Microarray technology in obstetrics and gynecology: A guide for clinicians

Kenneth Ward, MD*

Department of Obstetrics and Gynecology and Women's Health and the Pacific Research Center for Early Human Development, University of Hawaii, John A. Burns School of Medicine, Honolulu, HI

Microarrays can be constructed with dozens to millions of probes on their surface to allow high-throughput analyses of many biologic processes to be performed simultaneously on the same sample. Microarrays are now widely used for gene expression analysis, deoxyribonucleic acid re-sequencing, single-nucleotide polymorphism genotyping, and comparative genomic hybridization. Microarray technology is accelerating research in many fields and now microarrays are moving into clinical application. This review discusses the emerging role of microarrays in molecular diagnostics, pathogen detection, oncology, and pharmacogenomics.
© 2006 Mosby, Inc. All rights reserved.

Medical science continues to benefit from the technological revolutions occurring in nanotechnology, informatics, molecular biology, and many other disciplines. Since the mid-1970s, recombinant DNA methods, automated deoxyribonucleic acid (DNA) sequencing, the polymerase chain reaction, and many other breakthroughs have migrated onto the pages of this journal and into the hospitals and clinics at which we practice. The current issue of the *Journal* has several papers using or considering the role of a powerful new approach: microarray analysis. The purpose of this review is to serve as a primer for clinicians and investigators who are unfamiliar with this tool. This review will focus on DNA microarrays; common research and emerging clinical applications will be highlighted.

What are microarrays?

Microarray chips have dozens to millions of molecules (oligonucleotides, cloned DNA, antibodies, peptides, etc) arrayed on a surface; these attached molecules are used to probe a variety of biological phenomena simultaneously in a test sample. Manufacturers can reliably place picogram amounts of probe at each location and to place these probes just a few micrometers apart. Depending on the planned use for the microarray and the manufacturer, these molecular probes are attached to plastic, glass, nylon, or even silicon wafers. Each individual probe is placed at a precisely defined location on the array support, which is usually a flat, 2-dimensional surface. The identity of the molecule fixed to each spot

| **Table** Types of microarrays |
| --- |
| There are literally dozens of different designs and hundreds of different names that have been applied to microarrays. Gene expression, resequencing, CGH, copy number, and SNP arrays are discussed extensively in this review. This table briefly describes some of the other arrays being designed and used for high-throughput applications. |
| Antibody microarrays: Various antibodies with known ligands are arrayed and used to study gene regulation at the protein level.[66] |
| Bead arrays: These arrays are small beads with capture probes and fluorescent dyes attached. Bead arrays have faster hybridization kinetics and more flexibility in assay design than 2-dimensional arrays. |
| Carbohydrate microarrays (monosaccharide, oligosaccharide, polysaccharide, glycoconjugate, and glycoprotein microarrays): These arrays rapidly screen protein binding to carbohydrates. Most cell surface proteins are glycoproteins, and the carbohydrate attachments are critical to their function.[67,68] |
| Cell arrays: Living cells are placed at defined locations on chips and then tested for a variety of reactions to applied agents |
| Chemical microarrays: Chemical libraries of potential drugs are bonded to the array and protein affinities to these molecules are tested.[69] |
| GPCR microarrays: GPCR is an abbreviation for G protein-coupled receptors. These receptors are such an important drug target class that arrays have been developed to screen multiple GPCRs simultaneously.[70] |
| Electronic microarrays: These arrays consist of electrical circuits that can respond to presence biologic molecules. |
| Microfluidics chips: These chips are usually not manufactured as arrays. These chips contain tiny channels that control the movement of reagents over the chip, allowing various molecular assays to be performed entirely at a microscopic scale on the chip. The chips can save money when the reagents involved are particularly expensive.[71] |
| Protein microarrays: These chips are designed to measure changes in protein expression, protein-protein interactions, and the proteomic response to response to drugs and other stimuli.[65,72-74] |
| RNAi arrays: RNA interference (RNAi) is a method used to reduce the expression of specific target genes in cultured cells by delivering an RNA-blocking molecule. RNAi arrays use cell arrays to screen multiple cells against each RNA blocker in a high-throughput manner.[75-77] |
| Tiling arrays: Tiling arrays are used to refine the precise exon structures of genes in a genomic region of interest. Tiling arrays are built on the basis of nonrepetitive genomic DNA sequence. These chips allow unbiased interrogation of the genome because no assumptions are made about gene identities from the DNA sequence. Tiling arrays enable the discovery of regulatory elements and novel genes that were not predicted by current gene prediction software.[10,78-80] |
| Tissue microarrays: Tissue microarrays allow the simultaneous analysis of multiple samples of a tissue or cell line arranged in an array format to allow high-throughput molecular profiling of the tissue.[81,82] |
| Viral gene chips: These created chips can rapidly identify known viruses and classify new ones based on their genetic makeup. Viral gene chips recently identified that the cause for severe acute respiratory syndrome (SARS) was a novel coronavirus.[83-85] |

for any particular array design never changes. The microscopic scale of the array keeps assay costs lower and allows high-throughput "parallel" testing of samples as small as a single cell.

