Journal of Industrial Microbiology & Biotechnology (2002) 28, 180–185

www.nature.com/jim

MICROARRAY TECHNOLOGY
GEM™ Microarrays and drug discovery

Mark A Reynolds

Incyte Genomics, 6519 Dumbarton Circle, Fremont, CA 94555, USA

Incyte Genomics' GEM™ Gene Expression Microarray is a proven genomics tool used by a large number of pharmaceutical companies to speed up the drug discovery and development process. The development and integration of this technology, together with Incyte's sequence databases and clone resources, have resulted in GEM microarrays that span approximately 60,000 human genes as well as approximately 60,000 plant, rat, mouse, yeast, and bacterial genes. The technology underlying the use of these arrays and their application to the drug discovery process is highlighted.
Journal of Industrial Microbiology & Biotechnology (2002) **28**, 180–185 DOI: 10.1038/sj/jim/7000136

Keywords: microarray; gene expression; expression database; drug discovery

Introduction

Glass cDNA microarrays were first developed in the laboratory of Patrick O. Brown at Stanford University [3,7–9]. The idea was simple, yet elegant: If one could apply an array of polymerase chain reaction (PCR)-amplified cDNA clones (cDNAs) onto a glass surface at high density, one could assay the expression of large numbers of genes in a single hybridization experiment. This broke the common research paradigm where low numbers of specified gene candidates are studied through conventional blotting or PCR-based technologies. It now became possible to conduct massively parallel hybridization experiments, e.g., to identify novel genes or to associate genes within complex gene pathways. The first published example of a glass cDNA microarray included 48 genes from the small flowering plant, *Arabidopsis thaliana*, wherein it was estimated that the same printing technology could be used to conduct even larger-scale expression studies (up to 20,000 cDNAs per array) [7]. Several larger-scale studies (approximately 1000 cDNAs per array) were published the following year with yeast cDNAs [9] and human cDNAs of both verified and unverified sequences [3,8]. The first published genome-wide expression study from the yeast organism, *Saccharomyces cerevisiae* (comprising 6200 cDNAs), also utilized this technology [4].

cDNA microarray analyses are commonly conducted in a two-channel format. In this approach, RNA samples from two different tissues are labeled separately by reverse transcription with the incorporation of two different fluorescent dyes, typically cyanine-3 and cyanine-5. The resulting two probe populations are then pooled and hybridized together on the microarray. After hybridization, the microarrays are detected using a dual-laser scanner. This provides the ability to determine which genes are expressed differentially between the two tissues, a process that will be described in more detail in the *GEM Microarray Process* section.

In 1996, a biotechnology company named Synteni was established to develop the glass cDNA microarray approach as a commercial process. By accessing the UniGene database from the NCBI, Synteni successfully produced the first 10,000 human gene microarrays in 1997. Recognizing the potential of this approach for generating database quality gene expression information, Incyte Genomics acquired Synteni in January 1998. The first 10,000 human gene microarrays comprising proprietary cDNA clones (LifeGEM™ 1.0) were released in May 1998. Since that time, over 26 unique GEM microarray products have been released.

Over the past 4 years, Incyte's GEM microarray facility has been operated as a service business to provide high-quality expression data to customers and academic collaborators. Microarray operations have also been accelerated to provide compendiums of gene expression data for our LifeExpress™ database products. In the course of developing this business, it has been necessary to optimize and standardize every step of the process, from PCR amplification to hybridization and data analysis. Standard control measures were developed to ensure data quality and reproducibility. This review will summarize our current GEM microarray process and how it is being applied to elucidate disease-relevant gene pathways and identify new gene targets for therapeutic intervention.

GEM microarray process

Microarray fabrication

PCR-amplified cDNA clones can be chosen from sequence databases or random libraries. In our standard GEM microarray products, annotated genes are nominated and representative Incyte clones are selected using our proprietary LifeSeq® databases. Many of our customers also provided their own proprietary clones for custom microarray manufacturing. The resulting amplified cDNA targets are spotted robotically using proprietary printing technologies. Glass slides are manufactured on-site, which are chemically modified to promote DNA adhesion and allow *ca.* 100 pl droplets to be spotted at high density without merging. Typically, 10,000 cDNA elements are spotted per microarray (approximately 1.4×1.4 cm^2 area).

