MAMMALIAN CELL GENETICS

M. Wigler J. Alexander, A. Brady, K. Chang, M. Chi, J. Douglas, D. Esposito, S. Gass, H. Grasmo-Wendler, J. Healy, R. Lucito, A. Reiner, M. Riggs, L. Rodgers, S. Rostan, E. Thomas, J. Troge, J. West

It is a poor and unnecessary gamble to act as though either our theory or our knowledge of cancer is nearly complete. Future progress in detection, prognosis, and treatment of cancer will depend on the accuracy and completeness of our understanding of its specific molecular causes. This knowledge is likely to become increasingly important as cancers, or suspected cancers, are detected at earlier and earlier stages.

There are simple tests for the completeness of our understanding of how cancers survive in and kill their hosts. If our knowledge were complete, we would see a plateau in the number of genes commonly found mutated in cancers. If the principles were few, even advanced cancers with a large number of accumulated genetic lesions would show only a small number of commonly affected pathways. It follows from this that if mutation in a single gene were sufficient to affect a given pathway, then even advanced cancers would show only a small number of commonly affected genes, the remainder of lesions being more-or-less random.

We have developed a microarray-based method that can partially address these issues. We can readily identify loci in the genome that undergo amplification, deletion, and imbalanced breaks. Although there are many other possible mechanisms that alter critical genes, such as point mutations, balanced translocations, and possibly stable epigenetic changes, many if not most oncogenes and tumor suppressor genes will eventually be found in the types of lesions that we can readily detect. Moreover, if a region is commonly found altered in cancers, that region harbors a good candidate cancer gene. Therefore, the application of our method to a large series of cancers, and the comprehensive comparative analysis of such data, should reveal the position and number of candidate cancer genes in cancers.

Using our methodology, we have also discovered, as we predicted last year, that there are many examples of copy number polymorphisms in the human gene pool, i.e., large regions of the human genome that are present in individuals in unequal amounts. It remains to be shown that these regions of copy number variation are inherited as such, but it seems plausible to us that some of these regions will be shown to be associated with disease resistance and sensitivity. We have made great strides in developing our technology. We are now at the beginning stages of accumulating data. We also report on our progress at cancer gene identification, genome mapping techniques, and other genomic studies.

THE TECHNOLOGY

The major obstacle to the implementation of our technology has been financial. The principles and pilot demonstrations were established in previous years. Fortunately, due mainly to the contributions of many extraordinary philanthropic groups and individuals, and through additional support from corporations and the National Cancer Institute, we have been able realize our plans on a large scale only this year. But financial considerations aside, this work became possible because of the publication of the assembled human sequence.

The basis of our technology, representational oligonucleotide microarrays, or ROMA, has been explained over the past years. It involves making complexity-reducing representations of genomic DNA and hybridizing such to microarrays of oligonucleotide probes designed informatically from the published human genome assembly. The probes are chosen from the genome so that they are complementary to the parts of the genome that are in the representations and are further selected so that they have a minimal overlap with unrelated regions of the genomes.

The algorithms that minimize overlap were described briefly in last year's Annual Report, and were developed by John Healy in our group. They have now been submitted for publication. The algorithms allow counts of exact matches of sequences of any length throughout a sequenced genome and are based on a Burrows-Wheeler transform of the genome sequence and the construction of auxiliary data structures. The unintended uses for these algorithms are described below.

We use two forms of oligonucleotide microarrays, the printed form that we make ourselves, and a form in which oligonucleotides are synthesized in situ on the array surface using laser-directed photochemistry. A