Regardless of the array design, hybridization, the ability of 2 complementary molecules to lock together, is the central design element for microarray assays. For instance, DNA microarrays depend on Watson-Crick base pairing. Single-stranded DNA probes will hybridize, or stick, to the strands of DNA sample to be tested following the usual rules of base pairing (A to T, C to G). Complementary DNA sequences have incredibly high affinity for each other and thus the target DNA in a solution literally find and attach itself to the immobilized probe DNA. Probes as short as 20 nucleotides in length can be highly specific, whereas even a single mismatched base will greatly reduce the strength and likelihood of hybridization. Longer probes usually have greater sensitivity for the target molecule; numerous other modifications have been reported to change other probe characteristics. Usually probes are prepared by chemical synthesis or using the polymerase chain reaction. Other types of microarrays, such as arrays that use antibodies to probe for antigens or proteins to probe for protein interactions (Table), are more difficult to build,

but they also depend on the chemical and physical forces attracting complementary molecules.

DNA microarrays are basically reverse Southern blots. With Southern blots, DNA to be tested is arranged (by size) on a nylon membrane and a probe in solution is labeled with a detection molecule and applied to the membrane to generate a signal. With microarrays, the probe is fixed onto a surface and the nucleic acid to be tested is in a solution, which is applied to the array. Most microarrays use fluorescent tags as the means of identifying whether hybridization has occurred (whether the target molecule is stuck to the probe molecule on the array). Array scanners can rapidly detect very low levels of fluorescence and map the signal to its source on the array with great certainty. Usually the fluorescent tags are excited by a laser and the signal captured by a high-resolution digital camera. Most protocols improve sensitivity of detection by chemically attaching more than 1 copy of fluorescent tag per target molecule detected.

The upfront cost of microarray instrumentation is high, and most chips cost several hundred dollars each and they can be used only once. Fortunately, as manufacturers increase their sales and as competing products emerge, array prices are coming down. Costs
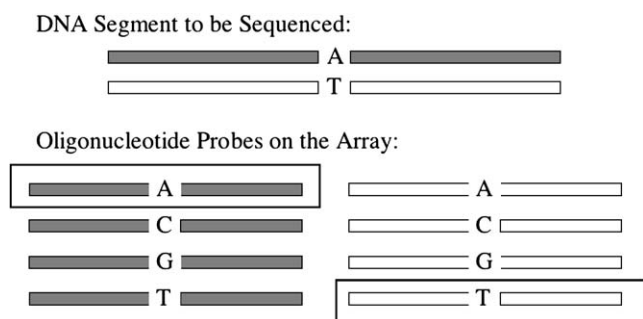
Figure 1   DNA resequencing using oligonucleotide probes. A large number of probes can be synthesized to interrogate any particular base pair of a known sequence. In this schematic, 8 oligonucleotide probes are depicted: 4 are complementary to the sense strand and 4 complement the antisense strand. Each of the 4 probes in either set has a different base at the critical position. Only 2 of these 8 probes (the 2 probes outlined by a *rectangle*) will hybridize to the DNA sequence being tested. This particular hybridization pattern occurs only when the unknown sequence has the sequence represented in the diagram.

considerations drive the development of statistical methods to rest as much valid data as possible from a small number of microarrays. Nonetheless, on a per-test basis, microarrays offer relatively inexpensive, rapid, simple testing, compared with other molecular methods. Certainly the increasing numbers, varieties, and subtleties of available arrays are accelerating the pace of biomedical discoveries. In my own laboratory, we can now generate genotype data in 1 day that 15 years ago it might have taken a year to produce. Our focus has shifted from producing data to the task of analyzing the voluminous data produced.

Whereas various DNA arraying methods (Southern blots and dot-blots) had been in use more than 2 decades earlier,[1] Fodor et al[2] launched the microarray era in 1991. They borrowed techniques from computer chip manufacturing (ie, photolithography) to allow parallel synthesis of a large number of oligonucleotides probes directly on silicon wafers. The company Fodor subsequently formed (Affymetrix) remains an innovator and is also the largest supplier of DNA microarrays. The variety of commercially available techniques, devices, and instrument systems keeps expanding. Homemade microarrays can be produced using inexpensive spotting devices, which work in a manner similar to ink-jet printers, allow researchers to develop their own probe sets that are then sprayed onto a precise position on a blank array.[3]

Research applications

DNA resequencing chips

DNA microarrays can be used to rapidly and accurately sequence known genes.[4,5] For any specific base pair in
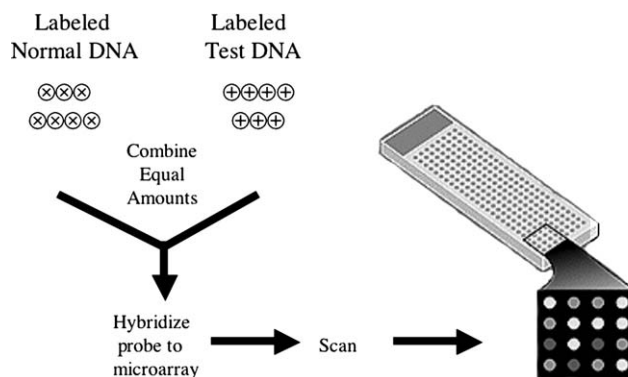
Figure 2   Comparative genomic hybridization on a microarray. The florescence ratios read off the array reflect whether chromosomal regions are deleted or duplicated in the test DNA sample.

