MAMMALIAN CELL GENETICS

M. Wigler J. Alexander, J. Brodsky, A. Buglino, K. Chang, D. Dong, J. Douglas, D. Esposito, E. Hatchwell, R. Lucito, V. Mittal, A. Reiner, M. Riggs, L. Rodgers, L. Serina, J. Stolarov, J. Troge, J. West, C. Yen

The pathophysiology of a cancer is mainly determined by its tissue of origin and the mutations it has accumulated. We do not have specific knowledge of the majority of genes that cause cancers. There are gaps in our knowledge of which pathways become deranged, and in the specific components of the known pathways that go awry. Without this knowledge, it is likely that our attempts to control cancer will be incomplete. Therefore, work in our group is centered upon cancer genomics. We have developed several methods for detecting genetic alterations in cancers, identifying the genes that are the targets of these alterations, and exploring the functional analysis of these genes by a variety of means. PTEN, a tumor suppressor gene isolated by our methods, has been the subject of our most intensive study. The major change in our program since last year is the inclusion of microarray analysis for both measuring gene copy number and profiling gene expression and the development and application of retrovirus-mediated gene induction systems for studying gene function. We are also applying our genomic methods to the analysis of spontaneous human genetic disease.

Discovery of Loci and Genes Altered in Cancer

C. Yen, R. Lucito, D. Esposito

The main engine for gene discovery in our lab has been representational difference analysis (RDA), a subtractive hybridization technique that allows us to search for amplified loci and homozygously deleted regions in cancers (Lisitsyn et al., *Science 259:* 946 [1993]; Lisitsyn et al., *Proc. Natl. Acad. Sci. 92:* 151 [1995]). We have applied RDA to the study of breast cancer, analyzing a total of 18 pairs of tumor/normal pairs. This has resulted in the discovery of numerous amplified loci, many of them previously characterized, such as the loci for ErbB2, c-myc, and cyclin D, but the majority are uncharacterized. Of these, the 8q11 region has been analyzed by Scott Powers and colleagues at Tularik for further analysis. Probes from this region detect gene amplification in 6 out of 70 primary breast tumors, and they have narrowed the candidate region to two overlapping BACs (bacterial artificial chromosomes). We have also discovered with RDA nine regions of confirmed homozygous deletion. One of these was the region on 10q23 containing the PTEN tumor suppressor; one was the region containing the tumor suppressor p16(INK4) on chromosome 9p (Serrano et al., *Nature 366:* 704 [1993]; Kamb et al., *Science 264:* 436 [1994]); one was the region on chromosome 3p14 that contains the FHIT candidate tumor suppressor gene (Ohta et al., *Cell 84:* 587 [1996]); and one mapped near the p53 tumor suppressor. Five homozygously deleted regions remain that are not yet fully characterized. Two of these reside on chromosome 8q22, within 10 Mb of each other, one on chromosome 4p16, one on 21p11, and one on chromosome 20p11. All but one of these loci have been observed deleted in a number of clinical specimens and/or cell lines and have been confined to regions from a few megabases to a few hundred kilobases.

In collaboration with Dr. Hamaguchi, two main candidate transcripts have been identified from one of the 8q22 loci. We are now in the process of obtaining full-length cDNAs and performing mutational analysis on these transcripts by a combination of techniques, including linked transcription-translation premature truncation and DNA sequencing. For a more comprehensive report on this locus, see Dr. Hamaguchi's research report in this section.

Development of Inducible Gene Expression Systems

J. Stolarov

In our early studies of the PTEN tumor suppressor, we used constitutively expressing retroviral vectors and immediately encountered the problem of great variability in the phenotypes of the clones that emerged. Clonal variability, and lingering questions about the cellular adaption to the chronic expression of PTEN,

prevented us from reaching clear conclusions. Thus, we sought to develop inducible expression systems in which, ideally, paired cultures could be compared for their acute responses to controlled levels of expression of the tumor suppressor.

In the ideal system, the inducer would be physiologically inert but would rapidly induce nearly 100% of modified cells to express close to endogenous levels of the target gene. Modified cells would not express this gene in the absence of inducer. Such a system would enable the observation of acute effects of expressing tumor suppressors in cells, both when growing in cell culture and when growing as a tumor in an animal. We chose a retroviral delivery system as it had the additional attraction that the expression system could be easily introduced into a variety of cell backgrounds. We chose the ecdysone system of Evans and coworkers (No et al., *Proc. Natl. Acad. Sci. 93:* 3346 [1996]) over the tetracycline system (Saez et al., *Curr. Opin. Biotechnol. 8:* 608 [1997]) because it has no background in the uninduced state and very swift kinetics of induction. In brief, we produce both amphotropic and ecotropic viruses using high-titer packaging lines that were derived by Beach and Hannon (Hannon et al., *Science 283:* 1129 [1999]) at Cold Spring Harbor Laboratory. Typically, we infect cells with the receptor viruses together, and the cells undergo double selection with puromycin and G418. Single-cell clones are selected, expanded, and tested for their ability to serve as a good "host," i.e., respond to inducer following infection and hygromycin selection with a third virus containing an inducible β-galactosidase gene.

