COPY-NUMBER ANALYSIS AND HUMAN DISEASE

M. Wigler G. Alex, C. Algieri, S. Chakraborty, K. Cook, D. Esposito, V. Grubor, J. Hicks, L. Hufnagel, R. Kandasamy, 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 that 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. The former are often seen in cancer, distinguish cancers from the normal cells of the body, and provide clues for the origin and behavior of the cancers. The latter distinguish individuals from each other and may be inherited, in which case, they are known as copy-number polymorphism, or CNPs, or they may arise spontaneously, in which case they serve as engines of human diversity, and can cause devastating genetic disorders, such as autism.

Our studies to date are largely based on a high-throughput high-resolution microarray technology developed at CSHL called ROMA (representational oligonucleotide microarray analysis), which itself was based on an earlier technology developed at CSHL called RDA (representational difference analysis) that was used for discovering tumor suppressors, oncogenes, and pathogens. ROMA is a form of a more general technology called CGH, or comparative genome hybridization. CGH is evolving, and part of our laboratory works on technical improvements and extensions, such as the use of a derivative technique called MOMA (methylation detection oligonucleotide microarray analysis) to study DNA methylation, and on more powerful statistical methods for data interpretation. Part of the lab uses copy-number data and DNA methylation status to study solid cancers, especially breast cancer and B-cell chronic leukemia (B-CLL). We seek to identify the genes most frequently mutated in cancers and leukemias and, in general, to determine whether genomic data can be used to predict the outcome of the disease and its response to therapy. Finally, part of the lab studies CNVs that may underlie autism, congenital heart defects, and other profound disorders of normal human development.

CANCER AND LEUKEMIA

In this past year, our studies on both breast cancer and leukemia have moved from the discovery phase toward clinical trials that will directly affect patient care. Our work to date indicates the superiority of CGH to the traditional clinical standard of genome analysis: fluorescent in situ hybridization (FISH). In the breast cancer area, we are participating in the evaluation of a clinical trial initiated at the Radium Hospital in Oslo, Norway that is designed to understand the role of bone marrow micrometastasis in the clinical outcome of breast cancer. The evaluation of copy-number data with clinical parameters in this trial is still in progress. We have also enrolled as collaborators with Memorial Sloan-Kettering Cancer Center (MSKCC) and Yale University in the retrospective molecular genomic analysis of cases from previous clinical trials designed to discover better markers for successful treatment with chemotherapeutic drugs.

We continue to focus on samples from two Scandinavian collaborations, with Anders Zetterberg at the Karolinska Institute, Sweden, and Anne-Lisa Borresen-Dale at the Radium Hospital, Norway, and with Larry Norton at 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. The set of breast cancer epicenters overlap with epicenters from lung cancer (data from Scott Powers and David Mu, CSHL) but are clearly a distinct set. In fact, we can distinguish breast cancers from lung cancers largely by the loci involved in amplification and deletion, a method that may be useful in a clinical setting. The epicenters are locations where many of the genes that drive the progression of malignancy reside, and hence, the elucidation of these loci may facilitate drug design and the production of focused and less-expensive tools for the evaluation of the clinical state of cancers.

We have begun a series of 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, proves that sampling one region of the cancer for molecular markers yields an incomplete picture, and provides

clues about the origins of breast cancer: the stages through which it passes and the speed with which it evolves.

Our studies of B-cell leukemia, a collaboration with Nick Chiorazzi of North Shore University Hospital, Manhasset, New York, uses ROMA to identify essentially all the known recurrent lesions that have been observed in that disease, as well as several new epicenters. We are in the process of designing B-CLL "tiling" arrays that will allow us to examine the leukemic epicenters at greater resolution, so that we will be able to narrow the gene candidates in each region and assess their recurrence with greater accuracy. Such a B-CLL chip may enable oncologists to rapidly assess the progression of the disease and guide decisions about therapy. As with breast cancer, we now have evidence of multiple clonogenic histories of leukemia in the same patient.

We have made significant progress in our studies of DNA methylation in cancer. Last year, we described results achieved, in collaboration with Rob Lucito here at CSHL, on a microarray platform (MOMA) designed to detect methylation of CpG islands. Now, we have identified genes that change methylation state during the initiation and progress of breast tumors. In our first efforts at determining the clinical significance of methylation, we have analyzed approximately 80 tumors from a clinical trial initiated at the Radium Hospital in Oslo, Norway, that is designed to understand the role of bone marrow micrometastasis in the clinical outcome of breast cancer. The results of methylation microarray analysis on this collection of patient samples along with a separate collection of breast tumors and adjacent normal breast tissue show that the methylation patterns clearly differ between normal tissue and tumor tissue. More surprising, however, is the result that the tumor DNAs naturally cluster into subgroups using methylation data alone. These subgroups match the clustering determined by gene expression. These studies will help us to understand the role of methylation in cancer etiology and facilitate clinical subtyping.

