Monitoring gene expression using DNA microarrays

Christina A Harrington*, Carsten Rosenow and Jacques Retief

The concurrent development of high-density array technologies and the complete sequencing of a number of microbial genomes is providing the opportunity to comprehensively and efficiently survey the transcription profile of microorganisms under different conditions and well-defined genotypes. Microarray-based studies are uncovering broad patterns of genetic activity, providing new understanding of gene functions and, in some cases, generating unexpected insight into transcriptional processes and biological mechanisms. One topic that has come to the forefront is how best to effectively manage and interpret the large data sets being generated. Although progress has been made, this remains a challenging opportunity for functional genomics research.

Addresses
Affymetrix, Inc., 3380 Central Expressway, Santa Clara, California 95051, USA
*e-mail: chris_harrington@affymetrix.com

Current Opinion in Microbiology 2000, **3**:285–291

1369-5274/00/$ – see front matter

Abbreviations

HCMV	human cytomegalovirus
MM	mismatch
ORF	open reading frame
PM	perfect match

Introduction

During the last half of the 20th century, the analysis of the regulation and function of genes has largely been driven by step-by-step studies of individual genes and proteins. In the past decade, a paradigm shift has emerged in which we are now able to produce large amounts of data about many genes in a highly parallel and rapidly serialized manner. An important tool in this process has been the development of DNA microarrays. These arrays consist of a highly ordered matrix of thousands of different DNA sequences that can be used to measure DNA and RNA variation in applications that include gene expression profiling, comparative genomics and genotyping [1–3,4••]. In this review we present an overview of DNA microarray technology and its application to the study of gene expression in microbial systems. We also discuss the looming question of how best to manage, share and interpret the large data sets generated by microarray experiments.

DNA microarray technology overview

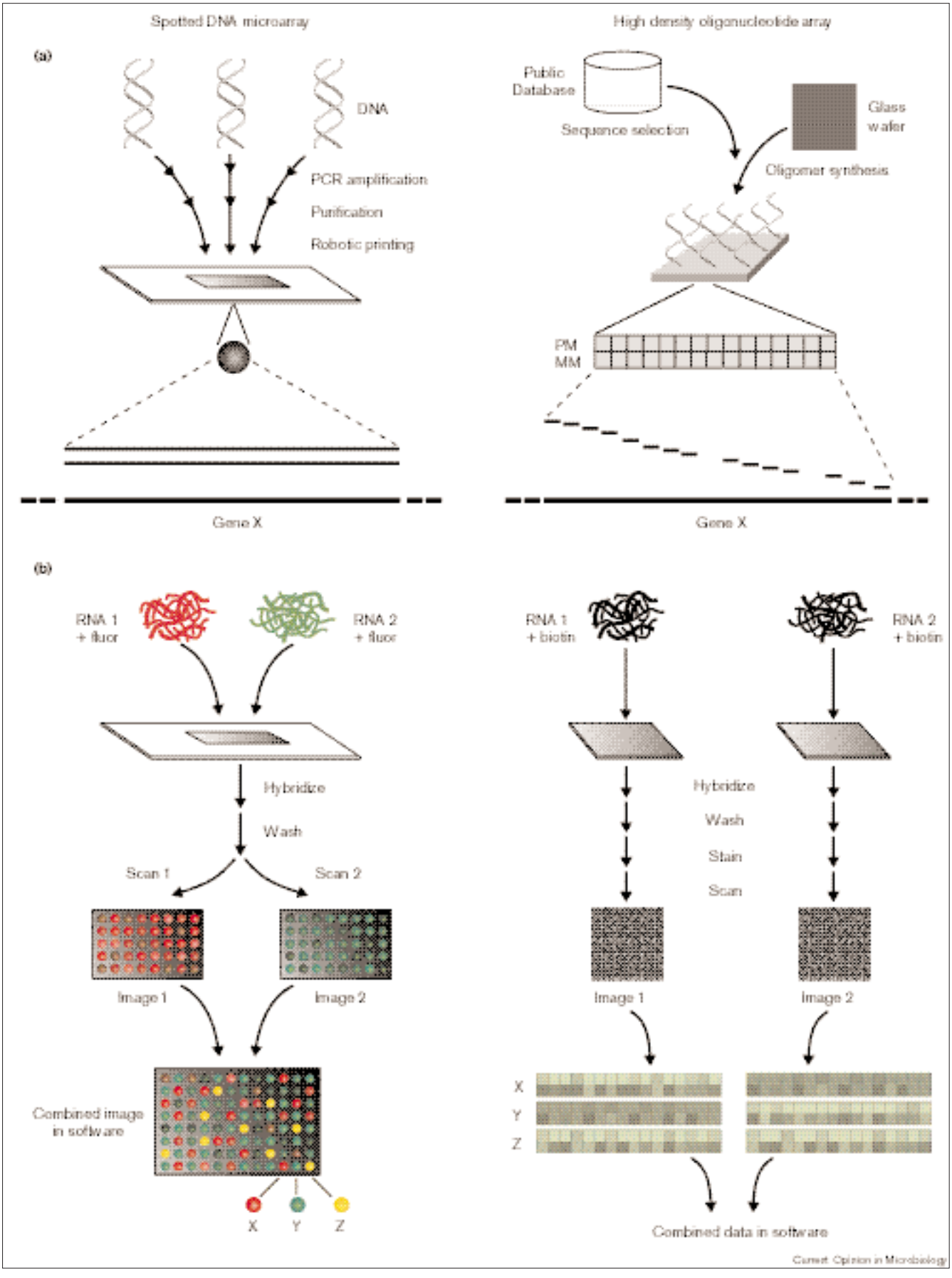
DNA microarrays provide a format for the simultaneous measurement of the expression level of thousands of genes in a single hybridization assay. Each array consists of a reproducible pattern of thousands of different DNAs (primarily PCR products or oligonucleotides) attached to a solid support, usually glass. Fluorescently labeled RNA or DNA prepared from messenger RNA is hybridized to complementary DNA on the array and then detected by laser scanning. Hybridization intensities for each DNA sequence on the array are determined using an automated process and converted to a quantitative read-out of relative gene expression levels. The data can then be further analyzed to identify expression patterns and variation that correlate with cellular development, physiology and function.

