

# DNA Microarray Analysis of Complex Biologic Processes

MANJULA KURELLA,\* LI-LI HSIAO,\* TAKUMI YOSHIDA,\*  
JEFFREY D. RANDALL,\* GARY CHOW,\* SATINDER S. SARANG,\*  
RODERICK V. JENSEN,<sup>†</sup> and STEVEN R. GULLANS\*

\*Renal Division, Brigham and Women's Hospital, Boston, Massachusetts; and <sup>†</sup>Department of Physics, Wesleyan University, Middletown Connecticut.

**Abstract.** DNA microarrays, or gene chips, allow surveys of gene expression, (*i.e.*, mRNA expression) in a highly parallel and comprehensive manner. The pattern of gene expression produced, known as the expression profile, depicts the subset of gene transcripts expressed in a cell or tissue. At its most fundamental level, the expression profile can address qualitatively which genes are expressed in disease states. However, with the aid of bioinformatics tools such as cluster analysis, self-organizing maps, and principle component analysis, more

sophisticated questions can be answered. Microarrays can be used to characterize the functions of novel genes, identify genes in a biologic pathway, analyze genetic variation, and identify therapeutic drug targets. Moreover, the expression profile can be used as a tissue or disease “fingerprint.” This review details the fabrication of arrays, data management tools, and applications of microarrays to the field of renal research and the future of clinical practice.

Recent advances in functional genomics have made possible new approaches to the diagnosis and management of a wide array of renal disorders. With new tools to explore gene expression and regulation, researchers can understand better the molecular basis of disease. Of the estimated 30,000 human genes, only 30% are functionally understood. Moreover, fewer than 3% of identified genes have been characterized in renal disease, underscoring the need for functional genomics techniques in renal research. Although recent research has uncovered the genetic basis of hereditary disorders, including familial focal segmental glomerulosclerosis and adult polycystic kidney disease, using traditional genetic mapping techniques, more comprehensive approaches are needed to identify the complex sequence of perturbations that underlie polygenic conditions such as hypertension and chronic renal insufficiency (1–3). Among the most powerful of these new tools are DNA microarrays, capable of genome-wide profiles of mRNA expression.

A DNA microarray, or gene chip, is a matrix of thousands of cDNA or oligonucleotides imprinted on a solid support (4,5). Labeled mRNA from the tissue of interest is hybridized to its sequence complement on the array to provide a measure of mRNA abundance in the sample. The hallmark of the microarray experiment is the expression profile, the pattern of gene expression produced by the experimental sample (Figure 1). Arrays composed of DNA fragments are not new (6,7). How-

ever, early arrays included only a small set of genes thought to be involved in the process being studied. Significant improvements in substrate materials, robotics, and signal detection have made possible miniaturization of arrays with the result that hundreds of thousands of oligonucleotides can be arrayed on a square-centimeter chip. This important feature makes it possible to study gene expression without specifying in advance which genes are to be studied. Thus, DNA microarrays permit systematic and comprehensive surveys of gene expression in an efficient manner.

## Principles of Microarray Technology

### Microarray Structure

Two variations of microarrays exist: (1) customized cDNA microarrays composed of cDNA or oligonucleotides and (2) commercially produced high-density arrays, *e.g.*, Affymetrix GeneChip (Affymetrix, Santa Clara, CA), containing synthesized oligonucleotides (8). The first type of array can analyze RNA from two different samples on a single chip but requires a source of genes to be spotted onto the chip, usually expressed sequence tag clones or oligonucleotides. High-density commercial arrays provide expression analysis over a larger number of genes (12,000 human genes in the case of the Affymetrix GeneChip) but can analyze only a single sample on one chip and at considerable cost, making them unsuitable for large-scale experiments in most academic laboratories. Both types of arrays produce sensitive and accurate expression data.

