

REVIEW
ARTICLE

Microarrays for microbiologists

S. Lucchini, A. Thompson and J. C. D. Hinton

Author for correspondence: J. C. D. Hinton. Tel: +44 1603 255352. Fax: +44 1603 255076.
e-mail: jay.hinton@bbsrc.ac.uk

Molecular Microbiology, Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, UK

Keywords: DNA microarray, gene expression profiling, microbial genotyping

Background

We are witnessing a remarkable change in the scale of molecular microbiological research and we are entering an era of ‘big science’. In the past decade we have moved from a time when entire research papers were based on the sequencing of a single gene or operon to a single paper describing the sequence of a whole genome. The completion of microbial genomes is continuing apace, with 37 genome sequences completed, and 142 in progress worldwide (<http://www.tigr.org/tdb/mdb/mdbcomplete.html>). The availability of this level of genetic information has spawned the terms ‘functional genomics’, ‘transcriptomics’ (Velculescu *et al.*, 1997) and ‘proteomics’ (Wasinger *et al.*, 1995), which describe the large-scale application of mass mutagenesis, gene expression profiling and global protein analysis. In this review we assess the role that gene expression profiling has begun to play in microbiology, discuss the potential for ‘genomotyping’ and consider future applications.

Assessment of transcription at the genomic scale has been achieved with DNA microarrays, which are glass slides containing an ordered mosaic of the entire genome as a collection of either oligonucleotides (oligonucleotide microarrays) or PCR products representing individual genes (commonly referred as cDNA microarrays).

The development of microarrays has been fuelled by the application of robotic technology to routine molecular biology, rather than by any fundamental breakthrough. The classical Southern and Northern blotting approaches for the detection of specific DNA and mRNA species (Southern, 1975; Alwine *et al.*, 1977, 1979) provided the technological basis for microarray hybridization with fluorescently labelled cDNA. The idea of depositing multiple DNA spots representing different genes onto a solid surface is also nothing new, having been used by Blattner’s group to investigate *Escherichia coli* gene expression on membranes (microarrays) as long ago as 1993 (Chuang *et al.*, 1993). Commercially available microarrays have continued to produce useful data, and should be considered before recourse to microarrays (Tao *et al.*, 1999). The recent

application of robotics to achieve high spotting densities of DNA on glass slides was innovative and facilitates the construction of microarrays containing up to 50000 genes on a single microscope slide (DeRisi *et al.*, 1996; Shalon *et al.*, 1996). This allows a single hybridization to be performed against multiple replicates of a single bacterial genome, or against copies of several unrelated genomes on a single glass slide. The development that has facilitated the reproducible comparison of gene expression between two samples, and hence between experiments, is dual fluorescent labelling (Schna *et al.*, 1995). Simultaneous hybridization of two cDNA populations labelled with the fluorescent dyes Cy3 and Cy5 allows accurate assessment of relative levels of gene expression, which is unaffected by hybridization variability or the differences between individual microarrays which can complicate microarray experiments.

Microarrays as a research tool

Microarrays allow us to produce a ‘gene expression profile’ or ‘signature’ for a particular organism under certain environmental conditions. Since the first report

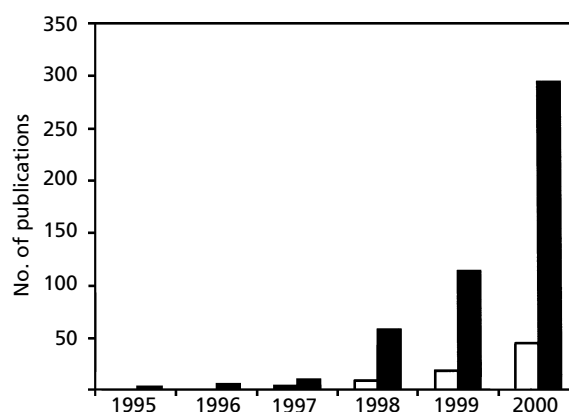


Fig. 1. The dramatic increase in the number of publications involving DNA microarrays. ■, All microarray papers; □, microbial microarray papers only.

Table 1. Applications of microarray technology to microbiological research

Application	Organism	Type of array*	Reference
Response to stress/environmental change			
Diauxic shift	<i>Sacch. cerevisiae</i>	cDNA glass slides	DeRisi <i>et al.</i> (1997)
Growth in minimal/rich media	<i>Sacch. cerevisiae</i>	HDA	Wodicka <i>et al.</i> (1997)
Heat shock, cold shock, growth with galactose/glucose	<i>Sacch. cerevisiae</i>	cDNA glass slides	Lashkari <i>et al.</i> (1997)
Effect of DNA-damaging agents	<i>Sacch. cerevisiae</i>	HDA	Jelinsky & Samson (1999)
Growth in minimal/rich media	<i>E. coli</i>	cDNA membranes	Tao <i>et al.</i> (1999)
Growth in minimal/rich media	<i>E. coli</i>	cDNA glass slides	Wei <i>et al.</i> (2001)
IPTG, heat shock	<i>E. coli</i>	cDNA membranes and glass slides	Richmond <i>et al.</i> (1999); Wei <i>et al.</i> (2001)
Aerobic/anaerobic growth	<i>Sacch. cerevisiae</i>	HDA	ter Linde <i>et al.</i> (1999)
Aerobic/anaerobic growth	<i>B. subtilis</i>	cDNA glass slides	Ye <i>et al.</i> (2000)
Osmotic stress	<i>Sacch. cerevisiae</i>	cDNA glass slides	Posas <i>et al.</i> (2000)
Growth in different carbon sources	<i>E. coli</i>	cDNA glass slides	Oh & Liao (2000)
Metal-ion tolerance	<i>E. coli</i>	cDNA membranes	Brocklehurst & Morby (2000)
Quorum sensing	<i>Streptococcus pneumoniae</i>	HDA	de Saizieu <i>et al.</i> (2000)
Response to changes that affect tryptophan metabolism	<i>E. coli</i>	cDNA glass slides	Khodursky <i>et al.</i> (2000a)
Cell-cycle-associated gene expression			
Transcriptional programme of sporulation in yeast	<i>Sacch. cerevisiae</i>	cDNA glass slides	Chu <i>et al.</i> (1998); Primig <i>et al.</i> (2000)
Identification of cell-cycle-regulated genes	<i>Sacch. cerevisiae</i>	cDNA glass slides	Spellman <i>et al.</i> (1998)
Analysis of mitotic cell cycle	<i>Sacch. cerevisiae</i>	HDA	Cho <i>et al.</i> (1998)
Transcriptional changes during competence development	<i>Strep. pneumoniae</i>	cDNA membranes	Rimini <i>et al.</i> (2000)
Identification of cell-cycle-regulated genes	<i>C. crescentus</i>	cDNA glass slides	Laub <i>et al.</i> (2000)
Dissection of regulatory circuitry			
Effects of mutation in RNA polymerase II components, determination of mRNA stability	<i>Sacch. cerevisiae</i>	HDA	Holstege <i>et al.</i> (1998)
Promotor sequence analysis of coexpressed genes	<i>Sacch. cerevisiae</i>	' <i>In silico</i> biology'	Wolfsberg <i>et al.</i> (1999); Zhang (1999)
Effects of nucleosome depletion on gene expression	<i>Sacch. cerevisiae</i>	HDA	Wyrick <i>et al.</i> (1999)
Effects of gene dosage	<i>Sacch. cerevisiae</i>	HDA	Galitski <i>et al.</i> (1999); Giaever <i>et al.</i> (1999)
Effects of gene dosage	<i>Sacch. cerevisiae</i>	cDNA glass slides	Hughes <i>et al.</i> (2000b)
Snf/Swi protein complex mutants and nucleosome structure	<i>Sacch. cerevisiae</i>	cDNA glass slides	Sudarsanam <i>et al.</i> (2000)
<i>pdr1p/pdr3p</i> mutants (transcriptional activators involved in pleiotropic drug resistance)	<i>Sacch. cerevisiae</i>	cDNA glass slides	DeRisi <i>et al.</i> (2000)
Regulation of iron uptake	<i>Sacch. cerevisiae</i>	cDNA glass slides	Yun <i>et al.</i> (2000)
<i>marA</i> mutants (transcriptional activator involved in drug resistance)	<i>E. coli</i>	cDNA membranes	Barbosa & Levy (2000)
Characterization of the Fap1p zinc-sensitive regulon	<i>Sacch. cerevisiae</i>	cDNA glass slides	Lyons <i>et al.</i> (2000)
Regulation of glycolytic enzyme genes	<i>Sacch. cerevisiae</i>	cDNA membranes	Lopez & Baker (2000)
Identification of copper regulon	<i>Sacch. cerevisiae</i>	cDNA glass slides	Gross <i>et al.</i> (2000)
Chromosomal position and gene expression	<i>Sacch. cerevisiae</i>	<i>In silico</i>	Cohen <i>et al.</i> (2000a)
Characterization of IHF regulon	<i>E. coli</i>	cDNA membranes	Arfin <i>et al.</i> (2000)
Function of histone deacetylase	<i>Sacch. cerevisiae</i>	cDNA glass slides	Bernstein <i>et al.</i> (2000)
Translational regulation	<i>Sacch. cerevisiae</i>	cDNA glass slides	Kuhn <i>et al.</i> (2001)

