

This is a preprint of an article accepted for publication in *Cytometry*
© 2001 Wiley-Liss Inc.

High-Density Microarrays for Gene Expression Analysis

Michael K. Deyholos and David W. Galbraith*

Department of Plant Sciences, University of Arizona, Tucson, Arizona

*Correspondence to: David W. Galbraith, Ph.D., Professor, Dept. of Plant Sciences,
University of Arizona, 1145 E 4th St., Tucson, AZ 85721. E-mail: galbraith@arizona.edu

Original date of submission: 12 June, 1999

Accepted for publication: 9 September, 2000

Revised manuscript submitted: 1 December, 2000

Running title: Production and analysis of microarrays

Key terms: microarrays; expression profiling; cluster analysis; functional genomics;
high-throughput analysis

Submitted to: *Cytometry*

INTRODUCTION

Changes in gene expression are associated with many important biological phenomena, including morphogenesis and aging, cancer and disease states, and adaptive responses to the environment. To detect these changes, molecular biologists have, over the years, developed a variety of techniques, including RNA blots (1), differential display (2), ribonuclease protection assays (3), and reverse transcription-PCR (4). Although these methods are effective, being both sensitive and accurate, their dependence on gel electrophoresis places an empirical limit on the number of samples that can be simultaneously analyzed. High-density microarrays, first described by Schena *et al.* (5), provide the important advantage of allowing parallel quantification of the expression levels of an almost unlimited number of genes from a given genome. This greatly accelerates the characterization of individual genes and, at the same time, provides the means to decipher complex, multigenic patterns of gene expression (6).

The importance of the high-throughput capability of microarrays is particularly relevant in the context of the recent explosion in genomic sequence acquisition. Each day, the human genome sequencing project alone deposits an estimated two million bases of DNA sequence into the GenBank database (7). For other eukaryotes, including *Saccharomyces cerevisiae* (8), *Caenorhabditis elegans* (9), and *Drosophila melanogaster* (10) the complete genomic sequence has already been reported. It is expected that by the end of the year 2000, the list of organisms having fully-sequenced genomes will comprise at least 36 different species (11), including the higher plant *Arabidopsis thaliana*. The rapid growth in accumulation of genomic sequences has been paralleled by the growth in collections of expressed sequence tags (ESTs), each of which contain the partial sequence

of an expressed gene (12). For genomic or EST sequences to be truly informative, the genes that they identify must be assigned biological functions. Expression profiling using microarrays provides a powerful tool for correlating gene functions with DNA sequences. It should be noted that DNA microarrays are also useful for identifying differentially expressed genes in organisms for which the available DNA sequence data is limited. This advantage is not shared by other high-throughput methods of expression analysis such as SAGE (serial analysis of gene expression) (13).

The use of microarrays for expression profiling is based on two fundamental principles. The first is that, for many genes, a predominant factor underlying changes in expression is an alteration in the abundance of the cognate mRNA. It is clear that post-transcriptional factors also can affect gene expression, however these factors are not considered in this review, since they are not generally amenable to microarray analysis. The second principle is that only DNA strands possessing complementary sequences can hybridize to each other to form a stable, double-stranded molecule. Microarrays exploit this property through the immobilization of millions of single-strand copies of a gene as individual array elements on a solid support surface. The array surface is then incubated with a mixture of labeled DNA molecules, which contains a proportional representation of all of the genes that are being expressed in a given tissue sample. Out of this mixture, only the labeled molecules that represent the same gene as the immobilized DNA elements can form heteroduplexes. By measuring the amount of label that is bound to each array element at the end of the hybridization reaction, a researcher can determine the relative transcript abundance level of each gene. Because each microarray comprises many elements, RNA abundance levels for thousands of genes can be measured in a

single experiment. By comparing abundance levels from several experiments, the investigator can then correlate patterns of gene expression with particular tissues or experimental conditions.

Despite being a relatively new field of research, some aspects of microarray nomenclature have already become ambiguous. In this review, we follow the convention established by the editors of Nature Genetics, who argue that the immobilized DNA should be termed the “probe” and the labeled DNA the “target”, due to similarities between microarraying and the older technique of reverse dot-blotting (14).

Two types of microarray are commonly used for expression profiling. These arrays differ primarily in the length of the probes that make up the array elements. For DNA microarrays, which are printed on membranes or glass, the array elements are large fragments of genes (~400-2000bp), typically produced through PCR amplification. In contrast, oligonucleotide-based microarrays (otherwise termed DNA “chips” (15)), have array elements comprising short (~20bp), synthetic DNA molecules. The production of oligonucleotide microarrays is usually done commercially, because it requires specialized equipment and procedures that are not typically available to research laboratories. In this review, we will therefore provide only a brief description of oligonucleotide microarrays, which will be followed by a detailed discussion of DNA microarrays. The analysis and applications of both types of microarrays will then be discussed.

