboration with Mike Schatz) into 2017. We also continued optimizing the Oxford Nanopore sequencing platform in 2016, seeing substantial improvements in yield, accuracy, and DNA amount required during the year. Last, we have continued the analysis of several of our ongoing projects in psychiatric disease and cancer genetics. Several of the projects in psychiatric disease genetics were published in 2016.

Whole-Genome Sequence Analysis of the Original DISC1 Pedigree

J. Lihm, M. Kramer, W.R. McCombie [in collaboration with D. Blackwood, D. Porteous, P. Thompson, N. Ryan, and I. Deary, University of Edinburgh; A. McRae, University of Queensland]

As part of our collaboration with the Porteous Lab in University of Edinburgh, we continued to analyze whole-genome sequencing (WGS) data of 47 individuals selected from the original DISC1 pedigree in order to detect novel risk regions and/or variants that affect psychiatric diseases. The 47 whole genomes consist of 24 affected and 23 unaffected family members. The affected group covers a spectrum of psychiatric diseases, including severe diagnoses such as schizophrenia (SCZ) and bipolar disorder (BD) to relatively moderate diagnoses such as major depressive disorder (MDD), anxiety, and conduct disorder.

We performed variant-level analysis for more detailed description of specific risk variants. Categorizing variants by functional types, we found that highly shared variants in pedigree-wide affected members are enriched in noncoding variants, whereas coding variants are shared by more closely related affected members—indicating a role for noncoding variants. To study the impact of coding variants, we annotated coding variants with ANNOVAR for functional category and deleteriousness using 13 tools. We selected variants with higher impact such as loss-of-function variants (stop gain and splice sites) and missense variants that are predicted deleterious by eight or more tools, totaling 150 genes. We also looked for the common properties shared by these 150 genes. Using permutation tests controlling for gene size and GC content, we calculated the significance of overlap with the publicly available database, GeneBook, which is a database of genes implicated in psychiatric diseases, including de novo and genome-wide association studies (GWASs). Our list of genes with functional exonic variants showed significant overlap with autism de novo variants and also postsynaptic density genes. In particular, genes detected from severely affected samples produced strong signals from these gene sets.

To search for larger structural variants with higher effect size, we built an in-house pipeline using a read depth approach and segmentation method developed by Chris Yoon to call copy number variations (CNVs). We called deletion regions genome-wide and selected deletion regions that are shared by affected samples but not by unaffected samples. We detected an intronic deletion in the *TENM4* gene on chromosome 11, 11 Mbp away from the translocation breakpoint that is shared by all translocation carriers.

In 2017, we are going to extend the single-nucleotide polymorphism (SNP) analysis to noncoding variants and build an in-house pipeline for prioritizing noncoding variants under pedigree structure. In addition to SNPs, we will explore the profiles of small indels that are called by Scalpel. CNV analysis will be further refined and extended to genome-wide scale. A manuscript detailing the results from the linkage analysis will be finalized in 2017.

The Epigenome of Fear

J. Lihm, S. McCarthy, W.R. McCombie [in collaboration with S. Ahrens, B. Li, and J. Gillis, CSHL]

The results of our previous collaboration showed that open chromatin regions called at genome-wide scale with ATAC-sequencing data contain many false-positive regions that make comparisons between samples difficult. In a continuing collaboration with the Li lab and the Gillis lab in which we do much of the computational analysis, we characterized the features of peaks from meta-analysis of ATAC-seq data and further strengthened our analysis pipeline using a read downsampling approach.

To characterize the distribution of peaks generated from ATAC-seq, we downloaded raw ATAC-seq reads from 13 studies that were published in early 2016. We processed 197 samples' bam files by a uniform computational pipeline. We found that peaks are located in TSS regions of 10,000 genes on average, nearly half the genes in the RefSeq database. Such a high number of genes detected impedes the statistical evaluation between samples and within samples due to the high false-positive rates. We hypothesized that a larger number of reads would help reduce the noise and increase the true signals. However, samples with a larger number of reads had a larger number of peaks, implying that higher coverage increases the noise level as much as the signals from true peaks. Next, we checked how many peaks were affected by a slight change in coverage and were possibly false-positive. Results from sampling 90% of reads, which is equivalent to losing only 10% of reads, showed that 10% of the peaks on average from unsampled original data are not called in the 90% sampled data. This implies that some peaks are sensitive to the coverage change and do not have robust signals.

Motivated by these results, we developed a novel approach using sampling of reads. We only sampled 10% of the original reads and called peaks from these to obtain peaks with strong signals. We applied this approach to the novel mouse data from amygdala and cortex generated by the Li lab. We looked for genes with differential chromatin accessibility between amygdala and cortex, fear-conditioned mouse and control mouse, and ErbB4 null mutant and wild-type mouse. Hypergeometric enrichment analyses were performed to find genes that were significantly present as accessible in one condition but not in the other. Twenty-two genes were found to be more accessible in cortex samples but less in amygdala samples (adjusted p-value < 0.05). We are currently finalizing the manuscript for this work.

Sequencing Parent–Sibling Quads with SCZ

O. Mendivil Ramos, S. McCarthy, M. Kramer, W.R. McCombie [in collaboration with D. Weinberger and R. Straub, Leiber Institute]

Continuing our collaboration with the Weinberger group, we have expanded the analysis of the exomes of 25 quads—two unaffected parents, one proband (affected with SCZ), and one unaffected sibling—to elucidate the role of de novo and transmitted mutations in the etiology of SCZ. The Weinberger group had selected this set of quads because they have the largest differences in IQs among affected and nonaffected siblings.

Last year we reported on our de novo SNV analyses. To expand our work in 2016, we used Scalpel (Narzisis et al., *Nat Methods 11:* 1033 [2014]) to call small indels in the exome data. Including the de novo indels did not alter the main findings. Enrichment analyses using DnENRICH still indicate an overrepresentation of functional de novo variants in the affected probands in gene sets associated with the activity-regulated cytoskeleton-associated protein complex (ARC complex) and with autism spectrum disorder (ASD). And NEUROCARTA shows suggestive association of genes harboring de novo mutations in probands with neuropsychiatric diseases. Highly deleterious de novo mutations in affected individuals include SHANK3 and other genes such as *MCM2*, which is involved in the development of the brain region that receives sound information, and *TENM2*, which is involved in the development of connectivity within the nervous system, as potential candidates.

We then expanded our pipeline to detect transmitted variants. Because of the large number of transmitted variants, we prioritized variants that are rare or novel, loss-of-function, and that have RVIS scores in the top 25% of the most intolerant genes. In the probands, we noted variants in *MYO16*, involved in the regulation of neuronal morphogenesis, and *GRIK5*, a gene that belongs to the glutamate-gated ionic channel family that has been associated with SCZ.

Furthermore, we were able to detect compound heterozygous mutations in both probands and siblings. In the probands, we noted ANK3, which previously had

been associated with intellectual disability and BD, and DLG2, which regulates synaptic stability at cholinergic synapses. We noted that the proband, which harbors the de novo mutation in SHANK3, also harbors a compound heterozygous mutation in the gene *C1orf94*, which has been associated with ASD and epilepsy and is expressed in testes and brain. In the siblings overall we noted fewer compound heterozygous variants—and the genes affected by those mutations are associated with muscular dystrophy rather than brain-related processes.

In 2017, we will extend this work to incorporate CNV and will aggregate this rare/novel variation with common variation and clinical data in an attempt to decipher a pattern of molecular pathogenesis.

Rare Variant Detection in Complex Disorders Using the Birthday Model

Y. Berstein, S. McCarthy, M. Kramer, W.R. McCombie

In an effort to identify rare causative variants in complex disorders such as SCZ and BD, we have improved the algorithm based on the birthday model (McKinney, *Am Math Mon 73:* 385 [1966]). The core of the algorithm is a general analysis of coincidences based on a popular probabilistic problem: the birthday problem. We have considered the probability of samples sharing a variant as analogous to the chance of individuals sharing the same birthday. To derive an optimal threshold of coincidence, we have incorporated the multiple testing algorithm by Westfall and Young (1993) to our core algorithm. The algorithm can be applied at both variant and gene resolutions. In addition to implementing this probabilistic method to published data on ASD, hypertriglyceridemia, and SCZ, we had fully implemented the algorithm on our case control study on BD (Goes et al. 2016), which detected rare variants with minor allele frequency of 1% or less in 1000 genomes, the Exome Variant Server (EVS), and Exome Aggregation Consortium (ExAC). The top results based on our method were Sanger validated. Several genes in the top results are associated with psychiatric disorders in published studies such as SCZ, autism, and BD. We have tested the robustness of the algorithm by evaluating the stability of the ranking when less data are available, meaning a smaller sample size than the original. It showed good performance even with 70% of the original data. Our model can help to estimate the significance of the finding, especially for studies that are intended to identify rare mutations. The core probability based on the birthday model is very sensitive to variants of relatively low recurrence, which generally do not provide enough signal in existing statistical tests. Given the insufficient sample size in current studies on complex disorders, our algorithm complements existing methods, specifically for the detection of the missing rare variants. Our approach provides a quantitative metric for evaluating whether rare findings, such as rare variants, merit additional investigation such as further biological experimental validation.

The Genetic Basis of Ethnic Disparity in Colon Cancer

S. Goodwin, M. Kramer, R. Wappel, S. Muller, W.R. McCombie [in collaboration with X. Wang, X. Yu, E. Li, and J. Williams, Stony Brook University]

Our previously reported colon cancer methylation work with colleagues at Stony Brook University led to a P20 grant to explore the increased rate and poorer prognosis of colon cancer in African–Americans (AAs) versus Caucasian–Americans (CAs). A major aim of this project is to establish standardized pipelines with best practices for both laboratory procedures and computational analysis of the genomic, transcriptomic, and epigenetic sequencing data from AA and CA patients.

To study coding variants, eight matched tumor/normal samples were captured with NimbleGen v3 exome probes and sequenced on an Illumina HiSeq2500. Reads were mapped to hg19 and then processed with VarScan2 (Koboldt et al., *Genome Res 22:* 568 [2012]) using the somatic variant calling mode. To increase confidence in the somatic variant calls, we required a minimum read depth of 10×, alternate allele frequency of at least 10%, and variant presence on both strands. Loss-of-function variants (stopgain, frameshift, or splice site) were detected in previously reported cancer drivers such as APC, TP53, ARID1A, and CASP8. However, only four out of 16 samples achieved coverage depth of ≥30-fold over at least 80% of the target. This coverage level is suboptimal for detection of variants at lower allelic fractions. The biased representation of the sample was

likely caused by the limited sample quantity that necessitated use of whole-genome amplification (WGA) to increase DNA for capture. A QIAGEN extraction protocol was then performed, which led to increased DNA yield and quality for 10 AA and six CA tumor/normal pairs. The new captures will be sequenced and analyzed in 2017.

To assess expression differences, 24 RNA-seq samples (including 10 tumor-normal pairs) were sequenced and trimmed with Trimmomatic (Bolger et al. 2013). Mapping with STAR aligner (Dobin et al. 2012) showed variability in the amount of uniquely mapped reads per sample, and a subset of samples showed persistent ribosomal contamination. Protocol modification at the ribo-erase step and stricter cutoff of RNA integrity scores may address these issues.

To survey methylation patterns, reduced representation bisulfite sequencing (RRBS) was performed on 15 samples. The reads were mapped to the bisulfite sequencing (BS) converted genome using BS-Seeker (Pellegrini et al. 2013). The MspI targeting was found to be deficient because of limited input DNA for the digestion. Additional material has been extracted, and the MspI digestion time will be increased to improve performance. Analysis of all components is ongoing, with emphasis on detection of novel cancer drivers in the underrepresented AA samples.

PacBio Sequencing and Analysis Optimization

S. Goodwin, M. Kramer, E. Antoniou, S. Mavruk Eskipehlivan, R. Wappel, W.R. McCombie

As the interest in PacBio sequencing increases, optimizations and new methods are needed to keep up with the dynamic needs of researchers. One such advance is the acquisition of the PacBio Sequel in February 2016. The Sequel instrument is the next generation of long-read technology available from Pacific Biosciences. Initially, the instrument performed at less than its potential, necessitating optimization. Rather than generating >5 Gb per SMRTcell, the instrument was yielding ~2 Gb per SMRTcell. This was accompanied by much shorter read lengths than what is provided in the RS II, ~3000 bp, which are not long enough to support our efforts in de novo genome assembly.

We began working closely with PacBio to optimize new methods aimed at increasing yield and read length. Two primary areas of improvement were identified: the carryover of free primer from the annealing stage and inefficient distribution of the DNA library across the ZMW. The first issue, primer carryover, was addressed by the incorporation of a column clean-up step following primer annealing. This reduced the amount of free primer being introduced to the SMRTcell, thus preventing the blockage of ZMWs by free primer. The second issue, inefficient loading of library, was addressed by the inclusion of PEG into the buffers used to deliver the DNA to the SMRTcell. The PEG allows the aqueous buffer to move more freely across the ZMW surface, thus improving DNA loading. This inclusion initially caused significant overloading of the SMRTcell, until it was discovered that significantly less DNA is required to load the SMRTcell in the presence of PEG. At this time, ~10× less DNA is needed to load a Sequel SMRTcell than an RS II cell; however, more DNA is still required for Sequel library prep. CSHL also tested various advances in SMRTcell and polymerase chemistry. During the course of this testing, the NC350 Maize genome was sequenced to 64× coverage with a mean read length > 6000 bp and N50 > 10353 bp. Although these values are somewhat less than what can be expected on the RS II, the average yield per SMRTcell was 2.5 Gb or ~3× the typical RS II yield.

In addition to our work optimizing the Sequel, we worked with Swift Biosciences to test and optimize a new, single-pass PacBio library preparation kit. Unlike the standard PacBio kits, the Swift method does not allow for circular sequencing. In general, long DNA molecules cannot take advantage of the circular consensus method from PacBio, so making long DNA molecules circular consensus compatible is an unnecessary, time-consuming step. The Swift prep is a one-day protocol that requires about two-thirds of the DNA needed to generate a circular consensus compatible library, which also allows larger fragments to be sequenced. Using this method, we were able to generate libraries with mean fragment lengths as high as 12 kb and N50s approaching 20 kb.

In 2017, we will be continuing to work on optimizing the Sequel and RS II instruments. Additional work is needed to increase the mean fragment lengths

on the Sequel and RS II while limiting the amount of DNA required. This will involve continued work with PacBio, Swift, and other groups to validate new chemistries and methods, including new means of fragmentation and size selection.

Oxford Nanopore MinION Sequencing and Analysis Optimization

S. Goodwin, R. Wappel, S. Iyer, W.R. McCombie [in collaboration with F. Sedlazeck and M.C. Schatz, Johns Hopkins; R. Levine and T. Baslan, Memorial Sloan Kettering Cancer Center]

Over the course of 2016, we have seen improvements in MinION flow cell performance and reliability, with the most recent flow cell version achieving a 95% accuracy rate in some cases. We have also observed a significant increase in throughput on the device from ~500 Mb in 2015 to as much as 7 Gb in 2016. These significant advances have led to the initiation of several new projects attempting to capitalize on the unique features on the Oxford Nanopore MinION.

The first of these projects focuses on the rapid and potentially low-cost characteristics of the Oxford Nanopore MinION. Previous work by our collaborators has shown that sparse sequencing (<0.1× coverage) is an effective means of detecting large CNVs in the human genome (Baslan et al. 2015). Detecting these events is important, as many diseases, including cancer, are characterized by large CNV events. To test the feasibility of using Nanopore technology for this application, we initially used the SKBR3 human cell line. This cell line has many large CNVs and has been extensively studied at CSHL. This portion of the study showed that ~50 k short reads (600–1000 bp) were sufficient for roughly even distribution over the genome, which is required for CNV calling. It is likely long reads can also be used in this fashion, but at the start of the study the MinION yield was such that short reads were needed to maximize total reads generated.

We tried this method on three clinically derived acute amyloid leukemia (AML) samples for which a karyotype and an Illumina CNV profile already existed. We were initially able to generate up to 300,000 reads per AML sample, of which ~60% uniquely mapped. These data sets showed very high concordance with MiSeq CNV profiles and identified nearly all of the same CNV events. We also used this method to test three clinical AML samples for which the karyotype failed. In these cases, we were using an improved version of the ONT chemistry and were able to generate >500,000 reads, of which ~80% uniquely mapped. As seen with the previous test of AML samples, there was very high concordance between the ONT and MiSeq data. We have ongoing testing with longer reads showing that similar performance can be seen if at least 50,000 reads are generated. Most recently, using NA12878, we have also shown that more than three million short reads are possible with the most current chemistries.

The second project we are working on is focusing on very long reads. Unlike the Pacific Biosciences instrument, the pores on the MinION are not inhibited by very long DNA. Rather, the technical difficulties lie in getting properly formed libraries to the pore. We have acquired a 75-kb bacterial artificial chromosome (BAC) clone coding for selected spider silk genes. This BAC will be digested with NOT1 and turned into an ONT library to determine the efficiency of sequencing consistent, very long molecules.

Finally, we are exploring the use of ONT technology as a single-cell platform. Unlike Illumina instruments, which require signal amplification to detect DNA, the MinION is a single-molecule sequencer. In theory, if a DNA molecule is present, it will eventually be sequenced by the MinION. We have attempted preparing libraries from as little as 500 pg of DNA, which is roughly equivalent to 100 single cells. This input generated more than 11,000 reads with an average read length of 10 kb. Although this is far below the typical output of a MinION run, this is quite similar to 100,000 1000-bp reads, well within the minimum threshold for CNV calling based on our previous work. We are working on optimizations that will increase yield and allow for lower input amounts.

Definition of the Wheat Epigenome

W.R. McCombie, S. Goodwin [in collaboration with R. Martienssen, M. Regulski, J. Han, C. LeBlanc, and M. Donoghue, CSHL; L. Gardiner, Earlham Institute; A. Hall and J. Kenny, University of Liverpool; K. Mayer and M. Pfeifer, Helmholtz-Zentrum Muenchen; M. Bevan, John Innes Centre Institute]

The focus of the ERA-caps project is to define the epigenome of the Chinese Spring Wheat 42 (CS 42)

and other elite wheat strains. This work consists of several experimental goals. The first phase will consist of high-throughput sequencing of bisulfite-converted DNA following pipelines developed and optimized at CSHL. Ultimately, 50× coverage of ~1000-bp paired-end sequencing data will be generated for CS 42. Transcriptome sequencing and small RNA sequencing of the same material will also be carried out. These steps will be used to correlate gene expression with epigenetic state and explore transposon methylation within proximity to coding regions of the genome.

The second experimental goal will entail the development of capture probes for methylation analysis. The design of this probe set will allow for the capture and high depth (~100×) sequencing of the wheat exome and flanking sequences (~24 Mb) while preserving methylation state.

Following preliminary analysis and method development for the CS 42 strain, these methods will be employed to enable several studies. These include defining the epigenome of hybrid progenitors and their germline, exploring the epigenome and sRNA changes in the newly formed hybrids, understanding the epigenomes and RNA changes during hybrid stabilization, and exploring the relationship between epigenetic modification and the environment.

In 2016, the initial experimental goal of defining the epigenome of CS 42 was initiated. This entailed developing and optimizing methods of DNA isolation and bisulfite treatment to generate consistent and high-quality DNA libraries for high-throughput sequencing. The preliminary results reveal that >60% of the reads align to the genome, and the alignments target ~20% of the genome; these are the expected results given the distribution of CpG islands. It was also observed that the coverage across the targeted region is about even, indicating very little bias in the library preparation. Moving forward, this optimized library preparation strategy will be used to generate libraries for CS 42 and other strains for sequencing on the Illumina HiSeq 2500 Rapid PE 250 format.

Development of the MaizeCODE Project

W.R. McCombie, S. Goodwin, M. Kramer [in collaboration with T. Gingeras, C. Davis, D. Jackson, R. Martienssen, M. Regulski, D. Micklos, M. Schatz, and D. Ware, CSHL]

The MaizeCode project is a large multi-institutional project to create a comprehensive encyclopedia of maize and teosinte, a maize ancestor. As one of the most economically and agriculturally important crops in the world, this will be an invaluable resource for future research into this essential crop. The initial phases of this project will focus on the generation of "platinum genomes" of important strains such as NC350 and W22, as well as teosinte. Further work for this genome will involve performing RNA-seq, CHiP-seq, synthetic long-read sequencing, and other methods on various strains and tissue types to map out the genomic and transcriptomic differences among relevant phenotypes. As data are generated, portions of the Cold Spring Harbor team will be developing analytical pipelines, data display methods, and educational outreach strategies to facilitate and simplify access for the agricultural community as a whole.

In 2016, the MaizeCode project completed long-read sequencing of the NC350 strain. Cumulatively, >100 Gb of data were generated, of which >50% was in reads of >10 kb. This is the first long-read genome completed on CSHL's PacBio Sequel instrument, acquired in 2016. This device is showing significantly better yield than our popular PacBio RS II device, with individual SMRTcells generating up 5 Gb. Draft versions of this data will be available through CyVerse. In addition to the long-read sequencing, 10× genomics sequencing was also completed for NC350. A draft assembly of these data is currently available on CyVerse.

PUBLICATIONS

Aguirre-Chen C, Kim N, Mendivil Ramos O, Kramer M, McCombie WR, Hammell CM. 2016. *C. elegans* PVD neurons: A platform for functionally validating and characterizing neuropsychiatric risk genes. *bioRxiv* doi: 10.1101/053900.

Baker EAG, Goodwin S, McCombie WR. 2016. SiLiCO: A simulator of long read sequencing in PacBio and Oxford NanoPore. *bioRxiv* doi: 10.1101/076901.

Goes FS, Piroozna M, Parla JS, Kramer M, Ghiban E, Mavruk S, Chen YC, Monson ET, Willour VL, Karchin R, et al. 2016. Exome sequencing of familial bipolar disorder. *JAMA Psychiatry* **73:** 590–597.

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CK, Severn-Ellis A, McCombie WR, Parkin IA, et al. 2016. The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat Commun* **7:** 13390.

Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet* **17:** 333.

Lee H, Gurtowski J, Yoo S, Nattestad M, Marcus S, Goodwin S, McCombie WR, Schatz M. 2016. Third-generation sequencing and the future of genomics. *bioRxiv* doi: 10.1101/048603.

Martin PM, Stanley RE, Ross AP, Freitas AE, Moyer CE, Brumback AC, Iafrati J, Stapornwongkul KS, Dominguez S, Kivimäe S, et al. DIXDC1 contributes to psychiatric susceptibility by regulating dendritic spine and glutamatergic synapse density via GSK3 and Wnt/β-catenin signaling. *Mol Psychiatry* **21:** 1644.

Roberts NJ, Norris AL, Peterson GM, Bondy ML, Brand R, Gallinger S, Kurtz RC, Olson SH, Rustgi AK, Schwartz AG, et al. 2016. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov* **6:** 166–175.

In Press

Mardis E, McCombie WR. 2017. Preparing plasmid subclones for capillary sequencing. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094581.

Mardis E, McCombie WR. 2017. Preparing polymerase chain reaction (PCR) products for capillary sequencing. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094599.

Mardis E, McCombie WR. 2017. Whole-genome sequencing: Manual library preparation. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094615.

Mardis E, McCombie WR. 2017. Whole-genome sequencing: Automated, nonindexed library preparation. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094623.

Mardis E, McCombie WR. 2017. Automated library preparation for DNA sequencing. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094631.

Mardis E, McCombie WR. 2017. RNA-Seq: RNA conversion to cDNA and amplification. *Cold Spring Harb Prot* doi: 10.1101/pdb.prot094672.

PLANT GENOME ORGANIZATION, EVOLUTION, AND FUNCTION

D. Ware M. Campbell, K. Chougule, Y. Jiao, V. Kumar, S. Kumari, Y.K. Lee, Z. Lu, M. Neves Dos Santos Leite, A. Olson, M. Regulski, J. Stein, M.K. Tello-Ruiz, J. Thomason, P. Van Buren, B. Wang, L. Wang, X. Wang, S. Wei, L. Zhang

The Ware lab has two primary goals: (1) research into plant genomics, with a focus on functional and comparative genomics of model plant systems and agriculturally important crop plants; and (2) development of tools, data sources, and resources for the genomics research community. The Ware lab also contributes actively to three large-scale, multi-institutional, cyberinfrastructure collaboratives designed to serve broad research and educational communities: the National Science Foundation (NSF)-funded Gramene and CyVerse projects and the U.S. Department of Energy (DOE)-funded Systems Biology Knowledgebase (KBase).

The Ware lab is a dynamic research group and its composition has continued to evolve, keeping pace with its research objectives. During this period, one of our computational developers, Joseph Mulvaney, went on to newer career opportunities, while Xiaofei Wang joined our team as a computational science analyst. Also, Dong Ding successfully concluded his collaborative research with the lab as a visiting scientist from Huazhong Agricultural University, China, and Erin DeNardo completed her Undergraduate Research Program (URP) during the summer.

PLANT GENOME RESEARCH

In the last decade, the sequencing and annotation of complete plant genomes has helped us understand plant function and evolution, as well as how to alter economically important traits. Efforts in many disparate disciplines are required to generate reference genomes. The work at the Ware lab often starts with laboratory scientists performing wet chemistry to generate the raw sequence data. Next, computational biologists and bioinformaticians kick off a series of computational steps to interpret the raw data. The process of interpretation involves the assembly of raw sequence reads into overlapping segments ("contigs"), which are combined to create a scaffold. This scaffold, in turn, discerns the position, relative order, and orientation of contigs within the chromosomes. The next step is annotation, the discovery and description of genes and other functional elements, and homologies (evolutionary relationships) with other genomes. This information must be faithfully communicated and visualized in web-based platforms, such as Gramene.

All of these activities are rapidly evolving with the advent of new sequencing technologies, algorithms, and data-handling requirements. For example, high-depth and low-cost sequencing of RNA transcripts is providing a vast stream of new evidence that informs genome annotation; this, in turn, has spurred the development of new software for modeling and performing genome annotation. Low-cost sequencing has also made it possible to ask whole new classes of questions, moving beyond the generation of single references for individual species and supporting the development of multispecies representation as a "pangenome." Ongoing projects within the maize, rice, and *Arabidopsis* research communities are now sequencing hundreds or thousands of genotypic backgrounds, chosen from carefully constructed populations, wild populations, and breeding germplasms in each species. Information about genetic variation is helping scientists to understand the genetic basis of phenotypic traits and the origins of domesticated crop species. New technologies are also driving research in epigenetics, the study of heritable variation not attributable to changes in the underlying genome sequence. Epigenetic mechanisms include modification of DNA by methylation and various forms of histone modification, which can cause changes in gene expression and other phenotypes. Both types of modifications can be studied using new sequencing technologies and analytical methods.

Updating the Maize B73 Reference Genome and Annotation Using Long Sequence Read Technologies

M. Campbell, Y. Jiao, S. Kumari, A. Olson, M. Regulski, J. Stein, B. Wang [in collaboration with The Maize Genome Sequencing Consortium: R. McCombie and S. Goodwin, CSHL; K. Guill and M. McMullen, Missouri University, USDA–ARS; J. Ross-Ibarra, UC Davis; K. Dawe, University of Georgia; D. Rank, Paul Peluso, and Tyson Clark, Pacific Biosciences; A. Hastie, BioNano Genomics]

Maize is an important feed and fuel crop, as well as a model system in developmental genetics. A complete and accurate reference genome is imperative for sustained progress in understanding the genetic basis of trait variation and crop improvement. The Ware laboratory has played a leading role in the development and stewardship of the maize reference genome since inception of the project a decade ago. The 2009 release of the B73 reference sequence was a milestone in plant genomics research because of the unprecedented size and complexity of the maize genome. Through several updates, this foundational resource has remained the principal genome reference for the maize research community. Yet, it continues to be a work in progress, with gaps and misassemblies that have defied available sequencing technologies, especially over the highly repetitive regions that are also the most dynamic and rapidly evolving.

This year marks a new milestone with the release of an entirely new reference assembly, designated B73 RefGen_v4. In collaboration with Pacific Biosciences, we sequenced maize to ~65× coverage using Single-Molecule Real-Time (SMRT) technology. These exceptionally long reads, typically >10 kb, were able to span not only genes, but also the extensive intergenic and repetitive regions rich in transposons. Using an optimized correction and assembly pipeline, we built a de novo assembly consisting of 2958 contigs and totaling 2.10 Gb, with more than half of this length in contigs of >1.2 Mb. This represents a 52-fold improvement in assembly contiguity compared with the previous bacterial artificial chromosome (BAC)-based maize assembly, while nucleotide agreement was maintained at >99.9%. Optical maps of the maize genome, developed at BioNano Genomics, enabled scaffolding of these contigs into the 10 chromosomes of maize, thereby placing >99% of genes. Many improvements in gene order and orientation were found. Just as important, the new sequence fills in many intergenic regions that harbor transposons and regulatory regions influencing gene expression. In addition, comparative optical mapping of two other inbreds revealed a prevalence of deletions in the region of low gene density and maize lineage-specific genes (Jiao et al. 2017).

Uncertainties about the complete structure of mRNA transcripts, particularly with respect to alternatively spliced isoforms, can be a limiting factor for research in the system. In addition to working on the reference genome sequence, we are using the same single-molecule sequencing technology to investigate the maize transcriptome. For this work we have sampled full-length cDNAs from six tissues of the maize inbred line B73; they were barcoded, pooled, size-fractionated (<1 kb, 1–2 kb, 2–3 kb, 3–5 kb, 4–6 kb, and 5–10 kb), and sequenced on the PacBio RS II platform with P6-C4 chemistry. We were able to capture 111,151 transcripts, representing ~70% of the genes annotated in the current maize RefGenV3 genome assembly. A large proportion of transcripts (57%) are novel. We were able to validate ~ 90% of the transcripts' splice-site-junctions within high-depth short reads generated from the matched tissues. In addition, we identified a large number of novel long noncoding RNAs (lncRNAs) and fusion transcripts. Our results show that the characterization of the maize B73 transcriptome is far from complete, and that maize gene expression is more complex than previously thought (Wang et al. 2016).

A Sorghum Mutant Resource as an Efficient Platform for Gene Discovery in Grasses

Y. Jiao [in collaboration with Z. Xin and J. Burke, USDA-ARS]

Sorghum (*Sorghum bicolor* L. Moench) is a versatile C4 crop and a model for research in the family Poaceae. High-quality genome sequence is available for the elite inbred line BTx623, but functional validation of genes remains challenging because of the limited genomic and germplasm resources available for comprehensive analysis of induced mutations. In this study (Jiao et al. 2016), we generated 6400 pedigreed M4 mutant pools from ethyl methane sulfonate (EMS)-mutagenized BTx623 seeds through single-seed descent. Whole-genome sequencing of 256 phenotyped mutant lines revealed more than 1.8 million canonical EMS-induced mutations, affecting >95% of genes in the sorghum genome. The vast majority (97.5%) of the induced mutations were distinct from natural

variations. To show the utility of the sequenced sorghum mutant resource, we performed reverse genetics to identify eight genes potentially affecting drought tolerance, three of which had allelic mutations and two that showed exact cosegregation with the phenotype of interest. Our results establish that a large-scale resource of sequenced pedigreed mutants provides an efficient platform for functional validation of genes in sorghum, thereby accelerating sorghum breeding. Moreover, findings made in sorghum could be readily translated to other members of the Poaceae via integrated genomics approaches (Jiao et al. 2016).

Pan-Genome Annotation

M. Campbell and E. DeNardo [in collaboration with W.H. Majoros, A.S. Allen, T.E. Reddy, Duke University; C. Holt, M. Yandell, University of Utah]

The accurate interpretation of genetic variants is critical for characterizing genotype–phenotype associations. Because the effects of genetic variants depend on their local genomic context, accurate genome annotations are essential. Furthermore, as some variants have the potential to disrupt or alter gene structure, variant interpretation efforts stand to gain from the use of individualized annotations that account for differences in gene structure between individuals or strains. We developed a suite of haplotype-aware software tools for predicting functional changes in gene structure that may result from sequence variants. ACE (assessing changes to exons) converts phased genotype calls to a collection of explicit haplotype sequences, maps transcript annotations onto them, and detects gene-structure changes and their possible repercussions, including several classes of possible loss of function. Using publicly available RNA-sequencing data, we showed that novel transcripts predicted by ACE are commonly supported by spliced RNA-seq. We also show that ACE predictions confirm earlier results regarding the quantitative effects of nonsense-mediated decay, and that predicted loss-of-function events are highly concordant with patterns of intolerance to mutations across the human population. ACE can be readily applied to diverse species, including animals and plants, making it a broadly useful tool in eukaryotic population-based resequencing projects, particularly for assessing the joint impact of variants at a locus. Future enhancements currently under development include a novel probabilistic model to predict the effects of genetic variants that alter splicing enhancers or that create pseudo-exons, and quantitative prediction of changes in stoichiometric isoform ratios. A manuscript describing this work was recently published (Majoros et al. 2016).

Gramene: Comparative Genomic Resource for Plants

M. Campbell, K. Chougule, Y. Jiao, S. Kumari, Y.K. Lee, J. Mulvaney, A. Olson, J. Stein, M.K. Tello-Ruiz, J. Thomason, B. Wang, S. Wei, L. Zhang [in collaboration with P. Jaiswal, Oregon State University; P. Kersey and R. Petryszak, EMBL-European Bioinformatics Institute; L. Stein, Ontario Institute of Cancer Research; C. Taylor, American Society of Plant Biologists; R. Wing, University of Arizona]

The Gramene project provides online reference resources for plant genomes and curated pathways to aid functional genomics research in crops and model plant species. Our website (www.gramene.org) facilitates studies of gene function by combining genome and pathway annotation with experimental data and cross-species comparisons. In other words, the data and tools in Gramene enable plant researchers to use knowledge about gene function in one species to predict gene function in other species. Drawing these connections facilitates translational research in plant development and physiology that influences economically important traits (e.g., grain development, flowering time, drought tolerance, and resistance to diseases). In the past year, the project accomplished several major milestones, culminating in data build 52 (November 2016), our 54th release since the inception of the project, with a significantly streamlined user interface and back-end functions. The Gramene team develops comparative genomics databases in collaboration with the Ensembl Plants project at the European Bioinformatics Institute (EMBL-EBI), and collaborates closely with the EBI's Expression Atlas project to provide manually curated, quality-controlled and analyzed transcriptomic data. We continued to host genome and pathway annotations via (1) the Ensembl genome browser and (2) the Plant Reactome pathways portal. Additions this year to our reference genome resource include five new species: sugar beet (*Beta vulgaris*), oilseed rape (*Brassica napus*), red clover (*Trifolium pratense*), and two red algae (*Galdieria sulphuraria* and *Chondrus crispus*). In addition, we made significant updates to corn

(B73 *Zea mays* RefGenV4; Fig. 1) and bread wheat (*Triticum aestivum* TGACv1) genome assemblies and updated the gene annotations for *Arabidopsis thaliana* (i.e., AraPort11). We incorporated new genetic variation data for rice (*Oryza sativa japonica*) and its wild relative, *Oryza glumaepatula*. In collaboration with USDA scientists, we released genotype data for a mutagenized population of *Sorghum bicolor* that was previously processed by our group (Jiao et al. 2016). Similarly to naturally occurring single-nucleotide polymorphisms (SNPs) and structural variants, the EMS-derived point mutation variants are displayed in the context of gene annotation and inferred functional consequences, which can be assigned to individual accessions within a sampled diversity panel (Fig. 2). In addition, we are now providing a new HMMER tool for fast protein sequence search using hidden Markov models. We added new track hubs for more than 900 public RNA-seq studies, totaling more than 16,000 tracks across 35 plant species and new noncoding RNA (ncRNA) alignments across all plant species. Curated baseline gene expression data from the Expression Atlas project (https://www.ebi.ac.uk/gxa/plant/experiments) are now available for 698 experiments in 17 plant species from both our genome and pathway browsers, as well as from a new search results view (Fig. 3). In addition, we expanded our Plant Reactome database by including more than 200 manually curated rice pathways and orthology-based pathway projections in 66 species.

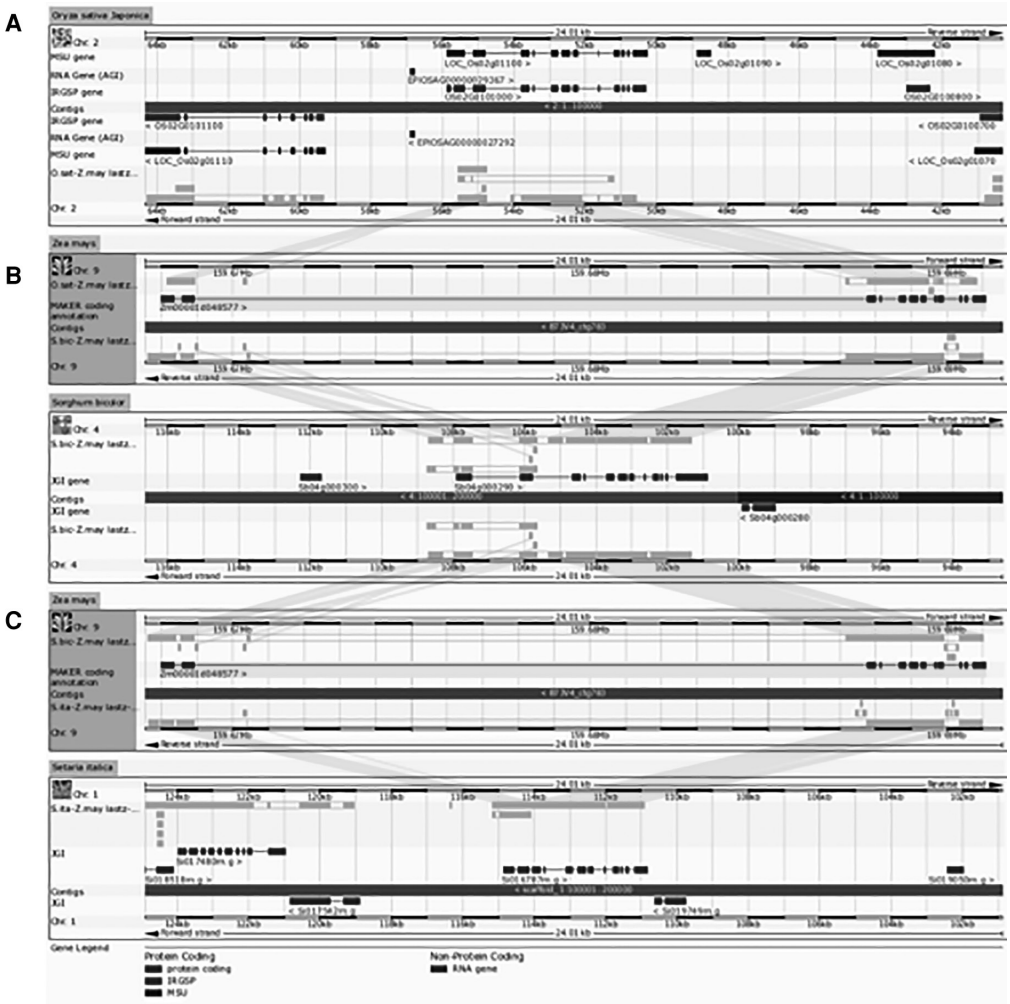


Figure 1. Whole-Genome Alignment between *Zea mays* v4 and each of (*A*) *Oryza sativa*, (*B*) *Sorghum bicolor*, and (*C*) *Setaria italica* in the Gramene genome browser build 52.

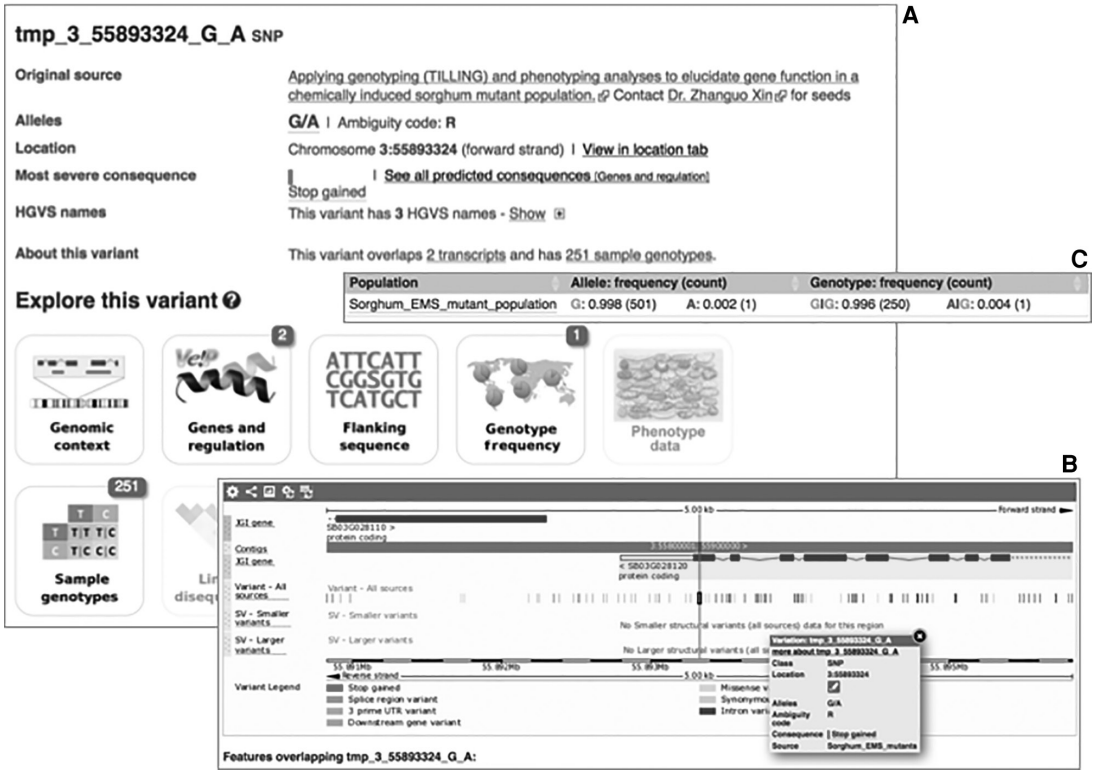


Figure 2. Gramene genome browser views (build 52) of sorghum ethyl methane sulfonate (EMS)-induced variation. (*A*) Exploratory view of each of ~1.8 M EMS-induced genetic variants from Jiao et al. (2016), (*B*) genomic context view, and (*C*) allele and genotype frequency views.

Gramene's integrated search database and modern user interface (http://search.gramene.org) leverage diverse annotations to facilitate finding genes through selecting autosuggested filters with interactive views of the results. The interface differs from most other search interfaces in that it guides the user to select filters based on gene identifiers, names, and descriptions, as well as structured annotations such as InterPro domains, Gene Ontology terms, and Plant Reactome pathways. Complex queries can be composed by adding more filters. A paginated list of matching genes is displayed in which each gene can be dynamically loaded to reveal detailed views, which can suggest additional search filters. Each gene match is presented with distinct possible tab views: Location, Expression, Homology, and X-refs (i.e., cross-references) (Fig. 2), and a new view for Pathway results is under development. Search results are also summarized via facet counting, which can rapidly calculate the genomic distribution of gene search results across all hosted genomes. Our new search API is hosted at data.gramene.org.

In the interest of rapid and broad dissemination of data resources to the plant community, we expedited the release of the updated maize B73 RefGen_V4 reference assembly and annotation before scientific publication via a dedicated—although provisional—genome browser resource (maize4.gramene.org). This early data release was performed in accordance with guidelines set forth by the Toronto agreement. Whereas Gramene's main site (www.gramene.org) is committed to supporting community-recognized annotations for a single reference genome per species, this resource features two maize assemblies: the old maize V3 and new V4 assemblies. To facilitate comparative and functional genomics research in corn, this resource also hosts a subset of key species, including sorghum, rice, soybean, *Setaria italica*, *Brachypodium distachyon*, tomato, banana, grape, and *Amborella trichopoda*. Featured annotations were used to build phylogenetic gene trees and define orthologous and paralogous relationships using the Ensembl Compara gene tree method. To compare these genomes

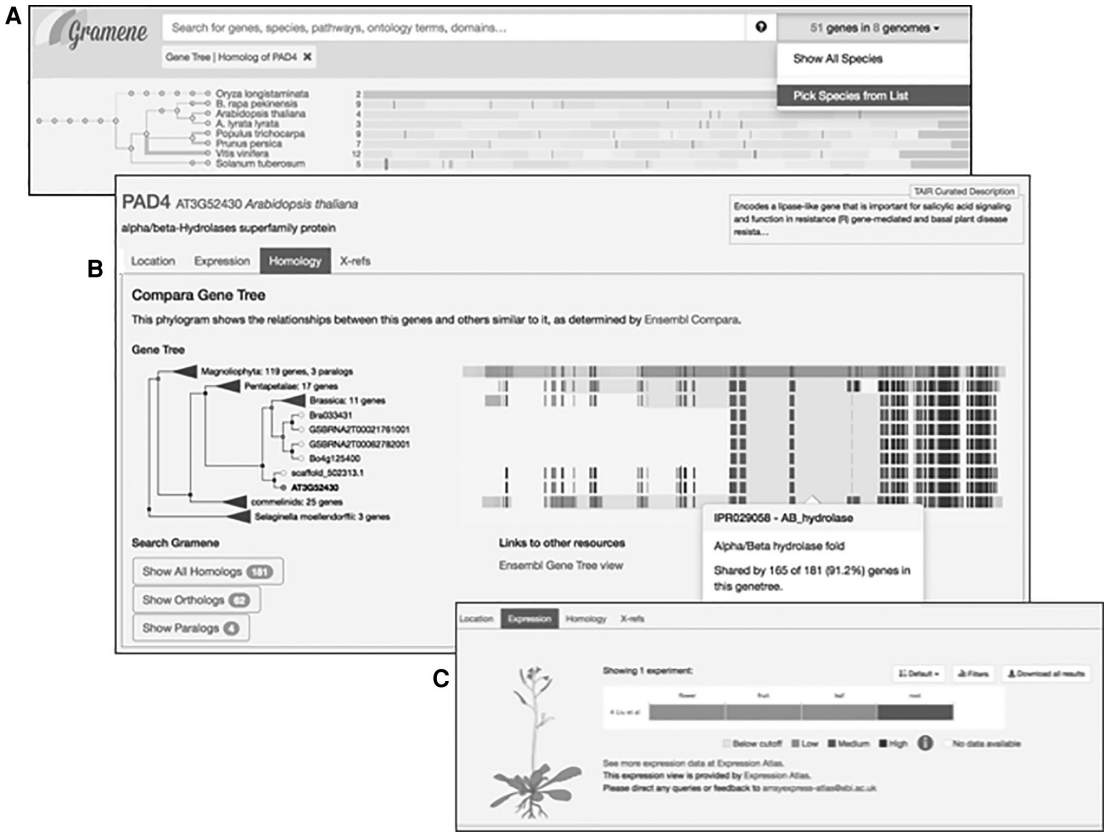


Figure 3. New features and views of Gramene's search interface. (*A*) The drop-down menu on the right side of the search box allows users to select species to be displayed in the Gene Tree (Homology) view. (*B*) The Homology view shows inferred evolutionary histories from Compara, integrated with functional domain information from InterPro. (*C*) The Expression view shows baseline expression in various experimental conditions, tissues, and from distinct sources/studies.

we applied systematic annotation and phylogenetic analysis of protein-coding genes. These results are yielding new insights into the taxonomic origin of genes and patterns of duplication, movement, and loss influenced by genome architecture. At present, the maize V3 browser is still available as legacy, but the V4 assembly and annotations have now been incorporated into the main Gramene genome browser (see Fig. 1).

Pathogen Immunity Genes in Wild Related Species of Rice

K. Chougule, J. Stein, S. Wei [in collaboration with R. Wing, University of Arizona; The International Oryza Map (I-OMAP) Consortium]

Pathogens such as rice blast (*Magnaporthe oryzae*) severely impact rice production and may pose an increasing threat as climate change alters the geographical range of pests in the future. Breeding for natural host resistance is a proven strategy, but is constrained by limited sources of variation. Wild relatives of rice, collected from around the world, provide a reservoir of resistance genes that can be transferred to cultivated rice by introgressive breeding. Taking advantage of an 11-species set of I-OMAP consortium reference genomes in OGEGramene (www.oge.gramene.org) that spans the *Oryzeae* tribe (including wild species in the *Oryza* and *Leersia* genera), we discovered more than 4500 NLR (nucleotide-binding domain and leucine-rich repeat-containing) genes in 28 families. Rapid diversification of complex haplotypes by gene expansion and loss is typical of NLR genes, contributing to disease adaptation. Applying phylogenetic reconciliation methods to gene trees in these 28 families, we found a 10-fold increase in

duplication rates in lineages leading to both Asian and African cultivated rice, consistent with selection for resistance traits before domestication. Most NLR genes were positionally clustered, often forming complex arrangements of distantly related genes. Yet, clear orthologous relationships and evidence of conserved underlying haplotype structures could be drawn, even in the most distantly related (~17 MY) species of *Leersia*. In rice and other plants, disease resistance is sometimes conferred by the required action of two neighboring NLR genes, whose products function as heterodimers. Examining all possible combinations of adjacent NLR pairs, we found that those reminiscent of functionally coupled NLR genes (i.e., belonging to heterogeneous families and configured in head-to-head [divergently transcribed] arrangement) are significantly more prevalent than those expected by chance and also more likely to be conserved across species than other arrangements. This finding suggests that evolution of haplotype structure may be constrained by underlying regulatory and functional interactions among such putative coupled NLR genes. Furthermore, we found a greater prevalence of putative integrated decoy domains among such pairs, which are thought to function in pathogen recognition by mimicking host targets of pathogenicity factors. Striking variation in domain structure suggests that swapping of various decoy domains contributes to the evolution of haplotype diversity and resistance specificity. This study has opened a treasure trove of potentially novel resistance functions that may help in the future development and sustainability of rice (Stein et al. 2017, currently under review).

PLANT SYSTEMS BIOLOGY

Exploring *Arabidopsis* Gene Regulatory Networks

Y.K. Lee, C. Liseron-Monfils, C. Noutsos, B. Wang, L. Zhang [in collaboration with S. Brady, UC Davis; M. Frank and B. Shen, DuPont-Pioneer, Inc.]

Plant microRNAs (miRNAs) play important roles in regulating plant development and stress responses by posttranscriptionally repressing the expression of their target genes. To identify upstream regulators of miRNA expression, we generated the *Arabidopsis* miRNA Gene Regulatory Network (ARMIG) using a yeast one-hybrid (Y1H) approach. Using a nearly complete root transcription factor (TF) library, we screened 180 miRNA promoters, their targets, and TFs that are highly connected within the network, and obtained 5376 protein–DNA interactions (PDIs). ARMIGs are characterized by highly connected genes as "hubs" in the network. The ZF-HD TFs were identified as hub genes within the miRNA network. An initial survey of six single loss-of-function mutants identified from existing T-DNA collections revealed no observable root phenotypes, suggesting that the ZF-HD TFs are functionally redundant. To test this hypothesis, we generated multiple loss-of-function mutants using different approaches, including genetic crosses to a single loss-of-function mutant, generation of amiRNA, and independent repressor lines (Fig. 4). Combining multiple loss-of-function mutants within a single *Arabidopsis* line, we observed several phenotypes, including altered flower structures and an increased vegetative branching. In addition, we performed transcriptome profiling of ZF-HD TF mutants to characterize differentially expressed genes. The gene-expression profiles and phenotypic properties of the mutants suggested a role for ZF-HD TFs as regulators of developmental transitions. Moreover, the results of this work show that the miRNA GRN can be applied more generally, beyond the root system.

To understand molecular mechanisms of ZF-HD TFs, we use RNA-seq analysis to reveal gene perturbation between wild-type and loss-of-function mutants, including the amiRNA transgenic line, *hb23/31/33/34* mutant, and two independent repressor lines, 35s:HB21:SRDX and 35s:HB31:SRDX. In the *hb23/31/33/34* mutant, 91% of differentially down-regulated genes are found in the amiRNA line, whereas 48% and 53% of DE genes are found in the HB31 and HB21 repressor lines, respectively. A similar pattern was observed for DE putative TF genes. To identify direct targets of ZF-HD TFs at genomic scale, we performed genome-wide chromatin immunoprecipitation (ChIP)-seq analysis of HB34 in young flower tissue. To find out the real target of ZF-HD TFs, we are trying to overlay with ChIP-seq and RNA-seq data. Collectively, these relationships constitute a direct coherent type IV feed-forward loop mediated by miR157d for regulation of plant architecture in *Arabidopsis*. To apply this in crop plants, we are screening mutant populations of maize and sorghum.

The bZIP TFs constitute another highly connected TF family. One member of the family, GBF2, binds many miRNA and miRNA target promoters. As with

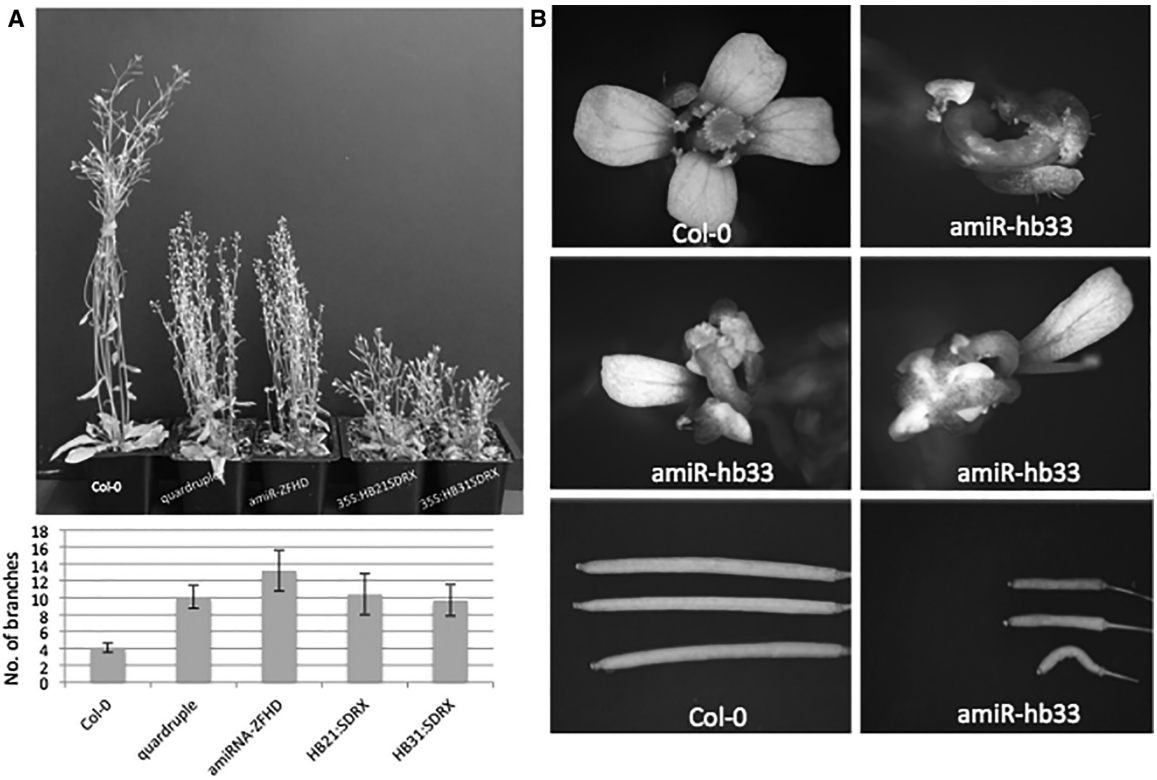


Figure 4. Phenotype of the ZF-HD TF mutant. (*A*) Branching phenotype of quadruple mutant, amiRNA, and two independent repressor lines compared with wild-type plant Col-0. (*B*) Lateral branching number of wild type and mutants.

the ZF-HD TFs, single-gene loss-of-function mutants do not have clear phenotypes. We are in the process of generating multiple loss-of-function mutants by performing genetic crosses between the available single-gene mutant lines. Initial screens for homozygous mutants from multiple loci suggest that these mutants may be lethal or have large fitness disadvantages.

Dissection of Gene Regulatory Networks Associated with Abiotic Stress Responses

K. Chougule, V. Kumar, S. Kumari, C. Liseron-Monfils, Z. Lu, A. Olson, L. Wang, X. Wang, L. Zhang [in collaboration with A.-M. Bagman, A. Gaudinier, S. Brady, UC Davis; M. Frank and B. Shen, DuPont-Pioneer, Inc.; J. Magalhães, Embrapa; L. Kochian, University of Saskatchewan]

Plants depend on an adequate supply of macro- and micronutrients from soil for sustained growth and development. Nitrogen and phosphorus are among the most critical macronutrients for optimum crop production. Plants absorb nitrogen primarily as nitrate and ammonium. Nitrogen is key to energy transfer processes and an essential component of nucleic acids, proteins, vitamins, etc. Phosphorus is absorbed primarily as orthophosphate (Pi). Phosphorus is critical to numerous plants processes such as energy transfer, photosynthesis, signal transduction, and regulation of protein activities—and is an essential component of nucleic acids, cellular membranes, etc. Under variable condition of nutrients (deficient, sufficient, and excessive), plants respond with adaptive mechanisms of nutrient uptake, assimilation, and distribution. The N and P uptake and assimilation are both adversely affected by abiotic stresses such as salinity, drought, and extreme temperatures. We have used a multipronged approach consisting of Y1H system, transcriptomics, and epigenetics assays to study and better characterize the gene regulatory networks that govern the process of N and P uptake, assimilation, use, remobilization, and transcriptional regulation. The data sets are currently under evaluation and in the next year we anticipate identification of TF families that regulate expression of gene response to abiotic stress.

Developmental Networks Controlling Inflorescence Architecture in Grasses

Y. Jiao, S. Kumari, Y.K. Lee [in collaboration with Z. Xin and J. Burke, USDA–ARS]

The goal of this work is to integrate genetics and genomics data sets to find molecular networks that influence the morphology (architecture) of grass inflorescences (flowers). Because inflorescences bear the fruits and grains that we eat, the genetic and regulatory factors that govern their formation are clearly relevant to important agronomic traits such as grain yield and harvesting ability. In addition to our previous work in maize, we have begun work on sorghum, an important emergent bioenergy crop, which is also used for human consumption in sub-Saharan Africa. The number of grains per panicle is a developmental trait that contributes to sorghum yield. Sorghum flowers comprise one fertile (sessile) and two sterile (pedicellate) spikelets with only the sessile spikelet producing seed. Using an EMS population, we identified independent multiseed (*msd*) mutants with both fertile sessile and pedicellate spikelets. A detailed dissection of developmental stages of wild-type and *msd1* mutant found the pedicellate spikelets in wild-type do not have floral organs, including ovary, stigma, filament, or anther, whereas the *msd1* mutants generate intact floral organ in the sessile spikelet. Using a bulk segregant analysis of an F2 individual, we identified the TCP transcriptional factor. The six causal SNPs found in *msd1* gene are highly conserved in grass species. The *TCP* gene was found to be differentially expressed during inflorescense development. To characterize the gene networks associated with the pedicellate spikelet fertility, we generated whole-genome expression profiling data of size tissues at four different sorghum inflorescense development stages in both the wild-type and the *msd1* mutant. Preliminary analyses suggest the *MSD1* gene may program cell death in pedicellate spikelets in wild-type through activation of hormone pathways.

In the next phase, we will prioritize candidates identified in this stage of the project by overlaying *Arabidopsis* regulatory network information (see the previous section). The resulting hypotheses (e.g., those related to stress responses) can be tested in *Arabidopsis* and ultimately translated into agronomic systems. Candidate genes that are maize- and/or grass-specific are also of high priority because they may contribute to the unique morphology of maize inflorescences and/or features shared among other grasses. In addition, we will use comparative genomics approaches, including both cross-species comparisons and integration of analogous RNA-seq data sets from closely related grasses such as sorghum, to identify candidate genes that may contribute to grass-specific aspects of inflorescence architecture.

CYBERINFRASTRUCTURE PROJECTS

CyVerse (formerly The iPlant Collaborative)

K. Chougule, Z. Lu, P. Van Buren, L. Wang, X. Wang [in collaboration with Cold Spring Harbor Laboratory, using more than 100 staff, and headquartered at the University of Arizona; PIs: A. Parker, D. Ware, N. Merchant, M. Vaughn, and E. Lyons; dozens of collaborators are located at more than 20 institutions]

Our world is changing rapidly. The human population is increasing, whereas arable land and fisheries are decreasing and food cultivation is being diverted for fuel production. Climate instability and energy sustainability are impacting agricultural and ecological systems, whereas concomitant changes in land-use patterns affect global biodiversity. To successfully address these issues, we need to understand how organisms' appearance, physiology, and behavior are shaped by the interactions between their genetic makeup and the environment. Although these global challenges are sobering, the efforts to respond productively will lead to exciting science—provided that the computational infrastructure is in place to handle the necessary datasets, analyses, interpretation of results, and dissemination of knowledge. Advances in biological research technology have enabled scientists to amass unprecedented amounts of data, and many researchers find themselves drowning in this sea of data. Foreseeing this major bottleneck in biological research, the U.S. National Science Foundation (NSF) established the iPlant Collaborative (http://iplantcollaborative.org) in 2008 to develop cyberinfrastructure for life sciences research and democratize access to U.S. supercomputing capabilities.

Having completed its first eight-year grant, the CyVerse Collaborative has made extensive progress toward meeting these goals. Work in the last year culminated in the development of multiple platforms that are now being used by the research and educational communities. Staff at CSHL contributed directly to

some of these cyberinfrastructure platforms or built on existing platforms to provide scientists and educators with ready access to needed software and analysis tools. Within the Ware lab, these platforms include the CyVerse Data Store, DE, and Atmosphere. In addition, the Ware lab has successfully built the first federated CyVerse system at CSHL, which supports a dozen high-performance computing (HPC) apps running on a local cluster and storage system. Leveraging the federation system, a workflow management system has been prototyped for supporting automation of complicated workflows.

The CyVerse Data Store was designed as a distributed storage system for hosting data on the cloud, providing convenient access. The distributed system allows us to "bring the infrastructure to the data" to enable fast computing and, at the same time, to reduce the bandwidth required to transfer large amounts of next-generation sequencing data. Currently, the CyVerse Data Store consists of two major storage systems with fast internet connections, one at the University of Arizona (UA) and another one at the Texas Advanced Computing Center (TACC). CSHL itself has relatively limited bandwidth relative to TACC and UA, making our institution an ideal case for validation of the CyVerse model of bringing the infrastructure to the data. Because it makes more sense to use local computing resources with CyVerse prebuilt analysis workflows via the CyVerse AGAVE API than to copy the data to outside clusters (e.g., it takes >10 min to move 1 GB of data from CSHL to the TACC clusters), the Ware lab at CSHL decided to migrate and has successfully migrated the important CyVerse platforms at CSHL. The first target is the data store, and a local resource server (WildCat) has been successfully added into the CyVerse storage system, allowing data transfer that is 80 times faster than with outside servers. For the second step, the Ware lab has successfully synchronized the resource server with CyVerse's AGAVE servers at TACC, which allows CyVerse's prebuilt workflows to be run locally on CSHL computing servers. The migration of these important platforms provided a proof-of-concept demonstration of the portability of CyVerse Platforms for enrolling more institutes for efficient data management and possibly use of national computing resources provided with good bandwidth.

The DE is the most visible portal for CyVerse tools and services. This web-based platform supports an "app store" model of user-extensible tools, automated workflows, and data storage. Users can take advantage of existing tools—integrated by CyVerse staff and the user community—or add their own tools to use privately or share. Although users may not be aware of this, the underlying infrastructure provides access to CyVerse's massive data store at UA and TACC. Computationally intensive tasks are handled by supercomputers located at TACC and other centers within the Extreme Science and Engineering Discovery Environment (XSEDE). So far, more than 300 tools have been integrated into the DE; these tools enable a broad range of research activities, including genome/transcriptome assembly, annotation, RNA-seq quantitation, variant detection, genome-wide association studies (GWASs), and phylogenetics. Members of the Ware lab have played important roles in contributing to workflow design, tool integration, validation, science tutorials, and documentation.

Atmosphere is CyVerse's configurable and cloud-enabled computational resource for the plant research community. From Atmosphere's web interface, users can launch a virtual machine (VM) with preconfigured working environments and precustomized, ready-to-use software. Users can also create their own applications and environments as VMs and share them with others via Atmosphere. As with the DE, Atmosphere is a gateway to CyVerse's core infrastructure resources, including the high-performance grid computing environment and big-data storage system. Using the Atmosphere platform, we created a VM to be used for image analysis based on high-throughput phenotyping.

The success of genome research depends on our ability to accurately assemble, annotate, and derive meaning from sequence data; however, the extremes of genome size, polyploidy, diversity, and repeat content push the limits of the algorithms, expertise, and computational power currently available to researchers. In response, CyVerse is fostering a community effort to identify best practices and state-of-the-art tools, install them, optimize their performance on the nation's most powerful supercomputers, and make them available as free online resources. Over the last three years, the CyVerse Discovery Environment has matured to provide a comprehensive set of tools and services for sequence handling, read alignments, RNA-seq profiling, and de novo genome and transcriptome assembly. To extend these capabilities, we have incorporated MAKER-P, a

standardized, portable, and easy-to-use plant-genome annotation engine with built-in methods for quality control. As part of this effort, MAKER-P was specifically optimized to take advantage of the parallel computing environment of the TACC Lonestar cluster and is now a supported module. Performance testing showed that MAKER-P provides high-quality, full-fledged annotation pipelines on even the largest plant genomes in a matter of hours. MAKER-P is currently available for use as an Atmosphere image.

A major mission of CyVerse is to promote adoption of the cyberinfrastructure through training workshops and outreach at academic institutions and scientific meetings. In 2016, members of the Ware lab participated as instructors in several "Big Data" workshops focused on transcriptomics and population genomics using the DE and Atmosphere platforms.

KBase: Department of Energy Systems Biology Knowledgebase

V. Kumar, S. Kumari, J. Thomason [in collaboration with U.S. DOE National Laboratories and led by PI A. Arkin, Lawrence Berkeley National Laboratory (LBNL) with co-PIs C. Henry, Argonne National Laboratory (ANL) and R. Cottingham (ORNL); C. Henry replaced R. Stevens as co-PI from ANL, one of the co-PIs, S. Maslov (BNL), transitioned out of the project. M. Schatz (CSHL) and D. Weston (ORNL) also transitioned out of the project during this year. As Plants Science Lead for KBase, D. Ware continues to informally serve as a co-PI on the project.]

The Systems Biology Knowledgebase (KBase; www.kbase.us) has two primary goals. The scientific goal is to produce predictive models, reference data sets, and analytical tools and to show their utility in DOE biological research relating to bioenergy, carbon cycle, and the study of subsurface microbial communities. The operational goal is to create the integrated software and hardware infrastructures needed to support the creation, maintenance, and use of predictive models and methods for the study of microbes, microbial communities, and plants.

KBase's computational infrastructure is supported by a distributed, high-performance, cloud-based system that includes more than 3 petabytes of storage, more than 12,000 cores for data processing, and 90 GBit/sec bandwidth over DOE's ESnet. The KBase data model supports more than 900 data types, including sequence reads, assemblies, genomes, annotations, expression data, quantitative phenotype data, and metagenomic profiles. The KBase data repository has integrated various datasets from public resources for thousands of microbes, hundreds of microbial communities, and tens of plants.

In addition, KBase supports more than 1000 analysis and access functions grouped into more than 30 integrated service modules, including sequence assembly, variation analysis, genome annotation, expression analysis, and various types of modeling (e.g., metabolic, regulatory, and flux balance analysis [FBA]). To capture complex, multistep analyses, users can build persistent and provenance-based workflows, which can capture complex, multistep analyses and mobilize rich annotations, visualization widgets, reusable workflows, and custom scripts. This capability empowers users to create science that is transparent, reproducible, and reusable.

Based on the feedback from the reverse site visit review in September 2015 and additional feedback from the scientific advisory committee and community, the primary focus of the project in 2016 was to upgrade the platform to improve reliability and usability of the infrastructure and to extend scientific functionality to some degree. This led to an update of the user interface to keep it compatible with the industry standard Jupyter framework, an update to the execution environment to support bulk operations, and the deployment of a software development kit that both internal and external developers can use to add functions to the system.

The plants team helped add tools for read management, bulk upload/download, and RNA-seq necessary for end-to-end gene expression profiling. In response to growing interest in RNA-seq and the challenges of large data sets, we expanded our initial RNA-seq pipeline further. FastQC and Trimmomatic apps were added for read quality control and refinement. The RNA-seq pipeline was used as a testbed for Bulk Operations, which included Bulk Execution (parallel implementation of the RNAseq pipeline), Bulk Input (uploading a large number of files and large files), and Narrative Big Data Handler (allowing users to drag and drop data directly into the Narrative from a staging area). The end goal of all of these improvements was to improve analysis speed, to facilitate upload of large files and large numbers of files, and to ensure ease of use in file manipulation. Bulk file upload and ease of use were in fact part of the key requests from both the users and the scientific advisory committee. Bulk execution lays some

of the basic groundwork that can be developed further into parallelizing other KBase workflows, as well as the early beginnings for HPC. A more effective, parallel, next-generation RNA-seq pipeline is in beta.

The KBase project also made major improvements to the usability and functionality of the metabolic modeling tool suite. The first is the introduction of numerous new tools in our metabolic modeling pipeline to expand functionality such as editing metabolic models and media formulations, comparison of flux profiles with gene expression profiles, and fitting model fluxes to expression data. The second major change is a set of significant enhancements to existing tools, including improved tools for building and gap-filling community metabolic models, improved tools for phenotype simulation to support reconciliation of models with phenotype data, and improved tools for model propagation to new genomes.

Extensive work was performed in 2016 to improve the reference data in KBase, as well as to advance our infrastructure for maintaining, updating, indexing, and querying reference data. This work focused on updating the KBase reference genomes to provide users with more data, adding taxonomy as a new type of reference data to help integrate species information within the system, and adding a service for maintaining, updating, indexing, and querying reference data, making it easier to stay up-to-date with new additions to the reference data.

Ontologies are an important mechanism in biology for maintaining controlled vocabularies of biologically meaningful concepts, phenomena, and functions. There are prevalent ontologies for functional annotations (e.g., GO, SEED), biochemistry (KEGG, EC), tissues (e.g., TO), and environments (e.g., EO). The project added support for maintaining and updating a database of important ontologies and offered users a mechanism for loading and applying their own ontologies. At this stage, we are primarily using ontologies as a systematic way to store consistent genome annotations made by our implementations of the RAST and InterProScan pipelines. Our ontology services also provide a means of mapping between these annotation ontologies.

KBase now offers a unique set of scientific functionality that allows users to go from reads to metabolic models of cells, analyze expression data, and perform some comparative genomics. The outlook for 2017 and beyond leverages these improvements by expanding and deepening science functionality aligned with the DOE's Biological and Environmental Research (BER) program objectives.

PUBLICATIONS

Adam-Blondon AF, Alaux M, Pommier C, Cantu D, Cheng ZM, Cramer GR, Davies C, Delrot S, Deluc L, Di Gaspero G, et al. 2016. Towards an open grapevine information system. *Hortic Res* **3:** 16056.

Gupta P, Naithani S, Tello-Ruiz MK, Chougule, K, D'Eustachio P, Fabregat A, Jiao Y, Keays M, Lee YK, Kumari S, et al. 2016. Gramene Database: Navigating plant comparative genomics resources. *Curr Plant Biol* **7-8:** 10–15.

He F, Yoo S, Wang D, Kumari S, Gerstein M, Ware D, Maslov S. 2016. Large-scale atlas of microarray data reveals the distinct expression landscape of different tissues in *Arabidopsis*. *Plant J* **86:** 472–480.

Jiao Y, Burke JJ, Chopra R, Burow GB, Chen J, Wang B, Hayes CM, Emendack Y, Ware D, Xin Z. 2016. A sorghum mutant resource as an efficient platform for gene discovery in grasses. *Plant Cell* **28:** 1551–1562.

Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. 2016. Ensembl Genomes 2016: More genomes, more complexity. *Nucleic Acids Res* **44**(D1)**:** D574–D580.

Majoros WH, Campbell MS, Holt C, DeNardo E, Ware D, Allen AS, Yandell M, Reddy TE. 2016. High-throughput interpretation of gene structure changes in human and nonhuman resequencing data, using ACE. *Bioinformatics* pii: btw799.

Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, Antin P. 2016. The iPlant collaborative: Cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biol* **14**: e1002342.

Naithani S, Preece J, D'Eustachio P, Gupta P, Amarasinghe V, Dharmawardhana PD, Wu G, Fabregat A, Elser JL, Weiser J, et al. 2016. Plant Reactome: A resource for plant pathways and comparative analysis. *Nucleic Acids Res* **45**(D1)**:** D1029–D1039.

Tello-Ruiz MK, Stein J, Wei S, Preece J, Olson A, Naithani S, Amarasinghe V, Dharmawardhana P, Jiao Y, Mulvaney J, et al. 2016. Gramene 2016: Comparative plant genomics and pathway resources. *Nucleic Acids Res* **44**(D1)**:** D1133–D1140.

Tello-Ruiz MK, Stein J, Wei S, Youens-Clark K, Jaiswal P, Ware D. 2016. Gramene: A resource for comparative analysis of plants genomes and pathways. In *Plant bioinformatics: Methods and protocols (methods in molecular biology)* (ed. Edwards D), Vol. 1374, pp. 141–163. Humana, New York.

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. 2016. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun* **7:** 11708.

Wang L, Van Buren P, Ware D. 2016. Architecting a distributed bioinformatics platform with iRODS and iPlant agave API. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 420–423.

QUANTITATIVE BIOLOGY

Gurinder S. "Mickey" Atwal and colleagues are applying insights from the physical and computational sciences to the study of population genetics and human disease. The Atwal lab has modeled the process by which genetic variants, or alleles, have evolved in the last 100,000 years of human history. This has recently led to surprising insights about the role of *p53*, a master tumor suppressor gene, in female fertility and furthered our understanding of how complex gene networks evolve. The lab has analyzed the comparative genomics and physical organization of cancer-related genes and their role in mediating tumorigenesis across numerous tissue types. Recently, they have begun to focus efforts on understanding cancer genome evolution on shorter timescales by analyzing nucleotide sequences from single cells.

Ivan Iossifov focuses on the development of new methods and tools for genomic sequence analysis and building and using molecular networks, and he applies them to specific biomedical problems. He studies the genetics of common diseases in humans using two main tools: next-generation sequencing and molecular networks representing functional relationships among genetic loci. These approaches, in combination, enable the kind of large-scale studies necessary for furthering our understanding of the complex etiology of disorders such as autism, bipolar disorder, and cancer.

Justin Kinney completed his Ph.D. in Physics at Princeton University in 2008 and began his term as a Quantitative Biology Fellow in 2010. His research focuses on developing next-generation DNA sequencing as a tool for dissecting the structure and function of large macromolecular complexes. Of particular interest to his lab is the biophysical basis of transcriptional regulation—how simple interactions between proteins and DNA allow promoters and enhancers to modulate genes in response to physiological signals.

In 2010, Kinney and colleagues published a paper pertaining to Sort-Seq, a novel sequencing-based method that can measure the functional activity of hundreds of thousands of slightly mutated versions of a specific DNA sequence of interest. Using a novel information-theoretic analysis of the resulting data, Kinney et al. were able to quantitatively measure, in living cells, the protein–DNA and protein–protein interactions controlling mRNA transcription at a chosen promoter. Kinney continues to develop this approach using a combination of theory, computation, and experiment. From a biological standpoint, Sort-Seq allows researchers to investigate important but previously inaccessible biological systems. Kinney's lab is currently using Sort-Seq to address open problems in transcriptional regulation, DNA replication, and immunology. These experiments also present new challenges for the field of machine learning, and a substantial fraction of Kinney's efforts are devoted to addressing the theoretical and computational problems relevant to the analysis of Sort-Seq data.

Alexander Krasnitz and colleagues use mathematical and statistical tools to discover key genetic elements involved in cancer and understand how cancer cells evolve. Array-based comparative genome hybridization, a technique honed in the Wigler lab—and, more recently, sequencing experiments—have revealed subtle patterns of frequent and widespread aberration in cancer genomes. Krasnitz hypothesizes that recurrent, aberrant genomic loci observed in a range of cancer types are under selection and therefore are enriched in important cancer genes. He has developed a novel, comprehensive methodology to discover such "cores" and has used it to analyze multiple genome data sets in breast, liver, ovarian, and prostate cancer. The results have been shared with cancer biology labs across CSHL, and they have been a key enabling agent of functional studies using mouse models and RNA interference. Krasnitz has begun to apply these novel statistical tools to

the latest generation of experimental data, which have characterized tumor samples down to the level of single cells. By interpreting single-cell genomes, he and colleagues seek to learn how specific tumors evolve and how cancer cells migrate to invade adjacent tissues and metastasize.

There is increasing evidence that rare and unique mutations have a significant role in the etiology of many diseases such as autism, congenital heart disease, and cancer. **Dan Levy**'s group develops algorithms to identify these mutations from large, high-throughput data sets comprising thousands of nuclear families. After earlier work with high-resolution comparative genomic hybridization (CGH) arrays, Levy's group now uses targeted sequence data. Levy has developed methods for identifying de novo mutations (i.e., those seen in a child but not in his or her parents) by simultaneously genotyping the entire family; the team is currently focused on building algorithms to detect copy number variants and multiscale genomic rearrangements. Although their copy number methods are based on "read" density, there are classes of mutations that require analysis at the level of the read. Thus, they are developing algorithms to identify insertions, deletions, inversions, transpositions, and other complex events. Other projects in the Levy lab include analysis of single-cell RNA, phylogenetic reconstruction from sparse data sets, and disentangling haplotypes from sperm and subgenomic sequence data.

Michael Schatz is a computational biologist and an expert at large-scale computational examination of DNA sequencing data, including the alignment, assembly, and analysis of next-generation sequencing reads. These methods have been used to reconstruct the genomes of previously unsequenced organisms, probe sequence variations, and explore a host of biological features across the tree of life. Recent improvements in sequencing technologies are challenging our capacity to store and analyze the huge volume of DNA sequence data being generated. Consequently, Schatz is particularly interested in capitalizing on the latest advances in distributed and parallel computing, especially cloud computing technologies, to advance the state of the art in bioinformatics and genomics. In a recent breakthrough, Schatz was able to create a hybrid software-based solution to eliminate errors in so-called third-generation sequencing. This makes it remarkably easier to compile, align, and analyze full-genome sequences.

Modern genomic technologies make it relatively easy to generate rich data sets describing genome sequences, RNA expression, chromatin states, and many other aspects of the storage, transmission, and expression of genetic information. For many problems in genetics today, the limiting step is no longer in data generation, but in integrating, interpreting, and understanding the available data. Addressing these challenges requires expertise both in the practical arts of data analysis and the theoretical underpinnings of statistics, computer science, genetics, and evolutionary biology.

Adam Siepel's group focuses on a diverse collection of research questions in this interdisciplinary area. Over the years, their research has touched on topics including the identification of recombinant strains of HIV, the discovery of new human genes, the characterization of conserved regulatory elements in mammalian genomes, and the estimation of the times in early human history when major population groups first diverged. A general theme in their work is the development of precise mathematical models for the complex processes by which genomes evolve over time and the use of these models, together with techniques from computer science and statistics, both to peer into the past and to address questions of practical importance for human health. Recently, they have increasingly concentrated on research at the interface of population genomics and phylogenetics, with a particular focus on humans and the great apes. Siepel's group also has an active research program in computational modeling and analysis of transcriptional regulation in mammals and *Drosophila*, in close collaboration with Prof. John Lis at Cornell University.

These investigators are all part of the **Simons Center for Quantitative Biology** (SCQB), CSHL's home for mathematical, computational, and theoretical research in biology. The central idea

behind the SCQB is to place researchers trained in mathematics, physics, computer science, and other quantitative fields on the front lines in biology, working shoulder to shoulder with experimentalists. In addition to collaborative work, SCQB researchers pursue independent research in algorithms, machine learning, statistical genetics, molecular evolution, and other areas. The ultimate goal of the Center is to promote the use of quantitative methods to enable groundbreaking research across a wide variety of biological domains, including human genetics, cancer, plant biology, and neuroscience. The SCQB was created in 2008 with a major donation from the Simons Foundation, as well as gifts from the Starr Foundation and Lavinia and Landon Clay. Adam Siepel was recruited to serve as Chair of the SCQB in 2014.

QUANTITATIVE BIOLOGY

G.S. Atwal R. Aboukhalil N. Sadagopan
J. Carter Y. Tanigawa
K. Grigaityte R. Utama
B. Harris

Fueled by data generated from recent technological developments in DNA sequencing, the Atwal lab primarily develops mathematical and computational methods to tackle problems in cancer genomics and machine learning. Recently, we have focused efforts on understanding the cellular heterogeneity and evolution of the tumor microenvironment and how we may leverage this knowledge to inform effective cancer immunotherapeutics in breast and pediatric cancers.

This year we have continued to press ahead with a number of cancer-related projects and initiated new studies on the single-cell genomics of the human adaptive immune system. We have been following up on the investigation of ectopic expression of germline genes, transcribing genetic loci exclusively expressed in testes/ovaries in various tumors. Previous work in our lab has established the pleiotropic and ancient role of the TP53 pathway in germline development. Preliminary results have also identified ectopic expression of germline piwi genes in samples of glioblastoma multiforme extracted from the Cancer Genome Atlas (TCGA). However, the landscape of ectopically expressed germline genes is unknown and their functional impact on cancer development remains elusive. We began an integrative study that sought to determine the ectopic expression of germline transcripts in glioblastoma multiforme and breast cancer, elucidating the predictive value of germline molecular programs as diagnostic markers and immunogenic therapeutic targets. Taha Merghoub and Jedd Wolchok at Memorial Sloan Kettering Cancer Center (MSKCC) continue to collaborate with us on this project.

Previous work in our lab in collaboration with colleagues at the Cancer Institute of New Jersey and Weill Cornell Medical College has established the association between single-nucleotide polymorphisms (SNPs) in the p53 tumor suppressor pathway and female infertility in mice and humans. This hypothesis was first generated through computational investigations of haplotypic diversity and positive selection in genes in the TP53 pathway, where we detected signatures of recent positive selection in SNPs present at high frequencies in various human populations. Selected alleles in SNPs in LIF, Mdm2, Mdm4, and Hausp genes, each of which lies in the p53 molecular network, were found to be enriched in women undergoing IVF treatment. Moreover, some of these SNPs have been shown to be associated with estrogen-driven cancer risk, highlighting the pleiotropic character of genetic variants in the p53 pathway. We have also sought association of SNPs in p63 and p73, and initial results are encouraging—although we await an increase in sample numbers before the results can be deemed to be statistically significant.

A major and recent research focus of our lab is the investigation of cancer evolution using sequencing data derived from single-cell genomics. Although the timescales vary by many orders of magnitude, the mathematical tools of population genetics, originally devised to model molecular evolution over millions of years, can be repurposed to understand single-cell evolution of tumors in the lifetime of an individual. In particular, we have been addressing the ubiquitous problem of how many cells and how much read coverage are needed before the inferred cell phylogeny accurately reflects the evolutionary history of the tumor. In addition, our lab has continued research in the use of information theory and other machine-learning tools in addressing the tsunami of data generated by next-generation sequencing.

Robert Aboukhalil graduated from the Watson School and took up a senior computational biologist position with Fluidigm. Kristina Grigaityte, with a background in experimental neuroscience, joined the lab as a member of the 2015 Watson School incoming graduate class. Narayanan Sadagopan, a first-year medical student from Hofstra University, joined the lab for the summer. Jason Carter, an M.D./Ph.D. candidate at Stony Brook University, rotated in the lab for the summer. Yosuke Tanigawa, a visiting

undergraduate researcher at Tokyo University, joined the lab for two months in early 2016 and has since accepted a position as a graduate student at Stanford University. Ben Harris, a URP student, continues to work on the summer research since returning to his home institution. Mickey Atwal won the 2016 Winship Herr Prize for Excellence in Teaching in the Watson School.

Ectopic Germline Gene Expression in Cancer

B. Harris [in collaboration with J. Wolchok, MSKCC]

We continued our research on an integrative pancancer study that sought to determine the ectopic expression of germline transcripts in tumors and elucidate the predictive value of germline molecular programs as diagnostic markers and therapeutic targets.

To determine the genome-wide landscape of restricted germline expression, we developed and calculated a gene expression exclusivity score for transcripts across approximately 1400 heterogeneous RNA-sequencing expression data sets from the GTeX consortium. Germline tissues were found to express a significantly large, but tightly knit, network of genes ($n = 1294$, q-value $= 10^{-2}$), which we inferred to be tissue-restricted, and the functionality of these genes is enriched in molecular pathway processes such as piRNA metabolism ($p = 10^{-6}$) and cell-cycle processes ($p = 10^{-36}$). After filtering genes with low expression (<2 TPM), we identified a set of approximately 250 high-confidence germline genes expressed exclusively in testes and ovaries. Analysis of 9000 samples from TCGA revealed that all tumors significantly ectopically expressed these genes, a concept we term "germness" of tumors. Ongoing efforts are focused on determining the peptides of these gene products, which may serve as immunogenic vaccines in mouse models of melanoma.

Ancient *trans*-Acting siRNAs Modulate Auxin Sensitivity in the Moss *Physcomitrella patens*

This work was done in collaboration with the Timmermans lab (CSHL).

Analysis of *P. patens* mutants perturbed in *trans*-acting small interfering RNA (tasiRNA) biogenesis reveals defects in the filamentous stage of moss development. These defects are the result of a spatial misregulation of the evolutionarily conserved targets of the tasiRNA pathway, the auxin response factors (ARFs). The similarity between these phenotypes and those of moss mutants defective in auxin signaling suggests that tasiRNAs may modulate plants' responses to this ancient hormone. Yevgeniy Plavskin modeled the effect of tasiRNA regulation of ARF levels on the output of the auxin response gene regulatory network in the context of network topology differences identified between *Arabidopsis* and *Physcomitrella*. The model predicts that tasiRNAs sensitize cells to auxin, and may have a complex effect on the robustness of the auxin response to noisy signals. This suggests that the robust, yet sensitive, auxin response that results from the structure of the auxin response network and its regulation by tasiRNAs may have favored its repeated cooption over the course of evolution.

Comprehensive Sequencing Analysis of High-Throughput Single T Cells in Humans

J. Carter, K. Grigaityte [in collaboration with Juno Therapeutics]

A diverse T-cell repertoire is a critical component of the adaptive immune system, providing protection against invading pathogens and tumors. T-cell receptors (TCRs)—the main signature of a T cell involved in antigen recognition—consist of a heterodimer of one α and one β chain. A diverse T-cell repertoire arises from numerous combinations of different α and β chains, potentially generating up to 10^{15} distinct TCRs. However, identifying αβ pairs and determining full TCR sequences in a high-throughput fashion is challenging. Recent studies of the T-cell repertoire have typically focused on characterizing either the α or β chain alone by bulk sequencing. Therefore, our understanding of statistical properties and selection of the T-cell pool in an individual is currently limited, hampering efforts to accurately quantify T-cell clonal changes in disease and immunotherapy.

This summer we began a study of three peripheral blood mononuclear cell (PBMC) samples, sorted by CD4 and CD8 T cells, and sequenced by our collaborators at Juno Therapeutics. Preliminary studies indicate a universal power-law distribution of the clone sizes among the three individuals, with significant

nonrandom associations between the α and β chains. These initial observations highlight the critical need to assay αβ pairing in single cells to accurately profile the landscape of the T-cell repertoire and monitor changes in the clonal distribution over time. Ongoing efforts are focused on acquiring more samples and investigating evidence of allelic inclusion in the human T cells.

Genomics of the Tumor Microenvironment in Breast Cancer

N. Sadagopan, R. Utama [in collaboration with City of Hope]

The goal of this project is to understand the role of stromal tissue genomic activity in mediating breast tumor microenvironment development and clinical outcome. This computational and mathematical study combined analyses of tissue samples from City of Hope and TCGA. The biological samples consisted of core biopsies obtained from normal, primary, and metastatic tumors from breast cancer patients enrolled in City of Hope. The tissue samples were cell-sorted using CD44 and epithelial cell adhesion molecule (EPCAM) markers for stromal and epithelial cells, respectively, and then cultured for up to two weeks, generating sufficient material for bulk RNA-poly(A) sequencing. In total, we obtained 35 stromal and epithelial library-prepared samples over the course of five months (July 2016–Dec. 2016), which were then sequenced at CSHL. The cDNA of these samples was sequenced using a NextSeq high-throughput Illumina machine generating approximately 60 million reads per sample. Initial expression analyses revealed more than 1000 genes differentially expressed between primary and metastatic stromal tissue, indicating a reprogramming of the expression activity of cancer-associated fibroblasts in concert with tumor development.

PUBLICATION

Plavskin Y, Nagashima A, Perroud P, Hasebe M, Quatrano RS, Atwal GS, Timmermans MCP. 2016. Ancient *trans*-acting siRNAs confer robustness and sensitivity onto the auxin response, *Dev Cell* **36:** 276–289.

GENETICS OF AUTISM

I. Iossifov S. Marks A. Munoz Jimenez

The bulk of our work in 2016 was in analyzing the large data set of whole-genome sequencing data generated from 510 families from the Simons Simplex Collection (SSC). These data are a rich resource that we use in numerous projects. In addition, we continue our analysis of the existing whole-exome and single-nucleotide polymorphism (SNP) genotyping data set available for the SSC and Autism Genetic Resource Exchange (AGRE) collection (a set of approximately 1000 multiplex families). Below is a list of the abstracts of four projects that are in submission or near submission status and that show our efforts in studying the role of de novo noncoding variants, rare structural rearrangements, and common variants in autism's etiology.

De Novo Indels within Introns Contribute to Autism Spectrum Disorder Incidence

I. Iossifov, A. Munoz Jimenez [in collaboration with B. Yamrom, Y.-H. Lee, P. Andrews, S. Marks, K.-T. Lin, Z. Wang, A. Krainer, M. Wigler, CSHL; B. Darnell, Rockefeller University, New York]

Whole-exome sequencing (WES) has allowed us, in the last 10 years, to make remarkable progress in our understanding of the genetics of autism. But there must be major genetic contributions to the disorder that are unexplained. Through whole-genome sequencing, there are additional types of genetic variants such as noncoding variants, small copy number variants, and structural rearrangements that might be observed. Because they are so abundant, analysis of which ones are functional is important but challenging. Here we analyzed whole-genome sequencing data from 510 of the SSC's quad families, and we focused our attention on intronic variants. In the introns of 546 high-quality autism target genes, we identified 63 de novo indels in the 510 affected children, whereas we found only 37 such events in their unaffected siblings. The difference of 26 events is significantly larger than expected (P-val = 0.01) and, using reasonable extrapolation, shows that de novo intronic events can contribute to as much as 20% of simplex autism. The significance increases if we restrict to the one-half of the autism targets that are intolerant to damaging variants in the normal human population and hence expected to be even more enriched for autism genes. For these 273 targets, the difference shrinks from 26 to 23 (P-val = 0.005). There was no significant difference in the number of de novo intronic substitutions in these genes, nor in de novo intronic indels or substitution in any of the control sets of genes analyzed. The excess of de novo intronic events was seen in introns that separate two coding exons and not in introns separating 5′ UTR exons.

Recurrent Variant Transmissions Contribute to Autism Spectrum Disorder

This work was done in collaboration with D. Levy, B. Yamron, and M. Wigler (CSHL); K. Ye (Albert Einstein, Bronx); and A. Buja and A. Krieger (University of Pennsylvania)

We develop a method of analysis (A2DS) that provides statistical evidence that transmission of shared variants contributes to a disorder. Using a standard measure of genetic relation, test individuals are compared with a cohort of discordant sib-pairs (CDS) to derive a comparative similarity score. We ask if the test individuals are more similar to the affected than to the unaffected siblings. Statistical significance is judged by randomly permuting the affected status in the CDS. In the analysis of published genotype data from the SSC and the AGRE cohorts of children affected with autism spectrum disorder (ASD), we find strong statistical significance that the affected are more similar to each other than to the unaffected (P-val < 0.0001). On the other hand, unaffected siblings, or parents, are not more similar to the unaffected siblings than they are to the affected siblings. These results do not depend on ethnic matching or gender.

MUMdex: Maximal Unique Match–Based Structural Variation Detection

I. Iossifov, S. Marks [in collaboration with P. Andrews, J. Kendall, Z. Wang, D. Levy, and M. Wigler, CSHL; L. Muthuswamy, New York Genome Center]

Standard genome sequence alignment tools primarily designed to find one alignment per read have difficulty detecting inversion, translocation, and large insertion and deletion events. Moreover, dedicated split read alignment methods that depend only on the reference genome may misidentify or find too many potential split read alignments because of flaws in the reference genome.

We introduce MUMdex, a maximal unique match (MUM)-based genomic analysis software package consisting of a sequence aligner to the reference genome, a storage-indexing format, and analysis software. Discordant reference alignments of MUMs are especially suitable for identifying inversion, translocation, and large indel differences in unique regions. Extracted population databases are used as filters for flaws in the reference genome. We describe the concepts underlying MUM-based analysis, the software implementation, and its usage.

We show via simulation that the MUMdex aligner and alignment format are able to correctly detect and record genomic events. We characterize alignment performance and output file sizes for human whole-genome data and compare with Bowtie 2 and the binary alignment map (BAM) format. Preliminary results show the practicality of the analysis approach by detecting de novo mutation candidates in human whole-genome DNA sequence data from 510 families. We provide a population database of events from these families for use by others.

A Platform for Access and Analysis of Genetic Variants in Phenotype-Rich Family Collections

S. Marks, I. Iossifov [in collaboration with Y.H. Lee and B. Yamron, CSHL; M. Cokol, Sabanci University, Istanbul; A. Nenkova, University of Pennsylvania; L. Chobradjiev, Seqpipe Ltd., Sofia]

WES, a technique that enables the inexpensive identification of genetic variants in the gene-encoding regions of the genomes of thousands of people, is quickly transforming human genetics. Particularly successful are the numerous studies that used WES in large collections of families to study the genetic architectures of human disorders with strong detrimental effect on fecundity, including autism, intellectual disability, schizophrenia, epilepsy, and congenital heart disease. These studies identified large numbers of genetic variants segregating in the families or arising de novo in children, gathered detailed phenotypic measurements of the studied individuals, and used the complex data sets to develop models of genotype and phenotype relationships.

There is an enormous amount of work that needs to follow the early success in the genetics of such complex disorders to develop effective treatment and early diagnostic strategies. A variety of future research projects will study in detail the effects of hundreds of genetic variants and genes at molecular, cellular, and organismic levels. Such projects will greatly benefit from the accumulated family WES data sets, but their large size and complex structure create a major obstacle for their efficient use. Here, we present the GPF (genotype and phenotype in families) system that manages such data sets and has an intuitive interface that makes it possible for the wider scientific community to benefit from the new collections.

PUBLICATIONS

Fang H, Bergmann EA, Arora K, Vacic V, Zody MC, Iossifov I, O'Rawe JA, Wu Y, Jimenez Barron LT, Rosenbaum J, et al. 2016. Indel variant analysis of short-read sequencing data with Scalpel. *Nat Protoc* **11:** 2529–2548.

Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, Raja A, Baker C, Hoekzema K, Stessman HA, et al. 2016. Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am J Hum Genet* **98:** 58–74.

SEQUENCE–FUNCTION RELATIONSHIPS AND OTHER QUANTITATIVE PROBLEMS IN MOLECULAR BIOLOGY

J.B. Kinney R. Adams A. Ayaz W.C. Chen K. Coshic T. Forcier A. Tareen

During my graduate and postdoctoral training, I developed an experimental method called Sort-Seq. Sort-Seq was the first massively parallel assay for dissecting the functional architecture of transcriptional regulatory sequences in living cells. A single Sort-Seq experiment can measure the transcriptional activities of ~10^5 different bacterial promoters. Through a quantitative analysis of the large data sets that Sort-Seq produces, it is possible to dissect the functional mechanisms operative at any promoter of interest.

Sort-Seq represents just one example of how ultra-high-throughput DNA sequencing can be used to develop precise quantitative models of complex biophysical/biochemical systems. To realize this vision, my laboratory is pursuing an integrated combination of theory, computation, and experiment. In our experimental work we continue to develop deep-sequencing-based assays for investigating biochemical systems. We are also developing computational tools for analyzing the data that these experiments produce. Our theory work is aimed at addressing problems in biophysics and machine learning that are highlighted by the availability of Sort-Seq and other massively parallel data sets.

Measuring the Sequence-Affinity Landscape of Antibodies with Massively Parallel Titration Curves

Despite the central role that antibodies play in the adaptive immune system and biotechnology, surprisingly little is known about how antibody-antigen-binding affinity depends on antibody sequence. For instance, we cannot predict the number of different antibody sequences that can bind an antigen of interest with a given affinity, how this density of states varies with binding affinity, and so on.

In collaboration with Aleksandra Walczak and Thierry Mora (ENS, Paris), my lab has developed a new experimental technique, called Tite-Seq, which is aimed at addressing these problems (Adams et al. 2016). Tite-Seq combines yeast display and fluorescence-activated cell sorting (FACS) in a way that allows one to measure the affinity of thousands of variant antibody proteins in a single experiment. In contrast to existing deep mutational scanning assays, which measure binding at only a single ligand concentration, Tite-Seq provides full binding titration curves. Having these titration curves eliminates the confounding factors present in deep mutational scanning assays, such as the sequence-dependence of antibody expression and stability.

Learning Quantitative Sequence-Function Relationships from Massively Parallel Data

Prior work of mine has shown that quantitative models of sequence–function relationships can be inferred from the data produced by Sort-Seq and other massively parallel assays. However, standard methods in statistical inference are not well suited for this modeling task. The reason is that standard inference methods rely on a quantity called "likelihood." To compute likelihood, one has to make strong assumptions about the quantitative form of experimental noise. The noise present in many massively parallel assays, however, is often difficult to precisely characterize.

Over the years I have published multiple papers showing how using a quantity called "mutual information" in place of likelihood can resolve this inference problem. Recently, Gurinder Atwal (CSHL) and I reviewed the theoretical underpinnings of this problem (Atwal and Kinney 2016). We further introduced the mathematical concept of "dual modes" in the space of noise models. Dual modes are closely related to the concept of "diffeomorphic modes" in parameter space, a concept that we introduced in a previous paper. To illustrate the difference between

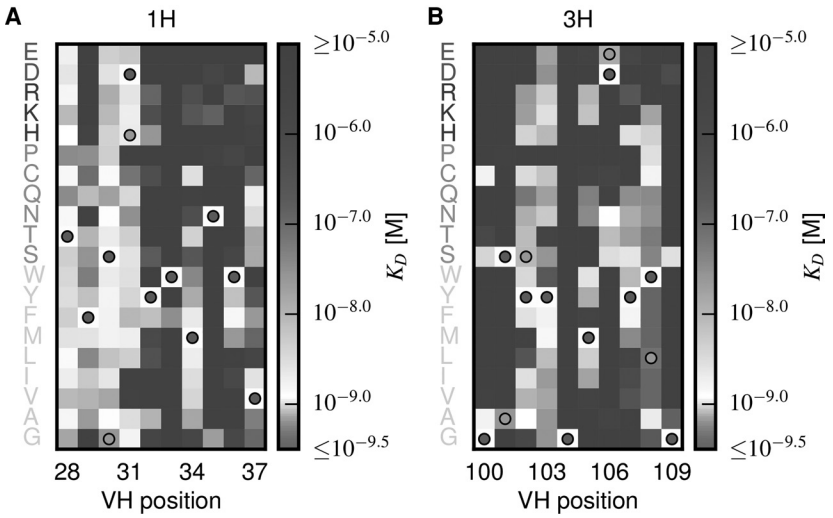


Figure 1. The sequence-affinity landscape of the CDR1H and CDR3H regions of the antifluorescein antibody studied in Adams et al. (2016). The affinity resulting from each possible single-residue mutation within these two 10-residue regions is shown. Dots indicate the wild-type residue at each position. Tite-Seq is the first deep mutational scanning assay to provide absolute affinity measurements such as these.

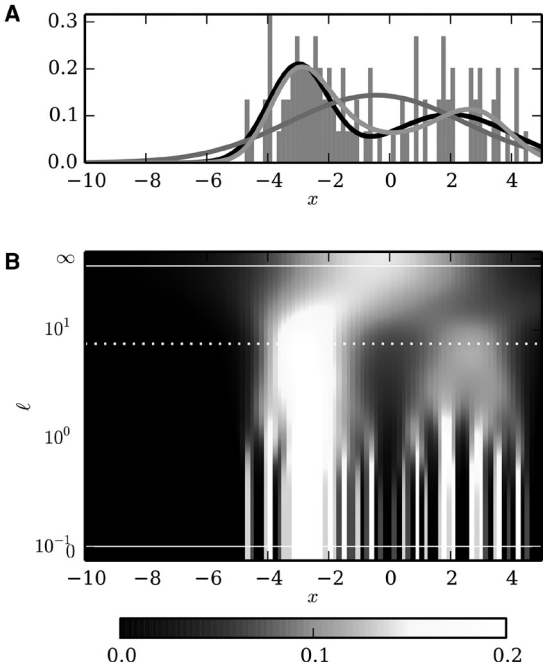


Figure 2. Density estimation using DEFT, as described in Kinney (*Phys Rev E* **92:** 032107 [2015]). (*A*) Data (gray) drawn from a "true" density (black) was simulated. Based only on these data, DEFT provides an estimate (orange) of the true underlying probability density. The corresponding MaxEnt estimate is shown in blue. (*B*) The DEFT estimate is part of a one-parameter family of density estimates, each parameterized by a different length scale *l*. Shown here are the density estimates across all length scales *l* from 0 to ∞. The dotted line indicates the optimal density estimate. The data histogram is recovered at $l = 0$, whereas the MaxEnt estimate is recovered at $l = \infty$.

mutual information- and likelihood-based inferences, we also described the first analytically tractable model (the Gaussian selection model) of inference from a massively parallel assay. In addition to their relevance to new biological experiments, these finding are applicable to a wide range of statistical regression problems.

However, performing mutual-information-based analysis of massively parallel data remains a major practical challenge. This is due to the lack of available software. To address this need, Bill Ireland and I have created a software package, called MPAthic, which is described in a preprint (Ireland and Kinney 2016). MPAthic provides simple command line methods that can be used to fit a variety of quantitative models to a variety of massively parallel data sets using a variety of different inference techniques.

Field Theoretic Methods for Estimating Low-Dimensional Probability Densities

Many data analysis efforts require the accurate estimation of continuous probability distributions (also known as probability densities). Despite the ubiquity of this task, the best way to estimate a continuous probability distribution from data remains unresolved, even in the case of one-dimensional data. The limitations of existing methods are especially acute when probability densities are estimated from small data sets.