Table 1 (cont.)

Application	Organism	Type of array*	Reference
Genomotyping			
Detection of point mutation polymorphism	<i>M. tuberculosis</i>	HDA	Gingeras <i>et al.</i> (1998)
Detection of gene deletions/insertions	<i>E. coli</i>	cDNA membranes	Riehle <i>et al.</i> (2001)
	<i>M. tuberculosis</i> / <i>M. bovis</i>	cDNA glass slides	Behr <i>et al.</i> (1999)
Strain comparison	<i>Sacch. cerevisiae</i>	HAD	Winzeler <i>et al.</i> (1999a)
	<i>Helicobacter pylori</i>	cDNA glass slides	Salama <i>et al.</i> (2000)
Species identification and detection of point mutations in <i>rpoB</i>	<i>M. tuberculosis</i>	HDA	Troesch <i>et al.</i> (1999)
Drug target characterization/identification			
Kinase inhibitors	<i>Sacch. cerevisiae</i>	HDA	Gray <i>et al.</i> (1998)
Effects of immunosuppressant FK506	<i>Sacch. cerevisiae</i>	cDNA glass slides	Marton <i>et al.</i> (1998)
Isoniazid	<i>M. tuberculosis</i>	cDNA glass slides	Wilson <i>et al.</i> (1999)
Effect of amino acid biosynthesis inhibitor sulfometuron methyl	<i>Sacch. cerevisiae</i>	cDNA glass slides	Jia <i>et al.</i> (2000)
Antifungal agents	<i>Sacch. cerevisiae</i>	HDA	Bammert & Fostel (2000)
Cellular response to bacterial infection			
Infection of monocytes (THP-1) with <i>Listeria monocytogenes</i>	Human	HDA, cDNA membranes, cDNA colony filters	Cohen <i>et al.</i> (2000b)
Infection of intestinal epithelial cells (HT-29 and T84) with <i>Salmonella enterica</i>	Human	cDNA membranes	Eckmann <i>et al.</i> (2000)
Infection of macrophages cells (RAW 264.7) with <i>Sal. typhimurium</i>	Mouse	cDNA membranes	Rosenberger <i>et al.</i> (2000)
Infection of epithelial cells (lung carcinoma A549 cell line) with <i>Pseudomonas aeruginosa</i>	Human	cDNA glass slides	Ichikawa <i>et al.</i> (2000)
Infection of intestinal epithelial cells (HeLa 229) with <i>Chlamydia trachomatis</i>	Human	cDNA membranes	Dessus-Babus <i>et al.</i> (2000)
Infection of bronchial epithelial cells (BEAS-2B) with <i>Bordetella pertussis</i>	Human	HDA	Belcher <i>et al.</i> (2000)
<i>In vivo</i> effects of a commensal bacterium on the gene expression of the ileal epithelium	Mouse	HDA	Hooper <i>et al.</i> (2001)
Others			
Identification of genetic markers	<i>Sacch. cerevisiae</i>	HDA	Winzeler <i>et al.</i> (1998)
Characterization of adaptive evolution	<i>Sacch. cerevisiae</i>	cDNA glass slides	Ferea <i>et al.</i> (1999)
RNA surveillance	<i>Sacch. cerevisiae</i>	HDA	Lelivelt & Culbertson (1999)
Prediction of protein subcellular localization	<i>Sacch. cerevisiae</i>	cDNA glass slides	Diehn <i>et al.</i> (2000)
Analysis of topoisomerase function at replication forks	<i>E. coli</i>	cDNA glass slides	Khodursky <i>et al.</i> (2000b)
Role of RNA polymerase II subunit Rpb9 on transcription elongation	<i>Sacch. cerevisiae</i>	cDNA glass slides	Hemming <i>et al.</i> (2000)
Prediction of gene function by comparing expression profiles of deletion mutants	<i>Sacch. cerevisiae</i>	cDNA glass slides	Hughes <i>et al.</i> (2000a)
Genome-wide location of DNA-binding proteins	<i>Sacch. cerevisiae</i>	cDNA glass slides	Ren <i>et al.</i> (2000); Vishwanath <i>et al.</i> (2001)

* HDA, high-density oligonucleotide array.

of DNA microarray technology in 1995 (Schena *et al.*, 1995; DeRisi *et al.*, 1996; Lockhart *et al.*, 1996) the potential of DNA microarrays has certainly captured the imagination of biologists worldwide. Naturally, we welcome the ability to monitor global gene expression in a single experiment rather than relying on the 'one gene