### Creating a cDNA Microarray

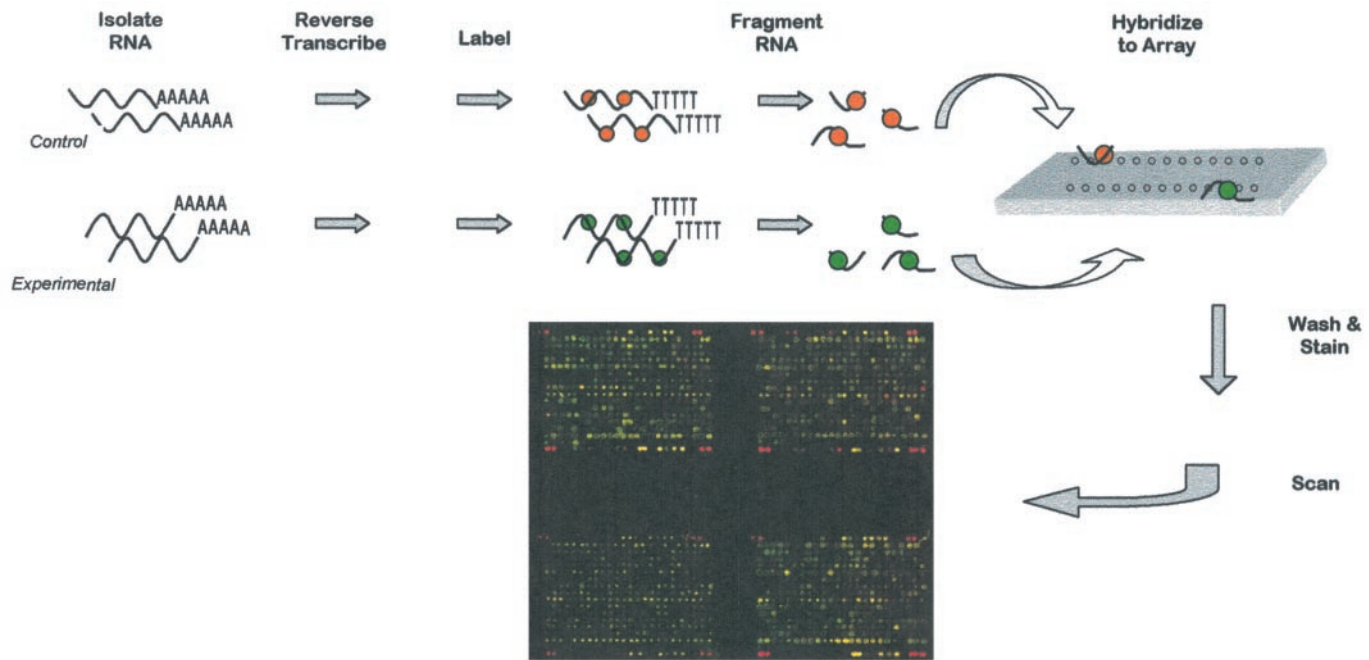
Customized cDNA microarrays are fabricated by first selecting the genes to be printed on the array from public databases/repositories or institutional sources. High throughput DNA preparation, usually done by robotics systems, consists of

Correspondence to Dr. Steven R. Gullans, Harvard Institutes of Medicine, 5th Floor, 77 Avenue Louis Pasteur, Boston, MA 02115. Phone: 617-525-5712; Fax: 617-525-5711; E-mail: sgullans@rics.bwh.harvard.edu

1046-6673/1205-1072

Journal of the American Society of Nephrology

Copyright © 2001 by the American Society of Nephrology



**Figure 1.** Performing a microarray experiment. To perform a microarray experiment, RNA from the experimental sample(s) is first isolated and purified. The purified RNA is then reverse-transcribed in the presence of labeled nucleotides. In the case of custom-made arrays, the fluorophores Cy3 and Cy5 typically are used. The two-color hybridization strategy permits simultaneous analysis of two samples on a single array (shown here). For high-density commercial arrays, nonfluorescent biotin labeled by staining with a fluorescent streptavidin conjugate typically is used. The labeled probe is fragmented and hybridized to the array, and then the array is washed and stained. Signal intensity, proportional to the amount of bound probe, is measured by scanning with a confocal laser. Background signal is subtracted from the average signal intensity for each spot on the array to generate a quantitative image. Because the sequence of each cDNA or oligonucleotide on the grid is known, the relative abundance of each transcript can be determined. Data are normalized across experiments by calculating the variance of all genes in the sample or of a known subset of unchanging, *e.g.*, maintenance, genes.

tens of thousands of PCR reactions. Purified PCR products representing specific genes are spotted onto a matrix. Spotting is carried out by a robot, which deposits a nanoliter of PCR product onto the matrix in serial order. Nylon filter arrays largely have been replaced by glass-based arrays, typically microscope slides, which have the advantage of two-color fluorescence labeling with low inherent background fluorescence. DNA adherence to the slide is enhanced by treatment with polylysine or other cross-linking chemical coating. Spotted DNA is cross linked to the matrix by ultraviolet irradiation and denatured by exposure to either heat or alkali. The Affymetrix GeneChip is produced by a novel photolithographic method in which thousands of different oligonucleotide probes are synthesized *in situ* on the array (8).