Field theoretic methods from physics can potentially solve this problem. Although such methods for estimating probability densities date back to the mid-1990s, lingering computational difficulties and complications due to boundary conditions have stymied their adoption. In Kinney (*Phys Rev E 92:* 032107 [2015]), I showed that the boundary conditions that had previously been thought to be necessary when performing field-theoretic density estimation are, in fact, unnecessary. Removing these boundary conditions has an important consequence: It causes well-known maximum entropy (MaxEnt) estimates to be recovered as a limiting case. This finding suggests a new way of testing the hypothesis that one's data comes from a probability distribution having a specified functional form. This powerful density estimation approach is in the process of being deployed as a software package called Density Estimation Using Field Theory (DEFT).

Modeling Multiparticle Complexes in Stochastic Chemical Systems

My work on transcriptional regulation requires building quantitative biophysical models of molecular interactions. Currently, the only way to mathematically define models of such interactions is to explicitly list all the possible states of a system. This task, however, becomes exponentially more difficult as the number of molecular components increases. This exponential explosion in difficulty reflects a shortcoming of the mathematical tools that are currently used for studying such systems.

To address this problem, Muir Morrison (Caltech) and I developed a mathematical formalism that represents stochastic chemical systems of multiparticle complexes in terms of component particles, interaction energies, and assembly rules. This mathematical formalism also has a diagrammatic representation that greatly aids in its use. In Morrison and Kinney (2016), we present this formalism and show how it can dramatically simplify the description of both equilibrium and nonequilibrium stochastic chemical systems. Our hope is that these methods will serve a purpose much like Feynman diagrams do in physics, bridging the gap between one's intuitive visual understanding of a system and rigorous quantitative models thereof.

PUBLICATIONS

Adams RM, Mora T, Walczak AM, Kinney JB. 2016. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* **5:** e23156.

Atwal G, Kinney JB. 2016. Learning quantitative sequence-function relationships from massively parallel experiments. *J Stat Phys* **162:** 1203–1243.

Ireland WT, Kinney JB. 2016. MPAthic: Quantitative modeling of sequence-function relationships for massively parallel assays. *bioRxiv* doi:10.1101/054676.

Morrison MJ, Kinney JB. 2016. Modeling multi-particle complexes in stochastic chemical systems. *bioRxiv* doi:10.1101/045435.

Sheu Y-J, Kinney JB, Stillman B. 2016. Concerted activities of Mcm4, Sld3 and Dbf4 in control of origin activation and DNA replication fork progression. *Genome Res* **26:** 315–330.

In Press

Ipsaro JJ, Shen C, Eri A, Xu Y, Kinney JB, Joshua-Tor L, Vakoc CR, Shi J. 2017. Rapid generation of drug-resistance alleles at endogenous loci using CRISPR-Cas9 indel mutagenesis. *PLoS ONE* **12:** e0172177.

Senturk S, Shirole NH, Nowak DD, Corbo V, Pal D, Vaughan A, Tuveson DA, Trotman LC, Kepecs A, Kinney JB, et al. 2017. Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat Commun* **8:** 14370.

IN SILICO CANCER GENOMICS

A. Krasnitz J. Song V. Zhygulin

Research in our group is focused on in silico cancer genomics. In the last several years, there has been explosive growth in the volume as well as quality, variety, and detail of cancer-related genomic data available for analysis. This development came about through rapid accumulation of public data sets generated by individual laboratories worldwide, through large-scale cancer genome projects such as TCGA, and with the advent of new experimental methodologies, especially next-generation sequencing and single-cell genomics. We see our goal as channeling this flood of data into a number of clinically relevant applications. These include discovery of genomic markers for clinical outcome and molecular classification of cancer; elucidating the clonal structure of tumors and its relation to progression, invasion, metastasis, and response to treatment; and pinpointing and prioritizing targets for functional analysis. All our work is performed in close coordination with experimental studies performed by the Wigler, Tuveson, Spector, Stillman, and Fearon laboratories at CSHL.

Diagnostics of Prostate Cancer: Single-Cell Genomics Complements Conventional Histopathology

Recently, we used sparse single-nucleus sequencing and the computational pipeline (described in the following) to reconstruct the cell population structure in samples from 11 sets of prostate biopsies, all with pathological grades quantified as Gleason scores. The grades spanned a range of Gleason scores from benign to 9. Eight of the biopsies were performed diagnostically, with 10 to 15 biopsy cores per patient. Among these eight patients, five went on to undergo radical prostatectomy. Remarkably, for three of these five patients, pathological evaluation of the resected tissue differed from that of the diagnostic biopsy.

We set out to assess the potential of single-cell genomics in resolving such ambiguities. Approximately 30 single-cell genomes were acquired by sparse sequencing from each core and processed using our single-cell computational pipeline to yield integer genomic DNA copy number profiles for each cell. Genealogical trees formed by these genomes were derived for each patient, and clones were identified. There are several key observations from our analysis of tumor cell genealogies. The first is that clones are unlikely to form in nonmalignant prostates. In prostates in which Gleason score indicated malignancy, cells in higher-grade regions displayed a well-defined clonal identity and, within each clone, shared a substantial number of genomic features. The second observation is that these clones mostly are formed by cells that share an anatomic location. At the same time, we do occasionally find evidence for clones spreading beyond their anatomic origins. Third, copy number profiles of cells in clonal populations grow progressively more complex with the Gleason scores, and the populations themselves grow more numerous in the prostatic tissue.

We introduced a number of simple clone-related genomic metrics to quantify these observations. For example, we defined clonal heterogeneity as the number of clones observed in the prostate. Importantly, we observed that such genomic measures are better correlated with the postsurgery Gleason score than conventional histopathological metrics at diagnosis, including the diagnostic Gleason score itself. These findings point to the potential utility of single-cell genomic analysis to complement conventional pathology.

Computational Pipeline for Single-Cell Genomics

The study of genomic and transcriptional properties of individual cells is now the focus of multiple research laboratories around the world. In application to cancer, this line of research has revealed the genomic complexity of the disease and the presence of multiple genealogically related cell populations in a tumor. Detailed knowledge of the clonal structure of a cancer potentially is of high clinical value: Multiplicity of

clones or lesions in most advanced clones is a possible measure of progression; spatial pattern of clone dispersal in a tumor may signal elevated propensity to invade; and lesions observed in individual clones but not in the bulk tissue may point to targets for therapy. DNA copy number profiling of cells from sparse sequencing is an accurate, economically feasible technological approach to the study of cancer subpopulation structure. Novel multiplex sequencing techniques, developed by the Wigler lab at CSHL, permit simultaneous sequencing of hundreds of single-cell DNA specimens and their subsequent copy number profiling at up to 50-kb resolution. Optimal use of this data form for robust reconstruction of cancer cell phylogenies is a challenging computational problem calling for new and robust informatic and statistical tools.

We responded to this challenge by developing a computational pipeline for single-cell genomics. The pipeline comprises two major modules, one for deriving integer-valued copy-number profiles of individual cells and one for establishing genealogical relations among the cells in a sample and identification of clones. The input into the pipeline consists of cell-specific sets of sequencing reads. These are first aligned with the genome. The sequencing read density as a function of genomic position is then used to derive integer-valued DNA copy number profiles for each cell. Before any further processing, these profiles are examined for evidence of extensive DNA damage or degradation, and profiles derived from damaged DNA are filtered out. Next, each of the remaining profiles is reduced to a set of copy number change points, and, for each change point, the genomic interval of its likely location and the sign of the change are specified.

From this point on we combine change-point reduced copy number profiles of the cells comprising the sample, with the ultimate goal of reconstructing their phylogeny. To this end, we first derive a minimal set of features to account for all the change points observed in the sample and construct a table to indicate, for each cell genome, the presence or absence of each feature. Pairwise similarity of cell genomes is quantified based on the number of features shared by the pair. We then examine the resultant similarity matrix for the presence of clones—that is, groups of cells whose unusually high number of shared features can only be explained by their descent from a recent common ancestor. Such clones may, in turn, contain subclones of even more closely related cells.

To maximize portability of these tools and their utility to the clinical and research community, we have used a recently developed container technology. Docker containers, in essence, are lightweight virtual machines that are easily deployed on a host computer. All the multiple, heterogeneous software tools constituting the pipeline, with all the dependencies, are installed and configured within the container. The containerized pipeline then becomes an easily distributable plug-and-play application.

Single-cell data as interpreted by the pipeline must be presented to cancer biologists and clinicians in a comprehensible way in order to have an impact on clinical outcome. In practice, this means that the data should be viewable in an organized way, with the viewer retaining the ability to change the organization. The end user must be able to see and navigate the phylogenetic organization. Moreover, the data and their interpretation must be viewed in the context of other parameters, such as anatomical sites along with their pathological assessment. Such examination of the data in aggregate offers the best chance to reveal the critical properties of the sample relevant to clinical assessment. An integrated user interface, capable of handling specialized forms of data that arise in single-cell research, is therefore essential. With this necessity in mind, we have created the single-cell genome viewer (SCGV). The viewer displays multiple single-cell genomic profiles in the chromosomal order, with colors encoding the sign and magnitude of copy number aberrations. The cells on display are ordered as leaves on a tree, reflecting the structure of the population being sampled. In particular, clonal and subclonal identities of the cells are indicated in the appropriate tracks. Other tracks provide information on the cell ploidy, the quality of the cell genomic data, and, importantly, the anatomic origin of the cell. Elements of the data can be examined in greater detail by a combination of selection and zoom-in. For example, any number of cells can be selected and the corresponding copy number profiles plotted in a separate interface. A link to the University of California, Santa Cruz (UCSC) genome browser is available for any genomic region displayed by the viewer to help put the findings in a broader genomic context. Finally, the viewer, to be released to the public in the course of 2017, can display, for each cell, a histopathological image of its tissue of origin.

PanCanAtlas: Stemness and Immune Response across Multiple Tumor Types

The Cancer Genome Atlas (TCGA) is a massive, comprehensive data repository for cancer research. More than 30 cancer types are currently represented in the atlas, with hundreds of patient cases per type. For each tumor somatic mutation, DNA copy number variation, DNA methylation, and mRNA and microRNA profiles are available, along with clinical annotation for the patient. By combining multiple data sets, TCGA consortium generated integrated molecular portraits of a number of cancer types. PanCanAtlas is a successor project to TCGA, seeking to take data integration one step further and across multiple cancer types. Our group has been part of this effort since June 2015, with two focal areas of interest. The first of these is concerned with quantifying the presence of stem-cell-like cell populations in tumors. In particular, loss of differentiation has long been recognized as a key feature of cancer pathology. We hypothesize that, on a cellular level, this loss of differentiation is accompanied by acquisition of stem-cell-like properties. In an exploratory study, we find that genes associated with loss of differentiation in multiple cancers frequently occur in published transcriptional signatures of embryonic stem cells.

In addition, we are interested in pancancer characterization of immune response to tumors with a particular focus on the role of cancer testis antigens (CTA)—that is, products of genes that are exclusively expressed in testes of healthy male individuals, where they are protected by immune privilege from exposure to the immune system. Expression of these genes in tumors should, therefore, elicit an immune response—in particular by cytotoxic T cells. Such response is an end result of a multistep process, involving mRNA expression and translation into proteins, followed by proteasome cleavage, antigen presentation, and recognition by the adaptive immune system. Our ongoing study aims at taking these multiple steps into account to quantify potential immune response to CTA in tumors documented by TCGA. With this ultimate goal in mind, we have collaborated with Seven Bridges Genomics to determine human leukocyte antigen (HLA) class I A, B, and C alleles for nearly 9000 TCGA patients with available RNA sequencing data. These results will be used as input for multiple additional projects—most prominently, identification of neoantigens for all patients in this massive cohort.

PUBLICATION

Diermeier S, Chang K, Freier S, Song J, El Demerdash O, Krasnitz A, Rigo F, Bennett C, Spector D. 2016. Mammary tumor–associated RNAs impact tumor cell proliferation, invasion and migration. *Cell Rep* **17:** 261–274.

COMPUTATIONAL GENETICS

D. Levy M. Wroten

The data-rich environment at CSHL generates a wealth of opportunities for the application of mathematics and computation to further our understanding of biology and genetics. The primary activities of our lab are algorithm and protocol development, data analysis, and genetic theory, with a focus on sporadic human diseases such as autism, congenital heart disease, and cancer.

SMASH Sequencing

Measuring copy number is accomplished by either array hybridization or standard next-generation sequencing (NGS). Given the exponential improvements in NGS, long-term trends favor the adoption of sequencing technologies. However, much of the gain in throughput observed in NGS derives from longer read lengths. These longer reads improve mapping and coverage, but do not increase the number of independent maps in a sequencing library. Typically, 35–40 bases are sufficient to uniquely map a fragment. Reads that extend 500 base pairs have considerably more information; however, all those extra bases do little to improve copy number resolution.

Together with Zihua Wang, Peter Andrews, and Michael Wigler, we devised a method to pack multiple short fragments into a single sequencing read and then an informatics method that unpacks the short maps, converting the results into read density and copy number measurements. We call the method SMASH sequencing (Wang et al. 2016). To pack multiple maps into a single read, we first fragment the DNA by mechanical and/or enzymatic methods to a mean size of 40 bp, ligate the fragments together, and then adapt for sequencing. Computationally, the reads are analyzed for all the longest possible sequences that map uniquely to the reference genome. These are called maximal unique matches or MUMs. We optimized a heuristic for selecting MUMs from each read and, from that point, applied standard methods for converting map density into copy number profiles.

We tested the method in its two primary use cases: identifying copy number variations (CNVs) in a family and identifying CNVs in a cancer cell line. In both applications, SMASH performance was equal to standard sequencing approaches on a map-for-map basis. With SMASH, we obtain about four to six maps per read such that at present read lengths, SMASH offers a significant savings in sequencing costs for CNV.

Transcriptomics

We are developing methods to look at the transcriptional profiles of cells within small sections taken from a frozen breast cancer tumor. The goal is to catalog the distinct cell types, determine how they colocalize in the tumor environment, and infer interactions based on deviations from expectation based on the cellular composition of the neighborhood. To achieve this aim, we are developing methods for identifying cell types from nuclear RNA, building a dictionary of transcriptional profiles. We will then sequence RNA from small neighborhoods and regress against the dictionary of transcriptional types to infer a likely composition. We can then adjust our dictionary of profiles with new information inferred from the neighborhoods, which include both nuclear and cytoplasmic RNAs. For the neighborhood sequencing, we are experimenting with both random fragmentation of the tumor and spatial transcriptomics protocols, which, at present, provide multicell "pixels" that span a cross section of a sample. If our approach is successful, these mixed approaches can be used for other situations in which extracting single cells is prohibitive.

Sensitive Detection

Current Illumina sequencing instruments generate high-throughput, high-quality sequence data with error rates of approximately 1 per 100 base pairs read. For many applications, such as determining heterozygous positions in a genome, this is a sufficient level of

precision—1% will not introduce a significant error in a measurement of 50 per 100. However, there are applications for which this level of error is unacceptable: in particular, those cases in which we wish to measure a variant that occurs at a rate of <1 per 100. These include important applications like measuring residual disease in cancer, identifying low-frequency subpopulations in a tumor, and determining rates of somatic variation.

To counteract the errors introduced by sequencing and polymerase chain reaction (PCR), together with Michael Wigler and Zihua Wang, we developed a protocol that adds a unique sequence identifier (or "varietal tag") to the initial template molecule. Because sequence error is sparse and independent of the template molecule, reads with the same varietal tag are unlikely to have the same sequencer errors. This provides a degree of resolution up to PCR error. To counteract PCR error, we introduce rounds of linear amplification and an optional second tag. Because PCR errors are unlikely to occur on distinct lineages, the second tag can be used to ensure that the read data samples independent lineages. The residual differences, then, account for either true variation or template damage. We estimate that this method offers a resolution of between 1 in 10^5 and 1 in 10^6, dependent on the sequence context.

With Alex Krasnitz, Peter Andrews, and Mona Spector, we are applying this method to measure residual disease in patients treated for acute myeloid leukemia (AML). By identifying mutations that are existent at presentation, we design probes to assess the presence of tumor-specific variants after treatment.

Genomic Rearrangements

Together with Peter Andrews, Michael Wigler, and Ivan Iossifov, we have developed an algorithm called MUMdex for identifying large genomic rearrangements—deletions, insertions, translocations, and inversions—by cataloging discontinuities in the mapping of reads against the reference genome. Taking advantage of memory-intensive computational methods, we are able to rapidly identify all MUMs between a read and a reference genome. This provides a scaffold for cataloging and indexing all of the sequence data such that we can quickly identify discontinuities: places where putting the reads on the scaffold bends or breaks the reference genome.

We developed a calculus for recording and storing these discontinuities that satisfies algebraic properties, which reveal aspects of the genomic event. By looking across all samples, we can identify common events, spurious rearrangements, and global discrepancies within the reference. With all reads indexed and anchored to the reference genome, we reassemble rare and de novo discontinuities within an individual or family. Analyzing large data sets requires an efficient and robust data format and processing algorithms, which we are currently deploying on both the 10,000 SSC exomes and thousands of whole-genome sequences.

Mutational Sequencing

Despite vast improvements in DNA sequencing, many problems of interpretation arise when trying to count or assemble molecules (templates) that are largely identical. In a recent paper, we showed that by randomly mutating DNA templates before amplification, many of these problems are resolved. We can accurately count the number of templates by counting the number of unique patterns. By introducing distinctive patterns onto otherwise identical spans, we enhance our ability to correctly assemble sequences. This idea has applications in RNA expression analysis, haplotype phasing, copy number determination, and genome assembly. Mutational sequencing (muSeq) solves counting problems and effectively generates long reads from short-read sequence data.

We developed a bench protocol and a set of informatics methods for counting and assembling templates. We are working to apply these techniques to cDNA and regions of the genome that are difficult to sequence using short-read technology, such as the human leukocyte antigen (HLA) locus. We are presently adjusting the bench protocol to work on beads, improving efficiency and allowing for a small number of templates to be sampled for downstream assembly.

PUBLICATIONS

Andrews PA, Iossifov I, Kendall J, Marks S, Muthuswamy L, Wang Z, Levy D, Wigler M. 2016. MUMdex: MUM-based structural variation detection. *bioRxiv* doi:10.1101/078261.

Wang Z, Andrews P, Kendall J, Ma B, Hakker I, Rodgers L, Ronemus M, Wigler M, Levy D. 2016. SMASH, a fragmentation and sequencing method for genomic copy number analysis. *Genome Res* **26:** 844–851.

COMPUTATIONAL SEQUENCE ANALYSIS

M. Schatz M. Amin M. Nattestad
H. Fang S. Ramakrishnan
T. Garvin F. Sedlazeck

Our lab develops novel computational analysis methods to study the structure and function of genomes, especially genomes with medical or agricultural significance. These include methods for assembling sequence data into complete genomes, aligning sequences to discover variations or measure transcription levels, and mining these data to discover relationships between an organism's genome and its traits. We then apply these methods to study several biological systems—for example, to discover mutations associated with human diseases such as cancer and autism, or to reconstruct the genomes of important agricultural crops and biofuels.

In the past year, we published several significant methods for de novo genome assembly, variation detection, and related high-throughput sequencing assays. These include an open-source de novo genome assembly algorithm called FALCON for long PacBio single-molecule real-time sequencing reads. Beyond most other genome assemblers, FALCON is able to create a phased assembly of diploid genomes to separate out the maternal and paternal chromosomes, enabling the analysis of allele-specific genome structure for the first time. A related advance was the development of Assemblytics, which is an online analytics tool to detect structural variations in a de novo assembly. With it we are finding tens of thousands of variants that had been missed using other approaches. We also introduced a new aligner for high error rate Oxford Nanopore sequencing, and new methods for finding indel mutations from Illumina sequencing. We applied these and other methods to study the genomes of several species, including tomato (*Solanum lycopersicum*), petunia (*Petunia hybrida*), *Plasmodium falciparum* (a parasite that causes malaria), Asian sea bass (*Lates calcarifer*), and bedbug (*Cimex lectularius*). In addition, Tyler Garvin successfully defended his Ph.D. thesis in Watson School of Biological Sciences at CSHL on "Algorithms to Aid Cancer Genomics with Advanced Sequencing Technology." After graduation, he moved to a startup technology company in San Francisco.

Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing

Although genome assembly projects have been successful in many haploid and inbred species, the assembly of noninbred or rearranged heterozygous genomes remains a major challenge. To address this challenge, we introduce the open-source FALCON and FALCON-Unzip algorithms (https://github.com/PacificBiosciences/FALCON/) to assemble long-read sequencing data into highly accurate, contiguous, and correctly phased diploid genomes. We generate new reference sequences for heterozygous samples including an F1 hybrid of *Arabidopsis thaliana*, the widely cultivated *Vitis vinifera* cv. Cabernet Sauvignon, and the coral fungus *Clavicorona pyxidata*—samples that have challenged short-read assembly approaches. The FALCON-based assemblies are substantially more contiguous and complete than alternative short- or long-read approaches. The phased diploid assembly enabled the study of haplotype structure and heterozygosities between homologous chromosomes, including the identification of widespread heterozygous structural variation within coding sequences.

Assemblytics: A Web Analytics Tool for the Detection of Variants from an Assembly

Assemblytics is a web application for detecting and analyzing variants from a de novo genome assembly aligned to a reference genome. It incorporates a unique anchor filtering approach to increase robustness to repetitive elements and identifies six classes of variants based on their distinct alignment signatures. Assemblytics can be applied both to comparing aberrant genomes, such as human cancers, to a reference or to identify differences between related species. Multiple interactive visualizations enable in-depth explorations of the genomic distributions of variants.

NanoBLASTer: Fast Alignment and Characterization of Oxford Nanopore Single-Molecule Sequencing Reads

The quality of the Oxford Nanopore long DNA sequence reads has been, to date, lower than other technologies, leading to great interest in developing new algorithms that can make use of the data. So far, alignment methods, including LAST, BLAST, BWA-MEM, and GraphMap, have been used to analyze these sequences. However, each of these tools poses significant challenges for use with these data: LAST and BLAST require considerable processing time for high sensitivity, BWA-MEM has the smallest average alignment length, and GraphMap aligns many random strings with moderate accuracy. To address these challenges we developed a new read aligner called NanoBLASTer specifically designed for long Nanopore reads. In experiments resequencing the well-studied *Saccharomyces cerevisiae* (yeast) and *Escherichia coli* genomes, we show that our algorithm produces longer alignments with higher overall sensitivity than LAST, BLAST, and BWA-MEM. We also show that the run time of NanoBLASTer is faster than GraphMap, BLAST, and BWA-MEM.

Indel Variant Analysis of Short-Read Sequencing Data with Scalpel

As the second most common type of variation in the human genome, insertions and deletions (indels) have been linked to many diseases, but the discovery of indels of more than a few bases in size from short-read sequencing data remains challenging. Scalpel (http://scalpel.sourceforge.net) is an open-source software package for reliable indel detection based on the microassembly technique. It has been successfully used to discover mutations in novel candidate genes for autism, and it is extensively used in other large-scale studies of human diseases. This protocol gives an overview of the algorithm and describes how to use Scalpel to perform highly accurate indel calling from whole-genome and whole-exome sequencing data. We provide detailed instructions for an exemplary family-based de novo study, but we also characterize the other two supported modes of operation: single-sample and somatic analysis. Indel normalization, visualization, and annotation of the mutations are also illustrated. Using a standard server, indel discovery and characterization in the exonic regions of the example sequencing data can be completed in ~5 h after read mapping.

The Evolution of Inflorescence Diversity in the Nightshades and Heterochrony during Meristem Maturation

One of the most remarkable manifestations of plant evolution is the diversity for floral branching systems. These "inflorescences" arise from stem cell populations in shoot meristems, which mature gradually to reproductive states in response to environmental and endogenous signals. The morphology of the shoot meristem maturation process is conserved across distantly related plants, raising the question of how diverse inflorescence architectures arise from seemingly common maturation programs. In tomato and related nightshades (Solanaceae), inflorescences range from solitary flowers to highly branched structures bearing hundreds of flowers. Because reproductive barriers between even closely related Solanaceae have precluded a genetic dissection, we captured and compared meristem maturation transcriptomes from five domesticated and wild species reflecting the evolutionary continuum of inflorescence complexity. We find that these divergent species share hundreds of dynamically expressed genes, enriched for transcription factors. Meristem stages are defined by distinct molecular states and point to modified maturation schedules underlying architectural variation. These modified schedules are marked by a peak of transcriptome expression divergence during the reproductive transition, driven by heterochronic shifts of dynamic genes—including transcriptional regulators with known roles in flowering. Thus, evolutionary diversity in Solanaceae inflorescence complexity is determined by subtle modifications of transcriptional programs during a critical transitional window of meristem maturation, which we propose underlies similar cases of plant architectural variation. More broadly, our findings parallel the recently described transcriptome "inverse hourglass" model for animal embryogenesis, suggesting both plant and animal morphological variation are guided by a mid-development period of transcriptome divergence.

Insight into the Evolution of the Solanaceae from the Parental Genomes of *Petunia hybrida*

Petunia hybrida is a popular bedding plant that has a long history as a genetic model system. We report the whole-genome sequencing and assembly of inbred derivatives of its two wild parents, *Petunia axillaris* and *Petunia inflata*. The assemblies include 91.3% and 90.2% coverage of their diploid genomes (1.4 Gb; $2n = 14$) containing 32,928 and 36,697 protein-coding genes, respectively. The genomes reveal that the petunia lineage has experienced at least two rounds of hexaploidization: the older gamma event, which is shared with most eudicots, and a more recent Solanaceae event, which is shared with tomato and other solanaceous species. Transcription factors involved in the shift from bee to moth pollination reside in particularly dynamic regions of the genome, which may have been key to the remarkable diversity of floral color patterns and pollination systems. The high-quality genome sequences will enhance the value of petunia as a model system for research on unique biological phenomena such as small RNAs, symbiosis, self-incompatibility, and circadian rhythms.

The garden petunia, *P. hybrida*, with its diversity of color and morphology, is the world's most popular bedding plant (with an annual wholesale value exceeding US$130 million in the United States alone). Petunia has a long history as a model species for scientific research. To the scientific community, petunia is best known for the discovery of RNA interference (RNAi). This breakthrough was the culmination of decades-long research on the synthesis and regulation of the floral pigments and, as a consequence, anthocyanin biosynthesis remains one of the best-known pathways of secondary metabolism in any plant species. Development, transposon activity, genetic self-incompatibility, and interactions with microbes, herbivores, and pollinators have also been active research topics using petunia as the model system.

The genus *Petunia* is a member of the Solanaceae family native to South America. It forms a separate and early branching clade within the family with a base chromosome number of $x = 7$ rather than the typical $x = 12$ found for most Solanaceae crown-group species, including important crops such as tomato, potato, tobacco, pepper, and eggplant. The commercial *P. hybrida* is derived from crosses between a white-flowered, moth-pollinated *P. axillaris*, and species of the *Petunia integrifolia* clade, a group of closely related bee-pollinated species and subspecies. The first hybrids were produced by European horticulturalists in the early 19th century, probably multiple times from different accessions of the two parent clades. The remarkable phenotypic diversity in today's commercial garden petunias is the result of almost two centuries of intense commercial breeding. Here, we present the genome sequences of *P. axillaris* and *P. inflata*, two inbred laboratory accessions representing the parents of *P. hybrida*.

Chromosomal-Level Assembly of the Asian Sea Bass Genome Using Long Sequence Reads and Multilayered Scaffolding

We report here the ~670-Mb genome assembly of the Asian sea bass (*Lates calcarifer*), a tropical marine teleost. We used long-read sequencing augmented by transcriptomics and optical and genetic mapping, along with shared synteny from closely related fish species, to derive a chromosome-level assembly with a contig N50 size of >1 Mb and scaffold N50 size of >25 Mb, which spans ~90% of the genome. The population structure of *L. calcarifer* species complex was analyzed by resequencing 61 individuals representing various regions across the species' native range. Single-nucleotide polymorphism (SNP) analyses identified high levels of genetic diversity and confirmed earlier indications of a population stratification comprising three clades with signs of admixture apparent in the Southeast Asian population. The quality of the Asian sea bass genome assembly far exceeds that of any other fish species and will serve as a new standard for fish genomics.

Genome Assembly and Geospatial Phylogenomics of the Bedbug *Cimex lectularius*

The common bedbug (*Cimex lectularius*) has been a persistent pest of humans for thousands of years, yet the genetic basis of the bedbug's basic biology and adaptation to dense human environments is largely unknown. Here, we report the assembly, annotation, and phylogenetic mapping of the 697.9-Mb *C. lectularius* genome, with an N50 of 971 kb, using both long- and

short-read technologies. An RNA-Seq time course across all five developmental stages and male and female adults generated 36,985 coding and noncoding gene models. The most pronounced change in gene expression during the life cycle occurs after feeding on human blood and included genes from the Wolbachia endosymbiont, which shows a simultaneous and coordinated host/commensal response to hematophagous activity. These data provide a rich genetic resource for mapping activity and density of *C. lectularius* across human hosts and cities, which can help track, manage, and control bedbug infestations.

Complete Telomere-to-Telomere De Novo Assembly of the *Plasmodium falciparum* Genome through Long-Read (>11-kb), Single-Molecule Real-Time Sequencing

The application of next-generation sequencing to estimate genetic diversity of *Plasmodium falciparum*, the most lethal malaria parasite, has proved challenging because of the skewed AT-richness (~80.6% [A + T]) of its genome and the lack of technology to assemble highly polymorphic subtelomeric regions that contain clonally variant, multigene virulence families (Ex: var and rifin). To address this, we performed amplification-free, single-molecule real-time sequencing of *P. falciparum* genomic DNA and generated reads of average length (12 kb), with 50% of the reads between 15.5 and 50 kb in length. Next, using the Hierarchical Genome Assembly Process, we assembled the *P. falciparum* genome de novo and successfully compiled all 14 nuclear chromosomes telomere-to-telomere. We also accurately resolved centromeres (~90%–99% [A + T]) and subtelomeric regions and identified large insertions and duplications that add extra *var* and *rifin* genes to the genome, along with smaller structural variants such as homopolymer tract expansions. Overall, we show that amplification-free, long-read sequencing combined with de novo assembly overcomes major challenges inherent in studying the *P. falciparum* genome. Indeed, this technology may not only identify the polymorphic and repetitive subtelomeric sequences of parasite populations from endemic areas, but may also evaluate structural variation linked to virulence, drug resistance, and disease transmission.

PUBLICATIONS

Amin MR, Skiena S, Schatz MC. 2016. NanoBLASTer: Fast alignment and characterization of Oxford Nanopore single molecule sequencing reads. In *Proceedings of the Computational Advances in Bio and Medical Sciences (ICCABS) Conference*, Atlanta, GA.

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M, Boersma MR, Borghi L, Bruggmann R, et al. 2016. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat Plants* **2:** 16074.

Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. 2016. Phased diploid genome assembly with single molecule real-time sequencing. *Nat Methods* **13:** 1050–1054.

Fang H, Grabowska EA, Arora K, Vacic V, Zody MC, Iossifov I, O'Rawe JA, Wu Y, Jimenez-Barron LT, Rosenbaum J, et al. 2016. Indel variant analysis of short-read sequencing data with Scalpel. *Nat Protoc* **11:** 2529–2548.

Lemmon ZH, Park SJ, Jiang K, Van Eck J, Schatz MC, Lippman ZB. 2016. The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res* **26:** 1676–1686.

Nattestad M, Schatz MC. 2016. Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32:** 3021–3023.

Rosenfeld JA, Reeves D, Brugler MR, Narechania A, Simon S, Durrett R, Foox J, Shianna K, Schatz MC, Gandara J, et al. 2016. Genome assembly and geospatial phylogenomics of the bed bug *Cimex lectularius*. *Nat Commun* **7:** 10164.

Vembar SS, Seetin M, Lambert C, Nattestad M, Schatz MC, Baybayan P, Scherf A, Smith ML. 2016. Complete telomere-to-telomere de novo assembly of the *Plasmodium falciparum* genome through long-read (>11 kb), single molecule, real-time sequencing. *DNA Res* **23:** 339–351.

Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, Singh S, Thevasagayam NM, Prakki SR, Purushothaman K, et al. 2016. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLOS Genet* **12:** e1005954.

POPULATION GENETICS AND TRANSCRIPTIONAL REGULATION

A. Siepel A. Blumberg, N. Dukler, B. Gulko, Y. Huang, M. Hubisz, E. Hutton, L. Pipes, A. Platts, R. Ramani

For the past several years, our research has focused on two major areas: human population genetics and transcriptional regulation in humans and *Drosophila*. The research in population genetics is performed either with publicly available genomic sequence data or with a variety of collaborators, whereas most of the work on transcriptional regulation is performed with our collaborator John Lis at Cornell. We also have smaller collaborative projects on topics ranging from comparative transcriptomics of primates (in collaboration with Chris Mason, Weill Cornell Medical College), to molecular evolution of microRNAs in *Drosophila* (in collaboration with Eric Lai, Memorial Sloan Kettering Cancer Center [MSKCC]), to prediction of the fitness consequences of mutations in rice and other crops (in collaboration with Michael Purugganan, New York University [NYU]), to studying the speciation process of recently diverged *Sporophila* songbirds (in collaboration with John "Irby" Lovette, Cornell; Ilan Gronau, Herzliya Interdisciplinary Center, Israel), to combined experimental and computational characterization of *cis*-regulatory sequences in the human genome (in collaboration with Barak Cohen, Washington University, St. Louis). We focus on theoretical and computational research and do not generate our own data, but we often work closely with experimental collaborators on projects that have substantial experimental as well as computational components. We are broadly interested in molecular evolution, population genetics, and gene regulation, as well as machine learning, probabilistic modeling, and Bayesian statistics, and our research projects cut a broad swath across these diverse areas. The group is highly interdisciplinary, with members trained in computer science, mathematics, physics, genetics, and biochemistry, among other areas.

Since our move to CSHL from Cornell in September 2014, we have expanded considerably, with four new members joining the group in 2015. The group is close to capacity now, but we were recently joined by a new postdoctoral associate, Amit Blumberg, who came to CSHL from Ben Gurion University in Israel. We have continued to push forward with three core research projects that we have been working on for several years, as discussed below. In addition, we have launched several exciting new projects that are still in early stages of development, but which we hope to discuss in future reports.

Reconstruction of Demographic History from Complete Genome Sequences

Several years ago, we developed a statistical method based on the theoretical framework of the coalescent for reconstructing the demographic history of complex, structured populations from DNA sequence data. Our method, called *G-PhoCS* (for Generalized Phylogenetic Coalescent Sampler), uses Markov chain Monte Carlo techniques to explore coalescent genealogies consistent with a particular population phylogeny, allowing for gene flow between designated populations. *G-PhoCS* produces Bayesian estimates of the key parameters that define these population phylogenies, such as the divergence times between populations, effective sizes of ancestral populations, and rates of postdivergence gene flow. We originally used the method to estimate the date of origin of one of the earliest branching extant human populations, the Khoisan hunter–gatherers of Southern Africa. More recently, we have used it in collaboration with other research groups to shed light on the demographic histories of dogs and wild canids and birds from the genus *Sporophila* (e.g., Campagna et al., *Mol Ecol 24:* 4238 [2015]).

G-PhoCS and methods like it "cheat" by considering only short, widely spaced genomic sequences and ignoring the difficult problem of modeling recombination. However, it would be preferable to consider not only the process of coalescence (finding common ancestry) at each locus in a genome, but also the manner in which historical recombination events alter these

genealogies along the genome sequence. This combined history of coalescence and recombination can be explicitly represented by a generalized representation known as an "ancestral recombination graph," or ARG. The problem of reconstructing an ARG from sequence data, however, is notoriously difficult, and ARG inference has not been widely used in applied population genomics. Recently, we developed an algorithm for sampling ARGs within an approximate framework known as the sequentially Markov coalescent (SMC). Our method, called *ARGweaver*, uses techniques from hidden Markov models to repeatedly "thread" individual sequences through an ARG, leading to a Gibbs sampler over the space of ARGs. *ARGweaver* is the first ARG inference method that is efficient enough to apply to complete mammalian genomes. We have shown that it works remarkably well on simulated data and that it reveals clear signatures of natural selection in real human genome sequences.

This year, together with collaborators from the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, we published a major paper on human evolutionary history that made heavy use of both *G-PhoCS* and *ARGweaver* (Kuhlwilm et al. 2016). Using these new computational tools, we were able to perform the first joint analysis of multiple archaic human genomes and multiple modern genomes and propose a single unified demographic model that explains all of the patterns of genetic variation observed in these sequences. This model includes previously reported gene flow events between archaic and modern humans, including the now-famous event, dated to 47,000–65,000 years ago, that is responsible for the 1%–3% of Neanderthal DNA found in human populations outside of Africa. Interestingly, however, this model also includes a much older gene flow event in the opposite direction, explaining the presence of early modern human genome sequences in one of the Neanderthal genomes. This event dates to more than 100,000 years ago and appears to require an earlier out-of-Africa migration of modern humans than the migration event 50,000–60,000 years ago that is mainly responsible for the peopling of Europe and Asia. Therefore, our findings support not only an additional interbreeding event between archaic and modern humans, resulting in gene flow in the opposite direction of previously reported events, but they also support a scenario with multiple migrations by early modern humans out of Africa. To show that this scenario is the most plausible explanation for the observed patterns in the data, we performed three separate genome-wide analyses, one using *G-PhoCS*, one using *ARGweaver*, and a third based on patterns of heterozygosity and genetic divergence in large windows along the genome. Our paper describing this analysis generated a great deal of interest in the scientific community and was widely reported in the press.

Analysis of Natural Selection on Regulatory Sequences in the Human Genome

We have a long-standing interest in characterizing the influence of natural selection on DNA sequences, particularly in noncoding regions of the genome. Most of our work in this area has involved comparisons of complete mammalian genomes and, hence, has considered evolutionary processes spanning tens to hundreds of millions of years. More recently, however, we have become interested in integrating this phylogenetic information with data on human polymorphism to gain insight into more recent evolutionary events. A few years ago, we developed a probabilistic model and inference method, called *INSIGHT*, that makes use of joint patterns of divergence and polymorphism to shed light on recent natural selection. We have used *INSIGHT* to show that natural selection has profoundly influenced transcription factor binding sites across the genome during the past five million years of evolution, with major contributions both to adaptive changes in humans and to weakly deleterious variants currently segregating in human populations.

INSIGHT provides an estimate of the fraction of nucleotides under natural selection in any given collection of genomic elements. This same estimate can, alternatively, be interpreted as probabilities that mutations falling in the given elements will have fitness consequences for the organisms that carry them. We recently realized that this property could be used to produce "fitness consequences" (fitCons) scores across the entire human genome. Using high-throughput data from the ENCODE project, we first partition the genome into classes of sites having characteristic functional genomic "fingerprints" in a given cell type. We then use *INSIGHT* to calculate a fitCons score for each fingerprint (Gulko et al., *Nat Genet 47:* 276 [2015]). Finally, we plot these scores along the genome

sequence. These fitCons scores turn out to be remarkably powerful for identifying unannotated regulatory elements in the human genome. We have also used fitCons scores to estimate that only ~7% of nucleotides in the human genome directly influence fitness.

In our latest effort in this area, we have addressed a major shortcoming of fitCons—namely, that it does not scale up for use with large numbers of functional genomic covariates. We developed an alternative approach that bypasses the need for clustering genomic sites and, instead, assumes a linear-logistic relationship between covariates along the genome and parameters of the *INSIGHT* model (Huang et al., in press). This method, called *LINSIGHT*, is extremely fast and scalable, enabling it to exploit the "Big Data" available in modern genomics. We have shown that *LINSIGHT* outperforms the best available methods in identifying human noncoding variants associated with inherited diseases (including fitCons). In addition, we have applied *LINSIGHT* to an atlas of human enhancers and shown that the fitness consequences of enhancers depend on cell type, tissue specificity, and constraints at associated promoters.

Work is under way on a deep-learning extension of *LINSIGHT*, called *DeepINSIGHT*, which both allows for nonlinearities and predicts allele-specific selection coefficients across the genome. In addition to their usefulness in human disease, these new methods are of interest to plant biologists in their efforts to guide breeding designs for crops to improve yield, drought tolerance, and other desirable traits. We have collaborations with Michael Purugganan (NYU) and Ed Buckler (Cornell University) to explore applications in plant breeding.

Characterization of Transcriptional Regulation Using GRO-seq

For several years, we have been working closely with John Lis's group on methods for interpreting data generated using their powerful GRO-seq (global run-on and sequencing) technology, which maps the positions of engaged RNA polymerases across the genome. It has gradually become clear that an unanticipated benefit of GRO-seq and derived technologies is that they are uniquely well suited for detecting so-called enhancer RNAs (or eRNAs) and, consequently, for identifying active enhancers and other regulatory elements in mammalian cells.

In our latest work in this area (Danko et al. 2016), we made use of PRO-seq to perform the first comparative study of nascent transcription in primates. We mapped actively transcribing RNA polymerases in resting and activated $CD4^+$ T cells in multiple human, chimpanzee, and rhesus macaque individuals, with rodents as outgroups. This approach allowed us to directly measure active transcription separately from posttranscriptional processes. We observed general conservation in coding and noncoding transcription, punctuated by numerous differences between species, particularly at distal enhancers and noncoding RNAs. Among other findings, we observed that transcription factor binding sites are a primary determinant of transcriptional differences between species, and rates of evolutionary change are strongly correlated with long-range chromatin interactions. This evolutionary work provides new insights into the ways in which eRNAs are gained and lost during evolutionary time.

PUBLICATIONS

Danko CG, Wang Z, Rice EJ, Chu T, Martins A L, Wojno ET, Lis JT, Kraus WL, Siepel A. 2016. Natural selection has shaped coding and non-coding transcription in primate $CD4^+$ T-cells. *bioRxiv* doi: 10.1101/083212.

Dukler N, Gulko B, Huang Y-F, Siepel A. 2016. Is a super-enhancer greater than the sum of its parts? *Nat Genet* **49:** 2–3.

Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, Gronau I, Silva PM, Galaverni M, Fan Z, Marx P, et al. 2016. Demographically-based evaluation of genomic regions under selection in domestic dogs. *PLoS Genet* **12:** e1005851.

Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, Fu Q, Burbano HA, Lalueza-Fox C, de la Rasilla M, et al. 2016. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530:** 429–433.

In Press

Huang Y-F, Gulko B, Siepel A. 2017. Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. *Nat Genet* **49:** 618–624.

COLD SPRING HARBOR LABORATORY FELLOWS

In 1986, Cold Spring Harbor Laboratory began a Fellows Program to encourage independent research by outstanding young scientists who, during their graduate studies, displayed exceptional promise of becoming leading scientists of the future. The purpose of this program is to provide an opportunity for these young scientists to work independently at the Laboratory for a period of 3 years on projects of their own choosing. Fellows are provided with a salary, research support, and technical assistance so that they can accomplish their goals free of distraction. The interactions among research groups at the Laboratory and the program of courses and meetings on diverse topics in biology contribute to a research environment that is ideal for scientific innovation by these Fellows. The CSHL Fellows Program has been tremendously successful and has served as a paradigm for several analogous programs at other institutions, most recently a Fellows Program sponsored by the National Institutes of Health.

The success of the program is apparent from the list of distinguished alumni. Most notably, Carol Greider—recipient of the 2009 Nobel Prize in Physiology or Medicine for her work on telomerase and telomere function—joined the Fellows Program in 1998. After completing her fellowship, Carol was on the CSHL faculty for 9 years, and she is currently the Daniel Nathans Professor and Director of Molecular Biology and Genetics at Johns Hopkins University School of Medicine.

The first CSHL Fellow, Adrian Krainer (1986), is currently a Professor at the Laboratory, and Chris Vakoc (2008) and Florin Albeanu (2008) are currently holding Associate Professor positions at CSHL. Scott Lowe (1995) is a Howard Hughes Medical Institute (HHMI) Investigator. After nearly 15 years on the CSHL faculty, he recently took on a Professorship at Memorial Sloan Kettering Cancer Center in New York City. Marja Timmermans (1998) was a member of the CSHL faculty for more than 17 years and recently accepted the Humboldt Professorship at the University of Tubingen. Eric Richards (1989) currently is the Vice President of Research and Senior Scientist at the Boyce Thompson Institute for Plant Research at Cornell University; David Barford (1991) is a Fellow of the Royal Society and Professor of Molecular Biology at the Institute of Cancer Research in London; Ueli Grossniklaus (1994) is Professor at the Institute of Plant Biology, University of Zürich, Switzerland; and Térence Strick (2000) left at the end of his fellowship to become a Group Leader at the Institut Jacques Monod in Paris. Lee Henry (2000) joined HHMI's Janelia Farm in Ashburn and joined a project headed by Thomas Südhof. Ira Hall (2004) is an Associate Professor and Associate Director of the Genome Institute at Washington University. Patrick Paddison, who had joined the Fellows Program in 2004, currently is an Associate Member at the Fred Hutchinson Cancer Research Center in Seattle, Washington.

Lingbo Zhang has been a Fellow at the Laboratory since 2013. He joined us from Harvey Lodish's laboratory at the Whitehead Institute of the Massachusetts Institute of Technology, where he studied the regulation of erythroid progenitor cell self-renewal. As a CSHL Fellow, Lingbo is conducting genetic and small-molecule screens to discover novel regulators of normal and aberrant stem cell biology. **Jason Sheltzer** has been a CSHL Fellow since 2015 after completing his graduate work in Angelika Amon's laboratory at MIT. His research focuses on studies of aneuploidy and how it impacts cancer progression.

ANEUPLOIDY AND GENE DOSAGE IMBALANCES IN CANCER

J. Sheltzer C. Giuliano N. Sayles
A. Lin A. Vasudevan

The Sheltzer Lab was founded in August 2015. We are working to combine in silico, in vitro, and in vivo approaches to address several questions in cancer genetics. In particular, we are deeply interested in the consequences of genomic imbalances on tumor development and progression. Genomic imbalances are found in nearly every cancer cell and can alter the dosage of hundreds of genes at once. Approximately 90% of solid tumors display whole-chromosome aneuploidy, whereas many tumors with diploid karyotypes nonetheless harbor segmental or arm-length aneuploidies that also result in significant gene copy number alterations. The prevalence of aneuploidy in cancer—first noted more than 100 years ago—has led to a widespread belief that genomic imbalances play a crucial role in tumor development. Indeed, in the early 20th century, Theodor Boveri speculated that abnormal karyotypes altered the balance between pro- and antiproliferative cellular signals and were therefore sufficient to induce transformation. "Boveri's hypothesis" has motivated nearly a century of research into the origins and consequences of aneuploidy, but the precise relationship between abnormal karyotypes and tumorigenesis remains unclear.

Single-Chromosome Aneuploidy Commonly Functions as a Tumor Suppressor

To test the link between aneuploidy and cancer, we transduced congenic euploid and trisomic fibroblasts with 14 different oncogenes or oncogene combinations, thereby creating matched cancer cell lines that differ only in karyotype. Surprisingly, nearly all aneuploid cell lines divided slowly in vitro, formed few colonies in soft agar, and grew poorly as xenografts, relative to matched euploid lines. Similar results were obtained when comparing a near-diploid human colorectal cancer cell line with derivatives of that line that harbored extra chromosomes. Only a few aneuploid lines grew at close to wild-type levels, and no aneuploid line showed greater tumorigenic capabilities than its euploid counterpart. These results show that rather than promoting tumorigenesis, aneuploidy can very often function as a tumor suppressor. Moreover, our results suggest one potential way that cancers can overcome the tumor-suppressive effects of aneuploidy: On prolonged culture in vitro or in vivo, cell lines with simple aneuploidies developed recurrent chromosomal aberrations that were absent from their euploid counterparts and were associated with enhanced growth. Thus, the genome-destabilizing effects of single-chromosome aneuploidy may facilitate the evolution of balanced, high-complexity karyotypes that are frequently found in advanced malignancies.

Mitotic Kinases in Breast Cancer

Chromosome segregation is controlled by the concerted action of a set of protein kinases that phosphorylate key components of the mitotic apparatus. These kinases tend to be overexpressed in cancer, and several small molecules that inhibit mitotic kinases have entered clinical trials for diverse malignancies. We have applied CRISPR-Cas9 mutagenesis to interrogate the functional consequences of losing these kinases. Our initial efforts have focused on the maternal embryonic leucine zipper kinase (MELK). MELK has been reported to be a genetic dependency in several cancer types, as small-molecule inhibitors of MELK and RNA interference (RNAi) against MELK block cancer cell proliferation. On the basis of these preclinical results, a MELK inhibitor is currently undergoing clinical trials in breast cancer. Surprisingly, we found that mutagenizing MELK with CRISPR-Cas9 had no effect on the fitness of cell lines from seven different cancer types. Cells that harbored null mutations in MELK showed wild-type doubling times, cell-cycle progression, and anchorage-independent growth. Furthermore, MELK-knockout lines remained sensitive to the MELK inhibitor in clinical trials, suggesting that this drug blocks cell division through an off-target mechanism. In total, our results undermined

the rationale for a series of current clinical trials and provide an experimental approach for the use of CRISPR-Cas9 in preclinical target validation that can be broadly applied.

Discovery and Characterization of Genes Affecting Survival Time in Cancer

In addition to our investigation of aneuploidy and chromosome segregation, we are working to uncover the molecular differences between benign and aggressive tumors. In human patients, certain tumors can be cured by surgery alone. Other tumors derived from the same tissue and classified at the same pathological stage will inevitably recur after surgery and often prove to be fatal. For instance, stage II colorectal cancer is typically treated by surgical resection with curative intent. However, ~20% of these tumors will recur following surgery, and patients with recurrent disease have a 5-year survival rate of only 30%. Although adjuvant chemotherapy can lower the risk of colorectal cancer relapse, the difficulty in identifying patients who would benefit from additional treatment, coupled with its debilitating side effects, has limited its use. In general, many clinical decisions are constrained by the poor understanding of the molecular features that differentiate potentially fatal and nonfatal human tumors. A greater understanding of the genes and biological pathways that drive tumor aggressiveness would improve patient risk stratification and allow for more accurate treatment decisions.

To uncover new cancer genes and increase our understanding of the molecular differences between fatal and nonfatal tumors, we are analyzing data derived from cancer survival studies. In these studies, investigators profile tissue from tumors surgically excised from patients and then link that data with clinical information on the patient's tumor, treatment regimen, and the length of time until death and/or the occurrence of another adverse event. Tumors included in these studies are typically matched for stage, grade, and treatment, and only pure samples that yield high-quality RNA (or other biological molecules) are analyzed. Genes whose expression consistently correlates with increased or decreased survival time in cancer may represent novel oncogenes, tumor suppressors, or mediators of the metastatic state. We are, therefore, deploying CRISPR technology to modulate the expression of these genes in primary cells and cancer cells to establish molecular links between these genes and cancer prognosis.

In Press

Lin A, Giuliano CJ, Sayles NM, Sheltzer JM. 2017. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. *eLife* **6:** e24179.

Sheltzer JM, Ko JC, Replogle JM, Habibe Burgos NC, Chung ES, Meehi CM, Sayles NM, Passerini V, Storchova Z, Amon A. 2017. Single-chromosome gains commonly function as tumor suppressors. *Cancer Cell* **31:** 1–16.

TARGETING SELF-RENEWAL PATHWAYS AND METABOLIC VULNERABILITIES IN NORMAL AND MALIGNANT HEMATOPOIETIC STEM AND PROGENITOR CELLS

L. Zhang W. Mei E. Wong
G. Trivedi L. Zhu
D. Tsang Hu

Self-renewal is a unique property of stem and progenitor cells, allowing them to undergo multiple cycles of cell division while maintaining an undifferentiated status. During normal development, self-renewal allows stem and progenitor cells to maintain tissue homeostasis. In cancer, malignant cells hijack self-renewal to support their abnormal proliferation.

Our research group is investigating molecular mechanisms underlying self-renewal of both normal and malignant stem and progenitor cells. We are probing the mechanistic divergence of self-renewal in normal versus malignant stem and progenitor cells to develop targeted therapy that selectively disrupts carcinogenesis, but not normal tissue homeostasis. As abnormal metabolism is emerging as a hallmark of cancer, we have a specific focus on how metabolic and energetic changes are involved in the regulation of self-renewal, with the ultimate goal to uncover metabolic vulnerabilities for cancer treatment. Our research focuses on the hematopoietic system, and we are currently studying two important cell types, the early erythroid progenitor and the leukemia stem cell. Utilizing both CRISPR-Cas9 functional genomics and forward chemical genomics approaches, we are developing small chemical compound–based novel therapeutic strategies for hematopoietic malignancies, including myelodysplastic syndrome (MDS) and acute leukemia.

Targeting Early Erythroid Progenitor Self-Renewal as a Cure for Refractory Myelodysplastic Syndrome

Myelodysplastic syndrome (MDS) is a form of lethal hematopoietic malignancy characterized by pancytopenia resulting from progressive bone marrow failure. The treatment options for MDS are limited. Although a small portion of MDS patients respond to erythropoietin (EPO) and lenalidomide, the only therapeutic option for the majority of MDS patients who do not respond to or who lose their initial response to EPO and lenalidomide is frequent blood transfusion, which causes iron overload toxicity and organ failure. We therefore focus our research on the highly self-renewing early erythroid progenitor population, whose failure causes EPO unresponsiveness, and we aim to harness its expansion and regeneration potential to treat refractory MDS.

We combined forward chemical genomic screening and in silico gene expression analysis, and identified multiple novel regulators of early erythroid progenitor self-renewal. One promising novel regulator is CHRM4, the muscarinic receptor for neurotransmitter acetylcholine of the parasympathetic nervous system. Inhibiting CHRM4 genetically or pharmacologically by nanomolar concentration of CHRM4-specific antagonist triggered expansion of early erythroid progenitors. We then tested the in vivo efficacy of CHRM4 antagonist in a preclinical conditional knock-in *Cre-Mx1 Srsf2 P95H/WT* MDS genetically engineered mouse model, an MDS disease model that carries the same mutation as ~15% of MDS patients and faithfully recapitulates all the essential pathological phenotypes of MDS. We demonstrated that injection of CHRM4 antagonist completely corrected MDS with sustainable long-term efficacy and extended the survival of MDS disease mice to the same level as control wild-type mice. In addition, we showed that injection of CHRM4 antagonist corrected anemia of aging. These results supported the view that CHRM4 antagonist is not just a treatment but a cure of MDS.

Mechanistically, we discovered that CREB, a downstream effector of muscarinic acetylcholine receptor signaling, preferentially binds to and triggers the upregulation of genes highly expressed in early

erythroid progenitor that are important for the maintenance of progenitor status. These genes include transcription factor *Gata2*, the haploinsufficiency of which causes MDS, and *Zfp36l2*, an RNA-binding protein that we recently identified as a crucial molecular switch for early erythroid progenitor self-renewal. Consistent with the recently proposed role of CREB in mediating young blood-induced neurogenesis, our work further established CREB as a central regulator of adult stem cell self-renewal and a "youth factor" to correct stem cell exhaustion in the aging population. Our work established an unexpected connection between muscarinic acetylcholine receptor signaling and self-renewal of the early erythroid progenitor and led us to propose a novel model of neural control of hematopoietic stem and progenitor self-renewal through a cholinergic-mediated "hematopoietic reflex." We have established a start-up company, ZhanGen Therapeutics, to translate this discovery into a therapeutic agent for refractory MDS via clinical trials conducted at Northwell Health System.

Targeting Metabolic Divergences between Normal and Malignant Hematopoietic Stem and Progenitor Cells as Novel Therapeutic Strategies for Acute Leukemia

Cancer cells undergo extensive reprogramming to fulfill metabolic and energetic requirements that support their abnormal proliferation. However, the lack of understanding of molecular details of these requirements precludes the development of novel therapies for cancer through targeting these metabolic changes. More importantly, normal stem and progenitor cells in adult tissues, such as hematopoietic stem and progenitor cells (HSPCs), also undergo regular expansion proliferation. Thus, the metabolic requirements of these cells are similar to those of cancer cells—including the extensive formation of biomacromolecules to generate building blocks for creating daughter cells. Consequently, one major side effect of most chemotherapies is the depletion of adult stem and progenitor cells. This side effect on adult stem and progenitor cells is even more severe for the hematopoietic system, because the system requires HSPCs to undergo constant expansion and proliferation to produce mature blood cells.

To overcome this severe side effect, our laboratory is interested in identifying metabolic divergences between normal HSPCs and malignant leukemic cells. Through analyzing gene expression profiles from $CD34^+$ HSPCs and leukemic cells, we found that global metabolic gene expression signature is capable of discriminating these two cell types. Importantly, the metabolic difference among different subtypes of acute myeloid leukemia cells carrying different genetic abnormalities is significantly smaller than the difference between them and HSPCs. This suggests that there are common metabolic vulnerabilities shared by multiple subtypes of acute leukemia, which potentially serve as a drug target to block their proliferation, while exhibiting no or minimal side effect on HSPCs. We identified approximately 300 metabolic genes that follow this pattern of high expression across multiple subtypes of leukemia in comparison to HSPCs.

To functionally test whether these leukemic cell–specific genes are functionally essential for leukemic cell proliferation, we performed a CRISPR-Cas functional genomic screening. A sgRNA library was designed to target these approximately 300 highly expressed genes, among which approximately 20 genes have been identified as positive hits from the screening. We then utilized both individual sgRNA and shRNA approaches to validate these positive hits in both leukemic cell culture systems and genetically engineered AML mouse models as novel drug targets to treat leukemia. Five positive hit genes involved in vitamin, amino acid, and fatty acid metabolism were validated. These genes, or the downstream and upstream components in the same pathway, are with the available tool compounds as inhibitors. We are currently validating therapeutic efficacies of these lead compounds in treating acute leukemia in genetically engineered AML mouse models in vivo. In conclusion, we have identified multiple genes involved in vitamin, amino acid, and fatty acid metabolism as malignant cell–specific metabolic vulnerabilities. These genes function as metabolic divergences to discriminate normal and malignant HSPCs, and molecular pathways related to these genes represent metabolic definitions of malignant status.

In Press

Trivedi G, Chung YR, Abdel-Wahab O, Zhang L. 2017. Muscarinic acetylcholine receptor regulates self-renewal of early erythroid BFU-E progenitor. *Nature Med* (in press).

AUTHOR INDEX

WATSON SCHOOL OF BIOLOGICAL SCIENCES

ADMINISTRATION

Alexander Gann, Ph.D., WSBS Professor and Dean
Carrie Cowan, Ph.D., Associate Dean (until 5/1/16)
Alyson Kass-Eisler, Ph.D., Postdoctoral Program Officer and Curriculum Director
Kimberly Creteur, M.Ed., M.S.Ed., Admissions and Recruitment Manager
Kimberley Geer, Administrative Assistant

EXECUTIVE COMMITTEE

Chair
Alexander Gann

Members
Mikala Egeblad
Alexander Krasnitz
Bo Li
David L. Spector (Director of Research)
David Stewart
Anthony Zador

Student Representatives
Paul Masset, WSBS
Brinda Alagesan, SBU

Secretary
Alyson Kass-Eisler

ADMISSIONS COMMITTEE

Chair
Adrian Krainer

Members
Alexander Gann (*ex officio*)
Bo Li
Zachary Lippman
W. Richard McCombie
Stephen Shea
Adam Siepel
Nicholas Tonks
Christopher Vakoc
Linda Van Aelst

Secretary
Kimberly Creteur

QUALIFYING EXAM COMMITTEE

Chair
Linda Van Aelst

Members
Molly Hammell
Stephen Shea
Christopher Vakoc

EXTERNAL ADVISORY COMMITTEE

Keith Yamamoto (Chair)
Executive Vice Dean, School of Medicine
University of California, San Francisco

Victor Corces
Professor, Department of Biology
Emory University
Professor, Howard Hughes Medical Institute

Gail Mandel
Senior Scientist, Vollum Institute
Oregon Health and Science University
Investigator, Howard Hughes Medical Institute

Marguerite Mangin
Academic Programs Director and Senior Research Associate
The Rockefeller University

Barbara Meyer
Professor of Genetics and Development
University of California, Berkeley
Investigator, Howard Hughes Medical Institute

Frank Solomon
Professor, Department of Biology and Center for Cancer Research
Massachusetts Institute of Technology

WATSON SCHOOL OF BIOLOGICAL SCIENCES DEAN'S REPORT

On May 1, 2016, we celebrated the Watson School's 12th graduation ceremony. Seven students were awarded Ph.D. degrees: Arkarup Bandyopadhay, Colleen Carlston, Silvia Fenoglio, and Lisa Krug from the Entering Class of 2010; Robert Aboukhalil and Anja Hohmann from the Entering Class of 2011; and Tyler Garvin from the Entering Class of 2012. Devinn Lambert from the Entering Class of 2014 was awarded a Masters degree. A posthumous Masters degree was awarded to Sophie Thomain (Entering Class of 2011), who died of leukemia in January. Honorary degrees were bestowed upon Senator Tom Harkin (retired) and Dr. John Tooze.

After earning a Ph.D. from King's College London, John Tooze spent 2 years doing postdoctoral work in Jim Watson's lab at Harvard, before returning to King's as a lecturer. By this time, he was also writing a weekly column on Cell Biology for the journal *Nature*, where he later became Deputy Editor. In 1970, Tooze moved to the Imperial Cancer Research Fund (ICRF) in London as Director of Research. In this capacity, he did much to rejuvenate ICRF and set it on the path to becoming a world leader in cancer research. In 1973, the European Molecular Biology Organization (EMBO) was looking for a new Executive Director, and Tooze spent the next 20 years in that role in addition to serving as Research Coordinator for the European Molecular Biology Laboratory, the large research facility also based in Heidelberg. In 1993, Paul Nurse had become the Scientific Director (and later Director General) of ICRF in London and suggested that Tooze join him as Director of Support Services. This included the period of the merger between ICRF and Cancer Research Campaign, the United Kingdom's two biggest cancer charities, after which he became Executive Director of Research Services at the resulting Cancer Research UK. From 2005 until his retirement in 2013, he served as Vice President for Scientific and Faculty Operations at The Rockefeller University, overseeing operations as varied as research facilities, building programs, and the Rockefeller University press. Beyond this extraordinary career in scientific administration, Tooze's influence is further seen in his record of scientific publishing, starting with his time at *Nature* and later as founding editor of *The EMBO Journal*. He edited the influential book *The Molecular Biology of Tumor Viruses*, published here at CSHL Press in 1974. He later collaborated with Jim Watson as co-author on *The DNA Story* and the undergraduate textbook *Recombinant DNA: A Short Course*, and, with Carl Brandon, of the textbook *An Introduction to Protein Structure*. For his remarkable and varied contributions to science, Tooze was elected to the Royal Society of London in 1994.

Tom Harkin was elected to Congress in 1974 from Iowa's Fifth Congressional District. After serving 10 years in the U.S. House of Representatives, Harkin won a seat in the United States Senate, to which he was reelected four times, becoming the first Iowa Democrat to win a fifth term in the Senate. As a young senator, Harkin was tapped by Senator Ted Kennedy to craft legislation to protect the civil rights of millions of Americans with physical and mental disabilities. The immensely important and influential Americans with Disabilities Act was the eventual end product of this process. Harkin has also worked tirelessly to advance the cause of scientific research and its applications to human health and well-being and delivered significant increases in funding for research into cardiovascular disease, cancer, Alzheimer's disease, and other diseases. Harkin also led the fight to lift restrictions on embryonic stem cell research. More broadly, Harkin was instrumental in leading the bipartisan push to double the National Institutes of Health (NIH) budget between 1998 and 2003, and he was engaged from the beginning in getting the Human Genome Project funded, working on this closely with Jim Watson while he was founding head of that project.

(*Left* to *right*) Robert Aboukhalil, Devinn Lambert, Lisa Krug, Anja Hohmann, CSHL President Bruce Stillman, CSHL Chairman of the Board Jamie Nicholls, honorary degree recipient Tom Harkin, Tyler Garvin, CSHL Chancellor Emeritus James Watson, Silvia Feoglio, honorary degree recipient John Tooze, Arkarup Banerjee, and WSBS Dean Alex Gann

In his convocation address, Senator Harkin spoke of the great value of scientific research to society. Citing the Human Genome Project as an example, he noted that government's investment of $3.4 billion had generated $965 billion in economic activity by 2012, as well as $293 billion in personal income and some 4.3 million job-years of U.S. employment. In addressing Jim Watson's key role in initiating the project, he said, "Our country, indeed the entire world, owes a great debt to Dr. James Watson for his vision and leadership in the Human Genome Project."

Turning to the problems facing American science today, Harkin talked forcefully about the need to end the anti-intellectualism and anti-science sentiments that dominate American political life these days, and which threaten America's dominance in the world of science research and could derail medical advances.

2016 WSBS DOCTORAL RECIPIENTS

Student	Thesis advisor	Academic mentor	Current position
Robert Aboukhalil	Mickey Atwal Michael Wigler	Josh Dubnau	Senior Computational Biologist, Fluidigm, South San Francisco, California
Arkarup Bandyopadhyay	Leemor Joshua-Tor	Christopher Hammell	Postdoctoral Fellow, New York University (Advisor: Michael Long)
Tyler Garvin	Michael Schatz	Zachary Lippman	Senior Software Engineer, Pict, San Francisco, California
Anja Hohmann	Christopher Vakoc	John Inglis	Senior Scientist, KSQ Therapeutics, Cambridge, Massachusetts
Matt Koh	Florin Albeanu	Bo Li	Data Science Fellow, Insight Data Science
Lisa Krug	Josh Dubnau	Stephen Shea	Scientist, Kallyope, New York, New York
Charles Underwood	Robert Martienssen	Michael Schatz	Postdoctoral Fellow, KeyGene, The Wageningen, Netherlands

2016 THESIS DISSERTATION DEFENSES

ENTERING CLASS OF 2010

Arkarup Bandyopadhyay, March 21, 2016

Identity and intensity encoding in the olfactory bulb.

Thesis Examining Committee

Chair:	**Anne Churchland**
Research Mentor:	**Florin Albeanu**
Academic Mentor:	**Zachary Lippman**
Committee Member:	**Anthony Zador**
Academic Mentor:	**Bo Li**
Committee Member:	**Glenn Turner**
External Examiner:	**Na Ji,** *HHMI, Janelia Research Campus*

Lisa Krug, January 15, 2016

The role of transposable elements in TDP-43-mediated neurodegeneration.

Thesis Examining Committee

Chair:	**Robert Martienssen**
Research Mentor:	**Josh Dubnau**
Academic Mentor:	**Stephen Shea**
Committee Member:	**Molly Hammell**
External Examiner:	**Robert Reenan,** *Brown University*

ENTERING CLASS OF 2011

Robert Aboukhalil, March 2, 2016

Elucidating cancer evolution using single-cell sequencing and comparative genomics.

Thesis Examining Committee

Chair:	**Alexander Krasnitz**
Research Mentor(s):	**Mickey Atwal**
	Michael Wigler
Academic Mentor:	**Josh Dubnau**
Committee Member:	**Mikala Egeblad**
External Examiner:	**Bud Mishra,** *New York University*

Brittany Cazakoff, December 7, 2016

State and learning-dependent changes in olfactory granule cells.

Thesis Examining Committee

Chair:	**Anne Churchland**
Research Mentor:	**Stephen Shea**
Academic Mentor:	**Christopher Hammell**
Committee Member:	**Glenn Turner**
External Examiner:	**Leslie Kay,** *University of Chicago*

Anja Hohmann, March 2, 2016

Function and inhibition of BRD9 in acute myeloid leukemia cells.

External Examiner:	**Thomas Cleland,** *Cornell University*

Matt Koh, July 25, 2016

A novel optical method for controlling and recording from neural circuits.

Thesis Examining Committee

Chair:	**Anthony Zador**
Research Mentor:	**Florin Albeanu**
Chair:	**Bruce Stillman**
Research Mentor:	**Christopher Vakoc**
Academic Mentor:	**John Inglis**
Committee Member:	**Christopher Hammell**
Committee Member:	**David Tuveson**
External Examiner:	**Emily Bernstein,** *Mount Sinai School of Medicine*

Charles Underwood, May 4, 2016

Epigenetic and genetic control of transposon silencing and meiotic crossover in Arabidopsis thaliana.

Thesis Examining Committee

Chair:	**David Jackson**
Research Mentor:	**Robert Martienssen**
Academic Mentor:	**Michael Schatz**
Committee Member:	**Ian Henderson,** *University of Cambridge*
External Examiner:	**Vincent Colot,** *L' Ecole Normale Supérieure*

ENTERING CLASS OF 2012

Tyler Garvin, March 1, 2016

Algorithms for the analysis of cancer genomics with advanced sequencing technologies.

Thesis Examining Committee

Chair:	**W. Richard McCombie**
Research Mentor:	**Michael Schatz**
Academic Mentor:	**Zachary Lippman**
Committee Member:	**David Tuveson**
External Examiner:	**Ben Raphael,** *Brown University*

Annabel Romero Hernandez, October 20, 2016

Molecular mechanisms of inhibition and subtype specificity of NMDA receptors.

Thesis Examining Committee

Chair:	**Bo Li**
Research Mentor:	**Hiro Furukawa**
Academic Mentor:	**Adrian Krainer**
Committee Member:	**Nicholas Tonks**
Committee Member:	**Lonnie Wollmuth,** *Stony Brook University*
External Examiner:	**Stephen Traynelis,** *Emory University*

DOCTORAL THESIS RESEARCH

Student	Academic mentor	Research mentor	Thesis research
ENTERING CLASS OF 2011			
Brittany Cazakoff *Edward and Martha Gerry Fellow* *NSERC Scholar*	Christopher Hammell	Stephen Shea	Dynamic granule cell processing of odor information. Thesis defended: December 7, 2016
Joaquina Delas Vives *La Caixa Fellow* *Boehringer Ingelheim Fonds Fellow*	Nicholas Tonks	Gregory Hannon	Functional role of long noncoding RNAs in hematopoiesis. Thesis defense: March 27, 2017
Justus Kebschull *David and Fanny Luke Fellow* *Genentech Foundation Fellow* *Boehringer Ingelheim Fonds Fellow*	Marja Timmermans	Anthony Zador	Grasping the brain. Thesis defense: March 31, 2017
Fred Marbach *Farish-Gerry Fellow*	Josh Dubnau	Anthony Zador	A study of auditory corticostriatal cells in the behaving mouse.
Onyekachi Odoemene *NIH Individual Fellowship* *William Randolph Hearst Scholar*	Stephen Shea	Anne Churchland	The role of neural inhibition in perceptual decision-making.
ENTERING CLASS OF 2012			
Talitha Forcier *NIH Predoctoral Trainee* *William Randolph Hearst Scholar*	Nicholas Tonks	Justin Kinney	Building the transcriptional regulatory code from the ground up.
Yu-Jui (Ray) Ho *David and Fanny Luke Fellow* *Taiwan Government Fellowship*	Michael Schatz	Molly Hammell	Methods development for low-input RNA-Seq analysis and application to breast cancer heterogeneity.
Paul Masset *Florence Gould Fellow*	Jan A. Witkowski	Adam Kepecs	Representations of decision confidence in the brain: From Bayes' rule to channelrhodopsin.
Annabel Romero Hernandez *Genentech Foundation Fellow*	Adrian R. Krainer	Hiro Furukawa	Molecular mechanisms of inhibition in NMDA receptors. Thesis defended: October 20, 2016
Abram Santana *Robert and Theresa Lindsay Fellow* *William Randolph Hearst Scholar*	Lloyd Trotman	David Tuveson	Pancreatic tumor biology and therapy.
ENTERING CLASS OF 2013			
Giorgia Battistoni *Starr Centennial Scholar* *Boehringer Ingelheim Fonds Fellow*	Christopher Hammell	Gregory Hannon	One-carbon metabolism and DNA methylome dynamics in pancreatic cancer.
Lital Charatifsky *Israeli WSBS Fellow*	John Inglis	Anne Churchland	Neural circuits for multisensory integration in normal and disease states.
Sanchari Ghosh *Charles A. Dana Fellow*	Josh Dubnau	Anthony Zador	Role of corticostriatal plasticity in learning an auditory discrimination task.
Michael Gutbrod *Bristol-Myers Squibb Fellow*	Zachary Lippman	Robert Martienssen	Small RNA and the RNAi pathway in transposable element regulation and differentiation in the preimplantation embryo.
Daniel Kepple *Crick-Clay Fellow* *NIH Predoctoral Trainee*	David Stewart	Alexei Koulakov	The human olfactory space.
Laura Maiorino *George A. and Marjorie Anderson Fellow* *Boehringer Ingelheim Fonds Fellow*	Nicholas Tonks	Mikala Egeblad	Understanding the role of the epithelial–mesenchymal plasticity in pancreatic cancer metastasis.
Maria Nattestad *John and Amy Phelan Student* *Genentech Foundation Fellow*	Linda Van Aelst	Michael Schatz	Algorithms for analyzing and assembling complex genomes. Thesis defense: January 26, 2017
Georgi Yordanov *Leslie C. Quick, Jr. Fellow* *Boehringer Ingelheim Fonds Fellow*	Leemor Joshua-Tor	David Tuveson	Role of c-Myc in pancreatic cancer.

(*continued*)

DOCTORAL THESIS RESEARCH (*continued*)

Student	Academic mentor	Research mentor	Thesis research
ENTERING CLASS OF 2014			
Emillis Bruzas *Starr Centennial Scholar*	Alea Mills	Mikala Egeblad	Investigation of mechanisms responsible for reawakening and chemoresistance in a breast cancer dormancy model.
Hamza Giaffar *Robert and Theresa Lindsay Fellow*	Jan Witkowski	Alexei Koulakov	The primacy model of olfactory coding.
Jacqueline Giovanniello *NIH Predoctoral Trainee*	Bruce Stillman	Bo Li	Disruption of central amygdala fear circuit in a 16p11.2 microdeletion model of autism.
Elizabeth Hutton *Elizabeth Sloan Livingston Fellow*	Molly Hammell	Adam Siepel	Functional variant prediction in noncoding regions.
Sashank Pisupati *Cashin Fellow*	Stephen Shea	Anne Churchland	Dissecting the circuits and mechanisms that support optimal multisensory integration in rodents.
Colin Stoneking *NIH Predoctoral Trainee*	Zachary Lippman	Anthony Zador	Neuronal mechanisms enabling decision-making to be learned.
Jue Xiang Wang *George and Marjorie Anderson Fellow* *Boehringer Ingelheim Fonds Fellow*	Mikala Egeblad	Hiro Furukawa	Impact of subunit composition and de novo mutations on NMDA receptor structure, channel function, and interactions.
Anqi Zhang *Starr Centennial Scholar*	Bo Li	Anthony Zador	From corticostriatal plasticity to a common pathway.
ENTERING CLASS OF 2015			
Benjamin Berube *NIH Predoctoral Trainee* *Elisabeth Sloan Livingston Fellow*	Zachary Lippman	Robert Martienssen	A single-cell assessment of germline epigenetic heterogeneity.
Kristina Grigaityte *Farish-Gerry Fellow*	John Inglis	Mickey Atwal	Computational analyses of high-throughput single-cell T-cell receptor sequences in health and disease.
Matt Lee *David H. Koch Fellow*	Nicholas Tonks	Lloyd Trotman	The road to metastasis: Defining the initial stages of prostate cancer progression.
Katarina Meze *Leslie C. Quick Jr. Fellow*	Jay Lee	Leemor Joshua-Tor	Structural and functional studies of RNA regulatory mechanisms mediated by Lin28.
Alexandra Nowlan *George and Marjorie Anderson Fellow*	Jessica Tollkuhn	Stephen Shea	Multisensory experience-dependent plasticity: Network dynamics in auditory processing following parturition.
Sofya Polyanskaya *Starr Centennial Scholar*	Alexander Krasnitz	Christopher Vakoc	Identification of fusion oncoprotein co-dependencies in cancer.
Ngoc Tran *Samuel Freeman Fellow*	Leemor Joshua-Tor	Alexei Koulakov	Predicting bioactivity using a data-driven representation of three-dimensional chemical structures.

Teaching Award

At the graduation ceremony this year, the School awarded its 10th Winship Herr Award for Excellence in Teaching to Mickey Atwal, the lead instructor of the Specialized Disciplines course in Quantitative Biology. Mickey, who also won the award in 2011, was chosen by the students for this award, based on his enthusiasm and his creativity in teaching. The following is some of what the students said about Mickey's teaching in their nominations: "Mickey always found ways to explain concepts that seem intimidating with intuitive and graphical examples." And, "I thought Mickey was the most enthusiastic teacher we had, who always aimed to make sure we understood

Mickey Atwal

Carrie Cowan and Alex Gann

everything before moving on to the next topic. He asked questions to challenge us and keep us engaged during the class and made quantitative biology thoroughly enjoyable."

Faculty and Administrative Changes

In May 2016, our Associate Dean, Carrie Cowan, left the School to take a position as Director of Pre- and Postdoctoral Programs at The Jackson Laboratory in Maine. We thank Carrie for her 3 years of service to the School, where she made a significant impact. In addition to leading us through our reaccreditation, she made improvements to our curriculum, developed a system for tracking student, faculty, and alumni data, and generally improved the quality of our programs.

One new faculty member joined the Watson School in 2016: Ullas Pedmale. Ullas earned his Ph.D. at the University of Missouri, Columbia. During his graduate training, he studied how phototropism is mediated by two key proteins in the pathway: phot1 and NPH3. As a postdoctoral fellow in Joanne Chory's lab at the Salk Institute, he worked to understand the molecular mechanisms underlying plant plasticity, or changes in plant architecture in response to variations in light quality. His CSHL lab is interested in determining the mechanisms behind how a plant perceives and successfully adapts to its environment. He also aims to understand how a plant must integrate intrinsic and extrinsic cues and "decide" how best to respond to environmental cues. Understanding how plants deal with and respond to a multitude of environmental signals could help to develop crops that cope with unfavorable growth conditions without significant changes in yield.

Ullas has already participated in WSBS activities, including being a guest lecturer in the Specialized Disciplines Course on Genetics and Genomics, giving Research Topics talks to the first-year students, and judging the poster session at the In-House Symposium. We look forward to his growing participation as a member of the faculty.

Admissions 2016

The School received 300 applications for the Entering Class of 2016 and is indebted to its Admissions Committee, whose members reviewed, interviewed, and selected candidates for our doctoral program. The Admissions Committee for the 2016 entering class comprised Adrian Krainer (Chair), Bo Li, Zachary Lippman, W. Richard McCombie, Stephen Shea, Adam Siepel, Nicholas Tonks, Christopher Vakoc, Linda Van Aelst, and myself.

Entering Class of 2016

On August 22, 2016, the WSBS welcomed the 17th incoming class, consisting of 11 new students: Brianna Bibel, Alberto Corona, David Johnson, Christopher Krasniak, Shaina Lu, Kathryn O'Neill, Luqun Shen, Olya Spassibjoko, Martyna Sroka, Ran Yan, and Chengxiang Yuan.

ENTERING CLASS OF 2016

Brianna Bibel, Saint Mary's College of California: B.S. in Biology (2016)
Academic Mentor: Hiro Furukawa

Alberto Corona, University of California, Riverside: B.S. in Biology (2016); HHMI EXROP Fellowship: David Anderson's lab, Caltech
Academic Mentor: David Jackson

David Johnson, Brown University: B.S. in Neurobiology (2016); HHMI EXROP Fellowship: Oliver Hobert's lab, Columbia University
Academic Mentor: Zachary Lippman

Christopher Krasniak, Colby College: B.A. in Neurobiology (2016)
Academic Mentor: Jan Witkowski

Shaina Lu, Swarthmore College: B.A. in Biology (2016); HHMI EXROP Fellowship: Matt Waldon's lab, Harvard University
Academic Mentor: Leemor Joshua-Tor

Kathryn O'Neill, University of California, Davis: B.S. Genetics and Genomics (2016); McNair Scholar
Academic Mentor: Camila dos Santos

Luqun Shen, University of Notre Dame; B.S. in Biology (2016)
Academic Mentor: David Stewart

Olya Spassibojko, Cornell University: B.S. in Biological Sciences (2016)
Academic Mentor: Jessica Tollkuhn

Martyna Sroka, University of Aberdeen, United Kingdom: B.S. in Biomedical Sciences (2016)
Academic Mentor: Christopher Vakoc

Ran Yan, Wright State University: M.S. in Biological Sciences (2013); Ocean University of China, B.S. in Biological Sciences (2010)
Academic Mentor: David Tuveson

Chengxiang Yuan, Imperial College London: B.S. in Biochemistry (2015); A*Star National Science Scholarship
Academic Mentor: Nicholas Tonks

2016 Entering Class: (*Top row, left to right*) Olya Spassibjojko, Ran (Rena) Yan, Martyna Sroka, Shaina Lu, Brianna Bibel. (*Bottom row, left to right*) Chengxiang (Charlie) Yuan, David Johnson, Kathryn O'Neill, Christopher Krasniak, Luqun Shen, Alberto Corona

Academic Mentoring

The Watson School takes great pride in our student mentoring. One example is our two-tiered mentoring approach, whereby each student chooses both an academic and a research mentor. The academic mentor is a critical advisor during the intensive coursework of the first term, during their rotations, and when identifying a suitable research mentor. Furthermore, the academic mentor continues to follow them throughout their doctoral experience, often serving as important advocates for the students. Entering students select, by mutual agreement, a member of the research or nonresearch faculty to serve as an academic mentor. This program continues to receive much support from the faculty who volunteer to be academic mentors, and it has rightfully become a vital ingredient in our success. The following are the Academic Mentors for the Entering Class of 2016.

STUDENT	MENTOR	STUDENT	MENTOR
Brianna Bibel	Hiro Furukawa	Luqun Shen	David Stewart
Alberto Corona	David Jackson	Olya Spassibojko	Jessica Tollkuhn
David Johnson	Zachary Lippman	Martyna Sroka	Christopher Vakoc
Christopher Krasniak	Jan Witkowski	Ran Yan	David Tuveson
Shaina Lu	Leemor Joshua-Tor	Chengxiang Yuan	Nicholas Tonks
Kathryn O'Neill	Camila dos Santos		

Recruiting Efforts

This year, we once again focused primarily on targeted visits to graduate fairs and minority conferences to identify prospective students for the School. In addition to these visits, the WSBS generated new mailing lists for the distribution of information to top undergraduate science departments around the world. A multiprogram booklet, incorporating the graduate, undergraduate, and postdoctoral programs, was designed for this recruitment season. Additionally, emails were sent to personalized contacts and an electronic mailing list of more than 50,000 individuals who receive information from the Cold Spring Harbor Laboratory Press or have attended meetings or courses at the Laboratory. We are grateful to these departments for sharing this contact list. We received 202 applications for the Entering Class of 2017, and it appears that many outstanding candidates have applied to the program.

Although we saw a decrease in the total number of applications to the program this year compared to last, the quality of the applicant pool was as high, if not higher, than in past years. Much of the decrease is likely attributable to changes we (deliberately) made to recruitment this year in an attempt to enrich the pool for the best candidates. The applicant pool is reviewed to identify the top 100 or so applications for further review by the admissions committee. It was clear during this process that the decrease in applicants this year was chiefly, if not exclusively, among those that would have been eliminated in the original triage.

WATSON SCHOOL OF BIOLOGICAL SCIENCES 2016 RECRUITMENT SCHEDULE

Event	Location	Date
American Association of Cancer Research, Annual Meeting	New Orleans, Louisiana	April 16–20
Vassar College, Visit and Information Session	Cold Spring Harbor Laboratory	April 28
Big 10+ Graduate School Expo (Purdue University), Graduate School Fair	West Lafayette, Indiana	September 25–26
University of Delaware National McNair/TRIO Scholars, Research Conference, Graduate School Fair	Newark, Delaware	October 13
Society for Advancement of Chicanos and Native Americans in Science (SACNAS), National Conference	Long Beach, California	October 13–15

(*continued*)

WATSON SCHOOL OF BIOLOGICAL SCIENCES 2016 RECRUITMENT SCHEDULE (*continued*)

Event	Location	Date
American Society for Human Genetics, Annual Meeting	Vancouver, British Columbia, Canada	October 18–22
Hunter College MARC and MBRS/RISE, Information Session	New York, New York	October 19
University of Maryland, Baltimore County, Meyerhoff Scholars Program, Graduate School Fair	Baltimore, Maryland	October 24
Midwest Association of TRIO and College Access and Success Professionals (EOA) Annual National Ronald E. McNair Research Conference, Graduate School Fair	Milwaukee, Wisconsin	October 28–30
National Association of African American Honors Programs (NAAAHP) Conference, Graduate School Fair	Nashville, Tennessee	October 29–November 1
Vanderbilt University, Information Session	Nashville, Tennessee	November 2
California Forum for Diversity in Graduate Education, Graduate School Fair	Los Angeles, California	November 5
Annual Biomedical Research Conference for Minority Students (ABRCMS), National Conference	Tampa, Florida	November 9–12
Society for Neuroscience Annual Meeting, Graduate School Fair	San Diego, California	November 12–16
American Society for Cell Biology, Annual Meeting	San Francisco, California	December 3–7

Students from Other Institutions

WSBS students account for approximately half of the total graduate student population here at CSHL; the other half comprises visiting graduate students from other universities who have decided to conduct some or all of their thesis research in CSHL faculty members' laboratories. A large fraction of students are from Stony Brook University (SBU), via a program established between CSHL and SBU more than 30 years ago. Over the years we have established relationships with other institutions around the world, enabling their students to conduct research here at CSHL. Currently, we have visiting students from institutions in China, France, Italy, Russia, Slovakia, The Netherlands, and the United States. The Watson School provides a contact person for the students and maintains relationships with the administrators from their home institutions. These students are fully integrated into the CSHL community and receive all the necessary assistance as they navigate the complexities of performing doctoral research away from their home institutions. The following students, listed below, joined us from SBU this year.

STUDENT	CSHL RESEARCH MENTOR	STONY BROOK UNIVERSITY PROGRAM
Batuhan Baserdem	Alexei Koulakov	Physics
Kush Coshic	Justin Kinney	Physics
William Galbavy	Z. Josh Huang	Neuroscience
Yixin Hu	Bruce Stillman	Molecular and Cellular Biology
Judith Mizrachi	Pavel Osten	Biomedical Engineering
Tzvia Pinkhasov	Adam Kepecs	Neuroscience/MSTP
Padmina Shrestha	Alea Mills	Molecular and Cellular Biology
Xiaoli Wu	Christopher Vakoc	Genetics
Wenbo Xu	David Spector	Molecular and Cellular Biology
Qian Zhang	Adrian Krainer	Molecular and Cellular Biology

Graduate Student Symposium

Each year, the students participate in two Graduate Student Symposia held at the Laboratory's Genome Research Center in Woodbury: one in May, the other in October. Each Symposium

consists of senior students giving short talks, with coffee breaks and lunches providing opportunities for more informal interactions. The prize for the best talk for the May session was awarded to Laura Maiorino (WSBS, Egeblad lab) and for the October session, it was awarded to Justus Kebschull (WSBS, Zador lab) and Allen Yu (SBU, Spector lab). We are grateful to the two student chairs—Tobiloba Oni (SBU) and Lital Chartarifsky (WSBS)—for their hard work and to WSBS's Kim Geer for helping make the Symposium a great success.

Graduate Student and Postdoctoral Fellow Departures

With each year come not only new arrivals, but also departures. The following graduate students and postdoctoral fellows left the Laboratory during 2016.

Postdoctoral Fellows

Jennifer Beshel
Yung-Heng Chang
Delphine Fagegaltier
Priscila Ferreira Papa
Jong-Jin Han
Erkan Karakas
Billy Lau
Fei Li
Siran Li
Uri Livneh
Tomas Malinauskas
Almudena Molla-Morales
Christos Noutsos
Jaclyn Novatt
Erik Olund
Juwon Park
Katherine Petsch
Mahshid Rahmat
Ileng Ramachandran
Meng-Fu Shih
Joanna Szczurkowska
Yang Yu
Wei Zheng

Graduate Students

Robert Aboukhalil
Arkarup Banerjee
Anand Bhagwat
Sudipto Chakrabortty
Na Hyun Cho
Miriam Fein
Tyler Garvin
Priyanka Gupta
Anja Hohmann
Ying Hu
Sean Kelly
Matthew Koh
Lisa Krug
Shanshan Li
Tomoki Nomakuchi
Jason O'Rawe
Priyadarshini Ravichandran
Ioana Rus
Chen Shen
Junwei Shi
Sophie Thomain
Charles Underwood
Robert Wysocki
Viacheslav Zhygulin

Executive Committee

The School's Executive Committee, in its monthly meetings, provides year-round direction for the School and its students through its invaluable policy recommendations. I thank faculty members Mikala Egeblad, Alexander Krasnitz, Bo Li, David Spector, David Stewart, and Anthony Zador. I also thank the student representatives Brinda Alagesan (SBU) and Paul Masset (WSBS), who contributed to discussions and provided useful suggestions and feedback from their colleagues.

The Watson School Continues to Benefit from Generous Benefactors

We are extremely grateful for the generous donors, whose one-time gifts or continued support made our 2016 programs possible, including The Annenberg Foundation, Lita Annenberg Hazen Foundation, Bristol-Myers Squibb Corporation, Mr. and Mrs. Landon Clay, Lester Crown, The Dana Foundation, Henriette and Norris Darrell, The Samuel Freeman Charitable Trust, The William Stamps Farish Fund, The Genentech Foundation, Mr. and Mrs. Alan Goldberg, Florence Gould Foundation, Gonzalo Río Arronte Foundation, William Randolph Hearst Foundation, Dr. and Mrs. Mark Hoffman, Annette Kade Charitable Trust, Mr. David H. Koch, Mr. and Mrs. Robert D. Lindsay and Family, Mr. and Mrs. David Luke III, Mr. and Mrs. William R. Miller, OSI Pharmaceuticals Foundation, Estate of Edward L. Palmer, Mr. and Mrs. John C. Phelan, The Quick Family, Estate of Elisabeth Sloan Livingston, The Starr Foundation, The Roy J. Zuckerberg Family Foundation, and anonymous donors.

We are also continuously grateful for our endowed lectureships: The John P. and Rita M. Cleary Visiting Lectureship, The George W. Cutting Lectureship, The William Stamps Farish Lectureship, The Martha F. Gerry Visiting Lectureship, The Edward H. Gerry Visiting Lectureship, The Edward H. and Martha F. Gerry Lectureship, The Susan T. and Charles E. Harris

Visiting Lectureship, The Klingenstein Lectureship, The Mary D. Lindsay Lectureship, The Pfizer Lectureship, The George B. Rathmann Lectureship, The Seraph Foundation Visiting Lectureship, The Sigi Ziering Lectureship, The Daniel E. Koshland Visiting Lectureship, The Michel David-Weill Visiting Lectureship, and The Fairchild Martindale Visiting Lectureship.

In addition, we are very fortunate to hold a prestigious National Research Service Award Predoctoral Training Grant from the National Institutes of Health, National Institute of General Medical Sciences.

Student and Alumni Achievements

To date, 92 students have received their Ph.D. degree from the WSBS. Twenty-two graduates currently hold tenure-track faculty positions, including eight who have attained the rank of Associate Professor. Our graduates have also moved into influential positions in publishing, medicine, government, consulting, and industry. In 2016, Hiroshi Makino started his own lab at the European Molecular Biology Laboratory in Rome; Saya Ebbesen joined the BGB Group, a medical communications agency in New York City, as a medical writer; Oliver Fregoso started his own lab at the University of California, Los Angeles (UCLA); Devinn Lambert is a Democratic Science and Technology Fellow, Senate Committee on Energy and Natural Resources, U.S. Department of Energy; Beth Nakasone started a residency in Internal Medicine at UCLA; and Susann Weissmueller is an Investor Relations Officer in the Strategic Partnering Department at Roche in Switzerland.

The WSBS students continue to impress us with their accomplishments. They publish their research findings in prestigious international journals (more than 350 papers to date), present their research at international meetings, and obtain fellowships to pursue their research interests.

In 2016, our current students and alumni were successful in receiving the following prestigious awards and fellowships:

- WSBS student Jue Xiang Wang was awarded a predoctoral fellowship from the Boehringer Ingelheim Fonds.
- WSBS graduate Ian Peikon was named one of Forbes magazine's "30 Under 30" in science for his graduate work on the connectome.
- WSBS student Giorgia Battistoni was named a finalist for the Regeneron Prize for Creative Innovation.
- WSBS graduate Monica Dus (currently an Assistant Professor at the University of Michigan) received a Rita Allen Foundation Award and the Klingenstein-Simons Fellowship for Neuroscience. Monica was also nominated for the Golden Apple teaching award at the University of Michigan, and she was also awarded the NIH Director's New Innovator Award.
- WSBS graduate Kaja Wasik was selected as a program participant at Entrepreneurship Lab (ELabNYC) Bio & Health Tech.

Prizes for the best posters by a postdoctoral fellow and by a graduate student were awarded at the Laboratory's annual In-House Symposium held in November 2016. The poster session provides a forum for the postdoctoral fellows and students to show off their research and gives the entire scientific community a chance to come together and share ideas beyond the walls of their individual laboratories. This year, Laura Maiorino, a Watson School student from Mikala Egeblad's laboratory, won the graduate student prize. The postdoctoral prize was shared by Rowan Herridge from Robert Martienssen's laboratory, Dhananjay Huilgol from Josh Huang's laboratory, Matthew Kaufman from Anne Churchland's laboratory, and Hillary Shiff from Bo Li's laboratory.

2016 WSBS STUDENT (CURRENT OR PREVIOUS) PUBLICATIONS

Bhagwat AS, Roe JS, Mok BY, **Hohmann AF**, Shi J, Vakoc CR. 2016. BET bromodomain inhibition releases the mediator complex from select *cis*-regulatory elements. *Cell Rep* **15:** 519–530.

Chin CS, Peluso P, Sedlazeck FJ, **Nattestad M**, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13:** 1050–1054.

Choi K, Reinhard C, Serra H, Ziolkowski PA, **Underwood CJ**, Zhao X, Hardcastle TJ, Yelina NE, Griffin C, Jackson M, Mézard C, McVean G, Copenhaver GP, Henderson IR. 2016. Recombination rate heterogeneity within *Arabidopsis* disease resistance genes. *PLoS Genet* **12:** e1006179.

Ebbesen SH, Scaltriti M, Bialucha CU, Morse N, Kastenhuber ER, Wen HY, Dow LE, Baselga J, Lowe SW. 2016. Pten loss promotes MAPK pathway dependency in HER2/neu breast carcinomas. *Proc Natl Acad Sci* **113:** 3030–3035.

Fagegaltier D, Falciatori I, Czech B, **Castel S**, Perrimon N, Simcox A, Hannon GJ. 2016. Oncogenic transformation of *Drosophila* somatic cells induces a functional piRNA pathway. *Genes Dev* **30:** 1623–1635.

Hohmann AF, Martin LJ, Minder JL, Roe JS, Shi J, Steurer S, Bader G, McConnell D, Pearson M, Gerstberger T, Gottschamel T, Thompson D, Suzuki Y, Koegl M, Vakoc CR. 2016. Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nat Chem Biol* **12:** 672–679.

Kebschull JM, Garcia da Silva P, Zador AM. 2016. A new defective helper RNA to produce recombinant Sindbis virus that infects neurons but does not propagate. *Front Neuroanat* **10:** 56.

Kebschull JM, Garcia da Silva P, Reid AP, **Peikon ID**, Albeanu DF, Zador AM. 2016. High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **S0896-6273:** 30421–30414.

Kim Y, **Perova Z**, Mirrione MM, Pradhan K, Henn FA, Shea S, Osten P, Li B. 2016. Whole-brain mapping of neuronal activity in the learned helplessness model of depression. *Front Neural Circuits* **10:** 3.

Liu Y, Chen C, Xu Z, **Scuoppo C**, Rillahan CD, Gao J, Spitzer B, Bosbach B, Kastenhuber ER, Baslan T, Ackermann S, Cheng L, Wang Q, Niu T, Schultz N, Levine RL, Mills AA, Lowe SW. 2016. Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature* **531:** 471–475.

Manchado E, **Weissmueller S**, Morris JP, Chen CC, Wullenkord R, Lujambio A, de Stanchina E, Poirier JT, Gainor JF, Corcoran RB, Engelman JA, Rudin CM, Rosen N, Lowe, SW. 2016. A combinatorial strategy for treating KRAS-mutant lung cancer. *Nature* **534:** 647–651.

Nattestad M, Schatz MC. 2016. Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32:** 3021–3023.

Park J, Wysocki RW, Amoozgar Z, **Maiorino L**, Fein MR, Jorns J, Schott AF, Kinugasa-Katayama Y, Lee Y, Won NH, **Nakasone ES**, Hearn SA, Küttner V, Qiu J, Almeida AS, Perurena N, Kessenbrock K, Goldberg MS, Egeblad M. 2016. Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. *Sci Transl Med* **8:** 361ra138.

Pisupati S, **Chartarifsky L**, Churchland AK. 2016. Decision activity in parietal cortex—Leader or follower? *Trends Cogn Sci* **20:** 788–789.

Plavskin Y, Nagashima A, Perroud PF, Hasebe M, Quatrano RS, Atwal GS, Timmermans MC. 2016. Ancient *trans*-acting siRNAs confer robustness and sensitivity onto the auxin response. *Dev Cell* **36:** 276–289.

Ren J, **Castel SE**, Martienssen RA. 2015. Dicer in action at replication-transcription collisions. *Mol Cell Oncol* **2:** e991224.

Romero-Hernandez A, Simorowski N, Karakas E, Furukawa H. 2016. Molecular basis for subtype specificity and high-affinity zinc inhibition in the GluN1-GluN2A NMDA receptor amino-terminal domain. *Neuron* **92:** 1324–1336.

Sanders JI, Hangya B, Kepecs A. 2016. Signatures of a statistical computation in the human sense of confidence. *Neuron* **90:** 499–506.

Tasdemir N, Banito A, Roe JS, Alonso-Curbelo D, Camiolo M, Tschaharganeh DF, Huang CH., **Aksoy O**, Bolden JE, Chen CC, Fennell M, Thapar V, Chicas A, Vakoc CR, Lowe SW. 2016. BRD4 connects enhancer remodeling to senescence immune surveillance. *Cancer Discov* **6:** 612–629.

Vembar SS, Seetin M, Lambert C, **Nattestad M**, Schatz MC, Baybayan P, Scherf A, Smith ML. 2016. Complete telomere-to-telomere de novo assembly of the *Plasmodium falciparum* genome through long-read (>11 kb), single-molecule, real-time sequencing. *DNA Res* **23:** 339–351.

Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, Singh S, Thevasagayam NM, Prakki SR, Purushothaman K, Saju JM, Jiang J, Mbandi SK, Jonas M, Hin Yan Tong A, Mwangi S, Lau D, Ngoh SY, Liew WC, Shen X, Hon LS, Drake JP, Boitano M, Hall R, Chin CS, Lachumanan R, Korlach J, Trifonov V, Kabilov M, Tupikin A, Green D, Moxon S, **Garvin T**, Sedlazeck FJ, Vurture GW, Gopalapillai G, Kumar Katneni V, Noble TH, Scaria V, Sivasubbu S, Jerry DR, O'Brien SJ, Schatz MC, Dalmay T, Turner SW, Lok S, Christoffels A, Orbán L. 2016. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genet* **12:** e1005954.

Watson School student is designated in boldface.

WSBS GRADUATES IN FACULTY POSITIONS (IN ORDER OF COMPLETION)

Name	Faculty Position
Amy Caudy	Assistant Professor, University of Toronto, Ontario, Canada
Ira Hall	Associate Professor, Washington University in St. Louis, Missouri
Niraj Tolia	Associate Professor, Washington University in St. Louis, Missouri
Patrick Paddison	Associate Member, Fred Hutchinson Cancer Research Center, Seattle, Washington
Elizabeth Bartom (née Thomas)	Assistant Professor, Northwestern University, Evanston, Illinois
Michelle Cilia	Research Molecular Biologist, U.S. Department of Agriculture, and Adjunct Assistant Professor, Cornell University, Ithaca, New York
Zachary Lippman	Associate Professor, Cold Spring Harbor Laboratory
Ji-Joon Song	Associate Professor, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
Elena Ezhkova	Associate Professor, Mount Sinai School of Medicine, New York, New York
Masafumi Muratani	Associate Professor, University of Tsukuba, Japan
Marco Mangone	Associate Professor, Arizona State University, Tempe
Elizabeth Murchison	Reader, Cambridge University, United Kingdom
Hiroki Asari	Group leader, EMBL Monterotondo, Rome
François Bolduc	Associate Professor, University of Alberta, Edmonton, Canada
Wei Wei	Assistant Professor, University of Chicago, Illinois
Christopher Harvey	Assistant Professor, Harvard University, Cambridge, Massachusetts
Tomas Hromadka	Project Leader, Slovak Academy of Sciences, Bratislava
Monica Dus	Assistant Professor, University of Michigan, Ann Arbor
Daniel Chitwood	Assistant Professor, Donald Danforth Plant Science Center, St. Louis, Missouri
Jeremy Wilusz	Assistant Professor, University of Pennsylvania, Philadelphia
Oliver Fregoso	Assistant Professor, University of California, Los Angeles
Yaniv Erlich	Assistant Professor, New York Genome Center, Columbia University
Colin Malone	Assistant Professor and Director of Genomic Analysis and Technical Operations, Columbia University, New York, New York
Hiroshi Makino	Assistant Professor, Nanyang Technological University, Singapore
Katherine McJunkin	Section Chief, National Institutes of Health

WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION)

Name	Current Position
Emiliano Rial-Verde	Director, Corporate Development, Bunge Limited, New York
Rebecca Ewald	International Business Leader, Ventana Medical Systems/Roche, Arizona
Catherine Seiler (née Cormier)	Manager, The Biobank Core Facility, St. Joseph's Hospital and Medical Center, Phoenix, Arizona
Darren Burgess	Senior Editor, *Nature Reviews Genetics*, United Kingdom
Rebecca Bish-Cornelissen	Senior Scientific Editor, *D.E. Shaw Research*
Allison Blum	Science Operations Manager, Howard Hughes Medical Institute, Columbia University, New York, New York
Keisha John	Director of Diversity Programs, University of Virginia, Charlottesville
Oliver Tam	Sequencing Analyst, ARUP Laboratories, Salt Lake City, Utah
Amy Rappaport	Scientist, Theranos, Emeryville, California
Frederick Rollins	Consultant, LEK Consulting, Boston, Massachusetts
Patrick Finigan	Associate II, Regulatory Affairs CMC, Sanofi, New Jersey
Elizabeth Nakasone	Resident Physician, UCLA Health, California
Maria Pineda	Co-Founder, CEO, Envisagenics, New York, New York
Felix Schlesinger	Bioinformatics Scientist, Illumina, Inc., San Diego, California
Paloma Guzzardo	Team Leader, Horizon Discovery, Vienna, Austria
Saya Ebbesen	Medical Writer, BGB Group, New York, New York
Joshua Sanders	Founder and C.E.O., Sanworks, L.L.C., Stony Brook, New York

(*continued*)

WSBS GRADUATES IN NONACADEMIC POSITIONS (IN ORDER OF COMPLETION) (*continued*)

Name	Current Position
Mitchell Bekritsky	Bioinformatic Scientist, Illumina, Inc., Cambridge, United Kingdom
Sang-Geol Koh	Scientist and Entrepreneur, {Mind}, South Korea
Susann Weissmueller	Strategic Partnering Associate, Roche, Switzerland
Ian Peikon	Lead Scientist, Kallyope, New York, New York
Jack Walleshauser	Research Scientist at Plexxikon Inc., Berkeley, California
Robert Aboukhalil	Senior Bioinformatics Software Engineer, Genapsys, California
Anja Hohmann	Senior Scientist, KSQ Therapeutics, Boston, Massachusetts
Tyler Garvin	Senior Software Engineer, Pict, California
Lisa Krug	Scientist, Kallyope, New York
Charles Underwood	Scientist, KeyGene, The Netherlands
Maria Nattestad	Founder, OMGenomics
Annabel Romero Hernandez	Associate Scientist, Regeneron Pharmaceuticals
Matt Koh	Data Science Fellow, Insight Data Science

In Memoriam

We would like to acknowledge the untimely passing of WSBS student Sophie Thomain (M.S. 2016) and WSBS graduates Joseph Calarco (Ph.D. 2013) and Michael Giangrasso (M.S. 2015). They will be much missed, and in memory of their time at CSHL, bricks in each of their names will be included in the Beckman Plaza.