= one postdoc' approach to eventually elucidate the function of all bacterial genes. However, there has also been a concern that this genome-wide approach might signal a move towards 'non-hypothesis-driven' or 'data-driven' science (Brent, 1999), a term that has been used rather pejoratively. It is clear that scientific inference

uses a combination of deductive and inductive reasoning (Kell & King, 2000). Functional genomics allows us to make new and unexpected links between the function of unrelated and hitherto uncharacterized genes, and suggest hypotheses, which must subsequently be tested by more traditional methods of molecular genetics and biochemistry (Hughes *et al.*, 2000a). An example lies with the increasing numbers of proteins which have unexpectedly been found to have dual cellular functions, such as enolase and aconitases: in addition to being a glycolytic enzyme enolase is also a vital constituent of the RNA degradosome (Py *et al.*, 1996), and aconitases, besides their catalytic role in the TCA cycle, have been shown to act as a post-transcriptional regulator by binding mRNA (Tang & Guest, 1999). Genomic-scale research may be termed 'non-hypothesis-driven' science, but we suggest it should be viewed positively as it is likely to reveal the function of many genes which have been missed by more conventional approaches. The need for this is apparent when considering the genome sequence of *E. coli*, which still contains 1632 (38%) FUN genes (of unknown function; Hinton, 1997), which remain to be functionally characterized (Nelson *et al.*, 2000). The role of FUN genes will not be discovered without the application of functional genomic technologies combined with creative experiments.

The explosive growth in the numbers of reviews discussing microarray technology has now been followed by many papers describing results obtained from gene expression microarray profiling (Fig. 1). Genomic and post-genomic approaches are likely to revolutionize our ability to understand how micro-organisms act, both in the laboratory and in the real world. But what effect has this new approach had on molecular microbiologists in general, and what difference is it likely to make in the future?

The answer depends on one's field of research. Yeast researchers have already embraced microarrays as a useful and productive tool, as exemplified by 60% of the publications in Table 1. In other areas of microbiology, the application of microarray technology to bacteria has been slower. *E. coli*, *Mycobacterium tuberculosis* and other bacteria were the subject of just 37% of the microbial papers involving microarrays. The slow application of microarrays to academic bacterial research probably reflects the complex nature of this technology. Many pharmaceutical and biotechnology companies are successfully using bacterial microarrays to drive programmes of novel drug development, and this industrial experience suggests that technical problems will not be a barrier. We hope that work in our own and other laboratories worldwide will soon produce a raft of informative data concerning bacterial gene expression.

Microbial gene expression profiling

Microarrays have already been used to perform high-quality experiments, which have improved our understanding of microbial environmental responses and

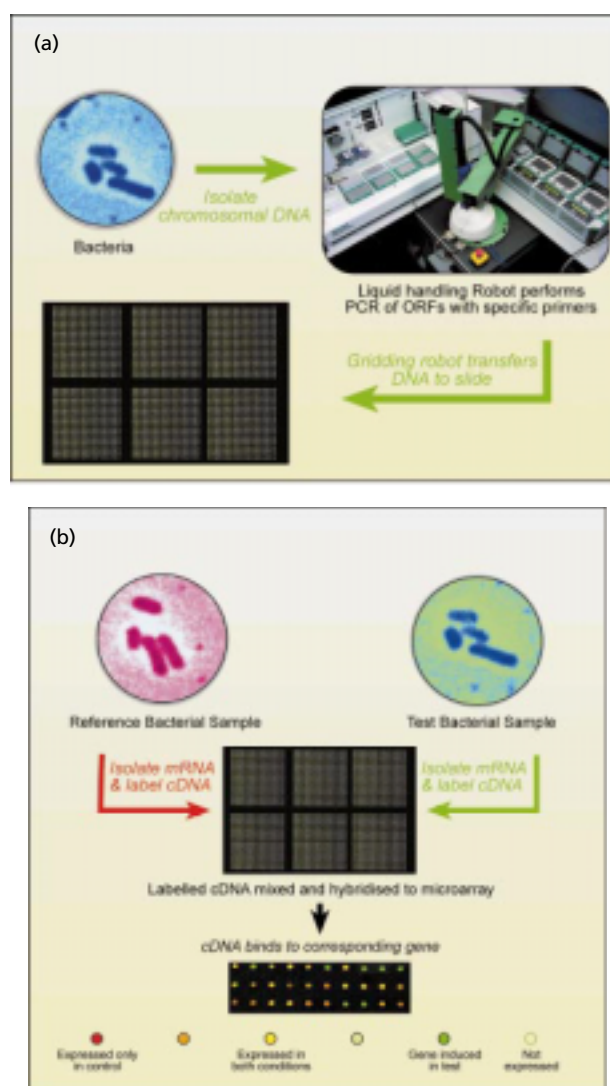


Fig. 2. Making (a) and using (b) DNA microarrays for gene expression profiling. cDNA labelling was performed using the indirect labelling method (<http://www.microarrays.org>).

global gene expression (Fig. 2; Table 1) and the concept of the 'global transcriptional response' as a detailed molecular phenotype is now gaining acceptance (Hughes *et al.*, 2000a). The first global transcriptional profile was obtained at the resolution of individual genes for *Saccharomyces cerevisiae*; Pat Brown's group did this with gene-specific microarrays (DeRisi *et al.*, 1997; Lashkari *et al.*, 1997) and Affymetrix utilized oligonucleotide 'gene chips' (Wodicka *et al.*, 1997). Eighteen months later, the first microarrays involving the whole genome of two prokaryotes were described: *M. tuberculosis* (Behr *et al.*, 1999) and *E. coli* (Richmond *et al.*, 1999; Tao *et al.*, 1999). As genome sequencing projects are being completed and new applications of microarrays are being developed, the potential of microarray analysis are being rapidly applied to other micro-organisms (Table 1). It is not always appreciated

that a complete genome sequence is not essential for a microarray project. Interesting results can be obtained from microarrays assembled from partial genome data, or from an uncharacterized gene library, where spots of interest are sequenced once their expression profile has been determined (Pennisi, 2000).

Yeast has been the micro-organism of choice for many research groups to analyse cell-cycle-associated gene expression and the effects of various environmental changes, such as osmotic shock, temperature shock, presence of DNA-damaging agents and growth in minimal or rich media (Table 1). The new level of analysis provided by whole-genome expression profiling has revealed the complexity of the cellular response to major changes in metabolism, exemplified by work on yeast diauxic shift (DeRisi *et al.*, 1997). The expression levels of 1840 genes (30% of a total of 6116 genes tested) were found to be affected by the transition from anaerobic to aerobic growth. Similar complexity of gene expression in the transcriptional programme was reported for yeast going through the mitotic cell cycle (Cho *et al.*, 1998; Spellman *et al.*, 1998) or sporulation (Chu *et al.*, 1998). Many of the responsive genes that were identified had previously been designated as FUN. To understand the large amount of data created by microarrays, Mike Eisen developed a computer program to cluster genes according to their expression profiles (Eisen *et al.*, 1998). Based on the important observation that functionally related genes show similar patterns of expression, identification of well-characterized genes that are co-expressed with FUN genes can give important clues towards function. Using this tool, Chu *et al.* (1998) defined seven sequential temporal classes of genes induced during yeast sporulation.