PRODUCTION OF OLIGONUCLEOTIDE ARRAYS

The majority of oligonucleotide arrays in current use are produced by Affymetrix Inc. (Santa Clara, CA; Fig. 1 A-E). As a first step in the production of Affymetrix

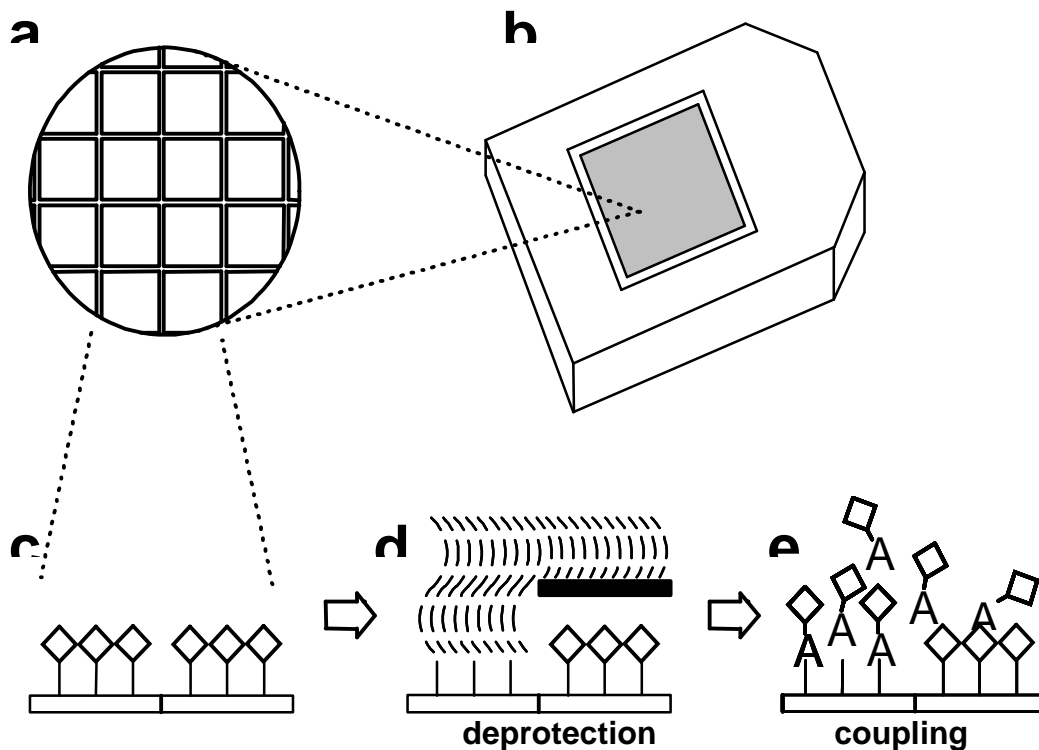


FIG. 1. Production of Affymetrix GeneChip® oligonucleotide microarrays. Each 1.28 x 1.28cm microarray consists of up to 260,000 square features (a). Each feature contains over 10^6 copies of a particular oligonucleotide probe. GeneChip arrays are packaged in cartridges (b) for handling during hybridization and analysis. Oligonucleotides are synthesized on linker molecules attached to the array substrate (c). The linker molecules are capped by a photolabile protecting group, represented here as a diamond. Protecting groups are removed in specific features by masking parts of the array and exposing it to light (d). Nucleotides, which also bear photolabile protecting groups (e), are flooded on to the array, where they bind to deprotected sites. The cycle of deprotection and coupling is repeated with different nucleotides and different masks to synthesize a different set of probes in each of the array's features.

GeneChips®, unigene EST databases are scanned for sequences representing genes of interest (16). Based on this sequence information, a set of 16 different complementary 25-mer oligonucleotides is designed for each of the selected genes. In addition, for each of these oligonucleotides a mismatch probe is also designed, which differs from the original sequence at only one base position, and therefore can be used during hybridization as a control for stringency.

All probes on Affymetrix chips are synthesized in parallel, on the array surface, using a photolithographic process (15, 17). This process begins with an array substrate that is coated with photolabile linker molecules (Fig. 1 C). Examples of photolabile protecting groups used in this context include DMBOC (5'-O-dimethoxybenzoincarbonate) and MeNPOC (((α -methyl-2-nitropiperonyl)-oxy)carbonyl)compounds (18, 19). Exposure to light removes the protecting group (Fig. 1 D) leaving a hydroxyl which is then available for coupling to reactive DNA monomers (i.e. MeNPOC or DMBOC-protected phosphoramidites of either adenosine, cytosine, guanosine, or thymidine) (Fig 1 E). Photolithographic masks are used to allow illumination only of those array locations at which a given monomer is to be attached. Unincorporated monomers are washed away, and deprotection is repeated using a different mask. This is followed by the coupling of the second of the four possible monomers at the newly deprotected sites. Subsequent cycles of washing, spatially-directed light exposure, and coupling are used to extend the oligonucleotides, thereby providing sequences that are predefined for each feature on the array. The result is the synthesis of approximately 10^6 copies of up to 260,000 different probes on a 1.28 x 1.28-cm array (16). Because each gene is represented by a total of 40 different probes

(including mismatch probes), the expression patterns of approximately 7,000 genes can be examined on the largest of the oligonucleotide arrays produced by Affymetrix.