Alexander Gann
WSBS Professor and Dean

SPRING CURRICULUM

Topics in Biology

ARRANGED BY **Alyson Kass-Eisler and Jan A. Witkowski**

Each year, invited instructors offer 7-day courses at the Banbury Conference Center exploring specialized topics outside the expertise of the Cold Spring Harbor Laboratory faculty. These courses include morning and evening lectures, as well as afternoon sessions during which students read assigned papers or work on problem sets and presentations. In Spring 2016, there was only one course: Evolution. The Microbial Pathogenesis course, which was set to begin May 1, was cancelled because of the unexpected death of instructor Ronald Taylor. Ron was a professor at Dartmouth University and a long-time instructor of the WSBS Topics course.

Evolution

March 20–26 **Attended by the entering classes of 2012 and 2013**

INSTRUCTOR **Nipam Patel,** University of California, Berkeley

VISITING LECTURERS **Rob DeSalle,** American Museum of Natural History
Kelley Harris, Stanford University
Mansi Srivastava, Massachusetts Institute of Technology
Jack Tseng, American Museum of Natural History

TEACHING FELLOWS **Adam Session**, DOE Joint Genome Institute
Rachel Thayer, University of California, Berkeley

Topics in Biology—Evolution 2016. (*Left to right*) Jue Xiang Wang, Onyekachi Odoemene, Jackie Giovanniello, Justus Kebschull, Brittany Cazakoff, Joaquina Delas Vives, Charlie Underwood, Colin Stoneking, Fred Marbach, Emilis Bruzas, Anqi Zhang, Hamza Giaffar, Devinn Lambert, Lizzie Hutton, Kelly Harris, Jenna Judge, Maria Nattestad, Nipam Patel, Sashank Pisupati

The field of evolutionary biology touches upon all other areas of the biological sciences, because every form of life and every biological process represent an ongoing evolutionary "experiment." Our aim in this course was to both discuss our understanding of the mechanisms of evolution and explore how evolutionary data can be used to further our understanding of various biological problems.

The course began with a discussion of the diversity of organisms that currently exist and methods for understanding the evolutionary relationships among these organisms. It then went on to study how paleontological data are collected and used to understand the history of life on earth and examined how DNA sequence data can be used to understand the evolutionary history of organisms, genes, and genomes. Within this molecular and genetic framework, the focus shifted to the mechanisms of evolutionary change and how variation within populations leads to the evolution of new species. Finally, the course ended with a discussion of how morphological changes are brought about through evolutionary changes in development and gene regulation. The hope was that this course would provide the student with a general overview of evolutionary biology and ways to use evolutionary data to gain further insight into all manner of biological problems. The course included a class favorite field trip and curator tour at the American Museum of Natural History in New York City.

SPECIAL COURSE
Optical Methods

January 7–9	**Attended by the entering class of 2015**
INSTRUCTOR	**Florin Albeanu**
GUEST LECTURERS	**Mikala Egeblad** **David Spector**
TEACHING FELLOWS	**Arkarup Banerjee** **Priyanka Gupta**

Optical imaging techniques are widely used in all areas of modern biological research. Our aim for this course was to give students an introduction into widely used basic and advanced optical methods. Given the experimental nature of the topic, a central aim of our course was to offer students a practical hands-on experience. This included both the use of commercially available systems and, more importantly, a primer on custom building and adapting optical setups to address specific biological needs.

Teaching Experience at the Dolan DNA Learning Center

Attended by the entering class of 2015

DIRECTOR	**David A. Micklos**	
INSTRUCTORS	**Amanda McBrien (Lead)**	**Erin McKechnie**
	Elna Gottlieb	**Bruce Nash**
	Katie McAuley	**Sharon Pepenella**

Science has an increasing role in society, and thus there is also an increasing need for biologists to educate nonscientists of all ages about biology. The WSBS doctoral program offers its students the opportunity to teach in the Laboratory's Dolan DNA Learning Center, where students teach laboratory courses to high school and middle school students. In so doing, they learn how to communicate with nonbiologists and to inspire and educate creative young minds. The teaching module entailed pairs of students teaching one morning or afternoon a week for 12 weeks. In the initial weeks, the Dolan DNA Learning Center instructors taught the Watson School students the didactic process; it was not until the fifth week that the graduate students taught on their own. At the end of the 12 weeks, the students were very excited about their teaching experience.

Laboratory Rotations

Attended by the entering class of 2015

ROTATION MENTORS	**Mickey Atwal**	**Leemor Joshua-Tor**	**Bo Li**	**Bruce Stillman**
	Anne Churchland	**Adam Kepecs**	**Zachary Lippman**	**Lloyd Trotman**
	Mikala Egeblad	**Justin Kinney**	**Robert Martienssen**	**Christopher Vakoc**
	Christopher Hammell	**Alex Koulakov**	**Stephen Shea**	**Anthony Zador**

The most important element of a doctoral education is learning to perform independent research. After the fall term courses, students participate in laboratory rotations. These provide students and faculty the opportunity to get to know each other and to explore possibilities for doctoral thesis research. At the end of each rotation, students make short presentations of their studies to the other students and their rotation advisors. These talks give students an opportunity to share their laboratory experiences and to practice giving scientific presentations. This year, 16 WSBS faculty members served as rotation mentors, some mentoring more than one student.

FALL CURRICULUM

The entering class of 2016 began the semester by attending boot camps in Molecular, Cellular, and Quantitative Biology to introduce them to the techniques and terminology that they will encounter in subsequent courses. The Molecular and Cellular Biology boot camp featured seven lectures from faculty members—Florin Albeanu, Hiro Furukawa, Dick McCombie, Pavel Osten, Lloyd Trotman, Linda Van Aelst, and Lingbo Zhang—and the Quantitative Biology boot camp featured lectures by Mickey Atwal and Justin Kinney.

CORE COURSES

The Leslie C. Quick, Jr. Core Course on Scientific Reasoning and Logic

INSTRUCTORS	**Linda Van Aelst (Lead)**	**Leemor Joshua-Tor**
	Alexander Gann	**Bo Li**
	Christopher Hammell	
GUEST LECTURERS	**Hiro Furukawa**	**Adrian Krainer**
	Z. Josh Huang	**Robert Martienssen**
	Justin Kinney	

In this core course, which forms the heart of the curriculum, students (1) acquire a broad base of knowledge about the biological sciences, (2) learn the scientific method, and (3) learn how to think critically. The initial four to five modules each focus on a different general theme. In each, students read an assigned set of research articles and, at the end of the module, provide written answers to a problem set that guides them through several of the articles.

Twice weekly, students attend lectures related to the module's topic that include concepts and fundamental information as well as experimental methods. The students meet among themselves to discuss the assigned papers not covered by the problem set. Each week, they spend an evening discussing the assigned articles with faculty. In the final module of the course, students participate in a mock study section in which real National Institutes of Health R01 grants are reviewed and critiqued. This allows students to evaluate the questions before the answers are known, evaluate routes toward discovery before knowing where they will end, and make critical judgments about how to proceed in the face of an uncertain outcome.

In 2016, the following were the module topics for this course.

Topic	**Instructor(s)**
Gene Expression	Alex Gann
Gene Regulatory Logic and the Construction of Multicellular Organisms: Insights from Flies, Plants, and Worms	Christopher Hammell
The Brain: Wiring, Plasticity, and Maladaptation	Bo Li
Macromolecular Structure and Function	Leemor Joshua-Tor
Study Section	Linda Van Aelst

The Darrell Core Course on Scientific Exposition and Ethics

INSTRUCTORS	**Mikala Egeblad (Lead)**	**David Jackson**
	Sydney Gary	**Charla Lambert**
GUEST LECTURERS	**Lisa Bianco**	**Alyson Kass-Eisler**
	Diane Esposito	**Richard Sever**
	Walter Goldschmidts	

VISITING LECTURERS

Keith Baggerly, M.D. Anderson Cancer Center
Robert Charrow, Greenberg Traurig, LLC
Graham Chedd, Alan Alda Center for Communicating Science
Susan Friedman, The Innocence Project
Avner Hershlag, North Shore University Hospital
Tung-Tien Sun, New York University

The 2016 Scientific Exposition and Ethics (SEE) core course was revamped this year with three distinct sections covering writing, oral communication, and ethics. As usual, the course hosted expert guest lecturers who covered topics including scientific presentations, ethical and legal responsibilities of scientists, DNA profiling and postconviction appeals, and reproductive genetics.

This core course offered instruction in the fundamental elements of scientific exposition—writing skills and public speaking—and ethics. The ability to communicate effectively and to appreciate the intricacies of ethical issues are essential skills for biologists; both subjects were taught in a series of example-based lectures and discussion groups. Writing skills included the fundamentals of modern scientific English and the organization and preparation of papers, research abstracts, and grant applications. Oral presentation skills were taught by instructors with different modes of presentation. Together with instructors, students critiqued formal seminar presentations at the Laboratory. Instruction and discussions about ethics included the ethical implications of biological discovery for society as well as the nature and boundaries of ethical behavior of scientists and their rights and responsibilities. A primary objective of the course was that students consider exposition and ethics integral aspects of scientific research.

Research Topics

ORGANIZERS	**Kimberley Geer**
	Alyson Kass-Eisler

This core course provided students with an in-depth introduction to the fields of research that the Laboratory scientists investigate. Students and faculty attended a weekly Research Topics seminar, at which faculty members and CSHL fellows presented their current research topics and methods of investigation each Wednesday evening over dinner. The students learned how to approach important problems in biology. These seminars, together with the annual fall Laboratory In-House Symposium, provided students with a basis for selecting laboratories in which to do rotations.

SPECIALIZED DISCIPLINES COURSES

The students in the Entering Class of 2016 took a total of four Specialized Disciplines courses this fall: *Genetics and Genomics, Quantitative Biology, Cancer,* and *Systems Neuroscience.*

Genetics and Genomics

September 6–23

INSTRUCTORS	**Zachary Lippman (Lead)** **Adam Siepel**
GUEST LECTURER	**Ullas Pedmale**
VISITING LECTURERS	**Jay Hollick**, Ohio State University **Guillaume Lettre**, University of Montreal

This course placed modern genetics and genomics into the context of classical genetics. History, technique, and perspective of genetic inference were described around four levels of analysis: forward genetics, natural genetic variants, gene interaction, and genomics. Emphasis was placed on integrating classical with modern questions of genetic analysis: How are genes mapped and "cloned"? How do gene mutations help to define biological processes? How are more complex traits genetically dissected into simpler (underlying) components? What concepts and techniques are used to organize genes into pathways and networks? What defines a gene, and what gene variation exists in natural populations? What are the functional consequences of gene variation, and how is it detected? How are genomes organized and coordinately regulated? How can genomic information be cataloged, organized, and mined? These questions and concepts were fleshed out using examples from the literature.

Quantitative Biology

September 26–October 21, November 14–December 9

INSTRUCTORS	**Adam Siepel (Lead)** **Mickey Atwal**
GUEST LECTURERS	**Justin Kinney** **Alexander Krasnitz** **Dan Levy**

Quantitative reasoning is a powerful tool to uncover and characterize biological principles, ranging from the molecular scale all the way to the ecological. With the advent of high-throughput technologies in genomics and neuroscience, it has become increasingly necessary for biological researchers to be able to analyze and interpret high-dimensional data and frame biological

hypotheses mathematically. To this end, this course aimed to equip the students with a basic training in computer programing, modern statistical methods, and physical biology. By the end of the course, students not only were able to answer many of the statistical questions that arise in data analyses, but also became familiar with the more complex techniques used by fellow computational biologists. Topics covered included probabilities, statistical fluctuations, Bayesian inference, significance testing, fluctuations, diffusion, information theory, neural signal processing, dimensional reduction, Monte Carlo methods, population genetics, and DNA sequence analyses. A common theme throughout the course was the use of probabilistic and Bayesian approaches.

Systems Neuroscience

October 2–29

INSTRUCTOR	**Stephen Shea (Lead)**
GUEST LECTURER	**Jessica Tollkuhn**

This course provides an overview of key aspects of neuroscience, with a focus on learning and plasticity from its cellular basis, through development, to systems and behavior. Both experimental and theoretical viewpoints were explored.

Cancer

October 11–27

INSTRUCTOR	**David Tuveson (Lead)**
GUEST LECTURERS	**Mikala Egeblad** **Raffaella Sordella** **David Spector** **Bruce Stillman** **Nicholas Tonks** **Lloyd Trotman** **Christopher Vakoc**
TEACHING ASSISTANT	**Lindsey Baker**

Cancer represents an increasing cause of morbidity and mortality throughout the world as health advances continue to extend the life spans of our populations. Although our basic understanding of cancer has increased considerably since 1971, when United States President Richard Nixon initiated the "War on Cancer," our ability to translate this knowledge into a health benefit for patients has been restricted to certain malignancies and often only temporarily. Importantly, specific

hypotheses developed from our knowledge of cancer biology can be tested in increasingly complex model systems ranging from cell culture to genetically engineered mouse models, and such investigations should prove invaluable in discovering new methodologies for the detection, management and treatment of cancer in humans.

At the conclusion of this course, students were able to elaborate an understanding of cancer as a pathobiological process that invades our bodies without offering any known benefit to the host; discuss how we diagnose cancer today; and contemplate how to replace the methods currently used to treat cancer. Students were also able to design tractable methods to investigate fundamental aspects of cancer biology and became familiar with translational approaches to defeating cancer. Topics covered in this course included biochemistry, epigenetics, immunology, resistance, growth control, microenvironment, noncoding RNA, and disease modeling. The implications of the biological findings on cancer prevention, diagnosis, and treatment were covered.

POSTDOCTORAL PROGRAM

PROGRAM DIRECTOR **Nicholas Tonks**

PROGRAM ADMINISTRATOR **Alyson Kass-Eisler**

The **Postdoctoral Liaison Committee** (PDLC), which is an elected group of postdoctoral fellows who communicate information and ideas between the administration and the postdoctoral community, continues to enhance CSHL's postdoctoral experience. The PDLC is essentially the voice of the community, and it holds regular meetings with Bruce Stillman, CSHL President. In 2016, new members Cristina Aguirre-Chen, Sonali Bhattacharjee, and Michael Regan joined returning PDLC members Sara Ballouz, Dhananjay Huilgol, Grinu Matthew, and Hillary Schiff. The PDLC organized a successful daylong retreat this year aimed at fostering networking and collaboration. The retreat took place at the Banbury Conference Center and included a presentation by Cesar Berrios-Otero of F1000Research, research talks by CSHL postdocs, a keynote address by Wei Ji Ma of the New York University Center for Neural Science Communication, and a "Mock Chalk Talk."

A daylong symposium called *Beyond Academia* was also organized by the PDLC and included panel discussions on the following: Commercializing Science; Professional Skill Building; Interviewing Skills; Careers in Consulting; Careers in Venture Capital and Equity Research; a Communication workshop hosted by Wei Ji Ma; Careers in Business Development in Big Pharma; and a Keynote: From Bench to Business.

This year, a new series, organized by the PDLC along with the CSHL Library, called *Perspectives on Science Careers*, was started and included the following guest speakers: Philipp Kaldis from the journal *Cell Division;* Keerthi Krishnan, University of Tennessee, Knoxville; Fatih Mercan, Scully, Scott, Murphy & Presser; JoAnna Klein, freelance journalist; Robert Weiss, Weill Cornell Medical School; and Andrew Whiteley, Enzo Biochem.

The PDLC also oversees and distributes funds provided by Dr. Stillman to two career development groups: the Career Development Program and the Bioscience Enterprise Club. These groups are primarily composed of postdoctoral fellows, but include some graduate students as well. Today's postdoctoral fellows face a number of challenges, including a very difficult and competitive job market. CSHL endeavors to prepare them to be competitive for a range of jobs beyond the merely academic. It is increasingly becoming CSHL's role to introduce the diversity of career opportunities available and to provide the tools postdocs need to prepare for these positions. As a result, a number of events were organized with the assistance of the PDLC and career development groups.

The **Career Development Program** (CDP) provides programming geared toward careers in academia. The successful *Conversations with Faculty* lecture series, where postdocs are provided career insights within an informal and interpersonal format, was held again and included a discussion on "Going on the Job Market—The Applicant Perspective," presented by current postdocs Keerthi Krishnan (Huang lab) and Michael Fegin (Tuveson lab). The CDP also organized two "Mock Chalk Talks" for postdocs preparing for a job interview in academia, and these were presented by Gonzalo Otazu (Albeanu lab) and Dawid Nowak (Trotman lab).

The CDP has also been bolstering connections with local colleges and universities to provide teaching opportunities for interested CSHL postdocs. Informal partnerships have now been established with Adelphi University, Molloy College, and Long Island University. In addition, selected postdoctoral fellows continue to participate in the Watson School as tutors either one-on-one or

in the classroom setting, thus providing CSHL postdocs with valuable experience in teaching and mentoring.

The **Bioscience Enterprise Club** (BEC) disseminates information about nonacademic careers to the CSHL postdoc community. Topic areas include biotechnology, intellectual property, law, regulatory affairs, and venture capitalism. This year, the BEC organized a visit to the Regeneron campus in Tarrytown, New York. Mr. Jonathan Kui, a scientist in the New York State Medical Examiners office and a Master's degree recipient from the Watson School, came to talk about careers in forensics. They also held talks and roundtable discussions with industry leaders including Dr. Ethan Perlstein, CEO of Perlara; Dr. George Yancopoulos, CSO at Regeneron; and Dr. Richard Ransohoff, VP at Biogen. They also held two panel discussions: The first panel was on World Health with Dr. Emilio Emini from the Gates Foundation, Dr. Dennis Carroll from USAID, and Dr. Marelize Gorgens from World Bank. The second panel was on Science Policy with Dr. Yvette Seger (FASEB) and Dr. Laurel Haak (ORCID).

An ice cream social was held on CSHL's campus to celebrate National Postdoc Appreciation week. These events provided a great opportunity for the community not only to join together for some fun, but also to network and learn about CSHL's ongoing programs.

All CSHL postdoctoral fellows and graduate students are enrolled in a special initiative of the New York Academy of Science (NYAS), known as the Science Alliance. The Science Alliance for graduate students and postdoctoral fellows is a consortium of universities, teaching hospitals, and independent research facilities in the New York City metro area that have formed a partnership with NYAS. The Alliance's aim is to provide career and professional development monitoring for postdoctoral fellows and graduate students in science and engineering, through a series of live events and a dedicated web portal. In addition, the Science Alliance offers graduate students and postdoctoral fellows the opportunity to network with their peers across institutions and with key leaders in academia and industry. The Science Alliance meetings and workshops this year discussed the following topics: Risky Business—The Future of Biopharmaceutical Innovation; Learn Basic Computing Skills to Be More Effective in the Lab; Scientists Teaching Science; Developing Scientists through Outreach: Defining Quality for the Scientist; "Speaking" Science: How to Communicate, Connect with Audiences; Making the Leap: A Non-Academic Career Planning and Job Search Boot Camp; Managing Difficult Situations in the Lab: The Essentials of Conflict Resolution; Career Fair for Scientists in Collaboration with INet NYC; Graduate Student and Postdoc Career Symposium at the NewYorkBio Conference; Minority Graduate Student Network Career Fair; The Ins and Outs of Scientific Publishing with Elsevier; Webinar: VISA Mechanisms for Scientists; Info session about the NSF Graduate Research Fellowship Program; From Scientist to CSO: Experiencing the Scientific Method as your Guide to Career Success; Clinical Research Management Course; Grantsmanship for Students and Postdocs: F30, F31, F32; and The Changing Landscape for Postdocs in the US: Potential Implications and Systemic Changes to Support Postdocs in the US Beyond the FLSA Ruling.

Since 2011, the Laboratory has held a Cancer Gene Discovery and Cancer Biology Postdoctoral Research Training Program with funding from the National Cancer Institute and National Institutes of Health (NIH). The program provides fellows with training in four integrated research areas: Cancer Genetics, Cancer Regulation and Cell Proliferation, Signal Transduction and Quantitative Biology. Last summer, we submitted a competitive renewal of this funding to the NIH. The grant received positive reviews and was renewed through 2021.

Finally, an important measure of our postdoctoral program's success is the ability of postdoctoral fellows to secure positions after they complete their training. In 2016, our fellows accepted positions at Cooley LLP; Vanderbilt University; University of Tennessee, Knoxville; SUNY Old Westbury; LIU Brooklyn; Umea University; and Institute of Biophysics, Chinese Academy of Sciences.

UNDERGRADUATE RESEARCH PROGRAM

PROGRAM DIRECTOR **Anne Churchland**

PROGRAM ADMINISTRATOR **Kimberly Creteur**

Established more than 50 years ago, the CSHL Undergraduate Research Program (URP) plays a major role in providing and setting the standard for meaningful hands-on undergraduate research training in biology. The 10-week program begins the first week of June. To ensure a smooth transition into the Laboratory community and research, during the first week, the students attend various orientations and receive a guided historical tour of campus and all the facilities and resources available to them. The URPs work, live, eat, and play among CSHL scientists, and they have a very busy academic and social calendar throughout the remaining 9 weeks of the summer. The students receive training in scientific research, science communication, career preparation, and bioinformatics and computational biology, all while interacting socially with fellow program participants and members of the CSHL community in formal and informal activities. Some of the 2016 activities included a pizza party with Dr. and Mrs. Watson, dinner with Dr. and Mrs. Stillman, volleyball games, designing the URP T-shirt, competing in the annual Scavenger Hunt, and the ever-famous URP versus PI volleyball match.

The students' scientific development is the most important component of the program. Therefore, in the beginning of the summer, the URPs write an abstract and present a talk on their proposed research. The URP participants work alongside scientists and become increasingly independent throughout the summer. Concluding the program in August, each URP student prepares a final report and presents his or her results in a 15-minute talk at the URP Symposium. As in previous years, the program directors and faculty mentors were highly impressed with the accomplishments of the URP students.

2016 Undergraduate Research Program participants

The following 20 students, selected from 1087 applicants, took part in the 2016 program:

Toby Aicher
Advisor: **Molly Hammell**
Funding: **National Science Foundation Scholar**
Investigating drug resistance in melanoma using single-cell RNA sequencing.

Daniel Barabasi
Advisor: **Anne Churchland**
Funding: **University of Notre Dame URP Scholar**
Extracting more, and more accurate, data from 2-photon calcium imaging.

Julia Bassell
Advisor: **Adrian Krainer**
Funding: **Libby Fellowship/William Shakespeare Fellowship**
5′-Splice-site selection in GT vs. GC splice sites.

Sara Blagburn
Advisor: **Florin Albeanu**
Funding: **National Science Foundation Scholar**
Developing a psychometric curve for odor intensity via a novel two-alternative forced choice protocol in head-fixed mice.

Alissa Castelberry
Advisor: **Thomas Gingeras**
Funding: **National Science Foundation Scholar**
Processing of Y5 RNA by cancer cell exosomes.

Debotri Chatterjee
Advisor: **David Jackson**
Funding: **Von Stade Fellowship/Burroughs Wellcome Fellowship**
Understanding the role of the G-protein β subunit in plant cell death.

Erin DeNardo
Advisor: **Doreen Ware**
Funding: **National Science Foundation Scholar**
Interpretation of gene structure changes in *Oryza sativa* from a single gene to a population.

Zhiwei Ding
Advisor: **Z. Josh Huang**
Funding: **James D. Watson Undergraduate Scholar/Dorcas Cummings Scholar**
Mapping the distribution of a genetically specified subpopulation of pyramidal neurons projecting to ventromedial striatum in mice.

Chris Giuliano
Advisor: **Mikala Egeblad**
Funding: **Robert H.P. Olney Fellow/Garfield Fellowship**
An antimetastatic role of lysyl oxidases through matrix metalloprotease inhibition in pancreatic cancer.

Benjamin Harris
Advisor: **Mickey Atwal**
Funding: **National Science Foundation Scholar**
Pan-cancer analysis of ectopic germline gene expression.

Daniel Hawkins
Advisor: **W. Richard McCombie**
Funding: **National Science Foundation Scholar**
Long-read sequencing and copy-number analysis.

Isaiah Holloway Jr.
Advisor: **Christopher Vakoc**
Funding: **William Townsend Porter Foundation Scholar**
Is the *CERS4* gene necessary for JAK2-mutated AML cell proliferation?

Ashley Kyalwazi
Advisor: **Stephen Shea**
Funding: **National Science Foundation Scholar**
Parvalbumin network and neuroplasticity in the auditory cortex.

Jingyi (Jenny) Ma
Advisor: **Camila dos Santos**
Funding: **Former URP Fund Scholar/30th Anniversary URP Scholar**
Using CRISPR-CAS9 to investigate the epigenetic regulation of mammary stem cells.

Ajay Nadig
Advisor: **Adam Kepecs**
Funding: **National Science Foundation Scholar**
Signatures of prediction error in cortical VIP interneurons.

Timothy Nolan
Advisor: **Alexei Koulakov**
Funding: **National Science Foundation Scholar**
Optimization of the short-time fourier transform spectrogram for machine learning objectives.

Sevahn Vorperian
Advisor: **Bruce Stillman**
Funding: **Alfred L. Goldberg Fellowship**
Using CRISPR screening to identify domain dependencies of ORC1 and CDC6 in diploid and cancerous cells.

Katelyn Wilensky
Advisor: **Jessica Tollkuhn**
Funding: **National Science Foundation Scholar**
Using sex differences to study the relationship between genes and behavior.

Kaitlin Williams
Advisor: **David Tuveson**
Funding: **Joan Redmond Read Fellowship/Howard Hughes Medical Institute Scholar**
Inhibition of Myc slows the proliferation of KRAS-driven pancreatic cancer organoids.

Lorna Wills
Advisor: **Lloyd Trotman**
Funding: **James D. Watson Undergraduate Scholar**
The effect of targeting Phlpp2 on cell proliferation and pAkt and Myc signaling pathways.

SUMMER RESEARCH INTERNSHIP FOR MEDICAL STUDENTS

PROGRAM ADMINISTRATOR **Jessica Gotterer**

Through the CSHL and Northwell Health affiliation, a summer internship program has been created to give students entering into their second year of medical school from the Donald and Barbara Zucker School of Medicine at Hofstra/Northwell the opportunity to spend a summer experiencing basic research in a CSHL lab and attending relevant seminars. To date, students have been offered positions in labs focusing on cancer and the genetics of human disease. Students commit 8–10 weeks (roughly June–August) during the summer following their first year of course work to full-time research in a CSHL lab. The students complete a research project and present their work at the annual "Scholarship Day" at Zucker School of Medicine the following fall.

The following students took part in the 2016 program:

STUDENT	CSHL MENTOR
Adam Ouellett	Camila dos Santos
Narayanan Sadagopan	Mickey Atwal
Alexis Tchaconas	Michael Wigler
Michael Wotman	David Tuveson

PARTNERS FOR THE FUTURE

PROGRAM DIRECTOR **David Jackson**

PROGRAM ADMINISTRATOR **Marygrace Navarra**

The Partners for the Future Program, established by Dr. James Watson in 1990, provides an opportunity for talented Long Island high school students to have hands-on experience in biomedical research at Cold Spring Harbor Laboratory. This highly competitive program is open to Long Island high school students in their junior year. Each high school science chairperson may nominate up to three students. The top candidates are interviewed by CSHL scientists. Students selected for the program are paired with a scientist mentor and spend a minimum of 10 hours per week, September through March of their senior year, conducting original research. At the conclusion, the students present their projects to an enthusiastic audience of the students, scientific mentors and colleagues, CSHL administrators, parents, and teachers. Although the students learn a great deal about modern biology and state-of-the-art research techniques, the main advantage of the program is that they are exposed to day-to-day life in a lab. Interacting with scientists and support staff, the students learn to define and pursue a research goal while solving problems that may occur along the way.

The 2016–2017 Partners for the Future were chosen from among 51 nominations. They are listed below.

Name	School	Lab	Mentor
Alexander Oruci	Chaminade High School	Robert Martienssen	Rowan Herridge
Anand Subudhi	Cold Spring Harbor High School	Adam Kepecs	James Sturgill
Annie Tao	Division Avenue High School	Christopher Hammell	Christopher Hammell
Jahan Rahman	Sayville High School	Camila dos Santos	Camila dos Santos
Jonah Wu	Syosset High School	Pavel	Eric Szelenyi
Kevin Chen	Farmingdale High School	Jessica Tollkuhn	Jessica Tollkuhn
Kimia Khaledi	Oyster Bay High School	David Tuveson	Tobiloba Oni
Linie Li	Massapequa High School	Robert Martienssen	Jean-Sebastien Parent
Owen McEvoy	Oyster Bay High School	David Tuveson	Ela Elyada
Philip Jang	Syosset High School	Raffaella Sordella	Raffaella Sordella
Roberta Geffrard	Uniondale High School	Alea Mills	Alea Mills
Shaina Zafar	H. Frank Carey High School	David Jackson	Qingyu Wu
Sonali Persaud	East Meadow High School	Christopher Hammell	Christopher Hammell
Urooba Abid	Farmingdale High School	Robert Martienssen	Sonali Bhattacharjee

MEETINGS & COURSES PROGRAM

ACADEMIC AFFAIRS

The Cold Spring Harbor Laboratory meetings and courses program attracts scientists from all over the world to communicate, learn, and critique the latest ideas and approaches in the biological sciences. The meetings and courses program at the Laboratory attracted strong attendance in 2016, with 7250 meeting participants and more than 1450 course participants (trainees, teaching, and support faculty). The Cold Spring Harbor Asia program included 19 conferences and attracted more than 3500 participants, bringing the anticipated year-end total for both U.S.- and China-based programs to almost 12,300.

The Laboratory held 27 academic meetings this year, which brought together scientists from around the world to discuss their latest research. The spring meeting season culminated in the 81st Cold Spring Harbor Symposium, which addressed Targeting Cancer, reflecting the enormous research progress achieved in recent years and providing a broad synthesis of the current state of the field. The Symposium attracted almost 490 participants, including notable scientists such as Angelika Amon, Joan Brugge, Gerard Evan, Ronald Evans, Tyler Jacks, Carl June, William Kaelin, Scott Lowe, Tak Mak, Elaine Mardis, Benjamin Neel, Kornelia Polyak, Charles Sawyers, Michael Stratton, Craig Thompson, Harold Varmus, Irving Weissman, and Eileen White, to name but a few. Dissemination includes the proceedings of the Symposium, published each year by the CSHL Press, and videotaped interviews with leading speakers by editors and journalists attending the Symposium, now available on our Leading Strand YouTube channel. The Symposium therefore reaches a much wider audience nationally and internationally than can possibly attend.

Cold Spring Harbor Laboratory meetings are unique in that organizers are encouraged to select talks from abstracts that are submitted on the basis of scientific merit, ensuring that the latest findings will be presented and that young scientists will have the chance to describe their work. The year 2016 saw the continuation of regular community meetings such as the Biology of Genomes, Retroviruses, Epigenetics and Chromatin, Translational Control, Axon Guidance, Synapse Formation and Regeneration, and Mechanisms of Aging. A successful new meeting on Transposable Elements was started, reflecting the long tradition of active research on this topic initiated by Barbara McClintock and the emergence of new concepts and approaches in the field. The meeting on Protein Homeostasis in Health and Disease included special talks by Costa Georgopoulos, F. Ulrich Hartl, Richard Morimoto, and Peter Walter to highlight the 25th anniversary of the first meeting in 1991 related to heat shock and the role of molecular chaperones. The CSHL Genentech Center Conferences on the History of Molecular Biology and Biotechnology this year addressed HIV/AIDS Research: Its History & Future. Notable speakers included David Baltimore, Francoise Barrè-Sinoussi, Emilio Emini, Anthony Fauci, Robert Gallo, Marty St. Clair, and Harold Varmus, among a program that included pioneers of the science of retroviruses, the discoverers of human retroviruses, the developers of HIV/AIDS therapeutics, and other key figures in the fight to understand and control this global pandemic, which, tragically, has persisted for more than 35 years. Partial support for individual meetings is provided by grants from the National Institutes of Health, the National Science Foundation, other foundations, and companies. Core support for the meetings program is provided by the Laboratory's Corporate Sponsor Program.

The Courses program covers a diverse array of topics in molecular biology, neurobiology, structural studies, and bioinformatics. Instructors update their course curricula annually, invite new speakers who bring a fresh perspective, and introduce new techniques and experimental approaches based on student feedback and progress in the field. New courses on Immersive Approaches to Biological Data Visualization and on Organotypic and Next-Generation Culture Methods were piloted this year in parallel with well-established courses in genetics, neuroscience, plant biology, and bioinformatics, while the Cellular Biology of Addiction course (now in its 16th year) was held at Gonville and Caius College in Cambridge, United Kingdom, for the first time. New techniques—for example, genome editing using tools such as CRISPR and super-resolution microscopy—are

introduced as methodologies develop and evolve. We strongly encourage each course to include the latest technical and conceptual developments in their respective fields. Instructors, lecturers, and assistants come from universities, medical schools, research institutes, and companies around the world to teach at Cold Spring Harbor. Students include advanced graduate students, postdoctoral trainees, and principal investigators and senior scientists from around the world.

Instructors, course assistants, and course lecturers come from universities, medical schools, research institutes, and companies around the world to teach at Cold Spring Harbor. Their excellence and dedication make the course program work so well. We would especially like to thank Drs. Michael Bereman, Michelle Cilia, Ileana Cristea, Mary Dunlop, Marc Gartenberg, Amy Herr, Gareth Howell, Dannielle Kennedy, Mark Lewandoski, Howard Salis, Dylan Taatjes, Marja Timmermans, Deneen Wellik, and Torsten Wittmann, whose exemplary teaching and leadership of their respective courses have benefitted so many young scientists.

Grants from a variety of sources support the courses. The core support grants provided through the Helmsley Charitable Trust and Howard Hughes Medical Institute are critical to our course program. The courses are further supported by multiple awards from the National Institutes of Health and the National Science Foundation, and additional support for individual courses is provided from various foundations and other sources. The courses also depend on equipment and reagents that are loaned or donated by a large number of companies, partnerships that are invaluable in ensuring that the courses offer training in the latest technologies (see below).

Now in its seventh year of operation, the Cold Spring Harbor Asia (CSHA) program, under the direction of Dr. Maoyen Chi, is headquartered at the Suzhou Dushu Lake Conference Center in the Suzhou Innovation Park high-technology suburb. In 2016, 17 scientific conferences were held in Suzhou and one summer school in Shanghai, while the program also arranged two meetings on immunology and plant biology in Awaji, Japan. CSHA's scientific program is designed for scientists from the Asia/Pacific region, who make up >75% of attendance, and include symposia and meetings, training workshops, and occasional Banbury-style discussion meetings. This program is described in more detail in a separate Annual Report.

Special events included the first Double Helix Day event on DNA Damage, Repair, and Mutation in February, Celebrating Francis—The Francis Crick Centennial in May, and a Pediatric Oncology symposium and Bioentrepreneur Networking event in October, as well as numerous local area one-day retreats, including several affiliated with the Feinstein Institute and Northwell Health. Although distinct from our regular academic program, these events attract significant numbers of leaders and individuals associated with biomedicine and bio-business from the tri-state area and beyond.

The meetings and courses program staff comprises a diverse team of talented professionals who handle the complexities of database design, programming, web and multimedia design, educational grants management, marketing and recruitment, conference and course administration, audiovisual and digital design services, and other activities. The year 2016 saw significant progress in ongoing efforts to update our systems, with the smooth introduction of integrated online credit card processing to accept fees. We said goodbye to Jim Duffy and Marge Stellabotte, after many years of service to CSHL, Katherine Bradley-Richardson to a new career in teaching, and course logistics coordinator Diana Monno to pastures new. We also welcomed several new staff in 2016, including Maggie Bernhardt, Meredith Cassuto, Cat Donaldson (in late 2015), Heather Johnson, Patricia Klieger, and Vivian Vuong.

David Stewart
Executive Director, Meetings and Courses Program
President, Cold Spring Harbor Asia

Terri Grodzicker
Dean of Academic Affairs

CSH ASIA SUMMARY OF CONFERENCES

Dates	*Title*	*Organizers*
April 18–22	Ubiquitin Family, Autophagy, and Diseases	Ivan Dikic, Jianping Jin, Xiaobo Qiu, Keiji Tanaka, Li Yu
April 25–29	Liver Diseases and Tumorigenesis: From Bench to Bedside	Gen-Sheng Feng, Hongyang Wang, Jessica Zucman-Rossi
May 9–13	Chromatin, Epigenetics, and Transcription	Genevieve Almouzni, Hiroyuki Sasaki, Yang Shi, Bing Zhu
May 16–20	Development, Function and Disease of Neural Circuits	Rui Costa, Zhigang He, Anatol Kreitzer, Minmin Luo, Yimin Zou
June 13–17	DNA Metabolism, Genomic Stability, and Human Diseases	Judith Campbell, Daochun Kong, Binghui Shen, Stephen West
September 5–9	Telomere and Telomerase	Eric Gilson, Ming Lei, Roger Reddel, Jerry Shay
September 19–23	Cancer and Metabolism	Navdeep Chandel, Sheng-Cai Lin, (Hongyuan) Rob Yang
September 26–30	Microbiology and the Environment	Dusko Ehrlich, Jack Gilbert, Nan Qin, Ting Zhu
October 3–6	Frontiers of Immunology in Health and Disease	Xuetao Cao, Richard Flavell, Tadatsugu Taniguchi
October 10–14	Systems Biology of Gene Regulation and Genome Editing	Xiaole Shirley Liu, Huck-Hui Ng, Len Pennacchio, Edward Rubin
October 17–21	Probing Neural Networks with Light: Imaging Structure and Function in the Living Brain	Mark Hübener, Mu-Ming Poo, Carlos Portera-Cailliau, Yi Zuo
October 24–28	Bone and Cartilage: From Development to Human Diseases	Hongwei Ouyang, Thomas Pap, Sakae Tanaka, Yingzi Yang
October 28–November 1	Systems Medicine Approach to Global Infectious Disease	Barry Bloom, Genhong Cheng, George Gao, Robert Modlin
November 7–11	Frontiers in Single-Cell Genomics	John Marioni, Nicholas Navin, Paul Robson, Fuchou Tang, Angela Wu
November 14–18	RNA Biology	Anne Ephrussi, Masatoshi Hagiwara, Adrian Krainer, Yijun Qi
November 28–December 2	Synthetic Biology	George Guo-Qiang Chen, Sang Yup Lee, Eriko Takano, Lingchong You
November 29–December 2	Latest Advances in Plant Development and Environmental Response	Philip Benfey, Wolf Frommer, Keiko Sugimoto, Jian-Kang Zhu, Cyril Zipfel
December 5–9	Novel Insights into Glia Function and Dysfunction	Shumin Duan, Mami Noda, Mengsheng Qiu, Bruce Ransom, William Richardson
December 12–16	Biology and Function of Extracellular Vesicles: Exosomes, Microvesicles, and Beyond	Yang Song Gho, Andrew Hill, Jan Lötvall, Kenneth Witwer, Xu Zhang

81ST COLD SPRING HARBOR LABORATORY SYMPOSIUM ON QUANTITATIVE BIOLOGY

Targeting Cancer

June 1–5 488 Participants

ARRANGED BY **Scott Lowe**, Memorial Sloan Kettering Cancer Center
Kornelia Polyak, Dana-Farber Cancer Institute
David Stewart, Cold Spring Harbor Laboratory
Bruce Stillman, Cold Spring Harbor Laboratory
Eileen White, The Rutgers Cancer Institute of New Jersey

Cancer is a deadly disease that will afflict one out of every two to three people in the developed world in their lifetime. Recent years have produced transformative discoveries revealing the mechanisms behind the development and progression of cancer and the implementation of targeted cancer therapies in the clinic based on these discoveries. We have defined the mutational landscape of many cancers and with that the incipient driver mutations. Sophisticated and elegant models now exist for many types of cancer, and we understand in great detail the changes in signaling and gene expression that promote cancer and have begun to unravel the antitumor immune response. Cell death and survival mechanisms used by cancer cells have been identified and therapeutically exploited, and we have revealed processes that dictate cancer cell growth, repair, and proliferation and have the means to inhibit these processes. We have also seen advances in the mechanistic understanding, implementation, and deployment of cytotoxic cancer therapy. Collectively, these advances are improving both the quality of life and the overall survival of cancer patients.

Lowe lab

C. June, B. Neel

L. Maiorino, E. Bruzas

The decision to focus the 81st Cold Spring Harbor Laboratory Symposium on Targeting Cancer reflected the enormous research progress achieved in recent years and provided a broad synthesis of the current state of the field, setting the stage for future discoveries and applications. Implications of how the underlying science can drive improvements in diagnostic, prognostic, and therapeutic approaches were a major theme throughout the Symposium. Previous Symposia that have included significant aspects of cancer research have occurred on a roughly five- to 10-year cycle, notably but not limited to Genes and Mutations (1951); Cellular Regulatory Mechanisms (1961); Transcription of Genetic Material (1970); Tumor Viruses (1974); Viral Oncogenes (1979); The Cell Cycle (1991); The Molecular Genetics of Cancer (1994); Biological Responses to DNA Damage (2000); and Molecular Approaches to Controlling Cancer (2005).

Major themes and topics highlighted at the 2016 Symposium included Cancer Genes and Genomes (genome stability, chromatin, and epigenetic regulation); Cancer Pathways (signaling pathways and networks); Tumor Cell Biology (cells of origin, metabolism, autophagy, and senescence); Cancer Growth and Progression (microenvironment, stroma/niche, tumor evolution, and metastasis); Innate and Adaptive Immune Responses (inflammation and immunotherapies);

J. Sheltzer, C. Gunes

V. Weaver, E. McKenna

M.C. Simon, C. Van Dang

D. Saffran, C. Fonseca

and Enabling Technologies (single-cell sequencing, imaging, genetic screens, genome editing, and organoids). Opening night speakers included Elaine Mardis (Nationwide Children's Hospital Research Institute) on translating cancer genomics into therapeutic target identification and vaccine design; Craig Thompson (Memorial Sloan Kettering Cancer Center), who spoke about an emerging treatment paradigm exploiting cancer metabolism; Carl June (University of Pennsylvania) on the use of genetic editing to generate synthetic lethal T cells; and Tyler Jacks (David H. Koch Institute for Integrative Cancer Research at MIT), who addressed engineering the cancer genome. Charles Sawyers (HHMI/Memorial Sloan Kettering Cancer Center) delivered an outstanding Dorcas Cummings Lecture on "Reflections on Precision Medicine and Cancer Moonshots" for Laboratory friends, neighbors, and Symposium participants prior to the annual Symposium dinners. Joan Brugge (Harvard Medical School) provided a masterful Summary at the conclusion of the Symposium immediately before the final banquet.

This Symposium was attended by almost 490 scientists from universities around the country and internationally, and the program included 55 invited presentations and more than 200 poster presentations. To disseminate the latest results and discussion of the Symposium to a wider audience, attendees were able to share many of the Symposium talks with their colleagues who were

M. Stratton

J. Brugge, R. Sever

U. Moll, G. Lozano, R. Iggo

J. Witkowski, W. Kaelin

unable to attend using the Leading Strand video archive. A collection of interviews by Gemma Alderton (*Nature Reviews Cancer*), Paula Kiberstis (*Science*), Mirna Kvajo (Cell Press), Elizabeth McKenna (*Cancer Discovery*/AACR), Richard Sever (CSHL Press), and Jan Witkowski (CSHL Banbury Center) with leading experts in the field were arranged during the Symposium and distributed as free video from the Cold Spring Harbor Symposium interviews website.

Major support was provided by the CSHL-Northwell Health Partnership, with additional support provided by Genomic Health and Fluidigm. Financial support from the corporate sponsors of our meetings program is essential for these Symposia to remain a success, and we are most grateful for their continued support.

PROGRAM

Introduction
B. Stillman, *Cold Spring Harbor Laboratory*

Microenvironment and Metastasis
Chairperson: A. Amon, *Massachusetts Institute of Technology, Cambridge*

RAS-Driven Cancers/Synthetic Lethals
Chairperson: M. Stratton, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom*

Breast Cancer Targets
Chairperson: G. Lozano, *University of Texas M.D. Anderson Cancer Center, Houston*

Cancer Stem Cells/Host Interactions
Chairperson: W. Kaelin, *HHMI/Dana-Farber Cancer Institute, Boston, Massachusetts*

Epigenetics
Chairperson: H. Varmus, *Weill Cornell Medical College, New York*

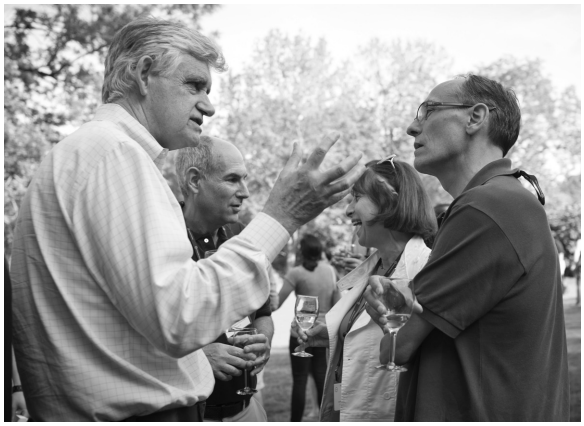
Screens and Resistance
Chairperson: M.C. Simon, *University of Pennsylvania Medical School, Philadelphia*

Dorcas Cummings Lecture
C.L. Sawyers, *HHMI/Memorial Sloan Kettering Cancer Center, New York*

Cancer Targets, CTCs, and Heterogeneity
Chairperson: K. Vousden, *Cancer Research UK Beatson Institute, Glasgow, United Kingdom*

Metabolism
Chairperson: M. McMahon, *University of Utah, Huntsman Cancer Institute, Salt Lake City*

Summary
J. Brugge, *Harvard Medical School, Boston, Massachusetts*



B. Stillman, W. Herr

MEETINGS

Double Helix Day: DNA Damage, Repair, and Mutation

February 29 75 participants

ARRANGED BY **David Stewart**, Cold Spring Harbor Laboratory
Bruce Stillman, Cold Spring Harbor Laboratory

This special annual celebration ("Double Helix Day") was intended to coincide with the actual date that James Watson and Francis Crick discovered the double helix structure of deoxyribonucleic acid (February 28, 1953) in Cambridge, England. Each year, a theme related to DNA science is explored through a series of review-style talks aimed at a broad scientific audience. How cells repair damaged DNA and how mutations arise were selected as the themes of the 2016 celebrations. The program commenced in the early afternoon on Monday, February 29, and concluded with an early evening reception and celebratory double helix feast in Blackford Hall.

D. Stewart, J. Watson

PROGRAM

Invited Speakers

J. Watson, *Cold Spring Harbor Laboratory:* Discovering the Double Helix: Going for Gold!

L. Samson, *Massachusetts Institute of Technology, Cambridge:* The pros and cons of DNA repair.

T. Kunkel, *National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina:* Generating and repairing mismatches during eukaryotic DNA replication.

J. Haber, *Brandeis University, Waltham, Massachusetts:* How cells fix their broken chromosomes.

S. Nik-Zainal, *Wellcome Trust Sanger Institute, United Kingdom:* The genome: An archaeological record of mutational processes.

S. Nik-Zainal, L. Samson

J. Haber, B. Stillman

Systems Biology: Global Regulation of Gene Expression

March 15–19 246 participants

ARRANGED BY **Barak Cohen**, Washington University School of Medicine in St. Louis, Missouri
Christina Leslie, Memorial Sloan Kettering Cancer Center, New York, New York
John Stamatoyannopoulos, University of Washington, Seattle

S. Pletcher, V. Gorbunova, M. Hansen

Systems biology is a discipline in which investigators aim to understand the properties of cellular networks by using high-throughput, technology-intensive, and computational approaches. One of the most actively researched areas in systems biology is the global regulation of gene expression, which coordinates complex metabolic and developmental programs in organisms. This 10th conference, like the ones in previous years, captured the continuing rapid progress in this exciting field. The 4-day meeting, which featured 65 talks and 119 poster presentations, covered a broad range of topics. The speakers, poster presenters, and other conference attendees were composed of a mix of students, postdocs, and PIs at all levels. Several attendees noted through their "tweets" the gender balance of the invited and selected speakers, which the organizers are particularly proud of. Two keynote speeches were exciting highlights of the meeting. The first was presented by Dr. Wendy Bickmore, who showed exciting new results documenting the link between genome organization and transcriptional regulation. Dr. Bickmore described how an enhancer located 2 million base pairs from its target gene loops to its cognate promoter. This combination of high-throughput genomics coupled to high-resolution microscopy promises to play a large part in our field in the years to come. Dr. John Lis gave the second keynote address and described his lab's work on polymerase pausing as a key step in the regulation of gene expression. Through innovative new techniques, such as the GRO-seq and PRO-seq methods, Dr. Lis has shown that pausing is both widespread in eukaryotic genomes and highly regulated. His insights into the dynamics of different steps of gene activation are an important reminder of the complexity of the systems the field seeks to understand.

As always, this year's meeting featured many new techniques utilizing next-generation sequencing. The maturation of massively parallel reporter gene assays (MPRAs) was demonstrated by many presentations that featured this methodology. Other exciting developments were methylase footprinting, which allows investigators to observe the different configurations of DNA-binding proteins on regulatory regions in vivo, and the continued progress toward two-step ChIP-seq protocols that allow investigators to assay the co-occurrence of histone modifications in vivo. Investigators also introduced innovative new computational frameworks for analyzing high-throughput data, particularly model-driven approaches for dissecting GRO-seq and PRO-seq data, and principled approaches to analyzing single-cell expression data. The rampant spread of genome editing by CRISPR-Cas9 technologies was another exciting development, one that is providing the power to test healthy numbers of the hypotheses generated from high-throughput studies.

Aside from technological developments, several scientific themes emerged that promise to loom large in the next few years. Several investigators raised the question of what the exact difference is, if any, between "enhancers" and "promoters." Opinions are still sharply divided on this important

D. Miller, L. Booth, A. Brunet, C. Murphy

T. von Zglinicki, M. Santra, A. de Graff

question. Likewise, the relationship of enhancers to lncRNAs was also discussed several times. There is likely to be robust debate as to whether lncRNAs constitute a distinct functional class of genes or whether they are simply active enhancers with nearby splice sites, or both. The explicit coupling of splicing to transcription came up in several talks and may be one of the mechanisms that provide directionality to nascent transcripts, which are produced bidirectionally in most places but are highly biased in one direction in cases with a functional neighboring splice site. These and other issues promise to keep this meeting an exciting and dynamic forum for the next several years.

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health.

PROGRAM

cis-Regulation
Chairpersons: R. Gordan, *Duke University, Durham, North Carolina;* N. Ahituv, *University of California, San Francisco*

Transcription Factors
Chairpersons: B. van Steensel, *Netherlands Cancer Institute, Amsterdam;* D. Schubeler, *Friedrich Meischer Institute, Basel, Switzerland*

Chromatin and 3D Architecture
Chairpersons: D. Arnosti, *Michigan State University, East Lansing;* A. Akhtar, *Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany*

Keynote Speaker
W. Bickmore, *University of Edinburgh*

Evolution and Variation
Chairpersons: B. Engelhardt, *Princeton University, New Jersey;* Y. Gilad, *University of Chicago, Illinois*

RNA and Its Regulation
Chairpersons: S. Eddy, HHMI/Harvard University, Cambridge, Massachusetts; J. Ahringer, *Wellcome Trust/ Cancer Research UK Gurdon Institute, Cambridge, United Kingdom*

Keynote Speaker
J.T. Lis, *Cornell University*

Cellular Decision-Making
Chairpersons: R. Mitra, *Washington University, St. Louis, Missouri;* A. Regev, *Broad Institute, Cambridge, Massachusetts*

Emerging Technologies

Chairpersons: W. Greenleaf, *Stanford University, California;* S. McCarroll, *Harvard Medical School, Boston, Massachusetts*

Networks and Global Analyses
Chairpersons: S. Liu, *Dana-Farber Cancer Institute, Boston, Massachusetts;* D. Pe'er, *Columbia University, New York*

Evolutionary Biology of *Caenorhabditis* and Other Nematodes

March 30–April 2 73 participants

ARRANGED BY **Scott Baird**, Wright State University, Dayton, Ohio
Marie Delattre, CNRS, Paris, France
Erik Ragsdale, Indiana University, Bloomington
Adrian Streit, MPI for Developmental Biology, Tübingen, Germany

This meeting was attended by researchers representing 10 countries. The keynote speaker was Ralf Sommer, who discussed the integrative approach used in his laboratory—combining evolutionary developmental biology, ecology, and population genetics to address questions in evolutionary biology using the example of nematodes of the genus *Pristionchus*. The scientific sessions and chairpersons are listed below under "Program." Among the investigators making oral presentations, there were several from cell biology and parasitology labs who had not previously attended an evolutionary biology meeting. These investigators expressed pleasure about the opportunity to present their work in an evolutionary context and indicated that they hoped to return to this meeting in the future. The duration of lunch and coffee breaks was ideal for discussions. One informal session was organized before the first afternoon session on Friday to discuss the sequencing strategies for *Caenorhabditis* nematodes, the proceedings of which were led by Mark Blaxter. In addition to the 46 oral presentations, 20 posters were presented in two poster sessions. The new facility for poster presentations was very highly appreciated by the participants and encouraged extensive discussions both at the posters and in small spontaneous round-table meetings on the veranda. During a meeting planning session, reasons for the low attendance, compared with the same meeting in 2012, were discussed. It appears that the tight funding situation of many labs and the unusually high number of other meetings with overlapping scopes this year were important factors. Furthermore, the high costs caused in part by the high airfares in the week after Easter and the strong U.S. dollar prevented many potential participants, in particular those from overseas,

M. Delattre, E. Ragsdale, S. Baird, A. Streit

A. Mortazav, M. Macchietto

J. Projecto Garcia, B. Wharam

from attending the meeting. Among the participants, extensive enthusiasm was expressed about the 2016 meeting at Cold Spring Harbor, new organizers were recruited, and it was agreed to meet again in 2018 at the Wellcome Genome Campus in Cambridge, England (as in 2010 and 2014) and then hopefully return to Cold Spring Harbor in 2020.

PROGRAM

Tools
Chairperson: A. Diaz, *University of Cambridge, United Kingdom*

Keynote Speaker
Chairperson: R.J. Sommer, *Max Planck Institute for Developmental Biology, Tübingen, Germany*

Symbiosis
Chairperson: Z. Zhao, *Hong Kong Baptist University, Hong Kong*

Body Plan Evolution
Chairperson: R. Sommer, *Max Planck Institute for Developmental Biology, Tübingen, Germany*

Genomics I
Chairperson: I. Yanai, *Technion Israel Institute of Technology, Haifa, Israel*

Early Embryonic Divisions and Evolution of Reproduction
Chairperson: B. Meyer, *HHMI/University of California, Berkeley*

Epigenetics and Small RNAs
Chairperson: H. Goodrich-Blair, *University of Wisconsin, Madison*

Competition and Adaptation
Chairperson: A. Smythe, *Virginia Military Institute, Lexington*

Gene Expression
Chairperson: A. Jex, *Walter & Eliza Hall Institute and University of Melbourne, Australia*

Reproductive Isolation
Chairperson: F. Landmann, *CRBM-CNRS UMR 5237, Montpellier, France*

Meiosis, Chromosome Biology
Chairperson: P. Sarkies, *Imperial College London, United Kingdom*

Pathogens and Parasites
Chairperson: H. Teotonio, *École Normale Supérieure, Paris, France*

Genomics II
Chairperson: E. Andersen, *Northwestern University, Evanston, Illinois*

Neuronal Circuits

April 6–9 203 participants

ARRANGED BY **Florian Engert**, Harvard University, Boston, Massachusetts
Carl Petersen, Ecole Polytechnique Federale de Lausanne, Switzerland
Kristin Scott, University of California, Berkeley

F. Engert, K. Scott, C. Petersen

The anatomical and functional connectivity of neurons underlies the simplest of behavioral decisions to the most complex cognitive tasks. Unprecedented advances in the ability to monitor the activity of large neural populations, to rapidly and precisely manipulate the activity of neurons with optogenetic approaches, and to trace connectivity with synaptic resolution are providing a new systems-level view of neural circuit processing. Coupling recent technological advances with computational modeling and quantitative behavioral approaches has the potential to reveal how the dynamic computations of the brain underlie thought and behavior. This meeting highlights recent progress in elucidating neural circuit function across nervous systems.

In keeping with the original goals of this meeting, researchers working in a variety of different organisms and systems were brought together to share their scientific and technological advances, focusing on recent advances in neuronal circuit processing. The meeting included six slide sessions covering a range of topics in sensory systems, motor control, behavior, cortical processing, methods, and plasticity—and a very interactive poster session.

For the sixth meeting of this kind, the response of the field was very enthusiastic. The meeting brought together 203 participants from all over the world, most of whom made either oral or poster presentations. Particularly impressive were the large numbers of students (38%) and postdoctoral fellows (29%) participating in the meeting. Invited talks were given by Susumo Tonegawa, Massachusetts Institute of Technology; Loren Frank, University of California, San Francisco; Gwyneth Card, Janelia Farm Research Campus/HHMI; Barry Dickson, Janelia Farm Research Campus/HHMI; Nirao Shah, University of California, San Francisco; Anatol Kreitzer, University of California, San Francisco; Gul Dolen, Johns Hopkins University; Kay Tye, Massachusetts Institute of Technology; Georg Keller, Friedrich Miescher Institute; David Kleinfeld, University of California, San Diego; Valentina Emiliani, CNRS and Paris Descartes University; Botond Roska, Friedrich Miescher Institute; Massimo Scanziani, University of California, San Diego; and Misha Ahrens, Janelia Farm Research Campus/HHMI. The meeting provided an important forum for the exchange of ideas and approaches and helped scientists in the field get the most up-to-date information, as well as enabling them to meet, network, and establish collaborations. Based on the uniformly enthusiastic comments of the participants, the intensity of the oral and poster sessions, the productive question and answer sessions and the large crowds that stayed up late to discuss science, the meeting was a great success.

A highlight of the meeting program is the Larry Katz Memorial Lecture, thanks to the generous donation of many colleagues in the field, which recognizes an outstanding research contribution by a graduate student. The overall quality of the 13 nominees was truly spectacular, representing an international cohort of exceptional graduate students. The selection committee was composed of the three current organizers and two past organizers, Liqun Luo, Stanford/HHMI, and Ed

S. Gandhi, G. Keller

K. Tschida, R. Peixoto

Calloway, Salk Institute. The committee unanimously selected Neir Eshel, Harvard University, as the 2016 Lecturer. He gave a wonderful presentation about his graduate research. A history of the Larry Katz prize, including background on Katz and a list of previous awardees, should be included in the 2018 program booklet.

This meeting was funded in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health.

PROGRAM

Learning and Plasticity
Chairperson: K. Tye, *Massachusetts Institute of Technology, Cambridge*

Behavior
Chairperson: K. Scott, *University of California, Berkeley*

Basal Ganglia
Chairperson: N. Shah, *University of California, San Francisco*

Sensorimotor
Chairperson: C. Petersen, *Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland*

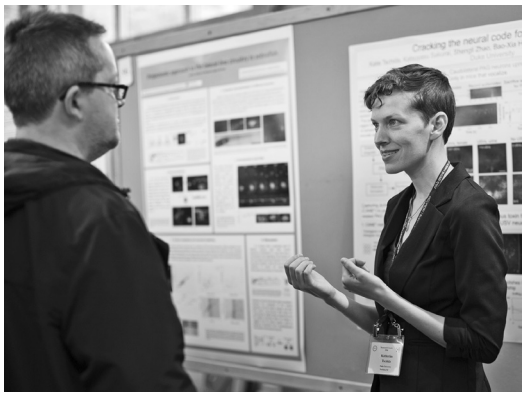
Sensory Processing I
Chairperson: F. Engert, *Harvard University, Cambridge, Massachusetts*

Sensory Processing II
Chairperson: F. Engert, *Harvard University, Cambridge, Massachusetts*

Technology
Chairperson: A. Kreitzer, *University of California, San Francisco*

Larry Katz Lecture
N. Eshel, *Harvard University, Cambridge, Massachusetts*

J. Clemens, T. Jovani



S. Shea, K. Tschida

The PARP Family and ADP-Ribosylation

April 13–16 105 participants

ARRANGED BY **Michael O. Hottiger**, University of Zurich, Switzerland
W. Lee Kraus, University of Texas Southwestern Medical Center, Dallas
Karolin Luger, University of Colorado, Boulder

W. Lee Kraus, K. Luger, M.O. Hottiger

This second meeting on PARPs and ADP-ribosylation followed the initial meeting titled The PARP Family and Friends: Gene Regulation and Beyond, which was held in 2014. The meetings in 2014 and 2016 built upon (1) previous international PARP meetings, which were organized without the support of a major conference organization, (2) growing interest in the basic biology of PARPs and ADP-ribosylation, (3) an expanded understanding of the PARP family, and (4) the growing clinical promise of PARP inhibitors as therapeutic agents.

The recent PARP meeting featured 36 talks (one keynote, 19 invited, 16 selected from the abstract; 15 women, 21 men, one underrepresented minority), as well as a poster session. Both the talks and posters provided a view of the field that differed considerably from those represented in the past, showing how far the field has advanced recently. The 2016 meeting focused much greater attention on the PARP monoenzymes and their diverse biology, away from the historical focus on nuclear PARP polyenzymes (e.g., PARP-1) and their role in DNA damage detection and repair. Other aspects of PARP biology that received greater attention than in the past included functional interactions with RNA, host–virus interactions, chromatin, and gene regulation. Methodological advances, including those related to PARP genomics and proteomics, were well represented in the presentations.

As in 2014, the faculty used this forum to meet and discuss issues important to the field, such as nomenclature (e.g., PARP vs. ARTD; monoPARP and polyPARP vs. PARP monoenzyme and

D. Filippov, J. Rudolph

W. Zhi, S. Zha

L. Mariotti, K. Jividen

G. Mbalaviele, M.O. Hottiger

PARP polyenzyme), promoting and expanding the field, recruiting and retaining trainees in the field, and the field's relationship to other NAD^+-related fields (e.g., NAD^+ metabolism, Sirtuins). Importantly, the faculty meeting led to an initiative to standardize the assays, protocols, and reagents used by the PARP field. Dr. Peter Bai, University of Debrecen, Hungary, is organizing this initiative, with the goal of producing guidelines for the study of PARPs and poly(ADP)-ribosylation. These guidelines should help to standardize experiments and analyses, allowing for more facile comparison of results generated by different labs. This meeting also led to the commissioning of a review article by the editors of *Genes and Development*, being written by W. Lee Kraus and colleagues, which will highlight the most novel results and directions of this changing field, as presented and discussed at the meeting.

Overall, the sense among the attendees (determined from informal surveys) was that this meeting marked an important turning point for the field and should be continued. The organizers decided that it would be good to have some turnover of the organizers, with a clear plan for succession. These discussions are ongoing.

This meeting was funded in part by Ribon Therapeutics and the National Cancer Institute, a branch of the National Institutes of Health.

PROGRAM

Structure and Molecular Mechanisms
Chairperson: W.L. Kraus, *University of Texas Southwestern Medical Center, Dallas*

Signaling and Molecular Mechanisms in Cancer
Chairperson: K. Luger, *HHMI/University of Colorado, Boulder*

Mono(ADP-Ribosyl) Transferases
Chairperson: M. Hottiger, *University of Zurich, Switzerland*

Keynote Speaker
S.-I. Imai, *Washington University School of Medicine, St. Louis, Missouri*

Enzymology, Metabolism, and the ADP-Ribosyl Proteome
Chairperson: J. Pascal, *Université de Montréal, Canada*

Nuclear Functions: Chromatin, Gene Regulation, and DNA Repair
Chairperson: F. Dantzer, *Université de Strasbourg, France*

Physiology, Disease, and Therapeutics
Chairperson: S. Smith, *Skirball Institute, New York University School of Medicine, New York*

Protein Homeostasis in Health and Disease

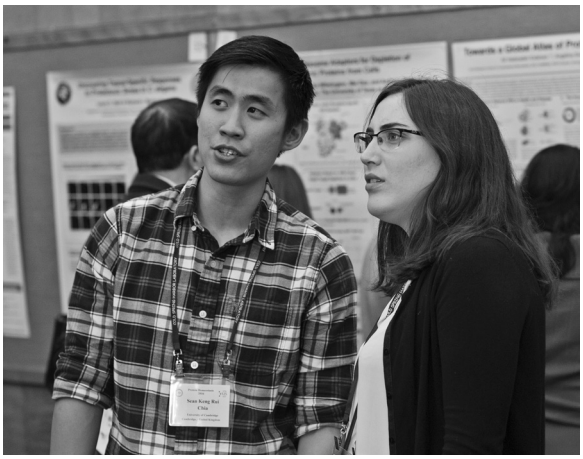
April 18–22 265 participants

ARRANGED BY **Judith Frydman**, Stanford University, California
F. Ulrich Hartl, Max Planck Institute for Biochemistry, Planegg, Germany
Harm Kampinga, University Medical Center/University of Groningen, The Netherlands
Richard Morimoto, Northwestern University, Evanston, Illinois
Peter Walter, HHMI/University of California, San Francisco

This was the 25th anniversary of this topic on the CSHL meeting schedule and the 35th anniversary of the very first meeting at CSHL on the topic of heat shock proteins. The topics and chairpersons presenting at the meeting are listed below under Program. In addition to the traditional poster sessions, three new aspects to this meeting were Rapid Fire Presentations of 2-minute single slide talks selected from the abstracts before the poster sessions, the Editorial Panel Discussion with representatives from *Science, Nature Structural & Molecular Biology, Genome Science,* and the *Journal of Cell Biology*, and lunch with the speakers. All three additions were well-attended and will continue in future meetings.

The high attendance, large number of posters, and lively audience participation in the sessions and in the informal activities that followed are testament to the success of the meeting and to the successful balance of the sessions. The field of molecular chaperones—protein folding in vivo and in the cell, proteostasis networks, and diseases of protein misfolding—is constantly evolving and this meeting has been central in leading the edges of this field to continue to be the cornerstone meeting for this rich and exciting field. The 2016 meeting reflected the many branches of biology and medicine that are now understood to be regulated by molecular chaperones and cell stress signaling pathways.

F.U. Hartl, P. Walter, J. Frydman, R. Morimoto, H. Kampinga



S. Keng, R. Chia

B. Dong, A. Jaeger

Among the cutting-edge themes explored in this meeting were presentations on events at the ribosome as an integrator of protein biogenesis at the birth of proteins communicating with the heat shock response to enhance chaperone levels and novel mechanisms to extract nascent chains with defects. Also exciting were systems-level studies focused on understanding the circuitry that maintains protein homeostasis within and between cells and tissues that impact on aging and disease. The integration of information between the mitochondria, the endoplasmic reticulum, the cytosol, and nucleus has revealed new mechanisms for spatial quality control in folding, aggregation, and degradation mechanisms. Talks on the effect of disruption of these networks by cancer or aging and the consequence of age-associated neurodegenerative diseases emphasized the delicate balance between folding and misfolding required to maintain a healthy proteome and the opportunities to enhance protein homeostasis to prevent aggregation.

This meeting was funded in part by the National Institute on Aging, a branch of the National Institutes of Health.

PROGRAM

Synthesis and Folding of Proteins
Chairperson: R. Morimoto, *Northwestern University, Evanston, Illinois*

Regulation and Properties of Proteostasis Networks
Chairperson: L. Sistonen, *Turku Centre for Biotechnology, Åbo Akademi University, Finland*

J. Lis (*center*), J. Joutsen (*right*)

A. Thapaliya (*left*), R. Yan (*center*)

G. Gaglia, C. Moore

Y. Cong, J. Frydman, T. Lopez, Y. Zang, M. Jin, K. Dalton

RAPID FIRE PRESENTATIONS

Spatial and Organellar Quality Control
Chairperson: J. Frydman, *Stanford University, California*

Degradation Mechanisms
Chairperson: J. Brodsky, *University of Pittsburgh, Pennsylvania*

Novel Properties and Functions of Chaperones I
Chairperson: E. Duerling, *Universität Konstanz, Germany*

Novel Properties and Functions of Chaperones II
Chairperson: U. Jakob, *University of Michigan, Ann Arbor*

Functional and Pathogenic Aggregates
Chairperson: A. Bertolotti, *MRC Laboratory of Molecular Biology, Cambridge, United Kingdom*

Aging and Diseases of Proteostasis
Chairperson: J. Kelly, *The Scripps Research Institute, La Jolla, California*

B. Adamson, J. Abisambra

J. Buchner, F. Tippel

Gene Expression and Signaling in the Immune System

April 26–29 259 participants

ARRANGED BY **Diane Mathis**, Harvard Medical School, Boston, Massachusetts
Stephen Nutt, The Walter & Eliza Hall Institute, Parkville, Victoria, Australia
Alexander Rudensky, Memorial Sloan Kettering Cancer Center, New York, New York

S. Nutt, D. Mathis, A. Rudensky

This meeting, held for the eighth time this year, registered more than 250 participants and included many first-time attendees, who participated in a highly engaging and interactive 4-day meeting. This meeting is unique in the calendar of immunology conferences in its focus on molecular and biochemical aspects of the development and function of the immune system in all its diversity. In addition, this meeting attracts a broad range of scientists who use the immune system as a model to study basic principles of biological regulation. Talks were presented by a mix of invited speakers and investigators selected from a group of more than 180 submitted abstracts, with more than 20 junior investigators being selected to give talks. As is the tradition of this meeting, most speakers focused almost exclusively on their unpublished work.

There were many exciting talks such as the presentation by Ajay Chawla (University of California, San Francisco) discussing the body's coping mechanisms for prolonged exposure to low temperature. The remarkable conclusion was that acclimatization to cold is exquisitely controlled by components of the innate immune system. Morgan Huse (Memorial Sloan Kettering Institute, New York) revealed how killer T cells use coordinated mechanical force to open perforin pores and lyse target cells. Another theme was the use of newer technologies such as single-cell RNA sequencing and cellular barcoding to investigate cellular heterogeneity and differentiation potential. One selected talk from Rahul Satija (New York University) applied single-cell genomics techniques on a massive scale, sequencing the RNA from 30,000 human cord blood progenitors in an attempt understand the differentiation process. The application of these techniques generated many questions from the audience and vigorous discussion in the coffee breaks. In research from a different perspective, Akiko Iwasaki (Yale) investigated how the body responds to viral infection in the nervous system and found that both CD4 T cells and antibodies were required for effective viral control. Most strikingly, the antibodies could not access the immune privileged neural tissue without help from IFNγ produced by CD4 T cells that relaxed vascular permeability, thus providing a new role for these helper cells in immune responses.

Such talks highlighted the relevance of the approaches discussed at this meeting, not just to immunology, but also to the broader scientific community. Oral presentations were supplemented by an afternoon poster session; the posters were extremely vibrant and the session provided a forum for all meeting participants (including many graduate students and postdocs) to share their most recent data. In addition, attendees took advantage of an engaging piano recital and an evening performance by the Yale-based music group The Cellmates, led by Richard Flavell.

This meeting was funded in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health.

F. Alt, M. Krangel

A. Rudensky, P. Cockerill, A. Rao

PROGRAM

Myeloid Lineage Cells and Innate Lymphocytes
Chairperson: M. Colonna, *Washington University School of Medicine, St. Louis, Missouri*

Transcriptional and Posttranscriptional Control of Gene Expression
Chairperson: S. Nutt, *The Walter and Eliza Hall Institute, Parkville, Australia*

Chromatin Regulation and Epigenetic Influences
Chairperson: A. Rao, *La Jolla Institute for Allergy and Immunology, California*

Lymphocyte Differentiation and Antigen Receptor Gene Expression
Chairperson: D. Schatz, *HHMI/Yale University School of Medicine, New Haven, Connecticut*

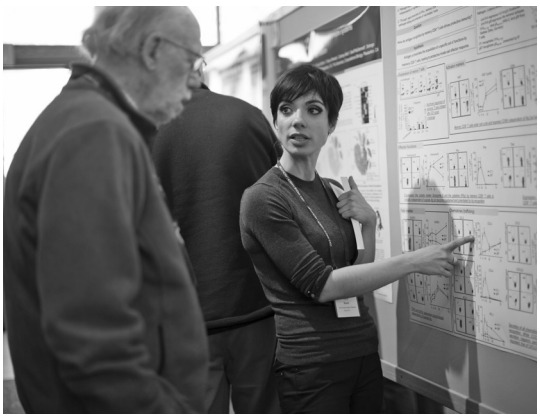
Signaling
Chairperson: A. Rudensky, *Memorial Sloan Kettering Cancer Center, New York*

Innate Sensors and Regulation
Chairperson: S. Hur, *Harvard Medical School, Boston, Massachusetts*

Regulation of Immune Cell Function
Chairperson: D. Mathis, *Harvard Medical School, Boston, Massachusetts*

Modes of Intra- and Intercellular Communications
Chairperson: R. Medzhitov, *HHMI/Yale University, New Haven, Connecticut*

S. Koralov, K. Rajewsky



M. Scharf, M. Boutet

Nuclear Organization and Function

May 3–7 225 participants

ARRANGED BY **Edith Heard**, Curie Institute, Paris, France
Martin Hetzer, Salk Institute for Biological Studies, La Jolla, California
David L. Spector, Cold Spring Harbor Laboratory

S. Wente, M. Hetzer

This meeting was one of the most exciting in the series, with important new breakthroughs on mechanistic relationships between nuclear organization, genome function and gene regulation. The presentations were all stimulating and followed by rich discussions, and the poster sessions were very animated. Many new insights into chromosome architecture and how it relates to gene expression were presented, based on chromosome conformation capture data as well as imaging. For example, HiC data at single-cell resolution was presented, revealing that topologically associating domains (TADs) can be dynamic during the cell cycle (Peter Fraser). The role of chromosome folding in long-range gene regulation by enhancers was also a recurrent theme, with key insights into how TAD organization enables precise spatiotemporal dynamics of gene expression (Job Dekker) and how perturbation of this organization can be linked to diseases such as cancer (Kadi Akdemir). The functional importance of the architectural protein, CTCF, in TAD structure formation using a powerful degron approach was presented. A novel role for the YY1 protein as a transcriptional regulator and mediator of long-range interactions was also proposed. Several new computational models for chromosome architecture based on "C" data were discussed, including a model for chromosome loop formation through "loop extrusion" via cohesin and CTCF (Geoffrey Fudenberg). Moving from chromosomes to nuclear structure, impressive progress in understanding of nuclear pore structure and nuclear lamina was reported at the meeting. Using combinations of X-ray crystallography (Andre Hoelz), super-resolution imaging (Ellenberg lab), and biochemistry (Karsten Weiss), as well as genetics approaches, several presentations provided unprecedented structural and functional information on nuclear pore complexes (NPCs) and nuclear lamins. Several talks provided insights into the precise protein–protein interactions and events underlying the structural organization (Hetzer lab) and assembly of NPCs (Martin Beck). There was also impressive work on the multiple different facets of the lamin proteins and how they form the architecture of the nuclear lamina, thanks to separate but overlapping filamentous meshworks (Robert Goldman). The participation of the nuclear lamina in tissue building and maintenance was discussed (Yixian Zheng). There was also an increasing appreciation of the interplay between RNA and nuclear organization. Joe Gall gave an inspiring overview of RNA, lampbrush chromosomes, and transcription in the oocyte. Exciting new insights into the dynamics of mRNA metabolism and transport through the nuclear pore were also presented (Yaron Shav-Tal). Recent powerful approaches exploring RNA structure in vivo, such as the SHAPE technique, were also presented (Weeks lab). Such information reveals how long noncoding RNAs such as XIST, which is involved in X-chromosome-wide silencing, consist of a mixture of both dynamic and stably structured regions that may have important implications for their

J. Brickner, A. Stephens

D.L. Spector, R. Goldman

protein and nucleic-acid-binding partners and their functions. John Lis presented exciting data using PRO-seq to characterize transcriptional changes that occur upon heat shock. The visualization and measurement of the dynamics of gene expression and regulation in living cells represented another of the exciting frontiers at this meeting. The frequency of promoter firing and how and when enhancers meet gene promoters in different contexts were explored using sophisticated imaging approaches as well as mathematical modeling (Jason Brickner, Dan Larson). For housekeeping genes, a continuum of transcriptional states was found, with a slowly fluctuating initiation rate converting the gene between different levels of activity, interspersed with extended periods of inactivity. This unexpected variation could provide a spectrum of possible expression states that can be explored in different contexts.

All in all, this meeting was one of the most stimulating to date on the topic of nuclear organization and function. It covered a wide range of topics, and yet there were many recurrent themes. It brought together a fantastic set of speakers, both established leaders in their fields as well as more junior participants, in a format that was conducive to active discussions.

PROGRAM

Chromatin Organization
Chairpersons: J. Dekker, *University of Massachusetts Medical School, Worcester*; P. Fraser, *Babraham Institute, Cambridge, United Kingdom*

Nucleocytoplasmic Transport and Nuclear Pore Complexes
Chairpersons: M. Beck, *European Molecular Biology Laboratory, Heidelberg, Germany;* Y. Zheng, *Carnegie Institution for Science, Baltimore, Maryland*

M. Capelson, M. Lusi

A. Vihervaara, A. Ford

J. Dekker, K. Maeshima

Y. Shav-Tal, R.H. Singer

Epigenetics
Chairpersons: W. Bickmore, *MRC Human Genetics Unit, Edinburgh, United Kingdom;* M. Merkenschlager, *Imperial College London, United Kingdom*

Biology of Nuclear RNAs and RNA Processing
Chairpersons: J. Gall, *Carnegie Institution for Science, Baltimore, Maryland;* Y. Shav-Tal, *Bar-Ilan University, Ramat-Gan, Israel*

The Nuclear Membrane and Nuclear Lamina
Chairpersons: R. Goldman, *Northwestern University Feinberg School of Medicine, Chicago, Illinois;* S. Wente, *Vanderbilt University Medical Center, Nashville, Tennessee*

Special Talk
R. Conroy, *National Institute of Biomedical Imaging and Biotechnology, NIH, Bethesda, Maryland*

Transcription
Chairpersons: J. Lis, *Cornell University, Ithaca, New York;* D. Larson, *National Cancer Institute, NIH, Bethesda, Maryland*

Nuclear Function in Disease and Development
Chairpersons: T. Misteli, *National Cancer Institute, NIH, Bethesda, Maryland;* J. Skok, *New York University School of Medicine, New York*

Gene Expression
Chairpersons: K. Plath, *University of California, Los Angeles*; A. Taddei, *Institut Curie/CNRS, Paris, France*

J. van Bemmel, R. McCord

I. Rempel

The Biology of Genomes

May 10–14 529 participants

ARRANGED BY **Ewan Birney**, EBI/EMBL, United Kingdom
Jonathan Pritchard, Stanford University, California
Molly Przeworski, Columbia University, New York, New York

E. Birney, M. Przeworski, J. Pritchard

This meeting marked the 29th annual gathering of genome scientists at the CSH Laboratory. Participants from around the world attended the meeting, with more than 350 abstracts presented describing a broad array of topics relating to the functional analysis, comparative characterization, and interpretation of genomes from diverse organisms. The scope and applicability of genome science continues to grow. Even after almost three decades of this meeting, the talks were fresh with fierce competition for abstracts. The meeting featured talks on a wide variety of genomics (see Program below). Session chairs ensured a reasonable balance of genders, and there was a strong focus on younger graduate students and postdoc presentations. The talks ranged from ecological/evolutionary biology (e.g., Nancy Chen on the Florida Scrub Jay) through to cancer genomics (Serena Nik-Zanial on breast cancer genomics) to common diseases such as schizophrenia (Jeff Barrett). Talks included engaging examples of genomics expanding in scope (e.g., Elinor Karlisson on dog compulsive disorders). The impact of genomics in translation studies was exemplified by talks from Heidi Rehm, Matt Hurles, and Daniel McArthur, in rare disease, and industry

B. Graveley, S. Salzberg, R. Green

T. Aneichyk

X. Zhao, A. Quinlan

D. Mishmar, L. Orlando

perspective from Sally John, Biogen. Three poster sessions allowed for comprehensive discussions for abstracts that did not make the talks; they were well attended.

These topics gave evidence of DNA sequence variation and its role in molecular evolution, population genetics and complex diseases, comparative genomics, large-scale studies of gene and protein expression, and genomic approaches to ecological systems. Both technologies and applications were emphasized.

All sessions were well attended, stretching the capacity of the CSHL facilities. The keynote presentations were from Emmanuelle Charpentier, who talked about her pioneering work behind CRISPR-Cas9, and Neil Shubin, who has bridged paleontology, developmental biology, and public understanding of science tied into his successful TV program.

The ELSI (Ethical, Legal, and Social Implications) panel was moderated by Dave Kaufman, and the area of discussion was "Participant Rights to Their Sequence Data: The Pros, Cons, and Pragmatics of Returning the Incidental Genome." Panelists included Misha Angrist, Duke University; Jason Bobe, Icahn School of Medicine at Mount Sinai; Mildred Cho, Stanford School of Medicine; and Wendy Chung, Columbia University.

This meeting was funded in part by the National Human Genome Research Institute, a branch of the National Institutes of Health; Fluidigm; Illumina; and Swift Biosciences.

C. Seoighe, M. Taylor

E. Quincy Rose, A. Cagen

J. Pritchard, D. Gaffney

E. Segal, C. Vinson

PROGRAM

Population Genomics
Chairpersons: N. Elde, *University of Utah, Salt Lake City;* J. Pickrell, *New York Genome Center, New York*

Functional Genomics
Chairpersons: W. Greenleaf, *Stanford University, California;* D. Winter, *Weizmann Institute of Science, Rehovot, Israel*

Computational Genomics
Chairpersons: E. Segal, *Weizmann Institute of Science, Rehovot, Israel;* A. Walczak, *CNRS/ENS, Paris, France*

Cancer/Medical Genomics
Chairpersons: S. Nik-Zainal, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom;* H. Rehm, *Harvard Medical School, Boston, Massachusetts*

ELSI Panel and Discussion
D. Kaufman, *National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland*

Evolutionary and Non-Human Genomics
Chairpersons: F. Jones, *Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany;* L. Orlando, *University of Copenhagen, Denmark*

Translational Genomics and Genetics
Chairpersons: M. Hurles, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom;* S. John, *Biogen Idec, Inc., Cambridge, Massachusetts*

Guest Speakers
E. Charpentier, *Max Planck Institute for Infection Biology, Germany;* N. Shubin, *University of Chicago, Illinois*

Genetics of Complex Traits
Chairpersons: C. McBride, *Princeton University, New Jersey;* S. Raychaudhuri, *Brigham and Women's Hospital, Boston, Massachusetts*

Celebrating Francis: The Francis Crick Centennial

May 16 90 participants

ARRANGED BY **Alex Gann**, Cold Spring Harbor Laboratory
Mila Pollock, Cold Spring Harbor Laboratory
Jan Witkowski, Cold Spring Harbor Laboratory

P. Moore, J. Witkowski

Francis Crick was born on June 8, 1916, in Weston Favell, a small village close to Northampton in England. Crick's contributions to science were many, not least with Jim Watson determining the double-helical structure of DNA. Through Watson, Crick had a long association with Cold Spring Harbor, and it was decided that so significant an anniversary should be celebrated here.

The meeting was intended to cover the broad sweep of Crick's life, including topics that are often neglected, dominated as his life was by the double helix, the genetic code, and studies of consciousness. Crick's war work for the Admiralty was on defense against mines, and Bob Olby described what Crick did and the importance of his contributions to the war effort. When he returned to the Cavendish Laboratory at Cambridge, he made seminal discoveries in the theory of X-ray crystallography, notably helical theory. Steve Harrison described this work and reminded us of the landmark paper Crick and Watson published on "The Structure of Small Viruses" in 1956. There followed two personal reminiscences from Jim Watson and Michael Crick, Francis's son. Watson described life in the Cavendish and Cambridge in the early 1950s, while Michael told us of the background of the remarkable letter Francis wrote to Michael on March 19, 1953, just after the discovery had been made. Matthew Cobb took us through the next epic contribution made by Crick, his role in deciphering the genetic code, and Tommy Poggio and Pat Churchland covered the long period that Crick spent at the Salk Institute. Finally, we returned to learning more about Crick's personality with a charming presentation by Kindra Crick, Francis's granddaughter, and closing remarks from Jim Watson.

Jim Watson addressing the gathering in Nichols-Biondi Hall

K. Crick, D. Casper, R. Oldby, M. Ridley

PROGRAM

M. Ridley: The life and times of Francis Crick.
R. Olby, *University of Pittsburgh, Pennsylvania:* Crick, World War II, and the Admiralty.
S. Harrison, *Harvard University, Cambridge, Massachusetts:* Helices, coiled-coils, and small viruses: Crick's adventures in x-ray crystallography.
J. Watson, *Cold Spring Harbor Laboratory:* Working with Francis.
Michael Crick, *Seattle, Washington:* On learning of the double helix.
M. Cobb, *University of Manchester, United Kingdom:* The genetic code.
T. Poggio, *Massachusetts Institute of Technology, Cambridge:* Crick's contributions to neuroscience.
P. Churchland, *University of California, San Diego:* Crick and the Salk Institute.
Kindra Crick, *Portland, Oregon:* My grandfather.
J. Watson, *Cold Spring Harbor Laboratory:* Closing remarks

J. Dahlberg, J.E. Darnell

E. Gardner, M. Crick

The Cell Cycle

May 17–21 175 participants

ARRANGED BY **Angelika Amon**, Massachusetts Institute of Technology, Cambridge
Jacqueline Lees, Massachusetts Institute of Technology Center for Cancer Research, Cambridge
David Toczyski, University of California, San Francisco

J. Lees, D. Toczyski, A. Amon

This 13th biannual meeting was held this year at Cold Spring Harbor Laboratory. This conference is the most important meeting on cell cycle control and is internationally recognized for its ability to bring together scientists who study cell cycle regulation in a broad range of model organisms and in humans. The meeting began with a keynote talk from David Morgan, a leader in the field of cell cycle control and author of the leading textbook in the field, *The Cell Cycle*. He provided an overview of the key discoveries of the cell cycle and his lab's recent discoveries regarding the regulation of the ubiquitin ligase APC, the central regulator of mitosis. Especially exciting were his lab's findings on the effects of cell size on the strength of the spindle assembly checkpoint, a surveillance mechanism that controls the activity of the APC. After the keynote talk, the evening session continued with five talks focused on cytokinesis and cell abscission. The remainder of the meeting was organized around six lecture sessions and two poster sessions, which focused on the major stages of the cell division cycle. Many of these sessions emphasized long-standing problems in cell cycle control, including cell size control of the cell cycle, as well as how surveillance mechanisms impinge on the cell cycle machinery. This year, there was also a significant focus on the intersection of development and cell division. Studies on mammalian cell

J. Lees, J. Skotheim, M. Kõivomägi, J. Turner, J. Benanti (*facing away*)

A. Seoane, A. Mizrak, D. Morgan, Y. Barral

cycle control and how deregulation of the cell cycle leads to cancerous transformation were also the topic of many talks and posters. There were several interesting reports about the mechanisms whereby oncogenes and tumor suppressor genes control the cell cycle, providing key insights into how genetic alterations in these genes promote tumorigenesis. The approaches and biological systems represented at this meeting were very broad and provided an excellent overview as to how the field attacks critical questions in this important area of research. As always, all model systems used for cell cycle analysis were represented, emphasizing the importance of basic research in studying cell cycle control and the high conservation of cell division control mechanisms. It was a landmark meeting for the cell cycle field, and the participants continue to look forward to equally exciting meetings in future years.

T. Martins, E. Logarinho

B. Mierzwa, J. Dare Shih

K. Engeland, C. Mages

T. Hawa-Racine

PROGRAM

Keynote Speaker
D.O. Morgan, *University of California, San Francisco*

Cytokinesis
Chairperson: D. Gerlich, *Institute of Molecular Biotechnology, Vienna, Austria*

Replication and DNA Damage
Chairpersons: H. Ulrich, *Cancer Research UK, London Research Institute, United Kingdom;* T. Paull, *HHMI/ University of Texas, Austin*

Chromosome Biology
Chairpersons: D. Koshland, *University of California, Berkeley;* C. Haering, *European Molecular Biology Laboratory, Heidelberg, Germany*

Cell Size and Cell Entry Decisions
Chairpersons: J. Skotheim, *Stanford University, California;* J. Walter, *Harvard Medical School, Boston, Massachusetts*

Metaphase to Anaphase
Chairpersons: S. Biggins, *Fred Hutchinson Cancer Research Center, Seattle, Washington;* R. Heald, *University of California, Berkeley*

G_1/S
Chairpersons: J. Sage, *Stanford University, California;* B. Schulman, *St. Jude Children's Research Hospital, Memphis, Tennessee*

Cell Growth and Development
Chairpersons: T. Orr-Weaver, *Whitehead Institute, Cambridge, Massachusetts;* I. Hariharan, *University of California, Berkeley*

Mitosis Regulation
Chairpersons: F. Chang, *Columbia University, New York;* S. Lens, *University Medical Centre Utrecht, The Netherlands*

Retroviruses

May 23–28 403 participants

ARRANGED BY **Kate Bishop**, The Francis Crick Institute, London, United Kingdom
Louis Mansky, University of Minnesota Twin Cities, Minneapolis

K. Bishop, L. Mansky

A characteristic feature of this meeting remains the high number of repeat attendees who come each year to initiate and maintain friendships and collaborations. Both of the organizers are regular attendees who, like many of the participants, have "grown up" at this meeting within its distinct community of scientists.

The organizers chose to recognize two prominent and highly respected senior retrovirologists as keynote speakers. Dr. Ronald Swanstrom presented an overview of his many contributions to basic retrovirology and HIV/AIDS research over the course of his career—with a sense of humor and fascinating personal anecdotes. His talk was very well received by the attendees, many of whom found it inspirational. Dr. Alan Rein also presented a very well received and humorous talk that included a picture from his first visit to the Cold Spring Harbor Labs in 1961 and highlighted work done over six decades of retroviral research. Both keynote talks discussed retroviral research in the context of wider virological and general biological research.

Three annual prizes, originally instituted and endowed by the community of regular attendees with assistance from CSH Meetings staff, were also awarded. These prizes memorialize three of our former colleagues who passed away unexpectedly in recent years (The Daniel Wolf Prize, The Uta von Schwedler Prize, and the Andy Kaplan Prize) and serve to encourage up-and-coming scientists at the graduate and postdoc/junior levels. The prizes are also a testament to the community nature of the Retroviruses meeting and the continuity of the attendees over the years. The award recipients this year were Michael Metzger (Andy Kaplan Prize), Anna Gres (Uta von Schwedler Prize), and Guangai Xue (Daniel Wolf Prize).

As previously, the meeting was organized into 13 sessions (10 oral and three poster). The presentations featured unpublished findings, with an emphasis on basic research studies. The sessions utilized an integrated approach, focusing on traditional areas such as specific aspects of the retrovirus life cycle, as well as on different types of host-cell defense mechanisms—in particular specific restriction factors and the viral proteins that counteract them. General sessions on Virus Entry and Spread and on Pathogenesis and Therapy opened and closed the meeting, respectively. A novel feature this year was to place an Evolution and Genetics session in the middle of the meeting on Wednesday evening to highlight its centrality to all topics. The organizers encouraged as many groups as possible to give talks by purposefully restricting each laboratory to no more than two talks at the meeting (and these two talks could not both be in the same session). Furthermore, there was a female co-chair for all but one session.

Scientific highlights of the meeting included the use of advanced imaging techniques to track viral particles from uncoating and nuclear entry through to virus particle assembly and updates on the mechanism of restriction factor inhibition. In addition, a novel capsid interacting protein, Sun2, was identified. A range of basic information regarding the retroviral life cycle was also

C. Richards, M. Galilee Zano

E. Poeschla, S. Kumar, S. Sarafianos, B. Kim

presented. As previously, the meeting also exemplified new techniques and technologies that could be applied both to the discipline of retrovirology and to other areas of research.

This meeting was funded in part by the National Institute of Allergy and Infectious Diseases, a branch of the National Institutes of Health.

PROGRAM

Virus Entry and Spread
Chairpersons: C. Jolly, *University College London, United Kingdom;* J. Munro, *Tufts University School of Medicine, Boston, Massachusetts*

Reverse Transcription and APOBEC
Chairpersons: L. Chelico, *University of Saskatchewan, Saskatoon, Canada;* V. Pathak, *National Cancer Institute, Frederick, Maryland*

CA and Early Postentry Events
Chairpersons: E. Campbell, *Loyola University Medical Center, Maywood, Illinois;* M. Naghavi, *Northwestern University, Chicago, Illinois*

Fifth Annual Uta von Schwedler Prize for Retrovirology
Presented by: W. Sundquist, *University of Utah, Salt Lake City*
Awarded to: A. Gres, *University of Missouri, Columbia*

Keynote Speaker
R. Swanstrom, *University of North Carolina, Chapel Hill*

VPX, VPR, and SAMHD1
Chairpersons: R. König, *Paul-Ehrlich-Institut, Langen, Germany;* I. Taylor, *The Francis Crick Institute, London, United Kingdom*

Evolution and Genetics
Chairpersons: W. Johnson, *Boston College, Chestnut Hill, Massachusetts;* M. Pizzato, *University of Trento, Italy*

T. Packard

J. Martin, L. Mansky, M. Meissner

W. Mothes, P. Uchil

A. Engelman, P. Bieniasz

Tenth Annual Andy Kaplan Prize
Presented by: W. Johnson, *Boston College, Chestnut Hill;*
Awarded to: M. Metzger, *Columbia University/HHMI, New York*

Nuclear Events and Integration
Chairpersons: A. Fassati, *University College London, United Kingdom;* G. Maertens, *Imperial College London, United Kingdom*

RNA Expression and Trafficking
Chairpersons: W.-S. Hu, *National Cancer Institute, Frederick, Maryland;* N. Sherer, *University of Wisconsin, Madison*

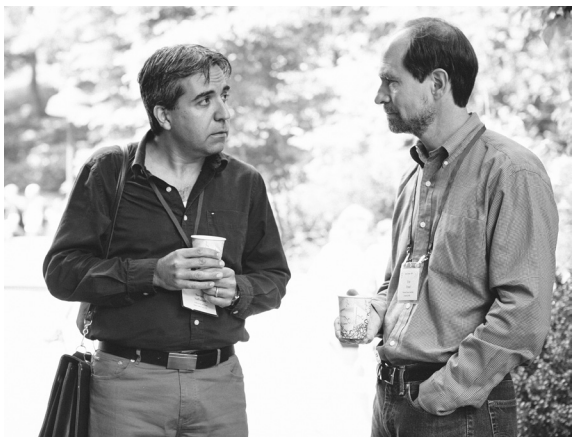
Keynote Speaker
A. Rein, *National Cancer Institute, Frederick, Maryland*

RNA Packaging and Virus Assembly
Chairpersons: E. Freed, *National Cancer Institute, Frederick, Maryland;* J. Lingappa, *University of Washington, Seattle*

Budding and Maturation
Chairpersons: F. Bouamr, *NIAID, National Institutes of Health, Bethesda, Maryland;* S. Neil, *King's College London, United Kingdom*

Sixth Annual Daniel Wolf Prize
Presented by: R Liberatore, *Aaron Diamond AIDS Research Center*
Awarded to: The best poster presentation, Guangai Xue

Pathogenesis and Therapy
Chairpersons: Z. Ambrose, *University of Pittsburgh, Pennsylvania;* F. Kirchhoff, *University of Ulm, Germany*



S. Saffarian, E. Freed

D. Ajasin, V. Prasad

Glia in Health and Disease

July 21–25 208 participants

ARRANGED BY **Marc Freeman**, HHMI/University of Massachusetts, Worcester
Kelly Monk, Washington University in St. Louis, Missouri

M. Freeman, K. Monk

Glial cells are as abundant as neurons in the nervous system, but the diverse roles played by these cells under physiological and pathological conditions are not well understood. In this summer's sixth such meeting, trainees and scientists from across the world gathered to discuss recent progress in this rapidly expanding field. More than half (~57%) of the speakers were women, and most speakers were selected from the abstracts with a focus on trainees and junior faculty. Speakers presented exciting new data on a variety of important topics including glial control of network activity and behavior, glial contribution to circuit assembly and nervous system plasticity, roles of glia in injury and repair, and regulation of the blood–brain barrier. The atmosphere was collegial and supportive, and there were lively discussions in oral sessions, poster sessions, and informal gatherings. Research involving vertebrate and invertebrate model systems was presented along with advances in state-of-the-art methodologies that are providing new insight into glial cell functions. A highlight of the meeting was the "trainee session," which supplanted a more traditional keynote session. During this session, four talented trainees (one graduate student and three postdoctoral fellows) using diverse experimental approaches presented their recent work. This session was convened in order to honor the long-standing contributions of the meeting's founder, Dr. Ben Barres, to the training of young scientists and his leadership in the field of glial cell biology. This session also served to highlight some of the many important advances made by trainees to the field. By the conclusion of the meeting, it was clear that information about properties and functions of glia is increasing exponentially, enabled by the many advances described at the meeting. These studies are yielding new insight into the

E. Kurant, B. Zhang

L. Zhang, X. He

K. Baldwin

S. Kanner, M. Shuo Chen

roles of distinct groups of glia in nervous system development, neural circuits, disease initiation and progress, and nervous system repair.

This meeting was funded in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health; and the Multiple Sclerosis Society.

PROGRAM

Control of Network Activity and Behavior
Chairpersons: D. Bergles, *Johns Hopkins University, Baltimore, Maryland;* R. Robitaille, *University of Montreal, Quebec, Canada*

CNS Myelination, Disease, and Repair
Chairpersons: M. Simons, *Max Planck Institute for Experimental Medicine, Göttingen, Germany;* A. Vanderver, *Children's National Health System, Washington, D.C.*

Reactive Microglia and Astrocytes
Chairpersons: A.V. Molofsky, *University of California, San Francisco;* M. Sofroniew, *University of California, Los Angeles*

Neural Circuit Assembly and Plasticity
Chairpersons: N. Allen, *Salk Institute for Biological Studies, La Jolla, California;* D. Schafer, *University of Massachusetts Medical School, Worcester*

C. Marion

S. Shaham, M. Freeman

Astrocyte Cell Fate, Morphogenesis, and Function
Chairpersons: H. Sontheimer, *Virginia Tech Carilion Research Institute, Roanoke;* E. Ullian, *University of California, San Francisco*

Peripheral Nerve Assembly and Repair
Chairpersons: P. Brophy, *University of Edinburgh, United Kingdom;* S. Kucenas, *University of Virginia, Charlottesville*

Communication Across the Blood–Brain Barrier
Chairpersons: R. Daneman, *University of California, San Diego;* C. Klambt, *University of Muenster, Germany*

Ensheathment and Support of Neurons
Chairpersons: M.L. Feltri, *State University of New York, Buffalo;* S. Scherer, *University of Pennsylvania, Philadelphia*

D. Schafer, M. Amin Sherafat

Genome Engineering: The CRISPR-Cas Revolution

August 17–20 409 participants

ARRANGED BY **Jennifer Doudna**, HHMI/University of California, Berkeley
Maria Jasin, Memorial Sloan Kettering Cancer Center, New York, New York
Jonathan Weissman, HHMI/University of California, San Francisco

M. Jasin, J. Doudna

The 2016 meeting was the second to be held at Cold Spring Harbor. Genome engineering (gene editing) involving the introduction of breaks in the DNA backbone has been possible for several years. The discovery of bacterial adaptive immunity and the co-opting of Cas9 and related components for programmable DNA recognition and cleavage have led to the widespread application of genome engineering, such that scientists working in many research areas and organisms can readily utilize the approach. Furthermore, Cas9 allows multiplexing to an extent that was not possible with previous approaches, and it is readily adapted for control of transcription and other applications, which were also highlighted at the conference.

Session titles included CRISPR Screens and Gene Drives, both of which were new this year (see Program below). A goal of this meeting was to bring together researchers working in diverse fields to stimulate discussions and ideas to further exploit Cas9 and related technologies for biological discovery and medical applications. Sixteen speakers were invited to cover these diverse topics. Another 38 speakers were chosen from submitted abstracts. Speakers

M. Gasperini, S. Vonesch, J. Smith, G. Findlay

M. Gundry

J. Haber, D. Gallaghar

represented institutions from the United States and abroad (Spain, United Kingdom, The Netherlands, and China) and also included five representatives from industry. Speakers chosen from submitted abstracts included staff scientists, postdoctoral fellows, and lab heads. Approximately 100 posters were presented in two sessions, complementing the oral presentations.

Talks throughout the meeting utilized molecular, cell, and computational biology in diverse model organisms, as well as economically important and some unconventional organisms. This included talks from investigators from DuPont (corn), the International Rice Research Institute in the Philippines (rice), Concordia University (*Apergillus*; speaker from Nigeria), Czech Republic (diplonemids), and Turkey (rare diseases, frequently from consanguinity). Much of the data presented was unpublished or only very recently published.