One of the most impressive examples of the use of microarrays for bacterial research has been provided by recent work on *Caulobacter crescentus* (Laub *et al.*, 2000). The definition of the cell cycle of *C. crescentus* by microarray analysis revealed that 572 of 2966 genes (19.3%) were cell-cycle-dependent. Not only were a number of classes of cell-cycle-induced genes identified, but also the proportion dependent upon the global cell cycle regulator CtrA was recognized for the first time. This study led to recognition of the role of 11 novel sensor kinases and 5 new sigma factors. The identification of cascades of gene expression during the *Caulobacter* cell cycle is an important landmark for bacterial research. We look forward to similar studies describing gene expression cascades during sporulation of *Bacillus subtilis*, and during *E. coli* cell division.

Definition of entire regulons

Because cell cycle and environmental changes are multifactorial, involving important metabolic changes, it is difficult to unravel the role played by specific global gene regulators. The definition of important regulons by the use of appropriate regulatory mutants provides the framework for a better understanding of complex cellular responses. This approach has led to the global

characterization of the IHF and MarA regulons of *E. coli* (Arfin *et al.*, 2000; Barbosa & Levy, 2000), but does not distinguish between direct and indirect effects of regulatory mutations. More recently, DNA microarrays have been used to combine gene expression profile analysis with the localization of binding sites for DNA-binding proteins on the yeast genome. This approach involved the formaldehyde cross-linking of proteins to the DNA, followed by a modified chromatin precipitation procedure. After cell disruption, proteins were immunoprecipitated to enrich for DNA fragments containing DNA-binding sites. The enriched DNA was then amplified, labelled and hybridized to a microarray containing all yeast intergenic regions. This new approach permitted the identification of genes which were directly controlled by the regulatory proteins Gal4 and Ste12 (Ren *et al.*, 2000).

Analysis of gene expression *in vivo*

The use of microarrays for the study of bacterial infection at the level of gene expression is in its infancy (Cummings & Relman, 2000; Hautefort & Hinton, 2000). We require improved technology for *in vivo* study of infection, both from the host's and the pathogen's point of view. Bacterial mRNA is less stable than eukaryotic mRNA, and is generally not polyadenylated, complicating mRNA purification (Sarkar, 1996). This does not cause problems for *in vitro* analyses, but does complicate the isolation of sufficient, good-quality bacterial mRNA from complex environments such as mammalian tissue. Advances have recently been made in applying the linear RNA amplification method (Eberwine *et al.*, 1992), which has been used to extract sufficient eukaryotic mRNA for microarray analyses from small amounts of mammalian tissue (Lockhart *et al.*, 1996; Luo *et al.*, 1999; Wang *et al.*, 2000). The successful application of this linear RNA amplification approach to polycistronic and non-polyadenylated bacterial mRNA remains to be demonstrated. An alternative way to decrease the amount of bacterial RNA required for the labelling of mRNA involves the new concept of genome-specific primers. Thirty-seven genome-directed primers (GDPs) were predicted to be sufficient to prime all genes in the *M. tuberculosis* genome, and were used to label mycobacterial RNA (Talaat *et al.*, 2000). The cDNA probes generated by GDPs showed improved sensitivity and specificity when compared with probes obtained by random priming, and allowed a degree of selective amplification of mycobacterial RNA from a mix containing mammalian RNA.

Mammalian gene microarrays have recently been used to study host-pathogen interactions from the viewpoint of the host, by identifying gene expression patterns induced by the presence of a pathogen (Manger & Relman, 2000). Several *in vitro* studies have explored the effects of infection on the mRNA expression profile of human cells (Table 1). The effects of *Listeria monocytogenes* (Cohen *et al.*, 2000b), *Salmonella enterica* (Eckmann *et al.*, 2000) and *Salmonella typhimurium*

(Rosenberger *et al.*, 2000) have recently been reviewed by Cummings & Relman (2000). Briefly, these studies reveal a specific host response, which is modulated by different host factors (Rosenberger *et al.*, 2000). Other groups are using a more complex approach by comparison of the human cellular transcriptional signatures of pathogenic strains carrying well-defined mutations to get a more detailed view of the mechanisms underlying pathogen clearance (Manger & Relman, 2000).

Genomotyping and microbial evolution

Phylogenetic classification based on rRNA/rDNA and signature sequences is usually able to provide an accurate classification of micro-organisms above the species level (Gupta, 2000). However, it has become evident that lateral gene transfer is an important mechanism of evolution for prokaryotes, complicating phylogenetic analysis based on a small number of genes. Differences can be considerable between closely related bacterial pathogens: certain serovars of *S. enterica* contain more than a megabase of sequence information than others (Ochman *et al.*, 2000). Preliminary analysis suggests that more than 700 ORFs are present in *Salmonella typhi* and not in *Sal. typhimurium* (P. O'Gaora, personal communication). Acquisition of new DNA is clearly an important mechanism for the adaptation of bacteria to new ecological niches, as shown by the example of the acquisition of pathogenicity islands, which encode virulence factors and were probably acquired at different times (Groisman & Ochman, 1997). Whole-genome based methods are thus required to determine the repertoire of virulence genes found in bacterial pathogens. Strain comparison by hybridizing genomic DNA to microarrays (genomotyping) is a more realistic approach than the whole-genome sequencing of dozens of strains. Gene-specific microarrays have been used to compare the entire genome of a *Mycobacterium bovis* vaccine strain with the closely related *M. tuberculosis* H37Rv, resulting in a 'gene-specific fingerprint' at a resolution of approximately 2 kb (Behr *et al.*, 1999). Alternatively, mutant alleles or single nucleotide polymorphisms (SNPs) can be detected with high-density oligonucleotide microarrays, which carry much shorter targets (typically about 25 nt on Affymetrix gene chips). A single nucleotide difference between target and probe can be sufficient to prevent hybridization, and has been used to identify 3000 polymorphisms between two strains of *Sacch. cerevisiae*. These polymorphisms were used as markers to map five uncharacterized loci to within 11–64 kb (Winzeler *et al.*, 1998). Complete characterization of SNPs can be achieved for each base of a sequence of interest by using a set of four oligonucleotides, one oligonucleotide corresponding to the wild-type and the remaining oligonucleotides to the three possible mutations (Lemieux *et al.*, 1998). Oligonucleotide-chip-based mutation analysis is limited by a lack of sensitivity for mutations in regions with high local A/T or G/C content or for small frameshift mutations (Favis *et al.*, 2000). Nevertheless, this method promises to be extremely powerful as a diagnostic tool,

as demonstrated by the identification of mutations in a 705 bp region of the *rpoB* gene, which caused rifampicin resistance in 44 clinical isolates of *M. tuberculosis* (Gingeras *et al.*, 1998).

Industrial applications of microarrays

Gene expression profiling is being used to determine the effects of antibiotics, agrochemicals and pharmaceutical products on different organisms, and is being used in the search for new antimicrobials. Strategies for drug-target validation and the identification of secondary effects have been described previously (Rosamond & Allsop, 2000). Following determination of the 'expression signature' of a wide range of compounds, the prediction of the mode of action of a novel compound becomes possible, simply on the basis of analysis of the transcriptional changes made by the drug. Large biotech companies are already using this approach to obtain cost-effective information, which avoids large-scale mode-of-action studies.