GENETIC DISORDERS

Since our seminal discovery that CNV is common in the human gene pool, we have studied the role of CNVs in human disease and, in particular, the role of spontaneous or de novo CNVs in autism. In a collaboration that continued with Jonathan Sebat here at cSHL, we tested the hypothesis that de novo CNV is associated with autism spectrum disorders (ASD). We performed CGH on the DNA of patients and unaffected subjects to detect copy-number variants not present in their respective parents. Candidate genomic regions were validated by higher-resolution CGH, paternity testing, cytogenetics, FISH, and microsatellite genotyping. Confirmed de novo CNVs were significantly associated with autism. Such CNVs were identified in about 10% of patients with sporadic autism, in about 2% of patients with an affected first-degree relative, and in only about 1% of controls. Most de novo CNVs were smaller than microscopic resolution. Affected genomic regions were highly heterogeneous and included mutations of single genes. These findings establish de novo germ-line mutation as a more significant risk factor for ASD than previously recognized and clearly point to a new approach for the further study of the genetic basis of this and other genetic disorders, such as schizophrenia and congenital heart disease.

Consonant with the finding of spontaneous CNVs in humans, studies in mice, a collaboration with Ira Hall here at CSHL, have demonstrated that de novo CNVs occur frequently in mice lineages. These studies indicated that certain loci are vastly more unstable than others and is consonant with the observation that many of the de novo changes in humans are recurrent.

Additional analysis of autism incidence data, a collaboration with Kenny Ye at Albert Einstein School of Medicine, has provided evidence for an important genetic model for the disorder. Autism families are divided into simplex (only one affected child) and multiplex (multiply affected children). In the latter case, we found by inspecting the records from the AGRE consortium that the risk to a male newborn in an established multiplex family is nearly 50%, the frequency expected of a dominant disorder. Our analysis of polymorphism data virtually rules out that this can be explained as inheritance on the X-chromosome. We find that overall autism incidence is consistent with a model in which new mutations arise that have a strong contribution to the disorder and are passed from a resistant carrier parent to a child in dominant fashion. One major source of resistance is in fact gender, and a singularly important fact about autism is its lower incidence in females.

We are now in the midst of preparing for a much larger-scale study of spontaneous mutation in simplex autism, dependent on a population of families being organized by the Simons Foundation. Hopefully, this 3-year study will lead to the identification of recurrent loci that contribute to the majority of autism risk and might thereby elucidate convergent pathways involved in the etiology of the disorder, possibly leading to treatments in some cases. One clear benefit of such a large study will be in the design of clinical tests for genetic coun-

seling, early detection, and perhaps 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 organization of data. This entails developing protocols for conducting microarray experiments; determining quality control, probe evaluation, signal extraction, and segmentation (the method of "observing" CNV); 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. Although we usually do not report our methods in this forum, we are in the midst of preparing several manuscripts in which our new statistical, mathematical, and computational ideas have a prominent role. Hopefully, we will detail them in next year's report.

PUBLICATIONS

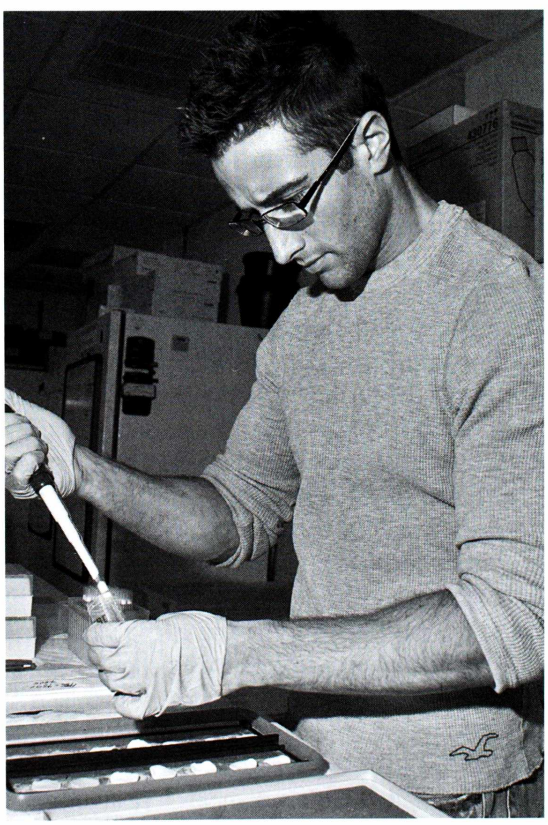
Egan C.M., Sridhar S., Wigler M., and Hall I.M. 2007. Recurrent DNA copy number variation in the laboratory mouse. *Nat. Genet.* **39:** 1384–1389.

Kendall J., Liu Q., Bakleh A., Krasnitz A., Nguyen K.C., Lakshmi B., Gerald W.L., Powers S., and Mu D. 2007. Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc. Natl. Acad. Sci.* **104:** 16663–16668.

Lucito R., Suresh S., Walter K., Pandey A., Lakshmi B., Krasnitz A., Sebat J., Wigler W., Klein A.P., Brune K., et al. 2007. Copy-number variants in patients with a strong family history of pancreatic cancer. *Cancer Biol. Ther.* **6:** 1592–1599.

Sebat J., Lakshmi B., Malhotra D., Lese-Martin C., Troge J., Walsh T., Yamrom B., Yoon S., Krasnitz A., Kendall J., et al. 2007. Strong association of de novo copy number mutations with autism. *Science* **316:** 445–449.

Zhao X., Leotta A., Qiu S., Kustanovich V., Lajonchere C., Geschwind D.H., Lord C., Sebat J., Ye. K., and Wigler M. 2007. A unified genetic theory for sporadic and inherited autism. *Proc. Natl. Acad. Sci.* **104:** 12831–12836.



Alex Garrick