Several methods have been described for producing microarrays and are comprehensively reviewed elsewhere [5•,6–9]. In this review, two basic types of array technology have been surveyed: spotted microarrays in which pre-synthesized single-strand or double-strand DNAs are bound, or 'printed', onto glass slides [10,11] and high-density oligonucleotide arrays (sometimes referred to as 'oligo chips') in which sets of oligomers are synthesized *in situ* on glass wafers using a photolithographic manufacturing process [12,13]. Spotted microarrays can be produced in-house or accessed through commercial service providers (see [4••] for a list of web sites with information on making and using DNA microarrays). High-density oligonucleotide arrays are available commercially. Although comprehensive expression profiles can be produced by both types of array, there are some fundamental differences between the two approaches (Figure 1).

On spotted arrays, genes are generally represented by single DNA fragments, greater than several hundred base pairs in length. The DNA samples hybridized to the array are, in most cases, labeled by incorporating fluorescently tagged nucleotides during oligo-primed reverse transcription of messenger RNA [7]. Different fluorophores (generally Cy3- and Cy5-dUTP) are used to label cDNAs from control (reference) and experimental (test) RNAs. The labeled cDNAs are then mixed together prior to hybridization to the array. Relative amounts of a particular gene transcript in the two samples are determined by measuring the signal intensities detected for both fluorophores and calculating signal ratios.

On GeneChip® oligonucleotide arrays (Affymetrix, Inc, Santa Clara, CA) a given gene is currently represented by 15–20 different 25-mer oligonucleotides that serve as unique, sequence-specific detectors. An additional control element on these arrays is the use of mismatch (MM) control oligonucleotides that are identical to their perfect match (PM) partners except for a single base difference in a central position. The presence of the mismatched oligonucleotide allows cross-hybridization and local background to be estimated and subtracted from the PM signal. In the GeneChip expression assay eukaryotic mRNA is converted to biotinylated cRNA from oligo-dT-primed cDNA [13]. (Options for labeling prokaryotic mRNAs are discussed below.) Each sample is hybridized to a separate

Figure 1


Spotted DNA microarray
High density oligonucleotide array
(a)
DNA
PCR amplification
Purification
Robotic printing
Public Database
Sequence selection
Glass wafer
Oligomer synthesis
PM
MM
Gene X
Gene X
(b)
RNA 1 + fluor
RNA 2 + fluor
Hybridize
Wash
Scan 1
Scan 2
Image 1
Image 2
Combined image in software
X
Y
Z
RNA 1 + biotin
RNA 2 + biotin
Hybridize
Wash
Stain
Scan
Image 1
Image 2
X
Y
Z
Combined data in software
Current Opinion in Microbiology