Microarray data analysis

Microarrays provide huge amounts of data. This can be an advantage and a disadvantage; the potential of genome-scale information is incredible, but often the results of microarray analyses have been limited to the production of a catalogue of the induction or repression ratios of particular genes. This restricted approach fails to exploit the true value of microarray data. To obtain interesting and reliable hypotheses, and hence results, good mathematical and statistical tools are needed for the intelligent interrogation, or 'mining' of microarray data. Consensus must be reached on the level of differential gene expression that is truly significant, and similar approaches must be used by the worldwide community. This will be complicated by a recent analysis of global gene expression in *E. coli*, which concluded that there was 'little correlation between the "fold difference" and the accuracy of differential gene expression levels'. Thus, the significance of differential gene expression measurements cannot be assessed simply by considering the magnitude of the difference between two experimental conditions (Arfin *et al.*, 2000). For a single microarray experiment involving 5000 measurements, it is predicted that 250 false positives could arise from a Gaussian distribution of data points, emphasizing the importance of experimental replication and careful assessment of statistical significance (Arfin *et al.*, 2000). To assess issues such as the number of assays required, the significance of changes in expression levels or within a cluster analysis will be essential (Zweiger, 1999). Fortunately, bioinformatic tools are being developed at great speed (Ochman *et al.*, 2000). As well as extracting all we can from microarray data, we also need to be able to directly compare experiments within our research community. Since whole-genome transcriptional responses are very complex, all source of noise must be minimized, and good standardization procedures must be applied in

terms of experimental design, the description of results and the format for data storage (Aach *et al.*, 2000). There is a strong argument for the use of universal controls for particular microarrays. For example, the proposed use of genomic DNA as a reference for all gene expression studies performed by members of the *Sal. typhimurium* microarray community should facilitate rapid exchange of meaningful data between laboratories.

A further step towards the prediction of gene function has been made by combining the high-throughput production of data provided by microarrays with a rigorous statistical analysis. Hughes *et al.* (2000a) have reported a large-scale approach that is intended to avoid problems of biological noise, and to build up a reference database or 'compendium' of patterns of gene expression profiles corresponding to 300 different mutations and chemical treatments in *Sacch. cerevisiae*. Two-dimensional hierarchical clustering of the obtained expression profiles identified several large groups of coregulated genes. Mutations in genes having similar known functions gave rise to similar profiles, which clustered together, giving an experimental basis for gene function prediction. This tactic has allowed small but coordinated differential gene expression levels to be observed across many different conditions, and to be related to gene function.

In the excitement of pursuing gene expression profiling for entire organisms, we must not lose sight of the fact that mRNA is only one intermediate between DNA and protein. Post-transcriptional and post-translational controls also play a major role in modulating protein expression. Transcriptional analysis may generate hypotheses, but more traditional molecular biological and proteomic approaches are still required to test these hypotheses.

The utility of microarrays now extends to the study of translational initiation. Kuhn *et al.* (2001) analysed translational regulation of specific mRNAs in yeast. Polysomal fractionation was used in conjunction with microarrays to study changes in translational initiation during diauxic shift. Although overall mRNA translation decreased, the authors identified one group of mRNA species (representing 610 out of 6275 genes examined) whose level of translational initiation was less affected by the change in carbon source. This group corresponded to the genes upregulated on diauxic shift, emphasizing the importance to the cell of mechanisms that ensure the translation of newly expressed genes.

Reliability of microarray data

High-level mathematics has taught us that it is important to prevent the introduction of systematic bias when working with large numbers of variables. This concept holds true for microarray data analysis. Errors may be introduced at many points between the production of microarrays and the analysis of hybridization signals. Two critical steps that may strongly affect the results are the isolation of RNA and the generation of fluorescently

labelled probes. RNA purification is inherently more difficult for bacterial than mammalian systems. The absence of polyadenylated mRNA means that cDNA labelling must be performed with total RNA, only approximately 3% of which is mRNA. Furthermore, bacterial mRNA is much less stable than eukaryotic mRNA. The half-life of mRNA molecules in *E. coli* can be as short as 30 s (Carpousis *et al.*, 1999). Since differential mRNA instability is an important mechanism in the control of gene expression, great care must be taken to obtain quality RNA that has not been degraded. The rapid stabilization of RNA by addition of chaotropic agents such as guanidinium isothiocyanate is one important tool (Cox, 1968). Alternative commercial products such as RNALater (Ambion) perform a similar function. The initial stabilization of bacterial RNA is critical; otherwise one is in danger of studying cold-shock genes induced during centrifugation of bacteria at 4 °C prior to RNA extraction! The cDNA synthesis step is also critical because the cDNA probe must reflect a representative population of labelled mRNA species. It has recently been shown that the reverse transcription of *E. coli* RNA using a pool of primers specific to the 3' regions of all mRNA molecules had a significant underrepresentation of about 30% of mRNAs when compared to the use of random hexamers (Arfin *et al.*, 2000). Another problem that can be associated with labelling is that different fluorescent dyes do not have the same incorporation rate during labelling. This can be controlled by performing 'dye swap' experiments (Wei *et al.*, 2001), but we recommend the 'indirect' labelling approach to avoid this problem (<http://www.microarrays.org>).

We have described some of the technical problems commonly encountered with microarrays. It is important to remember that the use of microarrays for gene expression profiling is a recent development, and some aspects of the approach are not completely understood. Therefore, important results obtained with microarrays must be confirmed with other techniques, such as real-time quantitative PCR or Northern blotting, until microarray-based methodologies are completely validated.

To build or to buy, that is the question

Many laboratories and research centres are considering whether to invest in microarray technology or to obtain pre-printed microarrays from commercial sources. A number of macroarrays (membrane-based arrays) and microarrays (glass slides) of interest to microbiologists are already available. Currently, the variety of arrayed microbial genomes on the market is restricted to relatively few organisms (Table 2), though on-going genome sequencing projects will rapidly expand this selection in the near future. However, current pricing is set at a high level (between £500 and £1000 per microarray), reflecting a lack of competition.

A common misconception is that gene expression profiling experiments will only require a small number

Table 2. Commercial 'pre-printed' microarrays (currently or soon to be available)

Micro-organism	Supplier
<i>Bacillus subtilis</i>	Sigma-Genosys, Eurogentech
<i>Candida albicans</i>	Eurogentech, Incyte
<i>Escherichia coli</i>	Sigma-Genosys, TaKaRa Biomedicals
<i>Helicobacter pylori</i>	Eurogentech
<i>Neisseria meningitidis</i>	Eurogentech
<i>Saccharomyces cerevisiae</i>	Corning, Eurogentech, Operon
<i>Staphylococcus aureus</i>	Incyte
<i>Streptococcus pneumoniae</i>	Eurogentech

of microarrays. Researchers must expect to use a significant number of microarrays to produce reliable data. Unfortunately, unlike radioactively labelled macroarrays, fluorescently labelled microarrays cannot be reused. A typical microarray experiment involving four time points performed in triplicate will require 12 microarrays for a single experiment, i.e. approximately £6000 at current prices plus the cost of labelling consumables. The majority of papers that have used microarray technology in the past year have described results from single experiments, without replicates or any indication of statistical significance. Such data require the 'suspension of disbelief' by the reader, and are unlikely to be acceptable in the future.

