MAMMALIAN CELL GENETICS

M. Wigler J. Alexander, J. Brodsky, A. Buglino, K. Chang, J. Douglas, D. Esposito, E. Hatchwell, J. Healy, M. Riggs, A. Reiner, L. Rodgers, S. Rostan, L. Serina, J. Stolarov, E. Thomas, J. Troge, J. West, C. Yen

Cancer is caused by mutations in critical genes. Some of these mutations may be inherited, but much more frequently, they are acquired during the lifetime of the individual. The critical genes that are at risk are still largely unknown, both in the general sense (i.e., which gene targets ever contribute to cancer) and in the specific sense (i.e., which genes are targeted in an individual's cancer). 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, continues to be the subject of our most intensive study. We continue to work intensively on developing tools to find additional genes, and explore their function, using microarray and inducible gene expression. All aspects of our work have now been strongly influenced by the publication of a working draft of the human genome. Recently, we added a genome informatics component in order to exploit this new resource. We are also applying the methods that we develop to analyze the spontaneous mutations that underlie some human genetic diseases, especially those that cause sporadic disease in children.

Discovery of Loci and Genes Altered in Cancer

C. Yen, D. Esposito [in collaboration with R. Lucito, M. Hamaguchi, R. McCombie, and M. Zhang, Cold Spring Harbor Laboratory, and S. Powers, Tularik Corporation]

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 it 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, *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 chromosome 21p11, and one on chromosome 20p11. All but one of these loci have been observed to be 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 the laboratories of Dick McCombie and Michael Zhang, the gene candidates from these regions are being identified using informatic tools.

In collaboration with Dr. Hamaguchi, seven transcripts have been identified from one candidate locus on 8q22. Further analysis, in collaboration with the

laboratory of Mary-Claire King at the University of Seattle, has helped us to focus on one candidate gene from this region, which encodes a heretofore unrecognized gene. See Dr. Hamaguchi's section for a more comprehensive report on this locus and its genetic content.

Development of Inducible Gene Expression Systems

J. Stolarov, K. Chang [in collaboration with V. Mittal, Cold Spring Harbor Laboratory]

In our early studies of PTEN, 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 co-workers (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 David Beach and Greg 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 have now extended this system to other human and mouse host cells, and streamlined the selection of inducible hosts, and are exploring the inducibility of other genes in known signal transduction pathways.

Unfortunately, in our first attempts to test the induction system in transplantable tumors, we observed that the U87MG hosts expressed PTEN constitutively. We are currently exploring whether this is a defect in the expression system, a defect of the particular host, or an unavoidable consequence of the loosening of transcriptional controls in dying cells.

Analysis of PTEN Tumor Suppressor

J. Stolarov, K. Chang [in collaboration with V. Mittal and N. Tonks, Cold Spring Harbor Labortatory, and D. Durden, University of Indiana]

In previous years, we collaborated with Nick Tonks and Mike Myers on the study of PTEN in animal cells. 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 confirmed by studies of 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 induction system described above for further studies on PTEN. Expression of PTEN was clearly controllable in

U87MG hosts with muristirone, an ecdysone analog, with induction levels up to 50-fold above a negligible background. At physiological levels of expression, induction of PTEN did not alter cellular growth rates. This result is in contrast to most reports in the literature, in which the levels of expression of PTEN were not so well controlled. Induction did cause a noticeably flattened cellular morphology, most clearly seen when cells were grown in low serum. An essentially identical change in morphology was induced by LY294002, a known small molecular inhibitor of PI3 kinase, the enzyme that produces PIP3,4,5. Even more striking results appeared following analysis of transcription patterns using cDNA microarrays. 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 essentially identical pattern of change in transcription was observed when U87MG cells were treated with LY294002. Enzymatically inactive mutants of PTEN produced essentially no transcriptional changes when induced. These results clearly confirm the hypothesis that PTEN action results from its perturbation of the PIP3,4,5 metabolism.

In collaboration with D. Durden, we have explored the effect of PTEN expression on tumorigenicity of the host-cell U87MG. PTEN expression suppresses the tumorigenicity of this cell line, reducing tumor growth and tumor vascularization. This happens without an apparent reduction of cell replication or increase in cell death within the tumor. Our present hypothesis is that the effect must be at the tumor margins, i.e., that PTEN retards invasiveness, and this is in keeping with the observation that loss of PTEN in glioblastomas is found most commonly in invasive tumors. The physiological mechanisms at play here are not at all clear, and warrant further investigation.

Microarray Analysis of Mutations in Tumors

J. West, A. Reiner, D. Esposito [in collaboration with R. Lucito and R. McCombie, Cold Spring Harbor Laboratory]

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 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 more than 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, although 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 adaptors are then ligated and used for a subsequent PCR amplification. The complexity of a representation based on *Bgl*II cleavage is approximately 5% 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 experi-

ments. 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. A summary of this work has been published recently (Lucito et al. 2000).