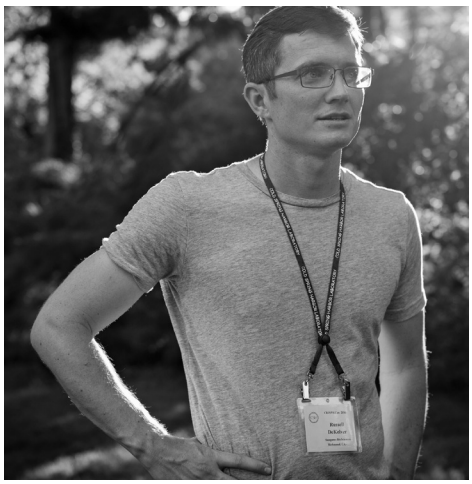
PROGRAM

Screens

Chairpersons: A. Regev, *Broad Institute of MIT and Harvard, Cambridge, Massachusetts;* D. Sabatini, *Whitehead Institute for Biomedical Research, Cambridge, Massachusetts*

Repairing DNA Breaks

Chairpersons: F. Storici, *Georgia Institute of Technology, Atlanta;* J. Haber, *Brandeis University, Waltham, Massachusetts*



R. DeKelver

B. Wienert

R. Chahwan, L. Higgins

F. Port, S. Schmidt

CRISPR
Chairpersons: J. DeRisi, *HHMI/University of California, San Francisco;* L. Marraffini, *The Rockefeller University, New York, New York*

Stem Cells
Chairpersons: W. Skarnes, *Wellcome Trust Sanger Institute, Hinxton, United Kingdom;* D. Huangfu, *Memorial Sloan Kettering Institute, New York, New York*

CRISPR Medicine
Chairpersons: F. Stegmeier, *KSQ Therapeutics, Cambridge, Massachusetts;* J. Smith, *Cellectis, Paris, France*

Gene Drive and Model Organisms
Chairpersons: E. Sontheimer, *University of Massachusetts Medical School, Worcester;* E. Bier, *University of California, San Diego*

Regulatory and Noncoding RNAs

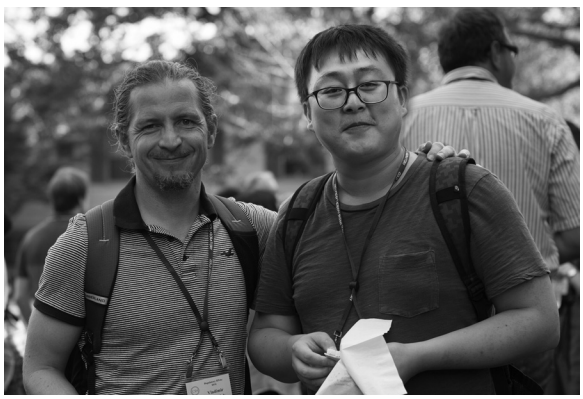
August 23–27 218 participants

ARRANGED BY **Victor Ambros**, University of Massachusetts Medical School, Worcester
Elisa Izaurralde, Max Planck Institute for Developmental Biology, Tübingen, Germany
Nicholas Proudfoot, University of Oxford, United Kingdom

V. Ambros, E. Izaurralde, N. Proudfoot

RNA is long established as a key feature of gene expression, where classically it was shown to act as the genetic messenger. Here it provides a blueprint of the protein sequence encoded in genes that is then translated by ribosomes into proteins. Added to this protein-coding function, RNA was soon after revealed to play critical structural and catalytic functions in its own right rather than simply encoding protein sequence. This so-called noncoding RNA has now been shown to act as a critical regulator of gene expression by playing central roles across numerous biological processes required for cell viability and function. In both prokaryotes and eukaryotes, short noncoding RNAs such as CRISPRs, microRNAs, and piRNAs act both to regulate absolute levels of protein production and to provide an elaborate defense mechanism to prevent the corruption of the genome by infecting viruses or transposons. Added to this, eukaryotic genomes are pervasively transcribed to generate thousands of long noncoding RNAs (lncRNAs), some of which are emerging as key regulators of diverse cellular processes. However, their precise physiological functions are still under intense scrutiny. In this summer's meeting, scientists at all career stages engaged in intense and enthusiastic discussion regarding the diverse and multifaceted roles of noncoding RNA across species and in a broad range of cellular processes. Sessions focused on recent, unpublished discoveries on both long and short noncoding RNAs. There were ample opportunities for discussion in the oral sessions, the poster sessions, and informal settings. There was a strong sense of community and the conviction that this meeting has great value. Thus, it provides an excellent framework to exchange knowledge and application of the rapidly advancing technologies of modern molecular and genome biology

C. Kam



V. Majerciak, H. Liu

R. Karni, D. Cazalla

P. Mouritzen, M. Mycko

and discuss how these can be harnessed to establish the roles of noncoding RNAs in gene regulation and function.

This meeting was funded in part by Arraystar, Inc., and the National Science Foundation.

PROGRAM

Keynote Speakers
N. Kim, *Seoul National University, Korea*
J. Mendell, *University of Texas Southwestern Medical Center, Dallas*

Transcriptional Gene Silencing
Chairpersons: D. Moazed, *Harvard Medical School, Boston, Massachusetts;* A. Ahktar, *MPI of Immunobiology and Epigenetics, Freiburg, Germany*

Nuclear and Cytoplasmic Silencing Mechanisms
Chairperson: Y. Tomari, *University of Tokyo, Japan*

miRNAs: Biogenesis and Function
Chairpersons: D. Bartel, *HHMI/Whitehead Institute, MIT, Cambridge, Massachusetts;* I. Bozzoni, *Sapienza University of Rome, Italy*

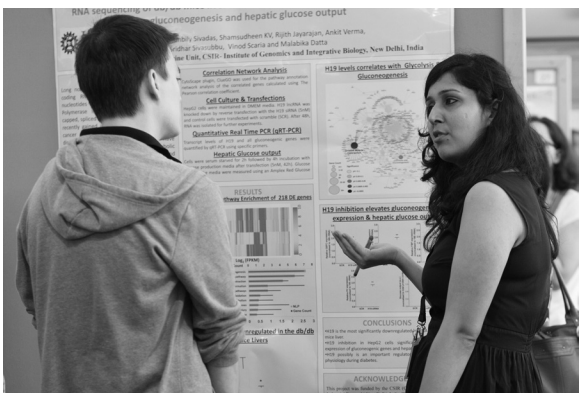
Biological Function of miRNAs
Chairperson: L. He, *University of California, Berkeley*

circRNA and piRNA Biogenesis
Chairpersons: M. Siomi, *University of Tokyo Graduate School of Science, Japan;* E. Miska, *University of Cambridge, United Kingdom*

Long Noncoding RNAs I
Chairperson: N. Proudfoot, *University of Oxford, United Kingdom*

Noncoding RNAs in Bacteria and Long Noncoding RNAs II
Chairperson: M. Terns, *University of Georgia, Athens*

K. Foster, J. Singh



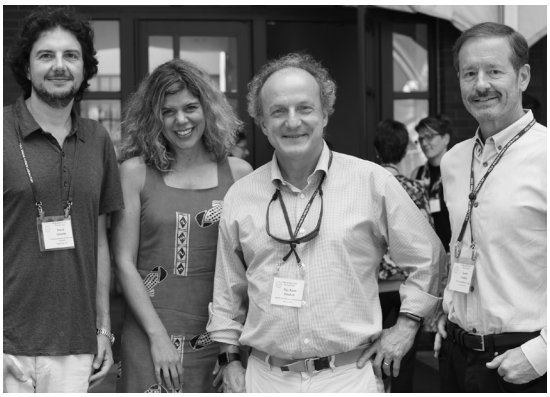
B. Takao Real Karia, N. Goyal

The PI3K–mTOR–PTEN Network in Health and Disease

August 30–September 3 112 participants

ARRANGED BY **Anne Brunet**, Stanford University, California
Lewis Cantley, Weill Cornell Medical College, New York, New York
Pier-Paolo Pandolfi, BIDMC and Harvard Medical School, Boston, Massachusetts
David Sabatini, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts

This sixth conference was a dynamic discussion forum bringing together top scientists studying the PI3K–mTOR–PTEN pathway using biochemical and molecular approaches as well as in vivo model systems. The meeting brought together a diverse group of scientists studying various biochemical, molecular, and genetic approaches to the analysis of the PI3K–mTOR–PTEN pathway at the cellular level and in development and diseases including cancer, metabolic disorders, and neurological disorders. Junior and senior researchers joined to discuss their latest research findings and technical approaches to the analysis of PI3K–mTOR–PTEN. Particular emphasis was given to model systems, disease pathogenesis, and therapy.



D. Sabatini, A. Brunet, P.P. Pandolfi, L. Cantley

A total of 43 investigators presented in the seven scientific sessions, with 43 platform and 40 poster presentations and 112 registered attendees. The seven platforms and two poster sessions were marked by cutting-edge dynamic presentations and enthusiastic exchanges of unpublished new results. During the scientific sessions, participants engaged in a vibrant discussion in the question and answer sessions. The poster sessions were also very well attended. The meeting sessions and chairpersons are listed below under "Program."

This meeting was funded in part by Avanti Polar Lipids.

M. Haigis, G. Hardie

P. Tomlinson

M. Costa Mattioli, C.J. Chen

K. Vousden, R. Shaw

PROGRAM

PTEN: Regulation and Targets
Chairperson: L. Trotman, *Cold Spring Harbor Laboratory*

mTOR: Regulation and Targets
Chairperson: M. Haigis, *Harvard Medical School, Boston, Massachusetts*

PI3K–mTOR–PTEN Network in Cancer and Therapeutics
Chairperson: K. Wellen, *University of Pennsylvania, Philadelphia*

Metabolism and Mitochondria
Chairperson: M. Hansen, *Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California*

AMPK/LKB1: Regulation and Targets
Chairperson: W. Mair, *Harvard T.H. Chan School of Public Health, Boston, Massachusetts*

Neurological and Neurodegenerative Diseases
Chairperson: N. Bardeesy, *Massachusetts General Hospital Cancer Center, Boston*

Aging and Longevity
Chairperson: M. Wang, *Baylor College of Medicine, Houston, Texas*

N. Steinbach, E. Stratikopoulos

S. Pors, J. Xu

Translational Control

September 6–10 328 participants

ARRANGED BY **Thomas Dever**, National Institutes of Health
Joel Richter, University of Massachusetts Medical School
Marina Rodnina, Max Planck Institute for Biophysical Chemistry, München, Germany

This meeting attracted participants from around the world and included Susan Ackerman, Holger Stark, and Aaron D. Gitler as keynote speakers, eight platform sessions, and three poster sessions that covered 249 abstracts.

T. Dever, M. Rodnina, J. Richter

Bringing together a wide variety of researchers with different areas of expertise, the meeting highlighted not only the molecular intricacies of protein synthesis, but also the impacts of misregulated protein synthesis on human disease including viral infection, cancer, and neurological disease. With an impressive mix of established and young investigators, the meeting maintained the central role of the Translational Control Meeting for researchers in the community.

This year, the program was organized into seven main topics (see Program below). A highlight of the program was the keynote address by Susan Ackerman, who described her forward genetic screens in mice that have uncovered roles for the translation apparatus in CNS development and have linked misregulated translation to neurodegeneration. Additional talks described nonconventional translation initiation in mRNAs containing nucleotide repeats associated with human diseases, and the impacts of mRNA poly(A) tail changes and mRNA cap binding proteins in regulating translation during development, differentiation, and cellular stress conditions.

Two additional keynote addresses were made by Holger Stark, who described his high-resolution cryo-electron microscopy studies elucidating how the ribosome incorporates the 21st amino acid, selenocysteine, and by Aaron Gitler, who described his studies on protein misfolding diseases including amyotrophic lateral sclerosis (ALS). In addition, 67 short talks were selected from among the submitted abstracts, and 179 abstracts were presented as posters divided among three

L. Marler

A. Wilczynska

M. Cargnello

O. Elroy Stein, M. Hatzoglou

N. Sonenberg, R. Dikstein

sessions on Wednesday afternoon, Thursday night, and Friday afternoon. The poster sessions were well attended in both Bush Hall and Nicholls Biondi Hall, and they prompted many discussions and exchange of ideas.

Highlights of the meeting included new insights into the mechanism and physiological regulation of protein synthesis, including how different translational control pathways contribute to either the prevention or promotion of cellular transformation in cancer. Ribosomal profiling studies examining the genome-wide distribution of ribosomes on mRNAs in a cell revealed how protein synthesis is modulated in response to different stresses, and a related technique to map translation initiation complexes on mRNAs provided mechanistic insights into how ribosomes are recycled following translation. Biochemical as well as ribosome profiling studies revealed the role of RNA helicases and nucleotide methylation in translation initiation and the roles of mRNA sequence and specific codons in regulating translation efficiency and mRNA turnover. An advancing area of investigation is the translation of specific mRNAs by specialized ribosomes including the role of rRNA modifications in generating pools of structurally and functionally distinct ribosomes. Advances in single-molecule and ensemble FRET techniques helped identify new translation intermediates and the kinetics of ribosome movements, and new cell biology methods enabled visualization of translation of single mRNAs in cells. Finally, studies revealing the resolution of stalled translational complexes by hydrolysis of the peptide-tRNA bond and by novel nontemplated addition of amino acids to stalled polypeptides revealed linkages to the cellular proteolysis machinery and provided new insights into ribosome quality control.

A. Jacobson, D. Schoenberg

P. Kumar, A. Goyal, M. Rodnina

In summary, this meeting continues the key role of the Translational Control meeting as the major focal point for the ongoing excitement and growth of this field of research.

PROGRAM

Development and CNS
Chairperson: L. Ranum, *University of Florida College of Medicine, Gainesville*

Keynote Speaker
S.L. Ackerman, *University of California, San Diego*

Initiation
Chairperson: C. Fraser, *University of California, Davis*

Turnover
Chairperson: J. Coller, *Case Western Reserve University School of Medicine, Cleveland, Ohio*

Elongation and Termination
Chairperson: A. Buskirk, *Johns Hopkins University School of Medicine, Baltimore, Maryland*

Ribosome
Chairperson: A. Mankin, *University of Illinois, Chicago*

Keynote Speaker
H. Stark, *Max Planck Institute for Biophysical Chemistry, Göttingen, Germany*

Regulation I
Chairperson: J. Pelletier, *McGill University, Montreal, Canada*

Virus and Disease
Chairperson: D. Ruggero, *University of California, San Francisco*

Keynote Speaker
A.D. Gitler, *Stanford University, California*

Regulation II
Chairperson: M. Hatzoglou, *Case Western Reserve University School of Medicine, Cleveland, Ohio*

J. Hershey (*left*), W. Merrick (*second from right*), C. Fraser (*far right*)

Epigenetics and Chromatin

September 13–17 362 participants

ARRANGED BY **Shelley Berger**, University of Pennsylvania, Philadelphia
Jürg Müller, Max Planck Institute of Biochemistry, Planegg, Germany
Yang Shi, Boston Children's Hospital, Massachusetts

J. Müller, S. Berger, Y. Shi

Following the successful start of this conference in 2012 and 2014, this third meeting brought together researchers working on different aspects of chromatin biology and epigenetics. The focus was on molecular mechanisms by which chromatin and chromatin-modifying protein complexes regulate gene transcription and the genome. The spectrum of the oral and poster presentations demonstrated that the field has been making progress through studies that range from atomic-resolution structure determination of chromatin proteins, to the epigenome and genomic architecture, to the functional analysis of regulatory networks at the organism level.

The high attendance, large numbers of posters (205), and the sustained involvement of the participants throughout the meeting illustrated the success of the meeting. Scientists from 22 countries attended the meeting, and for many of these it was their first visit to Cold Spring Harbor. Several of the oral presentations selected from abstracts and given by junior PIs, postdocs, or students turned out to be scientific highlights, showing the success of this meeting format as a presentation platform for junior researchers. The emergence of novel mechanistic themes in the field, the progress in methodologies and the presentations by junior scientists allowed for sustained interest by the large and diverse audience throughout the different sessions.

There were eight plenary sessions. The function and mechanisms of DNA and histone-modifying enzymes and readers of these modifications in gene activation and repression were covered in two sessions. Chromatin processes mediating three-dimensional genome organization were discussed in one session. Three sessions focused on chromatin modifications and remodeling controlling developmental processes at the organism level. One session focused on inheritance of chromatin state and one on metabolism and RNA.

N. Vander Schaaf

C. Petell, I. Sanli

D. Nicetto, M. Walker, L.M. Zink

N. Vander Schaaf, J. Pai

Among the highlights of the meeting were presentations of new approaches to utilize pharmacologics and genetics to target chromatin in disease including cancer, novel findings of how chromatin manifests transcriptional memory through DNA replication and mitosis, and how chromatin changes to mediate a "point of no return" during cell differentiation. There were fascinating novel findings regarding the role of chromatin in neurobiology, which remains an important frontier in epigenetics. Rising issues and technologies for single-cell and single-molecule chromatin biology were discussed. Important questions were brought forth regarding how chromatin regulatory complexes bind to chromatin and then effect transcription and other genomic processes. Key questions and approaches were discussed in relation to positioning of transcriptionally active genes within the nucleus and inheritance of these locations. Many issues regarding the role of chromatin regulation in organismal development, mutations of histones driving disease, and other issues encompassing metabolism and obesity were discussed. These numerous important and fascinating topics demonstrated the key role of chromatin in genome regulation—and the crucial questions that still await exploration and discovery.

PROGRAM

Chromatin State Inheritance
Chairperson: S. Henikoff, *HHMI/Fred Hutchinson Cancer Research Center, Seattle, Washington*

Keynote Speaker
R. Kingston, *Massachusetts General Hospital, Boston*

R. Gleason, P. Kopinski, E. Darrow

P. Kopinski, M. Couturier

N. Iglesias

S. Manickavinayahan

Nucleic Acids and Histone Modifications
Chairpersons: M.E. Torres Padilla, *IGBMC, Strasbourg, France*; W. de Laat, *Hubrecht Institute, Utrecht, The Netherlands*

Nucleosome Dynamics
Chairpersons: C. Wolberger, *Johns Hopkins University School of Medicine, Baltimore, Maryland;* M. Lazar, *University of Pennsylvania Perelman School of Medicine, Philadelphia*

Keynote Speaker
G. Almouzni, *Institut Curie, Paris, France*

Regulators of Development
Chairpersons: D. Moazed, *HHMI/Harvard Medical School, Boston, Massachusetts;* W. Bickmore, *MRC Human Genetics Unit, Edinburgh, United Kingdom*

3D Chromatin Organization
Chairpersons: J. Wysocka, *Stanford University, California;* D. Duboule, *University of Geneva and Ecole Polytechnique Fédérale, Lausanne, Switzerland*

Epigenetics in Germ Cells and Early Embryos
Chairpersons: C. Vakoc, *Cold Spring Harbor Laboratory;* A. Schaefer, *Icahn School of Medicine, Mount Sinai, New York*

Metabolism and RNA
Chairpersons: B. Strahl, *University of North Carolina School of Medicine, Chapel Hill;* G. Almouzni, *Institut Curie, CNRS, Paris, France*

Readers and Effectors
Chairpersons: P. Verrijzer, *Erasmus University Medical Center, Rotterdam, The Netherlands;* R. Kingston, *Massachusetts General Hospital/Harvard Medical School, Boston*

D. Acharya, Y. Zhang

K. Palozola, P. Kopinski, M. Couturier

Axon Guidance, Synapse Formation, and Regeneration

September 20–24 267 participants

ARRANGED BY **Greg Bashaw**, University of Pennsylvania, Philadelphia
Linda Richards, University of Queensland, Brisbane, Australia
Peter Scheiffele, Biozentrum University of Basel, Switzerland

P. Scheiffele, L. Richards, G. Bashaw

The field of neuroscience is undergoing a transformation with a number of brain projects under way or in their planning stages around the globe. These brain projects hope to uncover the mechanisms that generate and wire the nervous system into functioning circuits. This 10th meeting in the series was therefore both timely and important, as it provided a means for investigators in the field to exchange ideas and information that is generating new data and capitalizing on the global efforts currently under way to understand the brain. The meeting focused on the development of neural circuits, specifically focusing on molecular and cellular mechanisms of axon guidance and synapse formation and how these mechanisms could be applied to understanding axonal regeneration in the adult. A major challenge in the field is to retain and develop investigators who are equipped to identify the function of genes that have been associated with human diseases of the nervous system. The meeting brought together investigators working across a diverse range of model organisms, providing complementary approaches to understanding circuit formation and function and to uncover the role of specific genes in human neurological diseases.

Speakers were chosen from among the participants submitting abstracts by session chairs who are leaders in the field. Of all the participants, 40% were women—and this was reflected in the number of talks presented by women. Forty-nine abstracts in seven sessions were selected for talks, with the remaining abstracts being presented as posters. Advanced and starting assistant professors, postdoctoral fellows, and graduate students were well represented as speakers and participants. Session chairs at the meeting were well balanced between men and women, and the meeting had a clear international presence with participants from India,

C. Salmon, A. Hasan

D. Kulshrestha

J. Ziak, K. Luthy

N. Dominguez Iturza, A. Cruz Martin

Israel, Germany, France, Japan, China, United States, and Australia. Most of the talks this year were by graduate students.

All of the major areas of research in the field were covered, as were all of the major approaches (cellular, physiological, anatomical, molecular, dynamic imaging, biochemical, and genetic). In addition, there were two keynote addresses. The first was given by Professor Liqin Luo (Stanford University), a leader in the field of circuit formation in flies and mice. His outstanding studies using both transgenics and viral constructs have elucidated the fundamental mechanisms involved in the olfactory and other systems across species. The second keynote lecture was given by someone outside the field using cutting-edge technologies to understand cellular identity. The lecture was given by Professor Aviv Regev (Broad Institute), a computational biologist whose work is leading the field in using single-cell RNA sequencing to understand the molecular signatures of individual cells in complex systems such as the retina and hippocampus. Her laboratory has developed new techniques for sequencing RNA specifically in dividing cells and neurons by pre-sorting the cells before preparing the single-cell RNA libraries. Both lectures provided the audience with an amazing array of cutting-edge technologies that are helping to uncover the complexities of the nervous system.

This meeting also held several lunches targeted at professional development. Trainees were invited to interact with senior scientists to discuss topics as diverse as how to search for mentors (postdoc or faculty); prepare for job interviews (academic or industry); write grant proposals manage teaching and research; balance life and work; and handle paper reviews.

The students and postdocs expressed great appreciation for the close interaction and valuable advice they gained through this event. The business meeting was conducted with open discussion.

F. Fiederling, J. Loverde, A. Almoril Porras

F. Bisiak, J. McDermott

Overall, this meeting provided an important forum for ideas and approaches in developmental neuroscience and regeneration and helped scientists in the field get the most up-to-date information, as well as enabling them to meet, to network, and to establish collaborations. Indeed, many expressed an appreciation for the smaller size of the meeting in promoting more interactions, as well as the social events including the wine and cheese, banquet and informal lunches and dinners. The three poster sessions were extremely well attended. Based on the enthusiastic comments of the participants, the intensity of the oral and poster sessions, and the large crowds that stayed up late every night at the bar to discuss science, the meeting was truly a great success, and highly appreciated by each of the attendees.

This meeting was funded in part by the National Institute of Neurological Diseases and Stroke, a branch of the National Institutes of Health.

PROGRAM

Axon to Synapse I
Chairpersons: S. Butler, *University of California, Los Angeles;* R. Klein, *Max Planck Institute of Neurobiology, Martinsried, Germany*

Regeneration and Disease I
Chairpersons: F. Bradke, *German Center for Neurodegenerative Diseases, Bonn;* C. Collins, *University of Michigan, Ann Arbor*

Special Lecture
A. Regev, *Broad Institute of MIT and Harvard, Boston, Massachusetts*

Synapse to Circuit I
Chairpersons: K. Martin, *University of California, Los Angeles;* A. Villu Mariq, *University of Utah, Salt Lake City*

Regeneration and Disease II
Chairpersons: M. Granato, *University of Pennsylvania, Philadelphia;* M. Hilliard, *University of Queensland, Brisbane, Australia*

Axon to Synapse II
Chairpersons: R. Hiesinger, *Freie Universität Berlin, Germany;* G. Lopez-Bendito, *Instituto de Neurosciences, UMH-CSIC, Alicante, Spain*

Synapse to Circuit II
Chairpersons: G. Fishell, *New York University Medical School, New York;* K. Shen, *Stanford University, California*

Special Lectures
R. Adolphs, *California Institute of Technology, Pasadena;* L. Luo, *HHMI/Stanford University, California*

Axon to Synapse III
Chairpersons: E. Herrera, *Consejo Superior de Investigaciones Cientificas, Alicante, Spain*; I. Salecker, *The Francis Crick Institute, London, United Kingdom*

A. Poulopoulos, C. Salmon

D. Comoletti, J. Himanen

Mechanisms of Aging

September 26–30 294 participants

ARRANGED BY **Vera Gorbunova**, University of Rochester, New York
Malene Hansen, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California
Scott Pletcher, University of Michigan, Ann Arbor

S. Pletcher, V. Gorbunova, M. Hansen

Aging is the most prominent risk factor for developing a number of serious diseases, ranging from neurodegeneration to cancer. Research from a number of laboratories has shown that genetic makeup and the environment have prominent roles in the aging process. Among many significant findings is the identification and characterization of several molecular pathways that influence life span and susceptibility to age-related disorders in diverse species, from yeast to humans. This conference provided a highly stimulating and educational forum for presenting the latest research findings and for developing paradigms in aging research at the molecular, cellular, and organismal levels. The conference opened with a session on longevity genes and pathways, which provided an update on the latest findings on specific genes linked to the aging process and their age-related effects. The morning session on the second day highlighted research efforts on comparative biology and nonstandard animal models, such as naked mole rat and killifish. The afternoon session focused on stem cell biology, with several talks highlighting roles for basic cellular processes; it which was followed by an evening poster session of more than 60 posters. The third day of the conference featured two oral sessions on metabolism and epigenetics, respectively, describing novel molecular paradigms in both mammals and invertebrate systems. Again, the evening featured a poster session with more than 60 posters (125 posters in total). The fourth day of the conference began with a session about DNA repair and senescence and highlighted new mechanisms that drive cells to stop proliferating with time. The afternoon oral session was devoted to how cells maintain homeostasis (e.g., by engaging different cellular stress response pathways). The final

S. Ahmed, J. Tyler

S. Marttila, J.P. de Magalhães

T. von Zglinicki, M. Santra, A. de Graff

E. Santo, L. Booth

morning session of the fifth day focused on discussing nutrient sensing and interventions, and it highlighted the latest findings on noninvasive approaches to interfere with the aging process. The conference was attended by close to 300 researchers from many different fields, who collectively contributed to a highly interactive and exciting discussion forum focused on obtaining a better understanding of the molecular mechanisms of aging, so that better treatments can be developed against age-related diseases.

This meeting was funded in part by the National Institute on Aging, a branch of the National Institutes of Health; Calico Life Science; Glenn Foundation for Medical Research; and Macrogen Corp.

PROGRAM

Genes and Pathways
Chairpersons: B. Kennedy, *Buck Institute for Research on Aging, Novato, California;* S.-I. Imai, *Washington University School of Medicine, St. Louis, Missouri;* W. Mair, *Harvard T.H. Chan School of Public Health, Boston, Massachusetts*

Comparative Biology/Nonstandard Models
Chairpersons: D. Promislow, *University of Washington, Seattle;* J.P. de Magalhães, *University of Liverpool, United Kingdom.*

Stem Cells
Chairpersons: S. Chang, *Yale University School of Medicine, New Haven, Connecticut;* H. Jasper, *Buck Institute for Research on Aging, Novato, California*

Metabolism
Chairpersons: L. Partridge, *University College London, United Kingdom and Max Planck Institute for Biology of Ageing, Cologne, Germany;* M. Wang, *Baylor College of Medicine, Houston, Texas*

K. Kornfeld, A. Scharf

L. Booth, A. Webb

Epigenetics
Chairpersons: P. Adams, *University of Glasgow, United Kingdom;* J. Tyler, *Weill Cornell Medicine, New York*; A. Brunet, *Stanford University, California*

DNA Repair/Senescence
Chairpersons: N. Sharpless, *University of North Carolina School of Medicine, Chapel Hill;* T. Von Zglinicki, *Newcastle University, Newcastle-upon-Tyne, United Kingdom*

Homeostasis
Chairpersons: D. Walker, *University of California, Los Angeles;* K. Blackwell, *Joslin Diabetes Center, Boston, Massachusetts*

Interventions
Chairpersons: Y. Suh, *Albert Einstein College of Medicine, Bronx, New York;* D. Sinclair, *Harvard Medical School, Boston, Massachusetts;* P. Rabinovitch, *University of Washington, Seattle*

C. Murphy, L. Lapierre

N. Saini, A. Gurkar

Germ Cells

October 4–8 203 participants

ARRANGED BY **Robert Braun**, The Jackson Laboratory, Bar Harbor, Maine
Geraldine Seydoux, HHMI/Johns Hopkins School of Medicine, Baltimore, Maryland

This year's meeting marked the 10th anniversary of the biannual meeting on germ cells first held at CSHL in 1998. A long-standing hallmark of the meeting is its focus on comparative analysis of germ cell biology. This year's meeting continued this tradition and included talks on germ cells in crustaceans, flies, worms, zebrafish, medaka, frogs, sea urchins, mice, and humans. The approaches emphasized state-of-the-art applications of genetics, genomics, biochemistry, and imaging modalities. The meeting clearly met the historically high standards previously achieved for this conference. Meeting attendance was up from the previous meeting at 203 participants and 162 abstracts. Such participation in the currently dismal funding climate demonstrates how strongly the community supports this meeting and its format.

G. Seydoux, R. Braun

The co-organizers fashioned several new session themes (e.g., Transgenerational Inheritance, Chromosomal Behavior, Cell Biology) while retaining historically important sessions that continue to see significant advances. The meeting began with a keynote address by Dr. Lee Silver, who spoke on "The Future of Human Reproduction." His talk highlighted the changing landscape of human reproduction and the various shades of genetic choice that are sure to impact human reproduction in the coming years. Major advances were described in validating the phenomenon of transgenerational/intergenerational inheritance and in some cases elucidation of mechanism. The impact of the environment and aging on germ cells illustrated fascinating and unusual responses to different forms of stress across organisms. Future meetings could have entire sessions on these emerging topics.

R. Cox, M. Buszczak, M. Pepling, D. Drummond-Barbosa, L. Cooley

P. Rangan, F. Marlow, B. Draper, A. Melendez, H. Salz

The meeting was balanced for experimental systems, with 21 talks on mouse or human systems, 16 on *C. elegans*, 16 on the fruit fly, and eight on other models. This allowed participants to appreciate deeply conserved mechanisms and underscored a great strength of the meeting: significant inclusion of nonmammalian systems and a clear focus on basic biology. Additionally, the meeting participants represented a broad range of participants: 106 (52%) of the participants and 29/61 of the speakers were female. Ample opportunity was provided for early-career scientists, who illustrated the future strength and vitality of the field. Of all the participants, 53% were either students (23%) or postdocs (30%). The 61 talks included seven student, 21 postdoctoral/research scientist, 33 faculty representing all ranks (32 male, 29 female), and eight international speakers. There were two afternoon poster sessions that were extremely well attended.

The Germ Cells meeting thus continues to be a great success judged by the criteria of exemplary science, the breadth and depth of its discussion, and attendance. The co-organizers have always been drawn from the broader germline community, each with appreciation of the strengths of this meeting. Thus, David Page (Whitehead) and Yukiko Yamashita (University of Michigan), who eminently meet these criteria, were selected as the 2018 co-organizers. We are confident that David and Yukiko will organize a successful meeting that will continue to grow while retaining its participant, organism, and system diversity.

A. Fry, C. Baker

B. Capel, B. Lesch

J. Yang, A. Arkov

K. Webster

This meeting was funded in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, a branch of the National Institutes of Health; and the National Science Foundation.

PROGRAM

Transgeneration Inheritance
Chairpersons: G. Seydoux, *HHMI/Johns Hopkins University School of Medicine, Baltimore, Maryland;* R. Braun, *Janeway Distinguished Chair, The Jackson Laboratory, Bar Harbor, Maine*

Keynote Speaker
L. Silver, *Princeton University, New Jersey*

Germ Cell Origins
Chairperson: N. Patel, *University of California, Berkeley*

Stem Cells
Chairperson: X. Chen, *Johns Hopkins University, Baltimore, Maryland*

Gene Regulation
Chairpersons: R. Ketting, *Institute of Molecular Biology, Mainz, Germany;* T. Schedl, *Washington University in St. Louis, Missouri*

Chromosome Behavior and Meiosis
Chairperson: A. Villeneuve, *Stanford University, California*

Cell Biology of Gametogenesis
Chairperson: J. Schimenti, *Cornell University, Ithaca, New York*

Sex Choice and Germ Soma Interaction
Chairpersons: D. Page, *HHMI/Whitehead Institute, MIT, Cambridge, Massachusetts;* D. Zarkower, *University of Minnesota, Minneapolis*

Transgenerational Inheritance
Chairpersons: A. Brunet, *Stanford University, California;* M. Wang, *Baylor College of Medicine, Houston, Texas*

D.F. Conrad, S. Potter

B. Draper, F. Marlow

HIV/AIDS Research: Its History and Future

October 12–15 127 participants

ARRANGED BY **John Coffin**, Tufts University, Medford, Massachusetts
Robert Gallo, Institute of Human Virology, University of Maryland School of Medicine, Baltimore
Mila Pollock, Cold Spring Harbor Laboratory
Bruce Walker, Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts

The Cold Spring Harbor Laboratory (CSHL) Archives and Genentech Center for the History of Molecular Biology and Biotechnology, together with the meetings and courses department, hosted this meeting, which brought together some of the most important researchers on HIV/AIDS from the past and present. Never before have the pioneers of the science of retroviruses, the discoverers of human retroviruses, the developers of HIV/AIDS therapeutics, and key figures in important research findings come together to discuss the history and future of the field. This interdisciplinary group reviewed the key scientific, epidemiological, and clinical discoveries that created this field, delved into the present science of HIV/AIDS, and discussed and debated the paths to the future control of this global pandemic, which, tragically, has persisted for more than 35 years.

The venue was just as significant as the speakers. It was at a virology meeting hosted by CSHL in September of 1983—a time when AIDS was a mysterious disease and was only just beginning to enter the public consciousness—that Luc Montagnier made an important announcement about a virus he had discovered that would eventually become known as HIV.

HIV, or human immunodeficiency virus, is a retrovirus, or RNA virus, that attacks the immune system. By attacking the body's T cells, HIV weakens a person's ability to fight infections. The infection is permanent, but it can be controlled with antiretroviral drugs. If not treated, HIV infection can lead to AIDS, or acquired immunodeficiency syndrome, the last stage of HIV infection in which the human immune system is severely depleted. Today 37 million people are living with HIV.

Speakers at this meeting included seminal HIV researchers Ashley Haase, Robert Gallo, Françoise Barré-Sinoussi, and Anthony Fauci; retrovirologists Harold Varmus, Robin Weiss,

J. Coffin, M. Pollock, B. Stillman, B. Walker, E. Emini

B. Hahn, A.M. Skalka, B. Walker, R. Gallo

John Coffin, Myron Essex, Flossie Wong-Staal, and David Baltimore; and medical researchers and clinicians Michael Gottlieb, David Ho, Samuel Broder, and Marty St. Clair. They addressed the following topics:

- The Story of Animal Retroviruses
- The Pandemic Begins: Early Discoveries
- Antiretroviral Therapy
- Human and Primate Retroviruses, Origin of HIV
- The Extraordinary Virus: Molecular Biology
- Immunology and Prevention
- Prospects for an HIV Vaccine
- Pathogenesis and Prospects

The meeting ended with a public program and wrap-up panel.

This meeting was the sixth in a series of CSHL Genentech Center Conferences on the History of Molecular Biology and Biotechnology. Financial support was provided by the Bill & Melinda Gates Foundation, Gilead Sciences Inc., Merck & Co., and the National Institute of Allergy and Infectious Diseases. A website documenting the meeting's talks, panel discussions, posters, photos, debates, and interactions is available at http://library.cshl.edu/Meetings/HIV_AIDS.

A. Mahoney, J. Hamilton

E. Arnold, F. Barré-Sinoussi

J. Stoye, E. Arnold, J.J. Harrison, H. Varmus

D. Richman, S. Lewin

PROGRAM

Session I: The Story of Animal Retroviruses

Weiss, R., *University College London, United Kingdom:* Retrovirus history: Early searches for human retroviruses.
Coffin, J., *Tufts University, Boston, Massachusetts:* Origins of molecular retrovirology.

Session II: The Pandemic Begins: Early Discoveries

Volberding, P., *University of California, Global Health Sciences: San Francisco:* The first patients.
Curran, J., *Emory University, Atlanta, Georgia:* Deciphering the epidemiology of AIDS.
Harrington, M., *Treatment Action Group, New York:* The importance of activism to the U.S. response.
Gallo, R., *University of Maryland School of Medicine, Baltimore:* Discoveries of human retroviruses and their linkage to disease as causative agents and preparation for the future.
Barre-Sinoussi, F., *Insitut Pasteur, Paris, France:* Discovery of HIV.
Fauci, A., *National Institutes of Health, Bethesda, Maryland:* 35 Years of HIV/AIDS: Science and policy.

Session III: Antiretroviral Therapy

St. Clair, M., *ViiV Healthcare, Research Triangle Park, North Carolina:* Discovery of AZT as the first anti-HIV drug.
Broder, S., *Intrexon Corporation, Germantown, Maryland:* The first clinical trials of antiretroviral drugs.
Richman, D., *University of California, San Diego:* Antiviral drug resistance and combination ART.
Schinazi, R., *Emory University, Atlanta, Georgia:* Discovery and development of novel NRTIs.
Hazuda, D., *Merck Research Labs, Lansdale, Pennsylvania:* Discovery and development of integrase inhibitors.
Martin, J., *Gilead Sciences, Inc., Foster City, California:* Making it simpler: A single pill to treat HIV.

Session IV: Human and Primate Retroviruses, Origin of HIV

Desrosiers, R., *University of Miami, Florida:* The origin of SIVmac: Non-human primate models for HIV.
Peeters, M., *IRD Institut de Recherche pour le Development, Montpellier, France:* On the road to HIV: Primate lentiviruses.
Hahn, B., *Perelman School of Medicine, University of Pennsylvania, Philadelphia:* Apes to humans: The origin of HIV.
Worobey, M., *University of Arizona, Tucson:* Spread of HIV in the New World.

Session V: The Extraordinary Virus: Molecular Biology

Wong-Staal, F., Professor Emeritus, *University of California, San Diego:* Discovery of human retroviral *trans*-activators.
Sodroski, J., *Dana-Farber Cancer Institute, Boston, Massachusetts:* Primate host-specific selection of immunodeficiency virus Gag and Env proteins.
Malim, M., *King's College, London, United Kingdom:* Discovery of APOBEC restriction.
Berger, E., *NIAID, National Institutes of Health, Bethesda, Maryland:* Discovery of HIV co-receptors.
Rice, A., *Baylor College of Medicine, Houston, Texas:* Mechanism of Tat *trans*-activation.
Emerman, M., *Fred Hutchinson Cancer Research Center, Seattle, Washington:* Host-virus co-evolution.

Session VI: Immunology and Prevention

Hillier, S., *Magee-Womens Hospital of the University of Pittsburgh Medical Center, Pennsylvania:* Development and application of pre-exposure prophylaxis.
Burton, D., *The Scripps Research Institute, La Jolla, California:* How does HIV evade the antibody response?

Walker, B., *The Ragon Institute of MGH, MIT, and Harvard, Cambridge, Massachusetts:* Role of T cells in controlling HIV infection.
Haynes, B., Duke *University Medical School, Durham, North Carolina*: Development of HIV vaccine: Steps and missteps.
Emini, E., *Bill & Melinda Gates Foundation, Seattle, Washington:* Issues in HIV vaccine development: Will the future be any easier than the past?
Redfield, R., *University of Maryland School of Medicine, Baltimore:* The PEPFAR program to treat HIV in Africa.

PANEL DISCUSSION

Prospects for an HIV Vaccine

Moderator: Gray, G., *South African Medical Research Center, Cape Town, Africa*

Panelists

Franchini, G., *National Cancer Institute, Bethesda, Maryland*
Zolla-Pazner, S., *Icahn School of Medicine, Mount Sinai, New York*
Corey, L., *Fred Hutchinson Cancer Research Center, Seattle, Washington*
Barouch, D., *Beth Israel Deaconess Medical Center/Ragon, Cambridge, Massachusetts*

Session VII: Pathogenesis and Prospects

Mellors, J., *University of Pennsylvania, Pittsburgh:* MACS and beyond: Epidemiology, viremia and pathogenesis.
Ho, D., *The Rockefeller University, New York:* Understanding of HIV infection through dynamics.
Shaw, G., *Perelman School of Medicine, University of Pennsylvania, Philadelphia:* Transmitted/founder HIV genomes: What they teach us.
Siliciano, R., *Johns Hopkins University, Baltimore, Maryland:* The challenge of the HIV reservoir.
Lewin, S., *University of Melbourne, Australia:* Research to a cure: A possible goal?
Baltimore, D., *California Institute of Technology, Pasadena:* Bringing it to an end (and where are we going?).

Session VIII: Public Event/Panel

Cohen, J., *Science Magazine, Cardiff, California:* Responding to AIDS: A journalist's view.
Hildebrand, S., *The Face of AIDS, Karolinska Institute, Stockholm, Sweden:* Face of AIDS Project.
Harden, V., NIH Founding Director, Emerita, *National Institutes of Health, Bethesda, Maryland:* The future of the history of AIDS.

Public Wrap-Up and Discussion: What Have We Learned?

J. Stoye, O. Cingöz

L. Corey, D. Baltimore

Biological Data Science

October 25–29 217 participants

ARRANGED BY **Bonnie Berger**, Massachusetts Institute of Technology, Cambridge
Jeff Leek, Johns Hopkins University, Baltimore, Maryland
Michael Schatz, Cold Spring Harbor Laboratory

M. Schatz, J. Leek

This conference was the second session in the new biennial series focusing on the infrastructure, software, and algorithms needed to analyze large data sets in biological research. It brought together researchers from diverse areas of biology and diverse computational domains, with interest in large-scale data analysis in common. The goal of the meeting was for attendees to step outside their usual domain and glean lessons from other areas of research. In this regard, the meeting was an overwhelming success, with extensive discussion and exchange of ideas.

More than 200 researchers participated in the meeting, with 129 abstracts submitted from academic, government, and industry laboratories by researchers who regularly mine large data sets for genomics, imaging, translational, and clinical projects. There were 10 invited talks, two keynote lectures, and an education panel, as well as 34 additional talks selected from submitted abstracts and 83 poster presentations. The discussion topics are listed below under Program. Before the meeting, more than 25 attendees participated in a Biomedical Data Science Hackathon co-organized with NCBI, where four teams spent 3 days prototyping new computational tools for analyzing structural variations, for building and aligning to graph genomes, for studying cancer epitopes, and for ultrafast read mapping.

The keynote lecture was given by Xiaole Shirley Liu, and her talk, titled "Inference of Tumor Immunity and T-Cell Receptor Repertoire from TCGA RNA-Seq Data," described new computational and statistical approaches for understanding cancer's interaction with and influence on the immune system. The second keynote was a master lecture given by Olga Troyanskaya, who described the principles of deep learning and how they can be applied to biological data. The third special session was the Education forum moderated by Mark Gerstein of Yale University that discussed and debated which topics should be taught as part of a biological data science curriculum.

K. Hansen, B. Engelhardt

A. Groff

The master lecture, education forum, and premeeting hackathon were unique strengths of the meeting with unanimous support to continue in future years.

PROGRAM

Software for Biologists
Chairpersons: R. Irizarry, *Dana-Farber Cancer Institute, Boston, Massachusetts;* M. Morgan, *Roswell Park Cancer Institute, Buffalo, New York*

Machine Learning I
Chairpersons: B. Engelhardt, *Princeton University, New Jersey;* M. Hoffman, *University of Toronto, Canada*

Master Lecture
O. Troyanskaya, *Princeton University, New Jersey*

Machine Learning II
Chairpersons: B. Engelhardt, *Princeton University, New Jersey;* M. Hoffman, *University of Toronto, Canada*

Keynote Speaker
X.S. Liu, *Harvard School of Public Health, Boston, Massachusetts*

Compute Infrastructure
Chairpersons: N. Cox, *Vanderbilt University Medical Center, Nashville, Tennessee;* J. Goecks, *George Washington University, Ashburn, Virginia*

Algorithmics
Chairpersons: B. Langmead, *Johns Hopkins University, Baltimore, Maryland;* L. Pachter, *University of California, Berkeley*

Human Biology
Chairpersons: A. Battle, *Johns Hopkins University, Baltimore, Maryland;* B. Neale, *Massachusetts General Hospital, Boston*

Education Forum
Chairperson: M. Gerstein, *Yale University, New Haven, Connecticut*

Data Infrastructure
Chairpersons: B. Paten, *University of California, Santa Cruz;* V. Schneider, *National Center for Biotechnology Information, Bethesda, Maryland*

A. Tsherniak, P. Montgomery

T.H. Hwang, H. Hur

Transposable Elements

November 2–5 135 participants

ARRANGED BY **Rob Martienssen**, Cold Spring Harbor Laboratory
Phoebe Rice, University of Chicago, Illinois
Donald Rio, University of California, Berkeley

D. Rio, R. Martienssen, P. Rice

This first meeting held at the Cold Spring Harbor Laboratory brought together leading experts, junior and senior investigators, postdoctoral fellows, and graduate students working on many different aspects and systems of transposons. The six oral sessions and one poster session included microbial, animal, and plant transposon presentations. Material was presented at the structural, biochemical, biological, medical, and applied levels and spanned a variety of methods, approaches, and experimental systems. The session chairs were not instructed to give introductory comments at the beginning of each session, but some did and this was helpful because of the broad nature of the participants. There were approximately 150 people in attendance, with good numbers of female and younger, more junior, scientists participating. There was lively and extensive discussion following most of the talks and this was a highlight of the meeting, as was the poster session, which could have lasted 2–3 more hours. All in all, given the late advertising date and that this was the first time this meeting was held, it was a success.

PROGRAM

DNA Transposons
Chairperson: F. Dyda, *NIDDK, National Institutes of Health, Bethesda, Maryland*

Retrotransposons
Chairperson: J. Moran, *University of Michigan Medical School, Ann Arbor*

C. Lennon, W. Reznikoff

I. Arkhipova, P. Rice

P. Rocha (*left*)

R. Martienssen, T. Kakutani

Bacterial and Organellar Transposons
Chairperson: M. Belfort, *University at Albany, New York*

Transposon Control
Chairperson: K. Slotkin, *The Ohio State University, Columbus*

Host–Transposon Interactions
Chairperson: R. Lehmann, *New York University School of Medicine, New York*

Variety and Evolution of Transposable Elements
Chairperson: M. Chandler, *Université Paul Sabatier, Toulouse, France*

W. Reznikoff, M. Chandler

R. Lehmann, J. Boeke

Neurodegenerative Diseases: Biology and Therapeutics

November 30–December 3 188 participants

ARRANGED BY **Karen Duff**, Columbia University Medical Center, New York, New York
Richard Ransohoff, Biogen, Cambridge, Massachusetts
John Trojanowski, University of Pennsylvania School of Medicine, Philadelphia

K. Duff, R. Ransohoff, J. Trojanowski

Up to one-half of those aged 85 years or older will develop one or multiple debilitating neurodegenerative diseases (NDDs) of the central nervous system (CNS). These diverse NDDs include Alzheimer's disease (AD), Parkinson's disease (PD), frontotemporal lobe degeneration (FTLD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS), as well as combinations of these NDDs such as AD combined with PD and FTLD with ALS, whereas 50% of AD patients have TDP-43 inclusions that are associated with cognitive impairments. To reflect some of the newest and most relevant research while providing a diverse base of current advances in understanding these NDDs, the topics in the 2014 meeting included prion-like spread of diseases including AD, PD, and tauopathies, new biology of the ALS/FTLD *C9ORF72* mutation, new thoughts about biomarkers and co-morbidities, the role of RNA in the pathogenesis of neurodegenerative diseases, the role of glia (especially their potential as targets for therapy), and a number of therapeutic targets and approaches including immune therapy, small-molecule drugs, and gene inactivation strategies.

Although most forms of NDDS occur in the absence of obvious heritability or identifiable genetic mutations, it has been possible over the past 25 years to discover uncommon genetic mutations as well as risk-modifying DNA changes in some NDDs and predictable causative changes in others. For some of the NDDs, common genetic abnormalities lead to a spectrum of disease phenotypes, which is clearly seen for FTLD and ALS. New insights into disease mechanisms and possible overlapping therapeutic approaches are implicated. From these findings, transgenic technology has rapidly led to the development of mouse, fruit fly, and nematode model systems that partially recapitulate the clinical abnormalities of the human diseases, as well as some of the

V. Lee, R. Ransohoff

R. Ransohoff, R. Jaenisch

R. Jaenisch, V. Lee

S. Baltan, V. Hachinski

hallmark molecular and morphological pathology of the conditions. Of particular relevance in this respect is the emergence of new ideas about the spread of pathological disease proteins such as tau, α-synuclein, Aβ, and TDP-43 among others within the brain and from the periphery. This is similar to that seen in the prion diseases, which have become a dominant topic in AD, FTLD, and PD research, including the use of human postmortem brain-derived disease proteins as "proteopathic seeds" to propagate NDDs. New tools using human-induced pluripotent cells derived from these degenerative diseases were also discussed. This new and exciting field of research was well represented at the 2014 meeting.

This biannual meeting series started in 2000 under the organizing leadership of Sam Gandy, Harry Levine, and Marcie McDonald, and according to the plans for the launch of this new CSHL meeting, the explicit goals of the meetings are to focus on identifying disease pathways and facilitating the translation of "breakthrough" science into effective medicines by bringing together established and younger scientists as well as trainees from academia and pharmaceutical/biotech companies. At this year's meeting, numerous academic labs and commercial drug discovery organizations presented data on novel compounds, clinical trial results or new druggable pathways for AD, PD, FTLD, HD, ALS, spinal muscular atrophy, spinocerebellar ataxia, and prion diseases. Indeed, of the 179 participants, 19 came from industry. In recent years, various molecular, biochemical, and cell-based screens have led to the development of small-molecule, peptide, and oligonucleotide compounds that show promise in NDD models. Importantly, the emergence of gene inactivation strategies using stabilized antisense oligonucleotides has shown particular promise both as a tool for studying pathogenesis and, more importantly, as a clinically relevant therapeutic agent that could be used to knock down pathological proteins or pathways affected in several NDDs.

M.T. Heneka, J. Trojanowski

A. La Spada, V. Lee

S. Ebstein, V. Hachinski, D. Zamolodchikov

M. Diamond, L. Runum

For the 2016 meeting, platform sessions were organized around common technological themes (see attached Program). Chairpersons, invited speakers, and speakers selected from submitted abstracts were drawn from the academic and pharma sectors, representing 19 countries. The discussion of new, unpublished data was emphasized, and adequate time was left for discussion of each presentation. Poster presentations also covered a wide range of neurodegenerative disease pathways, new animal and insect models, and novel therapeutic insights. Posters were displayed for an extended period during the meeting, and poster viewing was especially encouraged during the cocktail hours before the evening meals. In general, the meeting was most notable for its breadth of coverage of different neurodegenerative diseases, insightful and novel presentations, and in-depth discussions of each presentation, and the insights gleaned from lively interaction among the diverse participants.

This meeting was funded in part by the National Institute of Neurological Disorders and Stroke, a branch of the National Institutes of Health, and by Biogen.

PROGRAM

Disease Propagation: Mechanisms and Therapeutic Opportunities
Chairperson: V. Lee, *University of Pennsylvania School of Medicine, Philadelphia*

Neuroinflammation
Chairperson: R. Ransohoff, *Biogen, Cambridge, Massachusetts*

RNA Metabolism in Neurodegenerative Disease
Chairperson: J.P. Taylor, *HHMI/St. Jude Children's Research Hospital, Memphis, Tennessee*

Proteostasis, Mechanisms, and Therapeutic Opportunities
Chairperson: D. Rubinsztein, *Cambridge Institute for Medical Research, United Kingdom*

New Neuronal Models: Ips, Organoids, Primary Cultures, Therapeutic Opportunities
Chairperson: S. Przedborski, *Columbia University, New York*

Vascular Aspects as a Cause or Contributor to Neurodegenerative Disease
Chairperson: V. Hachinski, *London Health Sciences Centre, Canada*

Imaging: Insights and Application
Chairperson: B. Dickerson, *Massachusetts General Hospital/ Harvard Medical School, Boston*

Gene Inactivation and Protein Lowering Strategies
Chairperson: T. Miller, *Washington University in St. Louis, Missouri*

V. Lee, M. Haney

Blood–Brain Barrier

December 7–10 132 participants

ARRANGED BY **Robert Bell**, Pfizer Inc., New York, New York
Chenghua Gu, Harvard Medical School, Boston, Massachusetts
Stefan Liebner, Goethe University Clinic, Germany

This meeting highlighted advances in the cellular and molecular aspects of blood–brain barrier (BBB) development, function, and disease. The conference was well attended, particularly by young scientists (graduate students and postdoctoral fellows), who also had a prominent role in giving oral presentations. This made for a lively discussion of unpublished data, giving foresight into the future of BBB research. Moreover, women scientists were well represented at the meeting. There were 57.9% female attendees, and many of them were session chairs and oral presenters.

R. Bell, C. Gu, S. Liebner

As in years past, the conference had a substantial representation of genetic model organisms, live imaging, and powerful molecular genetics highlighted in studying mammalian and nonmammalian BBB physiology and development. A major focus of the conference was the emerging field of molecular mechanisms underlying BBB permeability, with a focus on molecular mechanisms of transcytosis and transport functions. BBB modulation for drug transport and a balanced focus on the BBB in disease and therapeutic development were also discussed during the oral and poster presentations.

Two Keynote addresses were given by Berislav Zlokovic on BBB dysfunction and neurodegenerative diseases and by Britta Engelhardt on the interplay between BBB and immune cells, a relatively less understood process.

As is traditional at Cold Spring Harbor meetings, selection of material for oral and poster presentation was made by the organizers and individual session chairs on the basis of scientific merit.

V. Prevot, M. Rose

D. Irimia, A. Hartz

S. Storck, A. Iltzsche

S. Ramirez, J. Robert

PROGRAM

BBB Development
Chairpersons: C. Betsholtz, *Uppsala University, Sweden;* M. Blanchette, *University of California, San Diego*

Keynote Speaker
B.V. Zlokovic, *Zilkha Neurogenetic Institute, Keck School of Medicine of USC*

The Functional BBB
Chairpersons: R. Klein, *Washington University School of Medicine, St. Louis, Missouri;* V. Prevot, *INSERM, University of Lille, France*

Emerging BBB Model Systems
Chairpersons: D. Irimia, *Massachusetts General Hospital, Harvard Medical School, Boston*; A. Hartz, *University of Kentucky, Lexington, Kentucky*

Molecular BBB Transport
Chairpersons: I. Simpson, *Pennsylvania State University, Hershey, Pennsylvania;* M. Pizzo, *University of Wisconsin, Madison*

The Aging and Diseased BBB
Chairpersons: M. Schwaninger, *University of Lübeck, Germany;* A. Nordheim, *Eberhard-Karis University of Tübingen, Germany*

Keynote Speaker
B. Engelhardt, *University of Bern, Switzerland*

Functional Interactions at the Neurovascular Unit
Chairpersons: P. Carlsson, *University of Gothenburg, Sweden;* A. Brand, *The Gurdon Institute, University of Cambridge, United Kingdom*

R. Bell, J. Abbott

M. Hersom, D. Begley

POSTGRADUATE COURSES

Workshop on Leadership in Bioscience

March 11–14

INSTRUCTORS **C. Cohen**, Science Management Associates, Newton, Massachusetts
D. Kennedy, Worklab Consulting LLC, Cambridge, Massachusetts

This highly interactive 3.5-day workshop allowed students to develop the skills necessary to lead and interact effectively with others, in both one-on-one and group settings. The workshop focused on techniques, situations, and challenges that relate specifically to leading and managing in the scientific workplace. It emphasized learning by doing and involved role-playing, giving and receiving feedback, and group problem solving. Much of the learning was from peer to peer. Participants were expected to discuss their own experiences and listen to others as they discussed theirs. The workshop helped participants identify areas where they needed guidance and growth, as well as how to capitalize on areas of strength. Participants had the opportunity to share their experiences and challenges with others and received feedback and guidance from others with experience in leading scientists in a variety of settings. At the end of the course, participants linked through a unique online community in which they can continue learning from one another and from the course instructors.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Bohorquez, D., Ph.D., Duke University, Durham, North Carolina
Carvunis, A.-R., Ph.D., University of California San Diego, La Jolla
Chung, E.J., Ph.D., University of Chicago, Illinois
Cifuentes, D., Ph.D., Boston University, Massachusetts
Gribble, K., Ph.D., Marine Biological Laboratory, Woods Hole, Massachusetts
Guryanova, O., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Heier, C., Ph.D., Children's National Medical Center, Washington, D.C.
Hoban, M., Ph.D., Boston Children's Hospital, Boston, Massachusetts
Idoyaga, J., Ph.D., Stanford University, California
Izaguirre, M., Ph.D., Northern New Mexico College, Española
Joo, E., Ph.D., University of Illinois, Urbana
Kadlecova, Z., Ph.D., University of Texas Southwestern Medical Center, Dallas
Khan, O., M.S., University of Gothenburg, Institute of Medicine, London, United Kingdom
Lavelle, C., Ph.D., University of Florida, Gainesville
Lee, H.Y., Ph.D., Weill Cornell Medical College, New York
Lombard, Z., Ph.D., National Health Lab Service/Wits University, Johannesburg, South Africa
Malvankar, M., M.S., Ph.D., Yale University, West Haven, Connecticut
Neunuebel, J., M.S., University of Delaware, Newark
Niles, J., M.S., Ph.D., Susquehanna University, Selinsgrove, Pennsylvania
O'Donnell, M., M.D., Columbia University, New York, New York
Paduch, M., Ph.D., University of Chicago, Illinois
Perez de Arce, K., Ph.D., Tufts University, Boston, Massachusetts
Rister, J., Ph.D., New York University, New York
Romero Garcia, Z., Ph.D., University of California, Los Angeles
Shaw, D., Ph.D., University of Maryland, School of Medicine, Baltimore
Singh, S., Ph.D., Icahn School of Medicine at Mount Sinai, New York
Vogel, T., Ph.D., Washington University in Saint Louis, Missouri
Wang, L., Ph.D., University of Delaware, Newark

SEMINARS

C. Cohen, Science Management Associates, Newton, Massachusetts
D. Kennedy, Worklab Consulting LLC, Cambridge, Massachusetts

Session 1: Who we are.
Session 2: Leadership challenges: Case study overview.
Session 3: Difficult conversations and interactions: Fundamentals of negotiation.
Session 4: Case study analysis, Part 1.
Session 5: Case study analysis, Part 2.
Session 6: Scientific team and project meetings.
Session 7: Projecting leadership, Part 1.
Session 8: Projecting leadership, Part 2.
Session 9: Goal setting, feedback, and evaluation for scientists.
Session 10: Concluding group discussion.

Expression, Purification, and Analysis of Proteins and Protein Complexes

April 5–18

INSTRUCTORS **A. Courey**, University of California, Los Angeles, Los Angeles
S.-H. Lin, MD Anderson Cancer Center/University of Texas, Houston
M. Marr, Brandeis University, Waltham, Massachusetts
S. Nechaev, University of North Dakota School of Medicine, Grand Forks

ASSISTANTS **A. Abbas**, University of North Dakota School of Medicine and Health Sciences, Grand Forks
M. Chambers, University of California, Los Angeles
N. Clark, Brandeis University, Waltham, Massachusetts
M. Kim, University of California, Los Angeles
Y.-C. Lee, MD Anderson Cancer Center, Houston, Texas
A. Sawyer, Brandeis University, Waltham, Massachusetts
J. Warns, University of North Dakota School of Medicine, Grand Forks
T. Yau, University of California, Los Angeles

This course is for scientists, including graduate students, postdoctoral scholars, staff scientists, and principal investigators, who want a rigorous introduction to expression and purification of proteins as well as analysis of protein structure and function.

Through hands-on experience in the lab as well as extensive lectures and discussions, each student became familiar with key approaches in expression, purification, and analysis of soluble and membrane proteins and protein complexes from both natural sources and overexpression systems. The emphasis of the course was on the following.

1. *Approaches in protein expression*: Choosing the best bacterial or eukaryotic expression system tailored for the particular protein and experimental problem; determining how to optimize expression; understanding protein tagging: the advantages and pitfalls of various affinity and solubility tags.

2. *Approaches in protein purification*: Choosing the best strategy for a given protein including solubilization; bulk fractionation; liquid chromatography: including conventional methods (ion exchange, size exclusion, reverse phase, etc.) and affinity methods (e.g., MAC, DNA affinity, and immunoaffinity), as well as FPLC/HPLC.
3. *Approaches in protein analysis*: Introduction to common approaches for characterization of proteins including binding assays, activity assays, mass spectroscopy to identify protein interaction partners, and posttranslational modifications.

In addition to purification, students gained exposure to fundamental analytical approaches such as mass spectroscopy and protein structure determination (e.g., X-ray crystallography and cyro-EM).

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Arsiwala, A., B.Tech., Georgia Institute of Technology, Atlanta
Beguelin, W., Ph.D., Weill Cornell Medical College, New York, New York
Ezeuka, C., M.D., Emory University, Atlanta, Georgia
Kadgien, C., B.S., University of British Columbia, Vancouver, Canada
Kadzik, R., Ph.D., Princeton University, Pennsylvania
Kaushik, G., Ph.D., University of Kansas Medical Center, Kansas City
Kinney, J., Ph.D., Cold Spring Harbor Laboratory
Lin, Y.-C., M.S., University of Illinois at Urbana-Champaign, Urbana, Illinois
McFadden, K., B.S., Texas A&M Health Science Center, Bryan
Monti, D., Ph.D., University of Alabama, Birmingham
Orbach, R., Ph.D., Yale University, New Haven, Connecticut
Quezada, E., Ph.D., Yachay Tech, Urcuqui, Ecuador
Vyas, P., B.S., M.S., Technion-Israel Institute of Technology, Haifa, Israel
Wiley, P., Ph.D., National Cancer Institute, Bethesda, Maryland
Xu, J., B.S., McGill University, Montreal, Quebec, Canada

SEMINARS

Courey. A., University of California, Los Angeles: Welcome and introduction to course.
Marr, M., Brandeis University, Boston, Massachusetts: Introduction to protein purification.
Pappin, D., Cold Spring Harbor Laboratory: Introduction to mass spectrometry of proteins.
Jarvis, D., University of Wyoming, Laramie: An overview of the baculovirus-insect cell system for recombinant protein production.
Chappell, T., BioGrammatics, Inc., Carlsbad, California: Using the *Pichia pastoris* secretory pathway for heterologous protein expression.
Lin, S.-H., MD Anderson Center Center/University of Texas, Houston: Mediators of cross talk between tumor cells and their environment.
Marr, M., Brandeis University, Waltham, Massachusetts: Controlling gene expression in response to stress.
Love, J., DNA 2.0, Newark, California: Systematic approaches to engineering antibody and integral membrane protein expression.
Pappin, D., Cold Spring Harbor Laboratory: Quantitative approaches to mass spectrometry of proteins.
Nechaev, S., University of North Dakota, Grand Forks: A role of Pol II pausing in gene activation.
Courey, A., University of California, Los Angeles: Proteomic approaches to understanding SUMO.

Quantitative Imaging: From Cells to Molecules

April 5–18

INSTRUCTORS **H. Elliott**, Harvard Medical School, Boston, Massachusetts
J. Waters, Harvard Medical School, Boston, Massachusetts
T. Wittmann, University of California, San Francisco

CO-INSTRUCTOR **T. Lambert**, Harvard Medical School, Boston, Massachusetts

ASSISTANTS **T. Davies**, Columbia University Medical Center, New York, New York
A. Jost, Harvard Medical School, Boston, Massachusetts
D. Richmond, Harvard Medical School, Boston, Massachusetts
M. Weber, Harvard Medical School/Nikon Imaging Center, Boston, Massachusetts

Combining careful image acquisition with computational analysis allowed for extracting quantitative data from light microscopy images that is far more informative and reliable than what can be seen by eye. This course focused on advanced quantitative fluorescence microscopy techniques used for imaging a range of biological specimens, from cells to single molecules. The course was designed for cell and molecular biologists with little or no microscopy experience who wish to begin utilizing microscopy in their own research. Students gained a theoretical understanding of and hands-on experience with state-of-the-art equipment used in quantitative fluorescence microscopy, including wide-field fluorescence microscopy, laser-scanning and spinning-disk confocal microscopy, total internal fluorescence microscopy (TIRF), super-resolution methods (structured illumination, STED, STORM, and PALM), and digital image processing and analysis. Students learned how to design and implement a wide range of imaging experiments using these techniques. Students also learned fundamental image processing, segmentation, and analysis techniques using

a variety of commercial and open source software packages. Students used these image acquisition and analysis techniques to address specific quantitative questions and then discussed the results as a group, learning to troubleshoot the common problems that occur in the course of a quantitative imaging experiment. Among the lectures presented were quantitative microscopy basics, transmitted light microscopy, image segmentation, image analysis, CCD and sCMOS cameras, confocal microscopy, multiphoton microscopy, deconvolution, TIRF, imaging ratiometric "biosensors" (including FRET), light sheet microscopy, and super-resolution techniques. Students also learned guidelines for choosing fluorescent proteins and worked with live samples requiring environmental control.

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Aloisio, F., B.S., University of California, San Francisco
Canas Duarte, S., B.S., M.S., Harvard University, Boston, Massachusetts
Freed, E., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Hirsch, S., M.A., Columbia University, New York, New York
Indzhykulian, A., M.D., Harvard Medical School, Boston, Massachusetts
Kelley, M., B.A., The Rockefeller University, New York, New York
Kohrman, A., Ph.D., Stony Brook University, Stony Brook, New York
Kowal, A., Ph.D., Northwestern University, Chicago, Illinois
Lafranchi, L., Ph.D., Karolinska Institutet, Stockholm, Sweden
Mignardi, M., Ph.D., Stanford University, California
Opalko, H., Ph.D., Dartmouth College, Hanover, New Hampshire
Paz, E., Ph.D., University of California, San Diego, La Jolla
Quidwai, T., M.S., MRC Institute of Genetics and Molecular Medicine, Edinburgh, United Kingdom
Riestra, A., Ph.D., University of California, San Diego
Sari-Sarraf, F., M.S., Novartis Institutes for Bio Medical Research, Inc., Cambridge, Massachusetts
Shim, A., Ph.D., Northwestern University, Chicago, Illinois

SEMINARS

Waters, J., Harvard Medical School, Boston, Massachusetts: Quantitative microscopy basics; Objective lenses; Transmitted light microscopy; Fluorescence microscopy; Live-cell imaging; Quantitative confocal microscopy.
Lambert, T., Harvard University, Boston, Massachusetts: Digital imaging; Confocal microscopy: Theory and hardware.
Elliott, H., Harvard Medical School, Cambridge, Massachusetts: Image processing and analysis 101: Resolution, SNR, and diffraction-limited objects; Image processing and analysis 102: Image segmentation, binary operations, and tracking; Image processing and analysis 103: Advanced filtering and deconvolution; Quantifying fluorescence: Image arithmetic, image corrections, and ratiometry.
Shaner, N., The Scintillon Institute, San Diego, California: Fluorescent proteins.
Wittmann, T., University of California, San Francisco: Presenting quantitative microscopy data; Live confocal microscopy, FRAP, and photoactivation; TIRF.
Hillman, E., Columbia University, New York, New York: Multiphoton microscopy; Light sheet microscopy.
Huang, B., University of California, San Francisco: Super-resolution microscopy I; Super-resolution microscopy II.

Cell and Developmental Biology of *Xenopus*

April 5–18

INSTRUCTORS **M. Khokha**, Yale University, New Haven, Connecticut
K. Liu, King's College London, United Kingdom

ASSISTANTS **R. Huebner**, University of Texas, Austin
E. Mis, Yale University, New Haven, Connecticut
K. Pfister, University of Virginia, Charlottesville
B. Steventon, University of Cambridge, United Kingdom

Techniques covered in this course include microinjection and molecular manipulations such as CRISPR knockouts, transgenics, and mRNA overexpression. These can be combined with explant and transplant methods to simplify or test tissue-level interactions. To visualize subcellular and intercellular activities, we introduced a variety of imaging methods including time-lapse, fluorescent, and confocal microscopy. Additional methods included mRNA in situ hybridization and protein immunohistochemistry, as well as basic bioinformatic techniques for gene comparison and functional analysis.

This course was designed for those new to the *Xenopus* field, as well as for more advanced students who are interested in emerging technologies. We encouraged students to bring their own genes of interest and to tailor aspects of the course to enable them to initiate studies on their specific projects.

This course was supported with funds provided by the National Institute of Child Health and Human Development.

PARTICIPANTS

Baldwin, A., Ph.D., University of Texas, Austin
Borek, W., Ph.D., University of Edinburgh, United Kingdom
Chen, P., B.S., University of Wyoming, Laramie
Colleluori, V., B.S., Yale University, New Haven, Connecticut
Corkins, M., Ph.D., University of Texas Health, Houston
Dubiak, K., B.S., University of Notre Dame, Indiana
Khairallah, S., B.S., University of California, Santa Barbara
MacColl Garfinkel, A., B.A., Yale University School of Medicine, New Haven, Connecticut
Moreno, M., B.A., Federal University of Rio de Janeiro, Brazil
Nemes, P., Ph.D., George Washington University, Washington, D.C.
Pegge, J., B.S., King's College London, United Kingdom
Ramirez-Gordillo, D., Ph.D., University of Colorado Anschutz Medical Campus, Aurora
Sun, J., Ph.D., National Cancer Institute, Frederick, Maryland
Tingler, M., Ph.D., University of Hohenheim, Stuttgart, Germany
Verfaillie, A., M.A., Ph.D., The Francis Crick Institute, London, United Kingdom

SEMINARS

Keller, R., University of Virginia, Charlottesville: The integrated machines of *Xenopus* gastrulation.
Heald, R., University of California, Berkeley: Mechanisms of mitosis and size control in *Xenopus.*
Davidson, L., University of Pittsburgh, Pennsylvania: Cells and machines.
Cousin, H., University of Massachusetts, Amherst: Studying neural crest in *Xenopus*: Bringing molecular mechanisms to evo-devo questions.
El-Hodiri, H., Nationwide Children's Hospital Research Institute, Columbus, Ohio, Khokha, M., Yale University, New Haven, Connecticut, and Conlon, F., University of North Carolina, Chapel Hill: Proteomic-based approaches in *Xenopus* to development and disease.
Miller, A., University of Michigan, Ann Arbor: Regulation of localized RhoA activity and cell–cell junction remodeling in diving epithelial cells.
Wills, A., University of Washington, Seattle: Gene regulation during *Xenopus* regeneration.
Wallingford, J., University of Texas, Austin, and Thomsen, G., Stony Brook University, New York: *Xenopus* show and tell.

Workshop on Schizophrenia and Related Disorders

June 8–15

INSTRUCTORS **A. Abi-Dargham**, Columbia University College of Physicians & Surgeons, New York, New York
J. Hall, University of Cardiff School of Medicine, Cathays, United Kingdom
A. Sawa, Johns Hopkins University, Baltimore, Maryland

ASSISTANT **K. Delevich**, Cold Spring Harbor Laboratory

This workshop provided students with the most current understanding of the molecular, cellular, and neural systems underlying the disturbances in brain function in these devastating illnesses. During the 7-day workshop, students learned about the clinical aspects of schizophrenia, schizoaffective disorder, and bipolar disorder, and they explored in detail the genetic and neurobiological underpinnings of these complex psychiatric disorders. The Workshop included sessions focused on The Clinical Syndrome, Basic Neurobiology, Cognitive Neuroscience, Neuroimaging, Genetics and Genomics, Endophenotypes, Gene Expression, and Gene Modulation. In addition to hearing about the most recent research in these areas, controversial topics and challenges to basic assumptions in the field were explored and discussed. A diverse faculty brought the most up-to-date results and theories to the students, making this Workshop a valuable resource for young researchers starting out in this fast-moving and expansive field. Not only did it help them build the foundation for their future research, it also introduced them to many potential collaborators working to understand schizophrenia and similar disorders from different perspectives.

This course was supported with funds provided by Cold Spring Harbor Laboratory and the Stanley Foundation.

PARTICIPANTS

Ajnakina, O., Ph.D., Institute of Psychiatry, Psychology, and Neuroscience, London, United Kingdom
Alloza, C., B.S., Edinburgh University, United Kingdom
Brice, N., Ph.D., Takeda Cambridge, Cambridge, United Kingdom
Chang, E., Ph.D., Zucker Hillside Hospital, The Feinstein Institute, Manhasset, New York
Cooper, S., B.A., Temple University, Philadelphia, Pennsylvania
Furger, S., M.S., University of Bern, Switzerland
Just, D., M.S., KTH-Royal Institute of Technology/ SciLifeLab, Solna, Sweden
Moon, A., B.S., Cardiff University, United Kingdom
Neilson, E., M.S., University of Edinburgh, United Kingdom
Notter, T., University of Zurich, Switzerland
Pass, R., B.S., Cardiff University, United Kingdom
Quattrone, D., M.S., King's College London, United Kingdom
Sinkeviciute, I., M.D., Haukeland University Hospital, Bergen, Norway
Smolders, S., M.S., Hasselt University/Catholic University Leuven, Diepenbeek, Belgium
Stahnke, A., M.S., University of Bern, Switzerland
Stay, T., B.S., Baylor College of Medicine, Houston, Texas
Stevenson, A., M.S., Massachusetts General Hospital, Boston
Trotta, A., Ph.D., King's College London, United Kingdom
Tufvesson Alm, M., M.S., Karolinska Institute, Stockholm, Sweden
Vanes, L., B.S., Institute of Psychiatry, King's College London, United Kingdom
Zengaffinen, F., M.A., Psychiatric Services University, Bern, Switzerland

SEMINARS

Hall, J., Cardiff University, United Kingdom: Overview.
Murray, R., King's College, London, United Kingdom: The clinical syndrome.
Leiberman, J., Columbia University, New York, New York: Treatment.
Jones, P., University of Cambridge, United Kingdom: Epidemiology.
Cannon, T., Yale University, New Haven, Connecticut: The Prodrome.
Seidman, L., Harvard Medical School, Boston, Massachusetts: Lifespan developmental evolution of neurocognition in schizophrenia.
Fletcher, P., University of Cambridge, United Kingdom: Neuropsychology.
Owen, M., Cardiff University, United Kingdom: Genetics.
Weinberger, D., Lieber Institute for Brain Development, Baltimore, Maryland: Genetics to biology.
Kenny, P., Icahn School of Medicine, Mount Sinai, New York: Epigenetics.
Krystal, J., Yale University, New Haven, Connecticut: Glutamate system.
Law, A., University of Colorado, Denver: Gene expression to biology.
Sawa, A., Johns Hopkins University, Baltimore, Maryland: Translational approaches.
Anderson, S., University of Pennsylvania, Philadelphia: Molecular studies of the GABA system.
Moghaddom, R., University of Pittsburgh, Pennsylvania: Preclinical studies of the glutamate system.
Moore, H., Columbia University, New York, New York: Animal models in psychiatry.
Brennand, K., Mount Sinai School of Medicine, New York: iPSC cells in psychiatry.
Malhotra, A., The Zucker Hillside Hospital, Glen Oaks, New York: Pharmacogenetics.
Lewis, D., University of Pittsburgh, Pennsylvania, and Sohal, V., University of California, San Francisco: Circuit analysis and optogenetics.
Abi-Dargham, A., Columbia University, New York: PET imaging in schizophrenia.
Anticevic, A., Yale University, New Haven, Connecticut: MR methods and psychiatry.
O'Donnell, P., Pfizer, Boston, Massachusetts: Science in industry.
Heimer, H., Schizophrenia Research Forum, Providence, Rhode Island: Schizophrenia research forum.
All Instructors: Career discussion.

Single-Cell Analysis

June 8–21

INSTRUCTORS
A. Herr, University of California, Berkeley
M. McConnell, University of Virginia School of Medicine, Charlottesville
G. Yeo, University of California, San Diego, La Jolla

CO-INSTRUCTOR
I. Burbulis, University of Virginia, Charlottesville

ASSISTANTS
C. Aonbangkhen, University of Pennsylvania, Philadelphia
O. Botvinnik, University of California, San Diego, La Jolla
M. Haakenson, University of Virginia School of Medicine, Charlottesville
Y. Song, University of California, San Diego, La Jolla
J. Vlassakis, University of California, Berkeley
M. Wierman, University of Virginia, Charlottesville
M. Wolpert, University of Virginia McConnell Lab, Charlottesville
D. Wu, University of Pennsylvania, Philadelphia
K. Yamauchi, University of California, Berkeley

The goal of this 2-week course was to familiarize students with cutting-edge technologies for characterization of single cells. The modules of the course were taught by scientists with expertise in distinct areas of single-cell analysis. Topics covered included quantitative single-cell analysis by RNA-Seq, genomic DNA analysis, proteomics, and metabolomics. Multiple nucleic amplification methodologies including aRNA, MALBAC, and MDA were used. In addition, students were instructed in basic bioinformatic analysis of next-generation sequencing data.

Topics included single-cell genome, transcriptome, and proteome measurement, introductory next-generation sequencing data analysis, photoactivatable single-cell probes, and single-cell mass spectrometry/soft X-ray tomography.

This course was supported with funds provided by the National Institute of General Medical Sciences, the Howard Hughes Medical Institute, and the Helmsley Charitable Trust.

PARTICIPANTS

Brian, D., M.B., B.S., University College London, United Kingdom
Desalvo, A., M.S., B.S., University of Southampton, United Kingdom
DiSalvo, M., B.S., University of North Carolina, Chapel Hill
Duda, J., Ph.D., Ulm University, Germany
Goossens, E., M.S., Purdue University, West Lafayette, Indiana
Kaelberer, M., Ph.D., Duke University, Durham, North Carolina
Liedmann, S., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
Michailidis, E., Ph.D., The Rockefeller University, Columbia, Missouri
Niehaus, J., B.S., University of North Carolina Chapel Hill
Nunez, D., M.S., Ph.D., University of Michigan, Ann Arbor
Reardon, P., B.A., NIMH, National Institutes of Health, Bethesda, Maryland
Sala Frigerio, C., Ph.D., KU Leuven, Belgium
Sood, C., M.S., University of Virginia, Charlottesville
Trejo, C., Ph.D., Salk Institute for Biological Studies, San Diego, California
Varma, P., Ph.D., University of Texas Southwestern Medical Center, Dallas
Viragova, S., B.S., Columbia University Medical Center, New York, New York
Williams, E., Ph.D., Max-Planck-Institute for Developmental Biology, Tübingen, Germany
Yang, H., B.S., The Rockefeller University, New York, New York

SEMINARS

Allbritton, N., University of North Carolina, Chapel Hill: Micro-assays for the single cell.
Yeo, G., University of California, San Diego, La Jolla: Single-cell alternative splicing analysis by expedition.
Faulkner, G., University of Queensland, Brisbane, Australia: Somatic L1 retrotransposition in mammals.
McConnell, M., University of Virginia School of Medicine, Charlottesville: Mosaic copy-number variants in human neurons.
McCarroll, S., Harvard Medical School, Broad Institute, Boston, Massachusetts: Neurons, droplets, and the mammalian brain.
Chenoweth, D., University of Pennsylvania, Philadelphia: Chemical tools for spatial and temporal control in biology.
Herr, A., University of California, Berkeley: Targeted proteomics with single-cell and subcellular resolution.
Chun, J., The Scripps Research Institute, La Jolla, California: Genomic and transcriptomic mosaicism among single cells of the brain.
Liss, B., University of Ulm, Germany: Quantitative RNA and genomic DNA integrity analysis of individual neurons.
Bendall, S., Stanford University School of Medicine, Palo Alto, California: Massively multiplexed single-cell analysis in human systems.
Sims, P., Columbia University, New York, New York: Scalable microfluidics for linking single cell microscopy and sequencing.

Advanced Bacterial Genetics

June 8–28

INSTRUCTORS
L. Bossi, Institute of Integrative Biology of the Cell (I2BC), Paris, France
A. Camilli, Tufts University Medical School, Boston, Massachusetts
H. Merrikh, University of Washington, Seattle

ASSISTANTS
R. Balbontin Soria, Instituti Gulbekian de Ciencia, Oeiras, Portugal
N. Figueroa-Bossi, Institut de Biologie Integrative de la Cellule, Gif-Sur-Yvette, France
C. Merrikh, University of Washington, Seattle
P. Nugent, University of Washington, Seattle
C. Silva, Tufts University, Canton, Massachusetts

This course presented logic and methods used in the genetic dissection of complex biological processes in diverse bacteria. Laboratory methods included classical and cutting-edge mutagenesis using transposons, mutator strains, and error-prone PCR; recombineering with single- and double-stranded DNA; CRISPR-Cas genome editing; genome sequencing and assembly; detection of gene expression changes using reporter genes; mapping mutations using genetic and physical techniques; modern approaches to the generation and analysis of targeted gene disruptions and fusions using PCR and cloning methods; linker-scanning mutagenesis; site-directed mutagenesis; and fluorescence microscopy. Key components of the course were the use of sophisticated genetic

methods in the analysis of model bacteria (including *Escherichia coli, Salmonella, Bacillus subtilis,* and *Vibrio cholerae*), and the use of the wealth of new genomic sequence information to motivate these methods.

Invited lecturers presented various genetic approaches to study bacterial mechanisms of metabolism, development, and pathogenesis.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Bryan, L., D.V.M., Texas A&M University, CVM, College Station
DeAngelis, C., B.A., University of Toledo, Ohio
Dillard, R., B.S., Emory University, Atlanta, Georgia
Halvorsen, T., B.S., University of California, Santa Barbara, Isla Vista
Jagodnik, J., B.A., CNRS, Paris, France
Joshi, A., Ph.D., University of California, Santa Cruz
Karlsson, J., M.S., Karolinska Institutet, Solna, Sweden
Kelly, P., M.S., The Ohio State University, Columbus
Rao Krishnamurthi, V., Ph.D., University of Arkansas, Fayetteville
Seara, D., B.A., Yale University, New Haven, Connecticut
Silvis, M., Ph.D., University of California, San Francisco
Søndberg, E., B.A., University of Copenhagen-Faculty of Science, UCP, Copenhagen, Denmark
Srinivasan, S., Ph.D., Harvard University, Cambridge, Massachusetts
Tamar, E., M.S., Technion-Israel Institute of Technology, Haifa, Israel
Wadhwa, N., Ph.D., Harvard University, Cambridge, Massachusetts
Zhang, Y., Ph.D., University of Massachusetts Medical School, Worcester

SEMINARS

Miller, S., University of Washington, Seattle: Regulation of the *Salmonella* outer membrane.
Hochschild, A., Harvard Medical School, Boston, Massachusetts: Detecting protein–protein interactions in bacteria.
Hughes, D., Uppsala University, Sweden: Using genetic approaches to understand tuf sequence and regulation.
Hughes, K., University of Utah, Salt Lake City: How do bacteria build a flagellum?; Revised genetic code of life.
Lazazzera, B., University of California, Los Angeles: Cell–cell signaling and biofilm formation.
Slauch, J., University of Illinois, Urbana: Uncovering *Salmonella* biology through mutational analysis.
Vanderpool, C., University of Illinois, Urbana: Understanding the role of small RNAs during metabolic stress.
Court, D., National Cancer Institute, Frederick, Maryland: Recombineering: Genetic engineering in bacteria using homologous recombination.
Courcelle, J., Portland State University, Oregon: The usually faithful process of genome duplication.

Ion Channels in Synaptic and Neural Circuit Physiology

June 8–28

INSTRUCTORS **T. Branco**, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
I. Duguid, University of Edinburgh, Midlothian, United Kingdom
C. Schmidt-Hieber, University College London, United Kingdom

ASSISTANTS **K. Bittner**, Janelia Research, Ashburn, Virginia
L. Brosse, CNRS, Marseille, France
M. Jelitai, IEMHAS, Budapest, Hungary
K.V. Kuchibhotla, New York University School of Medicine, New York
P. Stemkowski, University of Calgary, Canada
V. Stempel, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
A. Tozer, MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
N. Wanaverbecq, Aix-Marseille University, Marseille, France

Ion channels are the fundamental building blocks of excitability in the nervous system. The primary goal of this course was to demonstrate, through lectures and laboratory work, the different biophysical properties of ion channels that enable neurons to perform unique physiological functions in a variety of neural systems.

Areas of particular interest included (1) voltage- and ligand-gated ion channels at central and peripheral synapses, (2) synaptic integration and plasticity, (3) neural circuit function in vitro and in vivo, and (4) optogenetic strategies for circuit manipulation. A typical day consisted of morning lectures followed by hands-on laboratory practical sessions in the afternoon and evening with guest lecturers available to give one-on-one practical advice.

The laboratory component of the course introduced students to state-of-the-art electrophysiological approaches for the study of ion channels in their native environments. The course provided students with hands-on experience in using patch-clamp electrophysiology to examine single-channel activity in cultured cells, ion channel biophysics in acutely dissociated neurons and synaptic integration, plasticity and circuit dynamics in in vitro slice and in vivo preparations. Different recording configurations were used (e.g., cell-attached, whole-cell dendritic and somatic patch, voltage- and current-clamp configurations) and the advantages and limitations of each method were discussed in relation to specific scientific questions. The course also provided practical experience in cellular and circuit manipulation techniques (i.e., pharmacological, electrophysiological, and optogenetic) both in vitro and in vivo.

Admissions priority was given to students and postdocs who could show a demonstrated interest and specific plans to apply these techniques to a defined scientific problem.

This course was supported with funds provided by the Helmsley Charitable Trust and the Howard Hughes Medical Institute.

PARTICIPANTS

Dorsett, C., B.S., University of North Carolina, Chapel Hill
Dube, S., B.S., Duke University, Morrisville, North Carolina
Gammons, J., Ph.D., University of Tennessee Health Science Center, Memphis
Huoviala, P., M.S., M.A., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom
Labouesse, M., Ph.D., ETH Zurich, Schwerzenbach, Switzerland
Lee, M., Ph.D., University of California, Berkeley
Orefice, L., Ph.D., Harvard Medical School, Boston, Massachusetts
Tay, A., B.S., University of California, Los Angeles
Antonietta Tosches, M.A., Ph.D., Max-Planck Institute for Brain Research, Frankfurt am Main, Germany
Traunmüller, L., M.S., Biozentrum University of Basel, Switzerland
Wang, Y., B.S., University of Utah, Salt Lake City
Wixted, E., B.S., University of Michigan, Ann Arbor

SEMINARS

Instructors: Excitable cells. Cells as an electrical circuit. Intro: Channel structure and function. ChRs compensation/amplifiers/bridge.
Larsson, P., University of Miami, Florida: Overview of channels.
Nimigean, C., Weill Cornell Medical College, New York, New York: K^+ channels.
Clark, B., University College London, United Kingdom: Na^+ channels.
Kammermeier, P., University of Rochester, New York: Ca^{2+} channels.
Plested, Leibnitz, A., Institute for Molecular Pharmacology, Berlin, Germany: Single channels.
Xu-Friedman, M., University of Buffalo, New York: EPSCs.
Farrant, M., University College London, United Kingdom: IPSCs.
Tinatin Brelidze, T., Georgetown University, Washington, D.C.: Ih.
Jeffrey Diamond, J., National Institutes of Health, Bethesda, Maryland: Synaptic integration.
Jesper Sjostrom, J., McGill University, Montreal, Quebec, Canada: Synaptic plasticity.
Adam Carter, A., New York University, New York: Circuits.
Nagel, G., Julius-von-Sachs-Institut, Wurzburg, Germany: ChR2- Biophysics.
Gradinaru, V., California Institute of Technology, Pasadena: ChR2 circuits, synaptic integration.
Cohen, J., The Johns Hopkins University, Baltimore, Maryland: In vivo—LFP/extracellular/probes.
Petersen, C., Ecole Polytechnique Federale de Lausanne, Switzerland: In vivo—Intracellular.
Schiller, J., Technion Institute of Technology, Haifa, Israel: Plenary.

Mouse Development, Stem Cells, and Cancer

June 8–28

INSTRUCTORS **M. Lewandoski**, National Cancer Institute, Frederick, Maryland
D. Wellik, University of Michigan Medical Center, Ann Arbor

CO-INSTRUCTORS **B. Allen**, University of Michigan Medical School, Ann Arbor
A. Ralston, Michigan State University, E. Lansing

ASSISTANTS **C. Cebrian**, University of Michigan, Ann Arbor
T. Frum, Michigan State University, E. Lansing
A. Lokken, Michigan State University, E. Lansing
L. Marty Santos, University of Michigan, Ann Arbor
J. Ortiz-Guzman, Baylor College of Medicine, Fresno, Texas
A. Parenti, Michigan State University, E. Lansing
J. Song, University of Michigan, Ann Arbor

This intensive lecture and laboratory course was designed for scientists interested in using mouse models to study mammalian development, stem cells, and cancer. The lecture portion of the course, taught by leaders in the field, provided the conceptual basis for contemporary research in embryogenesis, organogenesis in development and disease, embryonic, adult, and induced pluripotent stem cells, and cancer biology.

The laboratory and workshop portions of the course provided hands-on introduction to engineering of mouse models, stem cell technologies, and tissue analyses. Experimental techniques included genome editing by CRISPR-Cas9, pronuclear microinjection, isolation and culture/manipulation of pre- and postimplantation embryos, embryo transfer, embryo electroporation and

roller bottle culture, chimera generation, and generation and differentiation of mouse embryonic stem cells, as well as induced pluripotent stem cells and fibroblasts, vibratome and cryosectioning, in situ RNA hybridization, immunostaining, FACS sorting and analysis, vascular injections and casting, neural sterotaxic injections, skeletal preparation, organ explant culture, and fluorescent imaging, including live time-lapse microscopy.

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Hermosilla, V., M.S., Universidad de Concepción, Concepción, Chile
Jagoda, E., B.A., Harvard University, Cambridge, Massachusetts
Laumann-Lipp, N., M.S., Institute of Molecular Biotechnology, Vienna, Austria
Mahauad-Fernandez, W., Ph.D., University of Iowa, Iowa City
Manaf, A., B.S., Oslo University Hospital, Rikshospitalet Regnskap, Oslo, Norway
Navarro, A., B.S., Karolinska Institute, Stockholm, Sweden
Ngo, J., B.A., Case Western Reserve University, Cleveland, Ohio
Patel, S., Ph.D., Memorial Sloan Kettering Cancer Center/ Clarkson University, New York, New York
Rozhkova, E., Ph.D., Cold Spring Harbor Laboratory
Saito, A., Ph.D., Brazilian Biosciences National Laboratory, Campinas SP, Brazil
Saksena, J., B.S., Tulane University, New Orleans, Louisiana
Schanne, D., M.D., Massachusetts General Hospital, Charlestown
Singh, V., Ph.D., Stowers Institute for Medical Research, Kansas City, Missouri
Zafra Martin, M., Ph.D., Meyer Cancer Center/Weill Cornell Medicine, New York, New York

SEMINARS

Lewandoski, M., National Cancer Institute, Frederick, Maryland: Welcome and introduction to course
Threadgill, D., Texas A&M University, College Station: Historical overview of the mouse as a model system; 100+ years of the mouse in research with Mouse Genetics 101; Quantitative genetics applied to developmental biology.
Hadjantonakis, K., Memorial Sloan Kettering Cancer Institute, New York: Optical imaging of mouse embryos, early mouse development.
Mager, J., University of Massachusetts, Amherst: Epigenetic programming in the early mammalian embryo.
Ralston, A., Michigan State University, E. Lansing: Stem cell origins in the mouse blastocyst.
Capecchi, M., University of Utah School of Medicine, Salt Lake City: Mouse models of human disease from cancer to neuropsychiatric disorders.
Benjamin Allen, B., University of Michigan Medical School, Ann Arbor: Hedgehog signaling in embryogenesis and cancer.
Ondine Cleaver, O., University of Texas Southwestern Medical Center, Dallas: Morphogenesis of biological tubes: From blood vessels to pancreas.
Costantini, F., Columbia University, New York, New York: Branching morphogenesis during kidney development.
Tam, P., The University of Sydney, Westmead, Australia: From gastrulation to head formation: Tissue patterning, cell fates, and gene regulatory work.
Kopan, R., Cincinnati Children's Hospital Medical Center, Ohio, Trainor, P., Stowers Institute for Medical Research, Kansas City, Missouri, and Wellik, D., University of Michigan Medical Center, Ann Arbor: Mesenchymal stem cells: Hox regulation in development and disease.
Lovell-Badge, R., The Francis Crick Institute, London, United Kingdom: Sex determination and differentiation.
Vezina, C., University of Wisconsin, Madison: The developing prostate: A window to understanding a costly benign disease.
Arenkiel, B., Baylor College of Medicine, Houston, Texas: Synapses to satiety: Conditional genetic approaches to uncover brain circuits.
Maillard, I., University of Michigan, Ann Arbor: Hematopoiesis: Development, cancer, immunity; FACS analysis workshop.
Egeblad, M., Cold Spring Harbor Laboratory: Intravital imaging in living mice; visualizing drug resistance and metastasis.
Felsher, D., Stanford University School of Medicine, California: Modeling and predicting oncogene addiction.
Huppert, S., Cincinnati Children's Hospital Medical Center, Ohio, and Sun, X., University of Wisconsin, Madison: Linking lung development to physiology.
Cebrian, C., University of Michigan, Ann Arbor: Image processing and quantification workshop.
Selleri, L., University of California, San Francisco: ESCRT complex in morphogenesis.
Hinck, L., University of California, Santa Cruz: From milk to malignancy: The role of mammary stem cells.

Proteomics

June 15–28

INSTRUCTORS
M. Bereman, North Carolina State University, Raleigh
M. Cilia, Cornell University, Ithaca, New York
I. Cristea, Princeton University, New Jersey
K. Medzihradszky, University of California, San Francisco
D. Pappin, Cold Spring Harbor Laboratory

ASSISTANTS
M. Crow, Princeton University, New Jersey
T. Greco, Princeton University, New Jersey
E. Klement, Biological Research Centre Szeged, Hungary
K. Lum, Princeton University, New Jersey
A. Makarenko, Cold Spring Harbor Laboratory
K. Rivera, Cold Spring Harbor Laboratory
N. Saha Turna, Cold Spring Harbor Laboratory
J. Wilson, Cold Spring Harbor Laboratory

This intensive laboratory and lecture course was focused on cutting-edge proteomic approaches and technologies. Students gained practical experience purifying and identifying protein complexes and posttranslational modifications. In a section focused on quantitative whole-proteome analyses or top-down proteomics, students gained hands-on experience using two-dimensional gel electrophoresis and mass spectrometry analysis. Students used differential in-gel electrophoresis (DIGE) for gel-based protein quantification. Differentially expressed proteins were recognized by statistical methods using advanced gel analysis software and were identified using MALDI mass

spectrometry. For shotgun proteomic analysis sections or bottom-up proteomics, students used label-free and covalent isotopic-labeling quantitative approaches to differentially profile changes in protein complexes and whole proteomes. Students were trained in high-sensitivity microcapillary liquid chromatography coupled with nanospray-ESI and tandem mass spectrometry analysis. Students learned both single-dimension and multidimensional separation methods. In a section focused on targeted proteomics, students learned to analyze and process shotgun proteomic data for the development of SRM/MRM assays that accurately identify and quantify targeted proteins. Students designed transitions for selected peptides and performed SRM/MRM mass spectrometry assays. They learned to process and interpret the acquired data to measure changing quantities of targeted proteins in a variety of biological samples. For all sections of the course, a strong emphasis was placed on data analysis.

A series of outside lecturers discussed various proteomics topics including imaging by mass spectrometry, de novo sequence analysis, advanced mass spectrometry methods, protein arrays, and functional proteomics. The aim of the course was to provide each student with the fundamental knowledge and hands-on experience necessary for performing and analyzing proteomic experiments. The overall goal was to train students to identify new opportunities and applications for proteomic approaches in their biological research.

This course was supported with funds provided by the National Institute of Child Health and Human Development.

PARTICIPANTS

Allen, T., M.B.B.S., Duke University Hospital, Durham, North Carolina
Khurgel, M., Ph.D., Bridgewater College, Virginia
Kumar, A., Ph.D., University of Miami, Florida
Lampropoulos, A., M.S., Heidelberg University, Germany
Larsson, M., Ph.D., Uppsala University, Sweden
Maity, S., Ph.D., New York University, New York
Morozov, Y., Rowan University SOM, Stratford, New Jersey
Pathmasiri, W., Ph.D., RTI International, Research Triangle Park, North Carolina
Perez-Neut, M., B.S., University of Chicago, Illinois
Pineyro-Ruiz, C., M.S., University of Puerto Rico, Medical Science Campus, San Juan
Pollina, E., Ph.D., Harvard Medical School, Boston, Massachusetts
Post, S., B.S., Brown University, Providence, Rhode Island
Rider, M., Ph.D., Florida State University, Tallahassee
Tokmina-Roszyk, D., M.S., Florida Atlantic University, Jupiter
Tyson, K., B.S., Washington State University, Pullman
Vieira Parrine Sant'Ana, D., M.S., McGill University, Saint Anne de Bellevue, Canada

SEMINARS

Bruce, J., University of Washington, Seattle: The quantitative interactome: Finding edges in a functional landscape.
Claser, K., Broad Institute of MIT and Harvard, Cambridge, Massachusetts: Manual de novo peptide MS/MS interpretation for evaluating database search results.
Conlon, F., University of North Carolina, Chapel Hill: Proteomics applications to vertebrate development and disease.
Hawkridge, A., Virginia Commonwealth University, Richmond: Success and failure in mass spectrometry-based biomarker discovery; Protein digestion: Do tryptic peptide levels reflect original protein abundance?
Nesvizhskii, A., University of Michigan, Ann Arbor: Computational analysis of affinity purification-mass spectrometry data.
Vitek, O., Northeastern University, Boston, Massachusetts: Statistical methods for detecting differentially abundant proteins.

Statistical Methods for Functional Genomics

June 23–July 6

INSTRUCTORS
H. Bussemaker, Columbia University, New York, New York
S. Davis, NIH Center for Cancer Research Genetics Branch, Columbia, Maryland
R. Irizarry, Dana-Farber Cancer Institute, Boston, Massachusetts
T. Lappalainen, New York Genome Center/Columbia University, New York

ASSISTANTS
M. Brandt, New York Genome Center/Columbia University, New York
V. FitzPatrick, Columbia University, New York, New York
S. Kim-Hellmuth, New York Genome Center, New York
K. Korthuer, Dana-Farber Cancer Institute, Boston, Massachusetts
J. Kribelbauer, Columbia University, New York, New York
J. Paulson, Dana-Farber Cancer Institute, Boston, Massachusetts

Over the past decade, high-throughput assays have become pervasive in biological research because of both rapid technological advances and decreases in overall cost. To properly analyze the large data sets generated by such assays and thus make meaningful biological inferences, both experimental and computational biologists must understand the fundamental statistical principles underlying analysis methods. This course was designed to build competence in statistical methods for analyzing high-throughput data in genomics and molecular biology.

Topics included:

- the R environment for statistical computing and graphics
- introduction to Bioconductor
- review of basic statistical theory and hypothesis testing
- experimental design, quality control, and normalization
- high-throughput sequencing technologies
- expression profiling using RNA-Seq and microarrays

- in vivo protein binding using ChIP-seq
- high-resolution chromatin footprinting using DNase-seq
- DNA methylation profiling analysis
- integrative analysis of data from parallel assays
- representations of DNA-binding specificity and motif discovery algorithms
- predictive modeling of gene regulatory networks using machine learning
- analysis of posttranscriptional regulation, RNA binding proteins, and microRNAs

The format included detailed lectures and presentations by instructors and guest speakers combined with hands-on computer tutorials. The methods covered in the lectures were applied to example high-throughput data sets.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Biette, K., B.S., Harvard Medical School, Boston, Massachusetts
Capell, B., M.D., Ph.D., University of Pennsylvania, Philadelphia
Choi, J., Ph.D., Max-Planck-Institute for Biophysical Chemistry, Goettingen, Germany
Ciernia, A., Ph.D., University of California, Davis
Cortini, R., M.S., Ludwig Maximilian University of Munich, Germany
Dawicki-McKenna, J., Ph.D., University of Pennsylvania, Philadelphia
Domcke, S., Friedrich Miescher Institute, Basel, Switzerland
Ehmsen, J., M.D., Ph.D., Johns Hopkins School of Medicine, Baltimore, Maryland
Forno, E., M.D., University of Pittsburgh, Pennsylvania
Fuxman Bass, J., Ph.D., University of Massachusetts Medical School, Worcester, Hutchinson, K., Ph.D., Vanderbilt University, Nashville, Tennessee
Kundakovic, M., Ph.D., Fordham University, Bronx, New York
Lu, X., Ph.D., The University of Texas MD Anderson Cancer Center, Houston
Ma, J., B.S., Purdue University, West Lafayette, Indiana
Mellen, M., Ph.D., The Rockefeller University, New York
Normanly, J., B.A., Ph.D., University of Massachusetts, Amherst
Roa, S., Ph.D., Center for Applied Medical Research (CIMA-UNAV), Pamplona Navarra, Spain
Sapkota, D., Ph.D., Washington University, St. Louis, Missouri
Schiff, S., Ph.D., Penn State University, University Park, Pennsylvania
Sen, N.D., Ph.D., NICHD, National Institutes of Health, Bethesda, Maryland
Kumar Singh, P.K., National Institutes of Health, Bethesda, Maryland
Slansky, J., Ph.D., University of Colorado School of Medicine, Aurora
Sparks, R., M.D., National Institutes of Health, Bethesda, Maryland
Stelekati, E., Ph.D., University of Pennsylvania, Philadelphia

SEMINARS

FitzPatrick, V., Columbia University, New York, New York: Introduction to R.
Bussemaker, H., Columbia University, New York, New York: Normal distribution and multiple testing; Scoring gene ontology association using Fisher's exact testing; Introduction to eQTLs; Basic ChIP-seq analysis using Poisson statistics.
Cox, N., Vanderbilt University, Nashville, Tennessee: Data integration: Genome X transcriptome x EMR, theory, and applications.
Lappalainen, T., Columbia University, New York, New York: Basics of high-throughput sequencing; Testing for allele-specific expressions; Integrating transcriptome and genome sequencing to understand functional variation in human genome.
Siepel, A., Cold Spring Harbor Laboratory: Hidden Markov models for functional genomics.
Engelhardt, B., Princeton University, New Jersey: Linear regression; Structured sparsity and latent factor models in genomics.
Davis, S., National Institutes of Health, Columbia, Maryland: RNA-seq data analysis.
Kundaje, A., Stanford University, California: How to train your DRAGONN (deep regulatory genomic neural network).
Irizarry, R., Dana-Farber Cancer Institute, Boston, Massachusetts: Exploratory data analysis; High-dimensional data; Batch effects.

Advanced Techniques in Molecular Neuroscience

July 1–16

INSTRUCTORS
C. Lai, Indiana University, Bloomington
J. LoTurco, University of Connecticut, Storrs
A. Schaefer, Icahn School of Medicine, Mount Sinai, New York

ASSISTANTS
A. Battison, University of Connecticut, Storrs
J. Fang, University of Connecticut, Willington
E. Perez, Indiana University, Bloomington
M. Pozsgai, Indiana University, Bloomington
M. Sakai, Indiana University, Bloomington

This laboratory and lecture course was designed to provide neuroscientists at all levels with a conceptual and practical understanding of several of the most advanced techniques in molecular neuroscience. The course curriculum was divided into three sections: an extensive and up-to-date set of laboratory exercises, daily lectures covering the theoretical and practical aspects of the various methods used in the laboratory, and a series of evening research seminars. The informal and interactive evening lectures were given by leading molecular neuroscientists and served to illustrate the ways in which the various experimental approaches have been used to advance specific areas of neurobiology. In this year's course, the laboratory portion included topics such as an introduction to the design and use of animal virus vectors in neurobiology; the use of CRISPR genome editing and RNAi approaches for regulating the expression of specific genes in neurons; practical exercises in gene delivery systems including mammalian cell infection and transfection and electroporation techniques for targeted gene transfer in vivo; an introduction to overall strategies, use, and design of BAC transgenic vectors; RT-PCR analyses; assays of chromatin and chromatin structure in neurons; and mRNA isolation from specified neural subtypes by TRAP.