Our first experiments were performed with U87MG, a human glioblastoma cell line known to have no functional PTEN gene. After double infection with receptor viruses, and double selection, five candidate host clones were selected. These were then infected with a target gene (*lacZ*) under the control of ecdysone. For all five hosts, virtually all cells in the colonies were inducible, and none showed detectable background expression in the absence of inducer. We are examining the characteristics of this inducible system in other cell hosts and for responsiveness in animal hosts.

Analysis of PTEN Function

J. Stolarov, K. Chang, D. Dong

In previous years, we collaborated with Nick Tonks and Mike Myers here at the Laboratory on the study of PTEN. This work led to the observation that PTEN, predicted and shown to be a protein phosphatase, was also a PIP3,4,5 phosphatase (Myers et al., *Proc. Natl. Acad. Sci. 94:* 9052 [1997]). Since PIP3,4,5 is an important second messenger that mediates such fundamental cellular processes as response to growth factors and inhibition of cell death, this lipid phosphatase activity appeared to be a logical explanation of PTEN's function as a tumor suppressor. This was supported by studies for mutant PTEN (Myers et al., *Proc. Natl. Acad. Sci. 94:* 9052 [1997]; Myers et al., *Proc. Natl. Acad. Sci. 23:* 13513 [1998]).

We have utilized the retroviral system described above for further studies on PTEN. Expression of PTEN was clearly controllable in U87 hosts with muristirone, an ecdysone analog, with induction levels up to 50-fold above background. Cellular morphology changed radically, but not in vitro growth characteristics. Expression of several tumor markers were also affected, including thrombospondin, an inhibitor of angiogenesis, and TIMP3, an inhibitor of metalo-proteinases. Mutants of PTEN that lost either lipid phosphatase or all phosphatase activity were inert. Significantly, we observed virtually identical changes in cellular morphology with a small molecular inhibitor of the p110 PI3-kinase, the enzyme that generates PIP3,4,5, in keeping with the hypothesis that PTEN exerts its effects on cells through its lipid phosphatase activity. Additional details of the cellular responses of cells to PTEN are further discussed in the subsequent section on transcriptional profiling.

We have continued to study the properties of the PTEN homolog in the yeast *Saccharomyces cerevisiae*. In particular, we have shown that the yeast enzyme retains the biochemical activity of its mammalian homolog, and deletion of PTEN results in an impairment in sporulation.

Microarray Analysis of Expression

V. Mittal, J. Stolarov, R. Lucito, J. West, A. Reiner

We have begun implementing in our lab the technique of transcriptional profiling utilizing printed cDNA microarrays. In this technique, developed at Stanford in the laboratory of Pat Brown, arrays of cDNA probes are affixed to a glass surface, and the surface is hybridized to pairs of cDNA samples, prepared from the mRNA of two sources, each labeled with a different fluorescent dye (Schena et al., *Science 270:* 467

[1995]). The ratio of fluorescence detected at each probe gives a measure of the relative abundance of the mRNA corresponding to that probe. More than 25,000 expressed sequences, about one quarter of the estimated transcription units in the human genome, the IMAGE Consortium clones, are available for printing.

The reasons for implementing this technology in our lab are many. First, this is a potentially powerful tool for the functional characterization of new genes, especially for tumor suppressors and oncogenes, enabling their association with genes known to control certain pathways. Second, it is a powerful tool for the analysis of specific mutations, particularly for genes, such as RAS and PTEN, that encode proteins with multiple activities or multiple domains. Third, we can use this system in combination with our superb inducible expression system, and we are thus experimentally well positioned to carry out comparisons of well paired samples, allowing us to examine both up- and down-regulated genes, with time course and dosage controlled.

We printed moderate density (>6000 probes) cDNA microarrays within our own facility. We used as our paired samples U87MG cells induced or not induced to express PTEN and mutant PTEN genes, and treated or not with inhibitors of p110 PI3-kinase. The results of these experiments suggest that there are a very limited number of large changes in the transcriptional pattern when PTEN was induced. Of 6218 probes, 0.4% of transcripts appeared to be suppressed by greater than threefold upon PTEN induction, and 0.25% of transcripts appeared to be induced by greater than threefold upon PTEN induction. An overlapping pattern of change in transcription was observed when U87MG cells were treated with an inhibitor of PI3-kinase.

These results, although preliminary and requiring additional controls, strengthen our confidence that our model for the action of PTEN is at least partly correct and that our retroviral ecdysone-inducible system coupled with microarray analysis of transcription patterns is a powerful approach to the functional analysis of mammalian genes.

Detection of Gene Copy Number Fluctuations by Microarray Analysis of Genomic Representations

R. Lucito, J. West, A. Reiner

The search for tumor suppressors and oncogenes by RDA has been slow. The bottleneck has been the time and labor spent on "epicenter mapping." Once a locus has been identified, we need to demonstrate that the locus is affected in multiple cancers. Then, because these loci are large, we then need to find the minimum region affected so as to facilitate identification of candidate genes. This process, which we refer to as epicenter mapping, has required over 6000 individual quantitative polymerase chain reaction (Q-PCR) tests for the various loci we have identified. We therefore decided to develop a microarray-based method for gene copy measurements. Our expectation is that once fully established, this method will accelerate the rate of gene discovery by a factor of perhaps 1000-fold, as well as provide a detailed way of categorizing clinical cancers.