the human genome, an oligonucleotide probe can be constructed that matches the normal sequence on 1 strand perfectly. Three other oligonucleotides can be constructed, substituting an alternative base matching each of the 3 other possible nucleotides at that position. Similarly 4 probes can be constructed for either the normal base or the 3 substituted bases on the complementary strand (Figure 1).

Because these oligonucleotides can be designed with the base pair being probed at the 3 prime ends or the 5 prime ends or anywhere in between, a very large series of oligonucleotide probes can be designed. For instance, if each of the oligonucleotide probes is 25 nucleotides long, then it is possible to make 200 different oligonucleotides all assaying the same base pair in the genome (4 probes × 2 strands × 25 positions in the oligonucleotide). These 200 probes will have different hybridization affinities, but the signature of any base change should always be similar. Any missense mutation would affect hybridization to all 200 probes! It is also easy to predict how various microdeletions would alter the hybridization pattern. The redundancy this affords allows probes to be spread over different geographic areas of the microarray chip, thus lessening the chance that air bubbles or extraneous debris would interfere with the hybridization.

Comparative genomic hybridization/DNA copy number

In comparative genomic hybridization (CGH), DNA to be tested is labeled with one fluorescent dye, and it is added in equal amounts directly to DNA from normal reference DNA labeled with a second dye (Figure 2.). This mixed sample is then hybridized to DNA microarrays with probes from every region of the genome.[6] Relatively large DNA probes cloned in bacteria (such as bacterial artificial chromosomes) are frequently used for CGH arrays. Fluorescence ratios for every probe

spotted on the array are calculated. If any area of the genome is duplicated (as with trisomy), then the test sample will show an abnormally high signal for probes matching the duplicated region. If any areas are missing in the test sample (as with a microdeletion), then the test sample will show abnormally low signal in the affected regions.

With array-based CGH, it is possible to perform a molecular karyotype.[7] The resolution of a CGH array karyotype can be better than the resolution of conventional cytogenetics (deletions as small as 100,000 base pairs have been detected), and no culture time is required. Furthermore, probes can be included for all of the microdeletion regions commonly tested by ancillary fluorescent in situ hybridization tests (ie, Di George syndrome caused by a chromosome 22 microdeletion). Moreover, all of these common microdeletion regions can be tested in parallel rather than one at a time. The patterns obtained on arrays are easier to read by computer, compared with G-banded metaphase spreads. It is likely that CGH microarrays will become the preferred method for analyzing chorionic villi and amniocytes for chromosomal abnormalities. CGH microarrays cannot be used to detect balanced chromosomal changes or low-level mosaicism.

Expression profiling

Genes do not result in clinical phenotypes unless they are expressed as messenger ribonucleic acids (mRNAs) first and ultimately as proteins. DNA microarrays are used widely to study mRNA.[8,9] Because ribonucleic acid (RNA) is inherently unstable, mRNA is extracted from fresh cells or tissues to be studied and then reverse transcribed into a stable complementary deoxyribonucleic acid (cDNA) copy. Label is incorporated into the cDNA molecule as it is synthesized, and then the cDNA is placed on the array. DNA probes are now available for the cDNA for every known and predicted human gene and also for the genomes of all common experimental organisms. (However, recent experiments with tiling arrays show that gene predictions to date have underestimated the number of genes.[10]) These probes are arrayed as uniform sets, and the pattern obtained when the labeled test sample hybridizes to the array is the gene expression profile or signature for that starting test material. Probe sets exist that can examine every gene simultaneously, or the chip can be designed to focus on particular pathways (ie, inflammation).

Comparative profiling/differential expression

Usually cDNA arrays are used to screen for genes that are differentially expressed between 2 tissues (normal and diseased, treated and untreated). Comparative

analyses often consider what the genes are doing over time or after an intervention: which genes are upregulated and which genes are down-regulated.[11] The most common reason for doing these experiments is to try to find novel targets for drug development.[12] For example, by comparing the genes expressed in both normal and diseased ovaries, we might be able to find the genes, proteins, and pathways that are part of that ovarian disease process. An immediate result of this experiment might be the discovery of new biomarkers predicting clinical outcomes. Eventually, drugs may be discovered or developed targeted against these disease pathways (see for example recent work on breast cancer.[13]).

For most tissues and most known regulatory pathways, tremendous knowledge has already been gained, using many different investigational approaches. Microarray expression data can show whether particular gene pathways are regulated in a coordinated fashion or whether they are differentially regulated because of the disease process. Because thousands of pathways can be examined simultaneously, these analyses might suggest proteins that would be an appropriate drug target or which therapeutic compounds might have the lowest side effects (because they perturb the expression of the smallest number of genes). Expression profiles from tens of thousands of reference experiments are already available for most cell lines, tissue, physiological responses, pharmacologic response, etc to accelerate the analysis of any new data.

