Cell, Vol. 102, 9–15, July 7, 2000, Copyright ©2000 by Cell Press

# Biomedical Discovery with DNA Arrays

Review

**Richard A. Young***
Whitehead Institute for Biomedical Research
Nine Cambridge Center
Cambridge, Massachusetts 02142
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

DNA microarray technologies permit systematic approaches to biological discovery that have begun to have a profound impact on biological research, pharmacology, and medicine. The ability to obtain quantitative information about the complete transcription profile of cells promises to be an exceptionally powerful means to explore basic biology, diagnose disease, facilitate drug development, tailor therapeutics to specific pathologies, and generate databases with information about living processes. The new discipline of expression profiling will link many fields of biology and medicine with a shared knowledge base and a common mathematical language. However, major conceptual and technical challenges lie ahead in this rapidly evolving arena. This review highlights recent discoveries that come from the use of microarray technologies and considers the revolutionary changes and challenges that lie ahead in biology.

## DNA Microarray Technologies

Several different DNA microarray technologies are currently in use, some examples of which are shown in Figure 1. DNA arrays can be synthesized in situ with photochemical techniques or with ink jet technology (Marton et al., 1998; Lipshutz et al., 1999). Alternatively, prefabricated DNA molecules are printed in arrays on glass slides or nylon membranes using robots (DeRisi et al., 1997). Several of these technical platforms have been put to the test with the complete set of yeast genes and have produced expression data of sufficient quality for biological discovery. Microarrays produced by each of these technologies will likely accommodate the entire population of genes carried within the human genome, but improvements in microarray hardware, experimental design, and computational methods of analysis will need to be made before the full potential of expression profiling is obtained in mammalian cells. These issues will be discussed in more detail below.

## Transcriptional Programs

DNA microarrays have been used to estimate the levels of RNA species throughout the RNA population of living cells and to explore transcriptional programs. The population of RNA species in a cell, the transcriptome, has been estimated for yeast under certain growth conditions and, at least in part, for more complex eukaryotes (Holstege et al., 1998; Lockhart and Winzeler, 2000). Transcriptional programs reveal how global gene expression is remodeled during changes in cell growth, physiology, or environment. For example, the transcriptional program of the yeast cell cycle has been described in detail (Cho et al., 1998; Spellman et al., 1998). Similarly, a portion of the transcriptional program has been established for *Drosophila* development during metamorphosis (White et al., 1999). These expression profiles provide a rich source of information for identifying genes that may be coregulated and for modeling regulatory mechanisms and networks.

DNA array studies have also been designed to identify genes whose expression depends on a cell state or the functions of specific components of signaling or transcription apparatuses. In many cases, the nature of these genes has led to new insights into biological processes. For example, studies designed to identify genes whose expression depends on key components of the transcription initiation machinery in yeast led to the insight that the general transcription apparatus is physically remodeled when cells encounter limiting nutrients (Holstege et al., 1998). The set of genes regulated by genome ploidy in yeast revealed how ploidy can influence cell cycle progression and explained why greater cell size is associated with higher ploidies (Galitski et al., 1999). Specific genes associated with age-related phenotypes and diseases in humans have been identified, implicating mitotic misregulation in aging (Ly et al., 2000).

The value of such expression profiles has yet to be fully realized, even by the investigators who produced and published them. This limitation is due to the paucity of computational tools necessary to analyze large datasets, inadequate experience with modeling biological problems and systems using large amounts of data, and the lack of methods for systematic experimental examination of such models. Nonetheless, it is clear that expression profiles are remarkably robust signatures of biological conditions. These signatures are a precise molecular phenotype of the cell in a specific state. The importance and utility of these signatures is illustrated by recent microarray-based studies into gene function in yeast and classification of diseases in humans.