Probe labeling

Incyte has developed a novel mRNA labeling system that requires only 200 ng of PolyA$^+$ selected RNA without amplification. The

Correspondence: Dr. Mark A. Reynolds, Incyte Genomics, 6519 Dumbarton Circle, Fremont, CA 94555, USA
Received 16 November 2000; accepted 1 March 2001

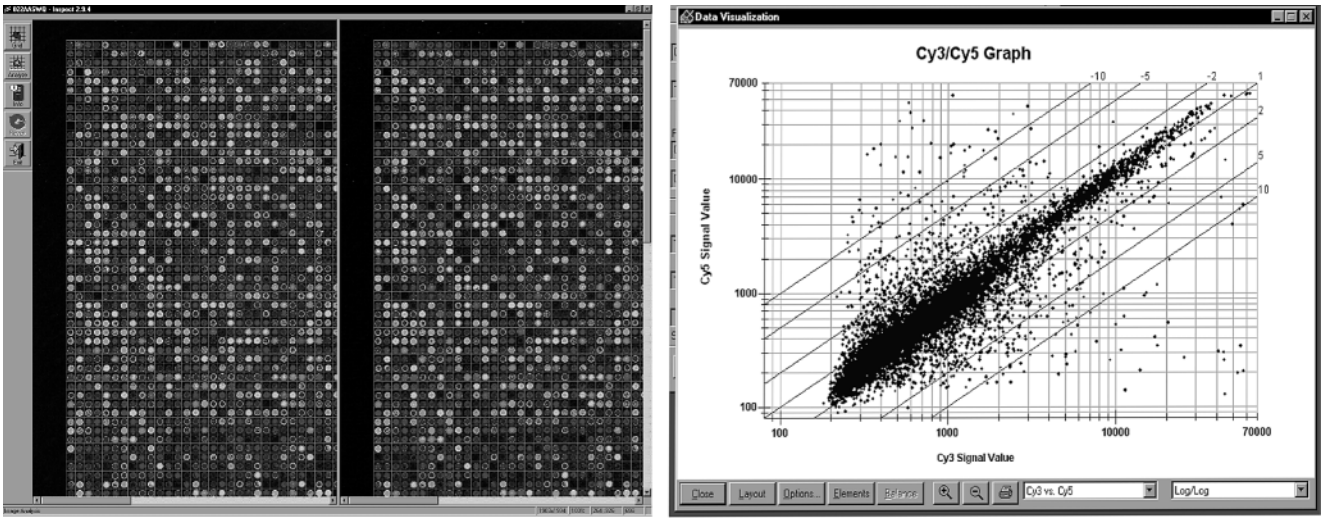

Figure 1 Examples of viewer windows available in Incyte's GEMTools™ microarray analysis software. Left: Scanned image of the gridded array. Right: Scatter plot showing genes with no differential expression (along the diagonal) and differentially expressed genes (above and below the diagonal).

labeling kit includes internal controls (e.g., RNA transcripts and labeled cDNAs derived from non-coding yeast control fragments) for quality control of the final hybridized GEM. These controls ensure, e.g., that the input RNA was of sufficient quality and that the reverse transcription reactions proceeded correctly (see below). In the standard production process, two different mRNA samples (e.g., matched normal *versus* diseased tissues) are labeled individually with either cyanine-3 or cyanine-5 fluorescent dyes. The resulting fluorescently labeled cDNAs are pooled, co-purified to remove unreacted dye and primers, precipitated, and then resuspended in hybridization buffer.

Hybridization and detection

The pooled fluorescent cDNA probe populations are hybridized competitively onto the GEM microarray in a custom-built chamber that is designed to ensure uniform heating and cooling and also to minimize evaporation. The GEM microarrays are then washed sequentially in low- to high-stringency buffers, dried, and scanned in a dual laser scanner (Axon Instruments, Union City, CA). The fluorescent intensities are corrected for local background and balanced using proprietary image analysis software. Signals are then filtered to remove values falling below signal-to-background and element area percentage values of 2.5% and 40%, respectively. The ratio of the cyanine-3 and cyanine-5 signal intensities provides a measurement of the relative expression levels between the two samples. This process routinely provides 1:100,000 sensitivity across most genes starting from 200 ng of input mRNA sample. Twofold differences in expression can be detected between the two mRNA samples over about a $3-\log_{10}$ dynamic range.