To date, our ability to generate comparative data is outpacing our capacity to draw the biologic meaning out of these experiments. The outputs from the microarrays and the underlying genetic pathways are extremely complex, but most analysis software packages perform only a limited number of tasks.[14] Large databases are used to store the expression data and to move the data between software packages. Replication and controls are critical. Positive controls can include spiking the test sample with DNA that should react with certain probes on the array, examining the expression of genes that are normally expressed in every cell (housekeeping genes), placing identical probes on different portions of the array, etc. For negative controls, DNA that should not hybridize can be included on the array and other positions can be left empty. The challenges related to data analysis will be considered only briefly in following text.

Genotyping: gene mapping/discovery

Most diseases have an intrinsic genetic component. Usually mutations or variants (polymorphisms) in several different genes contribute to complex disease risk. The mutations underlying common disease usually cause minor changes in gene expression or the amino acid structure of the encoded proteins. For most variant

alleles, the effect on the disease phenotype is rather weak. Furthermore, numerous minor genetic changes interact with nutritional, environmental, and other factors before resulting in the disease phenotype. Despite these challenges, microarrays that scan the entire genome in a single experiment have resulted in the discovery of dozens of important disease genes over the past year. The most powerful genome-wide approach uses single nucleotide polymorphisms (SNPs): DNA sequence variations that occur when a single nucleotide in the genomic sequence is altered. SNPs occur every 100 to 300 bases along the 3 billion-base human genome. For single-base changes to be considered a polymorphism and thus an SNP, the change must occur in at least 1% of the population.

For instance, the Affymetrix HuSNP arrays have either 100,000 or 500,000 SNPs selected on the basis of their location, their heterozygosity, and the likelihood that they are genetically linked to each other.[15] These microarrays are used to perform case-control association studies or relative pair studies. Subjects and controls are matched for the most readily apparent confounding factors (age, sex, known risk factors, etc).[16,17] Disease-associated alleles with modest relative risks (relative risk of 2 or more) can be detected with manageable sample sizes.[18,19]

As with resequencing arrays, the ability of SNP microarrays to discriminate perfect match versus mismatch hybridizations allows rapid and accurate typing of SNPs. Probes to adjacent nonpolymorphic sequences are included as controls. The SNP markers are less than 5000 base pairs apart on average for the 500,000 arrays. This marker density greatly enhances the chance of finding the actual disease-associated alleles and/or genetic markers strong linkage disequilibrium with the disease alleles.

When the SNP markers from a particular chromosomal region are very close together, the neighboring SNP markers do not give independent results (ie, certain SNP alleles are always observed in association with 1 particular allele for a neighboring SNP). This means that very little additional information is gained by testing SNPs that are right next to each other. Investigators are identifying the recombination hot spots in our genome that define the boundaries of blocks of SNPs, which tend to occur together as defined haplotypes (alleles strung together on the same DNA strand).[20] A limited number of common haplotype patterns account for most of the genetic variation in a block, and most of the variation in SNP allele frequencies can be accounted for by a small set of common haplotype patterns. The International HapMap Project is successfully describing the extent of these haplotype blocks in humans, promising greater efficiency in future disease gene discovery.[21,22] Current SNP chips are designed so that for every block in the human genome, there are enough SNP markers to determine the frequencies of these common haplotype patterns.

Because more than 500,000 case-control association studies can be tested simultaneously, corrections for multiple testing and very stringent significance levels need to be used when analyzing the results. False-positive association studies can occur. Most commonly this occurs if ethnicity in the cases and controls are not well matched or if there is hidden population stratification in the cases.[23] Founder effects reflecting the different racial or ethnic origin rather than any link to disease alleles can give rise to false-positive associations.[24] Various statistical methods are being developed to deal with this issue as we learn which of the SNPs tend to have more ethnic variation.[25,26] False associations can also be avoided by using of relatives as controls, ie, sibling pairs.[27]

When testing relatives, transmission disequilibrium testing, testing the frequency of transmission of a specific allele from a parent to an affected child, is compared with transmission of the other allele from the same parent, can also help avoid false conclusions. Ultimately replication in additional independent samples is the best way to be certain of the findings.

Clinical applications

Disease diagnosis/prognosis

As the relationships of individual genes and polymorphism to disease are discovered, this knowledge can be immediately used for diagnostic and prognostic tests using either microarrays or a variety of other formats for testing. In 2004 the Food and Drug Administration approved the first laboratory test using an actual microarray for medical use. The AmpliChip Cytochrome P450 genotyping test (Roche Molecular Systems Inc, Pleasanton, CA) analyzes 2 important cytochrome P450 genes. Enzymes encoded by these genes act in the liver to metabolize many commonly prescribed drugs (ie, beta blockers, many antidepressants, and some chemotherapy agents). Polymorphisms in cytochrome P450 genes can dramatically affect the rate of metabolism and thus the clinical effectiveness of these drugs. The Food and Drug Administration also approved the instrument to run this test, Affymetrix's GeneChip System 3000Dx chip reader as an in vitro diagnostic device. Both products have also been approved in Europe.

