MAMMALIAN CELL GENETICS

M. Wigler J. Alexander, M. Chi, D. Esposito, H. Grasmo-Wendler, V. Grubor, J. Healy, J. Hicks, E. Leibu, L. Muthuswamy, N. Navin, A. Reiner, M. Riggs, L. Rodgers, J. Sebat, J. Troge

This year marks a major turning point in the development of our laboratory. Traditionally, we have been a cancer research laboratory. Even during the period when the major problems we studied were embodied in model organisms—the yeasts, *Saccharomyces cerevisiae* and *Schizosaccaromyces pombe*—our effort was to understand abnormal growth regulation as a consequence of altered signal transduction. Although cancer is still one of our major efforts, we have taken on a new problem, the genetics of complex human diseases. Representational oligonucleotide microarray analysis (ROMA), the tool we developed for exploring "copy-number differences" between normal and cancer genomes, can equally well be used to explore the genetic differences among humans that may predispose individuals to medical disorders.

ROMA reveals that there are large differences in the number of genes in individual human genomes. One distinction of the complex metazoans is the size of gene families. Species evolve, at least in part, by gene duplication followed by specialization. It should not surprise us that we find this process alive and well in humans (Sebat et al. 2004). What does surprise us, given the intense scrutiny of the human genome by other methods, is that its extent was not previously appreciated. We expect to find differences in the number of copies of genes that account for disease susceptibilities, and we are determining if these differences contribute to autism.

BREAST CANCER AND LEUKEMIAS

We have now examined the abnormalities in a modest series of primary breast cancers. The bulk of our study comes from our collaborator Anders Zetterberg at the Karolinska Institutet in Stockholm, Sweden, who has framed the following question: Can we correlate the prognosis of a cancer with its genomic profile? The simple answer appears to be yes. There are more genomic lesions in the patients that succumb to their cancers than in those that survive, as well as specific loci that when unperturbed appear to correlate with survival. We seek to extend and confirm these observations in a larger series of experiments.

The same set of experiments have confirmed imbalances at well-known loci that contain oncogenes and tumor suppressor genes, but they have also revealed common imbalances at loci where there are not known candidate tumor genes. We are in the process of preparing to determine by correlation analysis specific patterns of imbalances that might suggest the pathways that are disturbed in cancer development, and the order in which these disturbances arise. Preliminary results indicate that a few very common lesions characterize the early stages of breast cancers.

In a collaboration with Nick Chiorazzi at North Shore–L.I.J. Research Institute, we have taken similar approaches to study chronic lymphocytic leukemia (CLL). There are fewer lesions present in CLL than in advanced stage breast cancer, but again, as in early stage breast cancer, there are a few common lesions. We wish to extend these studies to a larger series, to gain a foothold in predicting the outcome of the disease.

LARGE-SCALE POLYMORPHISM IN HEALTHY AND IMPAIRED HUMANS

As reported last year, we have applied our methods to the comparison of genomes from apparently healthy humans and humans with various disorders. There are a large number of extensive regions of copy-number variation between any two humans, approximately equal amounts of deletions and duplications. We have now examined on the order of several hundred individuals and find that we detect on the order of a dozen differences of 100 kb and larger between any two people. These regions contain one gene on average. We estimate that we see about half the lesions of this size or larger, and lesions that are smaller largely escape our present methods. Although copy-number polymorphisms exist throughout the chromosomes, they are conspicuously rare on the X chromosome and on the gene-rich chromosome 19.

A statistically significant difference appears to exist between autistic individuals and normal control populations: There is an increase both in the total amount of genome lost in autistics and in the size of their largest deletions. This is a property of the population as a whole, not one of individuals per se. Nevertheless, we do see genetic differences in autistics at specific loci, and in particular at loci that have been previously linked to mental disorders such as Tourrette's Syndrome and obsessive-compulsive disorders.

Considerably more effort is required to validate the significance of the differences we observe. We will soon apply similar methods to schizophrenia, congenital heart disease, and a variety of other human impairments.

TECHNOLOGY DEVELOPMENT

As before, we strive to improve both the resolution of our methods and their range of application. These efforts are largely focused on two objectives: increasing the density of probes that can be arrayed, and thus the total number of loci that can be observed, and diminishing the number of cells required for analysis. The former goal will enable us to screen smaller lesions that might upset gene function, and we are pursuing technology development in combination with NimbleGen Systems, the company that fabricates our arrays. A number of computational-algorithmic problems have needed to be solved, as well as the development of new laboratory protocols. In terms of "miniaturization," we have preliminary results indicating clear success with as few as 100 cells, and analysis of single cells is not beyond our reach. Such methods would enable us to examine the role of somatic rearrangement in aging, development, and in the evolution of the malignant phenotype.

All of our methods must transpire in a milieu of computationally based informatics processing, and so an increasingly large percentage of our effort is expended on creating mathematical tools for data analysis, and creating environments in which the large number of experiments and results of analysis can be viewed. Foremost in these efforts has been the development of reliable methods of "segmentation" (Daruwala et al. 2004; Olshen et al. 2004; Sebat et al. 2004). In these matters, we have been aided by collaborators (see authors of publications). The results of segmentation are validated by fluorescent in situ hybridization in collaboration with Dr. Anders Zetterberg's group at the Karolinska Institutet.

PUBLICATIONS

Daruwala R.-S., Rudra A., Ostrer H., Lucito R., Wigler M., and Mishra B. 2004. A versatile statistical analysis algorithm to detect genome copy number variation. *Proc. Natl. Acad. Sci.* **101:** 16292–16297.

Olshen A., Venkatraman E.S., Lucito R., and Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **4:** 557–572.

Sebat J., Muthuswamy L., Troge J., Alexander J., Young J., Lundin P., Maner S., Massa H., Walker M., Chi M., Navin N., Lucito R., Healy J., Hicks J., Ye K., Reiner A., Gilliam T.C., Trask B., Patterson N., Zetterber A., and Wigler M. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305:** 525–528.

Thomas E.E., Srebro N., Sebat J., Navin N., Healy J., Mishra B., and Wigler M. 2004. Distribution of short, paired duplications in mammalian genomes. *Proc. Natl. Acad. Sci.* **101:** 10349–10354.

In Press

Jobanputra V., Sebat J., Chung W., Anyane-Yeboa K., Wigler M., and Warburton D. 2005. Application of ROMA (representational oligonucleotide microarray analysis) to patients with known cytogenetic rearrangements. *Genet. Med.* **7:** 111–118.