## Gene Function and Drug Discovery

Two issues of interest to biologists who wish to exploit new genome sequence information have been examined in a paper in this issue of *Cell* (Hughes et al., 2000). The first issue addressed by this work is whether expression profiles of mutant cells can be used to accurately classify the functions of previously uncharacterized genes. The second issue explored by these investigators is whether the expression profile of cells treated with a pharmacological agent can be used to identify the target of that drug. The evidence indicates that the function of a gene can be accurately predicted from the expression profile of a cell with a mutation in that gene. The results also provide an important proof-of-principle for new approaches to pharmacological research and development.

*E-mail: young@wi.mit.edu

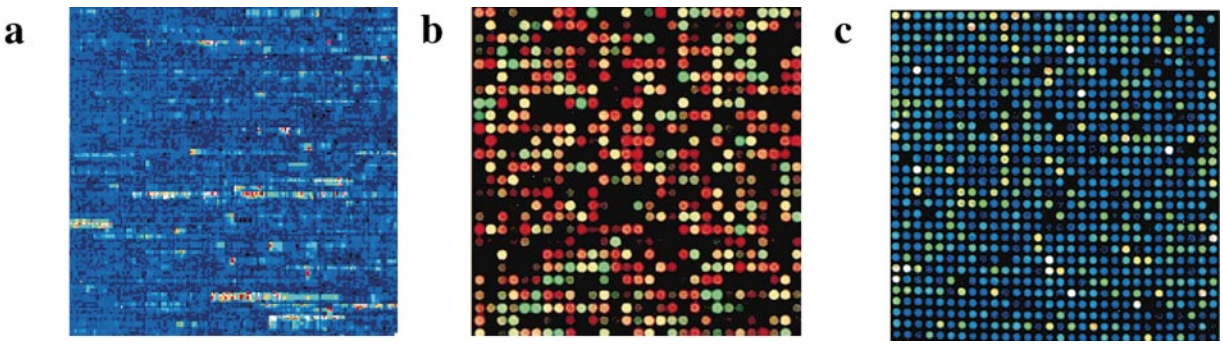



Figure 1. Selected DNA Microarrays

(a) Oligonucleotide array synthesized in situ with photochemical technology by Affymetrix. (b) Oligonucleotide array synthesized in situ with ink-jet technology (Image courtesy of Rosetta Inpharmatics). (c) DNA microarray printed on a glass slide (Image courtesy of Corning, Inc).

The recent acquisition of genome sequences from substantial numbers of organisms (see http://www.ncbi.nlm.nih.gov/Genomes/index.html) has led to considerable interest in methods that provide clues to the functions of newly discovered genes. Functional annotation of genes has been approached in several ways. Gene function can be inferred from sequence homology among predicted proteins, where the function of at least one protein is known (Chervitz et al., 1998). Creative computational methods have been developed that detect functional linkages among predicted proteins (Eisenberg et al., 2000). However, sequence analysis alone is insufficient to fully inform us about gene function. The microarray-based approach to this problem has been to cluster genes according to their expression behavior under a range of conditions and to assign function according to the set of known genes that fall into these clusters (Eisen et al., 1998; Tamayo et al., 1999). The premise with this guilt-by-association approach is that clustered genes may be coregulated and therefore be involved in similar functions.

The results described by Stephen Friend and colleagues (Hughes et al., 2000) indicate that gene function can be more accurately predicted from the expression profile of a cell with a mutation in the gene of interest. These investigators developed a reference database of yeast gene expression profiles for 300 different mutations and chemical treatments. Since most of these mutations and treatments affect well-understood cellular functions, the expression profiles could be assigned to functional groups. The cellular pathways affected by a mutation in an uncharacterized gene could be ascertained by using pattern matching to compare the expression profile of cells harboring the mutation with those in the reference database. They identified and experimentally confirmed functions for genes whose products are involved in sterol metabolism, cell wall function, mitochondrial respiration, and protein synthesis. This approach has the advantage that genes exhibiting little transcriptional regulation can still be functionally annotated if their mutation produces an expression signature that resembles signatures obtained with well-characterized genes. The power of the method is such that it should lead to a more rigorous classification scheme than those currently employed to annotate gene functions in public databases.