The scale of replication required to yield significant data could prove to be a significant barrier to the widespread application of microarray technology. And it is not yet possible for every medium-sized lab to design and print its own microarray slides. 'In-house' microarray technology is still expensive (especially at the level of consumables) and labour-intensive. The equipment for making and analysing microarrays is readily available, at a price. But fierce competition is pushing some companies to release machines before they are completely optimized. We would argue that it is worthwhile to rely on the robust homemade Stanford technology, which is responsible for the majority of microarray publications to date (Thompson *et al.*, 2001). The significant investment now being made in genomic and post-genomic centres throughout Europe should allow researchers at all levels to pursue functional genomic approaches, either independently or through collaboration, without needing to set-up 'in-house' facilities. Clearly, financial constraints can be overcome: Oh & Liao (2000) successfully used a small 'subarray' of 111 *E. coli* genes involved in central metabolism to investigate metabolic flux.

Caveat emptor!

We must emphasize that the use of microarrays for molecular microbiology is still not straightforward. The generation of reliable data requires an extremely rigorous approach, both at the technical and the microbiological levels. Appropriate experimental design is also essential. Simple basic errors, such as choice of the

wrong media or the stage of the growth phase used to compare mutant and wild-type strains, can dramatically affect gene expression! Changes in gene expression can be very transient. A 10 min exposure of yeast to 0.4 M NaCl resulted in the induction of 1300 genes, whereas only 172 induced genes were detected after 20 min (Posas *et al.*, 2000). Learning to use microarrays often takes months rather than weeks, and requires the support of an experienced laboratory. Microarray experiments should not be entered into lightly!

The future

Genomotyping should revolutionize our ability to distinguish bacteria. The combination of other 'chip technologies' with genomotyping has already produced a prototype capable of separating *E. coli* from blood and performing subsequent microarray analysis (Cheng *et al.*, 1998). Whether or not this will prove applicable to the diagnostics market depends upon cost considerations, and whether the technology can be made sufficiently robust to perform in the environment of a microbiology laboratory. We should remember that PCR promised to be a sensitive diagnostic tool, but has not led to many validated diagnostic approaches.

A significant area that needs to be investigated is the utility of microarrays for analysis of mixed bacterial communities. The application of gene expression profiling or genomotyping to obtain information about individual species within a natural community would prove invaluable for microbial ecology and for microbial systematics alike. Assuming that appropriate hybridization stringencies are employed, and given a sufficient microbial diversity within the population of interest, there is no theoretical reason for this approach to fail.

Microarrays can also be used to gain clues to gene function through looking at knockout mutants, particularly of predicted regulatory genes. This approach has already been successfully used by Winzeler *et al.* (1999b) to follow the growth of pools of 500 yeast knockout mutants under various environmental conditions. Each mutant was tagged with a unique oligonucleotide sequence (a 'molecular barcode') that was detected by hybridization to a custom-built microarray to determine growth conditions when certain mutants

were unable to grow. This methodology combined with a massive parallel analysis of mapped mutants (Ross-Macdonald *et al.*, 1999; Spradling *et al.*, 1999) offers a rapid route to determining the function of the FUN genes found in every microbial genome (Hinton, 1997).

The application of microbial gene expression profiling is only limited by our imagination! Bacteria have been used for decades as sensitive biosensors for mutagenicity (Maron & Ames, 1983), and this approach has recently been brought up to date. *E. coli* has been used to determine the effects of microwave radiation produced by mobile telephones. Macroarray analysis demonstrated that 13 genes were induced by a 2 h exposure to a commercial mobile telephone (A. Morby, personal communication). As we integrate the power of microarray analyses with our particular research interests, more creative applications are bound to arise.

We are moving from the period of genomics towards the post-genomic future and we are entering what is arguably the most exciting period in the history of microbiology. At last we have the potential to ask questions at a relevant scale, that of the whole genome and hence the whole organism. We are optimistic that swift progress will be made as we learn to implement microarray technology more effectively, and we look forward to the time when innovative ideas can be tested extremely quickly.

References

- Aach, J., Rindone, W. & Church, G. M. (2000). Systematic management and analysis of yeast gene expression data. *Genome Res* **10**, 431–445.
- Alwine, J. C., Kemp, D. J. & Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzoyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci USA* **74**, 5350–5354.
- Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R. & Wahl, G. M. (1979). Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzoyloxymethyl paper. *Methods Enzymol* **68**, 220–242.
- Arfin, S. M., Long, A. D., Ito, E. T., Toller, L., Riehle, M. M., Paegle, E. S. & Hatfield, G. W. (2000). Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factors. *J Biol Chem* **275**, 29672–29684.
- Bammert, G. F. & Fostel, J. M. (2000). Genome-wide expression patterns in *Saccharomyces cerevisiae*: comparison of drug treatments and genetic alterations affecting biosynthesis of ergosterol. *Antimicrob Agents Chemother* **44**, 1255–1265.
- Barbosa, T. M. & Levy, S. B. (2000). Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. *J Bacteriol* **182**, 3467–3474.
- Behr, M. A., Wilson, M. A., Gill, W. P., Salamon, H., Schoolnik, G. K., Rane, S. & Small, P. M. (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**, 1520–1523.
- Belcher, C. E., Drenkow, J., Kehoe, B., Gingeras, T. R., McNamara, N., Lemjabbar, H., Basbaum, C. & Relman, D. A. (2000). The transcriptional responses of respiratory epithelial cells to *Bordetella pertussis* reveal host defensive and pathogen counter-defensive strategies. *Proc Natl Acad Sci USA* **97**, 13847–13852.
- Bernstein, B. E., Tong, J. K. & Schreiber, S. L. (2000). Genomewide studies of histone deacetylase function in yeast. *Proc Natl Acad Sci USA* **97**, 13708–13713.
- Brent, R. (1999). Functional genomics: learning to think about gene expression data. *Curr Biol* **9**, R338–R341.
- Brocklehurst, K. R. & Morby, A. P. (2000). Metal-ion tolerance in *Escherichia coli*: analysis of transcriptional profiles by gene-array technology. *Microbiology* **146**, 2277–2282.
- Carpousis, A. J., Vanzo, N. F. & Raynal, L. C. (1999). mRNA degradation – a tale of poly(A) and multiprotein machines. *Trends Genet* **15**, 24–28.
- Cheng, J., Sheldon, E. L., Wu, L., Uribe, A., Gerrue, L. O., Carrino, J., Heller, M. J. & O'Connell, J. P. (1998). Preparation and hybridization analysis of DNA/RNA from *E. coli* on micro-fabricated bioelectronic chips. *Nature Biotechnol* **16**, 541–546.
- Cho, R. J., Campbell, M. J., Winzler, E. A. & 7 other authors (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* **2**, 65–73.
- Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O. & Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast *Science* **282**, 699–705 (and Erratum *Science* **282**, 1421).
- Chuang, S. E., Daniels, D. L. & Blattner, F. R. (1993). Global regulation of gene expression in *Escherichia coli*. *J Bacteriol* **175**, 2026–2036.
- Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. (2000a). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet* **26**, 183–186.
- Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinot-Chazel, C., Galiegue, S., Hadibi, E. H. & Casellas, P. (2000b). Monitoring cellular responses to *Listeria monocytogenes* with oligonucleotide arrays. *J Biol Chem* **275**, 11181–11190.
- Cox, R. A. (1968). The use of guanidinium chloride in the isolation of nucleic acids. *Methods Enzymol* **12**, 120–129.
- Cummings, C. A. & Relman, D. A. (2000). Using DNA microarrays to study host–microbe interactions. *Emerg Infect Dis* **6**, 513–525.
- DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. & Trent, J. M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet* **14**, 457–460.
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686.
- DeRisi, J., van den Hazel, B., Marc, P., Balzi, E., Brown, P., Jacq, C. & Goffeau, A. (2000). Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. *FEBS Lett* **470**, 156–160.
- Dessus-Babus, S., Knight, S. T. & Wyrick, P. B. (2000). Chlamydial infection of polarized HeLa cells induces PMN chemotaxis but the cytokine profile varies between disseminating and non-disseminating strains. *Cell Microbiol* **2**, 317–327.
- Diehn, M., Eisen, M. B., Botstein, D. & Brown, P. O. (2000). Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nature Genet* **25**, 58–62.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. & Coleman, P. (1992). Analysis of gene expression in single live neurons. *Proc Natl Acad Sci U S A* **89**, 3010–3014.
- Eckmann, L., Smith, J. R., Housley, M. P., Dwinell, M. B. & Kagnoff, M. F. (2000). Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in