In the coming decade, microarrays will allow rapid assessment of the fetal genotype in prenatal diagnosis, more accurate and extensive newborn screening,[28] better measures of viral loads and resistance or virulence factors, and more complete characterization of malignant lesions. Scientists are also studying physiologic

processes like endometrial receptivity using microarrays.[29,30] Bioterrorism concerns and newly emerging epidemics like severe acute respiratory syndrome have already caused development of microarrays for the rapid identification of infected individuals and rapid characterization of the threatening organism.[31] In industrial-scale diagnostics, microarrays may be used to detect genetically modified organisms or microbial contaminants in foods.[31] Systems have already been designed to allow point-of-care testing by staff with no molecular biology training.

Oncology applications

All cancers have genetic changes; microarray technology will be useful in assessing the degree of genetic damage in both the primary tumor and the surrounding issues, which could alert to the probability of tumor recurrence.[32,33] Even though tissue margins close to resected tumors may look microscopically normal, microarrays can be used to detect the genetic damage that crosses these histologic margins. All tumors will be analyzed by microarray technology to allow correct selection of primary and adjuvant treatment by predicting the sensitivity of each tumour to radiation and to various chemotherapeutic agents. Microarrays may be useful in predicting those cases of dysplasia or atypical benign lesions that will undergo malignant transformation.[34] Microarrays are beginning to unravel the mysteries underlying metastasis, and they are likely to be used in the future to predict the likelihood of metastasis.[35] Certainly expression profiles of melanoma[36] and breast cancers[37-40] have already led to advances in methods of staging and classifying these diseases. Patients with tumors can be subdivided into distinct groups based on their gene expression profiles, even though there were no obvious pathological differences between the tumors.[41]

Pharmacogenomics/toxicogenomics

Why do some drugs work better in some patients than in others? And why are some drugs highly toxic to certain patients? The field of pharmacogenomics is using microarrays to find correlations between therapeutic responses to drugs and the genetic profiles of patients.[42-45] A related field, toxicogenomics, seeks to find correlations between toxic responses to toxicants and changes in the genetic profiles of the objects exposed to such toxicants.[46-48] By identifying individuals with similar biological patterns, microarray analysis can assist drug companies in choosing the most appropriate candidates for clinical trials of new drugs. In the future, this technology my lead to personalized medicine, in which patients are prescribed drugs that are very likely to be effective and free of side effects, given their individual profile.

Pathogen detection[49]

Diagnostic assays for acute infections are rapidly changing from antibody detection to pathogen detection, from slower culture-based methods to rapid molecular methods, from clinical laboratory based to point-of-care–based tests, from detection of only a few types of organisms at a time to simultaneous detection of multiple pathogens. Microarrays have the ability to detect viruses, bacteria, and other microorganisms all on the same chip. Host studies are unraveling the development and activation of both innate and adaptive immunity; others are studying global gene expression of both the pathogen and the patient during progression. In the near future, virulence factors, resistance factors, and host response to the pathogen will all be monitored in parallel.[50,51]

Sequence-based tracking of pathogens has allowed more thorough evaluation of recent outbreaks such as monkey-pox or severe acute respiratory syndrome. Rapid point-of-care devices allow detection and surveillance of infections at ports of entry and will be very helpful in the event of a bioterrorism attack.

Problems and pitfalls

Microarrays are a tremendous advance, but they are still too costly and too difficult for widespread clinical use. Greater automation and increasing sophistication of the analytic paradigms are already on the horizon. The cost of arrays will decrease as patent protections expire.

As microarrays have improved, data analysis rather than data production became the critical issue. Microarray analysis has been addressed in a number of recent reviews.[14,29,52-59] As the reader can imagine, the first difficulty comes with analysis of the image off the array. Is any fluorescent spot identified a true spot? Is the detector aligned (recall that the spots are only micrometers apart)? Is the background variable? Is there evidence of uneven hybridization? Each feature has to be detected and localized correctly. Many variables can affect the signal obtained.

Massive amounts of data are produced by microarray experiments. With so many features being tested at once, corrections for multiple testing need to be applied to any tests of significance. Important, low-level signals can be missed because of the need for conservative statistical analyses.

Interpretation of gene expression has many more pitfalls than interpretation of genotyping or resequencing chips.[8,60] First of all, mRNA is an unstable molecule, so all current methods require fresh or frozen tissue. Rigorous quality control is essential and any tissue-processing protocol can introduce artifacts. Tissues are usually heterogenous and cancers even more so. Fine-needle aspirations, core-needle biopsies, and surgically resected specimens yield somewhat different transcriptional profiles from the same tumor.[61,62]

Statisticians and the software they write must make basic assumptions about the behavior of genes. Most of the analysis programs assume that genes that are expressed together in time and space are likely to participate in similar physiological processes.[63] We assume there is a link between the expression of a gene and its function; however, we know that genes are often expressed in settings in which no protein is subsequently produced, and therefore, expression may not have changed the physiology in a straightforward way.[64] Unfortunately, protein arrays that analyze downstream events lag well behind the nucleic acid arrays.

