COPY-NUMBER ANALYSIS AND HUMAN DISEASE

M. Wigler M. Chi, D. Esposito, V. Grubor, J. Hicks, J. Hicks, L. Hufnagel, J. Kendall, A. Krasnitz, Y-H. Lee, E. Leibu, A. Leotta, J. Meth, L. Muthuswamy, N. Navin, D. Pai, M. Riggs, L. Rodgers, J. Troge, B. Yamrom, S. Yoon

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 and distinguish cancers from the normal cells of the body, in which case, they provide clues for the origin and behavior of the cancers. The latter, germ-line CNVs, distinguish individuals from each other 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 creation of human diversity and also can cause devastating genetic disorders, such as autism.

Our studies are 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 that was used for genomic subtraction. 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 ROMA to study DNA methylation, and on more powerful statistical methods for data interpretation. Part of the lab uses the copy-number data and methylation status to study solid cancers, especially breast cancer, and leukemias, especially B-cell chronic leukemia (B-CLL). We seek to identify the genes most frequently mutated in cancers and leukemias, to distinguish solid cancers from each other, and in general to determine whether the genomic data that we generate can be used to predict the outcome of the disease and its response to therapy. Finally, part of the lab studies CNV that may underlie autism, congenital heart defects, and other profound disorders of normal human development.

CANCER

Our studies of breast cancer have largely drawn upon samples from two Scandinavian collaborations: Anders Zetterberg at the Karolinska Institute, Sweden, and Anne-Lisa Borresen-Dale at the Radium Hospital, Norway, and with Larry Norton at the Memorial Sloan-Kettering Cancer Center (MSKCC), New York. These studies have elucidated a set of loci, called epicenters, which 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 reside many of the genes that will make good tumor markers or therapeutic targets. Work in progress suggests that by dissecting tumors, we may be able to determine the time sequence of the amplifications and deletions, enabling us to determine the earliest genomic events. In a collaboration with the laboratories of Scott Lowe, Scott Powers, and Rob Lucito (all of CSHL), we have found that animal models will be a further source of information about these epicenters (Zender et al. 2006)

We documented a new form of genomic instability that is very common in breast cancer. There is a good correlation between the number of genomic events, and the type of genomic instability, with survival (Hicks et al. 2006). In working with Dr Zetterberg and pathologist Kiki Tan at MSKCC, we have shown an excellent correlation between CGH and fluorescent in situ hybridization (FISH), a mainstay of clinical pathology (see also Navin et al. 2006). CGH is in many ways more reliable than FISH, and we anticipate that FISH will be replaced by CGH, particularly for tests that oncologists use to decide on treatment.

In collaboration with David Botstein and Robert Pelham of Princeton Univesity, New Jersey, we have used mouse ROMA (Lakshmi et al. 2006) to determine how many of the stromal cells of human tumors transplanted in mice are likely to be both mutant and clonal (Pelham et al. 2006). We do not know whether these mutant stroma are selected by the tumor from

circulating cells or whether they become mutated during stromal growth. But the results open up a new window on host-cancer interactions.

Our studies of B-CLL are a collaborative effort with Nick Chiorazzi of North Shore University Hospital, Manhasset, New York. In these studies, we have used 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.

Utilizing a new microarray designed by Robert Lucito here at CSHL, we have begun to assess the methylation status of CpG islands in the genomes of cancer and leukemia cells. These islands are rich in sites for DNA methylation, and the status of methylation at these sites may be clues to cell lineage and cancer progression. Although still preliminary, the results suggest that the vast majority of CpG islands in all cells are partially methylated, and we can readily distinguish cell-specific patterns by the small number of recurrent sites that have become hypo- or hypermethylated. Only a few of the islands change status as the tumors progress. The clinical significance of these changes remains to be evaluated. Nevertheless, we expect that determination of methylation state in the DNA from a biopsy can be used to determine the origins of a tumor, as well as determine the cellular composition of a biopsy, both useful objectives in a clinical setting.

GENETIC DISORDERS

Since the seminal discovery that CNV is common in the human gene pool (Sebat et al., *Science 305:* 525 [2004]; Iafrate et al., *Nat. Genet. 36:* 949 [2004]), in collaboration with the laboratory of Jonathan Sebat here at CSHL we have focused on the role of CNVs in human disease and, in particular, on the role of spontaneous or de novo CNVs (see, e.g., Jobanputra et al., *Genet. Med. 7:* 111 [2005]). We tested the hypothesis that de novo CNV is associated with autism spectrum disorders (ASD). We performed CGH on the genomic 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 ($P = 0.0005$). Such CNVs were identified in 12/118 (10%) of patients with sporadic autism, in 2/77 (2%) of patients with an affected first-degree relative, and in 2/196 (1.0%) 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 and potentially rich avenue of approach for the further study of the genetic basis of the disorder (for details, see Sebat's report in this volume).

Consonant with the finding of spontaneous CNVs in humans, studies in mice (a collaboration with Ira Hall at CSHL) have demonstrated that de novo CNVs occur frequently in mice lineages (for details, see Hall's report in this volume).

PUBLICATIONS

Hicks J., Krasnitz A., Lakshmi B., Navin N., Riggs M., Leibu E., Esposito D., Alexander J., Troge J., Grubor V., Yoon S., Wigler M., Ye K., Børresen-Dale A.-L., Naume B., Schlicting E., Norton L., Hagerstrom T., Skoog L., Auer G., Maner S., Lundin P., and Zetterberg A. 2006. Genome novel patterns of genomic rearrangement and their association with survival in breast cancer. *Genome Res.* **16:** 1465–1479.

Lakshmi B., Hall I.M., Egan C., Alexander J., Leotta A., Healy J., Zender L., Spector M., Xue W., Lowe S.W., Wigler M., and Lucito R. 2006. Mouse genomic representational oligonucleotide microarray analysis: Detection of copy number variations in normal and tumor. *Proc. Natl. Acad. Sci.* **103:** 11234–11239.

Navin N., Grubor V., Hicks J., Leibu E., Thomas E., Troge J., Riggs M., Lundin P., Maner S., Sebat J., Zetterberg A., and Wigler M. 2006. Prober: Oligonucleotide FISH probe design software. *Bioinformatics* **22:** 2437–2438.

Pelham R.J., Rodgers L., Hall I., Lucito R., Nquyen K.C.Q., Navin N., Hicks J., Mu D., Powers S., Wigler M., and Botstein M. 2006. Identification of alterations in DNA copy number in host stromal cells during tumor progression. *Proc. Natl. Acad. Sci.* **103:** 19848–19853.

Zender L., Spector M.S., Xue W., Flemming P., Cardon-Cardo C., Silke J., Fan S-T., Luk J.M., Wigler M., Hannon G.J., Mu D., Lucito R., Powers S., and Lowe S. 2006. An integrative approach to cancer gene discovery: Identification and validation of new onogenes in liver cancer. *Cell* **125:** 1253–1267.