Since the method works to classify genetic perturbations, it should also be effective at identifying the functional consequences of chemical perturbations. Indeed, the expression profiles obtained for drug-treated cells are similar to profiles obtained when genes encoding the drug targets are mutated. The utility of this approach to pharmaceutical research and development was demonstrated by identifying a previously unknown target of the commonly used topical anesthetic dyclonine.

It is possible to envision this new approach to functional annotation being extended to more complex organisms. Mutagenesis screens in *Drosophila*, zebrafish, and mouse, and the identification of large numbers of rare single-gene disorders in man can contribute to functional annotation by this method. Chemical agents should be powerful probes to examine the functions of specific proteins and pathways in mammalian cells. Given the substantial number of bioactive small-molecule compounds currently available or under development for specific therapies, it is possible to imagine that "chemical genomics" will contribute significantly to our future understanding of the functions of some gene products.

Drug discovery in the last half century typically began with a biochemical pathway implicated in a pathophysiological process. With sufficient information about the enzymes that are critical for a functional pathway, biochemists and medicinal chemists collaborated to identify and optimize the therapeutic behavior of small molecules that bind to specific target enzymes. Clinical trials then identified the small molecules that have therapeutic efficacy in human beings. This process now costs $50–$500 million for each new drug brought to market. DNA microarray methods promise to make most steps in this process more precise and efficient. Indeed, microarray analysis is already being integrated into many steps in drug development, including target identification, target validation (demonstrating that affecting the enzyme or biochemical entity has therapeutic utility), optimizing efficacy and reducing toxicity (Debouck and Metcalf, 2000; Roses, 2000). DNA microarray analysis may also facilitate identification of clinical trial participants who best respond or adversely react to specific therapeutics.

**Disease Diagnosis, Prognosis, and Therapy**

Cancer and infectious diseases are major challenges to the health of human populations, yet the techniques

used to diagnose these diseases have changed little in decades. New approaches have emerged from the collaborative efforts of biologists, physicians, mathematicians, and computer scientists involved in DNA microarray–based research into these diseases. More accurate disease diagnosis and improvements in therapy will be the clinical benefit.

Current cancer classification techniques rely on highly subjective judgments of tumor histology by pathologists. Multiple studies have demonstrated that DNA microarrays are useful for classifying human cancers and have revealed that expression profiles are valuable both in cancer diagnosis and prognosis (Golub et al., 1999; Alizadeh et al., 2000). This global quantitative approach to classifying cancers and predicting outcomes will become a valuable clinical tool for pathologists and oncologists. DNA microarray analysis can allow physicians to follow the progression of disease even when there is no histological evidence of change. The ability to distinguish morphologically similar human cancers by the differences between their expression profiles is important because accurate identification of tumor types will facilitate matching the appropriate therapies, thereby maximizing therapeutic efficacy and minimizing toxicity. The information obtained through cancer classification will almost certainly contain valuable clues to cancer mechanisms and inspire new tactics to combat these diseases.

Infectious diseases are the major causes of global human morbidity and mortality. Indeed, infection has played a particularly important role in evolution and may be the most significant factor affecting the relative success of human populations in history (McNeill, 1976; Oldstone, 1998). The human population is continuously threatened with newly emerging and increasingly drug-resistant pathogens, many of which are difficult to culture and thus identify. Progress in understanding infectious diseases and developing new therapies has been limited by the time-consuming approaches that have traditionally been used to identify infectious agents, to discover their mechanisms of pathogenesis, and to explore potential therapeutic approaches. PCR has led to useful molecular diagnostic tools to identify new pathogens (Relman et al., 1991) and to confirm diagnoses (Rowley et al., 1990). The use of microarrays in infectious disease research and classification, however, is likely to produce revolutionary changes in this arena. It should be possible to obtain signatures for pathogens that are diagnostic, even when the etiologic agent is not known or cannot be easily cultured. To obtain signatures for specific pathogens, it will be necessary to capture their effects on host cell gene expression. The best approach to analyzing patient material remains to be identified, but expression profiles of human macrophages infected by various viral and bacterial pathogens in the author's lab and elsewhere suggest the feasibility of this approach (Manger and Relman, 2000). Information obtained through pathogen classification by using signatures from target cell populations will almost certainly contain valuable clues to pathogenesis and potential new approaches to combat these diseases.

