GENOMIC ANALYSIS OF CANCER AND HUMAN GENETIC DISORDERS

M. Wigler J. Alexander, P. Andrews, T. Baslan, M. Bekritsky, K. Cook, H. Cox, Y. Eberling, D. Esposito, I. Hakker, J. Hicks, J. Huang, R. Kandasamy, J. Katz, J. Kendall, A. Krasnitz, Y.-H. Lee, A. Leotta, D. Levy, S. Marks, J. McIndoo, J. Meth, N. Navin, M. Oswald, M. Riggs, L. Rodgers, M. Ronemus, J. Rosenbaum, J. Troge, Z. Wang, B. Yamrom, Z. Zhu

Our group uses methods for comparative genome analysis to study cancer and human genetic disorders. New mutations, somatic and germline, underlie the cancers and human genetic disorders. Our main focus to date has been studying mutations that change the numbers of copies of sections of the genome, causing regions of deletion and duplication, in cancer and pediatric disorders. Our research in cancer is a collaborative effort to discover mutational patterns that predict clinical outcome. Additionally, we have recently developed methods for the analysis of the genomes of single cells, which has led to new insights into tumor evolution and metastasis. This work has application to the early detection of cancer and its recurrence. Our work on human genetic disorders is predicated on the hypothesis that common disorders under strong negative pressure result from new and rare variants of high penetrance. Our methodology is therefore directed to their discovery. We have made headway in the discovery of the causative mutations in autism. The same underlying theory and methodology are being applied to other disabling genetic disorders, such as congenital heart disease and pediatric cancer. All of our programs are collaborative.

Cancer

Our work in cancer is divided into three parts. First, we do outcome research (Hicks et al., *Genome Res 16:* 1465 [2006]). We use copy-number profiling to find markers that predict which breast cancers respond to which therapies. In particular, we ask which patients will respond to drugs targeting the Her-2 receptor. We are also studying ductal carcinoma in situ (DCIS) to determine if there are reliable markers that reliably predict progression to invasive disease. This work has required finding solutions to the technological problems of extracting DNA from formalin-fixed and paraffin-embedded tissue and forming collaborations with clinical oncologists (e.g., Larry Norton at Memorial Sloan-Kettering Cancer Center and Lindsay Harris at Yale).

Second, we pursue studies to elucidate the ontogeny of cancer based on the analysis of population substructure. We used subpopulation partitioning to demonstrate the clonal heterogeneity of some breast cancers (Navin et al. 2010), and then developed single-cell sequencing techniques to study this in greater depth (Navin et al. 2011). The single-cell analysis revealed the population substructure of two cancers. One breast cancer was what we call monogenomic, with one major advanced subpopulation, and another cancer was what we called polygenomic, with multiple advanced subpopulations. No intermediate cells were found, suggesting that successful clones emerge suddenly during tumor formation, probably marking a phase change in the tumor growth pattern, and lead to metastasis. (There is an alternate hypothesis to explain the absence of intermediate cells, namely, that the cancers evolve off site, diverge greatly from their progenitors, and then reinvade the primary site.) One fundamentally surprising finding was the presence of a major subpopulation of pseudodiploid cells, each with many genomic markers, but none that resembled each other or the advanced tumor subpopulations. We suspect, but have not yet proven, that these cells are indicators of the presence of an early diploid progenitor of the tumor that benefits from, and may benefit, the advanced subpopulations.

The applications of single-cell genomic sequencing are potentially profound. Critical to all applications is developing the method so that it is inexpensive, enabling the sequencing of hundreds or even thousands of single cells in parallel for the cost of sequencing one genome in depth. Fortunately, to be useful, single-cell sequencing does not need deep coverage, so this goal is theoretically feasible. When we achieve this goal, applications include detecting cancer cells in blood and other

body fluids for the purpose of monitoring early occurrence or recurrence, extracting more information from biopsies, even directing the placement of surgical boundaries. This is the third part.

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 copy-number variants (CNVs) in human disease and, in particular, in the role of spontaneous (or de novo) germline CNVs. Our findings established that germline mutation is a more significant risk factor for autism spectrum disorders (ASD) 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. We also study the role of spontaneous mutation in congenital heart disease (a collaboration with Dorothy Warburton at Columbia University), rheumatoid arthritis (with Peter Gregersen at North Shore University Hospital), and pediatric cancers (with Ken Offit at Memorial Sloan-Kettering Cancer Center, New York).

One of the de novo events we identified in autism in our 2007 paper 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 of CSHL to engineer mice with the orthologous deletion on mouse chromosome 7, and she has continued to search for phenotypic consequences. 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. Autism incidence and sibling concurrence rates are consistent with a model in which new or recent mutations with strong penetrance explain the majority of autism in males and are consistent with a one-hit event.

We are now in the midst of a larger study of spontaneous mutation in autism, based on a population of simplex families (families with only one child on the spectrum, and at least one unaffected child) collected by the Simons Foundation. Early initial results confirm our previous findings, and we observe de novo (copy-number) mutation more frequently in children with autism than in their unaffected siblings. The statistical evidence is strong for deletion events, but much weaker for amplifications, an assessment that was not possible before because of lack of statistical power. 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.

Our analysis of autism families has revealed two additional major findings. The first is statistical evidence that inheritance of rare copy-number variation also has a role in contributing to autism risk. The second is the relative absence of inherited rare deletions affecting genes compared to inherited rare duplications affecting genes. The clear implication is that most deletions are under strong negative pressure and is consonant with the observation that de novo deletions are more strongly implicated in autism than are duplications.

Our study based on copy number does not pinpoint the genes that cause autism, because even the narrow events typically contain multiple genes. Pathway analysis, performed in collaboration with Ivan Iossifov of CSHL and Dennis Vitkup of Columbia University, does suggest a plausible set of interrelated genes. We are now pursuing our leads by sequence analysis of trios (mother, father, and child) from the Simons collection (a collaboration with the McCombie lab at CSHL). We are conducting a search for recurrent de novo point mutations that disrupt function in candidate genes. From our unified hypothesis, knowledge of the rate of de novo mutation in the germline, and the rate of autism in males, we estimate that there are on the order of 300–500 autism genes. In the 1000+ trios we expect to sequence, we predict to see a signal in the form of recurrent mutations only in the actual autism genes.

PUBLICATIONS

Navin N, Krasnitz A, Rodgers R, Cook K, Meth J, Kendall J, Riggs M, Eberling Y, Troge J, Grubor V, et al. 2010. Inferring tumor progression from genomic heterogeneity. *Genome Res* **20:** 68–80.

Russnes HG, Moen Vollan HK, Lingjærde OC, Krasnitz A, Lundin P, Naume B, Sørlie T, Borgen E, Rye IH, Langerød A, et al. 2010. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Science: Translational Medicine* **2:** 38ra47.

In Press

Kamalakaran S, Vinay Varadan V, Giercksky Russnes HE, Levy D, Kendall J, Janevski A, Riggs M, Banerjee N, Synnestvedt M, Schlichting E, et al. 2011. DNA methylation patterns in luminal breast cancer differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol Oncol* **5:** 77–92.

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. 2011. Tumor evolution inferred by single cell sequencing. *Nature* **472:** 90–94.

Jean McIndoo, Ravi Kandasamy