The project is currently being scaled-up. Our collection of *Bgl*II representational fragments is about 12,000 and 8,000 have been sequenced. Greater than 80% of these fragments map to the "golden path," the publicly assembled human genome. Many do not map, indicating gaps in the genome data, and many map to multiple chromosomes, suggesting either miss-assembly of the genome or multiple regions of large chromosome duplications. Many of the probes contain repeat sequences, and we are presently determining which probes will be useful. To accomplish this, we are using partitions of BAC libraries, which we can use as quality control reagents for determining which probes can detect deletions and which can detect amplifications.

At present, we do not know what percentage of probes will be useful for arraying. To produce a useful tool, we will need microarrays with coverage of 1 probe every 100 kb, on average. This will require about 30,000 probes, and to achieve this, we may need to clone and characterize greater than 60,000 probes. 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 in which we need to search for genes. 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

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 analyses 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. Currently, no general method exists for detection of submicroscopic genomic alterations. This project aims to develop methods that will detect hemizygous loss, one of the known mechanisms by which sporadic mutations arise. Using the techniques of microarray analysis, one pilot study has shown the feasibility of detecting half-copy gene copy in children with William's syndrome. We are presently testing "unknowns."

Array-based Genomic Mapping

J. West, A. Reiner, L. Serina [in collaboration with R. Lucito and V. Mittal, Cold Spring Harbor Laboratory, and B. Mishra, Courant Institute, New York University]

Microarrays have potential uses in genome mapping. We have designed two major algorithms and run sim-

ulations of them to guide us in experimental design. Additionally, we have conducted experiments with BAC pools and a new form of representation, called reflective representations, with which we have demonstrated the successful assignment of arrayed probes to BACs. In principle, these tools can be applied to confirmation of the assembled map of the human genome, correction of errors in assembly, and the identification of probes and BACs from gaps in the assembled genome. The sequencing of these BACs could accelerate the completion of a correctly assembled genome.

Computational Genomics

J. Healy, A. Reiner, E. Thomas [in collaboration with R. Sachidanandam, Cold Spring Harbor Laboratory, and J. Schwartz, Courant Institute, New York University]

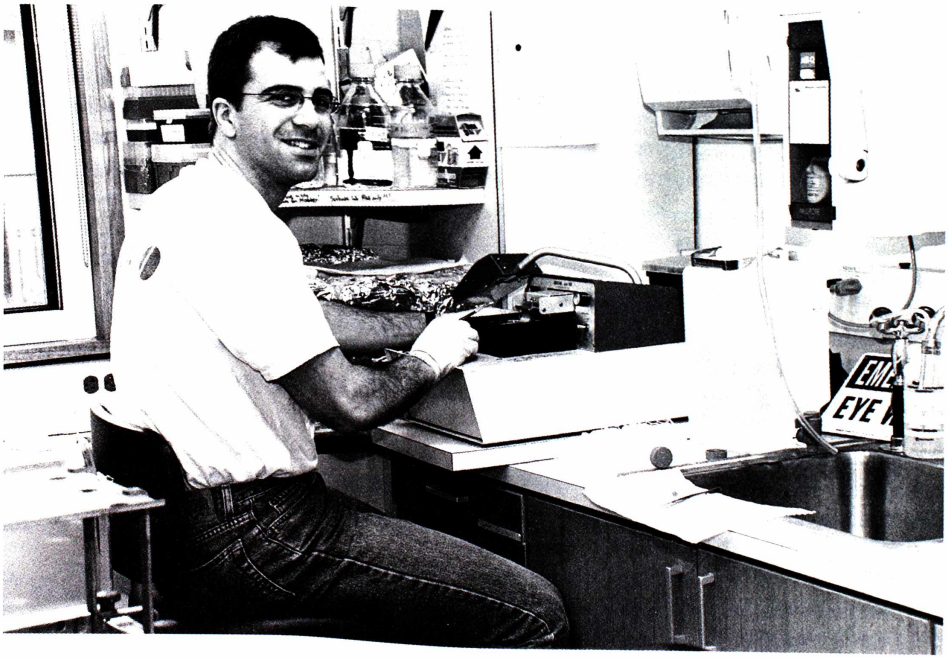
We have started a new informatics group within our lab group which is absolutely required for the tumor microarray project. However, we have begun to focus on other applications as well. One effort has been centered on probe characterization and the relation of probes to the human genome map and putative transcription units. The tools for this project have spun out other projects, including one to detect deletions/insertion alleles in the human genome, one to design a more efficient repeat masker, one to detect new repeats in the genome, and one to discover ancient segmental duplications.

PUBLICATIONS

Lucito R., West J., Reiner A., Alexander J., Esposito D., Mishra B., Powers S., Norton L., and Wigler M. 2000. Detecting gene copy number fluctuations in tumor cells by microarray analyis of genomic representations. *Genome Res.* **10:** 1726–1736.

In Press

Wen S., Stolarov J., Myers M.P., Su J.D., Wigler M.H., Tonks N.K., and Durden D.L. 2001. PTEN controls tumor-induced angiogenesis. *Proc. Natl. Acad. Sci.* (in press).



J. Trachtenberg