This course was supported with funds provided by the National Institute of Mental Health.

PARTICIPANTS

Albert Lim, J.-H., B.S., University of California, San Diego, La Jolla
Alsina, F., Ph.D., IBCN, CONICET-University of Buenos Aires, Argentina
Comer, A., B.S., Boston University, Massachusetts
Ferguson, B., B.A., Drexel University College of Medicine, Philadelphia, Pennsylvania
Foster, A., B.S., Oregon Health and Science University, Portland
Hardigan, A., M.D., Ph.D., University of Alabama, Birmingham
Holt, L., B.S., University of Alabama, Birmingham
Huang, S., B.S., University of Iowa, Iowa City
Joseph, N., M.S., The Scripps Research Institute, Jupiter, Florida
Leloup, N., Ph.D., Utrecht University, The Netherlands
Mirzaa, G., M.D., Seattle Children's Research Institute, Washington
Morelli, K., B.S., University of Maine/The Jackson Laboratory, Orono
Morrow, J., Ph.D., University of Michigan, Ann Arbor
Murta, V., Ph.D., Institute of Cellular Biology and Neurosciences, Buenos Aires, Argentina
Ramirez, V., Ph.D., Pontifical Catholic University of Chile, Santiago, Chile
Weiss, F., B.S., MRC/CSC, London, United Kingdom

SEMINARS

Darnell, R., The Rockefeller University/HHMI/NY Genome Center, New York: Applying advanced techniques in molecular neuroscience to the human brain.
Witkowski, J., Cold Spring Harbor Laboratory: Lecture on ethics.
Eberwine, J., University of Pennsylvania, Philadelphia: Multimodal single-cell analysis of live human neuronal cells and cellular subregions.
Rolls, M., Pennsylvania State University, State College: Cell biology of the nervous system.
Kenny, P., Icahn School of Medicine, Mount Sinai, New York: Pharmacology and systems therapeutic.
Allen, N., Salk Institute, La Jolla, California: Astrocyte regulation of synapse development and function.
Jaffrey, S., Weill Cornell Medical College, New York: Topic on pharmacology.
Harwell, C., Harvard Medical School, Boston, Massachusetts: Progenitor regulation of cortical neuron diversity and circuit assembly.
Silver, D., Duke University Medical Center, Durham, North Carolina: Molecular genetics and microbiology.
Gradinaru, V., California Institute of Technology, Pasadena: Clarity and optogenetics.
Casaccia, P., Icahn School of Medicine, Mount Sinai, New York: Topic on neuroscience.
Polleux, F., Columbia University, New York: Molecular approaches to study organelle-specific calcium dynamics in neurons.
Haas, K., University of British Columbia, Vancouver, Canada: Single-cell electroporation and imaging.
Maher, B., Lieber Institute, Johns Hopkins Medical Institute, Baltimore, Maryland: Methods for using neuronal transgenesis by in utero electroporation.
Cai, D., University of Michigan, Ann Arbor: New improved and brainbow technique module.
Luikart, B., Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire: Lentivirus module director.

Drosophila Neurobiology: Genes, Circuits, and Behavior

July 1–21

INSTRUCTORS **K. Kaun**, Brown University, Providence, Rhode Island
C.-H. Lee, National Institutes of Health, Bethesda, Maryland
S. Pulver, University of St. Andrews, Scotland

ASSISTANTS **J. Jonaitus**, University of St. Andrews, Scotland
N. Mei, Brown University, Providence, Rhode Island
P. Muthuirulan, NICHD, National Institutes of Health, Bethesda, Maryland
E. Petruccelli, University of Iowa, Iowa City
A. Spring, University of Iowa, Iowa City
C.-Y. Ting, NICHD, National Institutes of Health, Bethesda, Maryland

This laboratory/lecture course was intended for researchers at all levels from beginning graduate students through established primary investigators who wanted to use *Drosophila* as an experimental system for nervous system investigation. The 3-week course was designed to introduce students to a wide variety of topics and techniques, including the latest approaches for studying nervous system development, activity, and connectivity, as well as complex behaviors and disease models. Daily research seminars presented comprehensive overviews of specific subfields of nervous system development or function or focused on state-of-the-art techniques and approaches in *Drosophila* neuroscience. Expert guest lecturers discussed their findings and approaches and brought along their own assays and techniques for students to learn in the laboratory part of the course. The hands-on portion of the course was centered on inquiry-based projects, utilizing the different morphological and physiological measurements and behavioral paradigms learned through the course. This included molecular-genetic analyses, immunocytochemistry, recording of activity using electrophysiology and genetically encoded calcium indicators, optogenetic and

thermogenetic control of neural activity, as well as numerous quantitative behavioral measures. Collectively, the course provided a comprehensive and practical introduction to modern experimental methods for studying the neural basis of behavior in *Drosophila*.

This course was supported with funds provided by the National Institute on Drug Abuse, the National Science Foundation, the Helmsley Charitable Trust, and the Howard Hughes Medical Institute.

PARTICIPANTS

Currier, T., B.A., New York University, New York
Dombrovski, M., University of Virginia, Charlottesville
Eidhof, I., M.S., Radboudumc, Nijmegen, The Netherlands
Garaulet, D., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Ho, M., Ph.D., Johns Hopkins University School of Medicine, Baltimore, Maryland
Kang, Y., University of Houston-Downtown, Texas
Mezzera, C., Ph.D., Champalimaud Center for the Unknown, Lisbon, Portugal
Mihovilovic Skanata, M., Ph.D., New York University, New York
Mokashi, S., M.S., North Carolina State University, Raleigh
Palavicino-Maggio, C., Ph.D., Harvard Medical School, Boston, Massachusetts
Praschberger, R., M.D., B.A., University College London, United Kingdom
Woods, B., B.S., Harvard Medical School, Boston, Massachusetts

SEMINARS

Instructors: History and basics of *Drosophila* neurogenetics workshop. How to make a mutant and design a screen. Practical aspects of using neurogenetic tools and mosaic analysis and *Drosophila* as animal models to study drugs of abuse.
Wildonger, J., University of Wisconsin, Madison: Genome editing: Fly CRISPR development and use.
Harbison, S., NHLBI, National Institutes of Health, Bethesda, Maryland: Quantitative trait loci.
Auld, V., University of British Columbia, Vancouver; Serpe, E., NICHD, National Institutes of Health, Bethesda, Maryland: Development 101; Glia are great.
Serpe, E., NICHD. National Institutes of Health, Bethesda, Maryland: The synapse: development, anatomy and function.
Rolls, M., Penn State University, State College, Pennsylvania: Cell biology of the nervous system; Neuronal polarity.
Levitan, E., University of Pittsburgh, Pennsylvania: Microscopy, dense core vesicles, neurosecretion.
Frank, A., University of Iowa, Iowa City: Neuromuscular junction physiology I; Neuromuscular junction physiology II.
Fetcho, J., Cornell University, Ithaca, New York: Transparent vertebrates offer a direct view of the nervous system in action.
Zwart, M., HHMI Janelia Research Campus, Ashburn, Virginia: Central nervous system physiology I; Central nervous system physiology II.
Prinz, A., Emory University, Atlanta, Georgia, and Gunay, C., Georgia Gwinnett College, Lawrenceville: Computational modeling of *Drosophila* neurons and circuits.
Ruta, V., The Rockefeller University, New York, New York: Flexible and adaptive circuit processing in *Drosophila*.
Pulver, S., University of St. Andrews, Scotland: Optical imaging of central pattern generator output.
Heckscher, E., University of Chicago, Illinois: Control and assembly of motor circuits.
Louis, L., EMBL Center for Genomic Regulation, Spain: Larval sensory systems and behavior
Jayaraman, V., Janelia Research Campus, Ashburn, Virginia: Sensorimotor integration in the central complex.
McKellar, C., HHMI Janelia Research Campus, Ashburn, Virginia: Adult brain circuits and anatomy; Introduction to behaviors and genetic tools.
Keene, A., Florida Atlantic University, Jupiter: The genetic basis of sleep and feeding.
Lee, C.-H., NICHD, National Institutes of Health, Bethesda, Maryland: Visual circuit assembly and function; Correlating behavior with brain expression.
Brown, K.K., Brown University, Providence, Rhode Island: Using flies as a model to understand addiction.
Branson, K., Brown University, Providence, Rhode Island, and Robie, A., HHMI Janelia Research Campus, Ashburn, Virginia: In vivo-LFP/extracellular/probes.
Levine, J., University of Toronto, Mississauga, Canada: Social behavior and hierarchical interactions.
Dubnau, J., Cold Spring Harbor Laboratory, and Masek, P., SUNY Binghamton University, New York: Associative memory and conditioned taste aversion.

Frontiers and Techniques in Plant Science

July 1–21

INSTRUCTORS **S. Cutler**, University of California, Riverside
N. Provart, University of Toronto, Ontario, Canada
M. Timmermans, University of Tübingen, Germany

CO-INSTRUCTOR **U. Paszkowski**, University of Cambridge, United Kingdom

ASSISTANTS **D. Elzinga**, University of California, Riverside
L. Luginbuehl, John Innes Centre, Norwich, England
M. Reynoso, University of California, Riverside
J. Swift, New York University, New York

This course provided an intensive overview of topics in plant genetics, physiology, biochemistry, development, and evolution and hands-on experiences in molecular, analytical, computational, and high-throughput approaches to understanding plant biology. It emphasized recent results from model organisms including *Arabidopsis*, maize, and tomato, as well as a variety of other plants, and provided an introduction to current methods used in basic and applied plant biology, both theoretical and practical.

The seminar series included plant morphology and anatomy, development, evolution, light and circadian biology, hormones, small RNAs and epigenetic inheritance, biotic and abiotic interactions, plant biochemistry, crop domestication, and applications addressing current agronomic

problems. Speakers provided expert overviews of their fields, followed by in-depth discussions of their own work. The laboratory sessions provided exposure to cutting-edge experimental and computational techniques currently used in plant research. These included approaches for studying plant development, transient gene expression, applications of fluorescent proteins, genome editing, and chromatin immunoprecipitation. Students also gained hands-on experience with computational tools and environments for genome assembly, transcriptomics, construction of gene regulatory networks, identification of quantitative trait loci, mapping by sequencing, and mathematical modeling of development and hormone action. The course also included several short workshops on important themes in plant research. Throughout the course, students interacted individually and informally with the speakers to further enrich their learning experience.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Amorim, M., M.S., University of Tübingen (ZMBP), Germany
Brown, K., B.S., University of Kansas, Lawrence
Fischer, J., B.S., University of Kansas, Lawrence
Freund, D., Ph.D., University of Minnesota, Saint Paul
Kosentka, P., B.S., University of Tennessee, Knoxville
Lo Presti, L., Ph.D., Max-Planck-Institute for Terrestrial Microbiology, Marburg, Germany
Man, J., B.A., University of Massachusetts, Amherst, Hadley
Markham, K., Ph.D., University at Buffalo, New York
Mateluna, J., M.S., Eberhard Karls Universitat Tübingen, Germany
Nasti, R., B.S., University of Minnesota Twin Cities, St. Paul
Neto, C., M.S., Max-Planck-Institute for Plant Breeding Research, Cologne, Germany
Podolec, R., M.S., University of Geneva, Faculty of Science, Switzerland
Prywes, N., Ph.D., Harvard University, Cambridge, Massachusetts
Schulten, A., M.S., Ruhr University Bochum, Germany
Thomas, J., B.S., University of Minnesota Twin Cities, St. Paul
Wei Chee, W., Ph.D., Advanced Agriecological Research, Selangor, Malaysia

SEMINARS

Instructor Introductions

Student Presentation (Group 1)
Sinha, N., University of California, Davis: Lab 1: Introduction to plant biology. Lab 1: Analysis of plant anatomy. Talk: Leaves: Connecting form with function.

Student Presentations (Group 2)
Cutler, S., University of California, Riverside: Talk: Chemical and genetic dissection of plant hormone pathways.
Provart, N., University of Toronto, Canada, and Cutler, S., University of California, Riverside; Lab 2: Genetic mapping, mapping by sequence, bulk segregation analysis.
Timmermans, M., Cold Spring Harbor Laboratory: Lab 3: microRNA design lab. Talk: Leaf polarity.
Paszkowski, U., University of Cambridge, United Kingdom; Talk: Molecular genetics of root endosymbiosis.
Maloof, J., University of California, Davis: Lab 4: Introduction to R. Talk: QTL analysis: Introduction and examples. Lab 5: Identification of quantitative trait loci (QTL).
Harmer, S., University of California, Davis: Talk: Circadian clock. Lab 6: Affinity tagging.
Shiu, S.-H., Michigan State University, E. Lansing: Lab 7: Phylogenetic analysis. Talk: Genome evolution.
Law, J., and Palanca, A.M., Salk Institute, La Jolla, California: Introductory talk: Chromatin immunoprecipitation preceding Lab 8A. Lab 8A: Chromatin immunoprecipitation. Lab 8B: Chromatin immunoprecipitation.
Law, J., Salk Institute, La Jolla, California: Talk: Epigenetics and DNA methylation.
Girke, T., University of California, Riverside: Lab 9: RNA-seq analysis with R/bioconductor.
Gilroy, S., University of Wisconsin, Madison: Talk: Ca^{2+}, ROS, and stress: Biosensor approaches to visualize signaling systems in *Arabidopsis*.
Peck, S., University of Missouri, Columbia: Introductory talk to analysis of membrane proteins. Lab 10: Analysis of membrane proteins I. Talk: Non-self recognition between hosts and pathogens. Lab 11: Analysis of membrane proteins II.
Doebley, J., University of Wisconsin, Madison: Talk: Domestication of crops in general and of maize in particular.
Shukla, V., The Bill & Melinda Gates Foundation, Seattle, Washington: Talk: Translating basic science into agricultural development for smallholder farms.

Johnson, M., Brown University, Barrington, Rhode Island: Talk: Molecular dialogues between pollen and pistil. Lab 12: Confocal microscopy to determine GFP-tagged protein localization using pollen tubes as a model system.

Sage, R., University of Toronto, Canada: Talk: Photosynthesis: Limitations, evolution, and prospects for improvement.

Brady, S., University of California, Davis: Talk: Transcriptional regulation in *Arabidopsis thaliana*: The Interface. Lab 13, Part 1: Gene regulatory networks.

Jones, J., Sainsbury Laboratory, Norwich, United Kingdom: Talk: How plants succumb to or resist disease.

Bailey-Serres, J., University of California, Riverside: Lab 14: Julia Bailey-Serres. Talk: Abiotic stress solutions: From gene to field. Lab 15: Julia Bailey-Serres.

Baldwin, I., Max-Planck-Institute for Chemical Ecology, Jena, Germany: Talk: Defense is only one way of coping with herbivores.

Barbier de Reuille, P., University of Bern, Switzerland: Lab 16: Mathematical modeling of plant growth. Lab 16: Mathematical modeling of plant growth; MorphographX.

Kuhlemeier, C., University of Bern, Switzerland: Talk: Phyllotaxis: A quantitative developmental problem.

Provart, N., University of Toronto, Canada: Talk: Hypothesis generation with large data sets. Lab 17: Bioinformatic tools in plant research.

Voytas, D., University of Minnesota, Saint Paul: Lab 18: Genome engineering with the CRISPR-CAS system. Talk: Precise genome engineering with sequence-specific nucleases.

Dinneney, J., Carnegie Institution for Science, Stanford, California: Talk: Stressed! How roots cope through dynamics behaviors.

Metabolomics

July 8–21

INSTRUCTORS **A. Caudy**, University of Toronto, Ontario, Canada
E. Gottlieb, Cancer Research UK Beatson Institute, Glasgow, United Kingdom
A. Rosebrock, University of Toronto, Ontario, Canada

ASSISTANTS **K. Laverty**, University of Toronto, Ontario, Canada
G. MacKay, CRUK Beatson Institute, Glasgow, United Kingdom
N. Van Den Broek, Cancer Research UK Beatson Institute, Glasgow, United Kingdom
S. Zlitni, University of Toronto, Ontario, Canada

Metabolomics involves the systematic interrogation of the abundance of small chemical molecules (metabolites) within cells, tissues, organs, and organisms. In parallel with high-throughput technologies that facilitate genomic, transcriptomic, and proteomic analyses of cellular and organismal physiology, technologies for metabolite profiling represent an important source of information about the dynamic state of the cell or tissue that is relevant in both health and disease.

LC-MS metabolomics were the primary focus of the course and were applied for both targeted and untargeted analysis of endogenous metabolites and in vitro enzyme reactions. We used approaches for steady-state measurement of metabolite levels as well as assessment of metabolite flux. To complement these LC-MS analyses, there were experiments involving other methodologies. There were shorter-term activities with other methodologies, including GC-MS, polarimetric and Seahorse measurement of oxygen consumption, FRET sensors and/or MitoTracker measurements, enzymatic techniques for metabolite measurement, and uptake experiments.

The consistent and extended application of LC-MS reflected the emphasis of the course, and the exposure to other methodologies allowed students to appreciate the utility and complementarity of these methods.

Objectives for students:

- quantitative and qualitative analysis of LC-MS data using currently available tools (vendor software, Rosebrock tools, XCMS online, Agilent Profinder/Genespring)
- understanding of common interferences and limitations of LC-MS and GC-MS analysis
- recognizing key issues in experimental design and sample preparation for metabolomics
- awareness of major biochemical pathways active in commonly used cell types
- familiarity with methods for determining different types of oxygen consumption

Proposed lab exercises:

- Full-scan experiment on knockout/drug treatment. Students identify significantly changed metabolites and use MS/MS fragmentation and other methods for identification.
- Experimental treatments were selected in which the discovery of a phenotype was possible with only one of several analytic/separation methodologies.
- Identification of a significantly changed metabolite in blood/plasma/urine and development of a targeted method for its analysis by QQQ, including determination of LOD/LOQ/linearity and other appropriate method validation.
- Measurement of metabolite flux by pulse labeling (i.e., kinetic flux profiling).
- Enzyme assay to determine V_{max} and K_m (measure on MSD?).
- Enzymatic synthesis of a compound (e.g,. sedoheptulose bisphosphate or ribose-1-phosphate) and purification by mass-based or HPLC fractionation.
- Non LC-MS methodologies: Roche kits for lactate and glucose measurement in culture supernatant; seahorse experiment, including all drugs for uncoupled, etc.; FRET experiment (by cytometry and/or microscopy) for NADH levels and proton gradient sensitive Mitotracker staining; nutrient uptake by 14C (potentially amino acids?); glycolytic flux in mammalian cells by 3H.

Students received hands-on training on Agilent QTOF and Thermo Orbitrap, Q-exactive and Vantage triple quadrupole, and Waters SYNAPT G2-S mass spectrometers.

This course was supported with funds provided by the National Institute of General Medical Sciences, the Helmsley Charitable Trust, and the Howard Hughes Medical Institute.

PARTICIPANTS

Chi, H., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
Du, J., Ph.D., West Virginia University, Morgantown
Hubert, L., Ph.D., Baylor College of Medicine, Houston, Texas
Sofie Husted, A., M.S., University of Copenhagen, Denmark
Irnov, I., Ph.D., Yale University, West Haven, Connecticut
Kaminski, M., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
LaGory, E., Ph.D., Stanford University, California
Li, H., Ph.D., Indiana University, Bloomington
Linder, S., B.A., Massachusetts General Hospital, Boston
Petrova, B., M.S., Ph.D., Whitehead Institute for Biomedical Research, Cambridge, Massachusetts
Ringel, A., Ph.D., Harvard Medical School, Baltimore, Massachusetts
Snider, J., M.S., Stony Brook University, Setauket, New York
Tsang, S., M.D., Columbia University, New York
Wetzel, C., University of Cincinnati, Ohio
Ye, Y., Ph.D., Bayer CropScience, Durham, North Carolina
Yilmaz, L., Ph.D., University of Massachusetts Medical School, Worcester

SEMINARS

Caudy, A., University of Toronto, Canada: Lehninger and Stryer only tell part of the story: How metabolomics is rewriting biochemistry textbooks.

Rosebrock, A., University of Toronto, Canada: Direct biochemical phenotyping reveals unexpected metabolic responses to changes in cellular microenvironment.

Patti, G., The Scripps Research Institute, La Jolla, California: Reprogramming metabolism with two electrons.

Vander Heiden, M., MIT-Koch Institute/Broad Institute, Cambridge, Massachusetts: Determinants of metabolic dependencies in tumors.

Gottlieb, E., Cancer Research UK Beatson Institute, Glasgow, United Kingdom: Exploring and exploiting cancer's metabolic vulnerabilities.

Fan, T., University of Kentucky, Lexington: Exploring lung cancer metabolome: In vivo and ex vivo.

Cross, J., Memorial Sloan Kettering Cancer Center, New York, New York: Metabolomics strategies for microbiome research: Activation of innate immune cells by microbially derived ligands.

Metallo, C., University of California, San Diego, La Jolla: Tracing known and "unknown" metabolic pathways.

DeBerardinis, R., University of Texas Southwestern Children's Medical Center, Dallas: Understanding metabolic heterogeneity in cancer.

Gross, S., Weill Cornell Medical College, New York, New York: Untargeted metabolite profiling to discover mechanisms of drug actions and gene functions and screen for inborn errors of metabolism.

Computational Neuroscience: Vision

July 11–24

INSTRUCTORS **G. Boynton**, University of Washington, Seattle
G. Horwitz, University of Washington, Seattle
J. Pillow, Princeton University, New Jersey

ASSISTANTS **D. Popovkina**, University of Washington, Seattle
J. Yates, The University of Texas, Austin

Computational approaches to neuroscience will produce important advances in our understanding of neural processing. Prominent success will come in areas where strong inputs from neurobiological, behavioral, and computational investigation can interact. The theme of the course was to understand the computational problems, the constraints on solutions to these problems, and the range of possible solutions that can help guide research in neuroscience. Through a combination of lectures and hands-on experience with MATLAB-based computer tutorials and projects, this intensive course examined visual information processing from the retina to higher cortical areas, spatial pattern analysis, motion analysis, neuronal coding and decoding, attention, and decision-making.

This course was supported with funds provided by the Helmsley Charitable Trust and the Howard Hughes Medical Institute.

PARTICIPANTS

Bonnen, K., B.S., University of Texas, Austin
Calhoun, A., Ph.D., Princeton University, New Jersey
Christensen, A., B.S., Stanford University, California
De, A., B.Tech., University of Washington, Seattle
Duncker, L., M.S., Princeton University, New Jersey
Fan, Y., B.S., University of Pennsylvania, Philadelphia
Ghodrati, M., M.Eng., Monash University, Melbourne, VIC, Australia
Golden, J., Ph.D., Stanford University, California
Jozwik, K., Ph.D., University of Cambridge, United Kingdom
Kell, A., B.A., Massachusetts Institute of Technology, Cambridge
Kleene, N., B.A., Rutgers University, Somerset, New Jersey
Moreland, J., B.S., University of Washington, Seattle
Okazawa, G., Ph.D., New York University, New York
Park, W.J., M.A., University of Rochester, New York
Pottackal, J., B.A., Yale University, New Haven, Connecticut
Rajalingham, R., B.Eng., Massachusetts Institute of Technology, Cambridge
Ruda, K., B.A., Duke University, Durham, North Carolina
Schröder, S., Ph.D., University College London, United Kingdom
Semizer, Y., B.A., M.A., Rutgers University, Highland Park, New Jersey
Sha, L., B.A., New York University, New York
Stanciu, O., M.A., Central European University, Budapest, Hungary
Vo, V., Ph.D., University of California, San Diego, La Jolla
Walshe, R., Ph.D., The University of Texas, Austin

SEMINARS

Movshon, T., New York University, New York: What vision (and this course) is all about; Motion vision/MT.
Horwitz, G., University of Washington, Seattle: White noise analysis.
Simoncelli, E., New York University, New York: Image statistics, texture modeling, and vision.
Pillow, J., Princeton University, New Jersey: Statistical models for neural coding, GLMs.
Rieke, F., University of Washington, Seattle: Retina.
Chichilnisky, E.J., Stanford University, California: A Retina.
Palmer, S., University of Chicago, Illinois: Information theory, earl vision.
Gjorgjieva, J., Max-Planck-Institute for Brain Research, United Kingdom: Efficient coding.
Carandini, M., University of College London, United Kingdom, and Boynton, G., University of Washington, Seattle: fMRI, visual attention, visual prosthetics.
Brainard, D., University of Pennsylvania, Philadelphia: Color vision.
Rust, N., University of Pennsylvania, Philadelphia: v4 and IT.
DiCarlo, J., Massachusetts Institute of Technology, Cambridge: Object recognition.
Shlens, J., Google, San Francisco, California: Deep learning, convolutional neural networks, big data.
Fairhall, A., University of Washington, Seattle: Adaptive coding.
Read, J., Newcastle University, United Kingdom: Depth perception, stereopsis.
Huk, A., University of Texas, Austin: 3D motion perception.
Treue, S., German Primate Center, Germany: Physiology of attention.
Cohen, M., University of Pittsburgh, Pennsylvania: Attention and population coding.
Ma, W.J., New York University, New York: Psychophysics, modeling behavior.
Churchland, A., Cold Spring Harbor Laboratory: Multimodal processing.
Kiani, R., New York University, New York: Decision-making.

Genetics and Neurobiology of Language

July 25–31

INSTRUCTORS **S. Fisher**, Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
K. Watkins, Oxford University, United Kingdom

ASSISTANT **A. Tilot**, Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

Why are children able to acquire highly sophisticated language abilities without needing to be taught? What are the neurobiological and neurophysiological processes that underpin human speech and language, and how do they go awry in developmental and acquired disorders? Which genetic factors contribute to this remarkable suite of human skills, and are there evolutionary precursors that we can study in other species? This unique new CSHL course addressed core questions about the bases and origins of speech and language, through talks, interactive sessions, keynotes, and debates, involving leading experts from a range of disciplines. It integrated the state of the art from complementary perspectives, including development, cognitive models, neural basis, gene identification, functional genomics, model systems, and comparative/evolutionary studies.

This course was supported with funds provided by the Nancy Lurie Marks Family Foundation.

PARTICIPANTS

Argyropoulos, G., Ph.D., University College London Institute of Child Health, United Kingdom
Blanco-Elorrieta, E., B.A., New York University, New York
Bowling, D., Ph.D., University of Vienna, Austria
Roksan Caglar, D.R., M.A., Rutgers University, Newark, New Jersey
Doelling, K., B.A., New York University, New York
Fama, M., M.S., Georgetown University, Washington, D.C
Giustolisi, B., M.A., M.S., University of Milano-Bicocca, Milan, Italy
Gwilliams, L., M.A., New York University, New York
Lien, K., B.A., Arizona State University, Tempe
Lipkind, D., Ph.D., Hunter College, The City University of New York, New York
Martins, M., Ph.D., Max-Planck-Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
Mueller, K., Ph.D., The University of Iowa, Iowa City
Pinheiro, A., Ph.D., iMed-Faculty of Pharmacy, University of Lisbon, Braga, Portugal
Rios Lopez, P., M.S., BCBL, Donostia-San Sebastián, Spain
Theofanopoulou, K., M.A., Universitat de Barcelona, Spain
Truong, D., Ph.D., Yale School of Medicine, New Haven, Connecticut
Vinals, L., M.S., University of Cambridge, United Kingdom
Worster, E., B.A., University College London, United Kingdom
Zhukova, M., B.A., Saint-Petersburg State University, Russia

SEMINARS

Aslin, R., University of Rochester, New York: Speech perception and early word learning in infancy.
Werker, J., University of British Colombia, Vancouver, Canada: Experience and plasticity in speech perception.
Fisher, S., Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands: Translating the genome in human neuroscience.
Tomblin, J.B., University of Iowa, Iowa City: Individual differences in language development as phenotypes.
Newbury, D., University of Oxford, United Kingdom: Genetic mapping methods and applications to speech and language.
Emmorey, K., San Diego State University, California: The neurobiology of language from the perspective of sign languages.
MacSweeney, M., University College London, United Kingdom: The impact of deafness on the neurobiology of spoken language.
Scott, S., University of College London, United Kingdom: Neurobiology of links between speech perception and production.
Watkins, K., University of Oxford, United Kingdom: The neurobiology and genetics of stuttering.
Vargha-Khadem, F., University of College London, United Kingdom: Deep phenotyping of FOXP2 mutation cases.
Morgan, A., Murdoch Children's Research Institute, Melbourne, Victoria, Australia: Gene pathways to childhood motor speech disorders.
Pylkkanen, L., New York University, New York: Brain basis of semantic composition: What we know and what we don't.
Fedorenko, E., HMS/MGH:MIT, Belmont, Massachusetts: The language system and its place within the broader architecture of the human mind and brain.
Scharff Freie, C., Universitat Berlin, Germany: Animal models for language: What—if anything—can be "modeled."
Vernes, S., Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands: Understanding speech and language: From genes to bats and beyond.
Davis, M., Medical Research Council, Cambridge, United Kingdom: Predicting and perceiving degraded speech.
Lau, E., University of Maryland, College Park: Multimodal neuroimaging and predictive processing models.
Tchernichovski, O., Hunter College, CUNY, New York: Development of combinatorial ability and vocal coordination in songbirds.
Fitch, T., University of Vienna, Austria: A comparative approach to the neurobiology of language.
Slocombe, K., University of York, United Kingdom: Primate communication: Links to human language?

Chromatin, Epigenetics and Gene Expression

July 26–August 15

INSTRUCTORS **K. Adelman**, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina
L. Di Croce, ICREA and Centre for Genomic Regulation, Barcelona, Spain
G. Narlikar, University of California, San Francisco
A. Shilatifard, Northwestern University Feinberg School of Medicine, Chicago, Illinois

PART-TIME INSTRUCTOR **D. Taatjes**, University of Colorado, Boulder

ASSISTANTS **N. Gamarra**, University of California, San Francisco
T. Henriques, NIEHS, National Institutes of Health, Research Triangle Park, North Carolina
G. Mas Martin, Centre for Genomic Regulation, Barcelona, Spain
G. Shah, Nano Temper Technologies, San Francisco, California
C. Fant, University of Colorado, Boulder
C. Sze, Northwestern University, Chicago, Illinois

This course was designed for students, postdocs, and principal investigators who have recently ventured into the exciting area of gene regulation. Emphasis was placed on exposing students to a broad array of methodologies to study gene regulation and chromatin structure and dynamics, including both state-of-the-art and well-developed methods. Students performed widely used techniques such as quantitative RT-PCR, reporter assays of enhancer activity, and chromatin immunoprecipitation (ChIP) coupled with sequencing (ChIP-Seq). They generated high-throughput sequencing data to localize Pol II genome-wide and applied a basic informatics pipeline to analyze the results. Students

also isolated transcription factor complexes and assessed their activity in functional assays. They also used RNAi to knock down specific factors and evaluate the effects on gene expression.

This course provided the basic concepts for different methods to analyze the chromatin architecture of the genome. Students performed chromosome conformation capture (3C) experiments, together with other approaches aimed to interrogate the 3D organization of genomes. Moreover, the computational methods required to analyze these data were discussed.

Students learned how to assemble recombinant chromatin and use biophysical methods such as FRET to assay the activity of chromatin remodeling enzymes. They also learned principles of enzyme kinetics and applied these to quantify the remodeling reactions.

Given the broad biological roles for DNA-binding transcription factors and emerging roles of noncoding RNAs in transcription regulation, electrophoretic mobility shift assays (EMSAs) are again becoming widely used for assessing transcription factor binding to regulatory DNA or RNA elements. Students learned how to perform and interpret EMSA experiments using both microscale thermophoresis and gel-based methods. Students also performed in vitro transcription reactions that allowed them to determine whether specific factors (e.g., protein or ncRNA) directly affect transcription.

Experience with basic recombinant DNA techniques was a prerequisite for admission to this course. Lectures by the instructors covered the current state of the gene expression field, theoretical aspects of the methodology, and broader issues regarding strategies for investigating the regulation of gene expression in eukaryotes. Emphasis was placed on advantages and limitations of specific techniques, and data interpretation. The students were encouraged and expected to actively participate in these discussions.

Guest lecturers were experts in the field who discussed contemporary paradigms in eukaryotic gene regulation and technical approaches to interrogate new systems. From the guest lectures and discussions, students learned to design effective experiments, properly interpret their own data, and critically evaluate the gene expression literature.

This course was supported with funds provided by the National Cancer Institute.

PARTICIPANTS

Agarwal, P., M.S., University of Texas, Austin
Bardales, J., Ph.D., University of California, Berkeley
Bates, D., B.S., Brigham Young University, Riverton, Utah
Figueroa, C., M.S., Maine Medical Research Institute, Scarborough
Gañez Zapater, A., M.S., Stockholm University, Sweden
Hellmuth, J., M.D., Weill Cornell Medical College, New York, New York
Jordan III, W., B.S., Brown University, Providence, Rhode Island
Kumar, J., Ph.D., University of North Dakota, Grand Forks
Maharjan, M., Louisiana State University, Baton Rouge
Merrell, A., Ph.D., University of Pennsylvania, Philadelphia
Panday, A., Ph.D., Louisiana State University, Baton Rouge
Rodriguez, A., B.S., University of California, Los Angeles
Tanigawa, Y., B.S., The University of Tokyo, Kashiwa, Chiba, Japan
Wong, M., B.S., Cold Spring Harbor Laboratory
Zheng, C., B.S., M.S., University of Copenhagen, Denmark
Zhu, J., B.S., Indiana University, Bloomington

SEMINARS

Goodrich, J., University of Colorado, Boulder: RNA polymerase II transcription: General machinery, mechanism, and regulation.
Buratowski, S., Harvard Medical School, Boston, Massachusetts: Coupling transcription and chromatin and mRNA processing.
Lis, J., Cornell University, Ithaca, New York: Dissecting mechanism of transcription and its regulation in vivo using genome-wide strategies.
Neugebauer, K., Yale University, New Haven, Connecticut, and Stark, A., Research Institute of Molecular Pathology, Vienna, Austria: Decoding transcriptional reregulation in *Drosophila*; Decoding transcriptional reregulation in *Drosophila*.
Attardi, L., Stanford University, California: Deconstructing p53 transcriptional networks in tumor suppression.

Kingston, R., Massachusetts General Hospital/Harvard Medical School, Boston: The nucleosome as a direct regulator of gene expression and development.

Berger, S., University of Pennsylvania, Philadelphia: Epigenetic regulation in brain controlling behavior in ants and mammals.

Kadonaga, J., University of California, San Diego, La Jolla: Operating systems, apps, and a novel chromatin structure for the regulation of our genes.

Rao, A., La Jolla Institute for Allergy & Immunology, La Jolla, California: TET methyleytosine oxidases, immune responses, and cancer.

Heard, E., Curie Institute, Paris, France: Exploring facultative heterochromatin structure and function in the context of the inactive X chromosome.

Levine, M., Princeton University, New Jersey: Properties of transcriptional enhancers in development and evolution.

Kouzarides, T., University of Cambridge, United Kingdom, Lazar, M., University of Pennsylvania, Philadelphia, and Sheikhattar, R., University of Miami, Miami Beach, Florida: Enhancer RNAs and integrator complex in chromosome architecture.

Rinn, J., Harvard University, Cambridge, Massachusetts: Linking the noncoding genome to biology: A how-to guide.

Imaging Structure and Function in the Nervous System

July 26–August 15

INSTRUCTORS **F. Albeanu**, Cold Spring Harbor Laboratory
M. Orger, Champalimaud Foundation, Lisbon, Portugal
L. Palmer, University of Melbourne, Victoria, Australia

PART-TIME INSTRUCTOR **P. Tsai**, University of California, San Diego, La Jolla

ASSISTANTS **U. Lucas Boehm**, Paris School of Neuroscience (ENP), Paris, France
O. Braubach, Korea Institute of Science and Technology, Seoul, Seongbuk-gu, Republic of Korea
P. Garcia da Silva, Champalimaud Foundation, Lisbon, Portugal
M. Greaney, New York University Langone Medical Center, New York
F. Marbach, Cold Spring Harbor Laboratory
S. Renninger, Champalimaud Foundation, Lisbon, Portugal
N. Takahashi, Humboldt University, Berlin, Germany

Advances in light microscopy and digital image processing and the development of a variety of powerful fluorescent probes present expanding opportunities for investigating the nervous system, from synaptic spines to networks in the brain. This intensive laboratory and lecture course provided participants with the theoretical and practical knowledge to utilize emerging imaging technologies. The primary emphasis of the course was on vital light microscopy. Students learned the principles of light microscopy, as well as use of different types of cameras, laser-scanning systems, functional fluorophores, delivery techniques, and digital image-processing software. In addition to transmitted light microscopy for viewing cellular structure, the course examined a variety of molecular probes of cell function, including calcium-sensitive dyes, optogenetic and photoactivatable molecules. Particular weight was given to multiphoton laser-scanning microscopy and to biological fluorophores, especially green fluorescent protein (GFP) and its variants. We used a spectrum of neural and cell biological systems, including living animals, brain slices, and cultured cells.

This course was supported with funds provided by the Helmsley Charitable Trust and the Howard Hughes Medical Institute.

PARTICIPANTS

Brouwers, E., M.S ., VU University/CNCR, Amsterdam, The Netherlands
Cone, J., Ph.D., University of Chicago, Illinois
Diamanti, E., M.S., University College London, United Kingdom
Galliano, E., Ph.D., Harvard University, Cambridge, Massachusetts
Gomez, A., Ph.D., University of Basel, Switzerland
Guo, Y., B.S., Johns Hopkins University, Baltimore, Maryland
Hargil, H., B.S., Tel Aviv University, Israel
Marques, H., Ph.D., Champalimaud Foundation, Lisbon, Portugal
Pereda, A., M.D., Ph.D., Albert Einstein College of Medicine, Bronx, New York
Pierce, G., B.A., Columbia University, New York, New York
Shin, H., B.E., Brown University, Providence, Rhode Island
Taxidis, J., Ph.D., University of California, Los Angeles
Wood, K., Ph.D., University of Pennsylvania, Philadelphia
Yang, T., Ph.D., Cold Spring Harbor Laboratory

SEMINARS

Abrahamsson, S., The Rockefeller University, New York, New York, and Tsai, P., University of California, San Diego, La Jolla: Nature of light, coherence, wave description, polarization; Ray tracing, lenses; Huygens principle, law of refraction, diffraction, Abbe resolution limit; Kohler illumination, aperture, and field control; Incoherent emission pt source, PSF/Airy pattern, Fourier/OTF, aberrations; Contrast: phase, DIC, De Senarmont DIC, Dodt tube; Fluorescence/Fred's rocks, Jablonski diagrams, filters.
Tsai, P., University of California, San Diego, La Jolla: Introduction to optics bench lab.
Lichtman, J., Harvard University, Cambridge, Massachusetts: Part I: Principles and practice of confocal microscopy. Part II: Applications of confocal and other imaging approaches to connectomics.
Ma, J., Thorlabs, Newton, New Jersey: Resonant scanning systems.
Waters, J., Harvard Medical School, Boston, Massachusetts: Noise and detectors.
Judkewitz, B., Charité and Humboldt University, Berlin, Germany: 2P microscopy; lasers; Applications of 2P microscopy.
Emiliani, V., University Paris Descartes, Paris, France: Holographic photoactivation.
Wyart, C., Institut du Cerveau et de la Moelle épinière, Paris, France: In vivo optogenetics and circuit analysis in zebrafish spinal cord.
Orger, M., Champalimaud Foundation, Lisbon, Portugal: Introduction to light sheet microscopy.
Hillman, E., Columbia University, New York, New York: High-speed volume imaging with SCAPE microscopy.
Lin, M., Stanford University, California: Genetically encoded activity indicators 1; Genetically encoded activity indicators 2.
Oertner, T., University of Hamburg, Germany: Introduction to optogenetics.
Palmer, L., University of Melbourne, Victoria, Australia, and Waters, J., Allen Institute for Brain Science, Seattle, Washington: Calcium imaging/in vivo imaging in rodents.
Waters, J., Allen Institute for Brain Science, Seattle, Washington: Demo on analyzing calcium imaging data.
Paninski, L., Columbia University, New York, New York: Analysis of calcium imaging data and spike inference.
Albeanu, F., Cold Spring Harbor Laboratory: Intrinsic imaging.
Xu, C., Cornell University, Ithaca, New York: 3P imaging.
Smith, S., University of North Carolina, Chapel Hill: Two-photon imaging across large fields of view for systems neuroscience.
Holy, T., Washington University in St. Louis, Missouri: Light sheet microscopy.
Peterka, D., HHMI/Columbia University, New York: Applications for SLM.
Engert, F., Harvard University, Cambridge, Massachusetts: From maps to knowledge: A case against big data; Intro to building a homebrew 2P microscope including laser safety.
Deisseroth, K., Stanford University, California: Optogenetics: Recent advances/CLARITY.
Yasuda, R., Max-Planck Florida Institute for Neuroscience, Jupiter: FRET and FLIM.
Looger, L., Howard Hughes Medical Institute, Ashburn, Virginia: Latest developments in molecular tools.
Ji, N., Howard Hughes Medical Institute, Ashburn, Virginia: Superresolution imaging-structured illumination, PALM, STED; Deep imaging/adaptive hippocampus.
Dombeck, D., Northwestern University, Evanston, Illinois: Awake imaging in hippocampus.
Dieudonné, S., CNRS–Laboratoire de Neurobiologie, Paris, France: AODs and random access multiphoton imaging.
Denk, W., Max-Planck-Institute of Neurobiology, Martinsried, Germany: Block-face EM/connectome.

Yeast Genetics and Genomics

July 26–August 15

INSTRUCTORS **G. Brown**, University of Toronto, Ontario, Canada
M. Dunham, University of Washington, Seattle
M. Gartenberg, Robert Wood Johnson Medical School, Piscataway, New Jersey

ASSISTANTS **B. Taylor**, University of Washington, Seattle
E. Okeke, Robert Wood Johnson Medical School, Piscataway, New Jersey
B. Ho, University of Toronto, Ontario, Canada

This course was a modern, state-of-the-art laboratory course designed to teach students the full repertoire of genetic approaches needed to dissect complex problems in the yeast *Saccharomyces cerevisiae*. Combinations of classical and modern genetic approaches were emphasized, including the isolation and characterization of mutants, tetrad analysis, complementation, and mitotic recombination. Synthetic biology was explored through CRISPR-Cas9-directed engineering of heterologous biosynthetic pathways in yeast. Students learned array-based methods, next-generation sequencing, and genome-based methods of analysis facilitated by the yeast genome sequence, the deletion collection, and other genomic resources available to the community. Molecular genetic techniques, including yeast transformation, gene replacement by PCR, construction and analysis of gene fusions, and generation of mutations were also emphasized.

Students used classical approaches and modern whole-genome sequencing to gain experience in identifying and interpreting various kinds of genetic interactions, including suppression and

synthetic lethality (including SGA). Students were immersed in yeast genomics and performed and interpreted experiments using DNA arrays, whole-genome sequencing, and multiplexed DNA barcode sequencing. Students gained first-hand experience in modern cytological approaches such as epitope tagging and imaging yeast cells using indirect immunofluorescence, GFP-protein fusions, and a variety of fluorescent indicators for different subcellular organelles. Lectures on fundamental aspects of yeast genetics were presented along with seminars given by prominent experts in the field on topics of current interest.

This course was supported with funds provided by the National Science Foundation.

PARTICIPANTS

Celaj, A., Ph.D., University of Toronto, Ontario, Canada
Clark-Cotton, M.R., B.S., Duke University, Durham, North Carolina
Coughlan, A., B.S., University College Dublin, Ireland
DeLoach, D., M.S., University of Bath, United Kingdom
Dondi, A., M.S., European Institute of Oncology, Milano, Italy
Gohil, V., Ph.D., Texas A&M University, College Station
Laranjo, L., B.S., Brandeis University, Waltham, Massachusetts
Leon, D., B.S., University of Texas, Austin
Liu, Y., Ph.D., University of Texas Southwestern Medical Center, Dallas
Lukacisin, M., M.Bioch., Institute of Science and Technology, Klosterneuburg, Austria
Orner, E., B.S., Stony Brook University, Stony Brook, New York
Rieckh, G., Ph.D., Institute of Science and Technology, Klosterneuburg, Austria
Sharmeen, N., B.S., McGill University, Montréal, Quebec, Canada
Singh, A., M.Tech., Ludwig Maximilians University, Munich, Germany
Stieglitz, J., B.S., Tufts University, Medford, Massachusetts
Wehrs, M., M.S., LBNL/JBEI, Emeryville, California

SEMINARS

Schekman, R., University of California, Berkeley: Sorting and secretion of miRNAs in extracellular vesicles.
Nash, R., Stanford University, Palo Alto, California: Navigating data using SGD and YeastMine.
Boone, C., University of Toronto, Canada: Systematic genetics with yeast.
Tsukiyama, T., Fred Hutchinson Cancer Research Center, Seattle, Washington: Molecular basis for quiescent state (G_0).
Rosenzweig, F., Georgia Institute of Technology, Atlanta: Sweet are the uses of adversity: Insights into adaption and speciation using experimental evolution.
Fox, C., University of Wisconsin, Madison: Finding a good place to start: Analyzing chromosomal origins in budding yeast.
Hochwagen, A., New York University, New York: The surprisingly complex inner life of repetitive DNA arrays.
Cohen-Fix, O., National Institutes of Health, Bethesda, Maryland: Shaping and misshaping the yeast nucleus: Does size matter?
Hittinger, C., University of Wisconsin, Madison: Yeast biodiversity and genome evolution: The *Saccharomyces* genus and its subphylum.
Springer, M., Harvard Medical School, Boston, Massachusetts: Sensing sugar ratio.
McClean, M., University of Wisconsin, Madison: Elucidating principles of biology signal processing using microfluidic and optogenetic tools.
Marcotte, E., University of Texas, Austin: *Saccharomyces sapiens*.
Unal, E., University of California, Berkeley: Gametogenesis in budding yeast: Methods to uncover unique modes of chromosome segregation and organelle remodeling.
Hieter, P., University of British Columbia, Vancouver, Canada: Chromosome instability and synthetic lethality in yeast and cancer.
Boeke, J., New York University Langone Medical Center, New York: Synthesizing and restructuring genomes.

Cellular Biology of Addiction

July 31–August 7

INSTRUCTORS **D. Belin**, University of Cambridge, United Kingdom
C. Evans, University of California, Los Angeles
B. Kieffer, Douglas Research Centre at McGill University, Montreal, Quebec, Canada

Drug addiction is the most costly neuropsychiatric disorder faced by our nation. Acute and repeated exposure to drugs produces neuroadaptation and long-term memory of the experience, but the cellular and molecular processes involved are only partially understood. The primary objective of this workshop was to provide an intense dialogue about the fundamentals, state-of-the-art advances, and major gaps in the cell and molecular biology of drug addiction. Targeted to new or experienced investigators, the workshop combined formal presentations and informal discussions to convey the merits and excitement of cellular and molecular approaches to drug addiction research. With the advent of genomics and proteomics, an extraordinary opportunity now exists to develop comprehensive models of neuroadaptive processes fundamental to addiction, withdrawal, craving, and relapse to drug use and to brain function, in general. A range of disciplines and topics were represented, including noninvasive brain imaging to identify drug targets and adaptive processes; neuroadaptive processes at the molecular and cellular level, neural networks and their modulation, the relevance of genotype to susceptibility and drug response; tolerance and adaptation at the cellular level and approaches to exploiting the daunting volume of data generated by neuroinformatics. This workshop provided an integrated view of current and novel research on neuroadaptive responses to addiction, fostered discussion on collaboration and integration, provided critical information needed to construct a model of addiction as a disease, and novel molecular targets for biological treatments. Beyond the plane of scientific endeavor, the information is vital for formulating public policy and for enlightening the public on the neurobiological consequences of drug use and addiction. The workshop was designed to generate interest in this level of analysis, open conduits for collaborations, and present novel routes to investigating the neurobiology of addictive drugs.

This course was supported with funds provided by the National Institute on Drug Abuse.

PARTICIPANTS

Anastasio, N., M.S., Ph.D., University of Texas Medical Branch, Galveston
Bbosa, G., Ph.D., Makerere University College of Health Sciences, Kampala, Uganda
Carlyle, B., Ph.D., Yale University, New Haven, Connecticut
Dios, A., Ph.D., Massachusetts General Hospital/Harvard Medical School, Charlestown
Exton-McGuinness, M., Ph.D. University of Birmingham, United Kingdom
Faulkner, M., B.A., University of North Carolina, Chapel Hill
Gamble-George, J., Ph.D., University of Florida, Gainesville
Goldberg, L., B.A., Boston University School of Medicine, Massachusetts
Gross, K., B.A., University of Minnesota, Minneapolis
Hilderbrand, E., B.A., University of Illinois, Chicago
Kennedy, P., Ph.D., University of California, Los Angeles
Kober, H., Ph.D., Yale University, New Haven, Connecticut
Kuhn, B., M.S., University of Michigan, Ann Arbor
Lezak, K., M.A., Ph.D., Harvard University/McLean Hospital, Belmont, Massachusetts
Lopez, A., B.S., University of California, Irvine
Paris, J., Ph.D., Virginia Commonwealth University, Richmond
Rouzer, S., B.S., Binghamton University, New York
Savell, K., B.S., University of Alabama, Birmingham
Sepulveda-Orengo, M., Ph.D., University of North Carolina, Chapel Hill
Serrano, P., Ph.D., Hunter College, New York
Shah, A., Ph.D., Johns Hopkins University, Baltimore, Maryland
Sidhu, H., Ph.D., The Scripps Research Institute, La Jolla, California
Sithisarn, T., Ph.D., University of Kentucky, Lexington
Slosky, L., Ph.D., Duke University Medical Center, Durham, North Carolina

SEMINARS

Maldonado, R., Universitat Pompeu Fabra, Barcelona, Spain, and Kieffer, B., Douglas Hospital Research Centre, McGill University, Montreal, Canada: Opiates.
Goldman, D., Laboratory of Neurogenetics, National Institute of Alcohol Abuse and Alcoholism, Rockville, Maryland: Genetics.
Everitt, B., University of Cambridge, United Kingdom: Memory and addiction.
Belin, D., University of Cambridge, United Kingdom: From habits to compulsivity: Psychological and neural substrates.
Ungless, M., Imperial College London, United Kingdom: Dopamine neuron function.
Koob, G., The Scripps Research Institute, La Jolla, California: Hedonic allostasis/stress.
Mason, B., The Scripps Research Institute, La Jolla, California: Human laboratory studies and medications development.
Ersche, K., University of Cambridge, United Kingdom: Human/imaging/vulnerability.
Dalley, J., University of Cambridge, United Kingdom: Rodents imaging.
Spanagel, R., Central Institute of Mental Health, Germany: Alcohol.
Boutrel, B., Lausanne University Hospital, Switzerland: Lactate release from astrocytes to neurons and maladaptive plasticity underlying strengthened cocaine-related memories.
Picciotto, M., Yale University, New Haven, Connecticut: Molecular and cellular mechanisms underlying nicotine addiction.
Evans, C., University of California, Los Angeles, and Kenny, P., Icahn School of Medicine at Mount Sinai, New York: Food/nicotine.
Schumann, G., King's College London, United Kingdom: IMAGEN.

Brain Tumors

August 2–8

INSTRUCTORS **C. Gladson**, Cleveland Clinic Lerner Research Institute, Ohio
A. Lasorella, Columbia University, New York, New York
S. Majumder, University of Texas MD Anderson Cancer Center, Houston

ASSISTANT **J. Zhang**, Columbia University Medical Center, New York, New York

This 1-week discussion course provided a clinical overview of brain tumors and emphasized molecular mechanisms involved in the growth and development of brain tumors with special emphasis on neural differentiation, signaling mechanisms, DNA replication, chromatin modulation, stem cells, mouse models, genomics, imaging techniques, genetically modified mouse techniques, nanotechnology, mechanism-based therapeutic strategies, and biobanks and ethical concerns. Attendees were able to interact with senior investigators on a one-on-one basis in an informal environment. All nonfaculty students were considered for a generous stipend depending on stated need. Applications were invited from medical and graduate students, postdoctoral fellows, faculty, and clinicians and scientists.

Significant scholarship funds for individuals in need of financial support are available courtesy of the American Brain Tumor Association.

PARTICIPANTS

Banisadr, A., B.S., University of California, San Diego
Baskaran, S., M.S., Uppsala University, Sweden
Burns, T., M.D., Ph.D., Mayo Clinic, Rochester, California
Datta, M., B.S., Massachusetts General Hospital, Boston

Frijlink, E., M.S., University Medical Center Utrecht, The Netherlands
Ghosh, D., Cold Spring Harbor Laboratory
Hansberg, V., M.S., Universidad Nacional Autonoma de Mexico, Mexico City
Jameson, N., B.S., University of California, San Diego, La Jolla
Johansson, E., M.S., Lund University, Sweden
McAdam, R., B.S., University of Toronto, Canada
Michaelsen, S., M.S., Rigshospitalet, Copenhagen, Denmark
Muniz-Talavera, H., Ph.D., St. Jude Research Children's Hospital, Memphis, Tennessee
Oudenaarden, C., M.S., Lund University, Sweden
Parisian, A., B.S., University of California, San Diego, La Jolla
Qemo, I., B.S., University of Windsor, Ontario, Canada
Roos, A., Ph.D., Translational Genomics Institute, Phoenix, Arizona
Staberg, M., Ph.D., Copenhagen University Hospital, Rigshospitalet, Denmark
Susanto, E., Ph.D., Karolinska Institutet, Stockholm, Sweden
Tanjore Ramanathan, J., M.S., University of Helsinki, Finland
Vigneswaran, K., M.D., Emory University School of Medicine, Atlanta, Georgia
Yabut, O., Ph.D., University of California, San Francisco
Yu, H., Ph.D., The Hospital for Sick Children, Toronto, Canada
Zhou, Y., M.D., Ph.D., College of Life Sciences, Wuhan, China

SEMINARS

Sawaya, R., MD Anderson Cancer Center, Houston, Texas: Clinical aspects of brain tumors.
Fuller, G., MD Anderson Cancer Center, Houston, Texas: Clinical aspects of brain tumors.
Bondy, M., Baylor College of Medicine, Houston, Texas: Clinical aspects of brain tumors.
Reilly, K., National Cancer Institute, Bethesda, Maryland: Clinical aspects of brain tumors.
Mietz, J., National Cancer Institute, Rockville, Maryland: Clinical aspects of brain tumors.
Wechsler-Reya, R., Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California: Pediatric brain tumors, chromatin.
Weiss, W., University of California, San Francisco: Pediatric brain tumors, chromatin.
Roussel, M., St. Jude Children's Research Hospital, Memphis, Tennessee: Pediatric brain tumors, chromatin.
Baker, S., St. Jude Children's Research Hospital, Memphis, Tennessee: Pediatric brain tumors, chromatin.
Furnari, F., University of California, San Diego, La Jolla: Adult brain tumors: Mechanisms, metabolism, imaging, viro- and immuno-therapy.
Bachoo, R., University of Texas Southwestern Medical Center, Dallas: Adult brain tumors: Mechanisms, metabolism, imaging, viro- and immuno-therapy.
Pieper, R., University of California, San Francisco, Petaluma: Adult brain tumors: Mechanisms, metabolism, imaging, viro- and immuno-therapy.
Chiocca, E.A., Brigham and Women's Hospital, Boston, Massachusetts: Adult brain tumors: Mechanisms, Metabolism, imaging, viro- and immuno-therapy.
Reardon, D., Dana Farber Cancer Institute, Boston, Massachusetts: Adult brain tumors: Mechanisms, Metabolism, imaging, viro- and immuno-therapy.
Rich, J., Cleveland Clinic, Ohio: Stem cells, single-cell analysis, modeling, brain metastases.
Lasorella, A., Columbia University, New York, New York: Stem cells, single-cell analysis, modeling, brain metastases.
James, C.D., Northwestern University, Chicago, Illinois: Stem cells, single-cell analysis, modeling, brain metastases.
Brastianos, P., Massachusetts General Hospital, Boston: Stem cells, single-cell analysis, modeling, brain metastases.
Dirks, P., University of Toronto, Ontario, Canada: Stem cells, single-cell analysis, modeling, brain metastases.
Gutmann, D., Washington University School of Medicine in St. Louis, Missouri: Normal development and disease, angiogenesis, and cell death: Big data.
Majumder, S., University of Texas MD Anderson Cancer Center, Houston: Normal development and disease, angiogenesis, and cell death: Big data.
Snyder, E., Sanford Burnham Prebys Medical Institute, La Jolla, California: Normal development and disease, angiogenesis, and cell death: Big data.
Gladson, C., Cleveland Clinic, Ohio: Normal development and disease, angiogenesis, and cell death: Big data.
Holland, E., Fred Hutchinson Cancer Research Center, Seattle, Washington: Normal development and disease, angiogenesis and cell death: Big data.
Verhaak, R., MD Anderson Cancer Center, Houston, Texas: Big data and clinical trials.
Brennan, C., Memorial Sloan Kettering Cancer Center, New York: Big data and clinical trials.
Iavarone, A., Columbia University Medical Center, New York, New York: Big data and clinical trials.
Gilbert, M., National Institutes of Health, Bethesda, Maryland: Big data and clinical trials.

Synthetic Biology

August 2–15

INSTRUCTORS
C. Beisel, North Carolina State University, Raleigh
M. Dunlop, University of Vermont, Burlington
A. Khalil, Boston University/Wyss Institute at Harvard, Boston, Massachusetts
V. Noireaux, University of Minnesota, Minneapolis
M. Smanski, University of Minnesota, Falcon Heights
H. Wang, Columbia University, New York, New York

ASSISTANTS
S. Hsu, University of Minnesota, Falcon Heights
M. Luo, North Carolina State University, Cary
R. Marshall, University of Minnesota, Minneapolis
J. Park, Columbia University, New York, New York
N. Rossi, University of Vermont, Winooski
B. Wong, Boston University, Massachusetts

Synthetic biology is a discipline wherein living organisms are genetically programmed to carry out desired functions in a reliable manner. This field takes inspiration from our ever-expanding ability to measure and manipulate biological systems, and the philosophical reflections of Schrödinger and Feynman that physical laws can be used to describe and rationally engineer biology to accomplish useful goals. After all, cells are the world's most sophisticated chemists, and their ability to learn to adapt to changing environments offers enormous potential for solving modern engineering challenges. Nonetheless, biological systems are noisy, massively interconnected, and nonlinear, and they have not evolved to be easily engineered. The grand challenge of synthetic biology is

to reconcile the desire for a predictable, formalized biological design process with the inherent 'squishiness' of biology.

To learn techniques and perform research at the forefront of synthetic biology, this course focused on how the complexity of biological systems, combined with traditional engineering approaches, results in the emergence of new design principles for synthetic biology. The Course was centered around an immersive laboratory experience. Here, students worked in teams to learn the practical and theoretical underpinnings of cutting-edge research in the area of synthetic biology. Broadly, we explored how cellular regulation—transcriptional, translational, posttranslational, and epigenetic—can be used to engineer cells to accomplish well-defined goals. Specific laboratory modules covered the following areas: computational biology using ordinary differential equations models, gene circuit characterization using microfluidics, cell-free transcription and translation systems, engineering RNA molecules as biosensors, and high-throughput cloning techniques and genome engineering. Students first learned essential synthetic biology techniques in a 4-day "boot-camp," and then rotated through research projects in select areas.

In addition, students interacted closely with a panel of internationally recognized speakers who gave students a broad overview of applications for synthetic biology, including renewable chemical production and therapeutics, the current state-of-the-art techniques, and case studies in human practices and socially responsible innovation.

This course was supported with funds provided by the National Institute of General Medical Sciences, the Howard Hughes Medical Institute, the Helmsley Charitable Trust, and the National Science Foundation.

PARTICIPANTS

Bali, A., M.S., Biosyntia, Hoersholm, Denmark
D'Amore, R., Ph.D., GeneMill University of Liverpool, United Kingdom
Dixon, E., B.S., University of Maryland School of Medicine, Baltimore
Espada, R., M.S., CONICET/University of Buenos Aires, Argentina
Greiss, F., M.S., QBM-University of Munich Gene Center, Germany
Huenemann, J., M.BA., Dow AgroSciences, Indianapolis, Indiana
Izri, Z., Ph.D., University of Kyushu, Fukuoka, Japan
Kulkarni, V., Ph.D., University of Warwick, Coventry, United Kingdom
Ledesma-Amaro, R., M.S., Ph.D., INRA, Thiverval-Grignon, France
Li, X., Ph.D., Ohio University, Athens
Mohler, K., B.S., The Ohio State University, Columbus
Moseley, R., B.S., University of Tennessee/ORNL, Knoxville
Mukhitov, N., B.S., Florida State University, Tallahassee
Stoner, R., Synthego, Redwood City, California
Tiwari, R., Ph.D., Genspace/Columbia/Pepisco, Hawthorne, New York
Wang, S., Ph.D., University of Michigan, Ann Arbor

SEMINARS

DeLisa, M., Cornell University, Ithaca, New York: Bacterial glycoengineering: From designer enzymes and pathways to next-generation therapeutics and vaccines.
Densmore, D., Boston University, Massachusetts: Bio-design automation in synthetic biology.
Haynes, K., Arizona State University, Tempe: Synthetic chromatin in human cells.
You, L., Duke University, Morrisville, North Carolina: Programming bacteria in time and space.
Esvelt, K., Massachusetts Institute of Technology Media Lab, Cambridge: Gene drives, ecological engineering, and confidence in science.
McClean, M., University of Wisconsin, Madison: Elucidating principles of biological signal processing using microfluidic and optogenetic tools.
Liu, C., University of California, Irvine: Orthogonal replication for rapid mutation and synthetic biology.
Kalasekar, S. (in place of Kiani, S.), Arizona State University, Tempe: CRISPR genetic circuits for reprogramming cellular function.
Shetty, R., Ginkgo Bioworks, Boston, Massachusetts: The emerging organism engineering industry.
Ajo-Franklin, C., Lawrence Berkeley National Lab, Berkeley, California: Engineering microorganisms that create and communicate with materials.

X-Ray Methods in Structural Biology

October 10–25

INSTRUCTORS
W. Furey, V.A. Medical Center/University of Pittsburgh, Pennsylvania
G. Gilliland, Janssen Research & Development, LLC, Spring House, Pennsylvania
A. McPherson, University of California, Irvine
J. Pflugrath, Rigaku Americas, The Woodlands, Texas

ASSISTANTS
L. Boucher, Janssen Research & Development, LLC, Spring House, Pennsylvania
B. Hintze, Duke University Medical Center, Durham, North Carolina

X-ray crystallography has been the cornerstone of structural biology for half a century. This intense laboratory/computational course focused on the major techniques used to determine the three-dimensional structures of macromolecules. It was designed for scientists with a working knowledge of protein structure and function, but who are new to macromolecular crystallography. Topics covered included basic diffraction theory, crystallization (proteins, nucleic acids, complexes, and membrane proteins), synchrotron X-ray sources and optics, data collection and processing, structure solution by experimental phasing methods (SAD, MAD, MIR, and others) and molecular replacement, electron density maps improvement (solvent flattening, noncrystallographic averaging, etc.), model building and refinement, structure validation, coordinate deposition, and structure presentation. In addition, the course for the first time extended to the theory and computation for small-angle X-ray scattering (SAXS) and single-particle cryoelectron microscopy.

Participants learned through extensive hands-on experiments in fully equipped labs, crystallized multiple proteins, and determined their crystal structures by several methods while learning through extensive lectures on theory. Informal discussions behind the techniques were frequent, and students were responsible also for collecting questions to be answered in specific sessions.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Alexander, J.A., B.S., University of British Columbia, Vancouver, Canada
Dharadhar, S., M.S., Netherlands Cancer Institute, Amsterdam, The Netherlands
Fedry, J., B.S., Institut Pasteur, Paris, France
Gharpure, A., S.B., University of Texas Southwestern Medical Center, Dallas
Grisshammer, R., Ph.D., National Institutes of Health, Rockville, Maryland
Hayward, A., B.S., University of Minnesota-Twin Cities, Minneapolis
Kim, S.K., B.S., University of Pittsburgh, Pennsylvania
Matsumoto, M., Ph.D., Genentech, Inc., South San Francisco, California
Meagher, M., Ph.D., St. Jude Children's Research Hospital, Memphis, Tennessee
Omattage, N., B.S., Washington University in St. Louis, Missouri
Petrovic, S., Ph.D., California Institute of Technology, Pasadena
Shelton, K., B.S., University of Oklahoma, Norman
Silvas, T., B.S., University of Massachusetts Medical School, Worcester
Sweeney, E., Ph.D., University of Oregon, Eugene
Washington, E., Ph.D., Duke University Medical Center, Durham, North Carolina
Williams, K., B.S., Medical University of South Carolina, Charleston

SEMINARS

Gilliland, G., Janssen Research & Development, LLC, Spring House, Pennsylvania: The structure determination of biological macromolecules; Maximizing crystallization success through seeding.
McPherson, A., University of California, Irvine: Crystallization of macromolecules I and II; Symmetry, periodicity, unit cells, space groups, Miller planes, and lattices; Waves, vectors, and complex numbers; Fundamental diffraction relationships, and Bragg's law; Diffraction patterns, reciprocal space, and Ewald's sphere; Fourier transforms and the electron density equation; Atomic force microscopy of macromolecular crystals: Visualization of crystallization at the molecular level.
Caffrey, M., Trinity College, Dublin, Ireland: Membrane protein crystallization for structure–function studies using bicelles and lipid mesophases.
Pflugrath, J., Rigaku Americas, The Woodlands, Texas: X-ray data collection and processing; Cryo-crystallography; Data collection strategy roadshow: Native.
Sweet, R., Brookhaven National Laboratory, Upton, New York: A slightly different view of fundamental crystallography; X-ray sources and optics.
Borek, D., University of Texas Southwestern Medical Center, Dallas: X-ray data processing.
Holton, J., University of California, San Francisco: Beamline basics and tips and tricks for improving diffraction?
Furey, W., University of Pittsburgh, Pennsylvania: Patterson space and Isomorphous replacement; Density modification: Solvent flattening, phase combination, NCS; Anomalous scattering, SAD and MAD; Direct methods.
Emsley, P., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: Model-building tools in Coot.
Thorn, A., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: An introduction to SHELXC/D/E; The basics of twinning in crystals of macromolecules.
Smith, C., Stanford University, California: Synchrotron data collection and femtosecond crystallography.
Tronrud, D., Oregon State University, Corvallis: Macromolecular refinement; Electron density maps.
Hammel, M., Lawrence Berkeley National Laboratory, Berekley, California: Small-angle scattering.
Terwilliger, T., Los Alamos National Laboratory, New Mexico: Automated structure solution and model building.
Read, R., University of Cambridge, United Kingdom: Molecular replacement: New structures from old; Using SAD data in phaser.
Hendrickson, W., Columbia University, New York, New York: MAD and SAD phasing.
Kleywegt, G., European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom: The Gerard show.
Adams, P., University of California, Berkeley: Structure refinement; PHENIX overview.
Conway, J., University of Pittsburgh, Pennsylvania: Cryo-electron microscopy introduction; Cryo-electron microscopy of viral systems.
Richardson, J., Duke University, Durham, North Carolina: Structure presentation.
Scheres, S., MRC Laboratory of Molecular Biology, Cambridge, United Kingdom: Processing electron microscopy data: RELION; High-resolution structures of complexes from electron microscopy.
Carragher, B., New York Structural Biology Center, The Rockefeller University, New York: EM automation.

Programming for Biology

October 10–25

INSTRUCTORS **S. Prochnik**, DOE/Joint Genome Institute, Walnut Creek, California
S. Robb, Stowers Institute for Medical Research, Kansas City, Missouri

ASSISTANTS **J. Bredeson**, University of California, Berkeley
M.l. Campbell, Cold Spring Harbor Laboratory
W. Ingram, Geisinger Health Systems, Danville, Pennsylvania
J. Orkin, University of Calgary, Alberta, Canada
E. Ross, Stowers Institute for Medical Research, Kansas City, Missouri
D. Triant, University of Virginia, Charlottesville

Web-based tools are no longer enough for today's biologist, who must access and analyze large data sets from myriad sources in disparate formats. The need to design and implement custom analysis pipelines is becoming ever more important as new technologies increase the already-exponential rate at which biological data are generated. Designed for lab biologists with little or no programming experience, this course gave students the bioinformatics and scripting skills necessary to exploit this abundance of biological data. The only prerequisite for the course was basic knowledge of UNIX; some scripting experience was also helpful.

This course taught Perl, a scripting language that is easy to learn and efficient to use. Perl also has a vast array of ready-built tools such as BioPerl that are designed to solve common biological problems. The course began with one week of introductory coding, continued with a survey of available biological libraries and practical topics in bioinformatics, and ended with a final group project. Formal instruction was provided on every topic by the instructors, teaching assistants, and invited experts. Students worked together to solve problem sets covering common scenarios in the acquisition, validation, integration, analysis, and visualization of biological data. They learned how to design, construct, and run powerful and extensible analysis pipelines in a straightforward manner. For their final projects, students posed problems using their own data and worked with each other and the faculty to solve them. In the past, final projects have formed the basis of

publications and freely available resources (see, for example, the Deobfuscator module in BioPerl). Students were provided with a library of Perl reference books that they took home with them.

Note that the primary focus of this course was to provide students with practical programming experience, rather than to present a detailed description of the algorithms used in computational biology. For the latter, we recommend the Computational and Comparative Genomics course.

This course was supported with funds provided by the National Human Genome Research Institute.

PARTICIPANTS

Almiron Bonnin, D., B.A., Norris Cotton Cancer Center/ Geisel School of Medicine, Lebanon, New Hampshire
Amato, C., M.S., University of Colorado Denver AMC, Aurora
Chella Krishnan, K., Ph.D., University of California, Los Angeles
Drazer, M., M.D., The University of Chicago Medical Center, Illinois
Gluck-Thaler, E., B.S., The Ohio State University, Columbus
Henry, L., B.A., Princeton University, New Jersey
Hershberger, C., B.S., Cleveland Clinic Foundation, Ohio
Krautkramer, K., B.S., University of Wisconsin, Madison
Krogh, N., M.S., University of Copenhagen, Denmark
Lepeta, K., M.S., Nencki Institute of Experimental Biology, Warsaw, Poland
Lowenstein, P., Ph.D., University of Michigan, Ann Arbor
Luo, Y.-J., B.S., Okinawa Institute of Science and Technology, Onna, Japan
Lyons, J., Ph.D., University of California, Berkeley
Pacheco, N., University of Alabama, Birmingham
Patel, S., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Pocock, A., B.S., University of Queensland, Forest Hill, Australia
Quach, T., Ph.D., Feinstein Institute of Medical Research, Great Neck, New York
Salem, H., Ph.D., Emory University, Atlanta, Georgia
Saville, K., Ph.D., Albion College, Michigan
Webb, S., B.A., University of Calgary, Canada
Yamaji, M., Ph.D., HHMI/The Rockefeller University, New York, New York

SEMINARS

Prochnik, S., Joint Genome Institute, Walnut Creek, California: Unix logins, Perl I: Scripts; Perl III: Files; Perl IX: References; Perl X: Object-oriented programming; Web programming with CGI.pm; Perl pipelines.
Robb, S., Stowers Institute for Medical Research, Kansas City, Missouri: Perl II: Functions; Perl IV: Arrays and list; Perl V: Hashes; Perl VI: Regular expressions; Perl VII: Subroutines; Perl VIII: Modules; Bio Perl.
Marques-Bonet, T., Institut Biologia Evolutiva, Barcelona, Spain: Structural variation.
Cain, S., Ontario Institute for Cancer Research, Medina, Ohio: GMOD and Jbrowse.
Ross, E., Stowers Institute for Medical Research, Kansas City, Missouri: Introduction to NGS.
Haas, B., Broad Institute, Northbridge, Massachusetts: Transcript assembly I; Transcript assembly II.
Cantarel, B., University of Texas Southwestern Medical Center, Richardson: Metagenomics I; Metagenomics II.
Witkowski, J., Cold Spring Harbor Laboratory: Ethics and reproducibility in research.
Korf, I., University of California, Davis: BLAST I; BLAST II.

Workshop on Cereal Genomics

October 19–25

INSTRUCTORS **S. Hake**, University of California, Berkeley, Albany
D. Jackson, Cold Spring Harbor Laboratory
D. Ware, Cold Spring Harbor Laboratory/USDA/ARS

This 1-week workshop enabled participants to take advantage of emerging genetic tools and the completed cereal genome sequences of most of the major cereal crops. The workshop featured morning and evening lectures with afternoon lab exercises, including hands-on lab work and computer sessions in comparative anatomy, GWASs, next-generation expression analysis, whole-genome sequencing assembly, emerging model systems, genome editing, and phenomics. The faculty (instructors and invited lecturers) were active researchers in cereal genetics and genomics who have made significant contributions to the field, ensuring that the latest techniques and ideas were presented. The course was structured to provide time for informal discussions and exchange with instructors.

Topics included comparative anatomy and phylogeny; cereal genomes, assembly, annotation and synteny; genetics and databases; quantitative trait locus mapping and genome-wide association studies; genome-wide expression analyses; reverse genetics and genome editing; and phenomics

PARTICIPANTS

Alexander, M., Ph.D., University of California, Berkeley
Bayable, D., M.S., Purdue University, West Lafayette, Indiana
Bowerman, A., Ph.D., The Australian National University, Canberra, Australia
Brophy, J., Ph.D., Massachusetts Institute of Technology, Cambridge
Guthrie, K., B.S., University of Missouri, Columbia
Haining, B., B.S, University of California, Berkeley
Klein, H., Ph.D., University of Massachusetts, Amherst
Klein, S., B.S., The Pennsylvania State University, University Park
Liberatore, K., Ph.D., U.S. Department of Agriculture-ARS, St. Paul, Minnesota
Liu, H., B.S., HuaZhong Agricultural University, China

Poretsky, E., B.S., University of California, San Diego
Seeve, C., Ph.D., U.S. Department of Agriculture-ARS, Columbia, Missouri
Shen, X., B.A., HuaZhong Agriculture University, China
Smith, A., B.S., University of Wisconsin, Milwaukee
Willyerd, K., Ph.D., Oklahoma State University, Stillwater
Wittmeyer, K., Ph.D., Pennsylvania State University, University Park
Zhu, C., Ph.D., Donald Danforth Plant Science Center, St. Louis, Missouri

SEMINARS

Buckler, E., U.S. Department of Agriculture-ARS, Ithaca, New York: Breeding 4.0: Sorting through the adaptive and deleterious variants across the maize genome.
Paszkowski, U., University of Cambridge, United Kingdom: Genetics and genomics of endosymbiosis: Fundamental science and translation.
Connor, D., University of Cambridge, United Kingdom: Brachypodium tools and their application to comparative developmental biology.
Xin, Z., U.S. Department of Agriculture-ARS, Lubbock, Texas: Pedigree mutant library as a resource for forward and reverse genetics in sorghum.
Brown, P., University of Illinois, Urbana: Using sorghum to understand and improve stress resilience in cereals.
Devos, K., University of Georgia, Augusta: Use of genotyping-by-sequencing in the genomic analyses of nonmodel plants: From inbreeding diploids to outcrossing polyploids.
Topp, C., Danforth Plan Science Center, Saint Louis, Missouri: Reality is a perception: Using phenomics to explore the hidden world of plant phenotypes.
Gao, C., Institute of Genomics and Developmental Biology, Beijing, China: Genome editing in wheat and corn.
Hudson, M., University of Illinois, Urbana: Extracting biological insight from gene expression data; RNA-Seq methods and data analysis in complex genomes: Dealing with uncertainty, alternative splicing.
Eveland, A., Danforth Plant Science Center, St. Louis, Missouri: Extracting biological insight from gene expression data.
Harper, L., University of California, Berkeley: Cereals and their genomes.
Kellogg, E., University of Missouri, St. Louis: Introduction to the grasses and major subgroups; Evolution of ecologically dominant grasses.
Ware, D., Cold Spring Harbor Laboratory: Genome assembly.

Computational and Comparative Genomics

October 26–November 3

INSTRUCTORS **A. Mackey**, HemoShear, LLC, Charlottesville, Virginia
W. Pearson, University of Virginia, Charlottesville
J. Taylor, Johns Hopkins University, Baltimore, Maryland

ASSISTANTS **P. DeFord**, Johns Hopkins University, Baltimore, Maryland
O. Sabik, University of Virginia, Charlottesville
C. Seward, University of Illinois, Urbana
J. Troy, University of Illinois, Urbana

This course presented a comprehensive overview of the theory and practice of computational methods for the identification and characterization of functional elements from DNA sequence data. The course focused on approaches for extracting the maximum amount of information from protein and DNA sequence similarity through sequence database searches, statistical analysis, and multiple sequence alignment.

Additional topics included:

- alignment and analysis of "next-gen" sequencing data, with applications from metagenomic, RNA-Seq, and CHiP-seq experiments
- the Galaxy environment for high-throughput analysis
- regulatory element and motif identification from conserved signals in aligned and unaligned sequences
- integration of genetic and sequence information in biological databases
- genome browsers and genome features

The course combined lectures with hands-on exercises; students were encouraged to pose challenging sequence analysis problems using their own data. The course was designed for biologists seeking advanced training in biological sequence and genome analysis, computational biology core resource directors and staff, and for individuals in other disciplines (e.g., computer science)

who wish to survey current research problems in biological sequence analysis. Advanced programming skills were not required.

The primary focus of this course was the theory and practice of algorithms in computational biology, with the goals of using current methods more effectively for biological discovery and developing new algorithms.

This course was supported with funds provided by the National Human Genome Research Institute.

PARTICIPANTS

Aydin, B., B.S., New York University, New York
Barry, R., B.S., New York University Langone Medical Center, New York
Bendixsen, D., B.S., Boise State University, Indiana
Bozek, M., M.S., B.A., Ludwig Maximillians University, Munich, Germany
Gart, E., Ph.D., Texas A&M University, College Station
Gedling, C., B.A., The Ohio State University, Wooster
Guo, Z., Ph.D., AbbVie, Worcester, Massachusetts
Gupta, S., Ph.D., Duke University, Durham, North Carolina
Haenel, G., Ph.D., Elon University, North Carolina
Jaramillo Cartagena, A., B.A., The Rockefeller University, New York, New York
LaHaye, S., B.S., Nationwide Children's Hospital, Columbus, Ohio
Xiaoli, Ma, X., M.S., University of Tübingen, Germany
Ondracka, A., Ph.D., Universitat Pompeu Fabra, Barcelona, Spain
Ortega, V., Ph.D., Wake Forest School of Medicine, Winston-Salem, North Carolina
Pribluda, A., Ph.D., Genentech, South San Francisco, California
Risolino, M., Ph.D., University of California, San Francisco
Salazar, V., Ph.D., Cape Breton University, Sydney, Nova Scotia, Canada
Vis, M., Ph.D., Ohio University, Athens
Weaver, C., B.S., University of Kentucky, Lexington
Zhang, X., Ph.D., Columbia University, New York, New York

SEMINARS

Pearson, W., University of Virginia, Charlottesville: Protein evolution and sequence similarity searching; Practical sequence similarity searching; Multiple sequence alignment; Statistics for genomes I: Multiple tests–Similarity searching; PSSMs, HMMs, and phenotype prediction.
Mackey, A., HemoShear, LLC, Charlottesville, Virginia: Genome annotation (HMM basics); Gene lists to pathways; RNA-Seq.
Taylor, J., Johns Hopkins University, Baltimore, Maryland: Sequencing technologies—New genomics; Galaxy for high-throughput analysis; Variation and SNP discovery; Assembling genomes and transcriptomes; Galaxy visualization; Probing higher-dimension chromatin structure.
Hardison, R., Pennsylvania State University, University Park: Genomics of gene regulation 1: Epigenomics; Genomics of gene regulation 2: Feature integration.
Hawkins, D., University of Washington, Seattle: Chromatin states 1: Analysis of histone modifications; Chromatin states 2: Overlapping data sets.
Stormo, G., Washington University in St. Louis, Missouri: Modeling motifs: Collecting data; From motifs to regulatory networks.

Antibody Engineering, Phage Display, and Immune Repertoire Analysis

October 28–November 10

INSTRUCTORS **D. Siegel**, University of Pennsylvania School of Medicine, Philadelphia
G. Silverman, New York University School of Medicine, New York

ASSISTANTS **R. Güler**, Royal Institute of Technology, Stockholm, Sweden
D.N. Hernandez, Sackler institute of Biomedical Science at New York University, New York
M.A. Pohl, Tri-Institutional Therapeutics Discovery Institute, New York, New York

Recent advances in the generation and selection of antibodies from combinatorial libraries allow for the rapid production of antibodies from immune and nonimmune sources. This intensive laboratory/lecture course focused on the construction of combinatorial antibody libraries expressed on the surface of phage and selection of desired antibodies from the library. Students learned the theoretical and practical aspects of constructing combinatorial libraries from immune and nonimmune sources as well as the construction of synthetic antibody libraries. Antibodies were selected from the library by panning. Production, purification, and characterization of Fab fragments expressed in *E. coli* were also covered. In addition, epitopes were selected from synthetic peptide libraries and characterized.

The lecture series, presented by course faculty and a number of invited speakers, emphasized PCR of immunoglobulin genes, the biology of filamentous phage, and the utility of surface expression libraries, expression of antibodies in *E. coli* and mammalian cells, antibody structure and function, catalytic antibodies, directed protein evolution, retroviral and cell display libraries, the immunobiology of the antibody response, and recent results on the use of antibodies in therapy. The theory and practical implications for selection from phage-displayed libraries of

random peptides, cDNA products, and semisynthetic proteins were also explored. We also discussed principles and protocols for generation and analysis of immune repertoires determined by next-generation sequencing.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Bhatnagar, P., Ph.D., SRI International, Menlo Park, California
Reddy Devulapally, P.R., Ph.D., Max-Planck-Institute for Molecular Genetics, Berlin, Germany
Georges, A., M.S, Immunocore Ltd., Abingdon, United Kingdom
Hoseini, S.S., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York
Kudelka, M., B.A., Beth Israel Deaconess Medical Center, Boston, Massachusetts
Kwon, Y.D., Ph.D., Vaccine Research Center, Bethesda, Maryland
Martinez-Guzman, D., Ph.D., Grifols Diagnostic Solutions, Emeryville, California
Mener, A., B.A., B.S.H., Emory University School of Medicine, Atlanta, Georgia
Motley, M., B.S., Stony Brook University, New York
O'Malley, C., M.Res., Queen Mary University, London, United Kingdom
Radke, E., B.A., New York University School of Medicine, New York
Sawyer, A., H.N.C., The Jackson Laboratory, Bar Harbor, Maine
Strippoli, L., Ph.D., Diasorin, Gerenzano, Italy
Turner, A., B.S., Ipsen Bioscience, Inc., Lakeville, Massachusetts
Wawrzyniecka, P., M.S., University College London, United Kingdom
Xu, K., Ph.D., National Institutes of Health, Bethesda, Maryland

SEMINARS

Rader, C., The Scripps Research Institute, Jupiter, Florida: From phage display to cancer immunotherapy.
Payne, A., University of Pennsylvania, Merion Station: Lineage tracing of isotype-specific B-cell repertoires in pemphigus vulgaris.
Silverman, G., New York University School of Medicine, New York: Cellular and molecular features of the B-cell response.
Sidhu, S., University of Toronto, Ontario, Canada: Modulation of cell signaling with synthetic proteins.
Siegel, D., University of Pennsylvania, Philadelphia: Antibody repertoires, autoantibodies and CAR T.
Noren, C., New England Biolabs, Ipswich, Massachusetts: Top down and bottom up approaches to characterizing phage displayed peptide libraries.
Stahl, S., KTH Royal Institute of Technology, Sweden: Generating affibody molecules for therapy applications.
Boyd, S., Stanford University, California: B-cell repertoire responses to vaccination and infection.
DeKosky, B., NIH/NIAD, Washington, D.C.: High-throughput sequencing of paired antibody heavy and light chains.

Advanced Sequencing Technologies and Applications

November 8–19

INSTRUCTORS

M. Griffith, Washington University School of Medicine in St. Louis, Missouri
O. Griffith, Washington University School of Medicine in St. Louis, Missouri
E. Mardis, Washington University School of Medicine in St. Louis, Missouri
R. McCombie, Cold Spring Harbor Laboratory
A. Quinlan, University of Utah, Salt Lake City

ASSISTANTS

A. Farrell, University of Utah, Salt Lake City
S. Goodwin, Cold Spring Harbor Laboratory
J.S. Havrilla, University of Utah, Salt Lake City
M. Kramer, Cold Spring Harbor Laboratory
V. Magrini, Nationwide Children's Hospital, Columbus, Ohio
S. McGrath, Nationwide Children's Hospital in St. Louis, Missouri
T. Sasani, University of Utah, Salt Lake City
A. Wagner, Washington University in Saint Louis, Missouri
J. Walker, Washington University School of Medicine in St. Louis, Missouri
A. Ward, University of Utah, Salt Lake City
H. Abel, Washington University School of Medicine in St. Louis, Missouri

Over the last decade, massively parallel DNA sequencing has markedly impacted the practice of modern biology and is being utilized in the practice of medicine. The constant improvement of these platforms means that costs and data generation timelines have been reduced by orders of magnitude, enabling investigators to conceptualize and perform sequencing-based projects that heretofore were time-, cost-, and sample number–prohibitive. Furthermore, the application of these technologies to answer questions previously not experimentally approachable is broadening their impact and application. However, data analysis remains a complex and often vexing challenge, especially as data volumes increase.

This intensive 2-week course explored the use and applications of massively parallel sequencing technologies, with a focus on data analysis and bioinformatics. Students were instructed in the detailed operation of several platforms, including library construction procedures, general data processing, and in-depth data analysis. A diverse range of the types of biological questions enabled by massively parallel sequencing technologies was explored, including DNA resequencing of known cancer genes, de novo DNA sequencing and assembly of genomes, RNA sequencing, and others that were tailored to the students' research areas of interest.

Cloud-based computing was also explored. Guest lecturers highlighted unique applications of these disruptive technologies. We encouraged applicants from a diversity of scientific backgrounds including molecular evolution, development, neuroscience, medicine, cancer, plant biology, and microbiology.

This course was supported through funds provided by the National Human Genome Research Institute.

PARTICIPANTS

Braverman, J., Ph.D., Saint Joseph's University, Philadelphia, Pennsylvania
Burkhart, K., Ph.D., Massachusetts Institute of Technology/HHMI, Cambridge
Byrd, J.B., M.D., University of Michigan Medical School, Ann Arbor
Cameron, T., Ph.D., McGovern Medical School, Houston, Texas
Cornejo Castro, E., Ph.D., Leidos Biomedical Research Inc., FNL CR, Frederick, Maryland
Frades, I., Ph.D., Mount Sinai Hospital, New York
Heang, V., M.BA., U.S. Naval Medical Research Unit.2, Phnom Penh, Cambodia
Herre, M., B.F.A., The Rockefeller University, New York
Huang, M., Ph.D., University of California, San Francisco
Hunter, T., M.S., Peter MacCallum Cancer Centre, Melbourne, Australia
Kappelmann-Fenzl, M., Ph.D., University of Erlangen, Germany
Kopinski, P., B.A., Children's Hospital of Philadelphia, Pennsylvania
Mearls, E., Ph.D., Mount Holyoke College, South Hadley, Massachusetts
Neavin, D., B.S., Mayo Clinic, Rochester, Minnesota
Taylor, S., B.S., Rutgers University, Piscataway, New Jersey
Walavalkar, N., Ph.D., University of Virginia, Charlottesville
Walport, L., Ph.D., University of Tokyo, Japan
Wei, Y., Ph.D., Scripps Research Institute, La Jolla, California
Wilderman, A., M.S., University of Connecticut, Health Sciences, Farmington
Yasi, E., B.S., Georgia Institute of Technology, Atlanta

SEMINARS

Mardis, E., Nationwide Children's Hospital, Columbus, Ohio: NGS-based cancer genomics.
Marth, G., University of Utah, Salt Lake City: Variant discovery.
Sasani, T., University of Utah, Salt Lake City: Disease variant discovery with GEMINI.
Quinlan, A., University of Utah, Salt Lake City: Disease variant discovery with GEMINI; Genome arithmetic with BEDTools and challenge problems.
Hoffman, M., Princess Margaret Cancer Centre, Toronto, Ontario, Canada: Data integration techniques.
Griffith, M., Washington University School of Medicine in St. Louis, Missouri: Intro to RNA sequencing.
Haas, B., Broad Institute, Northbridge, Massachusetts: Transcript assembly.
Sedlazeck, F., Johns Hopkins University, Baltimore, Maryland: Single-cell CNV analysis.
Dewar, K., McGill University/Genome QC Innovation Centre, Montreal, Quebec, Canada: Intro to genome assembly of PacBio + Illumina data.
Moore, B., University of Utah, Salt Lake City: VAAST.
Elemento, O., Weill Cornell Medicine, New York, New York: Introduction to epigenomics analysis.
Hall, I., Washington University in Saint Louis, Missouri: SV detection and SpeedSeq pipeline
Goecks, J., Oregon Health and Science University, Portland: Analysis with Galaxy.

Scientific Writing Retreat

November 16–20

INSTRUCTORS **C. Lambert**, Cold Spring Harbor Laboratory
S. Matheson, *Cell Reports*, Cambridge, Massachusetts

WRITING COACHES **S. Gary**, Cold Spring Harbor Laboratory
E. Gaskell, Broad Institute, Cambridge, Massachusetts
C. Martin, Current Biology, Cambridge, Massachusetts
J. Rubin, Columbia University, New York, New York

The goal of this retreat was to have participants progress significantly on writing projects while improving their professional communication skills. The retreat included a mix of formal sessions and less structured writing time. The formal sessions covered publication writing for scientific journals from the perspectives of Cell Press and Cold Spring Harbor Press; writing clearly and conversationally about your research in ways that engage diverse audiences, a skill particularly useful when developing lay summaries for NIH and NSF proposals; and style tips and considerations for clear professional writing in all forms.

The less-structured sessions of the retreat included small writing groups and dedicated individual writing time. For the small group sessions, participants were preassigned to groups of three to four people for the purpose of soliciting peer feedback on writing samples they submitted ahead of time. For the individual writing sessions, coaches were on hand to work with participants

one-on-one. As with all CSHL meetings and courses, participants were required to respect the confidentiality of any unpublished research they may have read during the retreat.

Two of the formal sessions were led by instructors from the Alan Alda Center for Communicating Science at Stony Brook University. The Alda Center is dedicated to helping scientists communicate their research effectively to people outside their fields, including the general public, policy makers, the media, students, potential employers or funders, and prospective collaborators in other disciplines. Through discussion and practice, the Center's workshops focus on fundamental skills such as knowing your audience, connecting with it, communicating your research clearly, and explaining why it matters. The Center is a leader in efforts to improve science communication nationally and has led workshops for scientists at more than 130 institutions.

This course was supported with funds provided by the National Institute of General Medical Sciences.

PARTICIPANTS

Asare, A., Ph.D., Massachusetts General Hospital, Bethesda, Maryland
Bains, P., Ph.D., University of Southern California, Los Angeles
Barbier, M., Ph.D., West Virginia University, Morgantown
Burhans, M., Ph.D., Fred Hutchinson Cancer Research Center, Seattle, Washington
Fluegge, K., Ph.D., NYC Dept of Health and Mental Hygiene, Long Island City, New York
Jadavji, N., Ph.D., Carleton University, Ottawa, Canada
Kuscu, C., Ph.D., University of Virginia, Charlottesville
Lanning, B., Ph.D., Cold Spring Harbor Laboratory
Leone, J., Ph.D., The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut
Malvankar, N., M.S., Ph.D., Yale University, West Haven, Connecticut
Mayer, A., Ph.D., Harvard Medical School, Boston, Massachusetts
Quiros, M., M.S., Ph.D., University of Michigan, Ann Arbor
Scharf, A., Ph.D., Washington University in St. Louis, Missouri
Siegrist, S., Ph.D., University of Virginia, Charlottesville
Simpson, E., Ph.D., NYSPI/Columbia University, New York, New York
Zhang, X.-J., Ph.D., Memorial Sloan Kettering Cancer Center, New York, New York

SEMINARS

Marsh, S., Alan Alda Center for Communicating Science, Stony Brook, New York: Session on improvisation.
Lambert, C., Cold Spring Harbor Laboratory, and Matheson, S., *Cell Reports*, Cambridge, Massachusetts: Top 10 tips for strong professional writing.
Bass, L., and Chedd, G., Alan Alda Center for Communicating Science, Stony Brook, New York: Writing lay summaries for nonscientists.
Matheson, S., Cell Press, Cambridge, Massachusetts: Publications and manuscripts.

Immersive Approaches to Biological Data Visualization

December 1–10

INSTRUCTORS **K. Gaither**, Texas Advanced Computing Center, Austin
M. Vaughn, Texas Advanced Computing Center, Austin

CO-INSTRUCTOR **B. McCann**, University of Texas, Austin

ASSISTANTS **S.N. Yap**, University of Texas, Austin
A. Hall, University of Texas, Austin

This course explored the new frontier of immersive, interactive scientific visualization. It was aimed at researchers who wished to design and implement innovative visualization tools to help address biological or biomedical research questions. As part of their research programs, course participants had a need to visualize genomic or systems biology data, anatomical structures and models, scientific simulations, and/or other domains that generate complex, multidimensional, and often very large data sets.

Through a combination of detailed lectures and intensive hands-on computing exercises led by course instructors, students learned the principles of information and scientific visualization. However, rather than relying on simplified representations of complex data, students learned these visualization principles within the newly emerging context of immersive augmented and virtual reality. They worked with tools such as Unity 3D and Blender, as well as scripting environments such as Bash and Python, to craft data visualizations using the Oculus Rift virtual reality headset and Leap Motion gesture sensor. Students applied the concepts and techniques covered in class exercises to create an independent virtual reality data visualization project. Additionally, the

hands-on training was supplemented by guest lecturers from some of the leading experts working in data visualization today.

Some selected major topics were an overview of scientific visualization; perception and color; information visualization; scientific visualization; importing scientific data into an immersive visualization framework; and tool development using Oculus and Leap Motion.

This course was supported with funds provided by the Helmsley Charitable Trust.

PARTICIPANTS

Hebrard, M., Ph.D., RIKEN, Yokohama, Japan
Ho, Y.-J., Ph.D., Cold Spring Harbor Laboratory
Ko, K.D., Ph.D., National Institutes of Health, Bethesda, Maryland
Leone, J., Ph.D., The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut
Lepeta, K., M.S., Nencki Institute of Experimental Biology, Warsaw, Poland
Rauscher, B., M.S., German Cancer Research Center, Heidelberg, Germany
Rennerfeldt, D., B.S., Massachusetts Institute of Technology, Cambridge
Sheppard, T., B.A., Stanford School of Medicine, California
Werth, E., B.S., University of North Carolina, Chapel Hill

SEMINARS

Chen, M., Oxford University, Cambridge, United Kingdom: Information theory and visualization; Glyph-based visualization.
Varshney, A., University of Maryland, College Park: Visual computing with big data.
Ebart, D., Purdue University School of Electrical & Computer Engineering, West Lafayette, Indiana: Visual analytics for scientific and biological discovery.
Beyer, J., Harvard University, Cambridge, Massachusetts: Interactive visualization for connectomics research.

Organotypic and Next-Generation Cultural Methods

December 1–14

INSTRUCTORS **H. Clevers**, Hubrecht Institute, Utrecht, The Netherlands
D. Tuveson, Cold Spring Harbor Laboratory

CO-INSTRUCTOR **L. Baker**, Cold Spring Harbor Laboratory

ASSISTANTS **R. Arnes Benito**, University of Cambridge, United Kingdom
C. W. Chua, Columbia University Medical Center, New York, New York
D. Georgess, Johns Hopkins School of Medicine, Baltimore, Maryland
O. Kopper, Hubrecht Institute, Utrecht, The Netherlands
K. Kretzchmar, Hubrecht Institute, Utrecht, The Netherlands
B. Alagesan, Cold Spring Harbor Laboratory
G. Biffi, Cold Spring Harbor Laboratory
E. Elyada, Cold Spring Harbor Laboratory
D. Engle, Cold Spring Harbor Laboratory
C. Schoepfer, Cold Spring Harbor Laboratory
H. Tiriac, Cold Spring Harbor Laboratory
M. Yao, Cold Spring Harbor Laboratory

Recently, a number of technological advances have prompted the development of better methods to use cells cultured ex vivo as models for biological systems. These "next-generation" culture systems circumvent some of the limitations of traditional, two-dimensional cell culture and therefore hold great promise as models that will more accurately recapitulate biological processes and interactions. This course used both seminar and hands-on sessions to explore these systems, with a primary focus on 3D organoid culture. The course was designed for researchers with some cell culture experience, who wished to utilize 3D or conditionally reprogrammed culture methods in their research. Students need not have had any prior experience with 3D or conditionally reprogrammed cell culture methods.

Through seminars presented during the course, students developed an understanding of various organoid and next-generation culture systems. Students learned to compare and contrast different types of organoid models, including stem cell organoid and tissue organoid cultures. Important aspects of organoid culture such as tissue dissociation, culture media, and 3D culture matrices were also discussed. Students learned how co-culture of organoids with stromal cells can be used to model interactions between epithelial and mesenchymal cells. In addition, they learned about the generation and culture of conditionally reprogrammed cells, as an alternative to organoid culture. They also learned how these next-generation culture systems are currently being used to study development, disease progression, and tumorigenesis. The use of human patient-derived organoids for biomedical research was also discussed.

In the laboratory portion of the course, students gained hands-on experience generating, culturing, and passaging different organoid models, including stem cell organoids derived from pancreatic, liver, mammary, and intestinal tissue as well as mammary tissue organoids and prostate organoids. In addition, students learned to generate and passage conditionally reprogrammed cultures. They gained first-hand experience in methods for downstream analysis of organoid cultures, including IC_{50} analysis following chemical treatment and organoid imaging. Methods to transplant organoids into mice for in vivo analysis were also covered.

This course was supported with funds provided by the Helmsley Charitable Trust.

PARTICIPANTS

Afelik, S., M.S., Ph.D., University of Illinois, Chicago.
Bianconi, D., M.S., Medical University of Vienna, Austria
Bilal, F., M.S., Vall'd'Hebron Institute of Oncology, Barcelona, Spain
Campanale, J., Ph.D., University of California, Santa Barbara
Gorkin, D., Ph.D., University of California, San Diego, La Jolla
Harbig, J., M.S., University Hospital Tübingen, Germany
Hoh, H.H., M.D., Okinawa Institute of Science and Technology, Okinawa, Japan
Jo, S.-Y., B.S., Asan Institute for Life Sciences, Seoul, South Korea
Lee, B.R., M.S., Asan Medical Center, Seoul, South Korea
Li, J., M.D., The University of Hong Kong, China
Mohammad, J., B.S., North Dakota State University, Moorhead, North Dakota
Patel, S., Ph.D., Cold Spring Harbor Laboratory
Redondo, P., Ph.D., University College London, United Kingdom
Sharma, K., M.S., École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Suh, Y.-A., Ph.D., Asan Medical Center, Seoul, South Korea
Wu, P.-S., Ph.D., MacKay Memorial Hospital, New Taipei City, Taiwan

SEMINARS

Tuveson, D., Cold Spring Harbor Laboratory: Organoid models of ductal pancreatic cancer.
Ewald, A., Johns Hopkins School of Medicine, Baltimore, Maryland: Primary tissue organoids from normal mammary glands and breast tumors.
Huch, M., University of Cambridge, United Kingdom: Liver stem/progenitor cells and hepatic organoids to study liver biology and disease.
Agarwal, S., Georgetown University, Washington D.C.: Conditional cell reprogramming: Next-generation culture of normal and tumor cells.
Shen, M., Columbia University Medical Center, New York, New York: Organoid models for prostate and bladder.
Krawczyk, E., Georgetown University Medical Center, Washington D.C.: Conditional reprogramming and long-term expansion of normal and tumor cells.
Kim, C., Children's Hospital/Harvard Medical School, Boston, Massachusetts: Lung organoids: Modeling mechanisms of differentiation, disease, and therapy.
Clevers, H., Hubrecht Institute, Utrecht, The Netherlands: Organoids grown from adult stem cells.

The Genome Access Course

The Genome Access Course (TGAC) is an intensive 2-day introduction to bioinformatics that was held three times in 2016 and trained 99 participants in total. Registration for the course is open to all on a first-come, first-served basis subject to basic eligibility requirements. The core of TGAC covers the manipulation and analysis of sequence data using publicly available tools. The course is broken into modules designed to give a broad overview of a given topic, with ample time for hands-on exercises chosen by the instructors. Each module includes three parts: (1) a discussion of theory and methods, (2) coverage of software and Web resources, and (3) use of selected tools in practice. The modular design allows the instructors to tailor the curriculum to the interests of the students. Modules in 2016 included Genome Sequencing and Assembly, the UCSC Genome Browser, Ensembl, Comparative Genome Analysis, Functional Genomic Elements and The ENCODE Project, Gene Set Enrichment and Pathway Analysis, Sequence Polymorphisms, Next-Generation Sequence Data Analysis, The Galaxy Project, and RNA-seq Analysis Using R. Students were encouraged to bring questions and data from their own research projects to the course, and they were also encouraged to contact instructors with additional questions once they returned to their home institutions.

As in 2013–2015, one TGAC in 2016 was held in Manhattan at the New York Genome Center (NYGC, September 19–21). The course was one of several collaborative training projects with the NYGC, for which CSHL is an institutional founding member. It had a very high enrollment of 43 students, more than 50% of whom were from local New York institutions.

This course was supported with funds provided by the Helmsley Charitable Trust.