Sophisticated software is used to detect genes with different expression under different conditions. Because expression constellations can involve hundreds of individual genes, it can be difficult to determine whether and how a group of samples can be divided into 2 or more distinct groups. For many applications, expressed genes that meet certain quality criteria are first analyzed by unsupervised methods. These methods are less biased and more likely to discover previously unrecognized pathways based on the clustering of gene-expression profiles. Three-dimensional scatter plots are examined to determine whether related specimens cluster together. When they do, self-organizing groups can develop into evolutionary trees as multiple experiments are compared and as various types of regression analyses are performed. Some of the associations, although statistically significant, will not reflect actual pathways. The precision of the clustering analysis will increase as additional expression data become available.

Supervised analysis of expression data depends on known relationships between genes and the results of related expression studies. A number of public databases can provide pattern information useful for interpreting newly observed patterns of gene expression.

The National Center for Biotechnology Information, part of the National Library of Medicine, hosts the Gene Expression Omnibus (GEO). GEO (accessible at www.ncbi.nlm.nih.gov/geo) is a public repository that archives and distributes via the internet gene expression data submitted by the scientific community. GEO currently contains more than half a billion individual gene expression measurements, derived from more than 100 organisms, addressing a wide range of biological issues. Each submission includes data on the experimental design, the array design (each array used and each spot on the array), the preparation and labeling of the samples used, the hybridization procedures and parameters, the measurement methodology, and the types and characteristics of controls used. Various query and visualization tools are available on the GEO Web site.

Most other software packages can also query these databases. Unfortunately, when some patterns are compared, spurious patterns from early experiments may corrupt the analysis of new array data. A lack of standardization makes it difficult to compare data produced by different systems and has made it difficult to merge data.[63]

Conclusions

By providing global views of biological processes, microarrays enable systematic surveys of variations in DNA sequence and gene expression. Microarrays are fueling novel and expansive research. The current $2 billion per year market for microarrays in the United States is growing by more than 30% each year.[65] Patients are likely to benefit from this research activity as it leads to improved genetic diagnostics, personalized treatments, and more rapid and definitive testing of clinical specimens.

References

1. Ewis AA, Zhelev Z, Bakalova R, Fukuoka S, Shinohara Y, Ishikawa M, et al. A history of microarrays in biomedicine. Expert Rev Mol Diagn 2005;5:315-28.

2. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991;251:767-73.

3. Auburn RP, Kreil DP, Meadows LA, Fischer B, Matilla SS, Russell S. Robotic spotting of cDNA and oligonucleotide microarrays. Trends Biotechnol 2005;23:374-9.

4. Mandal MN, Heckenlively JR, Burch T, Chen L, Vasireddy V, Koenekoop RK, et al. Sequencing arrays for screening multiple genes associated with early-onset human retinal degenerations on a high-throughput platform. Invest Ophthalmol Vis Sci 2005;46:3355-62.

5. Davignon L, Walter EA, Mueller KM, Barrozo CP, Stenger DA, Lin B. Use of resequencing oligonucleotide microarrays for identification of Streptococcus pyogenes and associated antibiotic resistance determinants. J Clin Microbiol 2005;43:5690-5.

6. Vissers LE, Veltman JA, van Kessel AG, Brunner HG. Identification of disease genes by whole genome CGH arrays. Hum Mol Genet 2005;14(Spec No 2):R215-23.

7. Lapierre JM, Tachdjian G. Detection of chromosomal abnormalities by comparative genomic hybridization. Curr Opin Obstet Gynecol 2005;17:171-7.

8. Baggerly KA, Coombes KR, Hess KR, Stivers DN, Abruzzo LV, Zhang W. Identifying differentially expressed genes in cDNA microarray experiments. J Comput Biol 2001;8:639-59.

9. Imbeaud S, Auffray C. 'The 39 steps' in gene expression profiling: critical issues and proposed best practices for microarray experiments. Drug Discov Today 2005;10:1175-82.

10. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR. Applications of DNA tiling arrays for whole-genome analysis. Genomics 2005;85:1-15.

11. Broberg P. A comparative review of estimates of the proportion unchanged genes and the false discovery rate. BMC Bioinformatics 2005;6:199.

12. Dechering KJ. The transcriptome's drugable frequenters. Drug Discov Today 2005;10:857-64.

13. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 2006;439:353-7.

14. Marquez Mdel C, Perez PP, Lagunez-Otero J. An evolving neural network for the interpretation of gene expression patterns. Omics 2005;9:209-17.

15. Di X, Matsuzaki H, Webster TA, Hubbell E, Liu G, Dong S, et al. Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. Bioinformatics 2005;21:1958-63.

16. Dean M. Approaches to identify genes for complex human diseases: lessons from Mendelian disorders. Hum Mutat 2003;22:261-74.

17. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45-61.

18. Tabor HK, Risch NJ, Myers RM. Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3:391-7.

19. Hao K, Xu X, Laird N, Wang X, Xu X. Power estimation of multiple SNP association test of case-control study and application. Genet Epidemiol 2004;26:22-30.

20. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001;294:1719-23.

21. Morton NE. Linkage disequilibrium maps and association mapping. J Clin Invest 2005;115:1425-30.

22. Farrall M, Morris AP. Gearing up for genome-wide gene-association studies. Hum Mol Genet 2005;14(Spec No 2):R157-62.

23. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004;36:388-93.

24. Koller DL, Peacock M, Lai D, Foroud T, Econs MJ. False positive rates in association studies as a function of degree of stratification. J Bone Miner Res 2004;19:1291-5.

25. Olivier M. A haplotype map of the human genome. Physiol Genomics 2003;13:3-9.

26. Zhang K, Calabrese P, Nordborg M, Sun F. Haplotype block structure and its applications to association studies: power and study designs. Am J Hum Genet 2002;71:1386-94.

27. Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 1998;8:1273-88.

28. Green NS, Pass KA. Neonatal screening by DNA microarray: spots and chips. Nat Rev Genet 2005;6:147-51.

29. Hartmann O. Quality control for microarray experiments. Methods Inf Med 2005;44:408-13.

30. Schmidt A, Groth P, Haendler B, Hess-Stumpp H, Kratzschmar J, Seidel H, et al. Gene expression during the implantation window: microarray analysis of human endometrial samples. Ernst Schering Res Found Workshop 2005;52:139-57.

31. Cebula TA, Jackson SA, Brown EW, Goswami B, LeClerc JE. Chips and SNPs, bugs and thugs: a molecular sleuthing perspective. J Food Prot 2005;68:1271-84.

32. Bucca G, Carruba G, Saetta A, Muti P, Castagnetta L, Smith CP. Gene expression profiling of human cancers. Ann N Y Acad Sci 2004;1028:28-37.

33. Brentani RR, Carraro DM, Verjovski-Almeida S, Reis EM, Neves EJ, de Souza SJ, et al. Gene expression arrays in cancer research: methods and applications. Crit Rev Oncol Hematol 2005;54:95-105.

34. Cuperlovic-Culf M, Belacel N, Ouellette RJ. Determination of tumour marker genes from gene expression data. Drug Discov Today 2005;10:429-37.

35. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. Nature 2005;436:518-24.

36. Nambiar S, Mirmohammadsadegh A, Doroudi R, Gustrau A, Marini A, Roeder G, et al. Signaling networks in cutaneous melanoma metastasis identified by complementary DNA microarrays. Arch Dermatol 2005;141:165-73.

37. Chang JC, Hilsenbeck SG, Fuqua SA. Genomic approaches in the management and treatment of breast cancer. Br J Cancer 2005;92:618-24.

38. Fuqua SA, Chang JC, Hilsenbeck SG. Genomic approaches to understanding and treating breast cancer. Breast Dis 2004;19:35-46.

39. Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol 2005;23:7350-60.

40. Jeffrey SS, Lonning PE, Hillner BE. Genomics-based prognosis and therapeutic prediction in breast cancer. J Natl Compr Canc Netw 2005;3:291-300.

41. Brown RE. Morphoproteomics: exposing protein circuitries in tumors to identify potential therapeutic targets in cancer patients. Expert Rev Proteomics 2005;2:337-48.

42. Schmith VD, Campbell DA, Sehgal S, Anderson WH, Burns DK, Middleton LT, et al. Pharmacogenetics and disease genetics of complex diseases. Cell Mol Life Sci 2003;60:1636-46.

43. Burczynski ME, Oestreicher JL, Cahilly MJ, Mounts DP, Whitley MZ, Speicher LA, et al. Clinical pharmacogenomics and transcriptional profiling in early phase oncology clinical trials. Curr Mol Med 2005;5:83-102.

44. Luhe A, Suter L, Ruepp S, Singer T, Weiser T, Albertini S. Toxicogenomics in the pharmaceutical industry: hollow promises or real benefit? Mutat Res 2005;575:102-15.

45. Yengi LG. Systems biology in drug safety and metabolism: integration of microarray, real-time PCR and enzyme approaches. Pharmacogenomics 2005;6:185-92.

46. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 1999;24:153-9.

47. Gant TW, Zhang SD. In pursuit of effective toxicogenomics. Mutat Res 2005;575:4-16.

48. Shioda T. Application of DNA microarray to toxicological research. J Environ Pathol Toxicol Oncol 2004;23:13-31.

49. Bryant PA, Venter D, Robins-Browne R, Curtis N. Chips with everything: DNA microarrays in infectious diseases. Lancet Infect Dis 2004;4:100-11.

50. Longley DB, Johnston PG. Molecular mechanisms of drug resistance. J Pathol 2005;205:275-92.

51. McGuire K, Glass EJ. The expanding role of microarrays in the investigation of macrophage responses to pathogens. Vet Immunol Immunopathol 2005;105:259-75.