Currently, Affymetrix arrays are available for the expression analysis of genes from six species: human, rat, mouse, yeast, *Arabidopsis* and *E. coli* (<http://www.affymetrix.com>). These arrays are expensive relative to the DNA microarrays produced by robotic spotting technology described below, must be purchased from Affymetrix, and require an expensive proprietary fluorescence reader for analysis. A further disadvantage of Affymetrix chips is the lack of flexibility inherent in the process of synthesis. Thus, should modifications to the chips be required, based on on-going discoveries in the laboratory, a considerable lead time is required for the expensive process of probe and lithographic mask design, prior to production of the modified chips.

An alternative means of synthesizing oligonucleotide arrays, using ink-jet deposition, has recently been developed, and promises to reduce costs and increase the flexibility of oligonucleotide array production (Agilent Technologies, Palo Alto, CA). With this technology, the time from completion of the design to fabrication of the array can take less than a day. Current state of the art permits the creation of arrays with more than 25,000 features (individual oligos up to 60mers in length) on a 1" by 3" microscope slide (M. N. Kronick, personal communication).

PRODUCTION OF DNA MICROARRAYS

The production of DNA microarrays can be divided into three stages: (i) the preparation of probe DNA, (ii) the preparation of a suitable array substrate, and (iii) the robotic deposition of array elements. Preparation of probe DNA is one of the most

resource-intensive aspects of microarray production. Although linear DNA of any source in principle can be immobilized on microarrays (Fig. 2 A), probes are generally produced by PCR amplification under high throughput conditions (in 96 or 384-well plates) (Fig. 2 B). Inserts are typically amplified from cDNA and EST libraries. Library clones may be chosen randomly for arraying, particularly when the microarray is to be used as a screening tool for gene discovery. Alternatively, bioinformatics techniques may be employed during clone selection (20), for example to construct arrays containing all members of a specific gene superfamily. Vector-specific primers are commonly employed for large-scale amplifications, since the same pair of primers can be used to amplify every insert within the library, and because they do not require sequence information of the cDNA inserts. However, because parts of the vector sequences and of the poly(A) tail can be included in each of the probes that are amplified in this manner, the specificity of hybridization on the microarray may be diminished. An alternative is the use of unique pairs of gene-specific primers for the amplification of each probe. These primers can be designed to exclude the poly(A) tail as well as regions of sequence similarity found between members of a gene family. Gene-specific primers can also be employed for amplification of probes from genomic DNA, in some cases obviating the need for EST/cDNA libraries. The use of gene-specific primer pairs is limited, to date, by expense, and by the degree to which genome sequencing has been completed.

Prior to microarray printing, the amplified probes are purified, and analyzed by gel electrophoresis. DNA amounts are then sometimes also quantified by fluorometry (21). Electrophoretic analysis is used to confirm the presence of a single amplicon, as well as to provide an estimate of probe concentration. To ensure that microarray