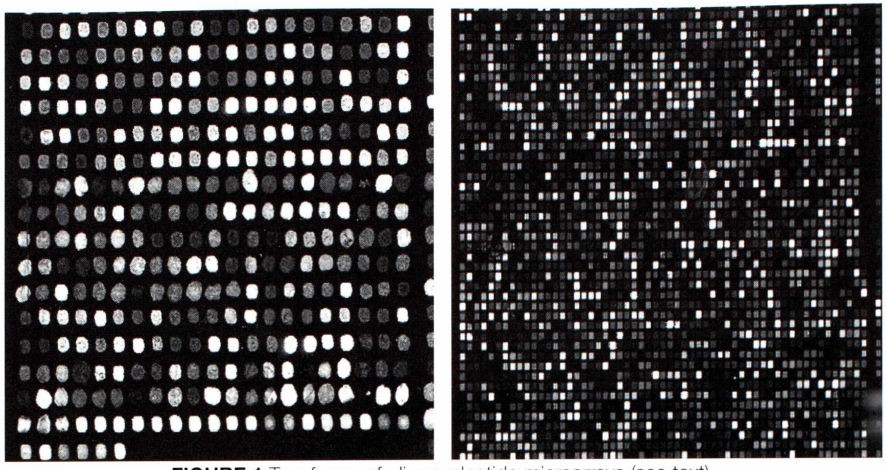

FIGURE 1 Two forms of oligonucleotide microarrays (see text).

company called Nimblegen makes the latter, and their technology has given us substantially greater flexibility in the design of arrays and the selection of representations. Pictures of these microarrays are shown in Figure 1, with the printed array on the left and the Nimblegen array on the right. With Nimblegen, we can array 85,000 probes (85K format), average one probe per 30 kb, and even greater densities could be achieved readily. We have shown that each format yields very similar measures, probe for probe. This work was conducted in collaboration with Robert Lucito here at CSHL.

CANCER LESIONS

We have applied our method to both tumor biopsies and cancer cell lines and have observed gross chromosomal copy number alterations, and highly localized amplification, imbalanced chromosome breaks, and deletions. In the latter case, we expect that we have observed both hemizygous and homozygous deletions. Representative examples of these lesions are illustrated in Figure 2. Clockwise from upper left, we have an imbalanced chromosome break, a complex series of amplifications, a homozygous deletion, and a presumed hemizygous deletion. Ratios of probe intensity are roughly proportional to copy number alteration and are plotted for each probe in the order that they occur on the genome. Averages of probe ratios for segments with similar behavior are indicated. In our data analysis, we have used algorithms for statistical segmentation designed by Adam B. Olshen and E.S. Venkatraman of the Memorial Sloan-Kettering Cancer Center. We observe a large number of lesions, of varying sizes, per cancer. We do not yet have enough examples of genomes analyzed at 85K resolution to allow us to make generalizations, but in the two genomes that we have analyzed in great detail, we observe on the order of 80 lesions each, a number greatly exceeding our expectation.

In collaboration with Rob Lucito here at CSHL, and L. Norton and W. Gerald of Memorial Sloan-Kettering Cancer Center (MSKCC), we have shown that detection of amplification at the ErbB-2 locus in breast cancer by ROMA is roughly comparable to its detection by fluorescence in situ hybridization (FISH). We therefore have initiated a far-ranging collaboration with Anders Zetterberg at the Karolinska Institute of Stockholm on a program to convert the findings of ROMA into FISH assays. Dr. Zetterberg is one of the world's foremost experts on FISH analysis of cancer.

Our future plans include a survey of a large number of breast, ovarian, and pancreatic tumors and cell lines, and leukemias and lymphomas, after which we will determine by epicenter mapping the minimal regions of frequently recurring amplifications and deletions and breakpoints. Candidate genes from these regions will then be further explored by functional analysis and sequencing. To complete this program, we need to automate data processing and gene annotation, so that our efforts for the next few years will focus on computational methods. In these efforts, we are currently collaborating with clinicians and researchers at the MSKCC, the Karolinska Institute, University of Toronto, Stony Brook University, Stony Brook, New York University, LI–North Shore University Hospital, and Johns Hopkins University School of Medicine.

NORMAL VARIATION

We have applied our method to the comparison of normal genomes and have discovered that there are a large

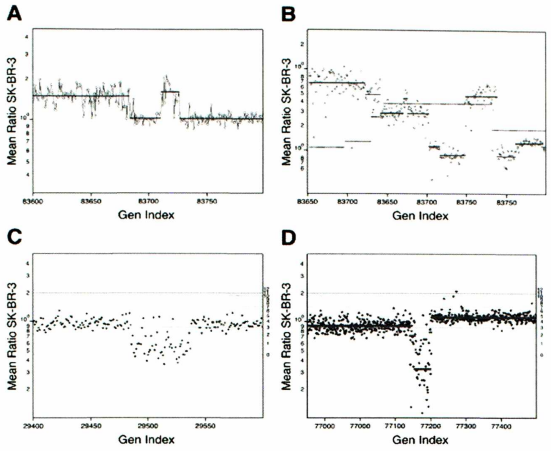

FIGURE 2 Representative examples of cancer lesions (see text).