Legend for Figure 1

Diagram of array preparation and expression assay for spotted DNA microarrays and high-density oligonucleotide arrays. **(a)** Array fabrication and gene representation. Spotted DNA arrays are made by printing amplified DNAs prepared from genomic or cDNA onto glass slides. Each spot on the glass slide corresponds to a contiguous gene fragment of several hundred bases or more. Pre-synthesized oligomers can also be spotted (data not shown). High-density oligonucleotide arrays are manufactured using a process of light-directed combinatorial chemical synthesis to produce thousands of different oligomer probes in a highly ordered array on a small glass chip. Genes are represented by 15–20 different oligomer pairs (PM, perfectly matched and MM, mismatched) on the array. **(b)** Expression assay. In the expression assay for spotted arrays, messenger RNAs from samples 1 (test) and 2 (reference) are either directly or indirectly labeled with different fluorophores, hybridized to DNA on the glass slide and subsequently scanned to independently detect both fluorophores. Colored dots labeled x, y and z at the bottom of the image correspond to hypothetical genes present at increased levels in sample 1 (x), increased levels in sample 2 (y), and similar levels in samples 1 and 2 (z). In the GeneChip® expression assay RNA is labeled in a two-step linear amplification process to produce biotinylated cRNA. After hybridization, biotin-cRNA bound to the array is stained with a fluorophore conjugated to avidin and detected by laser scanning. Sets of paired oligonucleotides for hypothetical genes present at increased levels in sample 1 (x), increased levels in sample 2 (y) and similar levels in samples 1 and 2 (z) are shown at the bottom of the diagram.

array. Transcript levels are calculated by reference to cRNA spikes of known concentration added to the hybridization mixture. Differences in mRNA levels between samples are determined by comparison of any two hybridization patterns produced on separate arrays of the same array type.

There are important differences in performance and information generated by the two types of arrays due to these differences in assay and gene representation. First, spotted arrays hybridized simultaneously with two distinctly labeled samples intrinsically normalize for noise and background in a pairwise comparison. The transcriptional read-out for these paired samples is provided as expression ratios and requires that different samples of an experimental set be hybridized with the same control or reference sample. The high-density oligonucleotide array assay, on the other hand, allows flexibility in sample comparisons and provides an estimate of the levels of gene transcripts in individual samples. Second, on oligonucleotide arrays the oligomer probes are designed to uniquely represent the cognate gene thus minimizing cross-hybridization between similar sequences. The potential for cross-hybridization between genes with significant levels of sequence similarity is high using spotted microarrays. In *Escherichia coli*, for example, out of a total of 4288 open reading frames (ORFs), 556 ORFs contain regions of >200 bp that share a minimum of 50% identity to at least one other ORF in the genome [14•]. Finally oligonucleotide arrays require gene sequence information for specifying the de novo synthesis of the oligomers on the array, whereas spotted arrays can be produced from both known and unknown cDNA and PCR fragments.

Application of microarrays to studies of microorganisms

The relatively small genomes of prokaryotes and simple eukaryotes such as yeast are well suited to global expression profiling with microarray technologies currently available. All ORFs and even intergenic regions can be represented, allowing unbiased assessment of an organism's expression profile. When array experiments are coupled with well-controlled experimental systems and well-characterized genotypes, a striking amount of highly interpretable information can be generated. The information can be used to assign function to unknown genes, enlarge our understanding of cellular processes, identify potential drug targets and generate genome-wide snapshots of transcriptional activity in response to any stimulus or developmental trigger.

Yeast

The availability of the complete genome sequence and well-characterized genetics has made *Saccharomyces cerevisiae* a popular focus of study with both spotted and oligonucleotide arrays. Work describing microarray analysis of gene expression across the entire yeast genome was first reported in 1997 [11,15,16]. Spotted arrays containing comprehensive sets of PCR-amplified ORFs were used to measure transcript changes accompanying the metabolic shift from fermentation to respiration [11] and changes induced by a variety of culture manipulations, including heat- and cold-shock [15]. Differences in transcription profiles for yeast grown in rich versus minimal media were determined using oligonucleotide arrays representing more than 6200 yeast genes [16]. A relatively small number of genes with dramatically different expression levels (approximately 3–5% of the ORFs surveyed demonstrated greater than four-fold changes in expression levels) were identified in these studies, while the majority of highly expressed genes were present at similar levels under the different growth conditions.