Microarrays, while widely utilized for expression profiling, cannot be effectively used for measuring gene copy numbers in a straightforward manner. Using the entire human genome for hybridization is inefficient due to its nucleotide complexity. To solve this problem, the microarray system we have developed is based on representations. A representation is a reproducible amplification of the genome with reduced complexity. In brief, a representation is produced by first cleaving the genome with a restriction enzyme, and then adaptors are ligated and used for a subsequent PCR amplification. The complexity of a representation based on *Bgl*II cleavage is approximately 3% of the genome, which is comparable to that of the complexity of transcripts within a cell. Single fragments from the representation are then cloned and arrayed on a slide. Representations from tumor and normal, prepared in the same manner as that of the array, are then hybridized to the array, and the results are analyzed to define regions that have undergone copy number changes such as deletion and amplification.

To determine the fidelity of our microarray format, we carried out several experiments. We made multiple *Bgl*II representations of the cell line SKBR-3 on different days. A representation derived from the cell line was labeled separately with Cy3 and Cy5, two fluorescent dyes in common use, and compared. We found that there was very little deviation from a ratio of one. These experiments validate the extreme reproducibility of representations and suggest that making well-controlled parallel representations introduces no more noise than is inherent in the measurements made by the system as we practice it.

We also examined the reproducibility of our microarray measurements between multiple experiments. Multiple representations of two different human breast cancer cell lines, SKBR-3 and MDA-MB-415, were compared by hybridization. The ratios obtained

from one experiment were compared to those obtained from another parallel experiment. There was excellent concordance between independent microarray measurements, and excellent agreement with measurements of gene copy number by Southern blot hybridization. These experiments again attest to the reproducibility of representations and also to the reproducibility of printing, labeling, and hybridization.

Due to the use of representations, we are able to use minute amounts of starting material. Thus, we are able to measure genomic changes in the sorted tumor and normal nuclei of a cancer biopsy. The results of the microarray hybridization completely corroborate the results obtained by other methods. This will allow us to use primary tumors that have been microdissected by some means into tumor and normal fractions.

The project is currently in a stage of scaling up. We are subcloning 30,000 *Bgl*II representational fragments and are presently characterizing them. This will produce a microarray with coverage of 1 probe every 100 kb, on average. This microarray will be used to scan the genomes of breast cancers to uncover regions that have undergone copy number changes. Our hope is that our scan will be done at such precise coverage, and against such a large number of tumors, that we will uncover new candidate loci for the location of tumor suppressors and oncogenes. Furthermore, by analyzing many tumors, we should also be able to limit the size of the region needed to look for candidates. This method also has other potential uses, including measuring mutational load in cancers, monitoring DNA methylation patterns, genome-wide genetic typing, and detection of de novo mutations in humans.

Analysis of De Novo Mutation in the Human Germ Line

E. Hatchwell, L. Serina

Approximately 1–2% of all children are born with some degree of learning difficulty. The vast majority are likely due to genetic factors. In many cases, the intellectual problems are part of a more global syndrome, including various physical abnormalities. Conventional approaches to diagnosis include biochemical and cytogenetic analysis, and, in specific circumstances, direct testing of suspect genes. When exhaustive testing has been performed, however, at least 50% of cases remain undiagnosed, either because the syndrome presentation is not specific enough or because the genetic abnormality is simply not detectable using conventional methods.

In recent years, a number of syndromes have been shown to result from submicroscopic genomic alterations (i.e., deletions, duplications, or inversions). In the vast majority, the knowledge gained depended heavily on serendipity. It is likely that many of the syndromes that are currently undiagnosed result from a similar general mechanism.

No general method yet exists for the detection of submicroscopic genomic alterations. This project aims to develop such methods, using the techniques of subtractive hybridization and microarray analysis. The main goal is to modify existing techniques (including RDA), in order to be able to scan the genome of affected children for de novo genomic alterations. We have carefully selected children with sporadic disease (the majority of those who remain undiagnosed) and assumed that their mutation is de novo. By direct comparison, using RDA and representation-based microarrays, we can compare the genome of the child with genomes of the child's parents. A de novo genomic alteration will be detected if it creates a novel junction fragment.

We have successfully conducted experiments on artificial controls and are currently using families with known de novo translocations to test out our method. If successful, we aim to test "unknowns" in the near future.

PUBLICATIONS

Dong D.L., Liu R., Sherlock R., Wigler M.H., and Nestler P.H. 1999. Molecular forceps from combinatorial libraries prevent the farnesylation of Ras by binding to its carboxyl terminus. *Chem. Biol.* **6:** 133–141.

Tu H. and Wigler M. 1999. Genetic evidence for Pak1 autoinhibition and its release by Cdc42. *Mol. Cell. Biol.* **19:** 602–611.