### Data Management

Once the hybridized chip is scanned, data flow through the following steps. Data are collected and saved as both an image and a text file. Of critical importance is that precise databases and tracking files be maintained regarding the spot configuration of all chips. These contain information on the location and names of genes arrayed on each chip. The saved files are imported to software programs that perform image analysis and statistical analysis functions. Finally, the data are mined for induced or repressed genes, patterns of gene expression, and

temporal relationships of expression under different experimental conditions. A significant challenge exists in making sense of the vast quantity of data generated by microarray experiments. There is no single tool that meets all of the needs of the microarray researcher. Collections of software programs are used to perform a multitude of tasks, including data tracking, image analysis, database storage, data queries, statistical analysis, multidimensional visualization, and interaction with public databases on the Internet. Basic spreadsheet programs can be adapted to answer questions regarding magnitude of change in gene expression. However, limitations often arise as a result of inadequate memory capacity for managing the enormous data sets. More sophisticated analytical tools, including cluster analysis, self-organizing maps, and principle component analysis, have been applied to biologic data to extract higher-order relationships embedded in expression patterns. The concepts that underlie these analytical methods are illustrated below.

### Applications of Microarrays

#### Gene Expression and Discovery

One of the most important and fundamental questions answered by the expression profile is, "Which genes are expressed and to what magnitude?" The expression profile represents the subset of gene transcripts or mRNA expressed in a

cell or tissue. The underlying assumption is that the relative abundance of mRNA transcripts represents the cellular response to a particular state. DNA microarrays have been used to study gene expression in a variety of organisms, including yeast, plants, and humans (9–11). Microarray experiments produce profiles of gene expression that reflect the transcriptional response of thousands of genes to a change in cellular state or in response to a pharmacologic stimulus. The typical goal of such experiments is the identification of new genes involved in a pathway, or diagnostic and/or prognostic expression markers that characterize a disease state. Such an approach permits a genome-wide survey in a *single* assay, without the need to identify potentially important genes *a priori*. In this sense, microarray experiments can be considered hypothesis generating rather than hypothesis driven. The expression profiles generated from microarray experiments can be used as a launching point to identify candidate genes for further study using traditional techniques such as Northern or Western blotting, reverse transcription-PCR, and gene transfection. As a general rule, 25 to 30% of genes present in the human genome are expressed in a given tissue, and expression of 10% of these genes is changed in response to a given stimulus. Two types of data can be generated from microarray experiments: static or dynamic. *Static data* refers to data that compare one sample with a second, independent sample, *e.g.*, disease tissue *versus* normal tissue. *Dynamic data* refers to data that are obtained temporally from a sample, *e.g.*, disease progression over time. The distinction is important because the experimental design and statistical methods differ for each type of experiment.

Several investigators demonstrated the utility of a global approach to expression profiling. Genes that are expressed preferentially in inflammatory disease states, such as inflammatory bowel disease and rheumatoid arthritis, have been identified using microarrays (12). In addition to identifying known genes, the use of microarrays allowed investigators to identify several novel genes that are expressed in these inflammatory conditions. Moreover, investigators often find that genes that are known to be important in another, unrelated context are unexpectedly involved in the process being studied. Expression profiles have also been used to identify genes that are important in tumorigenesis and to identify novel genes in multiple sclerosis, Alzheimer's disease, and viral hepatitis (13–18).

### Predicting Gene Function

In addition to providing a broad survey of gene expression, transcriptional profiling can reveal patterns of gene expression, which can be used to predict gene function. This is accomplished by grouping genes into sets, or clusters, with similar expression profiles produced over multiple experiments. This grouping can be performed either by visual inspection of the data or by using statistical methods. It is expected that genes that display similar expression patterns are functionally related such that genes in a pathway, *e.g.*, glycolysis, should be co-regulated under all experimental conditions. In a landmark study, DeRisi *et al.* (9) used differential gene expression to examine the temporal response of yeast undergoing the shift