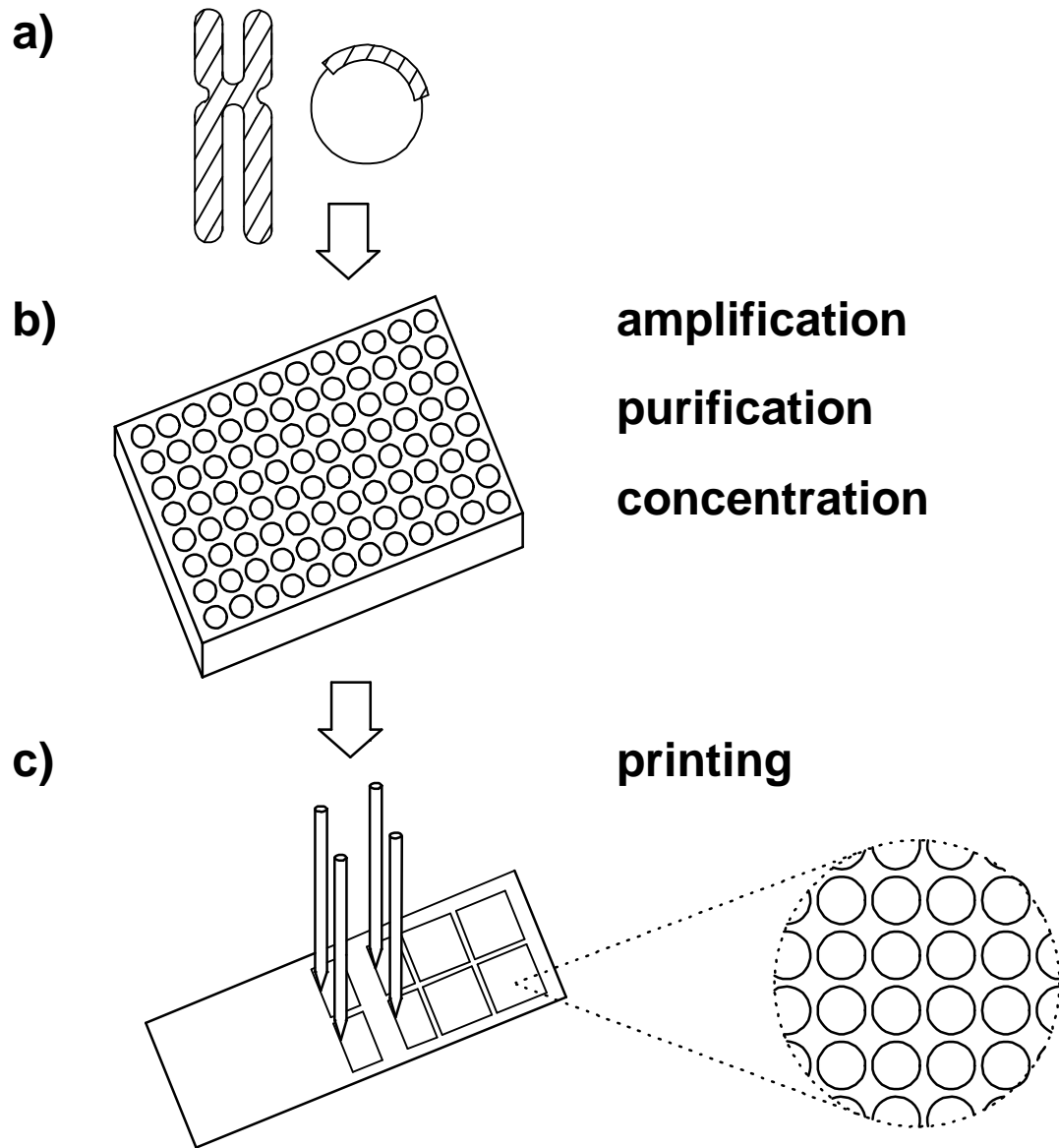


FIG 2. Production of DNA microarrays by microspotting. Starting with genomic DNA and/or cDNA and EST libraries (a), individual genes are amplified in multi-well plates (b). The PCR products are purified and concentrated, then robotically deposited on coated glass slides using devices such as slotted steel pins (c). Tens of thousands of spots can be deposited on each slide, and each spot contains roughly 10^8 copies of a specific gene fragment.

hybridization signal intensity is proportional to target abundance, it is essential that the final concentration of probe DNA be in excess relative to the expected concentrations of labeled target for each gene. In most published reports, probes for microarrays on glass are printed at concentrations between 100 and 500 ng/ μ L (22, 23). The minimum concentration of probe DNA required is likely to depend on factors including probe length, base composition, and the binding capacity of the array substrate. Heller *et al.* (24) examined the effects of printing various dilutions of different probes having lengths of 0.2 – 1.2 kb. Dilution of probe solutions up to eight-fold (i.e. a minimal DNA concentration of 125 ng/ μ L) affected hybridization signal intensities for only the shortest probes. Other reports indicate that short probes (<300bp) produce unreliable data in expression analyses (25).

Microarrays are formed by robotically depositing probes in a regular pattern upon a solid support (Fig. 2 C). Nylon filters can be used as substrates for arrays containing large numbers of elements (e.g., 36,000 elements; Incyte Microarray Systems, Fremont, CA) (26, 27). However, because of the deformability of these filters, array elements cannot be printed on nylon filters at high-density, meaning that a 36,000-element array necessarily occupies an area of at least 22 x 22 cm. High-density microarrays, which can be operationally defined as having element spacings (center-to-center) of 0.5 mm or less, are printed on solid substrates, typically glass microscope slides that have been coated to increase hydrophobicity and to enhance probe adhesion. The rigidity of glass, coupled to its desirable optical properties and the fact that it can be manufactured to exceptional surface flatness, means that arrays can be printed at element spacings as close as 60 μ m. In current practice, elements are generally arrayed at a minimal spacing of 150 μ m (28)

thus, more than 10,000 elements can be arrayed within the area of a standard (22 x 22 mm) coverslip. Minimizing the array area also minimizes the total volume required for the hybridization reaction. This is particularly important for experiments involving targets from low-abundance sources, for example tumor biopsies or plant meristems. Two other properties of glass that make it an excellent substrate for high-density microarrays are its low porosity, which helps minimize hybridization volumes, and its low inherent fluorescence, which minimizes background signals during analysis (29).