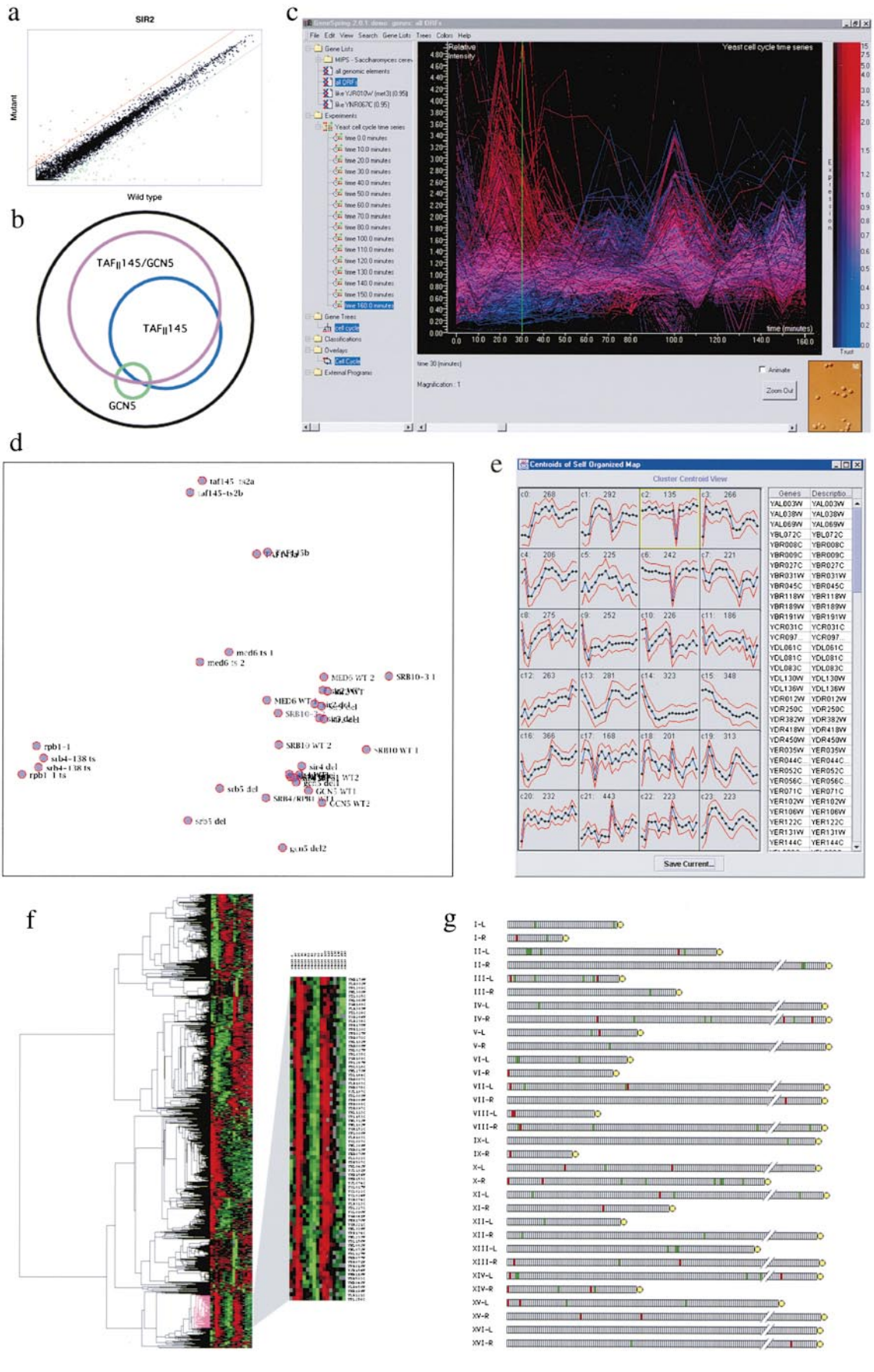
The approaches that are being used to classify cancer and other diseases with DNA microarray expression profiles underscore the importance of the collaboration between biology and computer science (Kahn et al., 1998; Alon et al., 1999; Golub et al., 1999; Perou et al., 1999; Alizadeh et al., 2000; Ross et al., 2000). There is a rich history of literature on classification algorithms that can be used for predicting qualitative properties from DNA microarray data (Bishop, 1995). Such qualitative properties can include diagnostic and prognostic indications. Classifiers require a training set of observations with known classifications to establish appropriate models and/or decision boundaries. The accuracy of a classifier can be established with a validation set of observations with known classifications, and the classifier can then be applied to a set of observations with unknown classifications. Clearly, the depth of our understanding of living processes stands to gain by the marriage of biology and computation.