Laboratory Information Management System (LIMS) and data analysis

GEMTools™ is an interactive software package that enables the researcher to examine and query their data. A false color scale is used to describe the signal intensities between the two samples (blue=lowest signals<green<yellow<red<white=highest signals). This helps the user to visually compare expression differences for genes of interest, represented either according to the original 96-well plates from which they were arrayed, or in tabular format. A typical scanned image and scatter graph from GEMTools is shown in Figure 1. The scatter graph represents expression ratios between the two tissues for all positive scoring target elements on the array. Data points falling along the diagonal correspond to genes exhibiting no significant differential expression (<twofold), whereas data points falling above or below the diagonal correspond to differentially expressed genes. Individual data points can be selected to provide more information about the gene, including access through the web to outside databases. These data can also be exported to other commercial applications for graphical or statistical analysis. GEMTools is also used internally at Incyte to manage data corresponding to all aspects of the production process, including GEM microarray fabrication, probe generation, hybridization, and data analysis.

Process and quality control

Process control

Figure 2 illustrates the current production process and the various control measures that are in place. As mentioned above, chemically modified glass slides are produced in-house according to proprietary silylation chemistry. Standard physical assays are used to monitor the uniformity and surface energy of these slides. Functional DNA binding assays were developed to establish specifications for high-density printing (MA Reynolds, manuscript in preparation). The integrity of each PCR-amplified cDNA target is examined on an agarose gel and any failed PCR product (e.g., double bands, dropouts) is annotated out (i.e., the signals corresponding to these target elements are masked in GEMTools). Novel fluorescence-based assays have been developed to ensure that the cDNA targets are above an acceptable concentration range for printing and that the final arrayed elements have sufficient DNA for optimal hybridization performance [11]. Fluorescence-based and chromatographic assays have

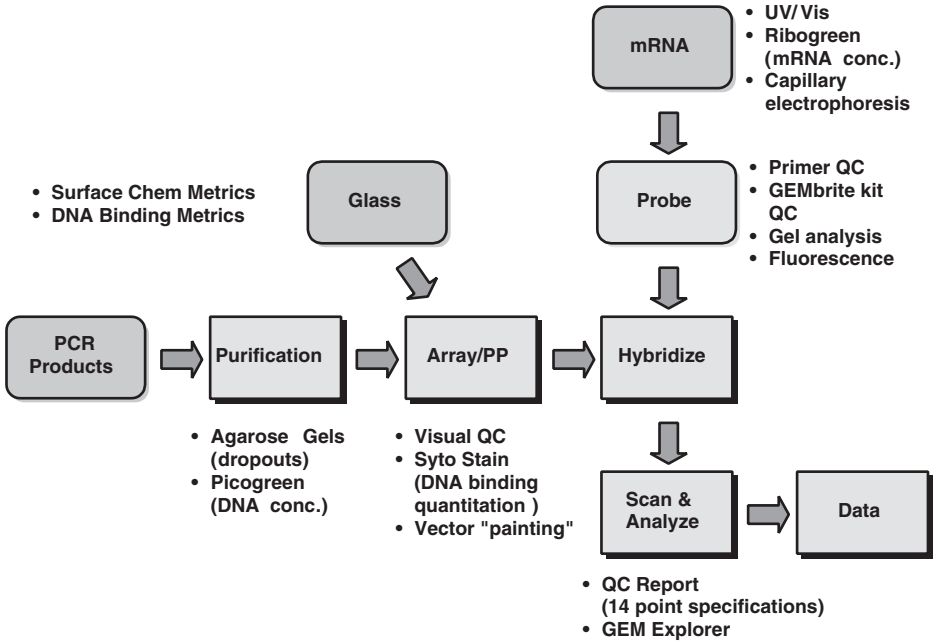

Figure 2 Schematic of Incyte's current GEM microarray production process. Quality control assays (such as the ones highlighted) ensure that each step of the process is followed according to standardized specifications that ensure data quality and reproducibility.

also been developed to control the concentrations of mRNA samples and to assess their purity before labeling [11]. After hybridization, each scanned GEM microarray is analyzed for different signal parameters, including hybridization signals for the spiked-in yeast control fragments (see below). A quality control report is generated to assist in grading and passing the GEM microarray.

Control plate

Hybridization data for the yeast control fragments that are included as internal standards in the probe labeling reaction are presented in Figure 3. These include dilution series and ratio controls for sensitivity, linearity, and channel balance. Complex targets and housekeeping genes are also included on the control plate to assess the quality of the biological mRNA samples that were labeled for hybridization.