Genome-wide expression profiles have been generated for cell cycle progression [17•,18•] and sporulation in budding yeast [19•]. These studies have been a rich source of new information leading to the identification of groups of genes co-regulated during the cell cycle or sporulation. Potential functions of many previously uncharacterized genes were suggested by their display of expression patterns similar to that of known genes or during functionally defined cell states. (Note: complete data sets from many of the studies cited here are available on the World Wide Web; see for example [17•–19•]).

In a study of cellular response to DNA damage, the yeast GeneChip array was used to examine transcriptional

responses to alkylating agents in *S. cerevisiae* [20•]. The catalogue of genes whose expression is induced by DNA damage was shown to increase by more than 15 fold. Northern blots confirmed the array results for 48 out of 50 ORFs identified as either responsive (42) or nonresponsive (8) to the alkylating agent. In other studies, several cellular functions have been analyzed by combining transcript profiling with genetics: transcription initiation [21]; nonsense-mediated mRNA decay [22]; and the mitogen-activated protein kinase cascade [23]. Use of microarrays to examine drug signature patterns and identify effects mediated through pathways independent of the putative drug target have been described [24,25]. Application of this kind of analysis to drug development programs should improve the odds of developing useful therapeutics [26].

Another example of ways in which expression profiling with yeast microarrays has been applied to date is the work by Ferea *et al.* [27] who used spotted arrays to look at adaptive evolution under natural selection. In this interesting study, the authors were able to identify similar alterations in expression in three independently evolved strains, which is consistent with proposed mechanisms for increased fitness under the selection regime.

Bacteria

One of the big challenges in prokaryotic expression analysis is the specific labeling of mRNA for microarray hybridization [28]. Unlike eukaryotic labeling strategies that rely on the presence of a polyA tail to enrich for mRNA, a reliable method to enrich or specifically label prokaryotic mRNA has not been available. In the first published report on bacterial gene expression monitoring using high density oligonucleotide arrays, de Saizieu *et al.* [29] showed that the presence of rRNA does not prevent detection of gene transcripts present down to a level of several copies per cell. In this study, a direct RNA labeling protocol was described that introduces label by photo-coupling with psoralen-biotin. Subsequent studies from Blattner and coworkers [14•] compared radioactive and fluorescent labeled cDNAs prepared from total *E. coli* RNA that were hybridized to arrays spotted on membranes and glass. On the basis of their results, they conclude that the microarray approach with fluorescent-labeled cDNA is more reproducible than radioactive hybridization to DNA dot blots.

The power of genome-wide bacterial expression analysis was demonstrated by Tao *et al.* [30•]. In this study the expression levels of 4290 genes representing the complete *E. coli* genome were determined with nylon macroarrays for bacteria grown in rich or minimal media. Similar to studies in yeast [16], this very simple experimental design delivered a large amount of information: 225 genes were expressed at significantly higher levels on minimal glucose medium, while the expression levels of 119 genes were increased in glucose-rich medium. In rich media the expression of genes involved in protein synthesis increased, whereas cells grown in minimal media turned on the expression of genes involved in biosynthetic pathways as well as the synthesis of stress tolerance proteins.

Wilson and coworkers [31••] explored drug-induced alterations in global gene expression in *Mycobacterium tuberculosis*. They monitored expression changes in response to the antituberculosis drug isoniazid. Several genes in a biosynthetic pathway relevant to the drug's mode of action were identified in the induced gene set. In addition genes were induced that are not directly related to the biosynthetic pathway. These newly identified genes may define new drug targets.

