TECH SIGHT DOT

Biochips for Gene Spotting

K. K. Jain

Biochips are currently one of the key technologies in biology. To develop them, scientists and engineers borrowed miniaturization, integration, and parallel-processing techniques from the computer industry to develop laboratory devices and procedures that fit on a wafer. Generally, the arrays that fit these microchips consist of orderly arrangements of samples such as cDNAs, oligonucleotides, or proteins. Macroarraying, which is also called gridding, is the process of organizing sample colonies on large nylon filters to ready them for screening by hybridization. In microarrays, however, the sample spot sizes usually are less than 200 micrometers in diameter and require microscopic analysis. Packing so much information onto a small space gives microarrays a clear advantage. A single DNA microarray plate can be a few centimeters square and can contain many thousands of samples, each representing a part of a single gene. It is possible, therefore, to design single chips containing the complete gene set of a complex organism, about 30,000 to 60,000 genes.

As the technology has developed, so too has the language used to describe it. At least 23 different terms are used to describe microarrays (*1*). Most commonly the microarray plates are known as "DNA chips." Although the term "gene chip" is sometimes used, GeneChip is a proprietary microarray for gene expression studies made by Affymetrix (see figure) that can pack up to 400,000 different oligonucleotides on a single array, or 40 sequences for each of 10,000 genes.

DNA microarray systems are versatile tools for mutational analysis, gene sequencing, and the study of gene expression. These systems combine DNA chips with the instruments to handle the samples, scanners to read the reporter molecules, and bioinformatics tools to analyze the data. The hybridization of nucleic acid–derived samples to the immobilized oligonucleotides in microarrays allows one to easily quantitate the expression of specific mRNAs on a genomic scale or to detect the occurrence of polymorphisms in genomic DNA.

Chip technology can be used to analyze proteins as well. A product like ProteinChip (Ciphergen Biosystems Inc., Fremont, California) can array proteins on a surface treated chemically or biologically with enzymes, receptor proteins, or antibodies. Unknown molecules can then be analyzed by their affinity with the surface of the chip, detached from this surface with the use of focused laser energy, and detected according to molecular weight. Such chips enable immunoassays, the study of protein-protein interactions, or the measurement of ligand binding. This article focuses on the uses of DNA chips; the various types of protein chips are described in more detail in a separate report on proteomics (*2*).

The author is at Blaesiring 7, CH-4057, Basel, Switzerland. E-mail: jain@pharmabiotech.ch

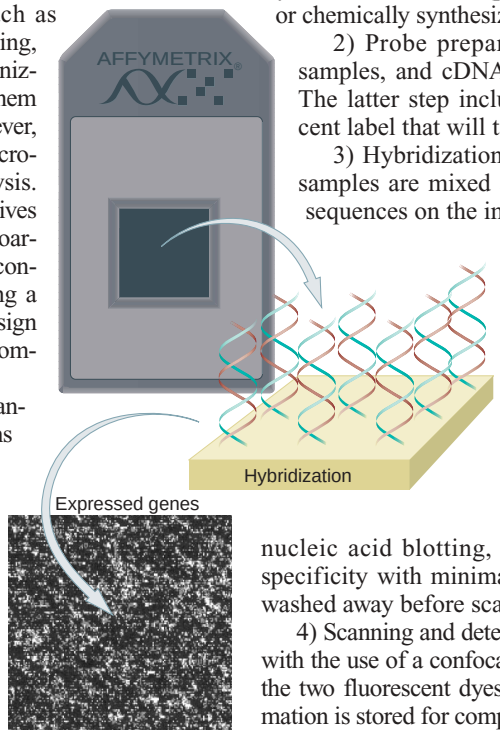
DNA microarrays provide a powerful method of simultaneously analyzing the expression of thousands of genes. Most commonly, microarray analysis uses immobilized sequences on the chip that are hybridized to fluorescently labeled cDNAs prepared from cellular RNA. Here, the term "probe" refers to the free, fluorescently labeled DNAs and "target" refers to the immobilized sequences. These definitions are sometimes reversed in other discussions, so the reader should use caution.

Microarray analysis of gene expression includes several steps as follows:

1) Array construction. The DNAs that are spotted onto the chip may consist of cDNAs (partial or complete), genomic DNAs, or chemically synthesized oligonucleotide sequences.