Process validation

An extensive validation study established the performance specifications of our current production process [11]. Replicate hybridizations with the same tissue in both channels (human placenta mRNA) showed system variability to be less than 1.4-fold. In order to examine channel bias, brain/placenta hybridizations were run with both dye combinations (i.e., brain mRNA was labeled with cyanine-3 in one data set and with cyanine-5 in the other). A comparison of differential expression ratios between these two data sets showed almost perfect axial symmetry (channel bias less than one part in 1000). A compilation of all yeast control fragment signals showed a coefficient of variance of 12%, and a sensitivity of 2 pg input RNA transcript at 2.5-fold signal-to-background. This corresponds to a sensitivity of about one copy of mRNA transcript in 10^7 cells. These results demonstrate that the GEM microarrays deliver a high level of sensitivity, dynamic range, and reproducibility.

Biological applications

Incyte offers two different types of human GEM microarray products. UniGEM V™ 2.0 contains 9182 verified clones representing 8616 unique genes/clusters and 8481 annotated genes/clusters from the public domain. Incyte clones were mapped to the UniGene database using BLAST and Smith–Waterman alignment analysis. For each gene of interest, a representative clone was selected containing the 5′-most information for inclusion on the microarray. The 5′-end of each clone is covered by a UniGene cluster sequence and each clone is sequence-verified. Human Genome GEM™ microarrays 1–5 are designed to provide the "best view" of transcribed gene products by combining public and proprietary content. Each Human Genome GEM microarray contains a unique set of more than 9600 verified clones representing both publicly available genes and Incyte proprietary ESTs. To reduce redundancy, nearly all of the clones selected correspond to genes with identified 3′-ends. These GEM microarrays are ideal for discovery of novel drug targets or previously uncharacterized genes. In addition, Incyte offers products for studies in animal models (Mouse GEM™ 2 and Rat GEM™ 3). Taken together, these represent one of the most comprehensive sets of microarray products available anywhere, creating exciting new possibilities for biological experimentation. For example, microarray experiments can now be conducted to identify new target genes for potential therapeutic intervention, and also to facilitate screening and optimization of new drug leads.

Gene expression profiling in invasive breast tumors

A large number of studies have been conducted to examine differential gene expression between diseased and non-diseased tissues. A typical example was conducted with breast infiltrating ductile carcinomas (unpublished results). mRNA was isolated from four non-diseased tissue samples and pooled for use as a