**Complex Biological Systems**

Deciphering the rules that govern complex systems is the greatest challenge to scientists. Complex systems in biology include the transcriptional regulatory networks of living cells, the patterns of cell division and death that lead to development of multicellular organisms, the communication networks among the 100 billion cells of the human brain, and the interactions between the vast numbers of living organisms that contribute to the ecology of the Earth's environment. Just as Linnaean classification of living species contributed the groundwork necessary for the theory of natural selection, classification of living processes based on expression profiles should contribute to profound improvements in our understanding of these processes.

The elucidation of genome sequences and the development of technology for monitoring global gene expression provide an opportunity to obtain information on complex systems in cells and to initiate entirely new fields of biological investigation. Reductionist approaches, first introduced by Rene Descartes in the mid-1600s and clearly among the most powerful intellectual instruments in modern scientific discovery, have helped biologists develop a fundamental understanding of an impressive array of basic features of cell and organismal biology. The powerful combination of genetic and biochemical approaches has served biologists well and will continue to do so. The coupling of systems analysis and reductionist methods should prove to be even more powerful. The models and hypotheses that emerge from analysis of systemic expression information can be tested with genetic and biochemical methods.

Hartwell, Hopfield, Leibler, and Murray have argued for recognition of functional modules as fundamental elements of biological organization and regulation because most biological functions arise from interactions among many components (Hartwell et al., 1999). A functional module, for example, might be the signal transduction system that converts the binding of pheromone at the yeast cell surface into the act of mating. Understanding the structure and function of modules requires more than analyzing properties of isolated components. The collection of precise information on global gene expression as cells go about their business will be critical to understanding functional modules. The product


a
SIR2
Mutant
Wild type
b
TAFII145/GCN5
TAFII145
GCN5
c
Yeast cell cycle time series
time (minutes)
d
e
Centroids of Self Organized Map
Cluster Centroid View
Save Current...
f
g

of this and other approaches should reveal general principles governing the structure and behavior of modules.

Another new arena involves elucidating the regulatory networks that govern gene expression. Just as knowledge of biochemical pathways has been fundamental to basic biology and to drug discovery, knowledge of the regulatory networks that govern gene expression will facilitate modeling of biological processes and efforts to develop therapies for disorders and diseases. The regulatory switch that governs the lysis versus lysogeny decision of bacteriophage lambda has been modeled (Arkin et al., 1998), and biologists and computer scientists are now exploring how to use genome-wide expression profiles to deduce models for the regulatory networks and circuitry of small genomes (Tavazoie et al., 1999). The principled comparison of competing models will be of critical importance as the complexity of such models grows to encompass tens or even hundreds of genes. We will need new tools that can provide probabilistic assessments of how well different models of complex genetic regulatory circuitry explain observed data.