Viruses

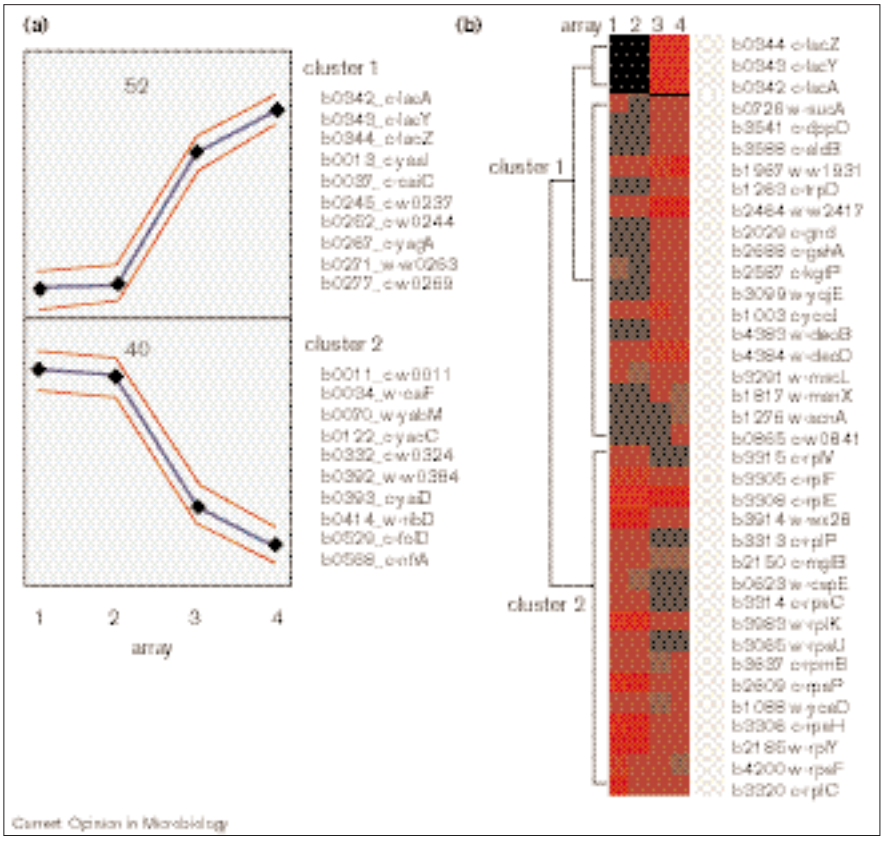
Microarray-based expression studies of viral systems have focused on both host and viral genomes. In a study of host cell response following infection with a viral pathogen, Zhu *et al.* [32•] examined mRNA changes in primary human foreskin fibroblasts 40 minutes, 8 hours, and 24 hours post infection with the human cytomegalovirus (HCMV). Using the Affymetrix HuGeneF1 array, they monitored the response of ~6800 human mRNAs following infection. After 24 hours, 364 mRNAs had changed in response to the virus by a factor of three or more. The authors speculate that several of the cellular genes whose mRNA levels change after infection of fibroblasts might profoundly influence HCMV replication and pathogenesis. Meanwhile, Chambers *et al.* [33] addressed the HCMV-host cell expression program from the other direction by examining changes in viral gene transcripts following infection.

Data analysis strategies

The data format produced by a microarray assay typically consists of a list of genes and corresponding values that represent relative RNA transcript levels. A high-density oligonucleotide array, such as the GeneChip yeast genome array (YG s98), generates 9337 data points for every hybridization reaction. In a small scale experiment consisting of only five samples with two replicates each such an array experiment will produce approximately 100,000 data points. It quickly becomes imperative that strategies are in place and computing resources available to manage the large quantities of data generated by microarray experiments [34].

Following acquisition and processing of the fluorescent array image, there are three basic steps required for efficient and effective data analysis: data normalization, data filtering, and pattern identification. To compare expression values directly, it is usually necessary to apply some sort of normalization strategy to the data, either between paired samples or across a set of experiments. Following this, data reduction can be done by filtering out uninformative genes; for example, genes that are expressed below a user defined threshold or genes that did not vary their expression level during the

Figure 2



Examples two types of clustering programs applied to the same data set. Samples from control (1 and 2) and isopropyl-β-D-thiogalactopyranoside (IPTG)-treated (3 and 4) *E. coli* K-12 (MG1655) cells were hybridized to high-density oligonucleotide arrays containing probes for more than 4200 *E. coli* ORFs. Expression values were determined using the GeneChip® software analysis program and scaled to a common, global average expression level. **(a)** Two of the 20 clusters produced by a self-organizing map program [36]. The red lines indicate the error bars. In cluster 1, 52 genes follow a similar pattern with increases in expression after induction (i.e. after IPTG treatment). Ten of the cluster members are shown in the list, including the three *lac* operon genes. In cluster 2, there are 40 genes with reduced expression after induction. Ten of the members are listed. **(b)** Two selected clusters produced by a hierarchical clustering program [35]. Cluster 1 generally follows a pattern of increased expression after induction. The three highly induced lac genes are grouped in a small sub-cluster. Cluster 2 is a group of genes whose expression levels are reduced after induction. Not all clusters are shown.