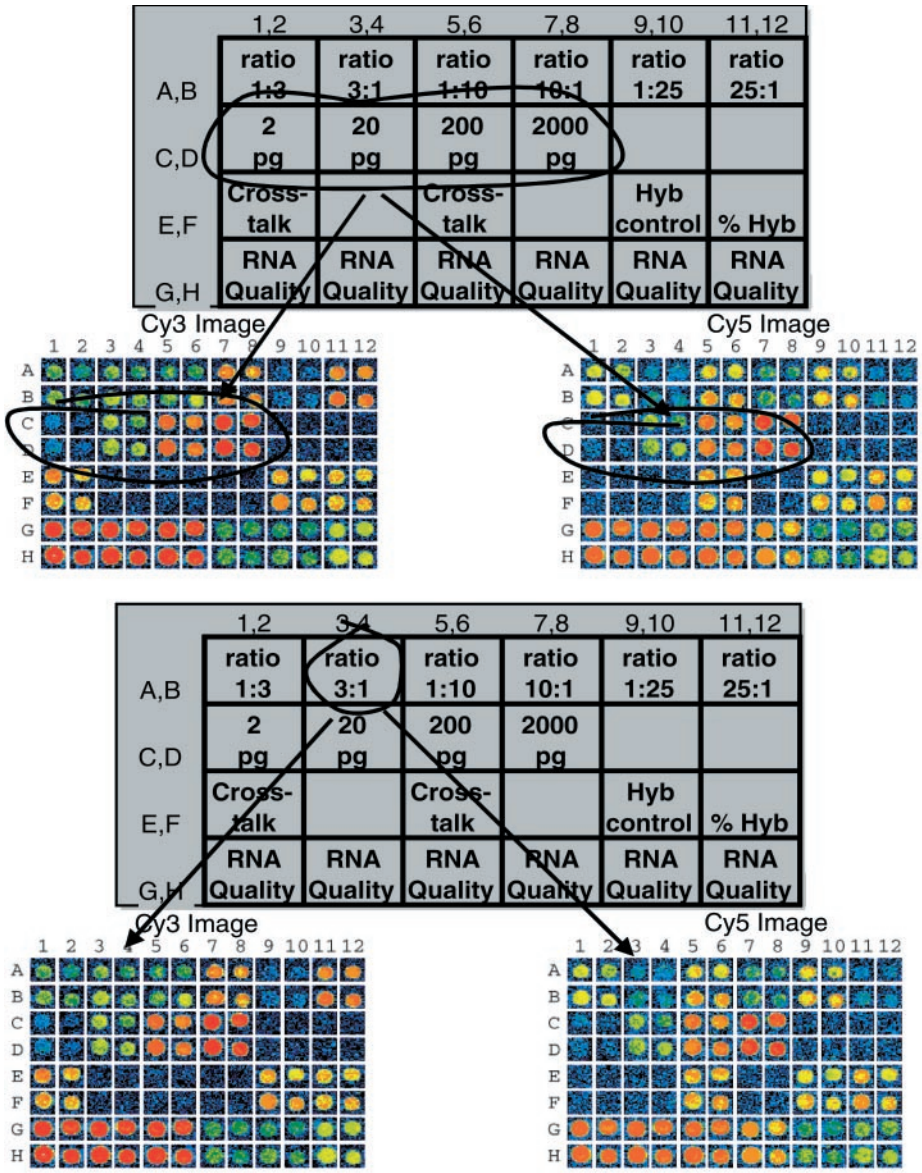

Figure 3 Example of a control plate used to assist Incyte's quality control scientists in grading and passing GEM microarrays. Top: Example of array elements corresponding to a dilution series of spiked-in yeast control fragment transcripts. Bottom: Example of array elements corresponding to yeast control fragment transcripts that are spiked-in at different ratios between the two channels.

common control (cyanine-5 channel). In a similar manner, mRNA was isolated from seven carcinomas and labeled individually (cyanine-3) for hybridization against the common control. Examples of genes that were downregulated in all diseased tissues included oncostatin, keratin, myosin, and a novel Incyte EST. Examples of genes that were upregulated included c-erb-B2 (HER2), fibronectin, and a T-cell receptor. This result is consistent with the HER2 gene's now-established role in breast cancer.

Toxicity screening

Microarrays with toxic marker genes are being developed as an early indicator of clinical safety [1]. A study conducted in collaboration with Tularik scientists illustrates this approach [2]. Tularik was interested in discovering new small molecule drugs against a particular target receptor. Using an *in vitro* screening assay coupled with microarray analysis, they found that their first lead candidate affected the expression of a number of other genes that had been associated with known toxic compounds. Subsequent rounds of medicinal chemistry and microarray screening enabled them to accelerate their drug discovery process rapidly without the need for expensive and time-consuming toxicity screening in animals.

Progressions from drug discovery to managed healthcare

Figure 4 summarizes the stages through which microarrays can be applied to accelerate the development of new and improved drug therapies. As illustrated in the previous examples, microarrays can facilitate the identification of new gene targets (sometimes referred

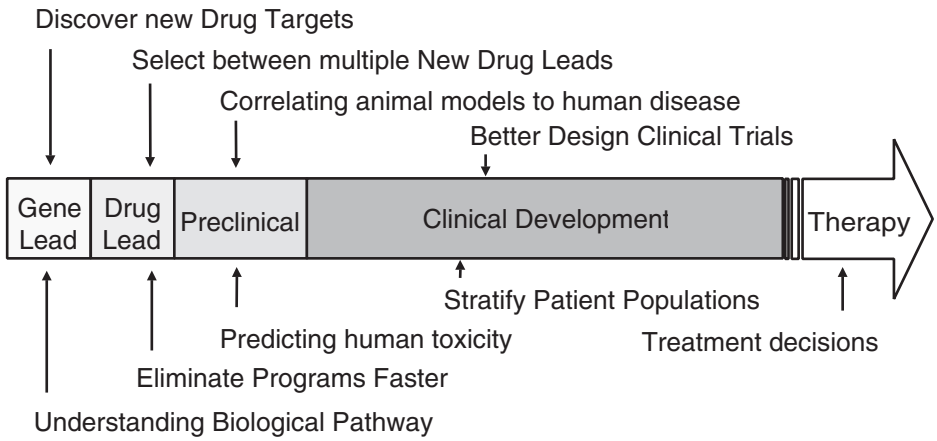

Figure 4 Diagram illustrating the various stages where microarrays can be used to facilitate drug discovery and development. Microarrays are envisaged to play a significant role in clinical diagnosis and drug therapy.