Glass slides are coated before printing, in order to increase the binding efficiency of probe DNA to the surface. A common coating material is poly-L-lysine, which binds to glass through non-covalent linkages involving the primary amines of the lysine residues. These amines also adsorb electrostatically to the phosphate groups of DNA to immobilize the probe on the slide (30). Other non-covalent interactions, including hydrophobic and hydrogen bonds also likely contribute to the coupling of DNA with poly-L-lysine. Amine-terminal silanes, which are a second class of coating, likewise provide positively charged amine-groups for binding probe DNA, but differ from poly-L-lysine in their covalent attachment to glass. A third type of surface coating, aldehyde-terminal silanes, bind covalently to both the glass surface and, via Schiff's base formation that is subsequently reduced, to the probe DNA (TeleChem Intl., Sunnyvale, CA). Schiff-base formation occurs between the aldehydes of the coating material and amine groups within the DNA molecules, either those occurring naturally in the bases of DNA, or those provided in the form of C6-amino modified PCR primers used during probe production (22). Although covalent attachment of DNA to the slide might be expected to diminish the loss of probes during denaturation and washing, the numerous published

reports that use slides coated with poly-L-lysine or amine-terminal silane indicate that non-covalent attachment of DNA to coated slides is sufficient for microarray production (23, 31).

Specialized robotic instruments are used to deposit probe DNA onto the coated glass slide. Some instruments use capillary tubes, piezoelectric ink-jets, solid pins, or pin-and-loop systems for probe deposition (32). However, the most common design uses slotted steel pins to transfer probes from their source plate to the slide. Much like a quill pen, these slotted pins collect the probe by capillary action, then deposit small aliquots of it each time they contact the slide surface. Each pin typically takes up 0.1 - 0.6 μ L of DNA solution from the source plate, and deposits approximately 0.3-1 nL as each array element (33, 34). This means that at least 400 spots can be printed from a single loading of the pin, although in practice this number may be limited by factors such as evaporation, or by the total numbers of slides that can be positioned on the printing platen. The printing times are typically 1-2 spots / s / pin, depending on the instrument and microarray parameters (26). This estimate does not include the time required for the rinsing, drying and pin sonication cycles which are usually programmed to precede each loading of the pin and change of source plates. Most robotic printers are capable of printing simultaneously with up to 32 pins, which greatly enhances the total printing rate. For example, GeneMachines (San Carlos, CA) reports that their OmniGrid™ equipped with 32 pins (Majer Precision Engineering, Tempe, AZ) can print a 34,000-element microarray on 100 slides in 17 hours.

TARGET PREPARATION AND HYBRIDIZATION

For the preparation of labeled target, RNA is extracted by standard methods from any tissue of interest. Either total cellular RNA (22, 35-37), or poly(A)+ mRNA (25, 31), or polysomal fractions of poly(A)+ mRNA (38) may be used. Hybridization to DNA microarrays typically requires 100 µg of total RNA, which is the approximate yield from 100mg of mammalian tissue, or from 300mg of Arabidopsis leaves. Poly(A)+ RNA normally comprises ~1-2% (w/w) of the total RNA (36). For 18mm x 18mm DNA microarrays printed on glass, hybridization can be successful with as little as 1 µg of poly(A)+ RNA (31, 39). RNA from microscopic amounts of tissue may also be successfully employed for microarray hybridization, if it is first amplified. Procedures for the linear amplification of RNA populations are based on *in vitro* transcription from an introduced T7 promoter, and can increase RNA abundance up to one million-fold (40, 41).

Prior to hybridization, target RNA is reverse transcribed into cDNA. For filter-based arrays, targets are generally labeled by incorporating radioactive nucleotides (27). In the case of oligonucleotide arrays and DNA microarrays printed on glass, targets are ultimately labeled with fluorescent dyes (42) (Fig. 3 A, B). Nucleotides that have been conjugated directly or indirectly to fluorochromes, such as fluorescein, lissamine, rhodamine, and phycoerythrin, have been used with both oligonucleotide and DNA arrays (42-44). The cyanine dyes Cy3 and Cy5 (Amersham Pharmacia Biotech, Piscataway, NJ) are currently the most popular fluorochromes for DNA microarrays because of their brightness, stability, and spectral separation. This spectral separation is important because it allows two different populations of labeled targets to be hybridized

