

COPY-NUMBER ANALYSIS AND HUMAN DISEASE

M. Wigler

G. Alex
C. Algieri
S. Chakraborty
K. Cook
Y. Eberling

D. Esposito
V. Grubor
J. Hicks
L. Hufnagel
R. Kandasamy

J. Katz
J. Kendall
A. Krasnitz
Y.-H. Lee
A. Leotta

D. Levy
S. Marks
J. Meth
L. Muthuswamy
N. Navin

D. Pai
M. Riggs
L. Rodgers

M. Ronemus
J. Troge
B. Yamrom

MAMMALIAN GENETICS

We study variations in the human genome and their association with disease and genetic disorders. The variations we study arise when a large segment of the genome is duplicated or deleted. Such copy-number variations, or CNVs, can arise somatically or in the germ line. Somatic variations are often seen in cancer, and they distinguish cancer cells from the normal cells of the body. They provide clues for the origin and behavior of the cancer and possibly its early detection. Germ-line variations distinguish individuals from one another and may be inherited, in which case, they are known as copy-number polymorphisms, or CNPs, or they may arise spontaneously, in which case, they serve as engines of human diversity or can cause devastating genetic disorders, such as autism.

Part of our lab uses copy-number data and DNA-methylation status to study solid cancers, especially breast cancer (Hicks et al., *Genome Res.* 16: 1465 [2006]) and B-cell chronic leukemia (Grubor et al. 2008). We seek to identify the genes most frequently mutated in cancers and leukemias and to determine whether genomic data can be used to predict the outcome of the disease, its response to therapy, and the early detection of its recurrence. Another part of our lab studies copy-number variation that may underlie autism, congenital heart defects, and other profound disorders of normal human development. A large portion of the lab is composed of quantitative/computational scientists who work on methods for data processing, visualization, and interpretation.

Until recently, our studies were largely based on a high-throughput, high-resolution, microarray technology developed in our lab called representational oligonucleotide microarray analysis (ROMA) (Lucito et al., *Genome Res.* 12: 2291 [2003]), which itself was based on an earlier technology developed in our lab called representational difference analysis (RDA) (Lisitsyn et al., *Science* 259: 946 [1993]). RDA was previously used to discover tumor suppressors, oncogenes, and pathogens. ROMA is one form of a more general technology called comparative genome hybridization,

or CGH. Methodology is constantly evolving, and part of our laboratory works on an extension to ROMA called methylation detection oligonucleotide microarray analysis (MOMA) used to study DNA methylation. We still use ROMA, for example, when the amount of a sample is limiting, but our CGH work is now mainly based on commercial platforms that have higher resolution than ROMA. We have recently been investigating the applications of high-throughput, single-molecule, DNA sequencing to discover and detect copy-number variation, even at the level of single cells.

CANCER AND LEUKEMIA

Our work in cancer applies genome analysis to both basic and clinical problems. This year, we continued our clinical studies on breast cancer, with an emphasis on marker discovery and detection, that we hope will directly benefit patient care. We participate with the Memorial Sloan-Kettering Cancer Center and Yale University in the retrospective genomic analysis of breast cancer cases from previous clinical trials designed to discover better markers for the successful response to chemotherapeutic drugs. Our present work suggests the superiority of CGH to fluorescent in situ hybridization (FISH) for the evaluation of responsiveness to Herceptin. We also have continued our collaboration with the Radium Hospital in Oslo, Norway, in clinical trials initiated to understand the role of bone marrow micrometastasis in the clinical outcome of breast cancer. We use both copy-number and DNA methylation to find markers for survival.

We continue our efforts to find cancer genes in breast cancer, in collaboration with two Scandinavian centers, led by Anders Zetterberg at the Karolinska Institute, Sweden, and Anne-Lisa Borresen-Dale at the Radium Hospital, Norway, as well as Cliff Hudis and Larry Norton at the Memorial Sloan-Kettering Cancer Center (MSKCC), New York. These studies have elucidated a set of loci, called epicenters, that are the recurrent sites for genome amplification and deletion in breast cancer. We and many others presume that the epicenters are

locations where the genes that drive the progression of malignancy reside. Hence, the elucidation of these loci may facilitate less-expensive tools for the evaluation of the clinical state of cancers and ultimately to gene discovery.

We have continued our studies on tumor heterogeneity. Clues to the progression of cancer, and its clinical profile, are missed if we assume that each individual's cancer is a single clone with a single molecular history. Work in progress indicates that at least half of breast cancers are composed of multiple clones. Our work proves that sampling one region of the cancer for molecular markers will often yield an incomplete picture of the state of the cancer. The careful analysis of clonal variety should yield insight into breast cancer: the order and mechanism by which it evolves and even if it has already metastasized.

Our studies on tumor heterogeneity have motivated us to attempt analysis of the genome of single cancer cells. Preliminary studies suggest that it is technically feasible to sequence the DNA of single cells. We use the DNA sequence reads as "counters" to determine the amplification/deletion pattern in the genome. We are preparing to use this method to detect the presence of cancer cells in biopsies, which could eventually lead to clinical tests for cancer recurrence and risk assessment.