to as "drugable genes"), and can also accelerate selection of lead drug compounds by coupling *in vitro* screening with microarray analysis. Microarrays can also help accelerate preclinical drug development by profiling the expression of genes associated with known toxicity, and by correlating gene expression data from animal models to human diseased tissues [1]. In the near future, it may be possible to use microarrays as surrogate markers to provide an early indication of drug efficacy. It may also be possible to use microarray data to stratify patient populations for clinical studies. This could potentially be used to design more effective drug therapies at the outset of diagnosis and have a profound impact on the quality of healthcare with significant cost savings.

Gene expression databases

Microarrays are receiving increasing interest for their ability to generate compendiums of gene expression data [5,6,10]. These can be used, e.g., to associate genes within complex gene pathways, to dissect and characterize the biochemical processes that lead to a particular disease state, and to validate genes as drug targets based on a variety of experimental data.

Incyte is systematically populating a compendium of microarray expression data designed to facilitate the drug discovery process. Human and animal tissues (diseased and non-diseased) are being processed for GEM and proteomic analyses to enable correlation between gene and protein expression. The present database product offering, LifeExpress™, is the world's first integrated gene and protein expression database. LifeExpress Target contains comprehensive expression data for a variety of human disease areas, including cancer, cardiovascular, central nervous system, immunology/inflammation, and metabolic diseases (diabetes, obesity, and osteoporosis). LifeExpress Lead contains expression data for drug compounds tested against human and animal tissues. This second product enables investigators to associate data from animal models as predictors of human efficacy. Taken together, Incyte's LifeExpress databases can help elucidate the potential efficacy and toxicity of lead drug candidates, thereby accelerating the drug discovery and development process. This database is growing rapidly, with hundreds of new tissues being processed and analyzed weekly.

Conclusions

Incyte's GEM™ microarray technology has expanded rapidly over the past several years. Extensive quality control measures are in place to ensure a high level of precision and reproducibility. Presently, these microarrays are configured for the examination of up to 10,000 genes simultaneously in a single hybridization experiment. It is also possible to run hybridization experiments across multiple GEM microarrays, including up to 60,000 annotated human genes not available anywhere else. Presently, nine GEM microarray products are available to outside investigators for their biological experiments. Incyte has also launched a comprehensive gene and protein expression database called LifeExpress™ that facilitates the drug discovery process.

Acknowledgements

I thank Lyle Arnold, Tod Bedilion, and Rick Johnston for providing the data summarized in this review. I also thank Scott Eastman and Tom Theriault for many useful scientific discussions, and Bruce Wang and Zora Modrusan for critically reviewing this manuscript.

References

1 Bartosiewicz M, M Trounstine, D Barker, R Johnston and A Buckpitt. 2000. Development of a toxicological gene array and quantitative assessment of this technology. *Arch Biochem Biophys* 376: 66–73.
2 Braxton S and T Bedilion. 1998. The integration of microarray information in the drug development process. *Curr Opin Biotechnol* 9: 643–649.
3 DeRisi J, L Penland, PO Brown, ML Bittner, PS Meltzer, M Ray, Y Chen, YA Su and JM Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14: 457–460.
4 DeRisi JL, VR Iyer and PO Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686.
5 Eisen MB, PT Spellman, PO Brown and D Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863–14868.

6 Hughes TR, CJ Roberts, H Dai, AR Jones, MR Meyer, D Slade, J Burchard, S Dow, TR Ward, MJ Kidd, SH Friend and MJ Marton. 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* 25: 333–337.
7 Schena M, D Shalon, RW Davis and PO Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470.
8 Schena M, D Shalon, R Heller, A Chai, PO Brown and RW Davis. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 93: 10614–10619.
9 Shalon D, SJ Smith and PO Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* 6: 639–645.
10 Staudt LM and PO Brown. 2000. Genomic views of the immune system. *Annu Rev Immunol* 18: 829–859.
11 Yue HH, PS Eastman, BB Wang, J Minor, MH Doctolero, RL Nutall, R Stack, JW Beeker, JR Montgomery, VA Vainer and R Johnston. An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression. *Nucleic Acids Res* 29: e41–e50.