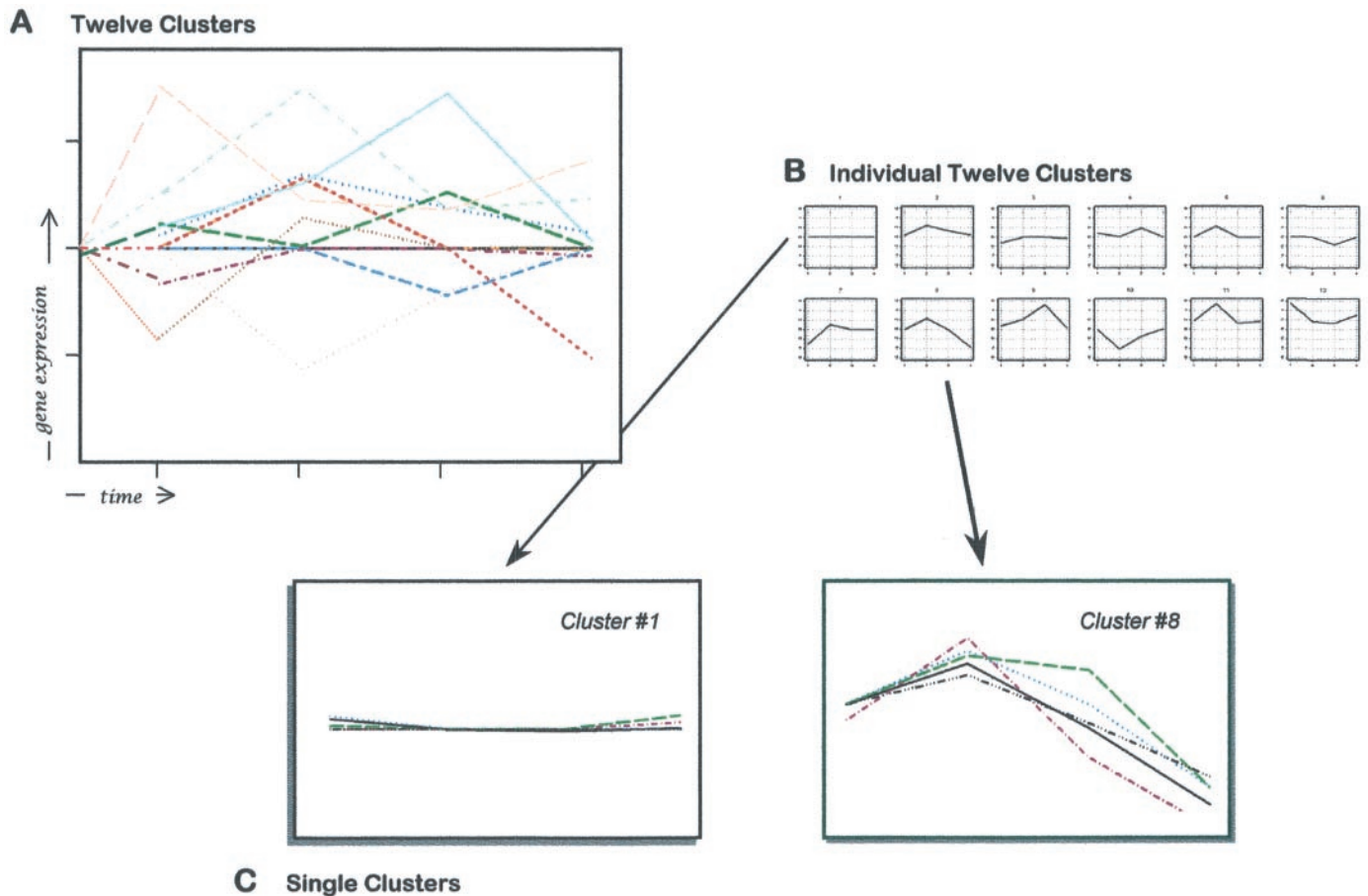
from anaerobic to aerobic metabolism, known as the diauxic shift. With the aid of clustering algorithms, distinct temporal patterns of gene expression were identified and genes were grouped on the basis of the similarity of their expression profiles. For example, cytochrome *c*-related genes, TCA/glyoxylate cycle-related genes, and genes involved in carbohydrate storage were coordinately induced during the diauxic shift. Importantly, temporal analysis revealed expression patterns in which families of genes with similar functions were discovered to be co-regulated. Thus, expression profiling using DNA microarrays can reveal co-regulated and therefore putative co-functional families of genes.