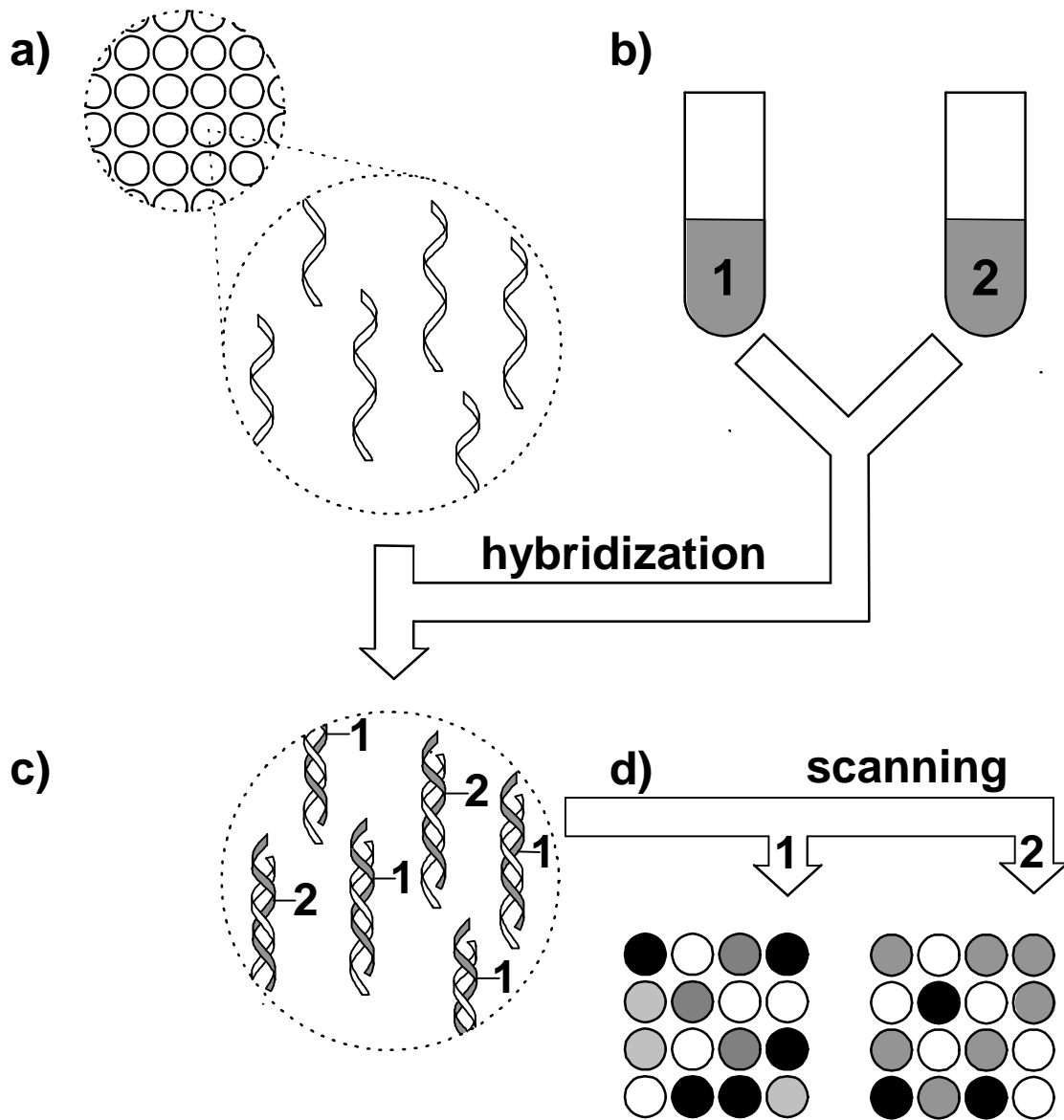


FIG 3. Hybridization and scanning of microarrays. Oligonucleotide or DNA microarrays consisting of single-stranded DNA attached to a solid substrate (a), are presented with the combined targets from experimental and reference RNA samples, which have been labeled with fluorochromes 1 and 2, respectively (b). Labeled target molecules hybridize to complementary probes, with a frequency proportional to their relative abundance in the original mixture of targets (c). Following hybridization and washing, the amount of the two labeled targets that has hybridized at each spot on the array is determined by laser scanning with CCD or confocal detection at wavelengths specific for the two fluorochromes.

simultaneously to a single DNA microarray. Affymetrix oligonucleotide arrays use analytical methods that do not permit dual-fluorochrome hybridizations.

When preparing targets for DNA microarrays printed on glass, fluorescent nucleotides can be incorporated directly into the first-strand cDNA product. However, some researchers prefer a two-step method in which labeled nucleotides are incorporated during a second-strand synthesis catalyzed by the Klenow fragment of DNA polymerase (<http://sgio2.biotech.psu.edu/protocols/klenow.txt>). Another two-step labeling method involves the incorporation of amino-allyl conjugated nucleotides during first-strand cDNA synthesis, followed by reaction of the first strand cDNA product with NHS-derivatized fluorochromes (<http://cmgm.stanford.edu/pbrown/>). Although direct incorporation is the fastest labeling method, it is thought that the two-step procedures eliminate a possible bias of some reverse transcriptases against the incorporation of Cy5-conjugated nucleotides, as compared to Cy3-conjugated nucleotides. It should be noted, however, that not all researchers consider this alleged bias to be significant.

Labeled targets for Affymetrix oligonucleotide arrays are prepared as cRNA, which is transcribed from approximately 1 μ g of first-strand cDNA product (45). Because many replicate cRNAs are transcribed from each cDNA template, this process amplifies the original mRNA population. Biotinylated nucleotides are incorporated during cRNA transcription, and these targets are visualized following hybridization by staining with streptavidin-fluorochrome conjugates (16, 45).