In our studies of B-cell chronic leukemia (CLL), a collaboration with Nick Chiorazzi of North Shore University Hospital, Manhasset, New York, we have used ROMA to identify essentially all of the known recurrent lesions observed in that disease and discovered several new epicenters in addition. As with breast cancer, we now have evidence of multiple clones of common origin in leukemia from the same patient. Using high-resolution microarrays, we also have found preliminary evidence for a set of signature genomic events in CLL at a set of transcriptional regulator genes. We are presently investigating whether these signature events are caused by deletion or are in some way artifacts of chromatin restructuring specific to CLL.

We have made significant progress in our studies of DNA methylation in cancer. We previously described a microarray platform (MOMA) designed to detect methylation of CpG islands (Kamalakaran et al., *Nucleic Acids Res.*, in press). In our studies of samples supplied from the Radium Hospital in Norway, we have identified CpG-rich islands that change methylation state during the initiation and progression of breast cancer. The genome-wide methylation patterns clearly distinguish normal tissue and tumor tissue and might be useful for the analytic deconvolution of complex tissue mixtures into their cellular components. The methylation patterns

cluster into subgroups that correlate with genome and expression profile clustering. In collaborations with Greg Hannon and Dick McCombie at CSHL, we have used DNA sequencing to demonstrate that there are DNA methylation switch zones. Detection of methylation switches in these zones may facilitate cancer detection and clinical subtyping.

GENETIC DISORDERS

After our discovery that copy-number variation is common in the human gene pool (Sebat et al., *Science* 305: 525 [2004]), we studied the role of CNVs in human disease and, in particular, in the role of spontaneous (or de novo) CNVs in the germ line. Our findings established that germ-line mutation was a more significant risk factor for autism spectrum disorders than previously recognized (Sebat et al., *Science* 316: 445 [2007]) and established a new approach for the further study of the genetic basis of this and other genetic disorders, such as schizophrenia and congenital heart disease.

One of the de novo events we identified was a deletion on 16p. This event has now been shown by two other groups to explain perhaps as much as 1% of autism. We assisted Alea Mills at CSHL to engineer mice with the orthologous deletion on mouse chromosome 7. We are hopeful that these mice will provide animal models suitable for understanding the underlying neuropathology of the condition and the search for palliative treatments.

Analysis of autism incidence in families, a collaboration with Kenny Ye at Albert Einstein School of Medicine, provided evidence for a unified theory of the genetic basis for the disorder (Zhao et al., *Proc. Natl. Acad. Sci.* 104: 12831 [2007]). Autism families are divided into simplex (only one affected child) and multiplex (multiply affected children). By inspecting the records from the AGRE consortium, we found that the risk to a male newborn in an established multiplex family is nearly 50%, the frequency expected of a dominant disorder. We compute that autism incidence and sibling concurrence rates are consistent with a model in which new mutations with a strong contribution arise sufficiently frequently to explain the majority of autism and are passed from a resistant carrier parent to a child in a dominant fashion. One major source of resistance is gender. Autism has a much lower incidence in females.

We are now in the midst of a larger study of spontaneous mutation in autism, based on a population of simplex families collected by the Simons Foundation. Early initial results confirm our previous findings, and we

observe de novo mutation more frequently in children with autism than in their unaffected siblings. Because our new studies are performed with higher-resolution microarrays, we also see many more examples of narrow new mutations (altering only a few genes), thus expanding our list of good candidate genes involved in the disorder. There is a male bias to the detection of narrow mutations, but we see little gender bias to detecting broad mutations (altering many genes). Because there should be no gender bias in the incidence of new mutation, the detection biases suggest to us that at least two contributory genes are targeted in the broad lesions, equalizing susceptibility in males and females. One clear benefit of our large study will be in the design of clinical tests for genetic counseling and early detection. Eventually, this knowledge will contribute to finding and matching patients with the most effective therapies.

DATA GENERATION AND ANALYSIS

The major part of our group's effort centers on the generation, analysis, and interpretation of high-volume data. This entails developing protocols for conducting microarray experiments, determining quality control, probe evaluation, signal extraction, and segmentation (the method of "observing" copy-number variation); comparisons of sets of experiments, including new statistical measures, data reduction, and data summary; and

construction of databases so that we can communicate our results to other investigators. Our novel contributions include methods to attenuate system noise in array hybridizations, parametrize hybridization performance, detect and correctly call regions of genetic polymorphism, detect de novo events, classify cancers for outcome, and define the epicenters of genetic change in cancers and leukemias.

PUBLICATIONS

- Alarcon, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H., et al. 2008. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82**: 150–159.
- Grubor, V., Krasnitz, A., Troge, J.E., Meth, J.L., Lakshmi, B., Kendall, J.T., Yamrom, B., Alex, G., Pai, D., Navin, N., et al. 2008. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* **113**: 1294–1303.
- Xue, W., Krasnitz, A., Lucito, R., Sordella, R., Van Aelst, L., Cordon-Cardo, C., Singer, S., Kuehnelt, F., Wigler, M., Powers, S., et al. 2008. DLC1 is a chromosome 8p tumor suppressor whose loss promotes hepatocellular carcinoma. *Genes Dev.* **22**: 1439–1444.
- Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., et al. 2008. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135**: 852–864.



K. Cook