course of the experiment. The next step is to find patterns and groups in the data that can be used to assign biological meaning to the expression profiles. The methods used for data mining and interpretation are varied, ranging from straightforward lists of increased and decreased genes based on user-defined thresholds to the implementation of sophisticated clustering and visualization programs, such as hierarchical clustering [35] and self-organizing maps also called k-means clustering [36–38] (Figure 2). Hierarchical clustering has been traditionally used in phylogenetic analysis and typically uses a progressive combination of elements that are the most similar, while self-organizing maps sample the complete data set to determine the distances of the data points from randomly chosen points, called centroids. Both methods have their own unique advantages, but generally succeed in grouping the genes or samples based on their expression pattern. (Table 1 provides a list of data analysis and visualization programs suitable for microarray data.)

The data mining strategy used depends on the experimental design and can be broadly divided into two categories: differential gene expression and coordinated gene expression [39••]. The differential gene expression approach generally consists of paired comparisons between normal/abnormal data such as from healthy and pathological specimens or wild-type and mutant genotypes. In a study of normal human colon and colon tumor tissue, hierarchical clustering was used to differentiate cancerous from normal tissue [40•]. Taking this approach, the majority of cancerous samples were identified correctly. The introduction of prediction strength criteria by Golub *et al.* [41••] significantly improved the accuracy of phenotype predictions based on microarray expression patterns because it allows the identification of uncertain calls. These authors set out to differentiate between acute myeloid leukemia and acute lymphoid leukemia. A set of predictor genes whose expression pattern strongly correlates with the disease was chosen. This set of 'predictor genes' was able to correctly identify the type of leukemia with a 100% success rate, because marginal calls could be identified.

Coordinated gene expression analysis involves the assessment of the expression levels of a large number of genes over a period of time or through a series of experimental conditions, such as the studies of the transcriptional program during sporulation in budding yeast [19•] or of cell cycle variation [17•,18•]. In these kinds of studies effort has to be put into normalizing the data set across samples [39••], because it is important to distinguish real, biological change from random noise or non-specific experimental variation. In this respect, the value of doing

Table 1

Resources and software for microarray data analysis.

Resources and software	URL
Spotfire Pro for Windows	http://www.spotfire.com/
DataDesk from Datadescription Inc.	http://www.datadesk.com/
GeneCluster	http://www.genome.wi.mit.edu/MPR/software.html
Cluster and Tree View	http://rana.stanford.edu/software/
Wisconsin Genetic Package (GCG) SeqLab	http://www.gcg.com/index.html
PHYLIP Phylogenetic inference package	http://evolution.genetics.washington.edu/phylip.html
GenExplore	http://www.applied-maths.com/home.html
GeneSpring	http://www.sigenetics.com/index.html

replicate experiments cannot be overestimated, as Claverie [39••] succinctly states: "… no data processing or elegant protocol can substitute for the requirement of multiple (at least two) independent determinations of the expression intensities."

The value of looking at the expression pattern of a complete genome is most fully realized when the data is linked to other databases. The public databases provided by NCBI (National Center for Biotechnology information, http://www.ncbi.nlm.nih.gov/) (GenBank, Entrez, BLAST and PubMed) are a valuable, integrated set of data and tools and KEGG (Kyoto Encyclopedia of Genes and Genomes, http://www.genome.ad.jp/kegg/) offers a comprehensive database of biological pathways. These tools make it possible to annotate and identify functions in genes with a common expression profile.

Conclusions

Expression profiling with DNA microarrays has only just begun to be applied in the microbiology community, yet a wealth of data has already been generated. Effective utilization of this data and what is to come will require several steps forward: accurate and up-to-date gene annotation, generally agreed upon rules for data standardization, and mechanisms for reporting and sharing complete files of microarray expression data. In the near future we should also expect to see new tools and statistically validated approaches to mining these large data sets.

Improvements currently being made in labeling protocols for prokaryotic mRNAs will allow for more sensitive expression assays, down to less than one copy of mRNA per cell (C Rosenow, G Miyada, personal communication). Increased standardization in labeling protocols will also help to minimize artefactual expression profile differences between data generated in different laboratories.

The ability to measure RNA expression profiles across entire genomes provides a level of information not previously attainable. Sophistication in analyzing these data effectively and comprehensively is continually growing. This, combined with the ability to conduct experiments in which the genotype and growth environment of microorganisms is carefully controlled, will fundamentally and dramatically advance our understanding of the basic processes of living organisms.