Labeled target molecules are hybridized to all types of microarrays under conditions similar to those conventionally used for northern blots (Fig. 3C). Buffers such as 2X SSC (300mM sodium chloride, 30mM sodium citrate) are employed to reduce

electrostatic repulsion between probe and target molecules. Competitive blocking reagents, such as yeast tRNA, CoT1 DNA, herring sperm DNA, or poly (dA) DNA are frequently also added to reduce non-specific hybridization. Arrays consisting of double-stranded probes are placed briefly in boiling water, to denature the probes. Filter-based arrays are hybridized in bottles or sealed bags. Hybridization on glass slides is performed under a coverslip, in a heated, humid chamber. Affymetrix oligonucleotide arrays are mounted in plastic cartridges, which also serve as hybridization chambers. On all types of arrays, as for standard northern and Southern blots, stringency can be increased by a decrease in the salt concentration, or by an increase in temperature during hybridization and washing. In certain cases, formamide is included, with a corresponding reduction in hybridization and post-hybridization temperatures.

Hybridizations in which stringency has not been optimized can lead to errors, due to cross-hybridization of genes that share small regions of sequence homology. One experiment with DNA microarrays indicated that short regions of similarity, even with up to 90% sequence identity, produced only low levels of cross-hybridization (24). However, when several of these short regions of homology were distributed throughout a longer DNA fragment, the incidence of cross-hybridization increased significantly. Other experiments, which used spotted 50-mer oligonucleotides as probes, demonstrated that sequences with <75% complementary sequence identity between probe and target did not contribute significantly to the hybridization signal, unless the entire region of identity was localized to a series of 15 or more contiguous bases (46). We found similar levels of specificity, when using PCR products as probes, during a systematic analysis of the

cytochrome P450 superfamily of *Arabidopsis thaliana* (W. Xu and D. W. Galbraith, unpublished observations).

Several types of negative controls for non-specific hybridization may be included as probes on DNA microarrays, including genes from distantly-related species, non-coding genomic sequences, or probes containing only poly(dA) or poly(dT). Non-specific hybridization on Affymetrix oligonucleotide arrays is assessed on a gene by gene basis, by comparing signals between probes and their corresponding “mismatch” probes.

SCANNING

Following hybridization and washing, the amount of target bound at each spot is quantified (Fig. 3 D). Radioactively labeled targets are detected using a phosphorimager. Fluorescently labeled targets are excited by laser or filtered xenon-bulb light sources, and are visualized using specialized instruments with charge-coupled device (CCD) or confocal detectors. Commercial instruments for reading arrays on glass slides currently scan at resolution as high as $5\mu\text{m}$ / pixel, with a sensitivity of <0.1 molecule fluor/ μm^2 , and a dynamic range of four orders of magnitude (47) (see for example, instruments from GSI Lumonics, Watertown, MA; Axon Instruments, Foster City, CA). Although some scanners are capable of excitation and detection in up to four different wavelengths, two-channel detection is currently most common. Images are obtained optically, then reported as a grey-scale bitmap, typically with a 16-bit range. Subsequently, the boundaries of each spot in the image must be identified, usually with the aid of image-processing software (48). This software quantifies the signal intensity at each spot, most often as either the median or mean of all pixel intensities within the spot boundary.

Many software programs for the analysis of DNA microarrays also measure the local background intensity surrounding the spot boundary. Mathematical filters may be applied during quantification, for example to exclude pixels with either very low or very high intensities, as these are often due to background fluorescence or dust. A detailed discussion of statistical methods for array quantification has been presented elsewhere (49).

NORMALIZATION

The goal of most microarray experiments is to correlate changes in gene-specific RNA abundance with particular treatments or conditions. As such, RNA samples are usually analyzed in pairs, with one sample being derived from a control state, and the other sample from the experimental state. The overall signal intensities from each sample must then be normalized, scaled, or standardized relative to each other, to eliminate systematic variations in intensity, which are not related to changes in relative RNA abundance for individual genes. Systematic sources of error between microarray experiments include differences in the amount of nucleic acid present in each target preparation, and variations in the efficiency of labeling, hybridization, and washing. The impact of some (but not all) of these errors is minimized by simultaneously hybridizing targets from two different sources to a single array, as is common with most DNA microarray experiments, but not with oligonucleotide array experiments (5) (Fig. 4).