response to infection with the invasive enteric bacteria *Salmonella*. *J Biol Chem* **275**, 14084–14094.

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

Favis, R., Day, J. P., Gerry, N. P., Phelan, C., Narod, S. & Barany, F. (2000). Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. *Nature Biotechnol* **18**, 561–564.

Ferea, T. L., Botstein, D., Brown, P. O. & Rosenzweig, R. F. (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci USA* **96**, 9721–9726.

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

Giaever, G., Shoemaker, D. D., Jones, T. W., Liang, H., Winzeler, E. A., Astromoff, A. & Davis, R. W. (1999). Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nature Genet* **21**, 278–283.

Gingeras, T. R., Ghandour, G., Wang, E. & 6 other authors (1998). Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res* **8**, 435–448.

Gray, N. S., Wodicka, L., Thunnissen, A. M. & 10 other authors (1998). Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **281**, 533–538.

Groisman, E. A. & Ochman, H. (1997). How *Salmonella* became a pathogen. *Trends Microbiol* **5**, 343–349.

Gross, C., Kelleher, M., Iyer, V. R., Brown, P. O. & Winge, D. R. (2000). Identification of the copper regulon in *Saccharomyces cerevisiae* by DNA microarrays. *J Biol Chem* **275**, 32310–32316.

Gupta, R. S. (2000). The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev* **24**, 367–402.

Hautefort, I. & Hinton, J. C. (2000). Measurement of bacterial gene expression in vivo. *Philos Trans R Soc Lond B Biol Sci* **355**, 601–611.

Hemming, S. A., Jansma, D. B., Macgregor, P. F., Goryachev, A., Friesen, J. D. & Edwards, A. M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J Biol Chem* **275**, 35506–35511.

Hinton, J. C. (1997). The *Escherichia coli* genome sequence: the end of an era or the start of the FUN? *Mol Microbiol* **26**, 417–422.

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

Hooper, L. V., Wong, M. H., Thelin, A., Hansson, L., Falk, P. G. & Gordon, J. I. (2001). Molecular analysis of commensal host–microbial relationships in the intestine. *Science* **291**, 881–884.

Hughes, T. R., Marton, M. J., Jones, A. R. & 19 other authors (2000a). Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126.

Hughes, T. R., Roberts, C. J., Dai, H. & 9 other authors (2000b). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genet* **25**, 333–337.

Ichikawa, J. K., Norris, A., Banger, M. G., Geiss, G. K., van't Wout, A. B., Bumgarner, R. E. & Lory, S. (2000). Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. *Proc Natl Acad Sci USA* **97**, 9659–9664.

Jelinsky, S. A. & Samson, L. D. (1999). Global response of

Saccharomyces cerevisiae to an alkylating agent. *Proc Natl Acad Sci USA* **96**, 1486–1491.

Jia, M. H., Larossa, R. A., Lee, J. M., Rafalski, A., Derose, E., Gonye, G. & Xue, Z. (2000). Global expression profiling of yeast treated with an inhibitor of amino acid biosynthesis, sulfometuron methyl. *Physiol Genomics* **3**, 83–92.

Kell, D. B. & King, R. D. (2000). On the optimization of classes for the assignment of unidentified reading frames in functional genomics programmes: the need for machine learning. *Trends Biotechnol* **18**, 93–98.

Khodursky, A. B., Peter, B. J., Cozzarelli, N. R., Botstein, D., Brown, P. O. & Yanofsky, C. (2000a). DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* **97**, 12170–12175.

Khodursky, A. B., Peter, B. J., Schmidt, M. B., DeRisi, J., Botstein, D., Brown, P. O. & Cozzarelli, N. R. (2000b). Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc Natl Acad Sci USA* **97**, 9419–9424.

Kuhn, K. M., DeRisi, J. L., Brown, P. O. & Sarnow, P. (2001). Global and specific translational regulation in the genomic response of *Saccharomyces cerevisiae* to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol Cell Biol* **21**, 916–927.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O. & Davis, R. W. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci USA* **94**, 13057–13062.

Laub, M. T., McAdams, T. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148.

Lelivelt, M. J. & Culbertson, M. R. (1999). Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol Cell Biol* **19**, 6710–6719.

Lemieux, B., Aharoni, A. & Schena, M. (1998). Overview of DNA chip technology. *Mol Breed* **4**, 277–289.

ter Linde, J. J., Liang, H., Davis, R. W., Steensma, H. Y., van Dijken, J. P. & Pronk, J. T. (1999). Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of *Saccharomyces cerevisiae*. *J Bacteriol* **181**, 7409–7413.

Lockhart, D. J., Dong, H., Byrne, M. C. & 8 other authors (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol* **14**, 1675–1680.

Lopez, M. C. & Baker, H. V. (2000). Understanding the growth phenotype of the yeast *gcr1* mutant in terms of global genomic expression patterns. *J Bacteriol* **182**, 4970–4978.

Luo, L., Salunga, R. C., Guo, H. & 8 other authors (1999). Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* **5**, 117–22 (and Erratum *Nature Med* **5**, 355).

Lyons, T. J., Gasch, A. P., Gaither, L. A., Botstein, D., Brown, P. O. & Eide, D. J. (2000). Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc Natl Acad Sci USA* **97**, 7957–7962.