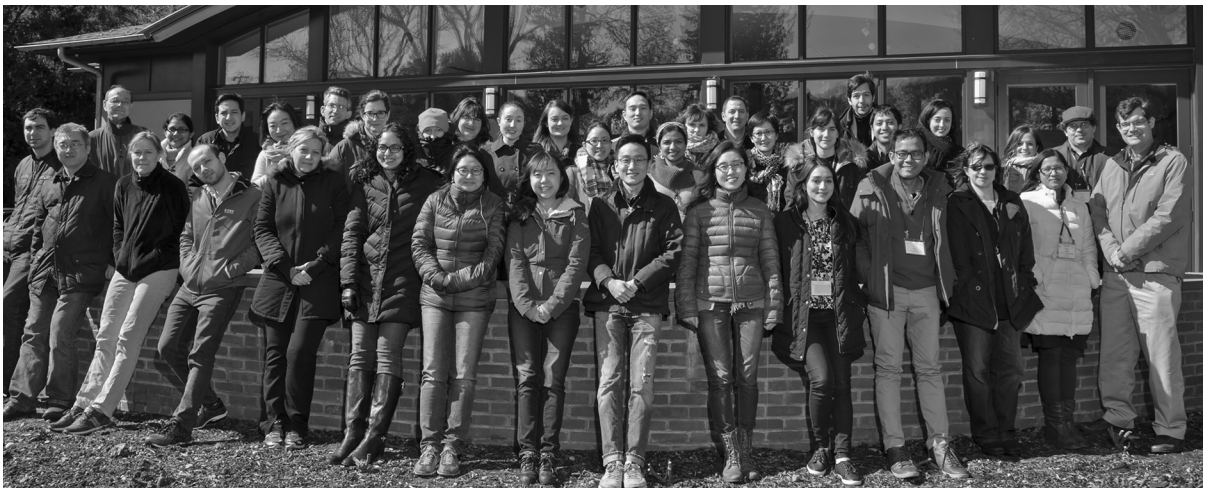
INSTRUCTORS
A. Gordon, New York Genome Center, New York
E. Hodges, Vanderbilt University School of Medicine, Nashville, Tennessee
G. Howell, The Jackson Laboratory, Bar Harbor, Maine
B. King, Mount Desert Island Biological Laboratory, Salisbury Cove, Maine
J. Ward, Middlebury College, Middlebury, Vermont

LECTURERS
T. Baslan, Memorial Sloan Kettering Cancer Center, New York, New York
S. Munger, The Jackson Laboratory, Bar Harbor, Maine

April 4–6

34 PARTICIPANTS (Cold Spring Harbor Laboratory)

Ahrens, S., Cold Spring Harbor Laboratory
Alter, E., York College/CUNY, Jamaica, New York
Batlevi, C., Memorial Sloan Kettering Cancer Center, New York, New York
Bausch-Jurken, M., Medical College of Wisconsin, Milwaukee
Bendapudi, P., Massachusetts General Hospital, Boston
Boire, A., Memorial Sloan Kettering Cancer Center, New York, New York
Chi, J., The Rockefeller University, New York, New York
Choi, C.H., The Rockefeller University, New York, New York
Choi, H., Memorial Sloan Kettering Cancer Center, New York, New York
Conway, C., BD GenCell, Limerick, Ireland
DeWille, J., Ohio State University, Columbus
Freedberg, D., M.D., Columbia University Medical Center, New York, New York
Hazra, R., Cold Spring Harbor Laboratory
Hsu, E., SUNY-Downstate Medical Center, Brooklyn, New York
Hui-Yuen, J., Feinstein Institute for Medical Research, CCMC, Lake Success, New York
Kim, Y.J., Cold Spring Harbor Laboratory
Kwak, Y., Kyungpook National University, Daegu, South Korea
Lai, M., University of Wyoming, Laramie
Miller, B., Ph.D., B.A., University of Florida, Gainesville
Motriuk-Smith, D., University of Wyoming, Casper
Nichols, A., Virginia Tech, Blacksburg
Saudale, F., Oklahoma State University, Stillwater



Schmandt, N., University of Chicago, Illinois
Schneider, J., Memorial Sloan Kettering Research Center, New York, New York
Shrestha, E., NYU School of Medicine, New York
Simpfendorfer, K., Feinstein Institute for Medical Research, Manhasset, New York
Singh., A., Columbia University Medical Center, New York, New York
Singh, S., Icahn School of Medicine, Mount Sinai, New York
Singh, S., Feinstein Institute for Medical Research, Manhasset, New York
Sun, P., Boehringer Ingelheim Pharmaceuticals, Ridgefield, Connecticut
Tamayo Orrego, L., McGill University, Montreal, Quebec, Canada
Tets, G., TGV-inhalonix, New York, New York
Vallanat, B., USEPA, Durham, North Carolina
Wang, G., Children's Hospital Oakland Research Institute, California

September 19–21

43 PARTICIPANTS (New York Genome Center)

Adli, M., University of Virginia, Charlottesville
Ally, R., Regeneron, Tarrytown, New York
Bayer Horowitz, L., New York University Langone Medical Center, New York
Bhat, V., McGill University, Montreal, California
Chorro, L., Albert Einstein College of Medicine, Bronx, New York
Gnuegge, R., Columbia University, New York, New York
Guemez Gamboa, A., The Rockefeller University, New York, New York
Hale., C., The Rockefeller University, New York, New York
Holzer, R., Rosetta Biomedical Research Institute, San Jose, California
Iuga, A., Hoboken, New Jersey
Jain, S., New York Blood Center, Manhattan
Jang, S.H., The Feinstein Institute for Medical Research, Manhasset, New York
Kaihami, G., Institute of Chemistry/University of Sao Paulo, Brazil
Kalin, T., Cincinnati Children's Hospital, Ohio
Kim, Y.J., Dongguk University, South Korea
Kulkarni, A., Albert Einstein College of Medicine, Bronx, New York
Lakhani, S., Yale University School of Medicine, New Haven, Connecticut
Liu, J., Northwell Health/Monter Cancer Center, Lake Success, New York
Llanos, A., Rutgers School of Public Health and CINJ, Piscataway, New Jersey
Lopez Soto, E., Brown University, Providence, Rhode Island
Maglic, D., Harvard University/Boston Children's Hospital, Massachusetts
Marin-Valencia, I., The Rockefeller University, New York, New York
Moerdler, S., Albert Einstein College of Medicine, Bronx, New York
Murphy, E., The Rockefeller University, New York, New York
Nandakumar, V., Altius Institute for Biomedical Research, Seattle, Washington

Niewiadomska, A.M., Somerset, New Jersey
Nowotschin, S., Memorial Sloan Kettering Institute, New York, New York
Orange, D., The Rockefeller University, New York, New York
Ottoz, D., Yale University New Haven, Connecticut
Pinzaru, A., NYU School of Medicine, New York
Radchenko, E., Tufts University, Medford, Massachusetts
Rollins, D., Weill Cornell Graduate School, New York, New York
Romberg, N., University of Pennsylvania, Philadelphia
Rossillo, M., NYU Medical Center, New York
Salvatore, J., Virginia Commonwealth University, Richmond
Truong, D., New York University Langone Medical Center, New York
Vele, O., Lenox Hill Hospital/Northwell Health System, New York, New York
Wang, Y., Columbia University Medical Center, New York, New York
Weinberg, R., Harvard Medical School, Boston, Massachusetts
Yang, J.Y., Columbia Presbyterian, New York, New York
Yang, W., Memorial Sloan Kettering Cancer Center, New York, New York
Zayas, G., New York City Department of Health, Brooklyn, New York
Zhang, Y., Columbia University, New York, New York

November 7–9

22 PARTICIPANTS (Cold Spring Harbor Laboratory)

Banks, L., Cold Spring Harbor Laboratory
Brennan, K., Alltech Inc., Nicholasville, Kentucky
Carlson, K.-S., Medical College of Wisconsin, Milwaukee
Cheong, H.I., Cleveland Clinic Lerner Research Institute, Ohio
Doerfler, P., St. Jude Children's Research Hospital, Memphis, Tennessee
Forsberg, C., University of California, Santa Cruz
Garcia, M., University of Rochester, New York
Gross, M., Ariel University, Israel
Heider, M., New England Biolabs, Inc., Ipswich, Massachusetts
Hwang, C.-I., Cold Spring Harbor Laboratory
Iwema, C., University of Pittsburgh, Pennsylvania
Jaitovich, A., Albany Medical College, New York
Jan, T., Massachusetts Eye & Ear/Harvard University, Boston
Kimonis, V., University of California, Irvine
Koirala, P., University of Mississippi Medical Center, Jackson
Rose, M., Brigham and Women's Hospital, Boston, Massachusetts
Schneider, K., SUNY Upstate Medical University, Syracuse, New York
Swartzman, E., Thermo Fisher Scientific, South San Francisco, California
Walter, K., University of Kansas Medical Center, Kansas City
Ward, J., MilliporeSigma, St. Louis, Missouri
Ward, S., MilliporeSigma, St. Louis, Missouri
Wu, J., University of Oklahoma Health Sciences Center, Oklahoma City

The Laboratory acknowledges the generosity of the following companies that loaned equipment and reagents to the various courses:

10X Genomics
AB SCIEX LLC
Advanced Illumination
Agilent Technologies Inc.
A-M Systems; Andor Technologies Inc.
Applied Scientific Instrumentation Inc.
BD Biosciences
Berthold Technologies USA, LLC
BioLegend
Bioline USA
Bio-Rad Laboratories
BioTek
Biotool
Bitplane
Bruker Daltonics Inc.
Cairn Research Ltd.
Charles River Laboratories Inc.
Chroma Technology Corporation
Coherent
ConOptics
Covaris, Inc.
CrystaLaser
Diagenode
Drummond Scientific Company
eBioscience
EMD Millipore Corp.
Epicentre Technologies Corp.
Eppendorf North America
Excelitas Technologies Corp.
Exiqon, Inc.
Fluidigm; Fotodyne Inc.
GE Healthcare
Hamamatsu Photonic Systems
Hamilton Drayage, Inc.
Hampton Research Harlan
Harvard Apparatus Inc.
Heka Instruments Inc.
Holoeye
Illumina Inc.
Intelligent Imaging Inc.
Invitrogen, Life Technologies
Kinetic Systems Inc.
Labcyte
Leica Biosystems
Leica Microsystems Inc.
LI-COR
Lumen Dynamics
Lumencor
MDS Analytical Technologies
Molecular Devices Corp.
Nanodrop Technologies
NanoTemper Technologies
Narishige International USA Inc.
Nasco; New England Biolabs Inc.
Newport Corporation
NightSea; Nikon Inc.
NSK America Corporation
Olympus American Inc.
PerkinElmer Life and Analytical Sciences
Photometrics
Physik Instruments
Precisionary Instruments
Prior Scientific
Promega Corporation
QIAGEN Inc.
Quantitative Imaging Corp.
RC Testing Services
Sage Science Inc.
Scientifica Ltd.
Singer Instruments
Sutter Instruments
SONY Biotechnology Inc.
Taconic Farms Inc.
Tecan Group US, Inc.
The Jackson Laboratories
Thermo Fisher Scientific
Thor Labs
Tokai Hit
TotalLab Ltd.
Waters Corporation
World Precision Instruments
Zenith Biotech

SEMINARS

INVITED SPEAKERS PROGRAM ("CSHL SEMINAR SERIES")

Each year, Cold Spring Harbor Laboratory invites speakers from outside the institution to present their findings. These weekly seminars keep the CSHL staff current on the latest scientific developments and broaden their perspectives. Graduate students and postdoctoral fellows meet with the seminar speakers for lunch immediately after the seminar, providing an opportunity for the exchange of ideas in an informal setting.

Speaker	Title	Host
January		
Barbara Conradt, Professor, Ludwig Maximilian University of Munich	Deadly dowry—How engulfment pathways promote apoptotic cell death.	Molly Hammell
Alexander Meissner, Harvard Department of Stem Cell and Regenerative Biology, The Broad Institute	Mechanisms of epigenetic regulation in stem cells and development.	Camila dos Santos
V. Narry Kim, Director, RNA Research Center, Institute for Basic Science, Professor, School of Biological Sciences, Seoul National University	MicroRNA processing and regulation.	Leemor Joshua-Tor
February		
Susan McCouch, Department of Plant Breeding and Genetics, Cornell University	Linking genome-wide association studies (GWASs) and genomic selection (GS) to better utilize natural variation in rice.	Dave Jackson
Ann Graybiel, Institute Professor and Faculty Member, Department of Brain and Cognitive Sciences–MIT McGovern Institute of Brain Research	The cortico-striosomal system and decision-making based on value.	WiSE
Joaquin Espinosa, Professor, Department of Pharmacology, University of Colorado Denver School of Medicine	Mechanisms of gene expression control in the p53 network.	Christopher Vakoc
Jan Karlseder, Professor, Donald and Darlene Shiley Chair, The Salk Institute for Biological Studies	Complex interactions between telomeres and the DNA damage response.	David Tuveson
March		
Loren Frank, Ph.D., Investigator, Howard Hughes Medical Institute, and Professor, University of California, San Francisco	Neural substrates of memories and decisions.	Josh Huang
Douglas Wallace, Ph.D., Professor and Director, Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia	A mitochondrial etiology of complex diseases.	Lloyd Trotman
Jonathan Weissman, Professor, Howard Hughes Medical Institute Investigator, University of California, San Francisco	Globally monitoring and controlling protein expression with ribosome profiling and CRISPRi/a.	Han Fang
Gerald Rubin, Ph.D., Vice President, Howard Hughes Medical Institute, and Executive Director, Janelia Research	A molecular geneticist's approach to understanding the fly brain.	Josh Huang
April		
Stavros Lomvardas, Ph.D., Professor, Biochemistry and Molecular Biophysics, Columbia University	Deciphering the mechanisms of monoallelic olfactory receptor expression.	Jessica Tollkuhn

Speaker	Title	Host
September		
Kevin Tracey, M.D., Feinstein Institute for Medical Research, Northwell Health	Reflex circuit mechanisms in immunity.	Tony Zador
October		
Joachim Frank, Ph.D., Professor, Howard Hughes Medical Institute, and Department of Biochemistry and Molecular Biophysics, Columbia University	A quick look at ribosome recycling and translation initiation using time-resolved cryo-electron microscopy.	Gholson Lyon
Grant Jensen, Ph.D., Professor of Biology and Biophysics and Investigator, Howard Hughes Medical Institute, California Institute of Technology	Structural biology in vivo through electron cryotomography.	Leemor Joshua-Tor
November		
David Katz, Professor of Neurosciences and Psychiatry, Case Western Reserve University School of Medicine	Rett syndrome: Crossing the threshold to clinical translation	Nick Tonks
Anirvan Ghosh, Ph.D., Chief Scientific Officer, E-Scape Bio	LRR proteins and the control of synaptic diversity.	Linda Van Aelst
Benjamin Neel, M.D., Ph.D., Professor of Medicine, Laura and Isaac Perlmutter Cancer Center Director, NYU-Langone Medical Center	New tricks for old PTPs.	Richard Sever
December		
John Lis, Barbara McClintock Professor of Molecular Biology and Genetics, Cornell University	Dissecting molecular mechanisms of transcription regulation in vivo using genome-wide strategies.	Jay Lee
Jeffrey Friedman, M.D., Ph.D., Investigator, Howard Hughes Medical Institute, and Marilyn M. Simpson Professor, Laboratory of Molecular Genetics, Rockefeller University	Leptin and the neural control of food intake and metabolism.	Tony Zador
Thomas Rando, M.D., Ph.D., Professor, Stanford University School of Medicine	Molecular regulation of stem cell quiescence activation.	Camila dos Santos

IN-HOUSE SEMINAR PROGRAM

Cold Spring Harbor Laboratory's In-House Seminars were initiated to provide a semiformal avenue for communication among the various research groups at the Laboratory. The seminars also afford a necessary opportunity for graduate students and postgraduate staff to develop their skills in organizing, presenting, and defending their research.

Speaker	Title
January	
Hongwu Zheng	Bad, but could be worse.
February	
Mikala Egeblad	Cancer cells moderate the debate with T cells, neutrophils, and ECM (macrophages are boycotting).
Ivan Iossifov	De novo and rare genetic variants in autism.
March	
Yang Yu	Panoramix: The missing link between piRNA pathway and general silencing machinery.
Chris Hammell	Time management and the making of an animal.
Dannielle Engle	Exploring glycosylation and discovering biomarkers for pancreas cancer.
Justin Kinney	Topics in the measurement and modeling of quantitative sequence–function relationships.
April	
Marcus Stephenson-Jones	Identifying the circuits for outcome evaluation.
Gaofeng Fan	New perspectives of tyrosine phosphorylation–dependent regulation in human disease.
October	
Hiro Furukawa	Obsession with NMDA receptors: Where we stand in the history of neuropharmacology and structural biology.
Sarah Diermeier	Mammary tumor–associated RNAs as novel drivers of tumor cell growth and metastasis.
November	
Lingbo Zhang	Targeting novel self-renewal pathways as therapeutics for hematopoietic malignancies.
December	
Justus Kebschull	Exploiting DNA sequencing technology for high-throughput neuroanatomy.
Benjamin Roche	RNAi is essential for cellular quiescence.
Gholson Lyon	Amino-terminal acetylation of proteins—Why?

BANBURY CENTER

BANBURY CENTER
EXECUTIVE DIRECTOR'S REPORT

I wrote in my first Banbury Center *Annual Report* (1988):

> I have been at the Banbury Center for just over one year, and I am finding it to be as enjoyable and fascinating as I had expected. I was enthusiastic about the aims of the Banbury Center before I arrived, and my experience of the variety of topics and the enthusiasm of participants demonstrates that the Banbury Center is a unique resource for exchanging scientific information.

My experience over the 29 years since then has confirmed in every respect my early expectations. In that time, Banbury has held almost 600 meetings with approximately 12,000 participants. There have been many memorable meetings, but none more significant than *DNA and Forensic Technology*, which helped in establishing the Innocence Project. Others were held at particularly auspicious moments in the development of a field: *The* Arabidopsis *Genome* and *Telomeres* in 1994, *RNA Silencing* in 2000, and many meetings on human genetic and psychiatric disorders. In particular, there were important series of meetings on the fragile X syndrome, prion disease, amyotrophic lateral sclerosis, and Lyme disease. Not all the meetings dealt with research. The Alfred P. Sloan Foundation provided many years of support for workshops introducing science journalists and congressional staff to important contemporary issues in biomedicine, and the Federal Judicial Center provided funding for a similar series of meetings for federal judges.

But, the time has come for me to step down and I am delighted to introduce Rebecca Leshan, who will be taking over from me. Rebecca's background is in molecular and integrative physiology. Her Ph.D. was on hypothalamic leptin receptor expression and action in the mouse with Martin Myers (University of Michigan, Ann Arbor). Rebecca did a postdoc with Donald Pfaff (the Rockefeller University). She decided to use her science background in a different field and joined the United Kingdom government's Science & Innovation Network (UKSIN), based at British Consulate General, Cambridge, Massachusetts. UKSIN promotes U.K. science in the United States, as well as fostering U.S.–U.K. collaborations and informing science policy. Her job involved writing reports and organizing meetings and seminars on a wide variety of science topics—skills that will be put to good use at Banbury.

Jan Witkowski and Rebecca Leshan

I have had a wonderful time as the Banbury director. I would never have learned about as many topics nor met such stellar scientists anywhere else. That the job was so enjoyable was also due to the support provided by many people. Since 2009, Hakon Heimer has provided advice and helped organize meetings on topics in neuroscience and mental health. It has been a pleasure to work with him, and our fruitful collaboration is exemplified by the 2016 meeting on Nordic genomics and healthcare. In the Banbury Center office, Bea Toliver, Janice Tozzo, and Ellie Sidorenko did a wonderful job, and Michelle Corbeaux and Pat Iannotti continue to do

Conversation at the coffee break

so. Katya Davey and Basia Polakowski welcomed participants as hostesses at Robertson House, while Barbara and Jennifer Gordon looked after participants' needs in the dining room. The Audiovisual crew from Meetings and Courses—Herb Parsons, Ed Campodonico, Bill Dickerson, Jonathan Parsons, Ken Orff, and James Whittaker—saw the transition from slides and overheads to digital and coped with all manner of emergencies. Chris McEvoy has acted as watchman as well as groundsman. The departments of Administration, Culinary Services, Facilities, and IT have all contributed to the smooth running of Banbury.

There are three people who have enabled the Banbury Center to flourish. The first, of course, is Charles Robertson, whose generosity in donating the Banbury estate to Cold Spring Harbor Laboratory is the foundation of what we have done. Jim Watson decided to use the estate as a conference center, and his advice to me over all these years has been invaluable. The Center could not have functioned at its high level without his continuing support. I am fortunate that Bruce Stillman shares Jim's enthusiasm for the Banbury program and has continued to support the Center.

2016 in Numbers

The Banbury Center continues to be a busy place. The Conference Room was used for 42 events in 2016, including Banbury Center meetings, courses from the Meetings & Courses Program, the Watson School of Biological Sciences, and CSHL scientists holding retreats at Banbury. There were 574 participants in the Banbury Center meetings, drawn from 30 states, with California, Maryland, Massachusetts, and New York leading the way. The proportion of non-U.S. participants was 18%, coming from 23 countries—a low number that continues to reflect concerns about travel expenses and that may be further affected by U.S. immigration policies. Thirty percent of participants were female, a percentage that has remained steady for a number of years.

Using the chalkboard

Looking at the year's meetings, I can see some of the themes that have run through my tenure at Banbury. The Center has held many meetings on what might be called basic research, although regrettably, these have declined recently because of funding difficulties. An example from 2016 was the meeting organized by Hyman Hartman and Temple Smith. Fifty years ago, the 31st Cold Spring Harbor Symposium on Quantitative Biology was devoted to *The Genetic Code*. It was, as Crick wrote in the Symposium volume, an historic occasion. However, he was less sanguine about progress in understanding the structure of the genetic code and its origin. Crick feared that they were heading for "a very unhealthy situation, in that theory will run far ahead of useful experimental facts," as had studies on the genetic code in the 1950s. What was needed was "some way of obtaining more experimental evidence." Crick's wish for more experimental evidence has been amply fulfilled in the 50 years since that historic occasion. Participants in the 2016 Banbury meeting on *Evolution of the Translational Apparatus: Implications for the Origin and History of the Genetic Code* had a wealth of data derived from new findings on the ribosome and the aminoacyl tRNA synthetases.

Banbury has also played a role in helping the development of new fields of research. Previous examples that come to mind are DNA bar coding and RNAi. Similarly, in 2016, *Ancient DNA and Archaeology*, generously supported by the Lehrman Institute, was held to encourage interactions between archaeologists, historians, ancient DNA specialists, and geneticists. One goal was to identify questions, regions, and time periods in which DNA studies would be particularly likely to yield insights not possible with other methods. It did not altogether succeed, but analogies were drawn with the introduction of radiocarbon dating, initially the province of just a few experts. Just as the "radiocarbon revolution" provided archaeologists with an accurate timescale for the past, the "ancient DNA revolution" has the potential to show how human remains—together with their archaeological contexts—relate to present and ancient populations.

Banbury has contributed to plant science by holding annual meetings funded by the contributions of Monsanto and Pioneer Hi-Bred over the years to the Laboratory's Corporate Sponsor

Program, with contributions from other companies. The meetings have covered a wide variety of topics, including plant genetics, genomics, and physiology. Crop breeders are hampered by the long time needed for traditional breeding, and the 2016 meeting on *Genomics-Enabled Accelerated Crop Breeding* examined how the new methods for manipulating genes can be used to accelerate the process. These new methods include TALENs and the CRISPR-Cas9 system. Participants reviewed the application of these techniques to a variety of crops, including cassava, lettuce, soybean, and tomato. The meeting was especially timely as the U.S. Department of Agriculture had recently declared that mushrooms modified using CRISPR-Cas9 would not be considered to be genetically modified organisms.

I was fortunate to arrive at Banbury at a time when mapping human disease genes was under way, and the human genome project was about to begin. My background was in human molecular genetics, and so I began a series of meetings in that field. Psychiatric genetics was an early topic; in the first flush of enthusiasm for restriction-fragment-length polymorphism (RFLP)-linkage analysis, it was hoped that genes involved with psychiatric disorders would soon be identified. The first of the meetings was *Genetic Approaches to Schizophrenia*, and others followed through the 1990s—but it became clear that psychiatric disorders were a very difficult problem. Nevertheless, progress has been made and genetic counseling is important and useful in psychiatry, even without full knowledge of all the genes involved. The meeting *Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era* was held to consider two questions: How can the ever-increasing understanding of the genetics of psychiatric disorders be translated into interventions that improve outcomes for patients and their families? What needs to be done to prepare for the time when exome or even whole-genome sequencing becomes the norm?

Meetings on cancer have been a feature of the Banbury Center from the early meetings on chemical carcinogenesis and environmental hazards. However, it is rather surprising, given the Laboratory's intensive research on DNA tumor viruses throughout the 1970s, that the first Banbury meeting on the molecular genetics of cancer was not held until 1984 (*SV40 Large-T Antigen*). Since then, the application of ever-more-sophisticated techniques has led to ever-increasing knowledge about the nature of cancer. In recent years, there has been an increasing emphasis on metabolic changes in tumor cells, focusing on redox pathways and targeting reactive oxygen species (ROS). This has been shown to be an effective strategy, and participants in the meeting *Making Oxidative Chemotherapy Less Toxic* reviewed the evidence for whether combination therapies may be more effective at killing cancer cells and less damaging to noncancer cells.

The first meeting of 2016 was *After UKCTOCS: Public Messaging on Screening and Early Detection for Ovarian Cancer.* UKCTOCS is the United Kingdom Collaborative Trial of Ovarian Cancer Screening study, which followed more than 200,000 women. The meeting was organized by the U.S. Ovarian Cancer Research Fund Alliance, and participants discussed the implications of the UKCTOCS findings for advising women on the use of multimodal screening for early detection of ovarian cancer. One of the questions posed to the participants was, "Do we believe that the UKCTOCS data are sufficiently strong to support a recommendation for population screening for ovarian cancer?" Three questions followed, depending on the answer:

1. If yes, should the recommendation be limited to the population studied in UKCTOCS or expanded or narrowed?
2. If no, should there be an alternative recommendation? What should it be?
3. If it's uncertain, what do we tell people?

The conclusion, published as an editorial in the journal *American Family Physician*, was the answer to question 3, that using multimodal screening was not yet justified.

If genomic medicine is to be successful, we need to know how genetic variants affect an individual's health. This requires sequencing tens of thousands of individuals and then examining

their medical records to correlate their health with their genetic variants. The former, although still not trivial, is no longer an obstacle—but relating genetic variants to health is a huge undertaking, requiring comprehensive and accurate medical records. Fortunately, the health systems of the Nordic countries have such records, and the title of the meeting *Studying the Genomic Variation That Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems* captures the essence of the discussions. Participants discussed how the countries could work together and some of the challenges—scientific, medical, and legal—to doing so. The meeting was a great success and led to an influential report and follow-up meetings.

These meetings exemplify the type of meeting for which Banbury is particularly suitable. The Banbury setting is ideal for intense discussions of perhaps controversial topics, discussions of a form that would not be possible in larger, public meetings. Other notable meetings in this category include those on scientific fraud, DNA fingerprinting, public mistrust of immunization, and end-of-life issues.

Acknowledgments

I have already thanked those who have helped me over the years, but there is every reason to repeat myself for those who worked hard in 2016 to keep Banbury running. Michelle Corbeau and Pat Iannotti hold the fort in the Banbury Center office, while Basia Polakowski continues to welcome and look after participants in Robertson House. Participants never fail to comment on the beauty of the estate, a tribute to the hard work of Jose Covera, Joe McCoy, and Saul Covera. Culinary Services, Facilities, and the Meetings Office play key roles in the operation of the Center. The meetings would not be the success they are without the contributions of organizers and participants, the generosity of the Laboratory's Corporate Sponsors and the other donors who fund our meetings, and the Laboratory's scientists who continue to support the Center.

Jan A. Witkowski
Executive Director

1988 2017

Jan Witkowski, 30 years at Banbury, 1987–2017

BANBURY CENTER MEETINGS

Date	*Title*	*Organizer(s)*
February 7–9	After UKCTOCS: Public Messaging on Screening and Early Detection for Ovarian Cancer	J. Boyd, S. DeFeo, A. Moran, M. Seiden
February 16–19	Studying the Genomic Variation That Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems	O. Andreassen, N. Freimer, L. Groop, H. Heimer, A. Palotie
February 26–March 2	Communicating Science	S. Schedler, C. Walther
March 6–9	Autophagy and Cancer	R. Amaravadi, A. Kimmelman, E. White
March 15–18	STAT3 in Cancer: How Can It be Inhibited?	J. Darnell, D. Levy, G. Stark
April 29–May 1	NIMH Brain Camp VIII	J. Chung, B. Cuthbert
May 15–17	Ancient DNA and Archaeology	T. Higham, S. Pääbo, D. Reich
July 5–8	Measuring and Modeling Quantitative Sequence–Function Relationships	D. Fowler, J.B. Kinney, A. Siepel
August 21–23	Can We Make Animal Models of Human Mental Illness? A Critical Review	H. Heimer, R. McCombie, E. Nestler
September 6–8	Mammalian Brain Cell Diversity and Census	A. Beckel-Mitchener, J. Huang
September 11–14	Making Oxidative Chemotherapy Less Toxic	A. Holmgren, R. Maki, D. Tuveson
September 18–21	Diagnostic Tests for Lyme Disease: A Reassessment	J. Branda, S. Schutzer
October 16–19	Genomics-Enabled Accelerated Crop Breeding	B. Staskawicz, D. Voytas
November 1	The Lustgarten Foundation: Vitamin D Day	R. Evans, P. Sharp, D. Tuveson
November 9–11	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	M. Brivanlou, H. Sauer, K. Sonnenfeld
November 13–16	Evolution of the Translational Apparatus and Implications for the Origin of the Genetic Code	H. Hartman, T. Smith
November 30–December 2	Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era	J. Austin, F. McMahon
December 4–7	Evolution and Revolution in Anatomic Pathology: Automation, Machine-Assisted Diagnostics, Molecular Prognostics, and Theranostics	J.M. Crawford, P. Mitra, M. Wigler
December 11–14	Developing Gene Editing as Therapeutic Strategy	C. Gersbach, J.K. Joung, A. Wagers

BANBURY CENTER MEETINGS

After UKCTOCS: Public Messaging on Screening and Early Detection for Ovarian Cancer

February 7–9

FUNDED BY **Ovarian Cancer Research Fund Alliance**

ARRANGED BY **J. Boyd,** Florida International University, Miami
S. DeFeo, Ovarian Cancer Research Fund Alliance, New York
A. Moran, Ovarian Cancer Research Fund Alliance, New York
M. Seiden, McKesson Specialty Health, The Woodlands, Texas

The United Kingdom Collaborative Trial of Ovarian Cancer Screening study (UKCTOCS) was designed to provide firm data that can be used as the basis for assessing the value of current methods of early detection of ovarian cancer. The findings were published online in *The Lancet,* December 2015, and will need to be communicated to patients, physicians, and payers—and a discussion about access and reimbursement will need to take place. The goals of this meeting were to review the findings of the UKCTOCS trial and discuss what recommendations Ovarian Cancer Research Fund Alliance might make to its constituency.

S. Skates, B. Levin, C. Chiuzan

Welcoming Remarks and Background: J.A. Witkowski, Cold Spring Harbor Laboratory, and **A. Moran,** Ovarian Cancer Research Fund Alliance, New York

Introduction and Background: J. Boyd, Florida International University, Miami
M. Seiden, McKesson Specialty Health, The Woodlands, Texas

SESSION 1: Addressing the Questions

Chairperson: I. Jacobs, University of New South Wales, Sydney, Australia

U. Menon, University College London, United Kingdom; and S. Skates, Massachusetts General Hospital, Boston: Update on the UKCTOCS trial.

S. Narod, University of Toronto, Canada: The UKCTOCS trial: A closer look.

B. Levin and C. Chiuzan, Columbia University, New York: UKCTOCS: Biostatistical perspectives.

Panel Discussion: Communicating Controversial Public Health Issues to Lay and Medical Audiences

Facilitator: M. Seiden, McKesson Specialty Health, Woodlands, Texas

Panel

M. Ebell, University of Georgia, Athens
A. Ellis, Ovarian Cancer Survivor, White Plains, New York
M. Eiken, Society of Gynecologic Oncology, International Gynecologic Cancer Society, Chicago, Illinois
R. Smith, American Cancer Society, Inc., Atlanta, Georgia
A. Moran, Ovarian Cancer Research Alliance, New York
C. Balas, Ovarian Cancer Research Fund Alliance, New York

Group Consideration of Three Major Questions

1. Do we believe that the UKCTOCS trial results prove that screening prevents deaths from ovarian cancer?

D. Barley

K. Gavin

2. Do we believe that the UKCTOCS data are sufficiently strong to support a recommendation for population screening for ovarian cancer?
3. If we support screening of a population of healthy women, do we support use of the ROCA-based algorithm and the use of Abcodia as the only legitimate screening strategy?

Continued Group Discussion

Group 1: Lay Community
Group 2: Medical Community

SESSION 2: Preparing the Communications

Studying the Genomic Variation That Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems

February 16–19

FUNDED BY **The Norwegian Research Council and NordForsk**

ARRANGED BY **O. Andreassen,** University of Oslo, Norway
N. Freimer, University of California, Los Angeles
L. Groop, Lund University, Malmo, Sweden
H. Heimer, Cold Spring Harbor Laboratory
A. Palotie, Broad Institute, Cambridge, Massachusetts

This meeting examined the possible coordination of medical genomics across Nordic countries, and how best to engage the health systems of the Nordic countries in such cross-national efforts. The meeting brought together data managers, clinical leaders, policy makers, and researchers from the Nordic countries; international researchers who collaborate on projects using Nordic health records; electronic records and data privacy experts, and representatives of funding agencies. The meeting explored critical questions about the potential of combining data sets that include as many as 20 million health records. In addition to scientific questions regarding study design, participants considered legal, ethical, and logistic issues concerning cross-national and international use of health records.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Origin of the Meeting: H. Heimer, Cold Spring Harbor Laboratory

C. Stoltenberg, P. Njølstad

K. Hveem, A. Palotie

SESSION 1: Opportunities in Genetics/Genomics

Chairperson: O. Andreassen, University of Oslo, Norway

N. Freimer, University of California, Los Angeles: Deep phenotype data obtained over decades: Can genomics turn past investments into future health?

A. Palotie, Institute for Molecular Medicine, Helsinki, Finland: Special features of Finland for developing genome medicine.

A. Metspalu, University of Tartu, Estonia: Deep sequencing of the Estonian population sample of 2400 subjects—first results.

K. Hveem, Norwegian University of Science and Technology, Levanger, Norway: Perspective from the Nordic cohorts and registries.

A. Addington, National Institute of Mental Health, Rockville, Maryland: Perspective from the NIH.

C. Fox, Merck & Co. Inc., Boston, Massachusetts: Perspective from pharma.

SESSION 2: Opportunities in Data Mining and Informatics

Chairperson: A. Palotie, Institute for Molecular Medicine, Helsinki, Finland

J. Kaprio, University of Helsinki, Finland: Twin registries as a resource in the Nordic countries.

C. Stoltenberg, Norwegian Institute of Public Health, Olso, Norway: MOBA—The Norwegian mother, father, and child cohort: A prospective, population-based health study with nearly 300,000 participants.

S. Brunak, University of Copenhagen, Denmark: Disease trajectories and time-ordered co-morbidities.

E. Hovig, Oslo University Hospital, Norway: Perspective from the Nordic cohorts and registries.

J. Palmgren, Karolinska Institutet, Stockholm, Sweden: Infrastructure for data: Integrating health, lifestyle, and molecular information.

J. Larkin, National Institutes of Health, Bethesda, Maryland: Perspective from the NIH.

M. Sogaard, Pfizer, Inc., New York: Perspective from pharma.

SESSION 3: Opportunities for Clinical and Translational Application of Genetics and Informatics

Chairperson: J. Kaprio, University of Helsinki, Finland

L. Groop, Lund University, Malmö, Sweden: Toward precision medicine in diabetes.

O. Andreassen, University of Oslo, Norway: Neuropsychiatric disorders: Opportunities for prediction and stratification.

N. Stitziel, Washington University School of Medicine, St. Louis, Missouri: Identifying and validating therapeutic targets for cardiovascular disease.

P. Njølstad, University of Bergen, Norway: Monogenic disease: Beacons for identifying therapy-relevant novel causes of complex disorders.

C. Jaquish, National Heart, Lung, and Blood Institute, Bethesda, Maryland: NHLBI precision medicine/whole-genome

A. Jalanko

sequencing program: NHLBI TOPMed (trans-omics for precision medicine).

Q. Li, Janssen Research & Development, LLC, Raritan, New Jersey: Perspective from pharma.

SESSION 4: Breakout Groups: Proposals for Opportunities for Collaborations Across Nordic Countries, between Nordic Countries, and U.S. Investigators and Funding Agencies; Public Private Partnerships

Group 1: Genetics/Genomics, M. Daly, Leader
Group 2: Data Mining/Informatics, E. Hovig, Leader
Group 3: Clinical/Translational, J. Dillner, Leader

SESSION 5: Presentation of Reports of Breakout Groups and Discussion

Chairperson: L. Groop, Lund University, Malmö, Sweden

Group 1: Genetics/Genomics. M. Daly, Leader
Group 2: Data Mining/Informatics, E. Hovig, Leader
Group 3: Clinical/Translational, J. Dillner, Leader

SESSION 6: Developing an Outline for a Position Paper and Discussion of Next Steps

Communicating Science

February 26–March 2

FUNDED BY **Boehringer Ingelheim Foundation for Basic Research in Medicine**

ARRANGED BY **S. Schedler,** Boehringer Ingelheim Fonds, Mainz, Germany
C. Walther, Boehringer Ingelheim Fonds, Mainz, Germany

The Boehringer Ingelheim Fonds has an international program of support for Ph.D. fellowships, and it first brought its fellows to the Banbury Center for their annual North American retreat in 2005. It has been a great pleasure to have them return, and their 2016 stay at Banbury was the ninth occasion that they have been here. At Banbury, the fellows receive intensive instruction in matters such as giving presentations and writing papers—topics usually learned by default, and often poorly, during graduate research.

Introductory Remarks: J.A. Witkowski, Cold Spring Harbor Laboratory

Opening Remarks and All About BIF—Part 1: C. Walther, Boehringer Ingelheim Foundation, Mainz, Germany

K. Achenbach, Boehringer Ingelheim Foundation, Mainz, Germany: Communication: Why and how?

N. LeBrasseur, DNA Medical Communications, New York: Writing techniques and how to structure papers; Writing assignment 1.

B. Tansey, Vanderbilt University, Nashville, Tennessee: Preparing and delivering a scientific talk; PowerPoint presentations, videotaped with replay and feedback.

N. LeBrasseur, DNA Medical Communications, New York: Discussion of writing assignment 1; Writing assignment 2.

N. LeBrasseur, DNA Medical Communications, New York: Image manipulation: Dos and don'ts—A short intro; Return and discussion of writing assignment 2.

J. Carlos Lopez, Roche Innovation Center: Career talk.

K. Ris-Vicari, Katie Ris-Vicari Graphic Design, Bethpage, New York: How to design figures.

C. Walther, Boehringer Ingelheim Foundation, Mainz, Germany: All about BIF: Part 2 and feedback.

Autophagy and Cancer

March 6–9

FUNDED BY **Astellas Pharma Inc., Millennium Pharmaceuticals, Inc., Merck Serono, Novartis, and Presage Biosciences, Inc.**

ARRANGED BY **R. Amaravadi,** University of Pennsylvania, Philadelphia
A. Kimmelman, Dana-Farber Cancer Institute, Boston, Massachusetts
E. White, Rutgers Cancer Institute of New Jersey, New Brunswick

Autophagy is a process of cellular self-cannibalization that captures intracellular proteins and organelles and degrades them in lysosomes. Autophagy plays a critical role in human disease, including cancer, and there is evidence that autophagy can be either a tumor suppression or promotion mechanism. There remain many important unanswered questions on the role of autophagy in cancer, and participants in the meeting focused on five: the role of autophagy in tumors; autophagy and nutrient sensing signaling; autophagy and metabolism; selective autophagy and nonmacroautophagy mechanisms; and translational/clinical aspects of autophagy modulation.

Introduction: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Goals and Objectives: E. White, Rutgers Cancer Institute of New Jersey, New Brunswick

K. Ryan

M. Lotze

SESSION 1: Role of Autophagy in Tumors

Chairperson: A. Kimmelman, Dana-Farber Cancer Institute, Boston, Massachusetts

J. Debnath, University of California, San Francisco: Autophagy in mouse models of breast cancer.
K. Ryan, Beatson Institute, Glasgow, United Kingdom: Pancreatic cancer autophagy.
N. Roy D'Amore, Takeda Oncology, Cambridge, Massachusetts: Atg7 and Vps34 inhibitors.
J. Moscat, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California: Control of p62 homeostasis by autophagy in cancer.

SESSION 2: Autophagy and Nutrient Sensing Signaling

Chairperson: K. Ryan, Beatson Institute, Glasgow, United Kingdom

E. White, Rutgers Cancer Institute of New Jersey, New Brunswick: Metabolic control of p53 by autophagy.
N. Cosford, Sanford Burnham Medical Research Institute, La Jolla, California: Ulk1 inhibitor development for cancer therapy.
N. Bardeesy, Massachusetts General Hospital, Boston: Lkb1 and nutrient sensing.
J.-L. Guan, University of Cincinnati, Ohio: The role of autophagy in cancer stem cells.
L. Shawver, Cleave Biosciences Burlingame, California: ERAD, UPR, and autophagy; experience with inhibitors of p97.

SESSION 3: Autophagy and Metabolism

Chairperson: E. White, Rutgers Cancer Institute of New Jersey, New Brunswick

A. Kimmelman, Dana-Farber Cancer Institute, Boston, Massachusetts: Ras and pancreatic cancer metabolism.
Y. Guo, Rutgers Cancer Institute of New Jersey, New Brunswick: Autophagy in lung cancer metabolism.
R. Perera, University of California, San Francisco: Transcriptional control of autophagy.
W. Harper, Harvard Medical School, Cambridge, Massachusetts: Autophagy networks.
M. Lotze, University of Pittsburgh, Pennsylvania: Autophagy, HMGB1/RAGE, Ras, and the tumor immune response.
R. Klinghoffer, Presage Biosciences, Seattle, Washington: Not all lysosomal inhibitors are created equal: Direct comparison of antitumor effects in canine sarcoma patients.

SESSION 4: Selective Autophagy and Nonmacroautophagy Mechanisms

Chairperson: J. Debnath, University of California, San Francisco

R. Youle, National Institutes of Health, Bethesda, Maryland: Mechanism of mitophagy.
K. Macleod, University of Chicago, Illinois: Autophagy promotes focal adhesion disassembly and cell motility of metastatic tumor cells through direct interaction of paxillin with LC3.
C. Dorsey, Eli Lilly, Indianapolis, Indiana: Targeting autophagy.
X. Jiang, Memorial Sloan Kettering Cancer Center, New York: Autophagy regulation.
J. Martinsson, Sprint Bioscience, Stockholm, Sweden: Development of selective Vps34 inhibitors.

SESSION 5: Translational/Clinical Aspects of Autophagy Modulation

Chairperson: R. Perera, University of California, San Francisco

R. Amaravadi, University of Pennsylvania, Pennsylvania: Clinical modulation of autophagy with HCQ.
J. Mehnert, Rutgers Cancer Institute of New Jersey, New Brunswick: Autophagy in melanoma.
A. Thorburn, University of Colorado, Denver: Autophagy in brain cancer.
V. Kirkin, Merck Serono, Darmstadt, Germany: How do we fill the current gap in validation of the concept of targeting autophagy in cancer?
J. Goodwin, Novartis, Cambridge, Massachusetts: Is autophagy a therapeutic target in cancer?

General Discussion and Closing Remarks

STAT3 in Cancer: How Can It Be Inhibited?

March 15–18

FUNDED BY **Boston Biomedical, Inc.**

ARRANGED BY **J. Darnell,** The Rockefeller University, New York
D. Levy, New York University School of Medicine, New York
G. Stark, Cleveland Clinic Foundation, Ohio

This meeting brought together an international group of researchers to review what is known of STAT3, its potential as a target in cancer, and what progress has been made in developing therapies. Discussions included talks on current research on the biology of STAT3 (e.g., cancer stem cells and STAT3; mitochondrial role of STAT3; STAT3 as a tumor suppressor) and current understanding of new and previously recognized targets for STAT3 inhibition. There were reports on the use of the newer anti-STAT3 compounds.

Welcoming Remarks: J.A. Witkowski, Banbury Center, and **J. Watson,** Cold Spring Harbor Laboratory

SESSION 1: Cancer Stem Cells and Natural STAT inhibitors

Chairperson: J. Darnell, The Rockefeller University, New York

I. Marie, New York University Medical Center, New York: What do loss-of-function studies teach us about the physiologic role of STAT3?

J. Rich, Lerner Research Institute, Cleveland, Ohio: STAT3 in brain tumor stem cells.

M. Venere, Ohio State University, Columbus: Converging on NF-κB to target cancer stem cells.

H. Rogoff and A. Yang, Boston Biomedical, Inc., Cambridge, Massachusetts: Targeting cancer stemness through blocking STAT3.

K. Shuai, University of California, Los Angeles: Mechanisms to inhibit STAT signaling through PIAS proteins.

G. Stark, K. Struhl

D. Levy, J. Bromberg, J. Turkson

J. Babon, Walter and Eliza Hall Institute of Medical Research, Victoria, Australia: Mimicking the action of SOCS3: A potent physiological inhibitor of STAT3 signaling.

SESSION 2: Mitochondrial Connection, Genetics, and Mutations

Chairperson: G. Stark, Cleveland Clinic Foundation, Ohio

T. Benveniste, University of Alabama, Birmingham: The role of CK2 and STAT3 in cancer: Impact on the tumor microenvironment.

G. Inghirami, Weill Cornell Medical College, New York: Activating mutations of the JAK/STAT3 pathway and T-cell transformation.

D. Levy, New York University School of Medicine, New York: The mitochondrial role of STAT3 in cancer.

J. Milner, National Institute of Allergy & Infectious Diseases, Bethesda, Maryland: Monogenic disorders due to germline mutations in STAT3 and other STATs: Lessons learned from gain, loss, and cross-talk.

J. Hart, Scripps Research Institute, La Jolla, California: Non-kinase inhibitors of STAT3.

C. Mertens, The Rockefeller University, New York: Mutations in the linker domain affect phospho STAT3 function and suggest targets for interrupting STAT3 activity.

SESSION 3: Pharmacologic Inhibitors: Old and New

Chairpersons: D. Levy, New York University School of Medicine, New York, and **J. Grandis**, University of California, San Francisco

J. McMurray, M.D. Anderson Cancer Center, Houston, Texas: Inhibition of STAT6 blocks aberrant Th2 signaling in allergic asthma.

J. Turkson, University of Hawaii Cancer Center, Honolulu: Targeting JAK/STAT signaling pathways for cancer therapy.

J. Grandis, University of California, San Francisco: A decoy oligonucleotide approach to STAT3 inhibition.

C. Li, Boston Biomedical Inc., Cambridge, Massachusetts: Clinical development of napabucasin (BB608), a first-in-class cancer stemness inhibitor that works by blocking STAT3.

C. Catapano, Institute of Oncology Research, Bellinzona, Switzerland: How to kill a cancer cell: Insights from novel small-molecule inhibitors of STAT3.

D. Placantonakis, New York University School of Medicine, New York: The overlap in basic ideas in management of glioblastoma multiforme.

P. McCoon, AstraZeneca Pharmaceuticals, Waltham, Massachusetts: Clinical biomarkers of a STAT3 antisense oligonucleotide, AZD9150, suggest an immune-modulatory role in tumors.

Y. Kanno, NIAMS, National Institutes of Health, Bethesda, Maryland: Targeting cytokine signaling by Jakinibs to control genomic switches.

D. Frank, Dana-Farber Cancer Institute, Boston, Massachusetts: Targeting the transcriptional function of STAT3: From the lab to clinical trials.

M. Jackson, Case Western Reserve University, Cleveland, Ohio: Cooperative STAT3-SMAD3 signaling drives cancer cell plasticity.

K. Struhl, Harvard Medical School, Boston, Massachusetts: Role of STAT3 and STAT3-mediated transcriptional regulatory circuits in cancer.

T. Miller, IC-MedTech, Las Vegas, Nevada: Can ROS prevent STAT3 phosphorylation?

SESSION 4: Caution and Discussion

Chairperson: J. Darnell, The Rockefeller University, New York

G. Stark, Cleveland Clinic Foundation, Ohio: Modulation of STAT3-dependent signaling by EGFR, induction of U-STAT3, and lysine methylation.

J. Bromberg, Memorial Sloan Kettering Cancer Center, New York: Targeting JAK and Stat 3 in solid tumors: Clinical and preclinical observations.

L. Kenner, Ludwig Boltzmann Institute Cancer, Vienna, Austria: IL-6/Stat3 in diagnosis and treatment of prostate cancer.

SESSION 5: General Discussion and Summary of Key Points for Further Research

J. Darnell, The Rockefeller University, New York: Thoughts and questions on the role of STAT3 in cancer.

I. Marie

J. Darnell, U. Vinkemeier

NIMH Brain Camp VIII

April 29–May 1

FUNDED BY **National Institute of Mental Health, NIH**

ARRANGED BY **J. Chung,** National Institute of Mental Health, Bethesda, Maryland
B. Cuthbert, National Institute of Mental Health, Bethesda, Maryland

Cold Spring Harbor Laboratory is renowned worldwide for its education programs, from high school level to the highest professional levels. One of the Banbury Center's contributions is to host the NIMH-sponsored "Brain Camp." The goal of the Brain Camp is to identify areas of neuroscience that are of interest and relevance to psychiatrists and to communicate these to a small group of outstanding psychiatry residents and research fellows. Some of the most distinguished and thoughtful neuroscientists in the country came as guest speakers to the meeting. The goal of the series of meetings is to develop a neuroscience curriculum that can eventually be shared with psychiatry training programs around the country.

SESSION 1

B. Cuthbert, National Institute of Mental Health, Bethesda, Maryland: Welcome and Introductions.
M. Pao, National Institute of Mental Health, Bethesda, Maryland: Brief history of the NIMH camp.
S. Hollingsworth Lisanby, National Institute of Mental Health, Bethesda, Maryland: From discovery to recovery: Transforming the practice of psychiatry through translation.

SESSION 2

A. Raznahan, National Institute of Mental Health, Bethesda, Maryland: A genetics-first approach to parsing the biology of neurodevelopmental disorders.
A. Molofsky, University of California, San Francisco: Psychiatric diseases from a glial cell perspective.
A. Etkin, Stanford University, California: Neural circuits as substrates of mental illness and targets for therapeutics.

C. Tamminga, University of Texas Southwestern Medical Center, Dallas: Psychosis as a learning and memory disorder: A dimensional approach.

J. Conn, Vanderbilt University, Nashville, Tennessee: Allosteric modulators of muscarinic acetylcholine receptors as a novel approach for treatment of schizophrenia.

SESSION 3

B. Stevens, Harvard University, Boston, Massachusetts: Pruning synaptic circuits: New mechanisms and implications in neuropsychiatric disorders.

Ancient DNA and Archaeology

May 15–17

FUNDED BY **Lehrman Institute, New York**

ARRANGED BY **T. Higham,** University of Oxford, United Kingdom
S. Pääbo, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany
D. Reich, Harvard Medical School, Boston, Massachusetts

Analysis of DNA extracted from ancient remains is transforming studies of the origins of modern humans, human migrations, and history, as well as related fields such as the domestication of plants and animals. The power of ancient DNA to supplement paleontological and archaeological studies may in some respects be compared with radiocarbon dating. Just as the "radiocarbon revolution" provided archaeologists with an accurate timescale for the past, the "ancient DNA revolution" has the potential to show how human remains—and by extension their archaeological contexts—relate to present and ancient populations. To realize this potential, it will be necessary over the next ten years to make DNA technologies readily accessible to archaeologists. This meeting was held to discuss how this can best be achieved.

Welcoming Remarks and Background: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

E. Banffy

S. Pääbo, N. Patterson

SESSION 1: The Big Questions

S. Pääbo, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany: The current state of ancient DNA: What can be done and what can't.

K. Kristiansen, University of Gothenburg, Göteberg, Sweden: What can archaeology contribute to genetics and vice versa?

Discussion

Moderators: T. Higham, University of Oxford, Oxford, United Kingdom, and **D. Reich**, Harvard Medical School, Boston, Massachusetts: What can we learn with archaeogenetics? What are the big questions?

SESSION 2: Moore's Law of Ancient DNA: The 2012–2015 Methods Revolution

M. Meyer, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany: Going deep in time: Denisova and Sima.

R. Pinhasi, University College Dublin, Ireland: Leveraging osteology, histology, and anatomy to optimize yields.

D. Reich, Harvard Medical School, Boston, Massachusetts: Industrial scale ancient DNA.

Discussion

Moderators: J. Krause, Max-Planck-Institute for the Science of Human History, Jena, Germany, and **D. Bradley**, Trinity College Dublin, Dublin, Ireland

SESSION 3: How Can Geneticists Provide Useful Information to Archeologists?

S. Pääbo, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany: What could a service facility for DNA look like?

T. Higham, University of Oxford, United Kingdom: How does a state-of-the-art radiocarbon service facility work?

J. Mountain, 23andMe, Mountain View, California: How does 23andMe make genetic results comprehensible?

Discussion

Moderators: D. Reich, Harvard Medical School, Boston, Massachusetts, and **D. Meltzer,** Southern Methodist University, Dallas, Texas: How can geneticists make ancient DNA an accessible tool for archaeologists? What can genetics provide to archaeologists on a routine basis? What is archaeologically useful? What would a useful report look like? Is there a way for archaeologists and geneticists to collaborate better at the outset?

SESSION 4: How Can Archaeologists Distinguish between What's Solid and Not?

M. Meyer, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany: The cases of Hoyo Negro and the early Neolithic British wheat.

D. Reich, Harvard Medical School, Boston, Massachusetts: Why many published ancient DNA findings are false.

Discussion

Moderators: D. Meltzer, Southern Methodist University, Dallas, Texas, **D. Anthony**, Hartwick College, Oneonta, New York, and **K. Kristiansen**, University of Gothenburg, Göteberg, Sweden: How can archaeologists distinguish between what's solid and not? What are the questions archaeologists should routinely ask of geneticists in regard to their analyses and results?

SESSION 5: Challenge Areas: A Genetics Perspective

J. Krause, Max-Planck-Institute for the Science of Human History, Jena, Germany: Ancient pathogen genomics.

B. Shapiro

B. Shapiro, University of California, Santa Cruz: Learning about human history using megafaunal ancient DNA.

G. Larson, University of Oxford, United Kingdom: Domestication.

Discussant-Led Conversation: N. Boivin, University of Oxford, United Kingdom, **D. Fuller,** University College London, United Kingdom, and **N. Patterson,** Broad Institute, Cambridge, Massachusetts: What opportunities exist for collaboration between geneticists and archaeologists?

SESSION 6: How Can Archeologists Help Geneticists?

Moderators: E. Banffy, German Archaeological Institute, Frankfurt, Germany, **D. Anthony,** Hartwick College, Oneonta, New York, and **D. Meltzer,** Southern Methodist University, Dallas, Texas.

Topics for Consideration:

How can we improve recovery from challenging areas?

How do we address issues of sampling, conservation, and destruction of material (e.g., petrous bones)?

Can archaeologists reduce contamination on site during excavation?

How can we best frame archaeological problems in a manner testable with DNA?

Pots aren't people/climate is not a deus ex machina: How can archaeologists help geneticists avoid archaeologically naïve interpretations and explanations?

SESSION 7: How Should We Overcome the Barriers between Disciplines?

Topics for Consideration: How should we overcome the barriers between disciplines (e.g., lack of comparable training, cross-disciplinary comprehension, few common journals, and sparse interaction) and improve communication?

What steps can be taken to make interactions between archaeologists and geneticists more productive?

Should we initiate a joint community project?

Moderators: T. Higham, University of Oxford, United Kingdom, and **D. Reich**, Harvard Medical School, Boston, Massachusetts: List and review ideas of all groups. Develop an outline for a white paper or opinion piece.

Final Comments

Measuring and Modeling Quantitative Sequence–Function Relationships

July 5–8

FUNDED BY **The Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory**

ARRANGED BY **D. Fowler,** University of Washington, Seattle, Washington
J.B. Kinney, Cold Spring Harbor Laboratory
A. Siepel, Cold Spring Harbor Laboratory

Understanding how DNA sequence relates to function is a fundamental problem in biology that is becoming increasingly acute as more organisms and individuals are sequenced. A variety of massively parallel assays now make it possible to measure sequence–function relationships with unprecedented resolution and quantitative precision. At the same time, advances in our theoretical understanding of sequence–function relationships have resulted in increasingly accurate models. However, current efforts are scattered across multiple disciplines, including gene regulation, protein science, and evolution. This meeting gathered leading experimentalists and theorists to discuss unifying disparate approaches for studying quantitative sequence–function relationships and to delineate important outstanding problems in this emerging area of biology.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Open Discussion I: Introduction and Meeting Goals
Chairperson: J. Kinney, Cold Spring Harbor Laboratory

J. Kinney, A. DePace

A. Keating, J. Thornton, M. Bulyk

SESSION 1

Chairperson: G. Stormo, Washington University School of Medicine, St. Louis, Missouri

M. White, Washington University School of Medicine, St. Louis, Missouri: The not-so-simple consequence of a simple *cis*-regulatory grammar.

M. Bulyk, Harvard Medical School, Boston, Massachusetts: Survey of variation in human transcription factors reveals prevalent DNA-binding changes.

R. Gordân, Duke University School of Medicine, Durham, North Carolina: Quantitative TF-DNA-binding models explain a large fraction of gene expression variation.

M. Maurano, NYU Institute for Systems Genetics, New York: Decoding human regulatory variation: Pinpointing trait associations and functional noncoding variant.

R. Rohs, University of Southern California, Los Angeles: Quantitative modeling of TF-DNA binding: Beyond DNA shape toward biophysical features.

SESSION 2

Chairperson: B. Frey, University of Toronto, Canada

G. Stormo, Washington University School of Medicine, St. Louis, Missouri: Transcription factor specificity and cooperativity.

H. Bussemaker, Columbia University, New York: Learning protein–DNA recognition models from sparse sequencing data.

J. Kinney, Cold Spring Harbor Laboratory: Quantitative modeling of sequence–function relationships.

SESSION 3

Chairperson: A. DePace, Harvard Medical School, Boston, Massachusetts

B. Frey, University of Toronto, Canada: Bridging the genotype–phenotype gap using quantitative sequence-to-molecular phenotype models.

R. Das, Stanford University School of Medicine, California: Testing computational models of RNA structure/function.

M. Noyes, NYU Institute for Systems Genetics, New York: Capturing the low end of affinity.

SESSION 4

Chairperson: S. Kosuri, University of California, Los Angeles

F. Roth, University of Toronto, Canada: Potential for exhaustive atlases of functional missense variation for most human disease genes.

J. Thornton, University of Chicago, Illinois: Evolutionary determinants of DNA recognition in an ancient transcription factor.

A. Keating, Massachusetts Institute of Technology, Cambridge: High-throughput, quantitative analysis of protein–protein interactions.

R. Sun, University of California, Los Angeles: Quantitative viral genomics at single-nucleotide resolution.

D. Fowler, University of Washington, Seattle: Large-scale functional assessment of variants for genome interpretation.

Open Discussion II: New Technologies and Needed Resources

Chairperson: D. Fowler, University of Washington, Seattle

SESSION 5

Chairperson: R. Phillips, California Institute of Technology, Pasadena

S. Kosuri, University of California, Los Angeles: How do we design the best 10,000 reporters to differentiate hypotheses for how sequence determines function?

A. Abate, University of California, San Francisco: High-density sequence function mapping of an enzyme with droplet-based microfluidics.

J. Taipale, Karolinska Institutet, Huddinge, Sweden: Genome-wide analysis of protein–DNA interactions.

SESSION 6

Chairperson: M. Laub, Massachusetts Institute of Technology, Cambridge

S. Kosuri

R. Phillips, California Institute of Technology, Pasadena: Discovering the rules of regulation in biology's best-understood organism.

G. Tkačik, Institute of Science and Technology Austria, Klosterneuburg, Austria: Evolutionary and biophysical constraints on the regulatory sequence.

A. Walczak, Ecole Normale Supérieure, Paris, France: High-throughput measurement of antigen-antibody affinity.

SESSION 7

Chairperson: F. Roth, University of Toronto, Canada

M. Laub, Massachusetts Institute of Technology, Cambridge: Mapping the sequence space of bacterial signaling proteins.

R. Ranganathan, University of Texas Southwestern Medical Center, Dallas: The evolutionary design of proteins.

J. Bloom, Fred Hutchinson Cancer Research Center, Seattle, Washington: Using measurements in the lab to understand evolution in nature.

A. DePace, Harvard Medical School, Boston, Massachusetts: Precision and plasticity in animal transcription.

D. McCandlish, University of Pennsylvania, Philadelphia: Comprehensible models of higher-order interactions.

A. Siepel, Cold Spring Harbor Laboratory: Inference of fitness consequences for regulatory mutations.

Open Discussion III: Big Challenges and Future Directions

Chairperson: A. Siepel, Cold Spring Harbor Laboratory

Can We Make Animal Models of Human Mental Illness? A Critical Review

August 21–23

FUNDED BY **The Stanley Research Foundation**

ARRANGED BY **H. Heimer,** Cold Spring Harbor Laboratory
R. McCombie, Cold Spring Harbor Laboratory
E. Nestler, Icahn School of Medicine, Mount Sinai, New York

The use of animal models in studies of psychiatric disorders is increasingly controversial. There are arguments, on the one hand, that although imperfect, they are indispensable for research and, on the other hand, that because they are imperfect, they are at best inadequate and at worst misleading. The participants in this meeting reviewed critically and dispassionately the state of this field and covered topics such as current models and their effectiveness and how to integrate genetic and environmental factors in animal models; discussed how the new gene-editing techniques might be used in this field; assessed arguments that only primate models are valid, and discussed the implications of this approach.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

L. Young

L. Monteggia, S. Morris

SESSION 1: Introduction to Critical Questions

Chairperson: H. Heimer, Cold Spring Harbor Laboratory

E. Nestler, Icahn School of Medicine at Mount Sinai, New York: Where have we gone wrong in the past that has limited our animal models?

S. Morris, National Institute of Mental Health, Bethesda, Maryland: Animal models of…what? The RDoC perspective.

SESSION 2: Social Processing

Chairperson: A. Grace, University of Pittsburgh, Pennsylvania

J. Crawley, MIND Institute, University of California Davis, Sacramento: Translational mouse models of autism to understand causes and discover therapeutics.

A. Mills, Cold Spring Harbor Laboratory: Modeling 16p11.2 copy-number variations.

L. Young, Emory University, Atlanta, Georgia: Oxytocin, social attachment, and empathy-related behaviors in monogamous prairie voles: Implications for autism.

L. Monteggia, University of Texas Southwestern Medical Center, Dallas: Mechanism of rapid antidepressant action.

Z.-L. Qiu, Institute of Neuroscience, Shanghai, China: The nonhuman primate for autism: What can we learn from monkey?

SESSION 3: Negative Valence Systems

Chairperson: B. Moghaddam, University of Pittsburgh, Pennsylvania

A. Grace, University of Pittsburgh, Pennsylvania: The MAM developmental disruption model of schizophrenia.

N. Kalin, University of Wisconsin, Madison: Translating molecular models in non-human primates to human anxiety disorders.

S. Russo, Mount Sinai School of Medicine, New York: Can we model domains of behavior relevant to personality disorders in mice?

B. Dias, Yerkes National Primate Research Center, Atlanta, Georgia: Using olfaction to study intergenerational influences of stress.

SESSION 4: Cognitive Systems

Chairperson: A. Mills, Cold Spring Harbor Laboratory

S. Haber, University of Rochester, New York: From primate anatomy to human neuroimaging: Linking circuits to psychiatric disease.

C. Kellendonk, Columbia University, New York: Using human brain imaging studies as a guide toward animal models of schizophrenia.

F. Lee, Weill Cornell Medical College, New York: Genetic mouse models of altered anxiety-related behaviors.

C. McClung, University of Pittsburgh, Pennsylvania: The Clock mutant mice: A complex model resembling bipolar disorder.

SESSION 5: Positive Valence Systems

Chairperson: C. Kellendonk, Columbia University, New York

E. Nestler, Icahn School of Medicine, Mount Sinai, New York: Reward circuitry in drug and depression models.

Y. Shaham, IRP-NIDA, Baltimore, Maryland: Incubation of drug craving after choice-based voluntary abstinence: Implications for current "gold standard" animal models of addiction.

R. Carelli, University of North Carolina, Chapel Hill: When a good taste turns bad: Modeling negative affect and natural reward devaluation by cocaine.

SESSION 6: Orthogonal Dimensions

Chairperson: S. Haber, University of Rochester, New York

Sex Differences

T. Bale, University of Pennsylvania, Philadelphia: Similar to cancer, thinking of neuropsych disease as multiple hits that may begin at the germ cell stage.

J. Becker, University of Michigan, Ann Arbor: Sex differences and rodent models of human mental illness.

Genes, Environment, Development

J. Waddington, Royal College of Surgeons, Dublin, Ireland: Closing the translational gap between animal models and the clinical reality of mental illness: The exemplar of dimensions of psychopathology, G × E and G × G interactions, in mutant mouse models of psychosis.

B. Moghaddam, University of Pittsburgh, Pennsylvania: How to integrate genetic and environmental factors in animal models.

M. Meaney, McGill University, Montreal, Canada: Gene x environment designs in animal models.

SESSION 7: Discussion Session

Chairperson: E. Nestler, Mount Sinai School of Medicine, New York

Review of important points from previous sessions.

Consensus statement or article?

C. Kellendonk, S. Haber

Mammalian Brain Cell Diversity and Census

September 6–8

FUNDED BY **National Institute of Mental Health, NIH**

ARRANGED BY **A. Beckel-Mitchener,** National Institute of Mental Health, Bethesda, Maryland
J. Huang, Cold Spring Harbor Laboratory

The objective of this meeting was to initiate discussions among international groups with common interests in identifying, classifying, and characterizing cell types in the vertebrate brain. Cells are essential components that make up the circuitry underlying complex function, and better classification of the functional cell classes that are present in the brain will yield valuable results providing an important foundation for systems-based studies. A detailed classification of the variety of cell types present will broaden our understanding of the brain and enable the manipulation of specific cells and circuits. The primary goal of the workshop was to discuss the potential for coordinating the production of broad reference cell catalogs for the vertebrate (mammalian) brain.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Introduction: A. Beckel-Mitchener, National Institute for Mental Health Bethesda, Maryland

G. Feng, J. Huang

W. Koroshetz, J. Eberwine

SESSION 1: Cell Type

Moderator: S. Hill, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland

C. Koch, Allen Institute for Brain Science, Seattle, Washington: Integrating distinct data modularities to derive cell types.

S. Linnarssen, Karolinska Institutet, Stockholm, Sweden: Cell-type discovery in mouse developing and adult nervous system.

J. Huang, Cold Spring Harbor Laboratory: Transcriptional definition of cortical GABAergic neuron types.

J. Ngai, University of California, Berkeley: Illuminating cellular diversity in the nervous system.

K. Harris, University College London, United Kingdom: New algorithms for scRNA-Seq data, applied to classification of CA1 and V1 interneurons.

A. Regev, Broad Institute of MIT and Harvard, Cambridge, Massachusetts: Case studies toward a cell atlas of neurons.

SESSION 2: Connectome

Moderator: Y. Yao, National Institute of Mental Health, Rockville, Maryland

P. Osten, Cold Spring Harbor Laboratory: Tools for automated mapping of brain cell density, morphology, and connectivity.

H. Dong, University Southern California, Los Angeles: Mouse Connectome Project: Bridging macro-, meso-, and micro-scales.

E. Callaway, Salk Institute for Biological Studies, La Jolla, California: Improved monosynaptic neural circuit tracing using engineered rabies virus glycoprotein variants.

H. Zeng, Allen Institute for Brain Science, Seattle, Washington: Multiscale, integrated connectomics among cell types in local and global circuits.

SESSION 3: Technology

Moderator: J. Huang, Cold Spring Harbor Laboratory

Q. Luo, Huazhong University of Science and Technology, Wuhan, China: Visible brain-wide networks at single neuron resolution with landmarks.

J. Eberwine, University of Pennsylvania, Philadelphia: Subcellular single-neuron genomics.

K. Zhang, University of California, San Diego, La Jolla: Methods for cell-type classification, annotation, and spatial mapping.

SESSION 4: Partnership: General Discussion

Discussion Leaders: W. Koroshetz, National Institute of Neurological Disorders and Stroke, Bethesda Maryland.

S. Hill, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland,

C. Koch, Allen Institute for Brain Science, Seattle, Washington, and

G. Feng, Massachusetts Institute of Technology, Cambridge

SESSION 5: Big Brain

Moderator: C. Koch, Allen Institute for Brain Science, Seattle, Washington

H. Zeng

T. Shimogori, RIKEN BSI, Saitama, Japan: Gene expression atlas of marmoset brain.

G. Feng, Massachusetts Institute of Technology, Cambridge: Genome-editing in primates.

E. Lein, Allen Institute for Brain Science, Seattle, Washington: Multimodal characterization and classification of cell types in human neocortex.

A. Kriegstein, University of California, San Francisco: Origins of cell diversity in the developing human neocortex.

C. Walsh, Harvard Medical School, Boston, Massachusetts: Cell-type-specific splicing regulates neurogenesis in developing cerebral cortex.

SESSION 6: Data Integration and Visualization

Moderator: K. Harris, University College London, United Kingdom

G. Ascoli, George Mason University, Fairfax, Virginia: Draft neuron census based on axonal/dendritic locations.

M. Hawrylycz, Allen Institute For Brain Science, Seattle, Washington: Digital atlases and resources for a mammalian brain cell census.

S. Hill, Ecole Polytechnique Fédérale de Lausanne, Geneva, Switzerland: A data-driven knowledge space for single cells.

Final General Discussion

Making Oxidative Chemotherapy Less Toxic

September 11–14

FUNDED BY **Northwell Health–Cold Spring Harbor Lab Partnership**

ARRANGED BY **A. Holmgren,** Karolinska Institute, Stockholm, Sweden
R. Maki, Northwell Health Cancer Institute, Lake Success, New York
D. Tuveson, Cold Spring Harbor Laboratory

Many effective anticancer drugs are known to induce cell cycle arrest, or kill tumors by increasing oxidative pressure on the tumor through the production of reactive oxygen species (ROS). ROS function as second messengers controlling cell proliferation and differentiation in cancer cells. Tight control of ROS is critical for biological processes in normal cells for regulating gene expression and protein translation, as well as protein–protein interactions and ATP production. However, oxidizing strategies useful in oncology, such as chemotherapy and radiation, lack selectivity, producing dose-limiting toxicities that prevent them from reaching their full therapeutic potential. Bursts of ROS that specifically target cancer cells could prove beneficial for patients if untoward toxicity can be minimized. This meeting discussed new strategies for making oxidative and other chemotherapies less toxic.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

SESSION 1: New Ways That Redox Regulates Cancer Cells

Chairperson: M. Espey, National Cancer Institute, Rockville, Maryland

D. Tuveson and C. Chio, Cold Spring Harbor Laboratory: NRF2 and mRNA translation.

L. Cantley, Weill Cornell Medical College, New York: ROS inhibits glycolysis.

SESSION 2: Metformin and Improved Cancer Treatments

Chairperson: D. Tuveson, Cold Spring Harbor Laboratory

K. Struhl, Harvard Medical School, Boston, Massachusetts: Can metformin permit the use of lower doses of oxidative chemotherapy?

M. Pollak, McGill University, Montreal, Quebec, Canada: Metformin and therapy.

SESSION 3: Balancing ROS Efficacy and Toxicity

Chairperson: R. Maki, Northwell Health Cancer Institute, Lake Success, New York

T. Pardee, Wake Forest Baptist Medical Center, Winston-Salem, North Carolina: Targeting the TCA cycle in cancer to increase response to therapy.

P. Roberts, G1 Therapeutics, Research Triangle Park, North Carolina: Protecting the bone marrow and immune system from cytotoxic drugs during cancer treatment.

C. Li, Boston Biomedical Inc., Cambridge, Massachusetts: STAT3 drugs in cancer.

P. Bingham, Stony Brook University, New York: Selectively targeting tumor mitochondrial metabolism synergizes with traditional oxidative chemotherapies.

General Discussion Highlighting Key Points

M. Espey, National Cancer Institute, Rockville, Maryland, and **R. Maki,** Northwell Health Cancer Institute, Lake Success, New York

SESSION 4: STAT3, Redox and Cancer

Chairperson: D. Frank, Dana-Farber Cancer Institute, Boston, Massachusetts

J. Bromberg, Memorial Sloan Kettering Cancer Center, New York: Reversing resistance to targeted therapies.

D. Frank, Dana-Farber Cancer Institute, Boston, Massachusetts: The effect of redox-active molecules on oncogenic signaling pathways.

D. Levy, New York University School of Medicine, New York, and M. Isabelle, New York University Medical Center, New York: Mitochondrial STAT3 and redox stress.

R. Pethig, The University of Edinburgh, Scotland: A summary of some studies (with Albert Szent-Györgyi) of the quenching of ascorbate/semiquinone free radicals by Ehrlich ascites tumor cells.

SESSION 5: Vitamin C, K, and Others in Cancer Therapy

Chairperson: A. Holmgren, Karolinska Institute, Stockholm, Sweden

S. Coutts, IC-MedTech, Santa Fe, New Mexico, and T. Miller, IC-MedTech, Las Vegas, Nevada: Apatone: Basic science to clinic.

J. Verrax, APB Belgian Pharmaceutical Association, Brussels, Belgium: Mechanisms involved in the anticancer properties of Apatone.

D. Neal, Summa Health, Akron, Ohio: The beginning for Apatone.

A. Holmgren, Karolinska Institute, Stockholm, Sweden: Apatone in cancer treatment: Role of replicative stress following oxidative effects on ribonucleotide reductase and its electron donors thioredoxin and glutathione.

L. Trotman, Cold Spring Harbor Laboratory: Apatone in metastatic prostate cancer.

General Discussion Highlighting Key Points

A. Holmgren, Karolinska Institute, Stockholm, Sweden

SESSION 6: Translation of Redox Therapies to the Clinic

Chairperson: G. Raptis, Northwell Health Cancer Institute, Lake Success, New York

D. Lamm, B.C.G. Oncology, P.C., Phoenix, Arizona: Urothelial carcinoma: The stepchild that could lead the way.

L.J. Hoffer, Lady Davis Institute for Medical Research, Montreal, Canada: Redox clinical trials.

T. Miller, IC-MedTech, Las Vegas, Nevada: Regulatory affairs of ROS clinical trials.

W. Isacoff, University of California, Los Angeles: Discussion of clinical trial development.

L. Cantley, J. Bromberg

D. Levy, C. Li, J. Watson

Diagnostic Tests for Lyme Disease: A Reassessment

September 18–21

FUNDED BY **Global Lyme Alliance, Greenwich, Connecticut**

ARRANGED BY **J. Branda,** Harvard University, Boston, Massachusetts
S. Schutzer, Rutgers, The State University of New Jersey, Newark

Lyme disease, caused by *Borrelia burgdorferi*, is the number one tick-borne disease in the United States and Eurasia. Accurate and unambiguous diagnosis of infections is not only important for the individual patient, but also essential for providing objective evidence of infections for subjects to be enrolled in clinical trials and to monitor the effectiveness of new therapies. The current diagnostic test was established in 1994 at the Dearborn Conference. However, new technologies for detecting microbial infections have been developed over the past 22 years, and this was an excellent time to review the current state of laboratory diagnosis of Lyme disease and to examine whether any of the more-recently developed techniques might be useful. This meeting brought experts in Lyme disease diagnostics together with experts developing tests for other emerging infections. The goal was to end the meeting with a clearer picture of what can be done to improve Lyme disease diagnosis.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Goals and Objectives: J. Branda, Massachusetts General Hospital, Boston
S. Schutzer, Rutgers New Jersey Medical School, Newark, New Jersey

SESSION 1: Historical Perspective and Current Approaches to Serologic Testing for Lyme Disease

Chairperson: J. Branda, Massachusetts General Hospital, Boston

A. Steere, Massachusetts General Hospital, Boston: Origin and overview of current serologic testing approach: Strengths of standard serologic testing.

R. Dattwyler, New York Medical College, New York: New-generation serologic tests.

SESSION 2: Recent Advances in Serologic Testing for Lyme disease

Chairperson: A. Steere, Massachusetts General Hospital, Boston

W. Robinson, Stanford University School of Medicine, California: Limitations of standard serologic testing; areas for improvement.

J. Branda, Massachusetts General Hospital, Boston: Variations on two-tiered testing.

M. Kintrup, Viramed Biotech AG, Planegg, Germany: Microchip platform in use for serologic testing.

SESSION 3: New Technologies and Approaches to Lyme Disease Diagnostics: Serologic Testing

Chairpersons: A. Marques, National Institute for Allergies & Infectious Diseases, Bethesda, Maryland, and **E. Fikrig,** Section of Infectious Diseases, Yale University School of Medicine, New Haven, Connecticut

M. Schriefer, Center of Disease Control and Prevention, Ft. Collins, Colorado: CDC experience with modified two-tiered testing protocols.

M. Gomes-Solecki, University of Tennessee Health Sciences Center, Memphis: Issues related to multiplexed assays, illustrated by lab-on-a-chip point of care device.

A. Steere, Massachusetts General Hospital, Boston: Detection of autoantibodies as biomarkers of *B. burgdorferi* infection.

SESSION 4: New Technologies and Approaches to Lyme Disease Diagnostics: Molecular Diagnostics

Chairpersons: M. Ilias, National Institute for Allergies and Infectious Diseases, Rockville, Maryland, and **S. Schutzer,** Rutgers New Jersey Medical School, Newark, New Jersey

T. Lowery, T2 Biosystems, Lexington, Massachusetts: High-sensitivity culture-free detection with T2MR for sepsis and Lyme disease.

T. Slezak, Lawrence Livermore National Laboratory Livermore, California: Targeted sequencing for microorganism detection.

E. Mongodin, University of Maryland, Baltimore, Maryland: Whole-genome sequencing to detect microorganisms.

L. Liotta, George Mason University, Manassas, Virginia, and S. Schutzer, Rutgers New Jersey Medical School, Newark: Sample concentration/enrichment technologies.

J. Boyle, Qiagen, Inc., Germantown, Maryland: Hot topics for further discussion.

SESSION 5: Adoption of New Diagnostic Methods for Other Infectious Diseases

Chairpersons: M. Gomes-Solecki, University of Tennessee Health Sciences Center, Memphis, and **A. Marques,** National Institute for Allergies & Infectious Diseases, Bethesda, Maryland

S. Wong, New York State Department of Health, Albany, New York: Public health diagnostic response to emerging pathogens.

M. Schriefer, Center of Disease Control and Prevention, Ft. Collins, Colorado: CDC experience with modified two-tiered testing protocols.

B. Branson, Centers for Disease Control and Prevention, Atlanta, Georgia: Adoption of new HIV testing strategies: How we did it.

SESSION 6: Pathways to Adoption of New Lyme Disease Diagnostic Approaches

Chairperson: M. Lewinski, Roche Molecular Systems, Inc., Pleasanton, California

P. Mead, Centers for Disease Control and Prevention, Atlanta, Georgia: Process for updating CDC recommendations.

K. Roth, U.S. Food and Drug Administration, Silver Spring, Maryland: FDA procedures for clearance of new tests for Lyme disease: New 2015 approach to ASRs.

B. Body, LabCorp, Burlington, North Carolina: How do regulatory issues influence the decision to adopt or develop new assays? Laboratory-developed tests versus FDA-cleared assays.

Further Discussion of Hot Topics

J. Branda, Massachusetts General Hospital, Boston, and **S. Schutzer**, Rutgers New Jersey Medical School, Newark, New Jersey

SESSION 7: Review of Highlights and Discussion about Potential White Paper

Chairpersons: J. Branda, Massachusetts General Hospital, Boston, and **S. Schutzer,** Rutgers New Jersey Medical School, Newark

M. Ilias, National Institute for Allergies & Infectious Diseases, Rockville, Maryland: New opportunities and initiatives at NIH.

Discussion of White Paper/Opinion Piece

Suggestions for Follow-Up Meetings

Genomics-Enabled Accelerated Crop Breeding

October 16–19

FUNDED BY **Monsanto Company (CSHL Corporate Sponsor Program) with additional funding from DuPont Pioneer, 2Blades Foundation, and Calyxt Inc.**

ARRANGED BY **B. Staskawicz,** University of California, Berkeley
D. Voytas, University of Minnesota, St. Paul

The application of genomics-enabled crop improvement is rapidly being adopted by both the academic and commercial sectors. The "next-generation" breeding tools are revolutionizing crop production and will also bring about profound changes in what are considered genetically modified organisms. Indeed, regulatory agencies in the United States are currently re-evaluating how these technologies will be regulated, and recommendations will be made toward the end of 2016. Over the years, the Banbury Center has held many meetings on rapidly changing fields, providing an opportunity to take stock of what is happening and to look to future developments. This meeting set out to do the same. Participants examined topics including the development and application of genome editing tools in plants, the future convergence of breeding and multiplex genome editing in crop plants, and what the regulatory landscape of genome-edited plants will be in the future. The meeting was international in scope and participants, covering a wide range of crops; participants were drawn from both academia and industry.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Techniques of Plant Genome Engineering: Past, Present, Future
H. Puchta, Karlsruhe Institute of Technology, Karlsruhe, Germany

R. Michelmore, Y. Qi

F. Zhang (MIT)

SESSION 1: Tools

Chairperson: C. Gao, Institute of Genetics and Developmental Biology, CAS, China

A. Britt, University of California, Davis: CenH3-mediated haploid induction.

G. Gocal, Cibus, San Diego, California: Precision genome-editing tools for nontransgenic trait development.

S. Jacobsen, University of California, Los Angeles: Targeted epigenetic modification.

N. Patron, The Earlham Institute, Norwich, United Kingdom: Improving the plant genome engineering toolbox.

Y. Qi, East Carolina University, Greenville, North Carolina: Multiplexing systems for plant genome editing and transcriptional regulation.

F. Zhang, Massachusetts Institute of Technology, Cambridge: Exploring bacterial diversity for genome engineering.

SESSION 2: Implementation

Chairperson: F. Zhang, Calyxt Inc., New Brighton, Minnesota

J. Corn, University of California, Berkeley: Mechanisms to improve sequence modification at defined genomic loci.

C. Gao, Institute of Genetics and Developmental Biology, CAS, China: Developing transgene-free genome-editing technologies in wheat.

D. Voytas, University of Minnesota, St. Paul: Optimizing gene targeting in plants.

Y. Yang, Penn State University, University Park, Pennsylvania: CRISPR/Cas9-enabled multiplex genome editing and precision crop breeding.

SESSION 3: Feeding the World

Chairperson: R. Bart, Donald Danforth Plant Science Center, St. Louis, Missouri

D. Horvath, 2Blades Foundation, Evanston, Illinois: Advancing crop-breeding strategies for disease resistance.

B. Mazur, DuPont Pioneer, Wilmington, Delaware: Providing advanced breeding-enabled crops to growers globally.

SESSION 4: Applications: Editing

Chairperson: A. Hummel, KWS Gateway Research Center, St. Louis, Missouri

R. Bart, Danforth Center, St. Louis, Missouri: Genome editing in cassava.

J. Jones, The Sainsbury Laboratory, Norwich, United Kingdom: Crisping in *Arabidopsis* and tomato for discovery and crop improvement.

R. Michelmore, University of California, Davis: Genome editing in lettuce.

R. Stupar, University of Minnesota, St Paul: Opportunities and obstacles for CRISPR in soybean.

B. Staskawicz, University of California, Berkeley: Genome editing for disease resistance in crop plants.

F. Zhang, Calyxt, Inc., New Brighton, Minnesota: Genome editing for crop and food improvement.

SESSION 5: Applications: Breeding

Chairperson: R. Stupar, University of Minnesota, St. Paul

R. Buell, Michigan State University, East Lansing Polyploid and clonally propagated crops: Challenges in genomics-enabled breeding.

S. Dellaporta, Yale University, New Haven, Connecticut: Genomic and computational pipelines for plant breeding populations.

R. Dirks, Rijk Zwaan Breeding BV, Fijnaart, The Netherlands: Chromosome substitution lines and libraries: Designer chromosomes, designer breeding.

S. Soyk, Cold Spring Harbor Laboratory: Using genome editing to create novel qualitative and quantitative variation for breeding.

S. Yang, Monsanto Company, St. Louis, Missouri: Accelerating breeding with molecular methods for trait discovery and deployment.

SESSION 6: The Future

Chairpersons: D. Voytas, University of Minnesota, St. Paul, and **B. Staskawicz,** University of California, Berkeley

D. Ware, Cold Spring Harbor Laboratory: Biology enabled crop breeding.

The Lustgarten Foundation: Vitamin D Day

November 1

FUNDED BY **Stand Up to Cancer and the Lustgarten Foundation**

ARRANGED BY **R. Evans,** Salk Institute for Biological Studies, La Jolla, California
P. Sharp, Massachusetts Institute of Technology, Cambridge
D. Tuveson, Cold Spring Harbor Laboratory

Synthetic, nonmetabolized vitamin D agonists promote drug delivery and response to chemotherapy in mouse models of pancreatic cancer. The working hypothesis is that this is due to reprogrammed pancreatic stellate cells and an altered tumor microenvironment. In the first human study, a neoadjuvant trial of gemcitabine + Abraxane + paricalcitol showed potential activity. The mechanism of response is not clear and may involve immune cells. The purpose of this meeting was to clearly articulate the known and unknown aspects of vitamin D clinical trials that are completed or under way, such that synergy and cooperation can occur.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Meeting Goals: P. Sharp, Massachusetts Institute of Technology, Cambridge
R. Evans, Salk Institute for Biological Studies, La Jolla, California
D. Tuveson, Cold Spring Harbor Laboratory

V. Balachandran

E. Furth

SESSION 1: SU2C Team 1 Results of Neoadjuvant Trial

Chairperson: D. Tuveson, Cold Spring Harbor Laboratory

Clinical Data: P. O'Dwyer, University of Pennsylvania, Philadelphia, and J. Drebin, University of Pennsylvania, Philadelphia

PSC Findings:

R. Evans, Salk Institute for Biological Studies, La Jolla, California

TME Analysis:

R. Vonderheide, Abramson Cancer Center, Philadelphia, Pennsylvania, and E. Furth, University of Pennsylvania, Philadelphia

SESSION 2: Other Vitamin D Approaches

Chairperson: D. Tuveson, Cold Spring Harbor Laboratory

"Grand Slam"

D. Von Hoff, Translational Genomics Research Institute, Phoenix, Arizona, and E. Borazanci, HonorHealth Research Institute, Scottsdale, Arizona: GAC + Paricalcitol + anti-PD1.

Convergence Team

J. Drebin, University of Pennsylvania, Philadelphia, and J. Wolchok, Memorial Sloan Kettering Cancer Center, New York: Anti-PD1 + vitamin D.

B. Vogelstein

LF Planned Trial in Stage-4 PDA

B. Wolpin, Dana-Farber Cancer Institute, Boston, Massachusetts: Paracalcitol + chemotherapy.

SESSION 3: Discuss Strategy to Expand Efforts

Chairperson: P. Sharp, Salk Institute for Biological Studies, La Jolla, California

Preclinical: Led by R. Evans and M. Truitt, Salk Institute for Biological Studies, La Jolla, California

Clinical: Led by P. O'Dwyer and J. Drebin, University of Pennsylvania, Philadelphia

D. Von Hoff, Translational Genomics Research Institute, Phoenix, Arizona, and R. Vonderheide, Abramson Cancer Center, Philadelphia, Pennsylvania: Neoadjuvant, adjuvant, advanced: First line versus second line; ImmunoTx.

SESSION 4: Plan for Future Trials (Milestones)

Chairpersons: D. Tuveson, Cold Spring Harbor Laboratory, and **P. Sharp,** Salk Institute for Biological Studies, La Jolla, California

SESSION 5: Assembly of a Vitamin D Task Force to Report Quarterly to SU2C/Lustgarten Foundation

Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects

November 9–11

FUNDED BY **Genentech, Inc., King & Spalding, LLP, DRI Capital Inc., McDonnell Boehnen Hulbert and Berghoff LLP, and Biotechnology Innovation Organization**

ARRANGED BY **M. Brivanlou,** King & Spalding LLP, New York
H. Sauer, Biotechnology Innovation Organization, Washington, D.C.
K. Sonnenfeld, King & Spalding LLP, New York

The Supreme Court's decisions in Mayo, Myriad, and Alice, and their broad interpretation by the USPTO, have significantly limited the subject matter eligible for patent protection related to diagnostic processes and preparations of naturally occurring substances and materials. Patent protection for subject matter that has been patentable for more than 100 years is now in question. The Federal Circuit itself has stated that it feels bound by the Supreme Court precedent to invalidate patents directed to commercially important discoveries used to create novel and nonobvious diagnostics. Furthermore, these actions have also created a growing anomaly in U.S. patent law, where biotech inventions that are patentable in most other industrialized countries are being denied patent protection in the United States, with attendant effects on trade and the cross-border flow of innovation. It is an appropriate time to examine whether well-intentioned Supreme Court decisions and their implementation by the USPTO have "overshot" their goals and given rise to an overcorrection in the law that is inconsistent with good trade and innovation policy.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Introduction: K. Sonnenfeld, King & Spalding, New York
M. Brivanlou, King & Spalding, New York
H. Sauer, Biotechnology Innovation Organization, Washington, D.C.

S. Knowles

SESSION 1: Where Have We Been? What We Have Gained? What We Have Lost?

Chairperson: K. Sonnenfeld, King & Spalding, New York

S. Knowles, Knowles Intellectual Property Strategies, Atlanta, Georgia: Framing the issues.
K. Noonan, McDonnell Boehnen Hulbert & Berghoff LLP, Chicago, Illinois: Is it time for Perestroika in U.S. Patent Law?
M. Moran, U.S. Patent and Trademark Office, Alexandria, Virginia: Evolution of USPTO eligibility guidance in the life sciences.

SESSION 2: The Law: How Did We Get Here? How Have the Courts and USPTO Reacted? How Will the Law Evolve?

Chairperson: I. Pleasure, Genentech, South San Francisco, California

K. Sonnenfeld, King & Spalding, New York: The expansion of Funk Brothers.
F. Chapinal, PharmaMar, Madrid, Spain: Under what circumstances can naturally occurring substances be patent-eligible?
K. Dow, Johnson & Johnson, Spring House, Pennsylvania: Patenting of method of treatment claims.
C. Coburn, Genentech Inc., South San Francisco, California: Therapeutic antibodies after Myriad.
J. Haley, Jr., Ropes and Gray LLP, New York: Claims that USPTO has issued in last 2–3 years to genes, natural products and diagnostics, and rationale.

SESSION 3: International Perspective: How Does Patentability of Genes, Natural Products, and Diagnostics Differ Abroad? What Can We Learn from the Comparison?

Chairperson: M. Brivanlou, King & Spalding, New York

H. Rainer-Jaenichen, Vossius & Partner, Munich, Germany: Limits to patentability of biotech and pharmaceutical inventions in the EPO.
J. Cherry, FPA Patent Attorneys, Melbourne, Australia: Myriad in Australia: The informational approach of the High Court.
G. Lewis, JA Kemp, London, United Kingdom: Patentable subject matter at the EPO with focus on antibodies and partner diagnostics.
C. Salsberg, Novartis, Washington, D.C.: International impact of U.S. Patent eligibility law.
J. Haley, Jr., Ropes and Gray LLP, New York: Reflections on the 1981 Banbury Center meeting.

H. Rainer-Jaenichen

T. Rea

SESSION 4: Policy: How Does the Patentability of Genes, Natural Products, and Diagnostics Impact Innovation, Investment, and Competition?

Chairperson: H. Sauer, Biotechnology Innovation Organization, Washington, D.C.

T. Rea, Crowell & Moring LLP, Washington, D.C.: Observations on recent studies and where we are today.

P. Alloway, DRI Capital Inc., Toronto, Canada: Challenges of investing in an antipatent environment.

D. Kappos, Cravath, Swaine & Moore LLP, New York: Approaches for cleaning up the 101 mess: Policy and practical.

SESSION 5: Panel Discussion: Where Do We Go from Here? What, If Any, Reforms Should be Made? How Could We Effect Them?

Moderator: G. Elliot, Retired USPTO, Alexandria, Virginia

R. Dreyfuss, New York University of Law, New York

R. Armitage, Consultant, IP Strategy & Policy, Marco Island, Florida

S. Michel, Google Inc., Washington, D.C.

H. Sauer, Biotechnology Innovation Organization, Washington, D.C.

SESSION 6: Review and Summary

Chairpersons: K. Sonnenfeld and M. Brivanlou, King & Spalding LLP, New York, and **H. Sauer,** Biotechnology Innovation Organization, Washington, D.C.

Evolution of the Translational Apparatus and Implications for the Origin of the Genetic Code

November 13–16

FUNDED BY **Cold Spring Harbor Laboratory Corporate Sponsor Program**

ARRANGED BY **H. Hartman,** Massachusetts Institute of Technology, Cambridge
T. Smith, Boston University, Massachusetts

The origin of the genetic code is one of the great challenges of biology. In the 50 years since the 1966 Cold Spring Harbor Laboratory Symposium on "The Genetic Code," there have been revolutionary advances in our understanding of the relationship between the genetic code and proteins and the insights this provides on how the genetic code has the form it does. The main evidence covered in this meeting was the origin and evolution of the translational apparatus, focusing on the ribosome, the aminoacyl-tRNA synthetases, and their tRNAs. Under the assumption that the origin of the code can be separated from the origin of life, participants also reviewed the pre-code biosynthesis of the monomers (e.g., the amino acids, lipids, sugars, nucleotides, and their early polymerization). The meeting concluded with discussions of the significance of these findings for our understanding of the origin and history of the genetic code.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Opening Remarks: T. Smith, Boston University, Massachusetts

M. Yarus, T. Steitz

K. Musier-Forsyth

SESSION 1: Ribosomal RNA Implications

H. Noller, University of California, Santa Cruz: The ribosome: Overview and origin.
T. Steitz, Yale University New Haven, Connecticut: What the ribosomal RNA structure tells us.
L. Williams, Georgia Institute of Technology, Atlanta: Evolution of the ribosome before LUCA.
A.S. Petrov, Georgia Institute of Technology, Atlanta: The LSU is from Mars, the SSU is from Venus.

SESSION 2: Early Peptide and RNAs from Pieces

A. Lupas, Max-Planck-Institute for Developmental Biology, Tubingen, Germany: Ribosomal proteins as documents of the transition from (poly)peptides to folded proteins.
L. Jaeger, University of California, Santa Barbara: RNA self-assembly, RNA structural evolution, and RNA nanomachines.
M. Yarus, University of Colorado, Boulder: Molecularity/catalysis/utility (MCU) theory and the primordial genetic system.

SESSION 3: Ribosomal Proteins

T. Smith, Boston University, Massachusetts: Protein taxonomic block structure.
G. Fox, University of Houston, Texas: Ribosome origins and subsequent evolution.

SESSION 4: Aminoacyl tRNA Synthetases

D. Soll, Yale University, New Haven, Connecticut: The evolution of genetic code deviants.
D. Moras, Institut Génétique Biologie Moléculaire Cellulaire, Illkirch, France: Specific structural features of class II synthetases.
L.R. de Pouplana, Institute for Research in Biomedicine, Barcelona, Spain: Functional limits of the genetic code.
K. Musier-Forsyth, Ohio State University, Columbus: Prolyl-tRNA synthetases: Aminoacylation and editing.
A. Torres-Larios, National Autonomous University of Mexico: Why two Glycyl-tRNA synthetases?

SESSION 5: tRNA Evolution

T. Steitz, Yale University, New Haven, Connecticut: Structure of CCA adding enzyme.
L. Aravind, National Center for Biotechnology Information, Bethesda, Maryland: The deep evolutionary links between cyclic nucleotide synthetases and nucleic acid polymerases.
C. Francklyn, University of Vermont, Burlington: Minihelices and the operational code.
M. Di Giulio, Institute Biosciences & Bioresources, CNR, Naples, Italy: The origin of the tRNA molecule.

SESSION 6: Origins

H. Jakubowski, Rutgers University, New Jersey Medical School, Newark: Thioester chemistry and the origin of coded peptide synthesis.
D. Segre, Boston University, Massachusetts: Richness and implications of a prephosphate metabolism.

SESSION 7: What Are the Implications for Our Understanding of the Origin and Evolution of the Genetic Code?

H. Hartman, Massachusetts Institute of Technology, Cambridge: Summary and thoughts on the evolution of the genetic code.

Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era

November 30–December 2

FUNDED BY **International Society of Psychiatric Genetics, Tennessee, National Society of Genetic Counselors, Illinois, Institute of Neurosciences, Mental Health and Addiction, University of British Columbia, Vancouver, and Cold Spring Harbor Laboratory**

ARRANGED BY **J. Austin,** University of British Columbia, Vancouver, Canada
F. McMahon, National Institute of Mental Health, Bethesda, Maryland

There is considerable misunderstanding of the term genetic counseling; in particular, it is often conflated with genetic testing and is often thought of as any simple interaction between a healthcare provider and a patient where genetic risk or testing is discussed. In fact, genetic counseling is a specialist healthcare discipline that involves helping clients to "understand and adapt to the medical, psychological, and familial implications of genetic contributions to disease." Fundamental questions remain regarding how best to use genetic counseling—and genetic testing—in psychiatry. Some of these questions were tackled by participants in this Banbury meeting with the aim of developing a framework to guide future developments in psychiatric genetic counseling.

Welcoming Remarks and Background: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Introduction: J. Austin, University of British Columbia, Vancouver, Canada
F. McMahon, National Institutes of Health, Bethesda, Maryland

SESSION 1: Use of Genetic Testing to Predict Risk for Psychiatric Disorders

D. Ledbetter, Geisinger Health System, Danville, Pennsylvania: Routine genetic diagnostic testing and prenatal diagnosis for psychiatric disorders.

C. Janssens, Emory University, Atlanta, Georgia: How predictive is our DNA? An overview of everything from Mendelian to polygenic diseases and traits.

J. Smoller, Massachusetts General Hospital, Boston: Implications of pleiotropy and/or issues in returning research results.

D. Alexis Carrere, McMaster University, Hamilton, Ontario, Canada: Direct-to-consumer genetic testing for bipolar disorder: Findings from the impact of personal genomics (PGen) study.

SESSION 2: Psychiatric Genetic Counseling, International Perspectives on Current Practice

E. Morris, University of British Columbia, Vancouver, Canada: Clinical applications of psychiatric genetics: Updates from Vancouver, B.C.

K. McGhee, Bournemouth University, Poole, United Kingdom: Bridging the gap between research and patient. Implementing PsyGC: The UK perspective.

R. Moldovan, Babes-Bolyai University, Cluj-Napoca, Romania: Evidence-based genetic counseling for psychiatric disorders.

F. Degenhardt, University of Bonn, Germany: Perspective from a clinical geneticist.

S. Hartz, Washington University Medical Center, St. Louis, Missouri: Pragmatic approaches to translate genetic findings into clinical care.

B. Biesecker, National Human Genome Research Institute, Bethesda, Maryland: Families rich with psychiatric disorders: Navigating communication, privacy, and a "need to know."

SESSION 3: Psychopharmacogenomics in the Clinic

J. Biernacka, Mayo Clinic, Rochester, Minnesota: Psychiatric pharmacogenomics.

J. Kennedy, Centre for Addiction and Mental Health, Ontario, Canada: Pharmacogenetic testing is becoming widely accepted: What are the issues?

A. Malhotra, Zucker Hillside Hospital, Glen Oaks, New York: Pharmacogenetics in psychiatry.

SESSION 4: Looking Forward/Next Steps

Questions for Break-Out Groups:

- What needs to be done to make genomic testing useful in the psychiatric clinic?
- What is the proper role of genetic counseling in psychiatric care? Consider relationships between psychiatrists and genetic counselors.
- What is the role(s) of professional societies in promoting the effective use of genetic testing and counseling, educating professionals, setting the research agenda as it relates to psychiatric genetics?

SESSION 5: Reports from Break-Out Groups, Summary, and Next Steps

Evolution and Revolution in Anatomic Pathology: Automation, Machine-Assisted Diagnostics, Molecular Prognostics, and Theranostics

December 4–7

FUNDED BY **Northwell Health–Cold Spring Harbor Laboratory Partnership**

ARRANGED BY **J.M. Crawford,** Northwell Health, Lake Success, New York
P. Mitra, Cold Spring Harbor Laboratory
M. Wigler, Cold Spring Harbor Laboratory

Revolutionary advances in machine intelligence, robotics, and genomics have taken place that should fundamentally improve the efficiency and quality of patient care through automated diagnostic algorithms and personalized medicine. However, implementing these new technologies will not be easy, and key hurdles to be overcome include closing nonautomated gaps in anatomic pathology; linking machine-learning to medical decision-making; data standardization and curation; and preparing for regulatory oversight and approval. The goal of the meeting was to facilitate this change by bringing together the different groups involved (scientists and engineers, medical professionals and leaders, FDA regulators, payers and insurance companies, and industry) to review two broad themes: enabling technologies, and how to integrate the information from these technologies into the practice of medicine.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

SESSION 1: Defining the Challenge for Anatomic Pathology

Chairperson: J. Crawford, Hofstra Northwell School of Medicine, Hempstead, New York

U. Balis, University of Michigan, Ann Arbor, and J. Tomaszewski, University of Buffalo, New York: Anatomic pathology as quantitative evidence for medicine: Defining the transformation that must occur for anatomic pathology.

B. Bastian, University of California, San Francisco: Tissue-based molecular diagnostics: State-of-the-art methods for analyzing the molecular evolution of cancers—From precursor identification of biomarkers to assisting in diagnosis and staging.

SESSION 2: Anatomic Pathology as Structural Molecular Diagnostics

Chairperson: K. Roth, Columbia University, New York

M. Loda, Dana-Farber Cancer Institute, Boston, Massachusetts: Anatomic pathology as molecular diagnostics: Comparing formalin-fixed paraffin-embedded versus 2D/3D molecular data.

R. Levenson, University of California, Davis: Data integrity and interoperability: Examining standards for generation of quantitative data from anatomic pathology and requirements for data transmission.

SESSION 3: Examining the Regulatory and Preanalytical Environment

Chairperson: R. Michel, The Dark Report, Spicewood, Texas

B. Gallas, U.S. Food and Drug Administration, Silver Springs, Maryland, and C. Compton, Arizona State University, Tucson, Arizona: Mitotic counting reproducibility/feature study with pathologists preanalytical processing: The biospecimen quality initiative.

SESSION 4: Expanding the Analytic Value of Anatomic Pathology

Chairperson: T. Chang, Hofstra Northwell School of Medicine, Hempstead, New York

J. Gilbertson, Massachusetts General Hospital, Boston: Informatics as empowerment for pathology to examine the informatics context for anatomic pathology.

T. Fuchs, Memorial Sloan Kettering Cancer Center, New York: Informatics as the pivot point for human diagnostics: To examine the broader perspectives of healthcare information.

M. Wigler, Cold Spring Harbor Laboratory: Cellular informatics: To examine the power of single-cell analysis and spatial data.

P. Mitra, M. Wigler

R. Levenson, J. Crawford

SESSION 5: Examining the Power of Digital Analytics 1

Chairperson: Y. Yagi, Memorial Sloan Kettering Cancer Center, New York

B. Perkins, Human Longevity, Inc., San Diego, California: The digital human: Looking to the bluest part of the sky.

M. Sivaprakasan, India Institute of Technology, Madras, India, and J. Joseph, Healthcare Technologies Innovation Center, Madras, India: The international context for automating pathology.