Some will argue that although global gene expression information is valuable, it does not provide a complete picture of the functions of a cell since proteins are the molecules that ultimately regulate and carry out cellular functions. Indeed, experimental tools that permit efficient systematic assessment of protein levels, modifications, and intracellular locations will need to be developed before the world of functional modules and other aspects of biology can be fully discerned (Pandey and Mann, 2000). However, two features of expression profiling make it the most productive approach to study biological systems for the immediate future. First, the present efficiency with which investigators can obtain global and quantitative information with DNA arrays exceeds that of proteomic techniques. Second, RNA expression profiles provide an extremely precise and reproducible signature of the state of the cell that probably reflects, albeit indirectly, the functional state of all proteins.

## Databases with Collective Knowledge

The development of databases containing expression profiles and other information on living processes will be a valuable asset for scientists involved in basic biological discovery, disease diagnostics, drug development and the new fields of research that will develop to explore such information. The biomedical community is already working to establish standards and expression database management systems to permit data sharing by scientists worldwide (Brazma et al., 2000). Shared databases will facilitate communication among scientists and physicians from different disciplines who have in the past used different experimental approaches, analysis, and language (i.e., jargon). A more universally understood language will evolve to discuss experiments, analysis, and interpretation of expression information.

Expression profiles will provide the basis for understanding how genes work together to guide the functions of cells and organisms. Deciphering these regulatory networks is attracting talent from many disciplines, including biology, mathematics, physics, engineering, and computer science. The task of deciphering regulatory networks will be substantial and will require shared databases that contain our partial knowledge about regulatory networks in a standard framework. Such shared databases will be essential for the collaborations necessary to elucidate regulatory networks and will be valuable for disseminating information to other investigators. Homology-based reasoning has been a powerful force in molecular biology to date. Capturing genetic pathways in computational form will extend this learning device and will likely have a profound impact on our understanding of how organisms and cells function because it will allow us to recognize conserved network structures within and across species.

The emergence of a world in which DNA microarrays allow scientists and physicians to explore and solve biological and medical problems with databases containing shared knowledge faces some formidable problems. If the powerful combination of expression profiles and mathematics is to be fully deployed to model biological behavior, then very accurate information must be available. The chief challenges for a database of shared information involve microarray quality and expense, experimental design, and information handling and analysis.

## Technical and Conceptual Challenges

Microarray hardware with sensitive and highly reproducible performance characteristics is necessary for the development of valuable databases, particularly those with collective knowledge on living processes. This will be especially challenging with the large number of genes that exist in mammalian genomes. DNA microarray technologies are already providing valuable expression information to biologists, but there is significant room for improvements. It is not yet clear how the various microarray technologies will compare when the complete human gene set is arrayed and expression profiles are obtained for a variety of different tissues. A direct

Figure 2. Computational Approaches to Classifying and Displaying Patterns in Expression Profile Data

Selected display tools illustrative of current approaches to visualizing expression data are shown. (a) A scatter plot identifies outlying genes whose expression differs meaningfully between the wild-type yeast and a *SIR2* mutant strain (Wyrick et al., 1999). (b) A Venn diagram showing the relationships among genes affected by mutations in two different yeast transcription factors ($TAF_{II}145$ and GCN5) when each of the mutations is present in separate cells or when both mutations are present in cells. The results indicate that many genes are affected in the double mutant that are unaffected by either single mutant, suggesting functional redundancy between the two transcription factors (Lee et al., 2000). (c) A screen shot of GeneSpring output (Silicon Genetics) showing expression patterns of all yeast genes over time during a cell cycle experiment. (d) Multidimensional scaling can illustrate the relative similarity between the expression signatures of different mutant yeast strains. (e) The GeneCluster program uses self-organizing maps (SOM) to group genes into coexpressed clusters (Tamayo et al., 1999). The SOM algorithm causes related clusters to be placed into adjacent positions in the grid. (f) Hierarchical cluster analysis of the yeast cell cycle using a program developed by Eisen et al. (1998). The highlighted cluster shows genes whose expression patterns follow the cell cycle. (g) A chromosome display of the *SIR2* mutant shows the distribution of changes in gene expression across the genome (Wyrick et al., 1999).