2) Probe preparation. RNA is extracted from the samples, and cDNA is copied by reverse transcriptase. The latter step includes the incorporation of a fluorescent label that will tag probes from each sample.

3) Hybridization of probe to array. Probes from two samples are mixed and allowed to find complementary sequences on the immobilized array. As with traditional nucleic acid blotting, conditions are optimized for high specificity with minimal background. Unbound probes are washed away before scanning.



Wondrous chip. GeneChip Technology (top), showing hybridization of single-stranded fluorescently labeled DNA target with oligonucleotides probe (middle). A magnified view (bottom) of expression shows thousands of genes that can be quantified. (Adapted from figure with permission of Affymetrix Inc., Santa Clara, California.)

4) Scanning and detection. The hybridized array is scanned with the use of a confocal laser scanner that can detect each of the two fluorescent dyes used to label the probes. This information is stored for computer analysis and image construction.

5) Normalization and analysis of data. The images produced during scanning for each fluorescent dye are aligned by specialized software so that the spots overlap. The number of spots and their individual intensity are also quantified, and background intensity is determined and subtracted. Controls, such as externally added sequences, reporter genes, or total fluorescence for each sample, help correct for differences in labeling and detection efficiency of the two fluorescent tags. Commonly, individual gene signals are scaled using the ratio of control signals in the two samples.

DNA array–based technologies provide relatively simple ways to measure differential gene expression, i.e., the relative levels of RNA transcripts in different cell or tissue samples, for all of the genes of an organism simultaneously. When the levels of specific RNAs from two different sources, such as control and diseased tissue, are measured, their differences can be easily represented if distinctly colored fluorescent labels—blue and yellow, for example—are used to make each sample's cDNA probe. If expression levels of a gene are the same in the two tissues, then a green color will be displayed by the computer. Changes in coloring that favor either blue or yellow show differences in gene expression. Parallel quantitation of large numbers of mRNA transcripts with the use of microarray technology

promises to provide detailed insight into cellular processes involved in the regulation of gene expression (*3*). DNA microarray information permits researchers to study changes in host-cell gene expression in disease states arising from viral infections or from cell transformation that leads to tumor formation. More complete understanding of these changes should lead to knowledge of mechanisms of virus replication and pathogenesis, as well as host antiviral responses.

Microarray analysis yields enormous datasets. For example, an array experiment with six samples involving 50,000 genes and 10 different experimental conditions will produce 3,000,000 pieces of primary information. Cross comparisons of sample images can multiply this total many times over. Such large collections of data necessitate large-scale information storage, analysis, and management. Systems for mining and warehousing of such biodata have entered the bioinformatics market recently. Software packages help researchers simplify the process of making sense of gene expression data from microarrays by assembling primary image data and quantitating and converting it into expression information for further analysis and visualization.

Analysis of patterns of gene expression has contributed to researchers' ability to determine the mechanism(s) of drug action. For example, cDNA microarrays showed that the action of ferrioxamine, an iron-chelating agent in the yeast *Saccharomyces cerevisiae*, largely regulates the uptake of iron by the transcription factor Aft1 because its genes are activated upon iron treatment (*4*). The results helped establish that *S. cerevisiae* has two sets of AFT-1–regulated genes that maintain different pathways for ferrioxamine-mediated iron uptake. Fet3p, a component of the high-affinity ferrous iron transport system, was expressed in the plasma membrane, whereas Arn3p, the siderophore-mediated transport system, was expressed in intracellular vesicles.

High-throughput gene expression analysis plays an important role in the drug discovery process (*5*). With the use of microarrays, genome-wide expression monitoring will enable the characterization of patterns of gene expression in diseased cells and the identification of promising new drug targets (*6*). Microarrays can also facilitate drug target validation and identification of secondary drug target effects by providing profiles to help identify drugs that have side effects or that interact with other drugs.