SESSION 6: Examining the Power of Digital Analytics 2

Chairperson: M. Lloyd, Inspirata, Inc., Tampa, Florida

P. Mitra, Cold Spring Harbor Laboratory: High-throughput histology pipeline and informatics.

Y. Yagi, Memorial Sloan Kettering Cancer Center, New York: Validating technologies for anatomic pathology.

SESSION 7: Technological Advances That Need to Occur: Machine Learning, Automation, Digital Imaging, Genomics, Biomarkers

Chairperson: R. Levenson, University of California, Davis: Moderated discussion.

SESSION 8: Nontechnological Events That Need to Occur: Regulatory, Payers, Clinical Workflow, Education of Pathologists

Chairperson: S. Cohen, Rutgers University–New Jersey Medical Center, New Brunswick: Moderated discussion.

SESSION 9: Identifying Hurdles, Making a Plan to Advance the Field

Chairpersons: P. Mitra, Cold Spring Harbor Laboratories, and **U. Balis,** University of Michigan, Ann Arbor: Moderated discussion.

SESSION 10: Implementing Banbury Outcomes after Banbury: White Paper Outline and Plan

Chairperson: J. Crawford, Hofstra Northwell School of Medicine, Hempstead, New York: Moderated discussion.

Developing Gene Editing as Therapeutic Strategy

December 11–14

FUNDED BY **Genentech, with additional funding from Pfizer and Cold Spring Harbor Laboratory Corporate Sponsor Program**

ARRANGED BY **C. Gersbach,** Duke University, Durham, North Carolina
J.K. Joung, Harvard Medical School, Charlestown, Massachusetts
A. Wagers, Harvard University, Cambridge, Massachusetts

Targeted genome editing has emerged as an exciting potential clinical strategy for many human diseases. New and ever-improving strategies for modifying mammalian genomes raise the imminent possibility that disease-causing mutations may be therapeutically recoded to provide long-term recovery of function. Yet, key challenges remain for realizing the full potential of genome editing in human patients. This meeting on genome editing brought together key leaders in this rapidly evolving field to discuss strategies to both anticipate and overcome challenges of clinical gene editing. Specific topics included the identification of disorders amenable to such approaches; the challenges of in vivo versus ex vivo genome editing; strategies for delivery of gene editing effectors to target cells; the influence of the immune response; approaches for increasing specificity and minimizing possible genotoxicity; and clinical and regulatory issues as informed by prior experience with other forms of gene therapy.

Welcoming Remarks: J.A. Witkowski, Banbury Center, Cold Spring Harbor Laboratory

Introduction and Goals of Meeting: A. Wagers, C. Gersbach, and J.K. Joung

SESSION 1: Technologies

Chairperson: C. Gersbach, Duke University, Durham, North Carolina

J.K. Joung, Harvard Medical School, Charlestown, Massachusetts: Defining and minimizing off-target effects of CRISPR-Cas nucleases.

F. Zhang, Massachusetts Institute of Technology, Cambridge: Therapeutic genome editing: Prospects and challenges.

A. Scharenberg, Seattle Children's Research Institute, Washington: Challenges for in vivo genome editing: Immunologic barriers to delivery of recombination templates.

D. Schaffer, University of California, Berkeley: Directed evolution of novel viral gene delivery vehicles for therapeutic gene delivery and genome editing.

K. Suzuki, Salk Institute for Biological Studies, La Jolla, California: In vivo genome editing via CRISPR-Cas9-mediated homology-independent targeted integration.

SESSION 2: T Cells

Chairperson: C. Dunbar, National Heart, Lung, and Blood Institute, Bethesda, Maryland

C. June, University of Pennsylvania, Philadelphia: Using synthetic biology to generate smarter T cells.

M. Sadelain, Memorial Sloan Kettering Cancer Center, New York: CAR T-cell editing for cancer immunotherapy.

Y. Zhao, Abramson Family Cancer Research Institute, Philadelphia, Pennsylvania: Use CRISPR/CAS9 gene editing to improve adoptive T-cell immunotherapy for cancer.

SESSION 3: Human Stem Cells

Chairperson: A. Wagers, Harvard University, Cambridge, Massachusetts

P. Cannon, Keck School of Medicine of University of Southern California, Los Angeles: Genome editing for HIV.

C. Dunbar, National Heart, Lung & Blood Institute, Bethesda, Maryland: Use of non-human primate models to optimize the safety and efficacy of hematopoietic stem cell gene editing.
D. Kohn, University of California, Los Angeles: Gene editing in human hematopoietic stem cells.
D. Bauer, Harvard University, Boston, Massachusetts: Genome editing for the hemoglobin disorders.

SESSION 4: In Vivo Genome Editing

Chairperson: M. Sadelain, Memorial Sloan Kettering Cancer Center, New York

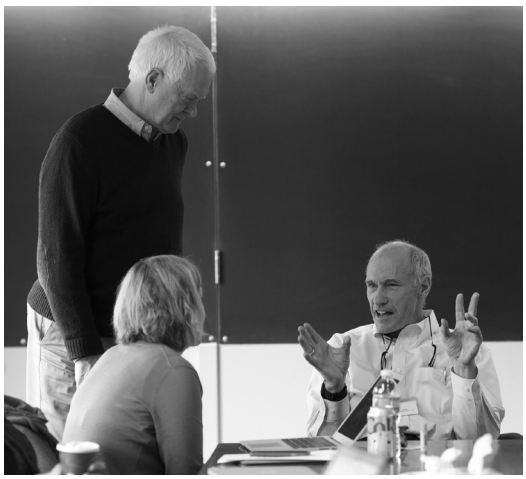
J. Wilson, University of Pennsylvania, Philadelphia: Challenges of in vivo genome editing with viral vectors.
A. Wagers, Harvard University, Cambridge, Massachusetts: Therapeutic gene editing in skeletal muscle and muscle stem cells.
C. Gersbach, Duke University, Durham, North Carolina: Genome editing for Duchenne muscular dystrophy.
D. Duan, University of Missouri Health, Columbia, Missouri: Large mammal translation.

SESSION 5: Industry Perspective

Chairperson: J.K. Joung, Harvard Medical School, Charlestown, Massachusetts

M. Holmes, Sangamo BioSciences, Inc., Richmond, California: Genome editing in primary human cells and organs: Toward the goal of engineering genetic cures.
M. Certo, bluebird bio, Cambridge, Massachusetts: Developing megaTALs for therapeutic genome editing.
C. Albright, Editas Medicine, Cambridge, Massachusetts: Advancing CRISPR medicines.
T. Barnes, Intellia Therapeutics, Inc., Cambridge, Massachusetts: Translating CRISPR/Cas9 genome editing into therapeutic reality.
S. Lundberg, CRISPR Therapeutics, Cambridge, Massachusetts: Gene editing to treat β-thalassemia and sickle cell disease.

Concluding Remarks



J. Witkowski, A. Wagers, C. June

D. Kohn, C. Dunbar

DNA LEARNING CENTER

DNA LEARNING CENTER EXECUTIVE DIRECTOR'S REPORT

ADMINISTRATION

Lindsay Barone
Lauren Correri
Mary Lamont
Valerie Meszaros
David Micklos
Carolyn Reid Faughnan

INSTRUCTION

Elna Carrasco
Heather Cosel-Pieper
Alison Cucco
Cristina Fernandez-Marco
Melissa Lee
Christine Marizzi
Kathleen McAuley
Amanda McBrien
Erin McKechnie
Bruce Nash
Michael Okoro
Sharon Pepenella
Keil Thomas
Emtiaz Uddin
Xiaoqun Catherine Zhang

BIOMEDIA

Cornel Ghiban
Susan Lauter
Joslynn Lee
Jason Williams
Chun-hua Yang

The DNALC occupies the middle ground on a continuum of science enterprise that spans from pure research to pure education. We adapt methods and concepts from research so that they can be done in educational settings. In this way, we empower students and teachers to experience cutting-edge research. In the last several years, most of our major grant projects explored this middle ground, trying to scale research methods to reach large numbers of students—especially in the context of undergraduate courses. The DNALC's task is to prepare undergraduate faculty so that they and their students can analyze the new data sets as they are released—at the same time and using the same tools as high-level researchers.

For example, our work with CyVerse (Cyber Universe) is bringing students and teachers into the world of biological "big data." This $100 million National Science Foundation (NSF) project supports an extensive computer infrastructure for scientific research and education. Leveraging the CyVerse infrastructure, our program in RNA sequence analysis (RNA-Seq) allows undergraduate faculty to generate a data set on all of the genes active in an organism of their choice. Using bioinformatics tools at our *DNA Subway* website, faculty work with students to analyze their data on the Extreme Science and Engineering Discovery Environment (XSEDE) national supercomputer system. This is likely the simplest graphical user interface to access the XSEDE system, which is typically reached via command line programming.

Over the last decade, six consensus publications, including reports by the American Association for the Advancement of Science, the NSF, and the National Academy of Sciences, have recommended similar reforms to undergraduate Science, Technology, Engineering, and Math (STEM) education.

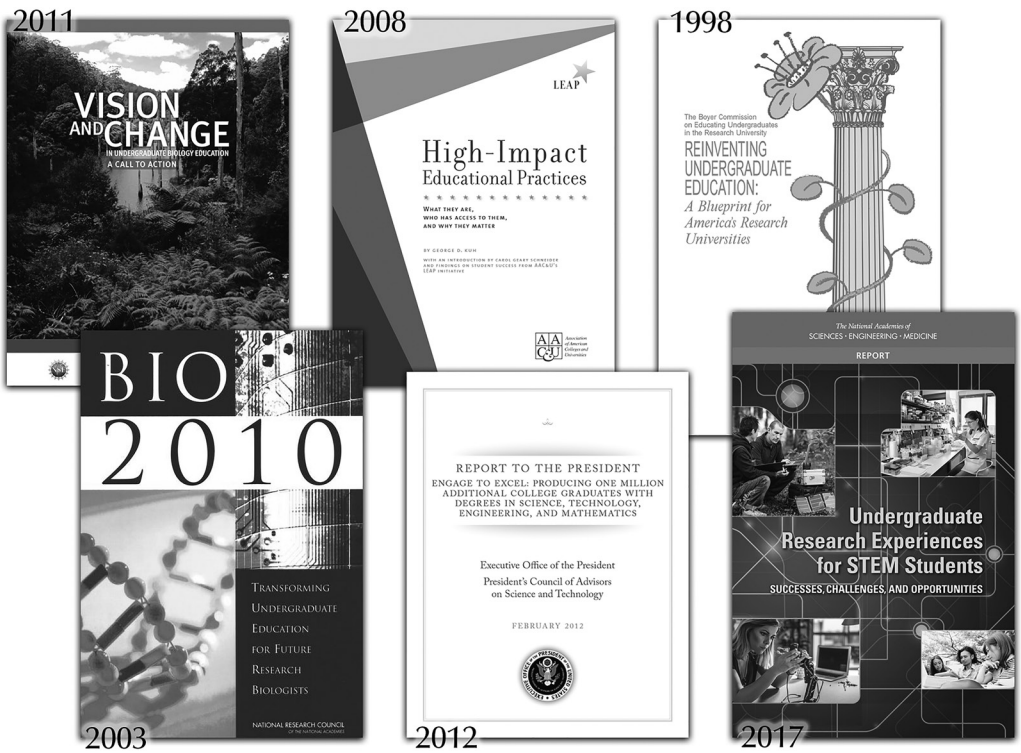
- Focus on the first 2 years of college education, which are critical to recruiting and retaining STEM majors.
- Foster conceptual understanding, higher-level thinking, and practice of STEM rather than memorization of terms, facts, and techniques.
- Adopt inquiry- and student-centered approaches that begin with students' own questions.
- Increase opportunities for interdisciplinary and collaborative work.

A study by the President's Council of Advisors on Science and Technology (PCAST) recommended "replacing standard laboratory courses with discovery-based research courses." Similarly, the 2015–2016 New York State Budget for the first time mandated "experiential

or applied learning activities"—including "faculty-supervised undergraduate projects"—as a graduation requirement at all City University of New York (CUNY) and State University of New York (SUNY) schools. The PCAST study calculated that increasing retention of STEM graduates by 10% would generate 750,000 additional STEM degrees over a decade. Past studies have shown that participation in independent research improves student persistence and academic performance. A new analysis of the Freshman Research Initiative (FRI) at the University of Texas at Austin showed that student participation in course-based undergraduate research experiences (CUREs) can *exceed* the PCAST challenge. Students who completed two semesters of course-based research had a 23% higher probability of graduating with STEM degrees than carefully matched controls. CURE participants also had a 17% higher 6-year graduation rate. These results have prompted many universities to adopt CUREs to provide authentic research opportunities to large numbers of freshman or sophomore students through courses for credit.

Developing genomics-focused CUREs is now feasible. In the last decade, a 50,000-fold decrease in DNA sequencing costs in combination with freely available bioinformatics software has made the analysis of whole genomes accessible to anyone with internet access.

The world of biological big data is creating myriad employment opportunities for students who are prepared to think in new and integrative ways. A 2003 report of an NSF blue-ribbon panel headed by Daniel Atkins popularized the term *cyberinfrastructure* to describe systems of data storage, software, high-performance computing (HPC), and people that can solve scientific problems of the size and scope presented by big data. The Atkins report described cyberinfrastructure as the means to develop a "knowledge economy." First on NSF's 2016 list of six future big research ideas is "harnessing data for 21st-century science and engineering." Third on the list is "understanding the rules of life" that determine phenotype (traits) from genotype



(genome information). According to *Science Careers*, "Big pharma, biotech, and software companies are clamoring to hire professionals with experience in bioinformatics and the identification, compilation, analysis, and visualization of huge amounts of biological and health care information."

Biology students and faculty realize that the world of big genome data is upon them. Nine out of ten graduate students, postdoctoral fellows, and research faculty we have surveyed at workshops and professional meetings say that they are currently analyzing large data sets or expect to soon. Bioinformatics and data science skills present significant barriers to faculty exploration of genomics in the classroom. Respondents to a 2016 CyVerse needs analysis of principal investigators (PIs) funded through the NSF's Biological Sciences Directorate said they have enough data storage and high-performance computing, but lack training to effectively use these resources to analyze big data. Through our role as outreach lead for CyVerse, we are helping faculty to grapple with big data science.

DNA Barcoding

In the last decade, the DNALC has developed several integrated biochemical and bioinformatics workflows that scale up discovery-based experiments for use in advanced high school and introductory college courses. These provide educators with the tools to easily and affordably provide authentic research experiences to their students. We think DNA barcoding in particular offers a complete solution for student research. ("DNA barcodes" are short DNA sequences that can be used to objectively identify almost any plant, animal, or fungal species.) DNA barcoding can be mastered in a relatively short time, can be applied to many different questions, and allows students to reach a satisfying research end point within a single academic course. Students can undertake individual projects to explore product mislabeling, or they can contribute to distributed efforts to explore a local ecosystem, museum collection, or conservation issue. These projects can stimulate independent thinking across different levels of biological organization—linking molecular genetics to ecology and evolution. DNA barcoding can readily scale for CUREs. Notably, 500 freshman students completed DNA barcode projects during the fall semester at James Madison University (JMU), one of the largest CURE implementations in the United States.

The DNALC runs three active programs that demonstrate different models for using DNA barcoding in student research. *Barcode Long Island (BLI)*, funded by the National Institutes of Health (NIH), involves students in collaborative "campaigns" to compare biodiversity across Long Island. In the *Urban Barcode Project (UBP)*, funded by the Thompson Family Foundation, and the *Urban Barcode Research Program (UBRP)*, funded by the Pinkerton Foundation, students work on independent projects to explore biodiversity in New York City. *BLI* and *UBP* students are mentored by classroom science teachers, whereas *UBRP* students are mentored by scientists from different New York City (NYC) research institutions.

UBP and *UBRP* involved 214 students from 46 schools; 29% of the participants were underrepresented minorities (URM) in science—African–American, Latino, or Native American. To complete their research projects in the spring, 73 students attended *Open Lab* sessions at *Harlem DNA Lab* and Genspace in Brooklyn, and 113 students used 16 equipment footlockers borrowed from the DNALC. Student teams developed 2200 barcode sequences, including 12 new GenBank entries, for studies of food fraud, biodiversity, conservation genetics, and forensics. Student teams competed at the annual symposium, held at the Borough of Manhattan Community College, with winning projects on invertebrate biodiversity on Staten Island and the spread of an invasive plant, *Corydalis incisa*, along the Bronx River. Winning team members from the Hostos-Lincoln Academy—Rosa Bermejo and Josiah Estacio—presented their

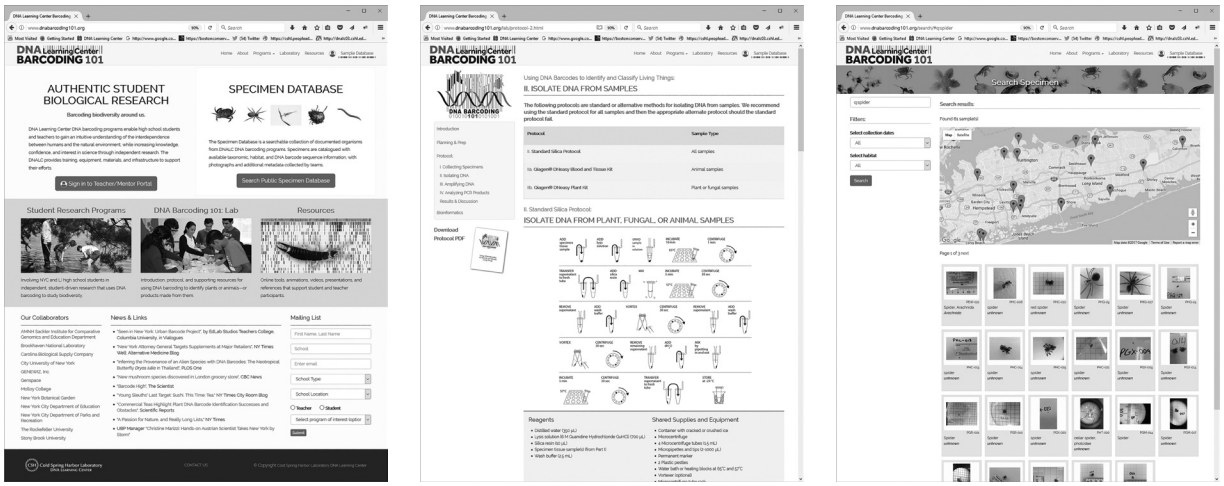
Students, mentors, and visitors explore project posters at the *UBP* and *UBRP* Symposium held at the Borough of Manhattan Community College.

research at Cold Spring Harbor Laboratory's (CSHL's) Double Helix dinner at the American Museum of Natural History. The NYC Department of Education continued its support for a 2-week barcoding course at its Environmental Study Center at Bergen Beach, Brooklyn. Four teacher-led teams of students contributed 66 DNA barcodes to ongoing study of biodiversity in Marine Park. Work measuring biodiversity during salt marsh restoration by a team from Forest Hills High School—Akansha Thakur, Ilona Petrychyn, Indu Puthenkalam, and mentor Camila Lock—was featured in the March 2016 *Barcode Bulletin* of the International Barcode of Life.

BLI enrolled 271 students from 29 schools in Nassau (14), Suffolk (13), and Queens (3), including 10% from underrepresented groups. To complete their projects, 141 students attended *Open Labs* at the DNALC, Brookhaven National Laboratory (BNL), and Stony Brook University (SBU), and 252 students used 34 footlocker kits. More than 1500 DNA barcodes were produced by biodiversity projects on plants (45%), invertebrates (42%), fungi (8%), and lichens (5%), including four novel barcode sequences published on GenBank. Mark Stoeckle, Senior Research Associate in the Program for the Human Environment at the Rockefeller University, gave the keynote talk at the annual *BLI* Symposium at CSHL (Mark launched educational barcoding when he encouraged his daughter and her friend to conduct studies of sushi collected from Manhattan restaurants.) Many *BLI* students also presented their research at regional, state, and national science competitions. Several students presented their work to New York State Commissioner of Education Mary Ellen Elia, Deputy Commissioner of Higher Education John D'Agati, and New York State Board of Regents member Roger Tilles during a visit to the DNALC on June 24.

Mark Stoeckle speaks to *BLI* Symposium attendees in Grace Auditorium on the CSHL campus.

Survey data from all three programs demonstrated the impact of barcoding research on high school students. Large majorities of students ($n = 204$) were proud of the research they had carried out (85.5%) and felt that the problem-solving experience was useful for future science courses (82.7%) and career paths (72.8%); they also felt that they were more capable of pursuing science (72.8%) and more likely to study science in college than



Our barcoding websites were merged into one comprehensive site that launched in November.

they were at the start of the project (61.3%). Among students with other research experiences ($n = 187$), majorities felt that DNA barcoding research provided as much or more opportunity for collaborative work with mentors (64.4%) and other students (68.0%) than their previous research projects. Notably, nearly 30% of students reported having discovered something that was unknown to scientists.

During the summer, we trained 31 *BLI* teachers, bringing the total to 181 teachers from 105 schools across Nassau County, Suffolk County, Brooklyn, and Queens who have been trained to mentor student projects. In the fall, proposals were accepted from 102 *BLI* teams (302 students), 43 *UBP* teams (172 students), and 19 *UBRP* teams (39 students).

Beyond Barcoding

With funding from an NIH Big Data to Knowledge supplement to *BLI*, DNA barcoding is being expanded to include metabarcoding—determining barcode sequences for all of the microbes in a mixed sample. Microbial diversity is immense and largely uncharacterized, providing students many opportunities for novel contributions. Moving from DNA barcoding to microbiomes perfectly embodies the conceptual transition from single gene to massively parallel genome analysis, introducing students to data science and high-throughput sequence (HTS) analysis. A single lane of an HTS machine can accommodate hundreds of student microbiome studies, supporting many projects and allowing comparisons between microbiomes from different environmental locations or conditions. After collection and DNA purification, PCR (polymerase chain reaction) is used to amplify a variable region of the 16S ribosomal RNA gene, and then HTS reads are used to identify the variety and abundance of microbial species.

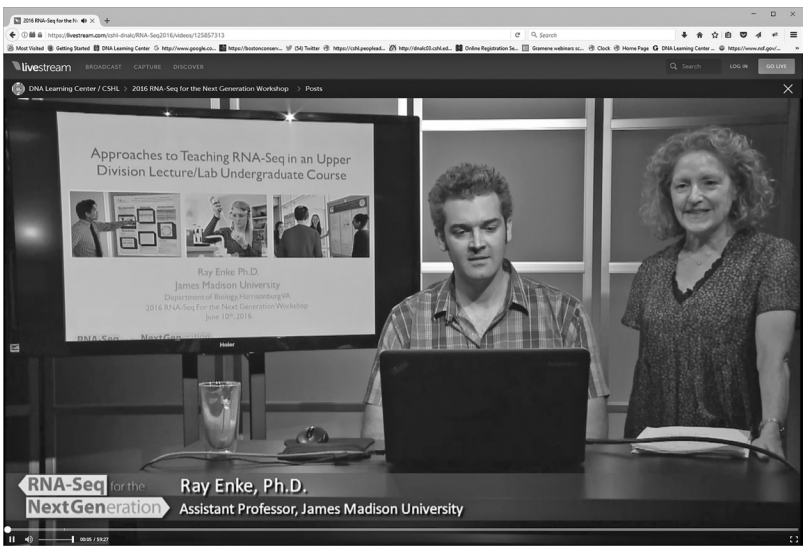
To support this new endeavor, we began to develop a laboratory and bioinformatics workflow to support student microbiome sequencing. This includes an inexpensive system to combine student samples during sequencing, a dedicated computer server, and programming notebooks to control microbiome analysis using a state-of-the-art bioinformatics tool, QIIME. Twelve *BLI* mentors were recruited from schools in Suffolk (six), Nassau (three), and Queens (three) to test the system. During the summer, mentors attended a 5-day workshop on the biochemistry and data science of microbiome analysis. Then, the mentors worked with 61 high school students (18% URMs) to study the effects of pollution, pathogens, pesticides, salinity, plant density, and boating on microbiomes from diverse samples around Long Island. In parallel efforts, one million

sequence reads were produced by microbiome research in the *UBRP*, while students from Cold Spring Harbor High School began a pilot project combining eDNA and microbiome analyses to study fish and microbes at the fish hatchery adjacent to CSHL. We are now using CyVerse resources to adapt QIIME as a new Purple Line of *DNA Subway*, which we hope will popularize microbiome analysis for student research.

RNA Sequence (RNA-Seq) Analysis

In 2016 we concluded our NSF grant, "Infrastructure and Training to Bring NGS Analysis into Undergraduate Education." This project created an extensible infrastructure and training program that empowers faculty to integrate NGS analysis and high-performance computing into undergraduate biology instruction. Free access to these tools democratizes big data, allowing faculty at any undergraduate institution to engage their students in cutting-edge biological research. Training materials, experiences, and insights from this project will help faculty bring other research technologies to scale in undergraduate classes.

The first 2 years of the RNA-Seq program involved in-person training workshops at universities on the East and West coasts. Faculty left workshops with preliminary analyses and then continued to explore the data with students during the ensuing academic year. This year's workshop made extensive use of the *Landeau Multimedia Studio* and the project website (http://www.rnaseqforthenextgeneration.org), which provides an easy-to-use interface to an extensive bank of faculty-developed tutorials, videos, lesson plans, and instructional use cases. Materials from the website were organized into a 2-week online curriculum that included live demonstrations, guest lectures, "office hours," and virtual faculty meet-ups combined with online tutorials and seminars. Forty participants from 22 different states enrolled in the 2016 workshop. Prior to the workshop, we coordinated with 12 participants to generate novel data sets, totaling 500 trillion nucleotides from seven eukaryotes. The remaining participants used private data sets or sets made available in the Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) by previous participants, identified by exploring detailed faculty profiles on the website. Several Year 1 and Year 2 faculty alumni joined the virtual workshop to share their data and classroom experiences. Nine of 10



Mona Spector (*right*) introduces Ray Enke (2014 RNA-Seq working group alumnus) in the DNALC *Landeau Multimedia Studio* during the virtual workshop.

participants reported they were "satisfied" or "very satisfied" with the virtual workshop—showing the cost-effectiveness of virtual training and of mining existing RNA-Seq data sets to answer novel questions.

Project goals and data were also disseminated through presentations by participating faculty and their students at 12 national or regional scientific conferences in 2016. We started this program with the goal of developing 30 RNA-Seq data sets that looked at novel research problems in eukaryotic genomics. We nearly doubled this output, producing a total of 53 novel data sets that have been integrated into a variety of undergraduate courses. To date, more than 900 students have been taught RNA-Seq by program participants in a variety of courses, including genetics, developmental biology, cell and molecular biology, and horticulture. During the 3 years of the project, 84 undergraduate faculty generated and analyzed RNA-Seq data sets totaling 600 trillion nucleotides from 36 eukaryotes.

CyVerse

During the year, the NSF-funded *iPlant Collaborative* was rebranded as CyVerse to reflect its expanded mission to provide high-level computation across all disciplines of biology. As lead for CyVerse Education, Outreach, and Training (EOT), the DNALC is nurturing the next generation of computationally savvy biologists through workshops, outreach at professional meetings, and support of faculty at primarily undergraduate-serving institutions (PUIs). As a direct indicator of this impact, students registered 70% of 10,000 new CyVerse accounts in 2016: 44% undergraduate, 18% graduate, and 14% precollege.

The DNALC's student-friendly bioinformatics interface, *DNA Subway,* saw a 14% increase to 63,460 visitors and a 25% increase in new projects to 36,045. Among the new registrants were the 500 freshman students at James Madison University (JMU) in Harrisonburg, who used the Blue Line during a semester-long course on DNA barcoding. The barcoding course replaced traditional freshman biology, uniting ecology, cell and molecular biology, and bioinformatics through a biodiversity study in the JMU Arboretum. Likely the largest CURE implementation of DNA barcoding in the country, the JMU course involved a team of 17 instructors teaching more than 20 lab sections.

We expanded our effort to integrate the training practices of *Data Carpentry* and *Software Carpentry* (the *Carpentries*), which focus on manipulating and visualizing data. Using the *Carpentries* approach, DNALC staff trained 366 researchers and educators at 1- and 2-day training events in four countries, including seven CyVerse *Tools and Services* workshops (targeted at researchers), two *Genomics in Education* workshops (targeted at educators), and four *Software* and *Data Carpentry* workshops, as well as other training events. As part of our commitment to diversity, we presented on CyVerse and *Data Carpentry* for research and learning at the 2016 annual meeting of the *Society for the Advancement of Chicanos/Hispanics and Native Americans in Science* (*SACNAS*).

MaizeCODE

At 1.017 billion tons, the worldwide harvest of maize (corn) tops that of rice (738 million tons) and wheat (711 million tons). Maize has assumed this position as the largest staple crop after centuries of careful genetic breeding to enhance many of its growth and nutritional properties. In 2016, the DNALC teamed with scientists at CSHL and New York University on MaizeCODE, a major research project to develop an "encyclopedia" of DNA elements that control gene action and influence the qualities that are selected by corn breeders. The project will generate 150 new data sets of RNA and DNA sequence that will shed light on the structure and function of the corn genome.

MaizeCODE builds on more than a century of pioneering maize research at CSHL. This began with George Shull's breeding of the first hybrid corn in the first decade of the 20th century and continued with Barbara McClintock's Nobel-winning discovery of transposable elements ("jumping genes") in the 1950s. Transposable elements constitute ~85% of the maize genome and contribute to the plant's adaptability under different growing conditions. Modeled after the successful human ENCODE project, MaizeCODE will be a critical resource to help scientists link maize genotypes to phenotypes that impact economic characteristics of maize. For example, changes in global climate and expanded cultivation in developing countries will require modifications of maize genes that govern traits such as drought resistance, salt resistance, and nutrient assimilation. The identification of the genetic variation that controls these and other traits will provide insights to breeders. The DNALC's task is to prepare undergraduate faculty so that they and their students can analyze the new data sets as they are released. Students will have access to virtual machines in the *CyVerse* cloud configured to run specific analyses, including measuring differential expression and splicing, identifying promoters and enhancers, and correlating transcription marks with population studies.

Even though the corn genome sequence was published in 2009, genes and other features have primarily been identified using computer predictions; surprisingly, the vast majority of the corn genome sequence has not been edited, or even carefully looked at, by researchers. The process of editing the genome sequence and adding detailed interpretations—including variations in the control, physical arrangement, and function of genetic information—is termed annotation. So, another key element of the outreach program is to improve the capability of *DNA Subway* for community annotation of the maize genome. The exercise in "genomic crowdsourcing" will provide a common tool for researchers and students to take a detailed look at families of related genes—and to understand how genes are regulated under similar and different situations. For these reasons, we hope that our work on MaizeCODE will show that good research and good education can be exactly the same thing.

Licensing

Since 2003, the DNALC has implemented a licensing program that offers institutions a formal affiliation and transparent access to intellectual property (IP) developed with $40 million in federal and private grants. The licensing program has founded DNA teaching labs at ten institutions, including those described below.

(*Left* to *right*) Bruce Stillman, Qing Yu, Melissa Du, and Dave Micklos during the December visit to Suzhou. Melissa Du stepped in as SIP DNALC Director and Qing Yu as Council Member following the management transfer.

DNALC Asia

In February 2015, CSHL and the Suzhou Industrial Park (SIP) entered into a 10-year agreement to develop DNALC Asia under the umbrella of the successful CSH Asia Meetings. By spring of 2016, renovations were completed on three pod-like structures in the biotechnology district of SIP—one pod is equipped with eight teaching labs and the others house an entrance/exhibit and cafeteria. After months of extensive training for instructional staff at Cold Spring Harbor, the first lab field trip was taught on April 6, Jim Watson's birthday. Later in the year, DNALC staff traveled to Suzhou to assist with the launch of summer DNA camps and to help organize academic-year field trips. Dr. Jessica Talamas

Dave and Bruce at SIP CSH Asia DNALC. DNALC Asia, from CSH Asia Meetings to the SIP Education Society, brings a strong connection to the Education Bureau and local schools.

spent an intensive 6 weeks at CSHL, preparing to take over the position of Education Director. With training and research experience at Johns Hopkins, University of California, San Diego, the Salk Institute, and the University of Pennsylvania, as well as experience coordinating human genome sequencing at the Broad Institute, Jessica brings scientific credibility to DNALC Asia. At the annual meeting of the governing Council, CSHL President Bruce Stillman and SIP Chairman Barry Yang ratified the transfer of management of DNALC Asia from CSH Asia Meetings to the SIP Education Society, bringing a strong connection to the Education Bureau and local schools.

Beijing DNALC

Since 2014, the DNA Learning Center has collaborated with Beijing 166 High School, located in the ancient Dongcheng District. Beijing 166 is a designated "Beacon School" in biology, giving us the chance to help it develop as a model for schools throughout Beijing and China. During the year, we focused on preparing for a large-scale program in DNA barcoding, modeled on our *Urban Barcode Project*. DNA barcoding teaching materials were translated into Chinese, and Beijing 166 faculty conducted summer workshops for students and teachers participating in the citywide science collaboration, Ao-Xiang. Despite snowy weather, 39 Beijing 166 students came to the DNALC in February to participate in courses on *DNA Science* and *DNA Barcoding*. During the summer, 53 middle school students visited for 3 weeks, completing *Fun with DNA*, *World of Enzymes*, *Green Genes,* and *Forensic Science* camps. DNALC instructor Catherine Zhang, a native of X'ian, visited Beijing 166 twice in 2016. In April, she conducted a workshop on *Human DNA Fingerprinting* for 85 high school students and three high school science teachers from the Beijing area. The workshop culminated in a field trip to Zhoukoudian, where they came face-to-face with Peking Man, an example of our ancient relative *Homo erectus*. In October, 119 middle school students and 10 teachers from the Dongcheng School District attended a workshop on *Genome Science*.

University of Notre Dame DNALC

During the academic year, 1800 students in grades 5–12 participated in hands-on labs, with some traveling as long as 2 hours to reach the center! During the summer, day and sleepover camps

Cristina Fernandez-Marco (*kneeling, third from left*) and Dave Micklos (standing, back row second from *right*) with workshop participants at Centro de Innovación y Desarrolio Agroalimentarion de Michoacán (CIDAM, which translates as Center of Innovation and Agri-food Development of Michoacán).

served 130 students in grades 6–10. Led by Dr. Amy Stark, more than 50 undergraduate and graduate volunteers assisted at Notre Dame DNALC labs and functions, including a DNA Day essay contest with prizes at the elementary, middle, and high school levels.

DNALC Mexico

We continued our collaboration with DNA Mexico, a consortium of clinical DNA testing companies, and the Mexican Ministry of Agriculture to develop a network of DNALCs around Mexico. Our key collaborators included Armando Barriguete, Hugo Scherrer, Francisco Gurria, Diego Ulibari, Alonso Sanz, and Manuel Rey. We continued to look for an inaugural site in Morelia, the capital of Michoacán State and the most important agricultural region in Mexico. In August, we visited Morelia to conduct workshops on *Human Molecular Genetics* for 72 high school teachers who came from as far as Mexico City—2.5 hours away. With the teachers, we tested a mass spectrometry system used for forensic DNA phenotyping, which uses DNA variation to predict the hair and eye color of crime scene suspects. Two lab instructors, Humberto Contreras-Cornejo and Ricardo Duran, received 4 weeks of intensive training at the DNALC. We also reconnected with former DNALC staff member Oscar Pineda-Catalan, who helped build the *Urban Barcode Project* in New York City. With plans to return to his native Mexico City, Oscar agreed to help us jump-start a DNA barcoding project in the capital city.

DNALC NYC

Experience from our licensing program—especially running *DNALC West* and *Harlem DNA Lab*—emboldened us to plan a large-scale center in New York City (NYC). In this way, we could extend our successful hands-on science education model to serve the city's vibrant and diverse communities, including substantial numbers of underrepresented minority and disadvantaged

students. After several years of searching sites around NYC, we were offered 17,500 square feet of space at the City University of New York (CUNY) College of Technology (City Tech) campus in downtown Brooklyn. At year's end, we awaited final approval from the CUNY Board of Trustees of our 30-year, no-cost lease of this property. The plan was initially developed with strong support from City Tech President Russ Hotzler and four Vice Chancellors of the CUNY system: Allan Dobrin (Chief Operating Officer), Gillian Small (Research), and Iris Weinshall and Judy Bergtraum (Facilities Planning, Construction, and Management).

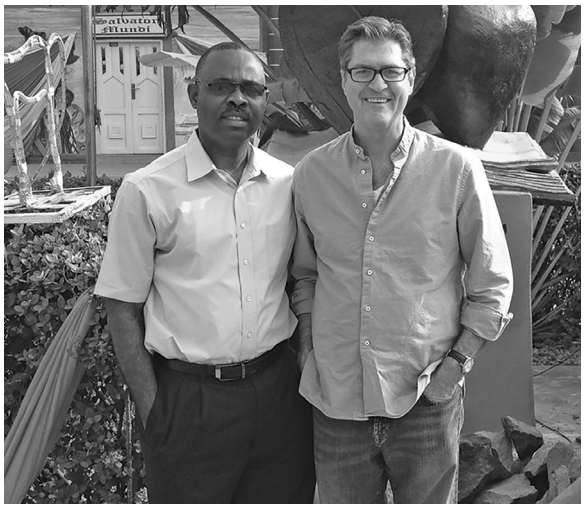
The City Tech site occupies the entire second floor of an academic building on the corner of Adams and Tillary Streets, at the foot of the access ramp to the Brooklyn Bridge. The department of dental hygiene is expected to move out of the space by December 2017, with our renovations beginning in January and ending in late in 2018. A conceptual plan developed by Centerbrook Architects and Planners includes six teaching labs, two bioinformatics labs, prep labs, a lunchroom, and a large exhibit space. Two of the teaching labs will be devoted to course-based research projects by CUNY students during the academic year.

The City Tech location fulfills our key requirements of visibility and ready access to the student populations of New York City. With signage on the façade of the building, the DNALC brand will be immediately visible to all traffic coming off the Brooklyn Bridge. Downtown Brooklyn is one of the most accessible parts of New York City, and there are eight subway lines within several blocks of City Tech. There is ample room for bus drop-offs on both Tillary and Adams Streets. Brooklyn is the fastest growing of the five boroughs and has the largest population of school-age children (~400,000) and students enrolled in public school (~300,000). Brooklyn also has the greatest number of charter schools (more than 50) and highest charter school enrollment (~18,000).

Bolstered by the lead giving of CSHL trustee and mentor Laurie Landeau, by year's end, we had raised $25 million to develop and partially endow the NYC center. Our business model achieves a balanced budget based only on income from tuition and endowment funds currently pledged to the project. The model provides scholarships for all CUNY students participating in research experiences and achieves our goal of providing scholarships for half of all precollege students participating in academic-year lab programs. Additional endowment or grant income will be used for student scholarships and new programs, limiting reliance on external funding.

Helping Out in Nigeria

Since 2013, we have collaborated with Nigerian plant biologist Dr. George Ude to set up a modern biotechnology lab at Godfrey Okoye University (GOU), a small Catholic institution in southeastern Nigeria. The school is located in Enugu, the capital of the Igbo tribe whose attempt to secede from Nigeria resulted in the Biafran Civil War of the 1960s. Enugu Province has never really recovered from the physical destruction of the war, the starvation of an estimated one million civilians, and the diaspora of millions of others. The transportation infrastructure and public administration buildings of Enugu are still in shambles. GOU is an oasis of peace and learning amidst the continuing chaos.



George Ude (*left*) and Dave Micklos at Godfrey Okoye University.

The biotechnology lab is a resource not only for GOU students but also for professors from throughout Nigeria, whose institutions lack modern lab equipment. Faculty training workshops in DNA barcoding held

Michael Okoro looks on as *Fun with DNA* camp students demonstrate DNA extraction techniques.

there in 2014–2015 identified many local plants with new sequence variations or that were new accessions to GenBank; four remain today as the only *rbc*L barcode records for that species. An NSF grant to Bowie State University in Maryland, where Dr. Ude is a full professor, allowed us to implement an exchange program for Nigerian students to do lab projects at Bowie and the DNALC. Michael Okoro was the first exchange student, spending 3 weeks here in 2014 to complete a project on DNA barcoding of indigenous plants of Nigeria.

We invited Michael back in 2016 as our first "DNALC Scholar," with the objective of further developing his lab and teaching skills to prepare him for entry into a U.S. graduate program. Michael arrived in spring, and by fall we had achieved our objective: Michael was accepted to the graduate program in molecular biotechnology at New York University (NYU)! Our excitement quickly turned to dejection when we realized that Michael had no resources at all to support his graduate education. Nigeria is in the grip of a continuing financial crisis, during which the national currency, the Naira, has lost 50% of its value since 2014. As retired civil servants, Michael's parents have not received any retirement payments during this time. Although Michael's university had promised him a scholarship for graduate school, they have had to lay off staff and could not afford to help.

The solution came, at the last moment, through an NSF training supplement to MaizeCODE. Under the supplement, Michael will be mentored by three MaizeCODE co-PIs: Dave Micklos and Doreen Ware at CSHL, and Ken Birnbaum, who fortuitously is also director of the NYU Ph.D. program and cellular biology core. Michael will split his time between CSHL and NYU, weaving maize cell genetics, bioinformatics, and genome analysis into his graduate program. As part of the outreach effort, he will assist undergraduate faculty in using MaizeCODE for student research projects. The supplement will pay Michael's NYU tuition, room, and board at CSHL and travel between the two sites.

Michael's traineeship aligns with NSF's Basic Research to Enable Agricultural Development (BREAD) program, funded in partnership with the Bill and Melinda Gates Foundation. The activities of MaizeCODE support BREAD's objective to develop genomics resources for breeding drought-, heat-, and disease-resistant varieties that can grow under the adverse conditions found throughout much of sub-Saharan Africa. Another important goal is to develop varieties that tolerate low-fertility conditions in the fields of smallholding farmers who cannot afford nitrogen fertilizers. (Smallholders, small plots that rely mainly on family labor, constitute ~80% of farmers in Nigeria and produce 98% of all food consumed there.) Michael's participation in MaizeCODE will, thus, help translate genetic gains into practical gains in the fields of family farms in Africa.

Breakthrough Prize

In 2015, we teamed with the *Breakthrough Junior Science Challenge*, in which precollege students produce short online videos that "explain a big scientific idea." The winner receives a $250,000 scholarship and the inspiring science teacher receives $50,000. Our task is to administer $100,000 to design and equip a *Breakthrough Science Lab* for the winning school. The first winner was Ryan Chester from North Royalton High School, outside of Cleveland, Ohio (https://breakthroughjuniorchallenge.org/winners/2015). During 2016, we worked with North Royalton teachers and administrators to design a lab with the look and feel of a DNALC lab—including our signature lab desks. In preparation for the new lab, we provided 3 days of teacher training in recombinant DNA technology and human

DNA polymorphisms. In the midst of planning with North Royalton, we received word of a crop of three 2016 winners from around the world (https://breakthroughjuniorchallenge.org/winners): Deanna See (Raffles Girls School, Singapore); Antonella Masini (Cambridge College, Lima, Peru); and Hillary Diane Andales (Philippine Science High School, Eastern Visayas, Philippines). The Philippine school is still recovering from a 7.2 magnitude earthquake and Super Typhoon Yolanda that devastated the campus in December 2013 and forced the relocation of students in 2014.

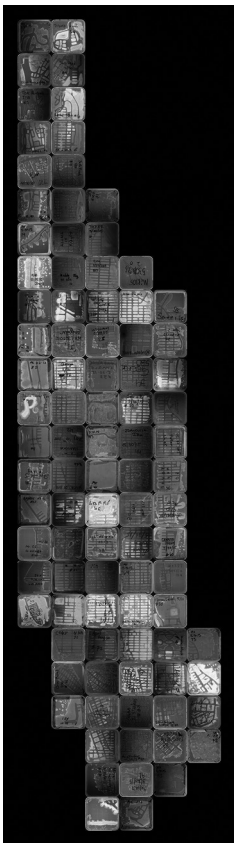
Student Programs

In 2016, 20,784 students participated in labs during field trips to our three facilities: Dolan DNA Learning Center, DNALC *West,* and *Harlem DNA Lab*. An additional 9232 students completed labs in schools led by DNALC staff, and 1451 students attended week-long summer camps. Footlocker kits were used by 1229 students, including 404 conducting DNA barcoding research through *UBP*, *UBRP*, or *BLI*. Grants from Bank of America and the National Grid Foundation supported programs for more than 2400 underserved students from Long Island public schools, including Brentwood, Malverne, Uniondale, William Floyd, Central Islip, Roosevelt, and Valley Stream. An additional 24 students received scholarships to attend *Fun with DNA* at the New World Prep Charter School in Staten Island, and 15 students from IS 59 in Queens participated in an ongoing collaboration with Northwell Health at DNALC *West*. Of the 2642 students who visited the *Harlem DNA Lab,* 73% received scholarships, partially funded by the William Townsend Porter Foundation. The NYC Department of Education (DOE) Office of School Programs and Partnerships subsidized camps at *Harlem DNA Lab* and the DOE's Environmental Study Center in Brooklyn for 45 students.

We continued long-standing partnerships to co-teach courses in *Molecular and Genomic Biology* at Cold Spring Harbor High School (12th year) and St. Dominic High School in Oyster Bay (5th year). These college-level experiences include research projects on bacterial genetics, DNA barcoding, bioinformatics, RNA interference (RNAi), CRISPR, human and plant genomics, and microbiomes. Five of 12 graduates of the CSH course intend to pursue undergraduate degrees in either biology or computer science. We also continued our collaboration with the CSHL Watson School of Biological Sciences to provide graduate students a teaching rotation during their first year. The 3-month training program includes observations and co-teaching that help the students develop effective teaching techniques as they work toward the final goal of independent teaching.

The NYC Partner Membership Program continued to provide custom programs to six independent schools in the tri-state region. The St. David's School curriculum came full circle this year, when eighth graders, who participated in the school's pilot year as fifth graders, surveyed the biodiversity of Cold Spring Harbor by barcoding plants and invertebrates collected on a field trip. The Chapin School, our longest-standing partner of 6 years, expanded its curriculum to include next-generation sequencing to study microbiomes. Similarly, the DNA barcoding teams from Convent of the Sacred Heart analyzed microbiomes from airplanes and parks. The eighth grade research students at Lycée Français de New York made formal presentations, including videos, of their barcoding research to classmates and parents. Marymount students also conducted DNA barcoding and genome science labs as part of a Molecular Genetics class. Riverdale Country School offered lab sequences for seventh, ninth, and 11th grade students, including labs on antibiotic resistance, gel electrophoresis, and DNA sequencing.

We joined with other educational groups in expanding STEM to STEAM (Science, Technology, Engineering, *Art*, and Math). As part of our collaboration with *Genspace*, 50 citizen scientists produced an "agar art" map of NYC made with glowing bacteria. Participants painted suspensions of bacteria on square Petri dishes prepared with stenciled portions of Manhattan's grid map, including outlines of streets, parks, and waterways. The resulting *NYC Biome MAP* was presented at the New Museum's Ideas City Festival NYC. The artwork came in second



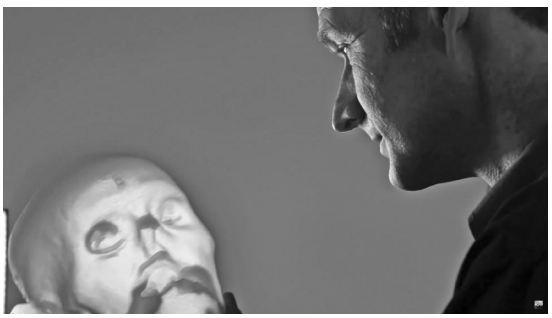
in the American Society of Microbiology (ASM) International Agar Art contest and was featured on the cover of ASM's first Agar Art Calendar. The project was covered in 160 news articles, including BBC, CNN, *USA Today*, and *Huffington Post.* At the prestigious Aspen Ideas Festival, 120 festival attendees worked to create collaborative microbial artwork depicting Aspen's iconic Maroon Bells mountain peaks. Half of the eight *Saturday DNA!* sessions, which drew 171 participants, also had a STEAM focus. Participants tried their hands at agar art, worked on projects with BOLD exhibit artist Joe Rossano, and made watercolor "portraits" of cells and their organelles. Other sessions explored what Ötzi carried in his first aid kit, how doctors are using molecular diagnostics, and how to isolate a "glowing" protein from bacteria.

Finishing touches were added to the Ötzi exhibition, and student guides and hands-on lab activities were developed in time to launch public tours that coincided with PBS's debut broadcast of NOVA *Iceman Reborn* on February 17. The documentary chronicled Gary Staab, a leading artist specializing in natural history and prehistoric life models for museums, as he created a three-dimensional (3D) replica of the Ötzi mummy and installed it at the DNALC. During the year, Ötzi drew 5000 visitors and an additional 400 registered for weekend interpretive tours, which included a new *Pollen Tells a Story* lab. Participants used compound microscopes to identify the types of pollen found in different parts of Ötzi's intestinal tract, deducing where he had been in the last 36 hours of his life: Pollen from flowering plants indicates lower elevations, whereas pollen from evergreen plants indicates higher elevations.

With sadness, *Long Island Discovery* ended its 23-year run at the DNALC, when Cablevision was purchased by Altice, a multinational telecoms group based in the Netherlands. Cablevision developed this 28-minute "electronic field trip" to help students explore the rich history and landscapes of America's longest offshore island. In 1993, Cablevision provided major funding to renovate our antique auditorium into a state-of-the-art "multitorium" with tiered seating, digital projection, and surround sound to support the multiscreen production. More than 100,000 visitors saw *Long Island Discovery* at the DNALC, including large numbers of fourth graders who study Long Island history as part of their curriculum. The multitorium continues to be used for showings of the short films *Cell Signals* and *Secret of Life* and presentations by students and scientists. Plans are under way to reclaim the rear projection booth to return a bank of windows and 14-foot ceilings to the lunchroom.

BioMedia

A record number of visitors accessed our suite of multimedia resources in 2016. Total website, YouTube, and smartphone/tablet app visitation reached more than 8 million, a 6.6% increase over 2015. Google Analytics counted 5,437,407 visits to 24 DNALC websites, just below the prior year's total. Our YouTube videos received 959,294 views (up 8.9% over 2015), and the *3D Brain*, *Weed to*



Wonder, and *Gene Screen* apps were downloaded 1,617,620 times (up 37%!). In-app purchases of 3D Brain HQ netted $8600.

The *BioMedia* Group worked hard to merge the three student barcoding sites—*UBP, UBRP*, and *BLI*—with online resources to support student research into *DNA Learning Center Barcoding 101* (http://www.dnabarcoding101.org). The updated site includes background on DNA barcoding and detailed instructions for isolating and amplifying DNA samples. Enhanced teaching resources include shared documents, slide sets, animations, videos, references, and links to other online resources. The bioinformatics tool, *DNA Subway*, was upgraded to support fungal (ITS) and microbial (16S rRNA) barcodes. Project management was improved with a cleaner and more intuitive user interface to allow teachers to enroll teams, review proposals, track team progress, register for Open Labs, and request equipment footlockers. All team information and proposals are now linked to sample and sequence information. A new mapping tool searches student barcode sequences by homology, habitat, and taxonomy. A symposium section provides a long-term record of student accomplishments—including project summaries and posters—which can be used to support college applications and mentor funding requests.

The RNA-Seq project website (http://www.rnaseqforthenextgeneration.org) provides an easy-to-use interface to an extensive bank of faculty-developed tutorials and videos, lesson plans, and instructional use cases. This format will allow the project to be inexpensively sustained. In addition, detailed faculty profiles provide ready templates to involve additional classes in research on the data sets developed during the project, including project abstract, research questions, experimental details, data set description, and teaching resources. The *BioMedia* Group supported the transition of the RNA-Seq project to virtual training by blending these website resources with synchronous (webinars, discussions, and instructor "office hours") and asynchronous components (seminars and tutorials). Streaming video, produced in the *Landeau Multimedia Studio*, was crucial to the online format. Using the Adobe Connect environment, 19 speakers presented lectures as if in a classroom, with DNALC staff switching between the presenter and a laptop displaying a slide show or live web demonstration. Adobe Connect also provided links and downloadable presentation materials, participant polls, a virtual whiteboard or notepad, screen sharing, and live chat to answer questions in real time. As a video alternative, we simultaneously streamed the webinars to the DNALC Livestream website, where the session was immediately available on demand for later viewing or review.

We have been fortunate to have Joseph Rossano's *BOLD the barcode of life* art exhibit at the DNALC since 2014. In the spring, *BioMedia* staff worked with DNALC educators to develop a *Saturday DNA!* session called "A BOLD Connection." Participants learned about the science of DNA barcoding and how it gives us a deeper understanding of the natural world, and then they explored with Joe how art can bridge the gap between scientists and the general public. Participants wrapped up the session by creating their own Joe-Rossano-inspired biodiversity art. The BOLD exhibition catalog was published in the fall. Seven essays—by artists, researchers, educators, and photographers—provide insight into the art and the barcoding science that inspired it. Unique to an art catalog, each "object" is described with a species name, Barcode of Life Database identification number, and DNA barcode sequence.

Staff and Interns

During the year, we built up new administrative and instructional capabilities. Jason Williams took on a new role as Assistant Director for External Collaboration to help with the management of our growing number of licensed centers. The position is a natural extension of his role as Education, Outreach, and Training Lead for CyVerse, which offers numerous training events around the country and world. Jason jumped in to support the start-up of DNALC Asia, and he will see more duty as we develop multiple centers in Mexico and Nigeria. Alison Cucco joined the New York City staff as lead teacher at *Harlem DNA Lab.* Her mentorship role in the NYC barcoding projects makes good use of her former faculty position at St. Francis College and her biology master's from Fordham University, where she investigated the effects of urbanization on ecosystems and plant productivity. Hired as a college intern in 2015, Keil Thomas was promoted to Middle School Educator after graduating with a bachelor's degree in biomolecular science from the NYU School of Engineering. Keil also manages the footlocker loan program for *Barcode Long Island*.

Alison Cucco

Keil Thomas

We were also saddened to say goodbye to several key staff members. Dr. Mona Spector returned to the CSHL main campus to continue her research on leukemia in Mike Wigler's lab. We were lucky when, in 2014, Mona decided to take a sabbatical from research to administer our NSF RNA-Seq project, the most technically advanced biochemistry and bioinformatics work we had ever attempted. Drawing on her extensive research experience with RNA preparation and NGS sequencing, she developed wet lab protocols and oversaw the production of more than 170 RNA-Seq transcriptomes for workshop participants. She also worked with the *BioMedia* staff to develop the project website and to improve the functionality of *DNA Subway's* Green Line. Pan Teng left to continue her career as a software engineer at Memorial Sloan Kettering Cancer Center. Pan joined the DNALC in 2015, increasing processing speed and expanding the organization of large-scale genomics data in *DNA Subway* and implementing GPS solutions for the *Barcode Long Island* website. Paul Donat joined the Ecology and Evolution Ph.D. program at SUNY Stony Brook in August, focusing on plant genome evolution. Beginning as a college intern in 2012, Paul became our first full-time lab technician in 2015 and was quickly promoted to instructor.

The DNALC relies on high school and college interns to support our day-to-day operations. An internship offers students the unique opportunity to gain real laboratory or design experience in an educational environment. The *BioMedia* Group also welcomes interns for summer or longer-term roles. In 2016, we worked with an amazing group of interns and said farewell as some left for college.

High School Interns

Duardo Akerle, Half Hollow Hills High School
Alyssa DiArrigo, Bethpage High School
Alec Haber, Syosset High School
Derek Lee, Elwood–John H. Glenn High School
Rahul Ranjan, Hicksville High School
Ben Rhee, Syosset High School
Bijia Wang, Syosset High School
GraceAnne Woods, Glen Cove High School

High School Interns Departing for College

Brady Anna, University of North Carolina Chapel Hill
Juliana Eastment, University of Richmond
John Messina, Hofstra University
Stefanie Montalbano, Fairfield University
Gabrielle Ramirez, University of Pennsylvania
Maria Urbina, Tufts University

College Interns

Benjamin Acosta, Boston College
Kathryn Bellissimo, The College of New Jersey
Gabrielle Blazich, Fordham University
Abigail Buckley, Dartmouth College
Brittany Coscio, College of the Holy Cross
Marie Jean Francois, City University of New York City College
Omotayo Ikuomenisan, Hunter College
Michaela Lee, State University of New York, Oneonta
William McBrien, Suffolk Community College
Pauline McGlone, Hunter College
Breanna Tahany, State University of New York, Binghamton

Sites of Major Faculty Workshops

Program Key: *Middle School* High School **College**

State	*Institution*	*Year(s)*
ALABAMA	University of Alabama, Tuscaloosa	1987–1990
	Hudson Alpha Institute, Huntsville	**2014**
ALASKA	**University of Alaska, Anchorage**	**2012**
	University of Alaska, Fairbanks	**1996**
ARIZONA	Arizona State University, Tempe	2009
	Tuba City High School	1988
	University of Arizona, Tucson	**2011**
	United States Department of Agriculture, Maricopa	**2012**
ARKANSAS	Henderson State University, Arkadelphia	1992
	University of Arkansas, Little Rock	**2012**
CALIFORNIA	**California State University, Dominguez Hills**	**2009**
	California State University, Fullerton	**2000**
	California State University, Long Beach	**2015**
	California Institute of Technology, Pasadena	**2007**
	Canada College, Redwood City	1997
	City College of San Francisco	2006
	City College of San Francisco	**2011, 2013**
	Community College of Denver	**2014**
	Contra Costa County Office of Education, Pleasant Hill	2002, 2009
	Foothill College, Los Altos Hills	**1997**
	Harbor-UCLA Research & Education Institute, Torrance	2003
	Los Angeles Biomedical Research Institute (LA Biomed), Torrance	**2006**
	Laney College, Oakland	1999
	Lutheran University, Thousand Oaks	1999
	Oxnard Community College, Oxnard	2009
	Pasadena City College	**2010**
	Pierce College, Los Angeles	**1998**
	Salk Institute for Biological Studies, La Jolla	2001, 2008
	San Francisco State University	**1991**
	San Diego State University	**2012**

	San Jose State University	**2005**
	Santa Clara University	2010
	Southwestern College, Chula Vista	**2014–2015**
	Stanford University, Palo Alto	**2012**
	University of California, Berkeley	**2010, 2012**
	University of California, Davis	1986
	University of California, Davis	**2012, 2014–2015**
	University of California, Long Beach	**2015**
	University of California, Northridge	**1993**
	University of California, Riverside	2011
	University of California, Riverside	**2012**
	University of California, San Francisco	**2015**
COLORADO	Aspen Science Center	2006
	Colorado College, Colorado Springs	1994, 2007
	Colorado State University, Fort Collins	**2013**
	Community College of Denver	**2014**
	United States Air Force Academy, Colorado Springs	**1995**
	University of Colorado, Denver	1998, 2009–2010
CONNECTICUT	Choate Rosemary Hall, Wallingford	1987
DELAWARE	**Jackson Laboratory, Farmington**	**2016**
	University of Delaware, Newark	**2016**
DISTRICT OF COLUMBIA	**Howard University, Washington**	**1992, 1996, 2009–2010**
FLORIDA	Armwood Senior High School, Tampa	1991
	Florida Agricultural & Mechanical University, Tallahassee	2007–2008
	Florida Agricultural & Mechanical University, Tallahassee	**2011**
	Florida SouthWestern State University, Fort Myers	**2015**
	North Miami Beach Senior High School	1991
	Seminole State College, Sanford	**2013–2014**
	University of Miami School of Medicine	2000
	University of Western Florida, Pensacola	1991
GEORGIA	Fernbank Science Center, Atlanta	1989, 2007
	Gwinnett Technical College, Lawrenceville	**2011–2012**
	Morehouse College, Atlanta	**1991, 1996–1997**
	Spelman College, Atlanta	**2010**
	University of Georgia, Athens	**2015**
HAWAII	Kamehameha Secondary School, Honolulu	1990
	University of Hawaii at Manoa	**2012**
IDAHO	University of Idaho, Moscow	1994
ILLINOIS	Argonne National Laboratory	1986–1987
	iBIO Institute/Harold Washington College, Chicago	2010
	Illinois Institute of Technology, Chicago	2009
	Kings College, Chicago	**2014**
	University of Chicago	**1992, 1997, 2010**
	University of Southern Illinois, Carbondale	**2016**
INDIANA	Butler University, Indianapolis	1987
	Purdue University, West Lafayette	**2012**
IOWA	Drake University, Des Moines	1987
KANSAS	University of Kansas, Lawrence	1995
KENTUCKY	**Bluegrass Community & Technical College, Lexington**	**2012–2014**
	Murray State University	1988
	University of Kentucky, Lexington	1992
	Western Kentucky University, Bowling Green	1992
LOUISIANA	Bossier Parish Community College	2009
	Jefferson Parish Public Schools, Harvey	1990
	John McDonogh High School, New Orleans	1993
	Southern University at New Orleans	**2012**
MAINE	Bates College, Lewiston	1995
	Southern Maine Community College	**2012–2013**
	Foundation for Blood Research, Scarborough	2002
MARYLAND	Annapolis Senior High School	1989
	Bowie State University	**2011, 2015**

	Frederick Cancer Research Center	1995
	McDonogh School, Baltimore	1988
	Montgomery County Public Schools	1990–1992
	National Center for Biotechnology Information, Bethesda	2002
	St. John's College, Annapolis	*1991*
	University of Maryland, School of Medicine, Baltimore	**1999**
MASSACHUSETTS	**Arnold Arboretum of Harvard University, Roslindale**	**2011**
	Beverly High School	1986
	Biogen Idec, Cambridge	2002, 2010
	Boston University	**1994, 1996**
	CityLab, Boston University School of Medicine	1997
	Dover-Sherborn High School, Dover	1989
	Randolph High School	1988
	The Winsor School, Boston	1987
	Whitehead Institute for Biomedical Research, Cambridge	2002
MICHIGAN	Athens High School, Troy	1989
	Schoolcraft College, Livonia	**2012**
MINNESOTA	**American Society of Plant Biologists, Minneapolis**	**2015**
	Minneapolis Community and Technical College, Madison	2009
	Minneapolis Community and Technical College, Madison	**2013**
	University of Minnesota, St. Paul	2005
	University of Minnesota, St. Paul	**2010**
MISSISSIPPI	Mississippi School for Math & Science, Columbus	1990–1991
	Rust College, Holly Springs	**2006–2008, 2010**
MISSOURI	**St. Louis Science Center**	**2008–2010**
	Stowers Institute for Medical Research, Kansas City	2002, 2008
	University of Missouri, Columbia	**2012**
	Washington University, St. Louis	**1989, 1997, 2011**
MONTANA	**Montana State University, Bozeman**	**2012**
NEBRASKA	**University of Nebraska-Lincoln, Lincoln**	**2014**
NEVADA	University of Nevada, Reno	1992, **2014**
NEW HAMPSHIRE	Great Bay Community College, Portsmouth	2009
	New Hampshire Community Technical College, Portsmouth	**1999**
	St. Paul's School, Concord	1986–1987
NEW JERSEY	Coriell Institute for Medical Research, Camden	2003
	Raritan Valley Community College, Somerville	2009
NEW MEXICO	Biolink Southwest Regional Meeting, Albuquerque	2008
	Santa Fe Community College, Santa Fe	**2015**
NEW YORK	Albany High School	1987
	American Museum of Natural History	2007, 2015
	Bronx High School of Science	1987
	Brookhaven National Laboratory, Upton	2015–2016
	Canisius College, Buffalo	2007
	Canisius College, Buffalo	**2011**
	City College of New York	2012
	Cold Spring Harbor High School	1985, 1987
	Cold Spring Harbor Laboratory	**2014–2015**
	Columbia University	**1993**
	Cornell University, Ithaca	**2005**
	DeWitt Middle School, Ithaca	*1991, 1993*
	Dolan DNA Learning Center	1988–1995, 2001–2004, 2006–2009, 2015–2016
	Dolan DNA Learning Center	**1990, 1992, 1995, 2000–2011**
	Dolan DNA Learning Center	*1990–1992*
	DNA Learning Center *West*	2005
	Environmental Science Center, Bergen Beach, Brooklyn	2015–2016
	Fostertown School, Newburgh	*1991*
	Harlem DNA Lab, East Harlem	2008–2009, 2011–2013, 2016
	***Harlem DNA Lab*, East Harlem**	**2015–2016**
	Huntington High School	1986
	Irvington High School	1986

	John Jay College of Criminal Justice	2009
	Junior High School 263, Brooklyn	*1991*
	Lindenhurst Junior High School	*1991*
	Mount Sinai School of Medicine	1997
	Nassau Community College, Garden City	**2013**
	New York Botanical Garden, Bronx	**2013**
	New York City Department of Education	2007, 2012
	New York Institute of Technology	2006
	New York Institute of Technology	**2006**
	Orchard Park Junior High School	*1991*
	Plainview-Old Bethpage Middle School	*1991*
	State University of New York, Purchase	1989
	State University of New York, Stony Brook	1987–1990, 2015, 2016
	State University of New York, Stony Brook	**2014, 2016**
	Stuyvesant High School	1998–1999
	The Rockefeller University	2003, 2015–2016
	The Rockefeller University	**2010**
	Titusville Middle School, Poughkeepsie	*1991, 1993*
	Trudeau Institute, Saranac Lake	2001
	Union College, Schenectady	2004
	United States Military Academy, West Point	**1996**
	Wheatley School, Old Westbury	1985
NORTH CAROLINA	CIIT Center for Health Research, Triangle Park	2003
	North Carolina Agricultural & Technical State University, Greensboro	**2006–2007, 2009–2011**
	North Carolina School of Science, Durham	1987
	North Carolina State University, Raleigh	**2012**
NORTH DAKOTA	**North Dakota State University, Fargo**	**2012**
OHIO	Case Western Reserve University, Cleveland	1990
	Cleveland Clinic	1987
	Langston University, Langston	2008
	North Westerville High School	1990
	The Ohio State University, Wooster	**2016**
OKLAHOMA	**Oklahoma City Community College**	**2000**
	Oklahoma City Community College	2006–2007, 2010
	Oklahoma Medical Research Foundation, Oklahoma City	2001
	Oklahoma School of Science and Math, Oklahoma City	1994
	Tulsa Community College, Tulsa	2009
	Tulsa Community College, Tulsa	**2012–2014**
OREGON	Kaiser Permanente-Center for Health Research, Portland	2003
	Linfield College, McMinnville	**2014**
PENNSYLVANIA	Duquesne University, Pittsburgh	1988
	Germantown Academy	1988
	Kimmel Cancer Center, Philadelphia	2008
RHODE ISLAND	Botanical Society of America, Providence	2010
SOUTH CAROLINA	**Clemson University**	**2004, 2015**
	Medical University of South Carolina, Charleston	1988
	University of South Carolina, Columbia	**1988**
SOUTH DAKOTA	**South Dakota State University, Brookings**	**2015**
TENNESSEE	NABT Professional Development Conference, Memphis	2008
TEXAS	**Austin Community College–Rio Grande Campus**	**2000**
	Austin Community College–Eastview Campus–Roundrock Campus	**2007–2009, 2013**
	Austin Community College–Roundrock Campus	**2012**
	Houston Community College Northwest	**2009–2010**
	J.J. Pearce High School, Richardson	1990
	Langham Creek High School, Houston	1991
	University of Lone Star College, Kingwood	**2011**
	Midland College	2008

	Southwest Foundation for Biomedical Research, San Antonio	2002
	Taft High School, San Antonio	1991
	Texas A&M University, College Station	**2013**
	Texas A&M University, Prairie View	**2013**
	Texas A&M, AG Research and Extension Center, Weslaco	2007
	Trinity University, San Antonio	**1994**
	University of Texas, Austin	**1999, 2004, 2010, 2012**
	University of Texas, Brownsville	**2010**
UTAH	**Brigham Young University, Provo**	**2012**
	University of Utah, Salt Lake City	1993
	University of Utah, Salt Lake City	**1998, 2000**
	Utah Valley State College, Orem	**2007**
VERMONT	University of Vermont, Burlington	1989
VIRGINIA	Eastern Mennonite University, Harrisonburg	1996
	Jefferson School of Science, Alexandria	1987
	Mathematics and Science Center, Richmond	1990
	Mills Godwin Specialty Center, Richmond	1998
	Virginia Polytechnic Institute and State University, Blacksburg	**2005, 2008–2009**
WASHINGTON	Fred Hutchinson Cancer Research Center, Seattle	1999, 2001, 2008
	Shoreline Community College	2011, 2012
	University of Washington, Seattle	**1993, 1998, 2010**
WEST VIRGINIA	Bethany College	1989
WISCONSIN	Blood Center of Southeastern Wisconsin, Milwaukee	2003
	Madison Area Technical College/Madison Area College	**1999, 2009, 2011–2014**
	Marquette University, Milwaukee	1986–1987
	University of Wisconsin, Madison	1988–1989
	University of Wisconsin, Madison	**2004, 2012**
WYOMING	University of Wyoming, Laramie	1991
PUERTO RICO	**Universidad del Turabo, Gurabo**	**2011, 2012, 2014**
	University of Puerto Rico, Mayaguez	1992
	University of Puerto Rico, Mayaguez	**1992**
	University of Puerto Rico, Rio Piedras	**1993**
	University of Puerto Rico, Rio Piedras	1994

AUSTRALIA	Walter and Eliza Hall Institute and University of Melbourne	1996
	EMBL/Australian Bioinformatics Resource, University of Melbourne	**2016**
AUSTRIA	**Vienna Open Lab, Vienna**	**2007, 2012**
CANADA	Red River Community College, Winnipeg, Manitoba	1989
CHINA	Beijing No. 166 High School, Beijing	2013, 2014–2016
	Ho Yu College, Hong Kong	**2009**
DENMARK	**Faroe Genome Project, Torshavn, Faroe Islands**	**2013**
GERMANY	Urania Science Center, Berlin	2008
IRELAND	**European Conference on Computational Biology/Intelligent System for Molecular Biology Conference, Dublin**	**2015**
ITALY	International Institute of Genetics and Biophysics, Naples	1996
	Porto Conte Research and Training Laboratories, Alghero	1993
MEXICO	ADN Mexico, Morelia	2016
	ASPB Plant Biology, Merida	2008
	Langebio/Cinvestav, Irapuato	**2016**
NIGERIA	Godfrye Okoye University, Enugu, Nigeria	2013
RUSSIA	Shemyakin Institute of Bioorganic Chemistry, Moscow	1991
SINGAPORE	National Institute of Education	2001–2005
SOUTH AFRICA	Singapore Science Center	2013
	North-West University, Potchefstroom	**2016**
	South African Bioinformatics Society, Durban	**2016**

SWEDEN	Kristineberg Marine Research Station, Fiskebackgkil	1995
	Uppsala University	2000
THE NETHERLANDS	**International Chromosome Conference, Amsterdam**	**2007**
	Wageningen University and Research Center, Wageningen	**2014**
UNITED KINGDOM	**The Genome Analysis Center, Norwich**	**2015**
	Wellcome Trust Conference Center, Hinxton	**2012–2013**
	University of Warwick, Coventry	**2013**

Workshops, Meetings, Collaborations, and Site Visits

January 7–8	NSF Data Science, Learning, and Applications to Biomedical & Health, "A Vision for Collaborative Training Infrastructure for Bioinformatics," New York Academy of Sciences, New York
January 9	*Saturday DNA!* "BioArt," DNALC *Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
January 9–12	International Plant and Animal Genome XXIV Conference 2016, "*DNA Subway*: Educational Challenger of Biological Big Data; Genomics in Education: DNA and Senses; Cultivating Broader Impacts," San Diego, California
January 12	NIH *Barcode Long Island* Open Lab, DNALC *West Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
January 19–20	NSF CyVerse *Data Carpentry Genomics* Workshop, Stony Brook University, New York
January 21	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
Jan 25–Feb 5	*DNA Barcoding* and *DNA Science* Workshops with students from Beijing 166 School, China, DNALC
January 28	NSF *Infrastructure and Training to Bring Next-Generation Sequence (NGS) Analysis into Undergraduate Education* Video Conference, DNALC
February 1	ExpandED Professional Development Workshop, *DNA Structure and Function*, *Harlem DNA Lab*
February 1–2	NSF CyVerse *Genomics in Education* Workshop, University of Delaware, Newark
February 2	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
February 2–5	NSF CyVerse *Software Carpentry* Workshop, New York Academy of Sciences, New York
February 3	NYCDOE "DNA Extraction and Bacterial Transformation," STEM Professional Learning, New World Preparatory Charter School, Staten Island, New York
	Site visit by Angelika Fleckinger, South Tyrol Museum of Archaeology, Bolzano, Italy, Emlyn Kostner, North Carolina Museum of Natural Sciences, Raleigh, and Heinz Reese, Museum Partners Consulting, LLC, Morristown, New Jersey, DNALC
February 3–4	NSF CyVerse *Tools and Services* Workshop, University of Delaware, Newark
February 6	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	NIH *Barcode Long Island* Open Lab, DNALC
	Saturday DNA! "Ötzi's First Aid Kit," DNALC
February 10	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
February 11	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
February 16	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
February 16–19	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
February 21	*Ötzi the Iceman* Tour, DNALC
February 25	Site visit by Dr. Armando Barriguete and Hugo Scherer, Mexico City, Mexico
February 27	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
	NIH *Barcode Long Island* Open Lab, DNALC
	Ötzi the Iceman Tour, DNALC
	STEP STEM Expo 2016, *Urban Barcode Research Program* booth, Bronx Community College, New York
March 2	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
March 3	Professional Development Workshop, Michael J. Petrides School, Staten Island, New York
March 3–5	NSF Mid-South Computational Biology and Bioinformatics Conference, "Overview of CyVerse," University of Memphis, Tennessee
March 5	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	Ötzi the Iceman Tour, DNALC
March 7–18	*Human Genome Science* Workshop, Beijing 166 School, China
March 8	*Ötzi the Iceman* Tour, DNALC
March 9	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
March 12	*Saturday DNA!* "Protein Aglow," DNALC

March 16	*Ötzi the Iceman* Tour, DNALC
March 17	NSF CyVerse Livestream Webinar, *Tools and Services* Workshop, The Ohio State University, Wooster
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
March 17–18	NSF CyVerse *Tools and Services* Workshop, The Ohio State University, Wooster
March 18	City College of New York NOYCE Teacher Academy, *DNA Extraction*, *Harlem DNA Lab*
March 19	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
	NIH *Barcode Long Island* Open Lab, DNALC
	Ötzi the Iceman Tour, DNALC
March 22	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
March 23	NSF "CyVerse Education, Outreach, and Training," iDigBio Webinar Series, DNALC
March 24	CSHL Public Presentation, "Asking the Wrong Questions about American Science Education," CSHL
March 29	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
March 29–30	NSF CyVerse *Genomics in Education* Workshop, Southern Illinois University, Carbondale
March 30	*Ötzi the Iceman* Tour, DNALC
March 31	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
Mar 31–Apr 1	NSF CyVerse *Tools and Services* Workshop, Southern Illinois University, Carbondale
April 2	NIH *Barcode Long Island* Open Lab, DNALC
	Ötzi the Iceman Tour, DNALC
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
April 6	Pinkerton Teen SciCafe and *Urban Barcode Research Program* Mentor Networking Event, American Museum of Natural History, New York
April 7	Site visit by Camille Santistevan, CUNY Advanced Science Research Center, New York, *Harlem DNA Lab*
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
April 9	NIH *Barcode Long Island* Open Lab, Stony Brook University, New York
	Ötzi the Iceman Tour, DNALC
April 11	NSF CyVerse Training Day, European Molecular Biology Laboratory/Australian Bioinformatics Resource, University of Melbourne, Australia
April 11–12	Cold Spring Harbor First Grade Science Fair, "Seaweed Surprise and Fingerprints," CSHL
April 12	NSF CyVerse Training Day, CyVerse and Atmosphere Overview, University of Melbourne, Australia
	Site visit by Regeneron Pharmaceuticals Planners, DNALC *West*
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
April 13	*Ötzi the Iceman* Tour, DNALC
April 16	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
	Ötzi the Iceman Tour, Long Island Real Estate Group, DNALC
	Saturday DNA!, "A BOLD Connection," DNALC
	34th Annual Conference of Empire State Association of Two Year College Biologists, *DNA Barcoding* Workshop, Middletown, New York
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
April 20	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
April 21	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
April 23	*Ötzi the Iceman* Tour, DNALC
April 25	*Ötzi the Iceman* Tour, DNALC
April 25–29	Pinkerton *DNA Barcoding* Workshop, The Rockefeller University, New York
April 26	CSHL *Gramene*: *A Resource for Comparative Plant Genomics* Webinar, DNALC
April 27	*Ötzi the Iceman* Tour, DNALC
April 27–29	American Association for the Advancement of Science: "Future of Undergraduate Stem Education: Research and Practice," Washington, D.C.
April 30	NIH *Barcode Long Island* Open Lab, Stony Brook University, New York
	Ötzi the Iceman Tour, DNALC
May 3	*Urban Barcode Project/Urban Barcode Research Program* Open Lab, *Harlem DNA Lab*
May 5	*Ötzi the Iceman* Tour, DNALC
May 7	NIH *Barcode Long Island* Open Lab, Stony Brook University, New York
	NIH *Barcode Long Island* Open Lab, DNALC
	Urban Barcode Project/Urban Barcode Research Program Open Lab, *Harlem DNA Lab*
May 9–12	NIH SEPA SciEd Conference, *Barcode Long Island: Exploring Biodiversity in a Unique Urban Landscape*, Washington, D.C.
May 11	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York

May 13	TEDxYouth at the Browning School *Biodiversity Day*, "DNA Barcoding," Browning School, New York
May 14	NIH *Barcode Long Island* Open Lab, Brookhaven National Laboratory, Upton, New York
May 17	New York Department of Education *Stemtastic* Event, New York Hall of Science, Corona, New York
May 19	Aguascalientes (AGS) Conference on Innovation, "Weed to Wonder: Maize in Mexico," Aguascalientes, Mexico
May 21	*Ötzi the Iceman* Tour, DNALC
	Manhattan Community School District 5 and Harlem Children's Zone, Science, Technology, Engineering and Math Expo, Harlem Armory, New York
May 23	NSF "CyVerse Atmosphere for Educators," Big Genomic Data Skills for Professors, Jackson Laboratory, Farmington, Connecticut
May 23–24	Site visit by Cheong Kim Fatt, Singaporean Ministry of Education, Queenstown, Singapore, DNALC and *Harlem DNA Lab*
May 24–26	NSF CyVerse *Genomic Science and Leadership Initiative* Workshop, J. Craig Venter Institute, Rockville, Maryland
May 25	*Ötzi the Iceman* Tour, DNALC
May 26–27	NSF CyVerse *Data Carpentry Genomics*, National Institutes of Health, Bethesda, Maryland
May 30–June 2	NSF CyVerse *Tools and Service/Data Carpentry* Workshop, Langebio Cinvestav, Irapuato, Mexico
May 31	NIH *Barcode Long Island* Open Lab, DNALC
June 1–2	NSF *Infrastructure and Training to Bring Next-Generation Sequence (NGS) Analysis into Undergraduate Education* Webinar, DNALC
	NSF CyVerse *Genomics Data Carpentry* Workshop, Langebio Cinvestav, Irapuato, Mexico
June 3	*Urban Barcode Project* and Pinkerton *Urban Barcode Research Program* Symposium, Borough of Manhattan Community College, New York
June 6–17	NSF *Infrastructure and Training to Bring Next-Generation Sequence (NGS) Analysis into Undergraduate Education* Virtual Workshop, DNALC
June 7, 9	*DNA Subway* Workshop and "Citizen Science at the DNALC," Ecsite Conference 2016, Natural History and Science Museum of the University of Porto, Portugal
June 10	NIH *Barcode Long Island* Symposium, CSHL
June 16	*Ötzi the Iceman* Tour, DNALC
June 20	QUBES – 2016 National Academies Special Topics Summer Institute on Quantitative Biology, University of North Carolina, Raleigh
June 21	23rd Annual Golf Outing, Piping Rock Club, Locust Valley, New York
June 22	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
June 22–26	Aspen Ideas Festival, Science/Art Workshop, "Painting with Microbes Brings an Unseen World into View," Aspen, Colorado
June 24	Site visit by Mary Ellen Elia, New York State Department of Education and University of the State of New York, John D'Agati, New York State Department of Education, Office of Higher Education, and Roger Tilles, New York State Board of Regents, Albany, New York, DNALC
June 26–July 15	*DNA Science, Genome Science,* and *Barcode Long Island* Workshops attended by student from Godfrey Okoye University, Enugu, Nigeria and students from Bowie State University, Maryland with George Ude, University System of Maryland Elkins Professor
June 27–July 1	NIH *Barcode Long Island* Workshop, Brookhaven National Laboratory, Upton, New York
June 27–July 1	*DNA Science* Workshop, DNALC (2 sessions)
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC
	Fun with DNA Workshop, DNALC *West*
June 28	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
June 29	National Congress of American Indians 2016 Mid-Year Conference and Marketplace, Tribal Leader Scholar Forum, "Should AI/AN communities consider genomic technologies in agricultural practices?" Spokane Convention Center, Washington
July 5–8	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	Forensic Detectives Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Genome Science Workshop, DNALC
	World of Enzymes Workshop, DNALC
	World of Enzymes Workshop, DNALC *West*
July 7	Site visit by Tom Flanagan, Girl Scouts of Suffolk County, Commack, New York
July 7–12	American Society of Plant Biologists/Plant Biology 2016, "Accelerating Plant Science with CyVerse," Austin, Texas
July 11	Site visit by Liz Baird, Christy Flint, Wendy Lovelady, and Imogen Hoyel, North Carolina Museum of Natural Science, Raleigh

July 11–15	NIH *Barcode Long Island* Workshop, DNALC
	DNA Science Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	DNA Science Workshop, Harlem DNA Lab
	DNA Science Workshop, DNALC *West*
July 14	*My Long Island TV* filming of *World of Enzymes* and *DNA Science* Workshops, DNALC
July 18	NIH *Barcode Long Island Microbiome Project* Planning Workshop, DNALC
July 18–22	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC *West*
	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
July 18–Aug 5	International Students Summer 2016, Beijing 166 School, China, DNALC
July 19	Site visit by David Stark, CSHL Association Member, and legal counsel for Teva Pharmaceuticals, Petach Tikva, Israel, DNALC
July 25–29	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	BioCoding Workshop, DNALC
	DNA Barcoding Research Workshop, DNALC
	Green Genes Workshop, DNALC
	Fun with DNA Workshop, DNALC *West*
July 30	NIH *Barcode Long Island* Microbiome Project Open Lab, DNALC
August 1–5	NIH *Barcode Long Island* Workshop, Stony Brook University, New York
	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
	DNA Barcoding Research Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Forensic Detectives Workshop, DNALC *West*
August 1–12	New York City Department of Education *DNA Barcoding* Workshop, Environmental Science Center, Brooklyn, New York
August 2	Site visit by Anthony Clarkson and Peter Turner, Clarkson University, Potsdam, New York
August 3–6	Network for Integrating Bioinformatics into Life Sciences Education Kick-off Meeting, University of Omaha, Nebraska
August 8–12	Pinkerton *Urban Barcode Research Program Conservation Genetics* Workshop, *Harlem DNA Lab*
	Backyard Barcoding Workshop, DNALC
	Being Human Workshop, DNALC
	Fun with DNA Workshop, DNALC
	World of Enzymes Workshop, DNALC *West*
August 8–13	ADN Mexico *Molecular Genetics Experiences for Students* Workshop, Morelia, Mexico
August 9	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
August 15–19	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Genome Science Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Backyard Barcoding Workshop, DNALC *West*
	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
August 21–22	"Leveraging Cyberinfrastructure to Scale Science and People," South African Bioinformatics Society, Durban, South Africa
August 22–26	*Forensic Detectives* Workshop, DNALC
	Green Genes Workshop, DNALC
	World of Enzymes Workshop, DNALC
	Fun with DNA Workshop, DNALC *West*
	NIH *Barcode Long Island* Microbiome Project Workshop, DNALC
August 23	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
	Site visit by Stefan Lutzmeyer, Gregor Mendel Institute, Vienna, Austria, *Harlem DNA Lab*
	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
August 23–26	*Urban Barcode Project*, *DNA Barcoding* Workshop, *Harlem DNA Lab*
Aug. 29–Sept. 2	*DNA Science* Workshop, DNALC
	Fun with DNA Workshop, DNALC
	Green Genes Workshop, DNALC

	World of Enzymes Workshop, DNALC *West*
September 14	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
September 15	NSF CyVerse SolGenomics Webinar, "CyVerse Virtual Demo," University of California, Davis
September 22	Pinkerton *Urban Barcode Research Program* featuring Irondale Theater Company, Irondale Theater, Brooklyn, New York
September 23	South African Bioinformatics Society, South African Genetics Society Joint-Conference, "Leveraging Cyberinfrastructure to Scale Science and People," Durban, South Africa
September 26–29	NSF CyVerse Data Carpentry Workshop, North-West University, Potchefstroom, South Africa
October 4	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
October 12	NSF CyVerse *Data Carpentry* Workshop, SACNAS (Advancing Chicanos/Hispanics & Native Americans in Science) National Conference, Long Beach Convention Center, California
	NSF CyVerse Webinar Series, "Getting Started with CyVerse," DNALC
October 15	*Saturday DNA!* "Cellular Portraits," DNALC
October 17–19	Breakthrough Prize, *Molecular Techniques for High School*, North Royalton High School, Cleveland, Ohio
October 17–27	*Genome Science* and *DNA Subway*, Beijing 166 School, China
October 19–25	NSF CyVerse *Cereal Genomics* Course, "Introduction to CyVerse," CSHL
October 21	Open Space Stewardship Program (OSSP) through Brookhaven National Laboratory, *A Day in the Life*, Peconic River, Riverhead, New York
October 26–29	Biological Data Science, "Required Parameters: What Does It Take to Bring Bioinformatics into the Classroom at the National Level?" CSHL
October 28	Site visit by Roberta Trapper, Principal, and Carol Yilmax, Founder, Long Island School for the Gifted, South Huntington, New York
November 5, 12	*Saturday DNA!* "Agar Art," DNALC
November 8	"One Native Scientist's Path from Red Valley to FLC MARC and Beyond!" Presentation, Fort Lewis College, Durango, Colorado
	Ötzi the Iceman Tour, DNALC
	NIH *Barcode Long Island Microbiome Project* Open Lab, DNALC
November 10–12	AISES National Conference, presentation on JCVI/David Jackson project, Minneapolis, Minnesota
November 12	*Ötzi the Iceman* Tour, DNALC
November 14–15	Science Media Exchange (SCIMEX) Presentation, "*Barcode Long Island* Student-Centered Biodiversity Research," Crest Hollow Country Club, Woodbury, NY
November 15	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
November 21	"Gleevec: Smart Bomb in the War on Cancer," Microsoft, New York
November 29–30	Indigenous Day Presentation "Balancing Diné and Keres Cultural Heritage in Academic Science," University of the Fraser Valley, Abbotsford, British Colombia
November 30	*Ötzi the Iceman* Tour, DNALC
December 2	Site visit by Dr. Sean Kassen, Dr. Mary Galvin, Allison (Maddux) Slabaugh, Dr. Amy Stark, University of Notre Dame, Indiana
December 3	*Saturday DNA!* "Molecular Diagnostics," (2 sessions) DNALC
December 6	CSHL *Gramene: A Resource for Comparative Plant Genomics* Webinar, DNALC
December 10	Pinkerton *Urban Barcode Research Program DNA Barcoding* Workshop, *Harlem DNA Lab*
December 17	NIH *Barcode Long Island* Open Lab, DNALC
December 19	Suzhou High School affiliated with Xi'an Jiaotong University, "What DNA Says About Our Human Family," Suzhou, China

COLD SPRING HARBOR
LABORATORY PRESS

PRESS PUBLICATIONS

Serials

Genes & Development, Vol. 30 (www.genesdev.org)

Genome Research, Vol. 26 (www.genome.org)

Learning & Memory, Vol. 23 (www.learnmem.org)

RNA, Vol. 22 (www.rnajournal.org)

Cold Spring Harbor Symposia in Quantitative Biology, Vol. 80: *21st Century Genetics: Genes at Work*, edited by Terri Grodzicker, David Stewart, and Bruce Stillman

Cold Spring Harbor Protocols (www.cshprotocols.org)

Cold Spring Harbor Perspectives in Biology (www.cshperspectives.org)

Cold Spring Harbor Perspectives in Medicine (www.perspectivesinmedicine.org)

Cold Spring Harbor Molecular Case Studies (www.molecularcasestudies.org)

Laboratory Manuals

Budding Yeast: A Laboratory Manual, edited by Brenda Andrews, Charles Boone, Trisha N. Davis, and Stanley Fields

CRISPR–Cas: A Laboratory Manual, edited by Jennifer Doudna and Prashant Mali

Fission Yeast: A Laboratory Manual, edited by Iain M. Hagan, Antony M. Carr, Agnes Grallert, and Paul Nurse

Textbooks

The Cytoskeleton, edited by Thomas D. Pollard and Robert D. Goldman

Monographs (Topic Collections from *Perspectives in Biology* and *Perspectives in Medicine*)

Microbial Evolution, edited by Howard Ochman

The P53 Protein: From Cell Regulation to Cancer, edited by Guillermina Lozano and Arnold J. Levine

Antibiotics and Antibiotic Resistance, edited by Lynn L. Silver and Karen Bush

Cilia, edited by Wallace Marshall and Renata Basto

Other

Enjoy Your Cells Coloring Book, by Fran Balkwill and Mic Rolph

Gene Machines Coloring Book, by Fran Balkwill and Mic Rolph

Germ Zappers Coloring Book, by Fran Balkwill and Mic Rolph

Have a Nice DNA Coloring Book, by Fran Balkwill and Mic Rolph

CSHL Annual Report 2015, Yearbook Edition

Banbury Center Annual Report 2015

E-books

The Cytoskeleton, edited by Thomas D. Pollard and Robert D. Goldman

Next-Generation DNA Sequencing Informatics, Second Edition, edited by Stuart M. Brown

Subcellular Fractionation: A Laboratory Manual, by Paul Pryor

Budding Yeast: A Laboratory Manual, edited by Brenda Andrews, Charles Boone, Trisha N. Davis, and Stanley Fields

At the Bench: A Laboratory Navigator, Updated Edition, by Kathy Barker

At the Helm: Leading Your Laboratory, Second Edition, by Kathy Barker

Statistics at the Bench: A Step-by-Step Handbook for Biologists, by Martina Bremer and Rebecca W. Doerge

Lab Dynamics: Management and Leadership Skills for Scientists, Second Edition, by Carl M. Cohen and Suzanne L. Cohen

Connecting with Companies: A Guide to Consulting Agreements for Biomedical Scientists, by Edward Klees, J.D. and H. Robert Horvitz, Ph.D.

Antibodies: A Laboratory Manual, Second Edition, edited by Edward A. Greenfield

Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition, by Richard Behringer, Marina Gertsenstein, Kristina Vintersten Nagy, and Andras Nagy

Mouse Models of Cancer: A Laboratory Manual, edited by Cory Abate-Shen, Katerina Politi, Lewis Chodosh, and Kenneth P. Olive

Navigating Metabolism by Navdeep Chandel

Lab Math: A Handbook of Measurements, Calculations, and Other Quantitative Skills for Use at the Bench, Second edition, by Dany Spencer Adams

CRISPR–Cas: A Laboratory Manual, edited by Jennifer Doudna and Prashant Mali

Lab Ref, Volume 2, A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench, edited by Albert S. Mellick and Linda Rodgers

Websites

Cold Spring Harbor Monographs Archive Online (www.cshmonographs.org)

Cold Spring Harbor Symposium on Quantitative Biology Archive (symposium.cshlp.org)

Services

bioRxiv, the preprint server for biology (www.bioRxiv.org)

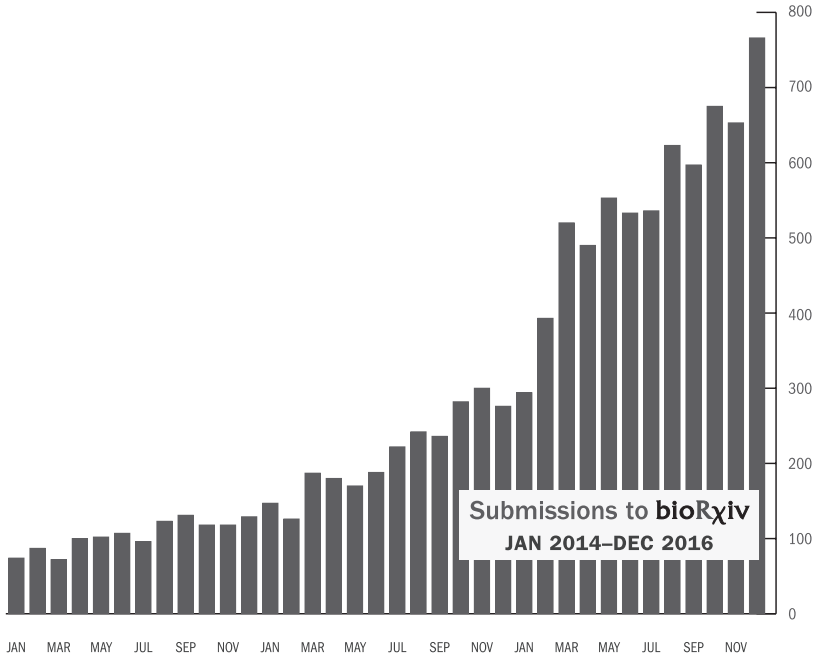
BioSupplyNet, scientific supply directory (www.biosupplynet.com)

COLD SPRING HARBOR LABORATORY PRESS EXECUTIVE DIRECTOR'S REPORT

Cold Spring Harbor Laboratory Press aims to provide scientists worldwide with authoritative, affordable, and appropriate information to further their research and career development. The Press carries forth the Laboratory's commitment to scientific communication, which first began more than 80 years ago at the Laboratory's first Annual Symposium.

The Press publishes eight journals and more than 200 books in print and electronic form. The long-established journals *Genome Research* and *Genes & Development* remain preeminent, with editorial teams adept at capturing the new ideas and technologies emerging in a broad range of disciplines. *RNA* and *Learning & Memory* continue serving their specialized research communities in valuable ways. The newer review journals, *CSH Perspectives in Biology*, *CSH Perspectives in Medicine*, and *CSH Protocols*, continue to advance in stature and financial success. *CSH Molecular Case Studies* had a steady increase in submissions in its first full year and was accepted for indexing by the National Library of Medicines' PubMed service. The Press journals overall had a record yearly download of more than 17 million full-text articles.

A highlight in 2016 was the remarkable growth of bioRxiv, the Laboratory's open preprint service that was founded and is operated by Press staff. bioRxiv enables scientists to make draft manuscripts of papers immediately available to the research community and receive feedback before submission to peer-reviewed journals. Manuscript postings in 2016 were 2.5 times higher than in the previous year. With nearly 9000 papers from 40,000 authors in more than 80 countries, bioRxiv is now the worlds' largest source of life science preprints and is accessed more than a million times each month. Established in November 2013, bioRxiv has fundamentally changed the communication practices of scientists in biology and ignited similar movements in other sciences. This impact was made possible by the Laboratory's vision and the generosity of Trustee Robert Lourie.



The Laboratory's bioRxiv is now the world's largest source of life science preprints, accessed more than a million times each month.

The book-publishing program added 13 new titles to its list of 200, including a timely manual on powerful CRISPR-based technologies. The bundling of print books with electronic editions for tablets, smartphones, and computers was warmly received by busy working scientists. The recent implementation of high-quality print-on-demand delivery for books now means that every title is always available to readers without the operational burden of costly inventory.

Staff

During the year, Maryliz Dickerson, a long-time and much valued colleague at CSHL Press and the Lab at large, passed away. Maryliz joined the Press in 1988, and in her 26 years at the Laboratory she managed some of our most complicated and successful book projects. Her dedicated, precise contributions to the work of our committed and demanding authors were immense, as the acknowledgment sections of so many of our books show. But above all she will be remembered for her warmth and generosity of spirit that endeared her to so many.

Heather Cosel-Pieper, Reviews Editor, *G&D*, left us to teach exclusively at the DNA Learning Center after 16 years at the Press. Peggy Calicchia, Administrative Assistant, *GR*, *L&M*, and *MCS*, retired following 19 years of service and moved to join her family in California. We also said farewell to Maria Ebbets, Production Assistant for two years, and welcomed Deborah Jarski in her place.

The mission of the Press is to create publications and services that help scientists succeed while supporting the Laboratory's finances and its reputation in scientific education and communication. We can do so only by engaging many of the world's most accomplished scientists, and these relationships are sustained by the skills and dedication of the Press staff. I thank in particular the individuals with leadership roles in our diverse activities: Assistant Director Richard Sever, journal editors Terri Grodzicker and Hillary Sussman, and departmental directors Jan Argentine, Wayne Manos, Stephen Nussbaum, Marcie Siconolfi, and Linda Sussman. And as always, I am extremely grateful for the cheerfulness, warmth, and efficiency of my Executive Assistant, Mala Mazzullo.

John Inglis
Executive Director and Publisher

Some members of the Press staff, January 2016; photo by Gina Motis.

FINANCE

FINANCIAL STATEMENTS

CONSOLIDATED BALANCE SHEET
December 31, 2016
(with comparative financial information as of December 31, 2015)

		2016	2015
Assets:			
Cash and cash equivalents	$	58,209,476	54,209,176
Grants receivable		8,066,970	9,545,355
Contributions receivable, net		47,353,618	30,100,043
Investments		472,977,094	449,931,993
Investment in employee residences		6,000,849	6,161,403
Restricted use assets		3,522,055	5,412,103
Other assets		6,323,460	10,270,883
Land, buildings, and equipment, net		229,093,660	230,619,980
Total assets	$	831,547,182	796,250,936
Liabilities and net assets:			
Liabilities:			
Accounts payable and accrued expenses	$	10,030,041	10,048,146
Deferred revenue		9,125,481	8,004,642
Interest rate swap		32,713,773	34,052,132
Bonds payable		95,675,157	95,608,887
Total liabilities		147,544,452	147,713,807
Commitments and contingencies			
Net assets:			
Unrestricted		364,233,745	342,262,835
Temporarily restricted		205,772,108	192,160,567
Permanently restricted		113,996,877	114,113,727
Total net assets		684,002,730	648,537,129
Total liabilities and net assets	$	831,547,182	796,250,936

CONSOLIDATED STATEMENT OF ACTIVITIES
Year ended December 31, 2016
(with summarized financial information for the year ended December 31, 2015)

	Unrestricted	*Temporarily Restricted*	*Permanently Restricted*	*2016 Total*	*2015 Total*
Revenue and other support:					
Public support—contributions and nonfederal grant awards	$ 27,723,061	41,207,207	883,150	69,813,418	31,083,872
Federal grant awards	32,371,500	–	–	32,371,500	31,750,274
Indirect cost allowances	29,781,859	–	–	29,781,859	27,286,692
Investment return utilized	18,455,646	–	–	18,455,646	17,887,633
Program fees	7,444,287	–	–	7,444,287	8,681,384
Publications sales	9,567,069	–	–	9,567,069	9,737,489
Dining services	4,596,996	–	–	4,596,996	4,819,543
Rooms and apartments	3,801,520	–	–	3,801,520	3,880,805
Miscellaneous	3,977,941	–	–	3,977,941	5,183,130
Net assets released from restrictions	27,665,440	(27,665,440)	–	–	–
Total revenue and other support	165,385,319	13,541,767	883,150	179,810,236	140,310,822
Expenses:					
Research	90,381,188	–	–	90,381,188	86,078,788
Educational programs	18,449,272	–	–	18,449,272	19,849,038
Publications	8,880,801	–	–	8,880,801	9,152,372
Banbury Center conferences	1,545,507	–	–	1,545,507	1,591,739
DNA Learning Center programs	2,385,791	–	–	2,385,791	2,001,720
Watson School of Biological Sciences programs	2,949,461	–	–	2,949,461	3,246,931
General and administrative	17,879,756	–	–	17,879,756	17,204,666
Dining services	5,949,534	–	–	5,949,534	5,924,858
Total expenses	148,421,310	–	–	148,421,310	145,050,112
Excess (deficiency) of revenue and other support over (under) expenses	16,964,009	13,541,767	883,150	31,388,926	(4,739,290)
Other changes in net assets:					
Investment return (loss) excluding amount utilized	2,668,542	69,774	–	2,738,316	(19,507,865)
Contribution reclassification due to change in donor intent	1,000,000	–	(1,000,000)	–	–
Change in fair value of interest rate swap	1,338,359	–	–	1,338,359	(428,579)
Increase (decrease) in net assets	21,970,910	13,611,541	(116,850)	35,465,601	(24,675,734)
Net assets at beginning of year	342,262,835	192,160,567	114,113,727	648,537,129	673,212,863
Net assets at end of year	$ 364,233,745	205,772,108	113,996,877	684,002,730	648,537,129

CONSOLIDATED STATEMENT OF CASH FLOWS
Year ended December 31, 2016
(with comparative financial information for the year ended December 31, 2015)

	2016	2015
Cash flows from operating activities:		
Increase (decrease) in net assets	$ 35,465,601	(24,675,734)
Adjustments to reconcile increase (decrease) in net assets to net cash provided by operating activities:		
Change in fair value of interest rate swap	(1,338,359)	428,579
Depreciation and amortization	13,694,691	13,808,887
Donated equipment	–	(1,880,032)
Amortization of deferred bond costs	66,270	66,269
Net (appreciation) depreciation in fair value of investments	(18,048,396)	4,238,813
Contributions restricted for long-term investment	(10,453,798)	(3,057,415)
Changes in assets and liabilities:		
Grants receivable	1,478,385	1,006,173
Contributions receivable, net	(10,729,807)	25,537,066
Restricted use assets	1,890,048	(284,288)
Other assets	3,947,423	(287,523)
Accounts payable and accrued expenses, net of financing activities	481,895	(1,848,350)
Deferred revenue	1,120,839	2,494,953
Net cash provided by operating activities	17,574,792	15,547,399
Cash flows from investing activities:		
Capital expenditures	(12,168,371)	(10,897,945)
Proceeds from sales and maturities of investments	40,841,045	44,262,648
Purchases of investments	(45,837,750)	(55,602,925)
Net change in investment in employee residences	160,554	(1,002,025)
Net cash used in investing activities	(17,004,522)	(23,240,247)
Cash flows from financing activities:		
Contributions restricted for long-term investment	883,150	2,939,438
Contributions restricted for investment in capital	9,570,648	117,977
(Increase) decrease in contributions receivable	(6,523,768)	3,149,150
Decrease in accounts payable relating to capital expenditures	(500,000)	(614,499)
Net cash provided by financing activities	3,430,030	5,592,066
Net increase (decrease) in cash and cash equivalents	4,000,300	(2,100,783)
Cash and cash equivalents at beginning of year	54,209,176	56,309,959
Cash and cash equivalents at end of year	$ 58,209,476	54,209,176
Supplemental disclosure:		
Interest paid	$ 3,816,392	3,978,881

FINANCIAL SUPPORT OF THE LABORATORY

Cold Spring Harbor Laboratory, Banbury Center, and the Dolan DNA Learning Center receive a substantial portion of funding through grants from the federal government and through grants, capital gifts, and annual contributions from New York State, private foundations, corporations, and individuals. The following section summarizes funding that occurred during 2016.