number of extensive regions of copy number variation between any two genomes. These regions could result from polymorphic gene duplications or deletions, and in further collaboration with Dr. Zetterberg, we are in the process of determining precisely this question. These regions must be categorized, otherwise they will be mistaken for recurring cancer lesions in our cancer surveys. Furthermore, these normal variations may be associated with inherited disease susceptibility or resistance. In a collaboration with Conrad Gilliam of Columbia University College of Physicians & Surgeons, we are planning to carry out a large survey of normal genomes and genomes from families with children with mental or psychiatric disorders.

CANCER GENES

We have continued our work on the relationship of PTEN expression and TGF-β, a growth factor. We have demonstrated that the expression of PTEN squelches the effects of TGF-β on gene expression and are trying to determine the mechanism of action. Work with collaborator Masaaki Hamaguchi here at CSHL has led to the discovery of the *DBC2* tumor suppressor (Hamaguchi et al. 2002). *DBC2* expression is lost in the majority of breast cancer cell lines, and in lung cancers as well. Its function is unknown, but its forced expression arrests the growth of cancer cell lines that have lost expression of the endogenous gene. For more details, see Hamaguchi's annual report. Work with the collaborator Scott Powers at Tularik has led to the discovery of a new class of oncogene, *KCNK9* (Mu et al. 2003). *KCNB* (a.k.a. *KCNK9*) encodes a potassium channel, and work from the Tularik group suggests that it is involved in resistance to death by oxygen limitation. *PTEN, DBC2*, and *KCNB* were all discovered through the use of representational difference analysis (RDA), a method that was developed by us in past years (Lisitsyn et al., *Nat. Genet. 6:* 57 [1994]; *Proc. Natl. Acad. Sci. 92:* 151 [1995]) and is now supplanted by ROMA.

CELL TALK

We continue to examine the question of whether we can use microarray expression analysis and other means to detect "conversations" between cancer cells and cancer cells and hosts. Our rationale is that cancer-host interaction remains a large unexplored chapter of cancer biology. We initiated with pairs of sarcoma-carcinomas and found evidence that the transcriptional state in one is affected by the presence of the other. We are now exploring the effect that these cells have on each other in co-tumorigenesis experiments. In collaboration with Jan Kitajewski at Columbia University College of Physicians & Surgeons, we are testing whether microarrays can be used to detect the interaction between the notch receptor and its ligands. These interactions depend on direct cell-cell contact.

GENOME MAPPING, ASSEMBLY AND EVOLUTION

In collaboration with Bud Mishra here at CSHL and the Courant Institute for Applied Mathematics at New York University, we have developed uses of microarray hybridizations for the mapping of probes in genomes. A graduate student, Joseph West, has recently collected a full set of data in a model organism, *Schizosaccharomyces pombe*, which will enable us to test these ideas. John Healy in our group and a Watson School graduate student, Elizabeth Thomas, have found that our exact matching algorithms have had some unexpected uses beyond probe design, namely, monitoring inconsistencies between human genome assemblies and the discovery of new genomic repeats.

PUBLICATIONS

Hamaguchi M., Meth J.L., von Klitzing C., Wei W., Esposito D., Rodgers L., Walsh T., Welch P., King M.-C., and Wigler M.H. 2002. *DBC2*, a candidate for a tumor suppressor gene involved in breast cancer. *Proc. Natl. Acad. Sci.* **99:** 13647–13652.

Wigler M. and Mishra B. 2002. Wild by Nature. *Science* **296:** 1408–1408.

In Press

Mu D., Chen L., Zhang X., See L.-H., Koch C.M., Yen C., Tong J.J., Spiegel L., Nguyen K.C.Q., Servoss A., Peng Y., Pei L., Marks J.R., Lowe S., Hoey T., Jan L.Y., McCombie W.R., Wigler M.H., and Powers S. 2003. Genomic amplification and oncogenic properties of the *KCNK9* potassium channel gene. *Cancer Cell* **3:** (in press).