Molecular diagnostics are yet another process that can be augmented with the use of microarrays. With the use of microbial genomic sequences as specific genetic markers for individual pathogens, the markers are arrayed on a glass support to create a DNA chip. Target DNA is then extracted from the clinical test sample, labeled with a fluorescent dye, and hybridized to the genomic sequences on the chip. Complementary sequences form duplexes with the probes on the chip, generating fluorescent signals at the site of the probe. Analysis of the pattern of fluorescence enables identification of the target sequences and, thus, of the microbes present in the specimen. Microarray technology has been used to monitor gene expression in cancerous cells and to detect mutations in the breast cancer gene BRCA1. A tissue microarray approach has been developed where samples from up to 1000 tumors can be analyzed simultaneously on one glass slide. This method has been used to rapidly identify molecular alterations associated with bladder cancer (*7*).

Completion of the human genome nucleotide sequence and availability of technologies for assessing sequence variations are likely to prompt studies of comparative genomic diversity in human populations across the globe (*8*). Uses of microarrays for gene expression profiling, genotyping, mutation detection, and gene discovery are al-

ready yielding insights into the function of thousands of genes previously known only by their sequence. Microarray-based comparative genomics is also a promising approach to exploring the molecular epidemiology of microbial infections. In one study, a *Mycobacterium tuberculosis* high-density oligonucleotide microarray was used to detect small-scale genomic deletions in well-characterized isolates of the microorganism, and the accumulation of certain mutations was found to be associated with a tendency toward diminished pathogenicity (*9*).

Small stretches of DNA that differ by only one base are called single nucleotide polymorphisms (SNPs) and can distinguish one individual's genetic material from that of another, and microarrays can screen for these. One protocol involves the binding of SNP-containing oligonucleotide targets to solid surfaces, such that many thousands of SNPs may be fixed to a single "chip," each at a separate, defined position. To determine which SNPs are present in an individual's genome, DNA from that individual is hybridized to the chip. The genome probe is labeled such that if a particular SNP is present, then the chip position representing that SNP will exhibit a positive signal. If the SNP is absent, no signal is produced. A significant disadvantage to this approach is that a separate microarray or chip must be used to genotype every single individual, an application that would be prohibitively expensive on a large scale.

With the OmniScan device (PolyGenyx Inc., Worcester, Massachusetts), the genomes of multiple individuals, rather than SNPs, are immobilized on a chip. Thus, SNP detection involves hybridizing SNPs instead of genomes to the solid surface, allowing an investigator to array genomic DNA from over 10,000 individuals on a single surface and analyze them simultaneously. This parallel processing approach is substantially more efficient than serial processing for large population studies.

SNP genotyping and the emerging field of pharmacogenomics may eventually lead to mass genetic screening and "personalized medicine," the prescription of specific treatments and therapeutics best suited for an individual's genotype (*10*). The genetic information of an individual, recorded on a biochip, could be used for determining that person's susceptibility to disease as well as the best way to treat it. For example, in the not-too-distant future, microchips could be incorporated into hand-held diagnostic instruments for use at the patient's bedside or in the doctor's office. Automated laboratory instruments could scan the genome of a patient and, after computer analysis, predict a patient's response to a specific drug. Thus, it may be possible to determine within minutes the ideal course of therapy for each patient. But before such technology could be used, the privacy and ethical issues that accompany the screening of genetic profiles for disease would have to be resolved.

A new age of medicine lies before us. And microarray technologies, with their screening and monitoring capabilities, will help realize integrated healthcare and make individualized therapy possible.

References and Notes

1. K. K. Jain, *Biochips & Microarrays: technology and commercial potential* (Informa Global Pharmaceutical Publications, London, 2000), p. 5.
2. K. K. Jain, *Proteomics: technologies and commercial opportunities* (Jain PharmaBiotech Publications, Basel, Switzerland, ed. 7, 2001), pp. 35–43.
3. A. Schulze, J. Downward, *Nature Cell Biol.* **3**, E190 (2001).
4. C. W. Yun *et al.*, *J. Biol. Chem.* **275**, 10709 (2000).
5. G. G. Lennon, *Drug Discovery Today* **5**, 59 (2000).
6. S. Braxton, T. Bedilion, *Curr. Opin. Biotechnol.* **9**, 643 (1998).
7. A. Nocito *et al.*, *J. Pathol.* **194**, 349 (2001).
8. A. Chakravarti, *Nature Genet.* **21** (suppl.), 56 (1999).
9. M. Kato-Maeda *et al.*, *Genome Res.* **11**, 547 (2001).
10. K. K. Jain, *Personalised Medicine*, (Informa Global Pharmaceutical Publications, London, 2001).