Expanding on these results, Hughes *et al.* (19) created a reference database, or compendium, of expression profiles in yeast cells corresponding to diverse genetic mutations and drug treatments. They showed that different mutants or treatments that affect similar cellular processes displayed similar expression profiles. Furthermore, they were able to identify cellular functions of unknown genes by comparing the expression profile of the corresponding deletion mutant with profiles of known mutants in the database that produced similar profiles. The strength of the compendium approach to functional discovery is that it relies solely on pattern recognition in the database of profiles. Knowledge of other genes in a pathway, regulatory elements, or even the complete sequence of the gene of interest need not be known to use this approach.

An example of cluster analysis generated from temporal gene expression data is shown in Figure 2. In this hypothetical experiment, gene expression in response to the experimental stimulus is measured at five time points. The data are plotted as expression level *versus* time (in this case, the log ratio of experimental to control expression level). In a cluster analysis, the expressed genes are grouped into clusters with similar expression patterns. Each line represents the average behavior of a discrete gene cluster containing 5 to 500 genes. The aggregate of gene clusters can be viewed together, as in panel A, or individually, as in panel B. In panel C, two of the individual clusters are magnified: clusters 1 and 8. In both graphs, the individual genes that compose each cluster are shown as a different line. In cluster 1, the genes show no change in expression level in response to the experimental stimulus. Such a profile might be expected for a cluster of housekeeping or maintenance genes. In cluster 8, the genes display an early wave of increased expression followed by a rapid decline. Such a profile might be expected for genes that are associated with transcription and translation regulation. Hence, cluster analysis is a powerful tool for identifying co-functional gene families.

### Linking Cell Pathways

Beyond studying the expression levels of individual genes under various conditions, the patterns produced by mRNA expression profiling can be exploited to study links among various cell pathways, the sequence of signaling within a pathway, and common regulatory mechanisms. In their study of differential gene expression in yeast cells, DeRisi *et al.* (9) identified transcription regulatory sequences as well. Because



**Figure 2.** Cluster analysis depicting co-regulated clusters of genes. The typical output of a cluster analysis of temporal gene expression data. Time is plotted on the x-axis, and the log of gene expression level is plotted on the y-axis. (A) All expressed genes are assigned a cluster, and the clusters are visualized on a single graph, with each line representing a different gene cluster. (B) The 12 clusters are seen individually. (C) Two of the individual clusters are magnified. Cluster 1 represents a group of genes whose expression is not changed in response to the experimental stimulus, *e.g.*, cell maintenance genes. In contrast, the group of genes that compose cluster 8 is induced early and then rapidly declines in response to the stimulus, *e.g.*, transcription factors.

the sequence of the entire yeast genome is known, they were able to examine the gene promoter region sequence of many of the genes within a co-regulated cluster and discovered that many shared common regulatory sequences. For example, seven of the genes that displayed a late induction profile during the diauxic shift were shown to possess a common upstream activating sequence, the carbon source response element. Similar observations were made in other gene clusters. Using more sophisticated informatics, Roth *et al.* (20) were able to identify distinct promoter regulatory motifs that are responsible for coordinated gene expression in yeast. Given these striking findings, some investigators have suggested that a compendium of expression behavior could be used to predict regulatory elements and thus obviate the need for conventional methods of studying genetic regulatory sequences using site-directed mutagenesis (21). Thus, by assembling profiles of deletion (or overexpression) mutants that are exposed to various physiologic stimuli, reasonable maps of genetic circuitry can be deduced.

The task of unraveling cell networks in eukaryotic cells is

considerably more difficult because of the enormity of the human genome, the complexity of intron/exon splicing, and the vast number of cell perturbations possible. Nonetheless, temporal analysis of gene expression profiles is a valuable tool, which can suggest the framework of cell pathways. Temporal patterns can reveal information on the coordinated regulation of genes involved in cell cycle, signal transduction, metabolism, transcription, and other cellular processes. The coordinated regulation of genes acting at different steps in a common cell process allows researchers to dissect complex cell pathways by examining temporal expression profiles. Iyer *et al.* (11) studied the temporal response of fibroblasts exposed to serum using this approach. They found that genes that are involved in programs of cell cycle and proliferation, inflammation, angiogenesis, tissue remodeling, and cytoskeletal reorganization each displayed distinct expression patterns.