52. Bretz F, Landgrebe J, Brunner E. Multiplicity issues in microarray experiments. Methods Inf Med 2005;44:431-7.

53. Curtis RK, Oresic M, Vidal-Puig A. Pathways to the analysis of microarray data. Trends Biotechnol 2005;23:429-35.

54. Krallinger M, Erhardt RA, Valencia A. Text-mining approaches in molecular biology and biomedicine. Drug Discov Today 2005;10:439-45.

55. Kreil DP, Russell RR. There is no silver bullet—a guide to low-level data transforms and normalisation methods for microarray data. Brief Bioinform 2005;6:86-97.

56. Qin L, Rueda L, Ali A, Ngom A. Spot detection and image segmentation in DNA microarray data. Appl Bioinformatics 2005;4:1-11.

57. Reimers M. Statistical analysis of microarray data. Addict Biol 2005;10:23-35.

58. Sherlock G, Ball CA. Storage and retrieval of microarray data and open source microarray database software. Mol Biotechnol 2005;30:239-51.

59. White CA, Salamonsen LA. A guide to issues in microarray analysis: application to endometrial biology. Reproduction 2005;130:1-13.

60. Armstrong NJ, van de Wiel MA. Microarray data analysis: from hypotheses to conclusions using gene expression data. Cell Oncol 2004;26:279-90.

61. Liang Y, Kelemen A. Associating phenotypes with molecular events: recent statistical advances and challenges underpinning microarray experiments. Funct Integr Genomics 2006;6:1-13.

62. Kanehisa M. The KEGG database. Novartis Found Symp 2002; 247:91-101; discussion 101-3, 119-28, 244-52.

63. Bracken MB. Genomic epidemiology of complex disease: the need for an electronic evidence-based approach to research synthesis. Am J Epidemiol 2005;162:297-301.

64. Dobson PD, Cai YD, Stapley BJ, Doig AJ. Prediction of protein function in the absence of significant sequence similarity. Curr Med Chem 2004;11:2135-42.

65. Stoll D, Templin MF, Bachmann J, Joos TO. Protein microarrays: applications and future challenges. Curr Opin Drug Discov Devel 2005;8:239-52.

66. Angenendt P. Progress in protein and antibody microarray technology. Drug Discov Today 2005;10:503-11.

67. Adams EW, Ratner DM, Bokesch HR, McMahon JB, O'Keefe BR, Seeberger PH. Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. Chem Biol 2004;11:875-81.

68. Hirabayashi J. Oligosaccharide microarrays for glycomics. Trends Biotechnol 2003;21:141-3; discussion 143.

69. Uttamchandani M, Walsh DP, Yao SQ, Chang YT. Small molecule microarrays: recent advances and applications. Curr Opin Chem Biol 2005;9:4-13.

70. Fang Y, Lahiri J, Picard L. G protein-coupled receptor microarrays for drug discovery. Drug Discov Today 2003;8:755-61.

71. Fiorini GS, Chiu DT. Disposable microfluidic devices: fabrication, function, and application. Biotechniques 2005;38:429-46.

72. Haab BB, Dunham MJ, Brown PO. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2001;2: RESEARCH0004.

73. Kersten B, Wanker EE, Hoheisel JD, Angenendt P. Multiplex approaches in protein microarray technology. Expert Rev Proteomics 2005;2:499-510.

74. Lueking A, Cahill DJ, Mullner S. Protein biochips: a new and versatile platform technology for molecular medicine. Drug Discov Today 2005;10:789-94.

75. Wheeler DB, Carpenter AE, Sabatini DM. Cell microarrays and RNA interference chip away at gene function. Nat Genet 2005; 37(Suppl):S25-30.

76. Taylor MF. Antisense oligonucleotides for target validation and gene function determination. IDrugs 1999;2:777-81.

77. Vanhecke D, Janitz M. Functional genomics using high-throughput RNA interference. Drug Discov Today 2005;10:205-12.

78. Bertone P, Gerstein M, Snyder M. Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Res 2005;13:259-74.

79. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, et al. Experimental annotation of the human genome using microarray technology. Nature 2001;409:922-7.

80. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet 2005;21:93-102.

81. Fedor HL, De Marzo AM. Practical methods for tissue microarray construction. Methods Mol Med 2005;103:89-101.

82. Tzankov A, Went P, Zimpfer A, Dirnhofer S. Tissue microarray technology: principles, pitfalls and perspectives-lessons learned from hematological malignancies. Exp Gerontol 2005;40:737-44.

83. Long WH, Xiao HS, Gu XM, Zhang QH, Yang HJ, Zhao GP, et al. A universal microarray for detection of SARS coronavirus. J Virol Methods 2004;121:57-63.

84. Wong CW, Albert TJ, Vega VB, Norton JE, Cutler DJ, Richmond TA, et al. Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. Genome Res 2004;14:398-405.

85. Striebel HM, Birch-Hirschfeld E, Egerer R, Foldes-Papp Z. Virus diagnostics on microarrays. Curr Pharm Biotechnol 2003;4:401-15.