Acknowledgements

The authors thank Tom Gingeras, Linda McAllister, Garry Miyada and Mark Durst for critical reading of the manuscript. We thank Andy Lau for excellent support in preparation of Figure 1.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Lander E: **Array of hope.** *Suppl Nat Genet* 1999, **21**:3-4.

2. Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays.** *Suppl Nat Genet* 1999, **21**:33-37.

3. Case-Green SC, Mir KU, Pritchard CE, Southern EM: **Analysing genetic information with DNA arrays.** *Curr Opin Chem Biol* 1998, **2**:404-410.

4. •• Ferea TL, Brown PO: **Observing the living genome.** *Curr Opin Genet Dev* 1999, **9**:715-722.
An overview of array technology with emphasis on spotted arrays. Includes table of web sites for making and using DNA arrays and useful information on pattern recognition and data visualization strategies.

5. • Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays.** *Suppl Nat Genet* 1999, **21**:20-24.
This review includes a summary of the manufacturing process for high-density oligonucleotide arrays and describes how these arrays are being applied, including gene expression monitoring and genotypic analysis.

6. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: **Expression profiling using cDNA microarrays.** *Suppl Nat Genet* 1999, **21**:10-14.

7. Eisen MB, Brown PO: **DNA arrays for analysis of gene expression.** In *Methods in Enzymology: cDNA Preparation and Characterization.* vol 303. Edited by Weissman SM. Academic Press; 1999:179-205.

8. Bowtell DDL: **Options available-from start to finish-for obtaining expression data by microarray.** *Suppl Nat Genet* 1999, **21**:25-32.

9. Winzeler EA, Schena M, Davis RW: **Fluorescence-based expression monitoring using microarrays.** In *Methods in Enzymology: Expression of Recominant Genes in Eukaryotic Systems,* vol 306. Edited by Glorioso J, Schmidt MC. Academic Press: London; 1999:3-18.

10. Schena M, Shalon D, Davis RW, Brown PO: **Quantitiative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1996, **270**:467-470.

11. De Risi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.

12. Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA: **Accessing genomic information with high-density arrays.** *Science* 1996, **274**:610-614.

13. 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* 1996, **14**:1675-1680.

14. Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR: **Genome-wide**
• **expression in *Escherichia coli* K-12.** *Nucleic Acid Res* 1999, **27**:3821-3835.
Changes in RNA levels after exposure of *E. coli* to heat shock or IPTG are compared using two different methods: hybridization of radioactive cDNA to spot blots on nylon membranes and fluorescence-based hybridization to glass microarrays.

15. Lashkari DA, DeRisi JL, McCusker JH, Namath AF, Gentile C, Hwang SY, Brown PO, Davis RW: **Yeast microarrays for genome wide parallel genetic and gene expression analysis.** *Proc Natl Acad Sci USA* 1997, **94**:13057-13062.

16. Wodicka L, Dong H, Mittmann M, Ho M-H, Lockhart DJ: **Genome-wide expression monitoring in *Saccharomyces cerevisiae.*** *Nat Biotechnol* 1997, **15**:1359-1367.

17. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A,
• Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman, Lockhart DJ, Davis RW: **A genome-wide transcriptional analysis of the mitotic cell cycle.** *Mol Cell* 1998, **2**:65-73. [http://genomics.stanford.edu]
One of the first studies to report genome-wide patterns of expression through a series of cell states. This study measures mRNA transcript levels during the yeast cell cycle using high-density oligonucleotide arrays.

18. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB,
• Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated gene of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297. [http://cellcycle-www.stanford.edu]
Study measuring genome-wide mRNA transcript levels during the yeast cell cycle using spotted DNA microarrays. Promoter elements predictive of cell-cycle regulation were identified.

19. Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO,
• Herskowitz I: **The transcriptional program of sporulation in budding yeast.** *Science* 1998, **282**:699-705. [http://cmgm.stanford.edu/pbrown/sporulation]
Global patterns of gene expression during a developmental program, sporulation in yeast, are presented. Temporal expression patterns suggest potential functions of many uncharacterized genes.

20. Jelinsky S, Samson L: **Global response of *Saccharomyces**
• ***cerevisiae* to an alkylating agent.** *Proc Natl Acad Sci USA* 1999, **96**:1486-1491.
This work demonstrates the ability of global expression profiling to dramatically increase information about specific cellular processes. It also includes a study correlating Northern analyses with microarray results.