#### *Detection of Mutations and Polymorphisms*

Variation in the human genetic code, *i.e.*, DNA sequence, has been studied using gene chips. Approximately 0.1%, or

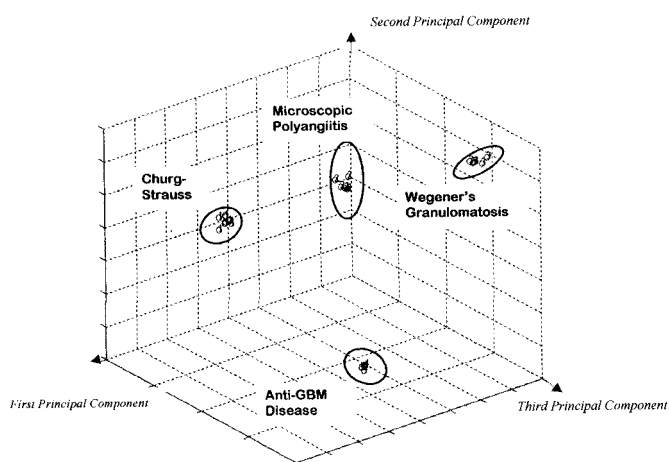
3,000,000, of nucleotides of the human genome is variant within the human population. Detecting these variations is critical to associating them with disease onset or therapeutic outcome. Recent studies illustrating the use of gene chips for this purpose include screening for mutations that lead to drug resistance in the HIV-1 genome, detection of heterozygous mutations in the BRCA1 breast and ovarian cancer gene, identification of mutations in the  $\beta$ -globin gene in  $\beta$ -thalassemia patients, and detection of polymorphisms in the human mitochondrial genome (22–25). By analyzing population-based genetic polymorphisms, clinicians could tailor therapeutic choices to individual patients. For example, hypertensive

therapy or an immunosuppressive regimen could be tailored to a patient's genotype profile. Ideally, then, therapeutic decisions would be made on the basis of the underlying pathophysiology in an individual patient, thereby limiting drug toxicities.

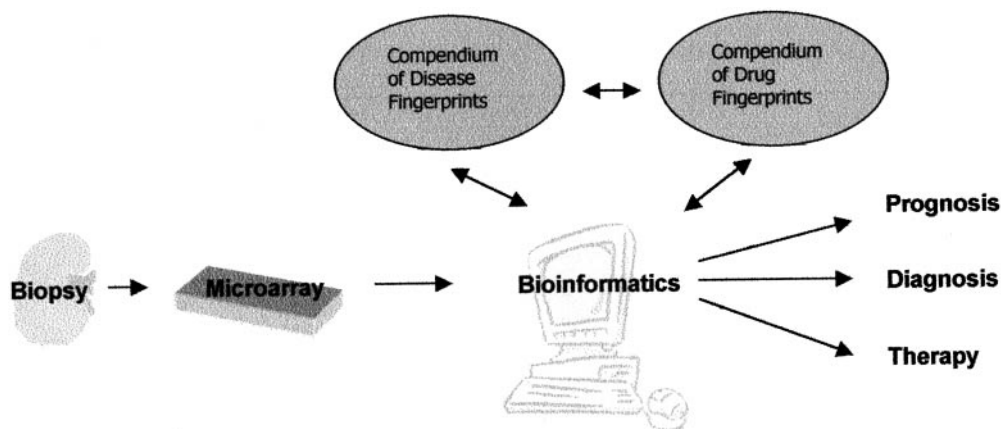
#### Expression Profile as a "Fingerprint" of Cellular or Disease Phenotype

The expression profile produced by microarray experiments represents the transcriptional response of a cell to a particular stimulus. As demonstrated by early microarray experiments, the response elicited is tightly regulated and highly distinct. Indeed, the pattern of gene expression could be considered the "fingerprint" of a cell or tissue in response to a specific stimulus. Such a molecular fingerprint could serve as a tool to infer the metabolic state of a cell, as a classification method for disease, or as a reference to compare the similarity between *in vitro* and *in vivo* experimental conditions. For example, tumors can be classified by their expression profile (26). Thus, a disease signature can be detected using microarrays. Our laboratory is in the process of compiling an index database of expression profiles of normal human tissues, which may serve as a fingerprint of tissue phenotype. These data are publicly available at [www.hugeindex.org](http://www.hugeindex.org). Though not yet complete, the database has already yielded important observations, including identification of a set of genes with similar expression levels in all human tissues, so-called maintenance genes, and differential gene expression within different regions of the kidney.