side-by-side comparison of the attributes of these different arrays has yet to be performed. Dynamic range, cross-hybridization, reproducibility, and cost must be assessed. As the size of the genome and amount of redundancy increase, aspects of array design that provide the most accurate signal for each gene, such as the optimal oligonucleotide length needed for high signal-to-noise ratios with minimal cross-hybridization, will have to be identified. It remains to be seen if DNA microarrays produced in academic settings can provide the sensitivity, dynamic range, and reproducibility that will be obtained with commercially produced arrays. DNA array expense will drop with large-scale production and commercial competition, and the anticipated increases in productivity and value due to these tools will need to be evident to the financial sponsors of biomedical research.

The design and implementation of microarray experiments demand special attention. Experimental designs that maximally exploit the full power of genome expression profiles need to be employed. Such designs need to exercise all of the components of a pathway to ensure that incorrect models are not supported by experimental data. Experimental manipulations need to be rigorously controlled, since a small difference in microenvironment can have a large effect on the expression profile. The experimental controls should include the DNA elements necessary to ascertain whether the behavior of a selected and relevant set of genes is consistent with information obtained from other experiments. For drug discovery and diagnostic applications, there is the challenge of designing efficient high-throughput microarray hardware, methods, and analytical processes.

Yeast cells are easier to study than human cells because all yeast genes are known and annotated with universally agreed-upon names, the yeast mRNA population is less complex, and yeast are less sensitive to the absence of their neighbors than are metazoan cells. Moreover, obtaining access to adequate quantities of normal and diseased human cells is a major challenge. Most DNA microarray experiments use RNA harvested from $10^6$–$10^7$ mammalian cells, yet many interesting and important biological problems can be studied only with small numbers of cells. Laser capture microdissection techniques have been developed that permit single cell capture (Luo et al., 1999), but the impact of this method on expression profiles is not yet clear. Amplification methods that permit analysis of small numbers of cells are under development, but the effect of these amplification processes on information quality remains to be fully established (Lockhart and Winzeler, 2000). Continued improvements in hardware and biological sampling methods will be needed to permit accurate and reproducible expression analysis using relatively small numbers of cells.

Analysis and interpretation of expression data benefit from comparison of many expression profiles from the same organism, but the experimental standards, expression database management systems, and computational tools that facilitate comparisons of large datasets are still under development. The computational tools available to most investigators consist of algorithms that cluster genes according to shared expression patterns and that display expression information using simple color schemes (Figure 2). Much more sophisticated approaches are needed to assist scientists in the examination and interpretation of expression profile datasets. Perhaps most importantly, it will be valuable to search for a body of mathematics that will serve as a natural language for gene expression information, one that parallels the mathematics that works so well to describe physical processes.

**Acknowledgments**

I am grateful to Gerry Fink, David Gifford, Alex Hartemink, Mark Johnston, Seema Kumar, Jerry Nau, Nicola Rinaldi, Phil Sharp, Robert Weinberg, and Peter Young for comments that improved this manuscript and Tom Volkert and Peter Young for assistance with figures.

**References**

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., and Levine, A.J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA *96*, 6745–6750.

Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. Genetics *149*, 1633–1648.

Bishop, C.M. (1995). Neural Networks for Pattern Recognition (Oxford: Oxford University Press).

Brazma, A., Robinson, A., Cameron, G., and Ashburner, M. (2000). One-stop shop for microarray data. Nature *403*, 699–700.

Chervitz, S.A., Aravind, L., Sherlock, G., Ball, C.A., Koonin, E.V., Dwight, S.S., Harris, M.A., Dolinski, K., Mohr, S., Smith, T., et al. (1998). Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science *282*, 2022–2028.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell *2*, 65–73.