21. Holstege F, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA: **Dissecting the regulatory circuitry of a eukaryotic genome.** *Cell* 1998, **95**:717-728.

22. Lelivelt M, Culberston M: **Yeast upf proteins required for RNA surveillance affect global expression of the yeast transcriptome.** *Mol Cell Biol* 1999, **19**:6710-6719.

23. Madhani HD, Galitski T, Lander ES, Fink GR: **Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatires of signaling mutants.** *Proc Natl Acad Sci USA* 1998, **96**:12530-12535.

24. Marton MJ, DeRisi JL, Bennet HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard, J, Slade D, Dai H *et al.*: **Drug validation and identification of secondary drug target effects using DNA microarrays.** *Nat Med* 1998, **4**:1293-1301.

25. Gray NS, Wodicka L, Thunnissen AWH, Norman TC, Kwon S, Espinoza FH, Morgan DO, Barnes G, LeClerc S, Meijer L *et al.*: **Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors.** *Science* 1998, **281**:533-538.

26. Lockhart DJ: **Mutant yeast on drugs.** *Nat Med* 1998, **4**:1235-1236.

27. Ferea TL, Botstein D, Brown PO, Rosenzweig RF: **Systematic changes in gene expression patterns following adaptive evolution.** *Proc Natl Acad Sci USA* 1999, **96**:9721-9726.

28. Gingeras T, Rosenow C: **High density oligonucleotide arrays and their use in the study of microbial genomes: opportunities unique to prokaryotic genomes.** *ASM News* 2000, in press.

29. De Saizieu A, Certa U, Warrington J, Gray C, Keck W, Mous J: **Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays.** *Nat Biotechnol* 1998, **16**:45-48.

30. Tao H, Bausch C, Richmond C, Blattner FR, Conway T: **Functional**
• **genomics: expression analysis of *Escherichia coli* growing on minimal and rich media.** *J Bacteriol* 1999, **181**:6425-6440.
Comparison of gene expression profiles using total RNA from *E. coli* cells grown in rich and minimal medium. A comprehensive analysis of differentially expressed genes is presented.

31. Wilson M, DeRisi J, Kristensen H-H, Imboden P, Rane S, Brown PO,
•• Schoolnik GK: **Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization.** *Proc Natl Acad Sci USA* 1999, **96**:12833-12838.
PCR-based microarrays of the entire *M. tuberculosis* genome were used to study gene induction after treatment of the bacterial cell with the drug isoniazid. Several induced genes were linked to proteins that are relevant to the drug's mode of action.

32. Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T: **Cellular gene**
• **expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays.** *Proc Natl Acad Sci* 1998, **95**:14470-14475.
Comprehensive analysis of human genes whose mRNA levels change after infection of primary fibroblast cells with the human cytomegalovirus.

33. Chambers J, Angulo A, Amaratunga D, Guo H, Jiang Y, Wan JS, Bittner A, Frueh K, Jackson MR, Peterson PA *et al.*: **DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression.** *J Virol* 1999, **73**:5757-5766.

34. Bassett DE, Eisen B, Boguski MS: **Gene expression informatics-it's all in your mine.** *Suppl Nat Genet* 1999, **21**:51-55.

35. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns**. *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.

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

37. Törönen P, Kolehmainen M, Wong G, Castrën E: **Analysis of gene expression data using self-organizing maps.** *FEBS Lett* 1999, **451**:142-146.

38. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-285.

39. Claverie JM: **Computational methods for the identification of**
•• **differential and coordinated gene expression.** *Hum Mol Genet* 1999, **8**:1821-1832.
Review of computational methods used to identify genes that are differentially expressed between normal and pathological samples or coordinately expressed across experimental conditions or through time. Also addresses the clustering of genes with coherent expression features, such as membership in biological pathways.

40. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ:
• **Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays.** *Proc Natl Acad Sci USA* 1999, **96**:6745-6750.
Two-way clustering was applied to microarray data from 40 tumor and 22 normal colon tissue samples. The clustering program separated cancerous from noncancerous tissue even when expression of individual genes varied only slightly between the tissues.

41. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M,
•• Mesirov-JP, Coller H, Loh ML, Downing JR, Caligiuri MA *et al.*: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286**:531-537.
This paper describes how two types of leukemia can be distinguished through use of an automatically derived predictor.