Phenotypes of disease or cellular states can be classified using principal component analysis (PCA) (27). PCA is an analytic method that identifies a subset of genes that are responsible for the majority of observed transcriptional differences and the distinct pattern underlying the differences. This technique aids visualization of multidimensional data by projecting it into a lower dimensional space. In other words, PCA structures a data set using as few variables as possible. Figure 3 shows a *theoretical* classification of pulmonary-renal syndromes using PCA. In this example, disease phenotypes are the variables and gene expression levels are the observations. First, the genes that compose the highest transcriptional variability between phenotypes are identified. The first principal compo-



**Figure 3.** Theoretical classification of pulmonary-renal syndromes using principal component analysis. Each principal component represents a unique set of genes among the entire set of thousands of genes surveyed on a gene chip. The sum of expression levels of the genes that compose this set, or principal component, is represented on each axis. Each point represents an individual patient. First, by plotting the first principal component on the x-axis *versus* the second principal component on the y-axis, the disease phenotypes of Churg-Strauss and microscopic polyangiitis group together and are distinct from one another. Repeating the process using additional principal components and multiple dimensions allows disease phenotypes to be separated and disease fingerprints to be visualized.



**Figure 4.** Paradigm of microarray applications in renal disease.

nent is the combination of gene expression that has the greatest variance among phenotypes. Each subsequent principal component is the combination of gene expression that has the greatest variance *and* is independent of defined components. Three principal components are used to define the observed phenotypes in this example. With the use of this method, transcriptional fingerprints that underlie phenotypic variation can be visualized easily. In addition, evaluation of the components can suggest the underlying factors that are responsible for phenotypic variation.

### Drug Discovery and Drug Target Validation

The introduction of DNA microarrays to the field of pharmacology has created new opportunities for drug discovery. Traditionally, drug discovery was accomplished by first identifying a target molecule within a biologic pathway and then developing an inhibitory compound against the intended target. With the aid of microarray technology, large-scale systematic approaches to drug discovery are possible. Comparing expression of thousands of genes between normal and diseased states can identify multiple potential drug targets without first knowing the biochemical pathway involved. Drug target validation can be accomplished using gene chips. The expression profile of drug-treated cells is screened against a database of deletion mutants to identify the profile that matches that of the drug-treated cells (19,28). Gene chips have important applications in toxicology as well. Secondary drug targets and potential undesirable side effects can be predicted using microarrays. Several excellent reviews have detailed in greater depth the use of microarrays in drug discovery and toxicology (28–31).

### Future Directions

Technical improvements in microarray technology continue to take place. Current areas of development include improving RNA amplification methods to allow analysis of smaller amounts of RNA and, eventually, single-cell expression analysis (8,32). At present, large amounts of RNA routinely are needed (10  $\mu$ g of total RNA or 500 ng of mRNA). Novel signal detection methods are being developed to permit more sensitive analyses (33). A critical need exists in the area of analytical tools for data mining. The use of such tools applied to complex biologic systems is still in its infancy, and many uncertainties still exist. There are no universal standards by which to analyze and compare microarray data, and analytical techniques are still untested. Furthermore, the vast data sets generated have strained most available computer resources. These issues must be addressed before microarray technology can achieve its full potential.

With improved accessibility to this technology and growing understanding of its full capabilities, microarrays will move from the research bench into the clinical arena. In the paradigm illustrated in Figure 4, comparison of a patient's expression profile to compendiums of disease profiles and drug response profiles will aid clinicians in diagnosing disease, identifying prognostic markers, and individualizing therapy. Thus, microarray technology complemented by bioinformatics repre-

sents an exciting new tool for biologic discovery in renal research.