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

Maron, D. M. & Ames, B. N. (1983). Revised methods for the *Salmonella* mutagenicity test. *Mutat Res* **113**, 173–215.

Marton, M. J., DeRisi, J. L., Bennett, H. A. & 11 other authors (1998). Drug target validation and identification of secondary

- drug target effects using DNA microarrays. *Nature Med* **4**, 1293–301.
- Nelson, K. E., Paulsen, I. T., Heidelberg, J. F. & Fraser, C. M. (2000).** Status of genome projects for nonpathogenic bacteria and archaea. *Nature Biotechnol* **18**, 1049–1054.
- Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000).** Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
- Oh, M. K. & Liao, J. C. (2000).** Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnol Prog* **16**, 278–286.
- Pennisi, E. (2000).** Microbial genomics. Culling genes early yields rich harvest. *Science* **287**, 1572–1573.
- Posas, F., Chambers, J. R., Heyman, J. A., Hoeffler, J. P., de Nadal, E. & Arino, J. (2000).** The transcriptional response of yeast to saline stress. *J Biol Chem* **275**, 17249–17255.
- Primig, M., Williams, R. M., Winzeler, E. A., Tevzadze, G. G., Conway, A. R., Hwang, S. Y., Davis, R. W. & Esposito, R. E. (2000).** The core meiotic transcriptome in budding yeasts. *Nature Genet* **26**, 415–423.
- Py, B., Higgins, C. F., Krisch, H. M. & Carpousis, A. J. (1996).** A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* **381**, 169–172.
- Ren, B., Robert, F., Wyrick, J. J. & 11 other authors (2000).** Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309.
- Richmond, C. S., Glasner, J. D., Mau, R., Jin, H. & Blattner, F. R. (1999).** Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res* **27**, 3821–3835.
- Riehle, M. M., Bennett, A. F. & Long, A. D. (2001).** Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc Natl Acad Sci USA* **9**, 525–530.
- Rimini, R., Jansson, B., Feger, G. & 8 other authors (2000).** Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high density DNA arrays. *Mol Microbiol* **36**, 1279–1292.
- Rosamond, J. & Allsop, A. (2000).** Harnessing the power of the genome in the search for new antibiotics. *Science* **287**, 1973–1976.
- Rosenberger, C. M., Scott, M. G., Gold, M. R., Hancock, R. E. & Finlay, B. B. (2000).** *Salmonella typhimurium* infection and lipopolysaccharide stimulation induce similar changes in macrophage gene expression. *J Immunol* **164**, 5894–5904.
- Ross-Macdonald, P., Coelho, P. S., Roemer, T. & 15 other authors (1999).** Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418.
- de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T. J., Keck, W., Amrein, K. E. & Lange, R. (2000).** Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol* **182**, 4696–4703.
- Salama, N., Guillemin, K., McDaniel, T. K., Sherlock, G., Tompkins, L. & Falkow, S. (2000).** A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* **97**, 14668–14673.
- Sarkar, N. (1996).** Polyadenylation of mRNA in bacteria. *Microbiology* **142**, 3125–3133.
- Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995).** Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
- Shalon, D., Smith, S. J. & Brown, P. O. (1996).** A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639–645.
- Southern, E. M. (1975).** Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503–517.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. & Futcher, B. (1998).** Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**, 3273–3297.
- Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverly, T., Mozden, N., Misra, S. & Rubin, G. M. (1999).** The Berkeley Drosophila Genome Project gene disruption project: single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135–177.
- Sudarsanam, P., Iyer, V. R., Brown, P. O. & Winston, F. (2000).** Whole-genome expression analysis of *snf/swi* mutants of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **97**, 3364–3369.
- Talaat, A. M., Hunter, P. & Johnston, S. A. (2000).** Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nature Biotechnol* **18**, 679–682.
- Tang, Y. & Guest, J. R. (1999).** Direct evidence for mRNA binding and post-transcriptional regulation by *Escherichia coli* aconitases. *Microbiology* **145**, 3069–3079.
- Tao, H., Bausch, C., Richmond, C., Blattner, F. R. & Conway, T. (1999).** Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181**, 6425–6440.
- Thompson, A., Lucchini, S. & Hinton, J. C. D. (2001).** It's easy to build your own microarrayer! *Trends Microbiol* **9**, 154–156.
- Troesch, A., Nguyen, H., Miyada, C. G., Desvarenne, S., Gingeras, T. R., Kaplan, P. M., Cros, P. & Mabilat, C. (1999).** *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. *J Clin Microbiol* **37**, 49–55.
- Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E. J., Hieter, P., Vogelstein, B. & Kinzler, K. W. (1997).** Characterization of the yeast transcriptome. *Cell* **88**, 243–251.
- Vishwanath, R. I., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. & Brown, P. O. (2001).** Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538.
- Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T. & Marincola, F. M. (2000).** High-fidelity mRNA amplification for gene profiling. *Nature Biotechnol* **18**, 457–459.
- Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A. & 7 other authors (1995).** Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**, 1090–1094.
- Wei, Y., Lee, J. M., Richmond, C., Blattner, F. R., Rafalski, J. A. & LaRossa, R. A. (2001).** High-density microarray-mediated gene expression profiling of *Escherichia coli*. *J Bacteriol* **183**, 545–556.
- Wilson, M., DeRisi, J., Kristensen, H. H., Imboden, P., Rane, S., Brown, P. O. & Schoolnik, G. K. (1999).** Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* **96**, 12833–12838.
- Winzeler, E. A., Richards, D. R., Conway, A. R. & 8 other authors (1998).** Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197.

- Winzeler, E. A., Lee, B., McCusker, J. H. & Davis, R. W. (1999a). Whole genome genetic-typing in yeast using high-density oligonucleotide arrays. *Parasitology* **118** Suppl., S73–S80.
- Winzeler, E. A., Shoemaker, D. D., Astromoff, A. & 49 other authors (1999b). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. & Lockhart, D. J. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol* **15**, 1359–1367.
- Wolfsberg, T. G., Gabrielian, A. E., Campbell, M. J., Cho, R. J., Spouge, J. L. & Landsman, D. (1999). Candidate regulatory sequence elements for cell cycle-dependent transcription in *Saccharomyces cerevisiae*. *Genome Res* **9**, 775–792.
- Wyrick, J. J., Holstege, F. C., Jennings, E. G., Causton, H. C., Shore, D., Grunstein, M., Lander, E. S. & Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418–421.
- Ye, R. W., Tao, W., Bedzyk, L., Young, T., Chen, M. & Li, L. (2000). Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J Bacteriol* **182**, 4458–4465.
- Yun, C. W., Ferea, T., Rashford, J., Ardon, O., Brown, P. O., Botstein, D., Kaplan, J. & Philpott, C. C. (2000). Desferrioxamine-mediated iron uptake in *Saccharomyces cerevisiae*. Evidence for two pathways of iron uptake. *J Biol Chem* **275**, 10709–10715.
- Zhang, M. Q. (1999). Promoter analysis of co-regulated genes in the yeast genome. *Comput Chem* **23**, 233–250.
- Zweiger, G. (1999). Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends Biotechnol* **17**, 429–436.