GRANTS January 1–December 31, 2016

COLD SPRING HARBOR LABORATORY

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
FEDERAL GRANTS					
NATIONAL INSTITUTES OF HEALTH					
Equipment Support	Dr. Furukawa	04/01/16	03/31/18	$ 190,000	*
	Dr. Huang	05/01/16	04/30/17	203,411	*
	Dr. Stillman	06/01/16	05/31/18	190,000	*
Program Project and Center Support	Drs. Stillman/Krainer/Spector/Vakoc	05/25/12	12/31/17	4,445,418	
	Drs. Stillman/Tuveson—Cancer Center Core	08/01/16	07/31/21	4,405,681	*
	Drs. McCombie/Tuveson	09/25/15	08/31/19	155,219	
Cooperative Research Agreement Support[2]	Dr. Gingeras	09/27/16	07/31/17	959,195	*
	Dr. Kepecs	09/30/15	08/31/18	394,924	
	Drs. Krasnitz/Wigler	03/01/15	02/28/18	314,160	
	Drs. Osten/Albeanu/Mitra	09/26/14	05/31/17	1,101,245	
	Drs. Powers/Krasnitz/Sordella/Tuveson	05/01/12	04/30/17	538,938	
	Dr. Tuveson	06/12/14	02/28/19	205,368	
	Dr. Zador	09/25/15	06/30/18	552,785	
Research Support	Dr. Albeanu	07/01/15	06/30/17	512,872	
	Dr. Albeanu	07/01/13	06/30/18	480,000	
	Dr. Albeanu	02/09/16	01/31/21	408,000	*
	Dr. Churchland	03/01/13	02/28/18	472,500	
	Dr. Dubnau	01/01/14	07/31/16	197,111	
	Drs. Dubnau/M. Hammell	07/01/15	06/30/17	240,000	
	Dr. Furukawa	03/01/15	02/29/20	510,384	
	Dr. Furukawa	05/08/14	03/31/18	476,797	
	Dr. C. Hammell	03/01/16	12/31/20	403,200	*
	Dr. Huang	07/01/13	03/31/18	753,224	
	Drs. Huang/Gillis	08/01/16	07/31/21	916,647	*
	Drs. Huang/Wigler	03/01/14	01/31/19	738,396	
	Dr. Joshua-Tor	06/10/16	03/31/20	345,600	*
	Dr. Kepecs	04/01/14	03/31/19	480,000	
	Dr. Kepecs	08/15/15	05/31/20	420,000	
	Dr. Kepecs	09/15/15	07/31/20	440,325	
	Drs. Koulakov/Zador	07/01/13	03/31/18	427,680	
	Dr. Krainer	04/01/12	03/31/17	746,014	

[1]Includes direct and indirect costs
[2]Cooperative research agreement funding amounts include only CSHL portion of award
*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]
	Dr. Lee	08/19/16	05/31/21	$ 480,000 *
	Dr. Li	03/01/14	02/28/19	623,589
	Drs. Li/Huang	09/28/15	06/30/20	660,707
	Dr. Martienssen	09/01/15	08/31/19	368,640
	Dr. Mills	09/16/14	08/31/19	634,545
	Dr. Mitra	09/30/16	06/30/19	450,322 *
	Dr. Osten	04/01/12	03/31/17	473,750
	Dr. Park	09/15/16	08/31/21	174,117 *
	Dr. Shea	12/08/15	11/30/20	477,244
	Dr. Sheltzer	09/18/15	08/31/20	470,800
	Dr. Siepel	09/01/14	12/31/17	340,039
	Dr. Siepel	04/24/15	01/31/18	192,000
	Dr. D. Spector	04/01/15	03/31/19	766,080
	Dr. Stenlund	05/01/15	04/30/20	504,989
	Dr. Tonks	05/14/15	04/30/20	762,314
	Dr. Tonks	01/15/15	12/31/18	448,735
	Dr. Trotman	07/30/14	06/30/19	398,400
	Dr. Tuveson	12/07/16	11/30/21	510,684 *
	Drs. Tuveson/M. Hammell/Pappin/Trotman	09/22/14	08/31/18	560,492
	Dr. Vakoc	04/01/13	03/31/18	557,762
	Dr. Van Aelst	08/01/13	05/31/18	532,426
	Dr. Van Aelst	04/01/13	03/31/18	489,466
	Dr. Zador	09/01/12	08/31/17	401,625
	Dr. Zador	07/01/14	04/30/19	420,000
Research Subcontracts				
NIH/Boston's Childrens Hospital Consortium Agreement	Dr. Mitra	09/23/15	06/30/18	120,632
NIH/Brandeis University Consortium Agreement	Dr. Mitra	09/26/14	05/31/17	298,565
NIH/Certerra, Inc. Consortium Agreement	Dr. Koulakov	01/01/13	12/31/16	108,750
NIH/Columbia University Consortium Agreement	Dr. Atwal	05/01/16	04/30/20	33,103 *
NIH/Emory University Consortium Agreement	Dr. Huang	04/01/14	02/28/18	37,800
NIH/Envisagenics, Inc. Consortium Agreement	Dr. Krainer	08/05/15	03/05/17	15,000
NIH/Johns Hopkins University Consortium Agreement	Dr. Joshua-Tor	03/01/15	02/29/20	36,960
NIH/New York Genome Center Consortium Agreement	Drs. Wigler/Iossifov/Levy/Siepel	01/14/16	11/30/16	231,857 *
NIH/New York University Consortium Agreement	Dr. Koulakov	06/01/14	05/31/19	176,513
NIH/The Research Foundation for the State of New York–Stony Brook Consortium Agreement	Dr. Koulakov	06/01/15	05/31/16	42,933
NIH/The Research Foundation for the State of New York–Stony Brook Consortium Agreement	Dr. Wigler	05/01/14	04/30/19	130,599
NIH/Sloan Kettering Institute for Cancer Research Consortium Agreement	Dr. Sordella	09/12/12	08/31/17	148,565
NIH/University of California–Los Angeles Consortium Agreement	Dr. Wigler	09/25/14	05/31/16	40,000

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
NIH/University of California–San Diego Consortium Agreement	Dr. Osten	01/01/16	08/31/16	$ 33,304	*
Fellowship/Career Development Support	Dr. Baker	02/01/15	10/31/17	58,002	
	Dr. Engle	08/01/16	07/31/18	128,235	*
	Dr. Hwang	03/01/15	02/28/18	59,970	
	O. Odoemene	01/01/15	10/31/16	35,472	
	Dr. Regan	12/01/15	11/30/18	56,042	
	Dr. Wang	05/02/16	05/01/19	52,542	*
	Dr. Wong	08/01/15	07/31/18	56,118	
Institutional Training Program Support	Dr. Gann/Watson School of Biological Sciences	07/01/12	06/30/17	252,696	
	Dr. Mills/Cancer Postdoctoral	09/01/11	08/31/16	154,311	
Course Support	Advanced Sequencing Technologies and Applications	04/10/12	06/30/18	70,160	
	Advanced Techniques in Molecular Neuroscience	07/06/15	03/31/20	105,668	
	Cell and Developmental Biology of *Xenopus*	05/05/14	03/31/19	76,284	
	Cellular Biology of Addiction	08/01/16	07/31/21	55,455	*
	Computational and Comparative Genomics Course	08/01/14	06/30/17	52,816	
	Empowering Nextgen Advanced Biomedical Leadership	06/01/15	02/29/20	521,040	
	Eukaryotic Gene Expression	04/01/12	03/31/17	99,312	
	Molecular Embryology of the Mouse	04/01/12	03/31/17	130,759	
	Neurobiology of *Drosophila*	07/15/12	06/30/17	35,000	
	Programming for Biology	09/01/14	08/31/17	80,120	
	Protein Purification and Characterization	04/01/12	03/31/17	89,541	
	Proteomics	08/01/12	04/30/17	116,661	
	Quantitative Imaging: From Cells to Molecules	04/01/16	03/31/21	102,233	*
	Statistical Methods for Functional Genomics	08/05/13	07/31/17	105,657	
	X-Ray Methods in Structural Biology	09/01/12	08/31/17	81,540	
Meeting Support	Axon Guidance, Synapse Formation, and Regeneration	04/01/16	03/31/17	16,000	*
	The Biology of Genomes	04/01/13	03/31/18	54,600	
	Blood–Brain Barrier	07/01/16	06/30/17	5,000	*
	Gene Expression and Signaling in the Immune System	04/01/16	03/31/17	9,000	*
	Germ Cells	07/01/14	06/30/19	6,000	
	HIV/AIDS Research: Its History and Future	09/01/16	08/31/17	50,000	*
	Molecular Genetics of Aging	03/15/14	02/28/19	30,972	
	The PARP Family and ADP Ribosylation	04/01/16	03/31/17	4,000	*
	Retroviruses	03/15/12	02/28/17	36,645	
	Systems Biology: Global Regulation of Gene Expression	02/22/16	01/31/19	29,951	*
NATIONAL SCIENCE FOUNDATION					
Multiple Project Award Support	Drs. Gingeras/McCombie/Martienssen/Jackson/Ware	06/15/16	05/31/19	1,833,950	*

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]
	Drs. Jackson/Lippman	09/01/16	08/31/20	$ 1,159,828 *
	Dr. Ware	06/01/12	05/31/17	1,835,951
Research Support	Dr. Jackson	06/01/15	05/31/18	197,347
	Dr. Kepecs	09/01/15	08/31/17	150,000
	Dr. Lippman	06/15/16	05/31/19	208,143 *
	Drs. McCombie/Martienssen	06/01/15	05/31/18	541,852
	Dr. Siepel	03/01/16	02/28/19	87,262 *
Research Subcontracts				
NSF–Cornell University Consortium Agreement	Dr. Timmermans	02/01/13	01/31/18	438,421
NSF–Cornell University Consortium Agreement	Dr. Ware	05/15/13	04/30/18	152,296
NSF–Iowa State University Consortium University	Dr. Jackson	03/01/13	02/28/18	413,554
NSF–University of Arizona Consortium University	Dr. Ware/Micklos	09/01/13	08/31/18	1,430,483
Fellowship Support	B. Bibel	09/01/14	07/31/19	46,000
	Dr. Lemmon (Direct)	07/01/15	06/30/18	72,000
Institutional Training Program Support	Drs. Churchland/Schatz	05/01/16	04/30/19	148,657 *
Course Support	Advanced Bacterial Genetics	01/15/14	12/31/16	85,000
	Drosophila Neurobiology: Genes, Circuits, and Behavior	07/01/15	06/30/17	25,000
	Frontiers and Techniques in Plant Science	05/01/15	04/30/18	104,579
	Synthetic Biology	07/01/15	06/30/18	33,300
	Yeast Genetics and Genomics	07/15/14	06/30/17	85,000
	Biological Data Science	05/01/16	04/30/17	5,625 *
	Epigenetics and Chromatin	05/01/16	04/30/17	6,305 *
	Evolutionary Biology of *Caenorhabditis* and Other Nematodes	05/01/16	04/30/17	5,625 *
	Germ Cells	05/01/16	04/30/17	6,945 *
	Organization and Function	05/01/16	04/30/17	5,625 *
	Regulatory and Noncoding RNAs	05/01/16	04/30/17	5,625 *
Meeting Support	Systems Biology: Global Regulation of Gene Expression	05/01/16	04/30/17	3,625 *
	Translational Control	05/01/16	04/30/17	5,625 *
UNITED STATES DEPARTMENT OF AGRICULTURE				
Research Support	Dr. Jackson	12/01/13	11/30/17	130,985
	Dr. Jackson	12/15/15	12/14/19	121,504
	Dr. Lippman	01/01/15	12/31/18	120,777
	Dr. Lippman	11/01/15	10/31/19	100,259
	Dr. McCombie	09/01/14	08/31/19	494,875
Course Support	Workshop on Cereal Genomics	11/01/16	10/31/17	25,000 *
UNITED STATES DEPARTMENT OF THE ARMY				
Research Support	Dr. Egeblad	06/15/14	06/14/19	936,990
	Drs. Trotman/Hicks	09/22/14	09/21/17	240,410

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
Research Subcontracts					
DOA/University of Southern California Consortium Agreement	Dr. Churchland	08/23/16	08/22/17	$ 36,000	*
UNITED STATES DEPARTMENT OF ENERGY					
Research Subcontracts					
DOE/Brookhaven National Laboratory Consortium Agreement	Dr. Ware	11/29/11	03/31/17	920,000	
DOE/New York University Consortium Agreement	Dr. McCombie	08/15/15	08/14/20	232,190	
UNITED STATES DEPARTMENT OF THE INTERIOR					
Research Subcontracts					
DOI/Harvard University Consortium Agreement	Drs. Zador/Koulakov	01/15/16	01/14/21	2,685,395	*
MISCELLANEOUS SOURCES OF FUNDING					
Equipment Support					
The Mary K. Chapman Foundation	Dr. Stillman	12/31/16	12/30/17	175,000	*
CSHL Translational Cancer Support	Dr. McCombie	11/01/15	10/31/17	75,000	
	Dr. Vakoc	01/15/16	01/14/19	473,800	*
The Claire Friedlander Family Foundation	Drs. Stillman/McCombie	10/14/16	06/30/18	190,000	*
Program Project Support					
Pfizer Inc.	Dr. Stillman	01/01/12	12/31/16	2,050,000	
The Simons Foundation/Autism	Dr. Wigler	01/01/12	12/31/16	4,000,000	
Research Support					
American Association for Cancer Research	Dr. Tuveson	06/01/15	05/31/17	127,000	
	Dr. Vakoc	07/01/16	06/30/18	100,000	*
Rita Allen Foundation	Dr. dos Santos	09/01/16	08/31/21	100,000	*
	Dr. C. Hammell	09/01/12	08/31/17	110,000	
	Dr. M. Hammell	09/01/14	08/31/19	110,000	
American Cancer Society	Dr. Trotman	07/01/14	06/30/18	198,000	
American Lung Association	Dr. Van Aelst	07/01/16	06/30/17	100,000	*
Anonymous	Dr. dos Santos	12/19/16	12/18/17	1,000	*
Austins Purpose Corporation	Dr. Furukawa	12/29/16	12/28/17	10,000	*
Beckman Research Institute of the City of Hope	Dr. Atwal	06/01/15	02/26/17	200,000	
Boehringer Ingelheim RCV GmbH & Co KG	Dr. Vakoc	09/15/15	09/14/18	343,559	
Brain & Behavior Research Foundation	Dr. Kepecs	09/15/15	09/14/17	50,000	
	Dr. Li	09/15/15	09/14/17	50,000	
Brandon Sales Corporation	Dr. Tuveson	11/17/16	11/16/17	3,500	*
The Breast Cancer Research Foundation	Drs. Wigler/Levy/Mitra	10/01/16	09/30/17	250,000	*
Calico Life Sciences LLC	Drs. Wigler/Levy/Mitra	09/01/16	08/31/17	647,299	*
California Table Grape Commission	Dr. Ware	03/15/16	03/14/17	20,000	*

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
Cornelia de Lange Syndrome Foundation, Inc.	Dr. Lyon	09/01/16	08/31/17	$ 15,000	*
Cedar Hill Foundation	Dr. Fearon	11/16/16	11/15/17	60,000	*
Columbia University/The Simons Foundation	Dr. Krainer	09/01/14	08/31/17	57,132	
CSHL Translational Cancer Support	Dr. Chang	06/01/15	05/31/18	340,988	
	Dr. Egeblad	06/15/16	06/14/19	504,483	*
	Dr. Egeblad	08/01/16	07/31/17	76,800	*
	Dr. Gingeras	01/15/16	01/14/19	621,651	*
	Dr. Joshua-Tor	06/01/15	05/31/17	85,885	
	Dr. Joshua-Tor	06/01/15	05/31/17	134,321	
	Dr. Krasnitz	07/01/16	06/30/17	42,928	*
	Dr. Lee	06/01/15	05/31/18	340,406	
	Dr. Levy	07/01/16	06/30/17	33,128	*
	Dr. Lyons	06/01/15	05/31/18	419,654	
	Dr. McCombie	11/01/15	10/31/17	800,857	
	Dr. Pappin	06/01/15	05/31/17	192,056	
	Dr. Schatz	11/01/15	10/31/17	137,098	
	Dr. Sordella	11/01/15	10/31/17	643,617	
	Dr. Spector	06/01/15	05/31/18	303,583	
	Dr. Tonks	06/01/15	05/31/17	808,232	
	Dr. Tonks	06/01/15	05/31/17	348,528	
	Dr. Trotman	06/01/16	05/31/19	503,039	*
	Dr. Tuveson	06/01/15	05/31/18	486,372	
	Dr. Tuveson	01/01/16	12/31/17	76,800	
	Dr. Vakoc	01/15/16	01/14/19	1,222,641	*
	Dr. Van Aelst	06/01/15	05/31/18	404,241	
	Dr. Wigler	07/01/16	06/30/17	381,887	*
	Dr. Yeh	06/15/16	06/14/19	911,385	*
	Dr. Zheng	06/01/15	05/31/19	382,764	
The Cure Starts Now Foundation	Dr. Krainer	09/08/16	09/07/17	100,000	*
Dart Neuroscience LLC	Dr. Dubnau	01/01/16	08/31/16	142,444	*
Oliver S. and Jennie R. Donaldson Charitable Trust	Dr. Tuveson	12/21/16	12/20/17	259,518	*
East–West International BV	Dr. Lippman	06/01/16	05/31/19	100,000	*
FibroGen, Inc.	Dr. Tuveson	06/01/15	05/31/18	236,717	
The Clark Gillies Foundation	Dr. Vakoc	09/16/15	09/15/17	50,000	
William Guy Forbeck Research Foundation	Dr. Vakoc	07/01/16	06/30/18	80,000	*
GlaxoSmithKline Intellectual Property Development Limited	Dr. Tonks	01/12/15	01/11/18	480,000	
Glen Cove C.A.R.E.S.	Dr. dos Santos	02/17/16	02/16/17	6,000	*
The GoGo Foundation	Dr. Li	09/29/16	09/28/19	40,000	*
Elijah Swartz Gordon	Dr. D. Spector	12/07/16	12/06/17	1,000	*
Gyeongsang National University/the Republic of Korea	Dr. Jackson	01/04/16	12/31/17	51,945	*
James Hicks, Ph.D.	Dr. Hicks/Anaparthy	12/27/16	12/26/17	3,000	*
The Hope Foundation	Dr. Tuveson	01/01/16	12/31/16	54,000	*
Howard Hughes Medical Institute —Gordon and Betty Moore Foundation	Dr. Martienssen	12/01/11	07/31/17	166,667	
Human Frontier Science Program	Dr. Lee	07/01/16	06/30/19	116,667	*
	Dr. Li	09/01/16	08/31/19	100,000	*
Indian Institute of Technology Madras	Dr. Mitra	01/01/15	12/31/19	27,283	

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
Ionis Pharmaceuticals, Inc.	Dr. Krainer	07/11/16	12/31/17	$ 768,000	*
	Dr. D. Spector	07/01/15	12/21/17	768,000	
F.M. Kirby Foundation, Inc.	Dr. Stillman	04/06/16	04/05/17	100,000	*
The Esther A. & Joseph Klingenstein Fund, Inc.	Dr. Churchland	07/01/14	06/30/17	75,000	
The Susan G. Komen Breast Cancer Foundation, Inc.	Dr. Hicks	10/22/13	10/21/17	129,423	
Carolyn M. Lahy	Dr. Wigler	04/14/16	04/13/17	50	*
Marilyn Leidner	Dr. D. Spector	04/29/16	04/28/17	1,300	*
The Leukemia & Lymphoma Society	Dr. Vakoc	07/01/15	06/30/20	110,000	
Long Island Bioscience Hub	Dr. Atwal	12/01/16	11/30/17	60,000	*
	Dr. Ronemus	05/15/16	05/14/17	118,584	*
	Dr. Zhang	12/01/16	11/30/17	60,000	*
Dr. Robert Lourie	Dr. Zador	08/01/16	07/31/17	1,000,000	*
The Lustgarten Foundation	Dr. Fearon	07/01/14	12/31/19	1,000,000	
	Dr. Tuveson	06/15/12	06/30/17	50,000	
	Drs. Tuveson/Krasnitz	05/01/16	04/30/18	1,575,742	*
	Drs. Tuveson/Pappin	09/01/12	08/30/17	1,000,000	
Manhasset Women's Coalition against Breast Cancer	Dr. dos Santos	12/15/16	12/14/17	65,000	*
Massachusetts General Hospital	Drs. Lyon/Gillis	01/01/16	12/31/16	164,241	*
The G. Harold and Leila Y. Mathers Charitable Foundation	Dr. Mitra	01/01/14	12/31/16	330,000	
Breast Cancer Awareness Day in Memory of Elizabeth McFarland	Dr. Wigler	01/01/16	12/31/17	50,347	*
The McKnight Endowment Fund for Neuroscience	Dr. Kepecs	02/01/14	01/31/17	100,000	
The Meier & Linnartz Family Foundation	Dr. Huang	06/01/16	05/31/17	50,000	*
The Don Monti Memorial Research Foundation	Drs. Stillman/Tonks	03/01/08	02/28/17	350,000	
Louis Morin Charitable Trust	Dr. Tuveson	12/23/16	12/22/17	100,000	*
New York State Department of Health	Drs. D. Spector/Trotman	11/01/15	10/31/17	146,652	
Omicia, Inc.	Dr. Lyon	11/17/14	11/16/16	7,500	
The Michelle Paternoster Foundation	Dr. Vakoc	09/16/15	09/15/17	50,000	
The Pershing Square Foundation	Dr. Trotman	07/01/14	06/30/17	200,000	
	Dr. Vakoc	09/09/16	07/31/19	200,000	*
The Pew Charitable Trusts	Dr. Churchland	08/01/14	07/31/19	60,000	
Seemon H. and Natalie E. Pines Foundation	Dr. D. Spector	12/20/16	12/19/17	5,000	*
The Reiss Family Foundation	Dr. Tuveson	10/04/16	10/03/17	10,000	*
Christina Renna Foundation Inc.	Dr. Vakoc	09/16/15	09/15/17	35,000	*
The Research Foundation for the State University of New York/American Society of Gastrointestinal Endoscopy	Dr. Tuveson	07/01/15	06/30/17	14,000	
RIKEN	Dr. Mitra	04/01/15	03/31/18	2,727	
Charles and Marie Robertson Foundation	Dr. dos Santos	03/01/16	02/28/17	40,000	*
	Dr. Egeblad	03/01/16	02/28/17	40,000	*
Diane Emdin Sachs Memorial Fund	Dr. Sordella	09/01/12	08/31/17	261	
Eleanor Schwartz Charitable Foundation	Dr. Tuveson	08/01/14	07/31/17	200,000	

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
Seven Bridges Genomics Inc.	Dr. Lyon	07/01/16	06/30/17	$ 10,000	*
The Simons Foundation	Dr. Churchland	07/01/14	06/30/17	300,000	
	Dr. Huang	08/01/14	07/31/17	72,271	
	Dr. Li	07/01/15	06/30/17	125,000	
	Dr. Mills	12/01/16	11/30/17	115,597	*
The Sontag Foundation	Dr. Zheng	10/01/13	09/30/17	150,000	
Starr Cancer Consortium	Dr. Krainer	01/01/15	12/31/17	166,000	
	Drs. Lee/M. Hammell	01/01/16	12/31/17	249,000	*
	Drs. Trotman/Hicks	01/01/15	06/30/17	249,000	
	Dr. Vakoc	01/01/16	12/31/17	149,400	*
	Dr. Wigler	01/01/16	12/31/17	49,800	*
Stoke Therapeutics, Inc.	Dr. Krainer	10/24/14	12/31/16	195,241	
Swim Across America	Dr. Sordella	12/29/16	12/28/17	70,000	*
The Swartz Foundation	Dr. Zador	01/01/16	12/31/16	88,737	*
Friends of the TJ Foundation Inc.	Dr. Vakoc	09/16/15	09/15/17	50,000	
University of Pennsylvania/The Simons Foundation	Dr. Mills	09/01/15	08/31/18	26,842	
U.S.–Israel Binational Agricultural Research and Development Fund	Dr. Lippman	12/01/15	11/30/18	47,000	
	Dr. Shea	09/01/16	08/31/20	16,667	*
	Dr. Shea	07/01/15	06/30/19	12,000	
The V Foundation	Dr. Atwal	01/19/16	01/31/19	81,281	*
	Dr. dos Santos	11/01/16	11/01/18	100,000	*
	Dr. Lee	11/01/15	11/01/17	100,000	
	Dr. Tuveson	11/01/16	11/01/19	200,000	*
Joan & Sanford I. Weill Medical College	Dr. Fearon	07/01/14	06/30/19	166,332	
Women's Partnership in Science	Dr. Van Aelst	01/01/16	12/31/16	168,822	*
Elisabeth R. Woods Foundation Inc.	Dr. Sordella	01/01/16	12/31/16	32,000	*
The WTFC Foundation, Inc.	Dr. D. Spector	06/01/16	05/31/17	10,000	*
The Bradley Zankel Foundation, Inc.	Dr. Zheng	01/18/16	01/17/17	15,000	*
Fellowship Support					
Agency for Science, Technology, and Research	Watson School of Biological Sciences	09/01/16	08/31/16	44,657	*
	C. Yuan	08/22/16	08/15/17	30,000	*
American–Italian Cancer Foundation	Dr. Tonelli	08/01/16	07/31/18	40,000	*
Arnold and Mabel Beckman Foundation	Dr. Mejia	12/31/15	08/31/17	64,024	
Boehringer Ingelheim Fonds	G. Battistoni	09/01/15	08/31/17	36,115	
	L. Maiorino	09/01/15	08/31/17	33,440	
	J. Wang	08/01/16	07/31/18	33,509	*
	G. Yordanov	09/01/15	08/31/17	33,440	
Brain and Behavior Foundation	Dr. Chae	01/15/15	01/14/17	35,000	
	Dr. Stephenson-Jones	02/01/15	01/31/17	35,000	
Terri Brodeur Breast Cancer Foundation	Dr. Albrengues	01/01/16	12/31/17	50,000	*
Carnegie Institution for Science	Dr. Jackson	03/07/16	03/06/17	24,157	*
CSHL Association Fellowship	Drs. Pedmale/Siepel	01/01/16	12/31/16	280,000	*
Damon Runyon Cancer Research Foundation	Dr. Chio	10/01/13	09/30/17	62,000	
Enzo Life Sciences	Watson School of Biological Sciences	01/01/15	12/31/16	1,500	

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]	
European Molecular Biology Organization	Dr. Albrengues	01/15/16	10/31/16	$ 26,663	*
	Dr. Soyk	01/01/15	10/31/16	38,881	
Genentech Foundation	Watson School of Biological Sciences	10/01/15	09/30/18	75,553	
German Research Foundation (DFG)	Dr. Knauer	05/01/14	04/30/17	43,175	
	Dr. Kuettner	09/01/14	08/31/17	57,799	
Lola A. Goldring	Dr. Stillman	10/01/16	09/30/17	100,000	*
Human Frontier Science Program	Dr. Biffi	06/01/15	05/31/18	53,640	
	Dr. Carnevale	01/01/16	12/31/18	51,720	*
	Dr. Elyada	04/01/14	03/31/17	64,980	
	Dr. Huilgol	04/01/14	03/31/17	55,620	
	Dr. Livneh	01/01/14	07/22/16	32,963	
	Dr. Xu	04/01/16	03/31/19	51,720	*
Annete Kade Charitable Trust	Watson School of Biological Sciences	12/23/16	12/22/17	25,000	*
The Susan G. Komen Breast Cancer Foundation	Dr. Albrengues	11/29/16	11/28/19	60,000	*
	Dr. Arun	12/18/15	12/17/18	60,000	
	Dr. Diermeier	09/07/16	09/06/19	60,000	*
Natural Sciences and Engineering Research Council of Canada (NSERC)	Dr. Bonham	07/01/15	06/30/17	33,493	
The Patrina Foundation	Watson School of Biological Sciences/Women in Science and Engineering Initiative	03/17/16	03/16/17	10,000	*
The Pew Charitable Trust	Dr. Rodriguez-Leal	08/01/16	07/31/20	30,000	*
Republic of China (Taiwan) Ministry of Education	Y. Ho	03/30/15	03/29/18	42,000	
The Research Foundation for State University of New York—Stony Brook	B. Alagesan	04/15/16	04/14/20	4,200	*
	J. Levine	09/15/15	09/14/19	4,200	
The Simons Foundation	Dr. Chen	03/01/15	02/28/18	78,000	
	Dr. Kaufman	03/01/15	02/28/18	78,000	
The Swartz Foundation	Drs. Albeanu/Bast	01/01/16	12/31/16	60,000	*
	Drs. Koulakov/Shuvaev	01/01/16	12/31/16	35,503	*
	Drs. Zador/Vaughan	01/01/16	12/31/16	60,000	*
Uehara Memorial Foundation	Dr. Funamizu	10/01/16	09/30/17	39,447	
U.S.–Israel Binational Agricultural Research and Development Fund	Dr. Hendelman	08/01/16	07/31/17	45,000	*
Vasterbottens Lans Landsting	Dr. Olund	01/01/15	12/31/16	17,437	
The Wolkoff and Green Families	Dr. Diermeier	03/17/16	03/16/17	1,000	*
Training Support					
Lita Annenberg Hazen Foundation	Watson School of Biological Sciences	05/01/08	04/30/18	10,000	
William Townsend Porter Foundation	Undergraduate Research Support	04/01/16	03/31/17	11,800	*
University of Notre Dame	Undergraduate Research Support	04/01/16	03/31/21	20,000	*
Course Support					
American Brain Tumor Association	Brain Tumors	01/27/16	08/01/16	25,000	*
The Leona M. and Harry B. Helmsley Charitable Trust	Course Program	02/01/15	01/31/18	1,500,000	
Howard Hughes Medical Institute	Course Program	08/01/15	07/31/19	600,000	

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

Grantor	*Program/Principal Investigator*	*Duration of Grant*		*2016 Funding*[1]
The Nancy Lurie Marks Family Foundation	Autism Summer Course: Genetics & Neurobiology of Language	07/01/16	07/31/17	$ 36,000 *
	Autism Summer Course: Autism Spectrum Disorders	07/01/16	07/31/17	39,000 *
International Brain Research Organization	Brain Tumor Course	07/01/16	06/30/17	12,690 *
Meeting Support				
AbbVie Inc.	Neurodegenerative Diseases: Biology and Therapeutics	09/01/16	08/31/17	5,000 *
	HIV/AIDS Research: Its History and Future	09/14/16	09/13/17	1,500 *
Alnylam U.S., Inc.	RNA & Oligonucleotide Therapeutics	12/21/16	12/20/17	10,000 *
Avanti Polar Lipids, Inc.	The PI3K–mTOR–PTEN Network in Health and Disease	05/27/16	05/26/17	2,000 *
CSHL Translational Cancer Support	81st Cold Spring Harbor Laboratory Symposium on Quantitative Biology: Targeting Cancer	03/15/16	03/14/17	144,000 *
	Making Oxidative Chemotherapy Less Toxic	07/01/16	06/30/17	109,440 *
	Evolution and Revolution in Anatomic Pathology	09/01/16	08/31/17	115,200 *
Bill and Melinda Gates Foundation	HIV/AIDS Research: Its History and Future	07/21/16	11/30/16	35,000 *
Genomic Health, Inc.	81st Cold Spring Harbor Laboratory Symposium on Quantitative Biology: Targeting Cancer	04/22/16	04/21/17	5,000 *
Glenn Foundation for Medical Research	Mechanisms of Aging	12/30/15	12/29/16	20,000
Merck Sharp & Dohme Corp.	HIV/AIDS Research: Its History and Future	08/18/16	08/17/17	10,000 *
Pfizer Inc.	Blood–Brain Barrier	11/09/16	11/08/17	5,000 *
Proteostasis Therapeutics, Inc.	Protein Homeostasis in Health and Disease	04/28/16	04/27/17	1,000 *
Ribon Therapeutics Incorporated	The PARP Family and ADP Ribosylation	04/05/16	04/04/17	2,500 *
Library Support				
The Ellen Brenner Memorial Fund		12/15/16	12/14/17	2,000 *
G.H. Arrow Co.		01/01/13	12/31/16	700 *
The New York State Education Department		07/01/16	06/30/17	4,180 *
Christine and Stephen Zinder and Charlotte and Michael Zinder in honor of Norton Zinder		02/16/16	02/15/17	25,000 *
Laboratory Press Support				
The Lourie Foundation	BioRxiv: The Preprint Server for Biology	01/07/2015	30/06/2016	150,000

[1]Includes direct and indirect costs

*New or competing renewal grants awarded in 2016

DNA LEARNING CENTER GRANTS

Grantor	*Program*	*Duration of Grant*	*2016 Funding*[†]
FEDERAL GRANTS			
National Institutes of Health	*Barcode Long Island*	7/14–3/19	327,310
National Science Foundation	Infrastructure and Training to Bring Next-Generation Sequence (NGS) Analysis into Undergraduate Education	9/13–8/16	130,480
National Science Foundation, University of Arizona	*iPlant Collaborative*: Cyberinfrastructure for the Life Sciences	9/13–8/18	690,344
National Science Foundation	Advanced Technology Education (ATE) Program: *Genomic Approaches in BioSciences*	4/11–3/16	5,901
National Science Foundation	TRPGR; Maize Cell Genomics: Resources for Visualizing Promoter Activity and Protein Dynamics using Fluorescent Protein Lines	10/14–9/17	81,801
National Science Foundation	MaizeCode—An Initial Analysis of Functional Elements in the Maize Genome	6/16–5/19	48,731
NONFEDERAL GRANTS			
Alfred P. Sloan Foundation	*DNA Center NYC* Start-Up	12/13–6/19	14,079
Bank of America Charitable Foundation	Genetics and Biotechnology Lab	11/15–11/17	15,000
Beijing No. 166 High School	Chinese Collaboration Agreement	5/14–6/17	263,872
Health Park	*Health Park Agreement*	12/15–12/20	12,125
National Grid Foundation	Scholarships for Minority and Underserved Students in the Central Islip Union Free School District	9/16–8/17	12,000
Pinkerton Foundation	*Urban Barcode Research Program*	1/13–5/17	151,461
William Townsend Porter Foundation	*Harlem DNA Lab for Underprivileged Students*	4/16–3/17	13,500
Teva Pharmaceuticals	DNA Learning Center STEM Access Fund to Support Usage of the DNALC by Underrepresented Minorities (URM) and Disadvantaged Students	10/16–10/17	5,000

The following schools and school districts each contributed $1,000 or more for participation in the *Curriculum Study* program:

Bellmore–Merrick Central High School District	2,000	Massapequa Union Free School District	4,500
East Meadow Union Free School District	3,000	North Shore Central School District	2,000
East Williston Union Free School District	1,500	Oceanside Union Free School District	2,000
Elwood Union Free School District	3,000	Oyster Bay–East Norwich Central School District	3,500
Fordham Preparatory School	2,000	Plainedge Union Free School District	2,000
Garden City Union Free School District	2,000	Plainview–Old Bethpage Central School District	2,000
Half Hollow Schools Central School District	2,000	Portledge School	3,000
Harborfields Central School District	2,000	Port Washington Union Free School District	2,000
Herricks Union Free School District	2,000	Ramaz Upper School	2,000
Huntington Union Free School District	3,500	Roslyn Union Free School District	2,000
Island Trees Union Free School District	2,000	Syosset Central School District	3,000
Locust Valley Central School District	3,500	Yeshiva University High School for Girls	2,000

The following schools and school districts each contributed $1,000 or more for participation in the *Genetics as a Model for Whole Learning* program:

Adelphi Science and Technology Entry Program	1,400	Floral Park–Bellerose Union Free School District	7,800
Baldwin Union Free School District	1,540	Garden City Union Free School District	2,880
Bellmore Union Free School District	2,400	Great Neck Union Free School District	8,250
Berkeley Carroll Middle School	1,300	Greenwich Country Day School	2,400
Cold Spring Harbor Central School District	12,600	Half Hollow Hills Central School District	19,095
Commack Union Free School District	1,700	Herricks Union Free School District	3,815
East Meadow Union Free School District	2,719	Hicksville Union Free School District	1,400
East Williston Union Free School District	1,875	Hofstra Science and Technology Entry Program	2,500
Edgemont Union Free School District	3,150	Holy Child Academy	3,411
Elwood Union Free School District	7,075	Horace Mann School	2,640

[†]Includes direct and indirect costs.

Huntington Union Free School District	2,600
Jericho Union Free School District	13,850
Lawrence Union Free School District	2,100
Lindenhurst Union Free School District	1,000
Locust Valley Central School District	9,330
Manhattan High School for Girls	1,440
Merrick Union Free School District	1,400
Mott Hall II Middle School, New York City	1,400
Mount Vernon City School District	1,400
New Rochelle City School District	1,050
North Bellmore Union Free School District	3,330
Oceanside Union Free School District	1,625
Oyster Bay–East Norwich Central School District	2,400
Port Washington Union Free School District	10,370
PS 144, New York City	10,590
Rockville Centre Union Free School District	6,240
Scarsdale Union Free School District	6,300
Smithtown Central School District	4,400
South Huntington UFSD	5,600
St. Dominic Elementary School	4,600
St. Patrick's School	1,950
Syosset Central School District	38,550
Three Village Central School District	4,690
Trinity Regional School	1,045
Yeshiva Darchei Torah	2,963

BANBURY CENTER GRANTS

Grantor	*Program*	*Duration of Grant*	*2016 Funding*
FEDERAL SUPPORT			
National Institute of Mental Health, National Institutes of Health	NIMH Brain Camp VIII	2016	$21,280
National Institute of Mental Health, National Institutes of Health	Mammalian Brain Cell Diversity and Census	2016	35,000
NONFEDERAL SUPPORT			
Astellas Pharma Inc.	Autophagy and Cancer	2016	19,985
AstraZeneca	STAT3 in Cancer: How Can It Be Inhibited?	2016	1,495
Biotechnology Innovation Organization	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	2016	5,000
Boehringer Ingelheim Foundation	Communicating Science	2016	66,481
Boston Biomedical, Inc.	STAT3 in Cancer: How Can It Be Inhibited?	2016	63,337
Calyxt Inc.	Genomics-Enabled Accelerated Crop Breeding	2016	5,000
Cold Spring Harbor Corporate Sponsor Program	Genomics-Enabled Accelerated Crop Breeding	2016	35,666
Cold Spring Harbor Corporate Sponsor Program	Evolution of the Translational Apparatus and Implication for the Origin of the Genetic Code	2016	46,100
DRI Capital	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	2016	5,000
DuPont Pioneer	Genomics-Enabled Accelerated Crop Breeding	2016	5,000
Genentech	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	2016	30,000
Genentech	Developing Gene Editing as Therapeutic Strategy	2016	10,000
Global Lyme Alliance	Diagnostic Tests for Lyme Disease: A Reassessment	2016	48,599
International Society of Psychiatric Genetics	Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era	2016	10,000
Institute of Neurosciences, Mental Health and Addiction, University of British Columbia	Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era	2016	3,133
King & Spalding, LLP	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	2016	5,000
Lehrman Institute	Ancient DNA and Archaeology	2016	47,397
The Lustgarten Foundation	The Lustgarten Foundation Scientific Meeting/Vitamin D	2016	26,556
McDonnell Boehnen Hulbert and Berghoff LLP	Patenting Genes, Natural Products, and Diagnostics: Current Status and Future Prospects	2016	5,000
Merck Serono	Autophagy and Cancer	2016	15,000
Millennium Pharmaceuticals, Inc.	Autophagy and Cancer	2016	5,000
National Society of Genetic Counselors	Genetic Counseling for Psychiatric Disorders: Challenges in the Genomic Era	2016	10,000
NordForsk	Studying the Genomic Variation that Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems	2016	27,475
Northwell Health–Cold Spring Harbor Lab Partnership	Making Oxidative Chemotherapy Less Toxic	2016	46,901
Northwell Health–Cold Spring Harbor Lab Partnership	Evolution and Revolution in Anatomic Pathology: Automation, Machine-Assisted Diagnostics, Molecular Prognostics, and Theranostics	2016	46,156
The Norwegian Research Council	Studying the Genomic Variation That Underlies Health and Disease: The Unique Contribution of the Nordic Health Systems	2016	26,742
Novartis	Autophagy and Cancer	2016	5,000
Ovarian Cancer Research Fund Alliance	After UKCTOS: Public Messaging on Screening and Early Detection for Ovarian Cancer	2016	36,467

BANBURY CENTER GRANTS (*Continued*)

Grantor	*Program*	*Duration of Grant*	*2016 Funding*
Pfizer	Developing Gene Editing as Therapeutic Strategy	2016	$10,000
Presage Biosciences, Inc.	Autophagy and Cancer	2016	2,500
The Simons Center for Quantitative Biology	Measuring and Modeling Quantitative Sequence-Function Relationships	2016	56,686
Stand Up to Cancer	The Lustgarten Foundation Vitamin D Day	2016	2,829
The Stanley Research Foundation	Can We Make Animal Models of Human Mental Illness? A Critical Review	2016	36,216
2Blades Foundation	Genomics-Enabled Accelerated Crop Breeding	2016	5,000

CORPORATE SPONSOR PROGRAM FOR MEETINGS SUPPORT

The Corporate Sponsor Program continues to provide critical funding for the vigorous meetings programs held at Cold Spring Harbor Laboratory, whether at Grace Auditorium on the main Laboratory campus or at the Banbury Center. Without the strong foundation provided by the Program, we could neither plan with confidence for the year's meetings nor introduce new and unusual topics.

Raising funds for this program continues to be challenging even as the economy improves. We are thus especially grateful to the companies that continued to support us in 2016:

CORPORATE SPONSORS

Agilent Technologies, Inc.
Bristol-Myers Squibb Company
Calico Labs
Celgene Corporation
Genentech, Inc.
Life Sciences Solutions (part of Thermo Fisher Scientific)
Merck & Co., Inc.
New England BioLabs
Pfizer Inc.

PLANT CORPORATE ASSOCIATES

Monsanto Company

DEVELOPMENT

It is with much gratitude and pride that I am writing this report as Cold Spring Harbor Laboratory recently completed its most ambitious fund-raising effort to date. The 125th Anniversary Campaign led by Campaign Chairman Marilyn Simons raised more than $275 million in 5 years for endowment, cancer therapeutics, and science education. I cannot thank all of our donors enough and especially Board Chairman Jamie Nicholls and the entire Board of Trustees, whose gifts represented 65% of the total goal. It is because of these generous gifts that the Laboratory is in a position to strengthen and grow our institution now and into the future.

New fund-raising milestones were reached in 2016 with a record $6.93 million raised for the annual fund, along with the most ever first-time donors, employee/alumni gifts, and planned gifts. Contributing to this success was the 125th Anniversary Brick Campaign, which raised $190,000 with 330 bricks purchased. This year's Double Helix Medals Dinner was another tremendous success. We were so pleased to honor two distinguished and influential people in the field of biomedical research: Alan Alda for inspiring better public understanding of science and P. Roy Vagelos for his leadership in the pharmaceutical and biotech industries and his commitment to philanthropy.

All of us are excited by the year-end news that the drug for spinal muscular atrophy resulting from the basic research of CSHL's Dr. Adrian Krainer was approved by the FDA. We look forward to 2017 with much hope that the work of the Laboratory will continue to make an impact in the education of the next generation of scientists and in the development of new therapies for cancer and other diseases that plague humanity.

Charles V. Prizzi, *Vice President for Development and Community Relations*

CSHL scientists with Phil Donahue, Marlo Thomas, Alan Alda, and Arlene Alda at the 2016 Double Helix Medals Dinner.

Cold Spring Harbor Laboratory Corporate Advisory Board

The Corporate Advisory Board (CAB) comprises prominent business leaders from the tristate community and is a vital source of funding and outreach for Cold Spring Harbor Laboratory. Board members are the driving force behind the Laboratory's annual golf outing at Piping Rock Club, which raises critical unrestricted funding for research and education programs. CAB president Eddie Chernoff chaired the 23rd annual CSHL outing, which honored CAB member Bob Isaksen. The CAB members also participate in other events and fund-raisers for the Lab and are instrumental "ambassadors" to the community.

Corporate Advisory Board
Edward A. Chernoff, MARS, *Chair*

Michael Aboff, Aboff's Inc.
Paul Amoruso, Oxford & Simpson Realty
Todd Andrews, Centerbrook Architects and Planners
Rocco S. Barrese, Dilworth & Barrese
Edward Blaskey, TD Bank
Thomas J. Calabrese, Daniel Gale Sotheby's International Realty
Christopher Callaghan
John D. Catalano, Catalano Enterprises, LLC
Richard A. Catalano, KPMG, LLP
Jonathan Connors, Wells Fargo
Marian Conway, Ph.D., New York Community Bank Foundation
Philip D'Avanzo, Cushman & Wakefield
J. Daniel Decker, Eppendorf North America
Gregory DeRosa, Roanoke Holdings
Robert Dickstein, Ph.D.
Brian Fox, McKinsey & Company
Lawrence Goodman, Curtis, Mallet-Prevost, Colt & Mosle, LLP
Mark W. Hamer, Harvest Real Estate Services, Inc.
Mary Jane Helenek
Richard W. Humann, H2M architects + engineers
Robert Isaksen, Bank of America
Alan L. Jakimo, Sidley Austin LLP
Patricia Janco-Tupper, Wells Fargo
John C. Kean III, Kean Development Company
Norman Kelker, Enzo Life Sciences
Amy Koreen, M.D.
Laurie J. Landeau, V.M.D.
Brian D. Lee, Newmark Grubb Knight Frank
David Lessing, Lessing's
Kyle Markland, Bethpage Federal Credit Union
Victoria Sagona Meagher
Stephen F. Melore, Farrell Fritz, P.C.
Robert Palatnick, DTTC
John G. Passarelli, M.D.
David Peikon
Patricia Petersen, Daniel Gale Sotheby's International Realty
Joseph Roberto, BankUnited, N.A.
Gareth Roberts, HSBC
Matthew Roberts, NBTY
Stephen Ross, Ph.D., Nikon Instruments
Don Saladino, Drive 495
Raju Sarwal, M.D.
Craig A. Weiss, T. Weiss Realty Corp.
Andrew Whiteley
Joseph M. Willen, Advantage Title

2016 Corporate Advisory Board

CSHL President Bruce Stillman presents an award to the 2016 CSHL Golf Honoree, Bob Isaksen of Bank of America.

Cold Spring Harbor Laboratory Association

2016 CSHLA Directors

The Cold Spring Harbor Laboratory Association (CSHLA) continues to be instrumental in its support of CSHL. Members continue to generously host outreach events and introduce friends and colleagues to the Lab to raise awareness of CSHL's cutting-edge research and education programs. CSHLA President Frank O'Keefe presided over the Association and its directors for a third year, and $6.9 million was raised. Some of this important funding resulted from events that were supported by the Association, its directors, community members, major donors, foundations, and corporate sponsors. These events included the 23rd annual golf tournament at Piping Rock Club, under the leadership of Eddie Chernoff, where longtime CSHL supporter Bob Isaksen was honored; the 15th annual Women's Partnership for Science luncheon; and the 11th annual Double Helix Awards Dinner honoring Alan Alda and P. Roy Vagelos for their efforts in bringing awareness to disease research.

2016 Directors

Frank O'Keefe, *President*
Michele Munn Celestino, *Vice President*
James A. Woods, *Treasurer*
Virginia Knott, *Secretary*

Honorary Directors

Mary D. Lindsay
Anne R. Meier
Cathy Cyphers Soref
Pearl F. Staller

Elizabeth Ainslie
Lori Bahnik
Hans E.R. Bosch
Barbara W. Callaghan
Edward A. Chernoff
Susan A. Cohen
Maria H. de Lesseps
Nelson DeMille
Gina Gherlone
David Goldring
Ronald A. Gottlieb
Nancy Lippman Israeli, M.D.
Joel J. Kassimir, M.D.
Kassahun Kebede
Peter J. Klein
Laureen Stanton Knutsen
Jeffrey A. Nagel
Timothy Q. O'Neill
Maria Emanuel Ryan, D.D.S., Ph.D.
Caroline Monti Saladino
Lonnie Shoff
Hope Geier Smith
Debbie Parmet Sondock
Mary Striano
Nancy M. Szigethy
Peter S. Tilles
Lara S. Trafelet
Marjorie J. van de Stouwe, M.D.
Heather Warren Whitman

Honor Roll of Donors

Lifetime contributions of $5 million +

Anonymous
Estate of Donald Everett Axinn
Dr. and Mrs. Lalit Rai Bahl
The Arnold and Mabel Beckman Foundation
Jamie Nicholls and Fran Biondi
David Boies, Chairman of Boies, Schiller & Flexner LLP
The Breast Cancer Research Foundation
Mr. and Mrs. Landon T. Clay
The Dana Foundation
Mr.* and Mrs. Norris W. Darrell
Dart Neurosciences LLC
Estate of Kathryn W. Davis
DeMatteis Family Foundation
Mr. and Mrs. Charles F. Dolan, Dolan Family Foundation
Estate of Robert B. Gardner, Jr.
Martha Farish Gerry*, The William Stamps Farish Fund
Jacob Goldfield
Kate Medina Guthart and Leo A. Guthart
Howard Hughes Medical Institute
W.M. Keck Foundation
Mr. and Mrs. David H. Koch
Laurie J. Landeau Foundation
Mr. and Mrs. Robert D. Lindsay and Family
The Lustgarten Foundation
Lucille P. Markey Charitable Trust
Nancy Abeles Marks
The G. Harold & Leila Y. Mathers Charitable Foundation
The Don Monti Memorial Research Foundation
GEN*NY*sis grant from New York State Dormitory Authority
New York State Empire State Development Corporation
Pershing Square Foundation
Cynthia Hazen Polsky and Leon Polsky, Hazen Polsky Foundation
The Quick Family
The Robertson Family
Drs. Marilyn and James Simons
The Simons Foundation
Alfred P. Sloan Foundation
St. Giles Foundation
Theodore R. Stanley*, Stanley Medical Research Institute
The Starr Foundation
The Swartz Foundation
Thompson Family Foundation, Inc.
Dr. and Mrs. James D. Watson
Mr. and Mrs. Henry Wendt III
Roy J. Zuckerberg Family Foundation

Lifetime contributions of $1 million + or a contribution of $100,000+ in 2016

Paul G. Allen Foundation
Angel Foundation
Anonymous
La Fundación Gonzalo Río Arronte
BioBay of Suzhou Venture Capital Group
Bristol-Myers Squibb Company
John Broven
Burroughs Wellcome Fund
Mary K. Chapman Foundation
Mr. and Mrs. Edward A. Chernoff, MARS
Ellen and Casey Cogut
Mr. and Mrs. Roderick H. Cushman
Davenport Family Foundation
Michel David-Weill
Ira W. DeCamp Foundation
Oliver S. and Jennie R. Donaldson Charitable Trust
Edward P. Evans Foundation
Charitable Lead Annuity Trust, Will of Louis Feil
The Samuel Freeman Charitable Trust
Claire Friedlander Family Foundation
Coleman Fung Foundation
Robert L. Garland*
Bill and Melinda Gates Foundation
Mr. and Mrs. Robert A. Gay
Genentech
Kenyon Gillespie*
Lola Goldring
The Florence Gould Foundation
The Grace Family
Michael J. Griffin
The Harrison Foundation and The Mary Anderson Harriman Foundation
Thomas Hartman Foundation for Parkinson's Research
Janet Strauss and Jeff Hawkins
Jo-Ellen and Ira Hazan
William Randolph Hearst Foundation
The Leona M. and Harry B. Helmsley Charitable Trust
Elizabeth McCaul and Frank Ingrassia
The Karches Foundation
Mr. and Mrs. Jeffrey E. Kelter
F.M. Kirby Foundation
Walter B. Kissinger
Dr. and Mrs. Daniel F. Klessig
Mr. and Mrs. John Klingenstein, The Esther A. and Joseph Klingenstein Fund
Elise H. Knight
The Forrest C. and Frances H. Lattner Foundation
Betsy and Bryan H. Lawrence
The Lehrman Institute
Mr. and Mrs. Stephen M. Lessing
Mary D. Lindsay
The Litchfield Charitable Trust
Estate of Elisabeth Sloan Livingston
Louis Morin Charitable Trust
Lisa Lourie
Robert Lourie
Mr. and Mrs. David L. Luke III*
Josiah Macy, Jr. Foundation

2016 Double Helix honorees Alan Alda and Roy Vagelos with CSHL President Bruce Stillman and Lesley Stahl

Bruce Stillman presents award to Rich Monti, Caroline Saladino Monti, and Artie Saladino to commemorate the 10-year collaboration between CSHL and the Don Monti Memorial Research Foundation.

Merck & Company
Gillian and Eduardo Mestre
William R. Miller
Dr. and Mrs. Howard L. Morgan
Estate of Thomas C. Nelson
William C. and Joyce C. O'Neil Charitable Trust
OSI Pharmaceuticals
The Pew Charitable Trusts
Amy and John Phelan
Dr. Gregory Prelich
William and Maude Pritchard Charitable Trust
Regeneron Pharmaceuticals
Fannie E. Rippel Foundation
The Robertson Foundation
Mr. and Mrs. David M. Rubenstein
Susan Wasserstein and George Sard
The Saunders Family, Ivor Charitable Fund
Alison Holtzschue and Douglas Schloss
Eleanor Schwartz Charitable Foundation
Edith and Alan Seligson
The Seraph Foundation
Jeanne and Herbert J. Siegel
Loren Eng and Dinakar Singh
Mr. and Mrs. Howard Solomon
Mr. and Mrs. Erwin P. Staller
Mr. and Mrs. James F. Stebbins
Lee Steinberg
Dr. and Mrs. James M. Stone
Waclaw T. Szybalski, D.Sc.
Mr. and Mrs. Paul J. Taubman
Dr. and Mrs. Stuart T. Weisbrod
Whitehall Foundation, Inc.

Contributions of $50,000+

Anonymous
Cedar Hill Foundation
Dr. N. Anthony and Mrs. Robyn Coles
Mr. and Mrs. Mark Eichorn
Friends of TJ Foundation
The Clark Gillies Foundation
Joni Gladowsky Breast Cancer Foundation
Drs. Nancy Lippman Israeli and Ron Israeli
Mr. and Mrs. David M. Knott
Manhasset Women's Coalition Against Breast Cancer
The Meier and Linnartz Family Foundation
Michelle Paternoster Foundation
Robertson Family Foundation
Charles and Marie Robertson Foundation
Mr. and Mrs. Theodore Smith III, The Geier Foundation
Swim Across America Nassau/Suffolk
Dr. and Mrs. P. Roy Vagelos
The Wasily Family Foundation

Contributions of $25,000+

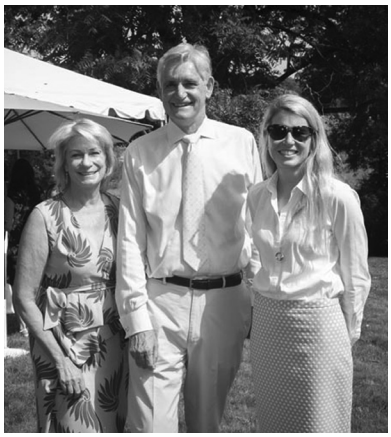
Lee and Elizabeth Ainslie
Anonymous
The Bahnik Foundation
Mr. and Mrs. Hans E.R. Bosch
David P. Brown
Calico Life Sciences LLC
Michele and Jim Celestino
Christina Renna Foundation
Mr. and Mrs. Geoffrey de Lesseps, Alpha Omega Charitable Foundation
Dr. Lester Dubnick
Ike, Molly and Steven Elias Foundation
The Go Go Foundation
Dr. and Mrs. Philip Goelet, Francis Goelet Charitable Trust
Mr. and Mrs. Ronald A. Gottlieb
Mr. and Mrs. Oliver R. Grace, Jr., OBX, Inc.
The Marc Haas Foundation
Mr. and Mrs. Mark W. Hamer
Herman and Jacquelyn Heinemann
Annette Kade Charitable Trust
Breast Cancer Awareness Day in Memory of Elizabeth McFarland
The Morningside Foundation
Mr. and Mrs. Douglas P. Morris
Mr. and Mrs. Frank T. O'Keefe
Patricia J. Petersen, Daniel Gale Sotheby's International Realty
Pfizer
William Townsend Porter Foundation
Thomas C. Quick
Kristin Olson and Dustin Smith
David M. Stark, Teva Pharmaceuticals
Stony Brook Foundation
Dr. and Mrs. Michael Wigler
Elisabeth R. Woods Foundation

Contributions of $10,000+

Austin's Purpose
Mr. and Mrs. W. Dillaway Ayres, Jr.
Broad Hollow Bioscience Park, Inc.
Judith and George Carmany
Estate of Joan L. Clarkson
Dr. and Mrs. Bayard Clarkson
Columbia University Medical Center
Barbara J. Amonson and Vincent J. Della Pietra
John and Laura Desmarais
Drs. Miki and David Donoho
Eppendorf North America
Mr. and Mrs. Alexander Farman-Farmaian
Fortunato Sons Contracting, Inc.
Putney Cloos and Brian Fox
Carolyn and Leonard Gero
Mr. and Mrs. Albert Gherlone
The Randall and Mary Hack Foundation
Susan T. Harris

Cold Spring Harbor student Elijah Gordon presents a check to David Spector for his breast cancer research. Elijah raised money donated through his fund-raising effort Kick for Cancer.

*Deceased



CSHL Board Vice Chair Marilyn Simons and Chair Jamie Nicholls with President Bruce Stillman

HRD-C Foundation
Yaoqing & Shouying Huang Family Foundation
Joan M. Hutchins
Kappa Alpha Order
Dr. and Mrs. Joel J. Kassimir
Katheryn C. Patterson and Thomas L. Kempner, Jr.
Andrea B. and Peter D. Klein
Mr. and Mrs. Ragnar Meyer Knutsen
Gordon E. Lamb, Jefferson Family Charitable Foundation
Mr. and Mrs. Thomas H. Lister
Mr. and Mrs. John Macaskill
Mrs. David J. Mahoney
Rachel and Thomas Maniatis
Edward E. Matthews
Dr. Marcia Kramer Mayer
Mr. and Mrs. Edward C. Meagher, Carter, DeLuca, Farrell & Schmidt LLP
Jennifer L. Mercer
Mr. and Mrs. Jeffrey A. Nagel
National Grid Foundation
Northwell Health
University of Notre Dame
Gerry Ohrstrom
Sue Olin
Wendy O'Neill
Mr. and Mrs. George D. O'Neill
David P. Pearson
Mrs. Richardson Pratt, Jr. and Family, The Spionkop Charitable Trust
Mr. and Mrs. Preston V. Pumphrey
Mr. and Mrs. Thomas W. Purcell
Rauch Foundation
Mr. and Mrs. John R. Reese
Reiss Family Foundation
Alexander J. Roepers
Stephen Ross, Nikon Instruments
Pamela and Richard Rubinstein Foundation
Sebonack Golf Club
Mr. and Mrs. Michael Shoff
Laura and Harry Slatkin
Deborah K. Solbert
Mr. and Mrs. Vincent Striano, Striano Electric
Stand Up For Suzanne
Dr. and Mrs. Bruce W. Stillman
The Lauri Strauss Leukemia Foundation
Nancy Szigethy
TD Bank, N.A.
Doris M. and Peter S. Tilles Foundation
Mrs. Robert W. Tilney
Mr. and Mrs. Edward Travaglianti
Mr. and Mrs. Howard J. Tytel
Wilhelmus Johannes Van der Goes
Paul and Robin Vermylen
West Islip Breast Cancer Coalition for L.I.
Mr. and Mrs. James A. Woods
Ann Eden Woodward Foundation
The WTFC Organization
The Bradley Zankel Foundation
Charlotte and Michael Zinder
Christine and Stephen Zinder

Contributions of $5,000+

Bank of America
Geoffrey Beene Foundation
The Benevity Community Impact Fund
Lisa and Tim Broadbent
Mr. and Mrs. Thomas J. Calabrese, Jr.
Keri and Eric Carlstrom
Louise Parent and John Casaly
Mr. and Mrs. John D. Catalano, Garagetek-175
Centerbrook Architects and Planners, LLC
Rita M. Cleary
Susan and Glenn Cohen
Drs. Isobel Coleman and Struan Coleman
Nathalie and Helene Comfort
Mr. and Mrs. Jonathan Connors
Mr. and Mrs. Elliot S. Conway
Marian Conway, New York Community Bank Foundation
Denise R. Coyle, The Benjamin Companies
Dau Family Foundation
Philip D'Avanzo, Cushman & Wakefield
Sandy and Nelson DeMille
Lee MacCormick Edwards Charitable Foundation
Diane and Bob Fagiola
Farrell Fritz P.C.
Michael L. Focazio, Carissa Maringo Fund
Glen Cove C.A.R.E.S.
Pierre and Paula Gonthier
H2M Architects and Engineers
Robin Hadley
The Hastings Foundation
Mr. and Mrs. Alan L. Jakimo
Patricia Janco and Rebecca Creavin, Wells Fargo Private Bank
Johnston Family Foundation
Kean Development Company, Inc.
Rita and Art Levinson
Robert T. Lindsay
Robert L. Marcus
Masthead Cove Yacht Club
Mr. and Mrs. Philip L. Milstein
Matthew Roberts, NBTY
Mr. and Mrs. Timothy Q. O'Neill
Robert and Lauryl Palatnick
Dr. and Mrs. John G. Passarelli
Seemon H. and Natalie E. Pines Foundation

CSHL scientists with Women's Partnership for Science speaker Evelyn Witkin

*Deceased

Laurie Landeau, Kate Freidman, Diane Fagiola, Connie Goldman, and John Friedman at the Double Helix Medals

Dr. Amy Rappaport Koreen and Dr. Judith Rappaport
Joe Roberto, BankUnited, N.A.
Gareth Roberts, HSBC Bank USA
Lari and Frank Russo
Maria and Charles Ryan
Dr. and Mrs. Raju Sarwal
Mr. and Mrs. Ronald Smith
Mr. and Mrs. Douglas S. Soref
J.B. Vermylen
Craig A. Weiss, T. Weiss Realty Corp.
Mr. and Mrs. W. Fifield Whitman
Joseph M. Willen, Advantage Title Agency
Mr. and Mrs. Steven Jay Winick
Mr. and Mrs. Richard H. Witmer, Jr.

Contributions of $2,500+

16W Marketing, LLC
Michael and Nancy Aboff, Aboff's
Paul Amoruso, Oxford & Simpson Realty Services, Inc.
Eileen and William Araskog
Debra and Scott Arenare
Fred H. Baron
Mr. and Mrs. Dominic P. Bencivenga
Kathy and Gene Bernstein
Brandon Sales Corp.
Dr. Steven Brenner
Frank and Deenie Brosens
Peter B. Cannell
Nancy Parmet Cook and Jim Cook
Lawrence Goodman, Curtis, Mallet-Prevost, Colt & Mosle LLP
Mr. and Mrs. Claudio Del Vecchio
Dr. Peter DeLuca
Gregory DeRosa
Dilworth & Barrese, LLP
Anna Evans
Fidelity Investments
G & M Earth Moving Inc.
The Alice Giffuni Foundation
David Goldring
James B. Hicks, Ph.D.
Mr. and Mrs. Mark Hoffman
The Hotchkiss School
Mr. and Mrs. Barclay G. Jones III
Kassahun Kebede
Mr. and Mrs. Richard P. Kleinknecht
Bernadette Ladisa and Sam Kliger
KPMG
Lisa LaRocca
The Litwin Foundation
Jennifer and Bradley Marsh
Mr. and Mrs. Michael Maturo
Michael McGlone
The Elena Melius Foundation
Newmark Grubb Knight Frank
Mrs. William H. Osborn, Jr.
Tracy Dellomo and Clifford Packingham
Betsy and Jeffrey Pash
Patricia J. Petersen
Dr. and Mrs. Anthony F. Piazza
Mrs. Joseph D. Posillico, Jr.
Nicole and Charlie Prizzi
Mr. and Mrs. John Pufahl
Drs. Marjorie J. van de Stouwe and Scott Ratner
Miriam K. Rothenberg
Diane Rothstein
Mr. and Mrs. J. Wright Rumbough, Jr.
Daniel T. Russo
Debbie Parmet Sondock
Stand Up To Cancer
Samuel L. Stanley, Jr. and Ellen Li
Carl J. Studer and Michele M. Miroff
Syosset Limousine Service, Inc.
Susie and Stanley Trotman
Kathryn and William Tyree
UBS Financial Services
United HealthCare Services, Inc.
Universe Kogaku (America), Inc.
Mr. and Mrs. Douglas A. Warner III
Jordan Woldenberg
The Richard H. Yearick Foundation

Contributions of $1,000+

Mr. and Mrs. Robert J. Adams
Anonymous
Brett Arabie
Tricia and Hal Avidano
Erica and Henry Babcock
Mrs. Robert H.B. Baldwin
William W. Bancroft and Thomas M. Bancroft Families, Norwood Foundation
Peter Baranay, Baranay Family Foundation
Carmelita S. Bautista
Elizabeth Beach
Mr. and Mrs. Anthony J. Berejka
Dr. L. Elizabeth Bertani
Mr. and Mrs. Donald Bettencourt
Bétèrre Skin + Care
Elizabeth Borden, Katharine Bradford Foundation
Drs. Renee A. Fitts and David Botstein
Ronald J. Brecht
G. Morgan Browne
Mr. and Mrs. Christopher C. Callaghan
Mr. and Mrs. William J. Candee III
Stephen and Carol Canter
Maria S. Carey

Mary Lindsay and Mardi Matheson at CSHL's annual Helix Society member lunch

*Deceased

Members of CSHL's Sarcoma Group with Associate Professor Chris Vakoc, President Bruce Stillman, and Karen Orzel

Mr. and Mrs. Gregg Carlin
Richard and Mary Catalano
Dr. Eleanor Chernick
Ann K. Thompson and James C. Childress
Jane H. Choate
Mr. and Mrs. Jeffrey T. Colon
Cornelia Corbett
Cosel-Pieper Family Foundation
Angela N.H. Creager
Marilyn F. Cunniff
The Cyphers Family
Mr. and Mrs. Jack W. Dolce
Dr. and Mrs. Thomas J. Dowling, Jr.
Mrs. James A. Eisenman
Electronix Systems Central Station Alarms, Inc.
Frederick W. Engel
Enzo Life Sciences
Mr. and Mrs. Hector Felbarg
Dr. James R. Feramisco
Food Authority
James C. Ford
Mr. and Mrs. W. Regis Fox
Christine and Douglas Fox
Bill and Rosemarie Friedman
E. Maxwell Geddes, Jr.
Drs. Joan and Alfred Goldberg
Dr. and Mrs. Walter L. Goldschmidts
Dr. Joseph L. Goldstein
Elijah Gordon
Mr. and Mrs. John S. Grace
Mr. and Mrs. Austen T. Gray, Jr.
Green Ink of Skaneateles
Mr. and Mrs. William H. Grover
Thomas C. Gsell
Reverend Linda Peyton Hancock
Dr. Gregory Hannon
Margaret M. Hargraves
Mary Beth Pearlberg and Edward Harlow
William Allan Herzog
Bob and Cindy Higginson
Susan Hirschfeld
Robert Honeyman
Frederic B. Ingraham
Bridgette and Adin Johnson
Charles and Jane Klein Family Fund
Victor R. Klein, M.D.
Mr. and Mrs. Daniel R. Knopf
Steven Lampe
Mr. and Mrs. James M. Large
Mr. and Mrs. Michael L. Leidner
Jodi and Stephen Lerner
Mr. and Mrs. Carl Lombardi
Long Island 2 Day Walk to Fight Breast Cancer
Jane N.P. Mallinson
Mary Stolfa Cancer Foundation
Mr. and Mrs. Donald S. Mazzullo
Mr. and Mrs. Charles G. McCurdy
Mr. and Mrs. Neil T. McGoldrick
Mr. and Mrs. Clarence F. Michalis
The Leo Model Foundation
Sheila and Walter Mondschein
Dr. Kathleen A. Mooney
Vicki S. and Stephen V. Murphy
Richard P. and Faye Nespola Charitable Foundation
Patricia and Hugh O'Kane
Mr. and Mrs. Richard S. Overton
Rhoda Parmet
Lisa Puntillo and Robert Pascucci
Mr. and Mrs. Nicholas B. Paumgarten
Drs. Judith and Thoru Pederson
Peto Maintenance and Development, LLC
Mr. and Mrs. Robert A. Pilkington
Ludmila Pollock
Ride For Life
Dr. and Mrs. Richard J. Roberts
Mr. and Mrs. William D. Roche
Linda K. Rodgers
Dr. and Mrs. Gerald M. Rubin
RXR Realty
Dr. Walter Schaffner
Mr. and Mrs. Peter Schiff
Robin and Enrique Senior
SH Landmark College Park Venture
Drs. Mona S. and David L. Spector
Mr. and Mrs. James L. Spingarn
Judith and Irwin Tantleff

Michele Celestino, Terry Lindsay, and Ashley O'Keefe at the Women's Partnership for Science lunch

*Deceased

Brian Taylor
Mrs. David S. Taylor
Jeremy W. Thorner
Pamela M. Thye
Sarah Vermylen Trust
Dr. and Mrs. Fredrick Valauri
Richard John Walk
Walter B. James Fund No. 2
Gerald I. White
Isabella G. Witt
Jim Wolfensohn, Wolfensohn Family Foundation
Mr. and Mrs. Stephen Worth

Corporate and Foundation Matching Grant Contributions

Air Products Foundation
Bank of America
Bristol-Myers Squibb Foundation
DTCC Giving Program
iStar Financial, Inc.
Merck Partnership for Giving
Morgan Stanley
Oxford Commercial
Verizon Foundation
Wells Fargo Foundation

Contributions in honor

Jamie Biondi
Jean Bobick
Ed Chernoff
Casey Cogut
Kristina Perkin Davison
Samantha Hallman
Robert Isaksen
Ginny and David Knott
Emma Larson
Mary D. Lindsay
Robert Lindsay
Alea Mills
Mind Over Mood
Elizabeth Moran
Justin Rosati
Susan Wasserstein and George Sard
Nikita Shoff
Dr. Bruce Stillman
Ann Maria Studer
Marjorie von Stade
Elizabeth Watson
Dr. Michael Wigler
Evelyn Witkin

Contributions in memory

Sandy Aranoff
TJ Arcati
Daniel Arena
Florence Baron
Jack Bohrer
Elias Boulukos
Jack & Mildred Bowers
Carol A. Buonaiuto
Robert Cafferkey
Joe Calarco
Vincent James Caputo
Vinny Carey
John Paul Cleary
Mary Cyphers
Norris Darrell
Mrs. Phyllis DeVita
Maryliz Dickerson
Tina Duque
Thomas English
Fred Erichsen
Claire Martin Fairman
Dominic Fontana
Ardell Bierman Fox
Nancy Slaughter Gay
E. Maxwell Geddes
Martha Gerry
Joni Gladowsky
Dr. Christopher G. Goff
Elayne Goldman
Teresa Haire
Charles E. Harris
Rhoda Holzer
Dr. Andy Kaplan
Domenick Lembo
Antonina Lipari
Susan Petersen Lubow
David Luke
Carol Marcincuk
Carissa Maringo
Dr. Egon Mayer
Elizabeth McFarland
Florence McNeill
Harold F. McNeill
Frank Morelli
Monica Mowery
Kevin O'Connor
Gerald Olin
Marilyn Palatnick
Anna Paternoster
Michelle Paternoster
Dr. Herbert Perr
Merrill Poliak
Lydia and Christopher Reeves
Christina Renna
Elizabeth Roberts
Barbara Rogers
Diane Edmin Sachs
Mr. Perry Sylvan Samuels
Mr. Peter Scalia
Charles Schmall
Helen Schmall
Andrew Seligson
Lois Ann Sepez
Suzanne Shea
Abe Siev
Adelaide Simpson
Caryl Sircus
Edward Starr
Jason Stewart
David S. Taylor
John W. Taylor
Vicky Terenzi
Dr. Uta von Schwedler
Evelyn Claire Hickey Walkowski
Dr. Daniel Wolf
Dr. Norton Zinder

*Deceased

LABORATORY MANAGEMENT

Bruce W. Stillman, Ph.D., President & CEO
W. Dillaway Ayres, Jr., Chief Operating Officer

RESEARCH

David L. Spector, Ph.D.
Director of Research

Walter L. Goldschmidts, Ph.D.
Vice President, Sponsored Programs

Sydney Gary, Ph.D.
Director, Research Operations

Professors

Douglas Fearon, M.D.
Thomas Gingeras, Ph.D.
Professor and Head, Functional Genomics
Z. Josh Huang, Ph.D.
David Jackson, Ph.D.
Leemor Joshua-Tor*, Ph.D.
Adam Kepecs, Ph.D.
Alexei Koulakov, Ph.D.
Adrian Krainer, Ph.D.
Program Chair, Cancer and Molecular Biology
Robert Maki, Ph.D.
Robert Martienssen*, Ph.D.
Program Chair, Genomics and Plant Biology
W. Richard McCombie, Ph.D.
Alea Mills, Ph.D.
Partha Mitra, Ph.D.
Adam Siepel, Ph.D.
Program Chair, Simons Center Quantitative Biology
Marja Timmermans, Ph.D.
Nicholas Tonks, Ph.D.
Deputy Director, CSHL Cancer Center
David Tuveson, M.D., Ph.D.
Deputy Director, CSHL Cancer Center
Linda Van Aelst, M.D., Ph.D.
Michael Wigler, Ph.D.
Anthony Zador, Ph.D.
Program Chair, Neuroscience

Associate Professors

Dinu Albeanu, Ph.D.
Gurinder Atwal, Ph.D.
Anne Churchland, Ph.D.
Mikala Egeblad, Ph.D.
Hiroyasu Furukawa, Ph.D.
Christopher Hammell, Ph.D.
Alexander Krasnitz, Ph.D.
Bo Li, Ph.D.
Zachary Lippman, Ph.D.
Pavel Osten, M.D., Ph.D.
Stephen Shea, Ph.D.
Raffaella Sordella, Ph.D.
Arne Stenlund, Ph.D.
Lloyd Trotman, Ph.D.
Christopher Vakoc, M.D., Ph.D.

Assistant Professors

Camila dos Santos, Ph.D.
Jesse Gillis, Ph.D.
Molly Hammell, Ph.D.
Ivan Iossifov, Ph.D.
Justin Kinney, Ph.D.
Je Hyuk Lee, M.D., Ph.D.
Dan Levy, Ph.D.
Gholson Lyon, M.D., Ph.D.
Vincent Pedmale, Ph.D.
Jessica Tollkuhn, Ph.D.
Hongwu Zheng, Ph.D.

Research Professors

James Hicks, Ph.D.
Darryl Pappin, Ph.D.
Scott Powers, Ph.D.

Research Assistant Professors

Kenneth Chang, Ph.D.
Scott Lyons, Ph.D.
Michael Ronemus, Ph.D.
Tsung Han Yeh, Ph.D.
Seungtai Yoon, Ph.D.

Senior Computer Scientist

Boris Yamrom, Ph.D.

Computer Scientists

Osama El Demerdash, Ph.D.
Vivek Kumar, Ph.D.
Ashwin Prakash, Ph.D.

Senior Research Investigators

Carrie Davis, Ph.D.
Zihua Wang, Ph.D.

Research Investigators

Joan Alexander, M.D.
Christopher Faehnle, Ph.D.
Carla Gauss, Ph.D.
Gilbert Henry Jr., Ph.D.
Navasona Krishnan, Ph.D.
Yoon-Ha Lee, Ph.D.
Youngkyu Park, Ph.D.
Michael Regulski, Ph.D.
Yi-Jun Sheu, Ph.D.
Mona Spector, Ph.D.
Priya Sridevi, Ph.D.
Asya Stepansky, Ph.D.
Herve Tiriac, Ph.D.
Vasily Vagin, Ph.D.
Huiqing Zhan, Ph.D.

Clinical Research Manager

Jessica Gotterer

Professor Emeritus

James D. Watson, Ph.D.
Oliver R. Grace Professor Emeritus/Chancellor Emeritus

Adjunct Professors

Grigori Enikolopov, Ph.D.
Gregory Hannon, Ph.D.
Scott Lowe, Ph.D.
Larry Norton, M.D.
Yi Zhong, Ph.D.

Adjunct Associate Professors

Michael Schatz, Ph.D.
Glenn Turner, Ph.D.
Doreen Ware, Ph.D.
USDA Adjunct Associate Professor

Visiting Scientists/ Collaborative Scientists

Taimour Baslan, Ph.D.
Amit Blumberg, Ph.D.
Ebru Demir, Ph.D.
Delphine Fagegaltier, Ph.D.
Bingxing Huo, Ph.D.
Yinping Jiao, Ph.D.
Xiangning Li, Ph.D.
Hysell Oviedo, Ph.D.
Natalia Peunova, Ph.D.
Ileng Ramachandran, Ph.D.
Heather Read, Ph.D.
Jason Tucciarone, Ph.D.
Kenny Ye, Ph.D.

Visiting Professors

Matthew Ridley, Ph.D.
Stuart Weisbrod, Ph.D.

Visiting Scholar

Bhubaneswar Mishra, Ph.D.

Director, Gene Targeting Shared Resources

Leyi Li, Ph.D.

Database Manager, Genomics

Christopher Zaleski

Senior Fellow

Dannielle Engle, Ph.D.

Clinical Fellow

Kenneth Yu, M.D.

CSH Fellows

Jason Sheltzer, Ph.D.
Lingbo Zhang, Ph.D.

Postdoctoral Fellows

Sandra Ahrens, Ph.D.
Jean Albrengues, Ph.D.
Xu An, Ph.D.

*Employee of the Howard Hughes Medical Institute

Gayatri Arun, Ph.D.
Lindsey Baker, Ph.D.
Balasooriya Balasooriya, Ph.D.
Sara Ballouz, Ph.D.
Leah Banks, Ph.D.
Walter Bast, Ph.D.
Yael Berstein, Ph.D.
Kuhulika Bhalla, Ph.D.
Sonali Bhattacharjee, Ph.D.
Giulia Biffi, Ph.D.
Christopher Bonham, Ph.D.
Christian Bravo-Rivera, Ph.D.
Robert Bronstein, Ph.D.
Michael Campbell, Ph.D.
Michelle Carmell, Ph.D.
Federico Carnevale, Ph.D.
Irene Casanova Salas, Ph.D.
Hong Goo Chae, Ph.D.
Cristina Chen, Ph.D.
Xiaoyin Chen, Ph.D.
Yushu Chen, Ph.D.
Quentin Chevy, Ph.D.
Iok In Chio, Ph.D.
Monika Chugh, Ph.D.
Tae Mun Chung, Ph.D.
Hannes Claeys, Ph.D.
Megan Crow, Ph.D.
Filipe de Sousa Borges*, Ph.D.
Edgar Demesa Arevalo, Ph.D.
Sarah Diermeier, Ph.D.
Max Doerfel, Ph.D.
Roman Dvorkin, Ph.D.
Elad Elkayam*, Ph.D.
Ela Elyada, Ph.D.
Gaofeng Fan, Ph.D.
Michael Feigin, Ph.D.
Wesley Frey, Ph.D.
Akihiro Funamizu, Ph.D.
Debarati Ghosh, Ph.D.
Luana Guerriero, Ph.D.
Priyanka Gupta, Ph.D.
Rasmani Hazra, Ph.D.
Xueyan He, Ph.D.
Anat Hendelman, Ph.D.
Rowan Herridge, Ph.D.
Manzar Hossain, Ph.D.
Longwen Huang, Ph.D.
Yifei Huang, Ph.D.
Yuhan Huang, Ph.D.
Dhananjay Huilgol, Ph.D.
Aman Husbands, Ph.D.
Change-Il Hwang, Ph.D.
Jonathan Ipsaro, Ph.D.
Anbalagan Jaganathan, Ph.D.
Byoung Il Je, Ph.D.
Ashley Juavinett, Ph.D.
James Jung*, Ph.D.
Matthew Kaufman, Ph.D.
Virginie Kergourlay, Ph.D.
Munenori Kitagawa, Ph.D.
Steffen Knauer, Ph.D.
Victoria Kuettner, Ph.D.
Choon-Tak Kwon, Ph.D.
Bryan Lanning, Ph.D.
Seung Cho Lee, Ph.D.
Young Koung Lee, Ph.D.
Zachary Lemmon, Ph.D.
Jiayun Li, Ph.D.
Qian Li, Ph.D.
Shujing Li, Ph.D.
Siran Li, Ph.D.
Jayon Lihm, Ph.D.
Kuan-Ting Lin, Ph.D.
Bodu Liu, Ph.D.
Lei Liu, Ph.D.
Bin Lu, Ph.D.
Jiangteng Lu, Ph.D.
Michael Ludwig, Ph.D.
Wai Kit Ma, Ph.D.
Grinu Mathew, Ph.D.
Katherine Matho, Ph.D.
Luis Mejia, Jr., Ph.D.
Olivia Mendivil Ramos, Ph.D.
Koji Miyabayashi, Ph.D.
Mehrab Modi, Ph.D.
Rodrigo Munoz Castaneda, Ph.D.
Adriana Munoz Jimenez, Ph.D.
Simon Musall, Ph.D.
Farzaneh Najafi, Ph.D.
Shinichi Nakamuta, Ph.D.
Arun Narasimhan, Ph.D.
Gal Nechooshtan, Ph.D.
Dawid Nowak, Ph.D.
Kin On*, Ph.D.
Gonzalo Hugo Otazu Aldana, Ph.D.
Torben Ott, Ph.D.
Jean-Sebastien Parent, Ph.D.
Sapan Patel, Ph.D.
Katherine Petsch, Ph.D.
Hyun Jae Pi, Ph.D.
Arnaud Pommier, Ph.D.
Mohammad Rahman, Ph.D.
Michael Regan, Ph.D.
Ashlan Reid, Ph.D.
Jie Ren, Ph.D.
Benjamin Roche, Ph.D.
Daniel Rodriguez-Leal, Ph.D.
Jae Seok Roe, Ph.D.
Nikolay Rozhkov, Ph.D.
Elena Rozhkova, Ph.D.
Thiago Santos Gouvea, Ph.D.
Nalani Sachan, Ph.D.
Juergen Scharner, Ph.D.
Hillary Schiff, Ph.D.
Andrea Schorn, Ph.D.
Mario Shields, Ph.D.
Atsushi Shimada, Ph.D.
Damianos Skopelitis, Ph.D.
Timothy Somerville, Ph.D.
Sebastian Soyk, Ph.D.
Carlos Stahlhut Espinosa, Ph.D.
Sarah Starosta, Ph.D.
Marcus Stephenson-Jones, Ph.D.
James Sturgill, Ph.D.
Xueqin Sun, Ph.D.
Yilin Tai, Ph.D.
Nami Tajima, Ph.D.
Julian Taranda, Ph.D.
Yusuke Tarumoto, Ph.D.
Ante Tocilj, Ph.D.
Claudia Tonelli, Ph.D.
Raditya Utama, Ph.D.
Alexander Vaughan, Ph.D.
Bo Wang, Ph.D.
Bor-Shuen Wang, Ph.D.
Minghui Wang, Ph.D.
Zhikai Wang, Ph.D.
Yiliang Wei, Ph.D.
John Wilson, Ph.D.
Sze Wong, Ph.D.
Qingyu Wu, Ph.D.
Cao Xu, Ph.D.
Fang Xu, Ph.D.
Xiaosa Xu, Ph.D.
Tao Yang, Ph.D.
Zhaolin Yang, Ph.D.
Kai Yu, Ph.D.
Dinar Yunusov, Ph.D.
Qiangqiang Zhang, Ph.D.
Xian Zhang, Ph.D.

Graduate Students

Brinda Alagesan
Naishitha Anaparthy
Batuhan Baserdem
Giorgia Battistoni
Benjamin Berube
Brianna Bibel
Anna Lien Bouhuis
Emilis Bruzas
Brittany Cazakoff
Kung-Chi Chang
Lital Chartarifsky
Derek Cheng
Hsiang-Chen Chou
Alberto Corona
Kush Coshic
Maria Joaquina Delas Vives
Han Fang
Miriam Fein
Talitha Forcier
William Galbavy
Nicholas Gallo
Sanchari Ghosh
Hamza Giaffar
Jacqueline Giovanniello
Kristina Grigaityte
Michael Gutbrod
Abram Handly-Santana
Yu-Jui Ho
Yixin Hu
Elizabeth Hutton
Dong-Woo Hwang
David Johnson
Justus Kebschull
Daniel Kepple
Young Jin Kim
Navyateja Korimerla
Christopher Krasniak
Matthew Lee
Jesse Levine
Shaina Lu
Laura Maiorino
Fred Marbach
Paul Masset
Katarina Meze
Judith Mizrachi
Maria Nattestad
Alexandra Nowlan
Onyekachi Odoemene
Kathryn O'Neill
Tobiloba Oni
Debjani Pal
Tzvia Pinkhasov
Sashank Pisupati
Sofya Polyanskaya
Ricardo Raudales
Annabel Romero Hernandez
Luqun Shen
Lei Sheng
Nitin Shirole
Padmina Shrestha
Sergey Shuvaev
Dongyan Song
Junyan Song
Olya Spassibojko
Martyna Sroka
Colin Stoneking
Eric Szelenyi
Ngoc Tran
Jue Xiang Wang
Xiaoli Wu
Yiyang Wu, M.D.
Wenbo Xu
Yali Xu
Ran Yan
Georgi Yordanov

*Employee of the Howard Hughes Medical Institute

Allen Yu
Chengxiang Yuan
Anqi Zhang
Qian Zhang
Siwei Zhang

Computational Science

Peter Andrews, Ph.D.
Mihail Bota, Ph.D.
Kapeel Chougule
Alexander Dobin, Ph.D.
Evan Ernst
Daniel Ferrante, Ph.D.
Elena Ghiban
Yuan Hao, Ph.D.
Ying Jin, Ph.D.
Jude Kendall
Melissa Kramer
Sunita Kumari, Ph.D.
Wen-Wei Liao
Zhenyuan Lu
Steven Marks
Andrew Olson
Ritika Ramani
Joshua Stein, Ph.D.
Marcela Tello-Ruiz, Ph.D.
James Thomason III
Kannan Umadevi Venkataraju
Liya Wang, Ph.D.
Xiaofei Wang, Ph.D.
Xuehong Wei
Matthew Wroten, Ph.D.
Miu ki Yip

Research Associates

Carmelita Bautista
Barry Burbach
Hilary Cox
Jorg Drenkow
Nour El-Amine, Ph.D.
Hyun Soo Kim, Ph.D.
Lisa Manche
Saikat Nandi, Ph.D.
Anirban Paul, Ph.D.
Diana Ravens
Suzanne Russo
Patricia Wendel
Caizhi Wu, M.D.
Melody Wu, Ph.D.
Scilla Wu

Laboratory Facility Managers

Sabrina Boettcher
Amy Brady
Patricia Brady
Julie Cheong
Robert Eifert
Stephanie Goldsmith*
Sara Goodwin, Ph.D.
Stephen Hearn
Pamela Moody
Timothy Mulligan
Keith Rivera
Joseph Simorowski*
Spencer Teplin
Ming Wang

Scientific IT Support

David Trimboli
Peter Van Buren

Scientific Administrators

Sadie Arana*
Lorraine Baldwin
AnnMarie Ballew
Christine Bedell
Patricia Bird
Barbara Cascone
Becky Dong
Susan Fredericks
Irene Gill
Carol Gundersen
Lisa Kimoto
Idee Mallardi
Sandra Moran
Elizabeth Panagot
Kylie Parker
Dorothy Tsang Hu
Madeline Wisnewski

Research Technicians

Kelli Aibel
Alexandra Ambrico
Ariana Amendolara
Andalus Ayaz
Brian Benz
Eleni Boikoglou, Ph.D.
Denise Cahn
Jacqueline Cappellani
Lauryn Cartagine
Yon Chang
Ojasvi Chaudhary
Xin Cheng
Joshua Cohen
Jonathan Crain
Brandon Da Silva
Justin Dalrymple
Cassidy Danyko
Martin Davis
Alexa DeAngelis
Amanda Epstein*
Margaret Falkowski
Christy Felice
Inessa Hakker
Kaarina Hanington
Ga-Ram Hwang
Andrew Krainer
Eunjung Lee
Ying Hsiu Liu
Kathleen Lucere
Beicong Ma
Alfonso Macias
Senem Mavruk Eskipehlivan
Nicoletta Memos
Joseph Milazzo
Kristin Milicich
Madison Miller
James Mulhern
Stephanie Muller
Mariana Neves Dos Santos Leite
Hien Nguyen
Kevin Nicholas
Aigoul Nourjanova
Jie Ouyang
Ramesh Palaniswamy
Paola Pisterzi
Umamaheswari Ramu
Michael Riggs
Julie Rosenbaum
Polona Safaric Tepes
Nikita Saha Turna
Stephen Savoia
Nicole Sayles
Christina Schoepfer
Kacey Schwartz
Regina Shaw
Om Shrestha
Noriko Simorowski
Tara Skopelitis
Yu-Chi Sun
Alexander Tolpygo
Gaurang Trivedi
Maria Cristina Velasquez Cobos
Jing Wang
Robert Wappel
Kaitlin Watrud
Shanshan Xia
Melissa Yao
Candice Megan Young
Lifang Zhang, Ph.D.

Laboratory Facility Staff

Alexander Aguilera
Damir Begonja
Lisa Bianco
Oral Bryan
Amy Burrows
Michael Cahn
Manuel Canas
Jodi Coblentz
Kristin Cox
Charlene De Poto
Maria Demott
Jared Downing
Rhonda Drewes
Eileen Earl
Wilson Garciaguirre
Jill Habel
Tristan Jarrett
Alyssa Molloy
Steven Mora
Maria Mosquera
Gustavo Munoz
Aigoul Nourjanova
Jason Pellegrini
Ann Smith
Laura Stonebridge
Walter Vigil Garcia
Peter Wimett Jr.

Media Makers/Farm Assistants

Nancy Bolanos
Terrance Chisum
Adriana Hincapie
Kelly Ann Holfester
Amy Ji
Vincent Meschan
Michele Murrell
Mariana Neves Dos Santos Leite
Hong Jie Shen
Shouyun Qiao
Sarah Vermylen

EDUCATION

Banbury Center

Jan A. Witkowski, Ph.D.
Executive Director
Michelle Corbeaux
Patricia Iannotti
Rebecca Leshan
Barbara Polakowski

CSHL Press

John R. Inglis, Ph.D.
Executive Director
Barbara Acosta
Janet Argentine
Cynthia Blaut
Carol Brown
Kathleen Bubbeo
Kathleen Cirone
Laureen Connell, Ph.D.
Marie Cotter

*Employee of the Howard Hughes Medical Institute

Laura DeMare, Ph.D.
Corrisa Farmer
Bibiane Garite
Michael Henigman
Nancy Hodson
Deborah Jarski
Catherine Kelly
Tara Kulesa
Wayne Manos
Mala Mazzullo
Joanne McFadden
Mary Mulligan
Stephanie Novara
Stephen Nussbaum
Dorothy Oddo
Robert Redmond
Theodore Roeder
Mary Sabala
Eric Sawey, Ph.D.
Diane Schubach
Richard Sever, Ph.D.
Inez Sialiano
Marcie Siconolfi
Maria Smit, Ph.D.
Hillary Sussman, Ph.D.
Linda Sussman
Jesse Ulloa
Denise Weiss

DNA Learning Center

David A. Micklos
Executive Director
Lindsay Barone, Ph.D.
Elna Carrasco
Lauren Correri
Jenna Dorey
Cristina Fernandez Marco, Ph.D.
Cornel Ghiban
Mary Lamont
Susan Lauter
Joslynn Lee, Ph.D.
Melissa Lee
Christine Marizzi, Ph.D.
Kathleen McAuley
Amanda McBrien
Pauline McGlone
Colleen McKechnie
Valerie Meszaros
Eric Nash, Ph.D.
Sharon Pepenella, Ph.D.
Carolyn Reid
Keil Thomas
Emtiaz Uddin
Jason Williams
Chun-hua Yang
Xiaoqun Zhang

Library and Archives

Ludmila T. Pollock
Executive Director
Paula Abisognio
Thomas Adams
Clare Clark
Matthew Covey, Ph.D.
Elizabeth Pessala
Stephanie Satalino
Gail Sherman
Rhonda Veros
Robert Wargas

Meetings & Courses Program

David J. Stewart, Ph.D.
Executive Director
Stephen Alexander
Alfred Baman
Margaret Bernhardt
Gerard Burlandi
Edward Campodonico
Owen Caren
Joseph Carrieri
Meredith Cassuto
Reginald Cayo
Katiria Ceballos
Maoyen Chi, Ph.D.
Wendy Crowley
Matthew Denninger
William Dickerson
Jane Donaldson, Ph.D.
James Duffy
Alicia Franco
Bradley Frey
Constance Hallaran
Heather Johnson
Patricia Klieger
Julia Kuhl
Charla Lambert, Ph.D.
Rachel Lopez
Dorothy McGee
Andrew Mendelsohn
Nicholas Moore
Maureen Morrow
Michele Navarro
Andrea Newell
Kenneth Orff
Valerie Pakaluk
Jonathan Parsons
Victor Ryder, Sr.
Stephen Silver
Mary Smith
Samuel Stewart
Vivian Vuong
James Whitaker
Barbara Zane

Office of the President

Terri I. Grodzicker, Ph.D.
Dean of Academic Affairs
Editor, Genes & Development

Watson School of Biological Sciences

Alexander Gann, Ph.D.
Lita Annenberg Hazen *Dean*
Kimberly Creteur
Kimberley Geer
Alyson Kass-Eisler, Ph.D.

ADMINISTRATIVE DEPARTMENTS

Comptrollers Office and Accounting

Lari C. Russo
Chief Financial Officer
Sharon Bense
Karine Boyce
Damian Desiderio
Maryanne Detweiler
Karen Filasky
Mary Ellen Goldstein
Dieu Huynh
Evelyn Imbriani
Mary Mastropaolo
Alison McDermott
Denise Miller
Kristine Murphy
Carolyn Nowicki
Patricia Penner
Nicole Solomon
Patricia Urena
Ann Williams

Central Administration

Margie Bell
Maureen Berejka
Bridgette Johnson
Carol Link
Denise Roberts, Ph.D.

Culinary Services

James Hope
Director
Doris Abrego
Inolka Abreu
Alan Aguirre
Jeannette Amato
Eliseo Amaya
Ingrid Amaya
Yamic Ambra
Francis Bowdren
Maryam Brinkwart
Elmer Canales
William Carmona
Dilicia Castro Najera
Francis Cayamcela
James Charvis
Marvin Chavez
Erica Douglas
Julie Ehrlich
Suzanne Esposito
Bartola Garcia
Luis Garcia
Andres Gonzalez
Silvia Gonzalez
Barbara Gordon
Jennifer Gordon
Deborah Koslosky
Pamela Lancellotti
Jennifer LoSordo
Juan Magana
Jose Maradiaga Gonzalez
Thomas McIlvaine
Laurie Moller
Jason Moss
Frederik Munter
Briana Naccarato
Joel Pena Corvera
Jose Reyes
Susana Roman
Marlenis Romero
Patricia Sawyer
Robert Scherer
Karen Scola
Shannon Silva
Jose Suarez
Jose Velasquez
Hector Vigil Garcia
Maria Villatoro
Maria Vilorio
Tracy Woolfson
Dennis Zelaya
Jennifer Ziegler
Lidia Zuleta

Daycare Center

Deborah Keller
Director
Michele Bauman
Cassandra Bense
Margie Davis
Emma DeDora
Stephanie Dennis
Karen Flore
Aileen Gaines

*Employee of the Howard Hughes Medical Institute

Sara Genovese
Colleen Hewson
Ivonne Huda
Jamie John
Stefanie Levitz
Cynthia Lyon
Stephanie Magana Contreras
Paula Messinger
Lucinda Mow
Wendy Reyes Ortez
Meryl Rosen-Wells
Helen Sacklow
Kathleen Sausto
Judith Smith
Shannon Smith
Keri Spielberger

Development

Charles V. Prizzi
Vice President, Chief Development Officer
Jean DeForest
Diane Fagiola
Karen Orzel
Wendy Schneps
Bailey Stark
Jill Stone

Facilities

Arthur Brings
Vice President, Chief Facilities Officer
Robert Albinson
Brian Allen
Leslie Allen
Steven Amatrudo
Edward Anderson
Peter Anderson
Jerry Armstrong
Oshane Bell
William Bishop
Keith Brown
Caitlin Carroll
Dessie Carter
Daniel Chapman
Anthony Ciccarella
Christopher Clark
Robert Collins
Drew Comer
Gabriela Cruz
Matthew D'Amico
Peter Dale
John Damico
Nicholas Dassau
GianPaolo Dato
James Dean
Marcus Diaz
Robert Dickerson
Glen DiMaria
Steven Dlugolonski
Kevin Donohue
Michelle Doran
Sarah Drayzen
Wesley Dreusike
Paul Edwards
Elias Eifert
Rafael Estrella
Timothy Foley
Robert Gensel
Robert Gorney
Erick Greene
Leo Greene
Stephen Gregorovic
Wilman Gutierrez
Wayne Hamilton
Gerard Holler
Joseph Houser
Matthew Hoyer
Christopher Hubert
Louis Hunter
Michael Hutchinson
Randal Jones
Daniel Jusino
Kevin Keenan
Jeffrey Klaverweiden
Oscar Lastra
Joseph Leonard
Walter Leute
Jian Lin
Daniel Lopez
Charles Mandell
Gregory Marotta
Patricia McAdams
Tahid McCoy
Christopher McEvoy
Kevin McHugh
Riley McKenna
Daniel Miller
David Miritello
Oscar Moreira
Guido Mosquera
Mary Muno
Michael Murphy
Kenneth Nastri
Herbert Neazer
Michael Nolasco
Michael Noone
Wayne Pav
Christopher Pearson II
Jose Pena Corvera
Saul Pena Corvera
John Pisciotta
Manuel Ramones
Wilson Ramones
Ronald Ranaldo Jr.
Steven Randall
Jaqueline Reynolds
Alfredo Rivera Jr.
Ronald Romani
Frank Russo
Richard Rutherford
Eric Sappell
Steven Sardone
John Schindler
Claudia Schmid
Lauren Schmidt
Peter Schuman
Cesar Sisalima
Steven Tang
Hock Yew Teo
Albaro Trinidad
Leigh Tyrrel
James Van Wie
Fredy Vasquez
Trinidad Velasquez
Wilfredo Velasquez
Benjamin Veneable III
Alvin Watson
Kyle Welton
Leslie Wenzel
Tracey Anne White
Graham Wildt
Daniel Zimmer

Human Resources

Katherine G. Raftery
Vice President
Andres Alarcon
Wendy Alexander
Camille Cava-Belluscio
Patricia Crawford
Jessica Heckel
Lesley Inglis
Laura Magri
Laura Moran
Marygrace Navarra
Elsy Padro
Victoria Panebianco
Marjorie Ryan
Patricia Schoeffler
Lauren Warmbrand

Information Technology

Hans Erik Aronson
VP Information Technology/CIO
Alex Barton
Edward D'Azzo-Caisser
Darian Defalco
Daniel Drayzen
Frank Falco
Kelvin Flores
Todd Heywood, Ph.D.
Thomas Karow
Leila Kelly
Sean Kelly
Caryn Koza
Michael Malave
Louis Malfi
Bhargavkumar Patel
Barbara Peters
Vincent Reale
Salvatore Serafino
Michael Skuthan
David Jonathan Torres
Steven Zielinski

Procurement

Michael Marchesiello
Vice President
Deborah Aufiero
Kimberly Bronson
Carol Brower
Susan De Angelo
Jeffrey DuPree
Elizabeth Janow
Christopher Oravitz
Barbara Purcell
Jorge Ramirez
Krystyna Rzonca
Theresa Saia

Public Affairs

Dagnia Zeidlickis
Vice President, Communications
Andrea Alfano
Margot Bennett
Jessica Giordano
Philip Renna
Susan Runkowski
Brian Stallard
Peter Tarr, Ph.D.

Research Operations

Diane Esposito, Ph.D.
Research Compliance/Research Investigator
Robert Gerdes
Jaclyn Jansen, Ph.D.
Julie Sutherland

*Employee of the Howard Hughes Medical Institute

Sponsored Programs

Walter L. Goldschmidts, Ph.D.
Vice President
Irina Armstrong
Sara Jane Beck
Rachel Desrochers
Carol DuPree
Diane Errico
David Garner
Donna Gavin
Jill Hemish, Ph.D.
Joanne Janelli
Ruta Ledins
Andrea McCarthy
Dave Neale
Catherine Perdikoylis
Jaclyn Tenaglia

Technology Transfer

Teri Willey
Vice President, Business Development/Technology Transfer
Fauzia Chaudhary, Ph.D.
Elizabeth Cherian-Samuel
Vladimir Drozdoff, J.D., Ph.D.
Shiu Yie Hsieh
Lorraine McInerny

Legal

Debra Arenare, J.D.
Vice President, General Counsel
Alexander Sosnovski

RESEARCH STAFF DEPARTURES DURING 2016

Associate Professor

Joshua Dubnau

Senior Fellow

Olga Anczukow-Camarda

Visiting Professor

Kathleen Rockland

Research Investigators

Eric Antoniou
Simon Knott
Keerthi Krishnan

Visiting Scientists/ Collaborative Scientists

Martin Akerman
Pedro Alexandre Garcia Da Silva
Sassan Azarian
Shi Bai
Saikat Chatterjee
Dong Ding
Ahmed Elewa
Daniel Grimanelli
Zhou Han
Balazs Hangya
Alexandra Hoermann
Tomas Hromadka
Jaikishan Jayakumar
Hyun Yong Jeon
Duda Kvitsiani
Deguang Liang
Lisa Prazak
Junwei Shi
Mikkel Terp
Iuliu Vasilescu
Min Zhang

Postdoctoral Fellows

Rhys Adams
Jennifer Beshel
Yung-Heng Chang
Priscila Ferreira Papa
Jong-Jin Han
Erkan Karakas
Billy Lau
Fei Li
Uri Livneh
Tomas Malinauskas
Almudena Molla-Morales
Christos Noutsos
Jaclyn Novatt
Erik Olund
Juwon Park
Mahshid Rahmat
Ileng Ramachandran
Meng-Fu Shih
Joanna Szczurkowska
Yiqun Yao
Yang Yu
Wei Zheng

Computational Science

Gregor Havkin
Ferenc Mechler
David Molik
Joseph Mulvaney
Srividya Ramakrishnan
Sotirios Raptis

Research Associate

Linda Rodgers

*Employee of the Howard Hughes Medical Institute