Debouck, C., and Metcalf, B. (2000). The impact of genomics on drug discovery. Annu. Rev. Pharmacol. Toxicol. *40*, 193–207.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein. D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Eisenberg, D., Marcotte, E.M., Xenarios, I., and Yeates, T.I. (2000). Protein function in the post-genomic era. Nature *405*, 823–826.

Galitski, T., Saldanha, A.J., Styles, C.A., Lander, E.S., and Fink, G.R. (1999). Ploidy regulation of gene expression. Science *285*, 251–254.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531–537.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. (1999). From molecular to modular cell biology. Nature *402*, C47–C52.

Holstege, F.C.P., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. Functional discovery via a compendium of expression profiles. (2000). Cell *102*, this issue, 109–126.

Kahn, J., Simon, R., Bittner, M., Chen, Y., Leighton, S.B., Pohida, T., Smith, P.D., Jiang, Y., Gooden, G.C., Trent, J.M., and Meltzer, P.S. (1998). Gene expression profiling of Alveolar Rhabdomyosarcoma with cDNA microarrays. Cancer Res. *58*, 5009–5013.

Lee, T., Causton, H.C., Holstege, F.C.P., Shen, W.-C., Hannett, N., Jennings, E.G., Winston, F., Green, M.R., and Young, R.A. (2000). Redundant roles for SAGA and TFIID in global transcription. Nature *405*, 701–704.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. Nat. Genet. *21*, 20–24.

Lockhart, D.J., Winzeler, E.A. (2000). Genomics, gene expression and DNA arrays. Nature *405*, 827–836.

Luo, L., Salunga, R.C., Guo, H., Bittner, A., Joy, K.C., Galindo, J.E., Xiao, H., Rogers, K.E., Wan, J.S., Jackson, M.R., and Erlander, M.G. (1999). Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat. Med. *5*, 117–122.

Ly, D.H., Lockhart, D.J., Lerner, R.A., and Schultz, P.G. (2000). Mitotic misregulation and human aging. Science *287*, 2486–2492.

Manger, I.D., and Relman, D.A. (2000). How the host 'sees' pathogens: global gene expression responses to infection. Curr. Opin. Immunol. *12*, 215–218.

Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., et al. (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat. Med. *4*, 1293–1301.

McNeill, W.H. (1976). Plagues and Peoples (New York: Doubleday).

Oldstone, M. (1998). Viruses, Plagues, and History (New York: Oxford University Press).

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. Nature *405*, 837–846.

Perou, C.M, Jeffrey, S.S., van de Run, M., Rees, C.A., Eisen, M.B, Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C.F., et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers, Proc. Natl. Acad. Sci. USA *96*, 9212–9217.

Relman, D.A., Falkow, S., LeBoit, P.E., Perkocha, L.A., Min, K.W., Welch, D.F., and Slater, L.N. (1991). The organism causing bacillary angiomatosis, peliosis hepatis, and fever and bacteremia in immunocompromised patients. N. Engl. J. Med. *324*, 1514.

Roses, A.D. (2000). Pharmacogenetics and the practice of medicine. Nature *405*, 857–865.

Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet. *24*, 227–235.

Rowley, A.H., Whitley, R.J., Lakeman, F.D., and Wolinsky, S.M. (1990). Rapid detection of herpes-simplex-virus DNA in cerebrospinal fluid of patients with herpes simplex encephalitis. Lancet *335*, 440–441.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell *9*, 3273–3297.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. (1999). Interpreting patterns of gene expression with self-organzing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA *96*, 2907–2912.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999) Systematic determination of genetic network architecture. Nat. Genet. *22*, 281–285.

White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of *Drosophila* development during metamorphosis. Science *286*, 2179–2184.

Wyrick, J.J., Holstege, F.C.P., Jennings, E.G., Causton, H., Shore, D., Grunstein, M., Lander, E.S., and Young, R.A. (1999) Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature *402*, 418–421.