### References

1. Kaplan JM, Kim SH, North KN, Renne H, Correia LA, Tong HQ, Mathis BJ, Rodriguez-Perez JC, Allen PG, Beggs AH, Pollak MR: Mutations in ACTN4, encoding alpha-actinin-4, cause familial focal segmental glomerulosclerosis. *Nat Genet* 24: 251–256, 2000
2. Winn MP, Conlon PJ, Lynn KL, Howell DN, Slotterbeck BD, Smith AH, Graham FL, Bembe M, Quarles LD, Pericak-Vance MA, Vance JM: Linkage of a gene causing familial focal segmental glomerulosclerosis to chromosome 11 and further evidence of genetic heterogeneity. *Genomics* 58: 113–120, 1999
3. Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan JL, Gamble V, Harris PC: The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. *Nat Genet* 10: 151–160, 1995
4. Schena M, Heller RA, Thierault TP, Konrad K, Lachenmeier E, Davis RW: Microarrays: Biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 16: 301–306, 1998
5. Bowtell DD: Options available—from start to finish—for obtaining expression data by microarray. *Nat Genet* 21[1 Suppl]: 25–32, 1999
6. Lennon GG, Lehrach H: Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7: 314–317, 1991
7. Southern EM, Maskos U, Elder JK: Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models. *Genomics* 13: 1008–1017, 1992
8. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14: 1675–1680, 1996
9. DeRisi JL, Iyer VR, Brown PO: Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686, 1997
10. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470, 1995
11. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO: The transcriptional program in the response of human fibroblasts to serum. *Science* 283: 83–87, 1999
12. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW: Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci USA* 94: 2150–2155, 1997
13. Bubendorf L, Kononen J, Koivisto P, Schraml P, Moch H, Gasser TC, Willi N, Mihatsch MJ, Sauter G, Kallioniemi OP: Survey of gene amplifications during prostate cancer progression by high-throughput fluorescence in situ hybridization on tissue microarrays. *Cancer Res* 59: 803–806, 1999 [published erratum appears in *Cancer Res* 59: 1388, 1999]
14. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4: 844–847, 1998
15. Villaret DB, Wang T, Dillon D, Xu J, Sivam D, Cheever MA, Reed SG: Identification of genes overexpressed in head and neck

- squamous cell carcinoma using a combination of complementary DNA subtraction and microarray analysis. *Laryngoscope* 110[3 Pt 1]: 374–381, 2000
16. Whitney LW, Becker KG, Tresser NJ, Caballero-Ramos CI, Munson PJ, Prabhu VV, Trent JM, McFarland HF, Biddison WE: Analysis of gene expression in multiple sclerosis lesions using cDNA microarrays. *Ann Neurol* 46: 425–428, 1999
  17. Han J, Yoo HY, Choi BH, Rho HM: Selective transcriptional regulations in the human liver cell by hepatitis B viral X protein. *Biochem Biophys Res Commun* 272: 525–530, 2000
  18. Hanzel DK, Trojanowski JQ, Johnston RF, Loring JF: High-throughput quantitative histological analysis of Alzheimer's disease pathology using a confocal digital microscanner. *Nat Biotechnol* 17: 53–57, 1999
  19. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburttty K, Simon J, *et al.*: Functional discovery via a compendium of expression profiles. *Cell* 102: 109–126, 2000
  20. Roth FP, Hughes JD, Estep PW, Church GM: Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol* 16: 939–945, 1998
  21. Lockhart DJ, Winzler EA: Genomics, gene expression and DNA arrays. *Nature* 405: 827–836, 2000
  22. Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP: Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* 19: 442–447, 1995
  23. Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS: Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* 14: 441–447, 1996
  24. Yershov G, Barsky V, Belgovskiy A, Kirillov E, Kreindlin E, Ivanov I, Parinov S, Guschin D, Drobishev A, Dubiley S, Mirzabekov A: DNA analysis and diagnostics on oligonucleotide microchips. *Proc Natl Acad Sci USA* 93: 4913–4918, 1996
  25. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP: Accessing genetic information with high-density DNA arrays. *Science* 274: 610–614, 1996
  26. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286: 531–537, 1999
  27. Raychaudhuri S, Stuart JM, Altman RB: Principal components analysis to summarize microarray experiments: Application to sporulation time series. *Pac Symp Biocomput* 12: 455–466, 2000
  28. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, Bassett DE Jr, Hartwell LH, Brown PO, Friend SH: Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 4: 1293–1301, 1998
  29. Debouck C, Goodfellow PN: DNA microarrays in drug discovery and development. *Nat Genet* 21[1 Suppl]: 48–50, 1999
  30. Afshari CA, Nuwaysir EF, Barrett JC: Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res* 59: 4759–4760, 1999
  31. Lennon GG: High-throughput gene expression analysis for drug discovery. *Drug Discov Today* 5: 59–66, 2000
  32. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P: Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA* 89: 3010–3014, 1992
  33. Stears RL, Getts RC, Gullans SR: A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol Genomics* 3: 93–99, 2000