Normalization may be achieved by adjusting detection sensitivity during scanning, or by mathematical adjustment of raw data following scanning. Various strategies can be used determine the amount of adjustment required (48, 50, 51). Each of

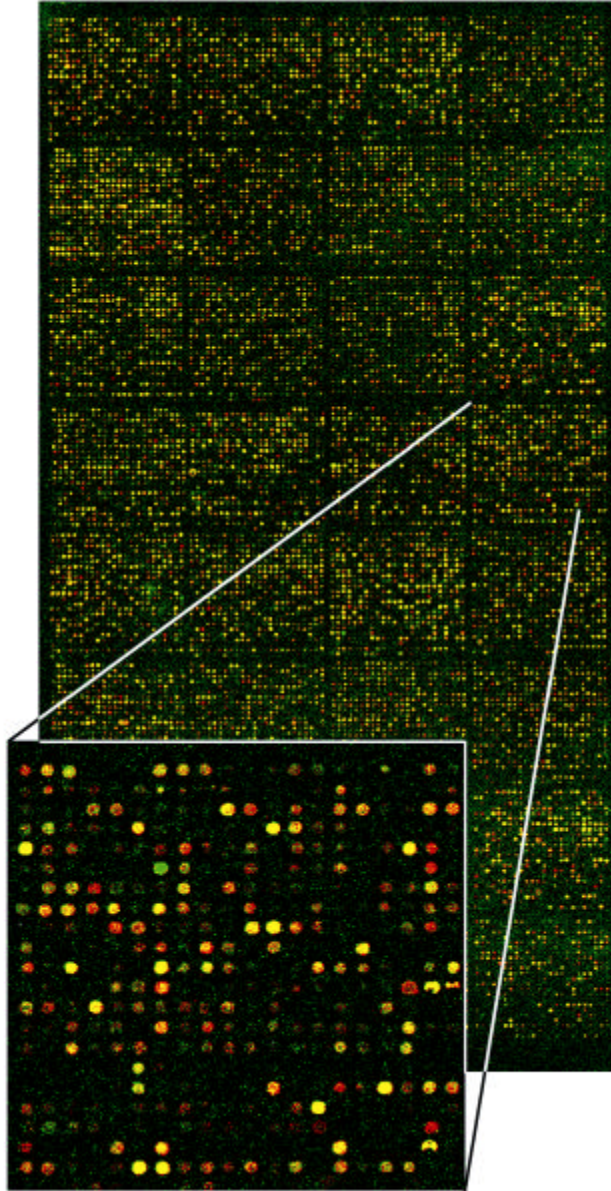


FIG. 4. Two-color hybridization to an 18,000 element DNA microarray. ESTs representing over 9,000 Arabidopsis genes were spotted in duplicate arrays at 180 μ m center-to-center spacing on an aminoalkylsilane-coated microscope slide. This slide was hybridized with targets derived from NaCl-treated (Cy3-labeled) and untreated control (Cy5-labeled) Arabidopsis plants. Spots that hybridized most strongly with the Cy-3 or Cy-5 labeled targets appear red or green, respectively, and may represent genes that are differentially expressed in response to NaCl-treatment. Where an EST hybridized equally to both two targets, the spots appear yellow or brown. The inset in this image shows a close-up of part of the microarray.

these strategies depends of the selection of a set of reference signals, whose levels are not expected to change as a result of the experimental treatment. For example, some researchers pre-select a suite of housekeeping genes, which are not expected to be differentially expressed under most conditions (44). One problem with this method is its dependence on the accuracy of assumptions made about the expression patterns of the housekeeping genes. As an alternative, a set of probes from heterologous genes, having no sequence similarities to other array elements, can be included as part of the microarray. Equal quantities of *in vitro* RNA transcripts corresponding to these cDNAs are then added to the control and experimental RNA samples prior to labeling (36). The normalization factor is then determined by comparing the mean signal intensities of these heterologous controls in the two different samples. This method is limited by the accuracy with which proportional quantities of the heterologous genes can be transferred into each of the RNA samples, and is less popular than other methods, perhaps because of the additional resources and experimental manipulations it requires. A third, and more common, method of normalization uses the entire set of signals from each channel or experiment as the reference set. This global method relies on the assumption that only a small proportion of genes represented on a given array will be differentially expressed. Accordingly, when the two complete data sets are compared, the systematic variation is assumed to equal the overall variation measured between the two sets of data (52). The data sets can therefore be normalized by multiplying each signal in the appropriate channel by a factor that is inversely proportional to the overall variation. This factor is normally determined by linear regression of the two populations, or simply by taking the quotient of the medians of all signals from the two data sets. We recommend the global

method of normalization as a general approach for all DNA microarray experiments, except for cases in which the selection of probes is small or else strongly biased towards one set of experimental conditions, as may occur when ESTs are chosen from subtracted libraries.

ACCURACY AND PRECISION

Validation of microarray accuracy is generally done by comparing results from a subset of microarray data, with results produced by other methods of RNA analysis. In most cases, these comparisons have been qualitative, demonstrating, for example, that genes that were up- or down- regulated on microarrays were regulated in a similar manner when detected by RNA blotting (37, 53), RT-PCR (28), or SAGE (23, 54). In a quantitative comparison of DNA microarray and RNA blot experiments, Schena *et al.* (22) reported that the expression ratios measured by the two procedures varied by less than two-fold, although absolute expression levels varied up to five-fold between the techniques. In terms of precision, our research has shown a small but significant variation in signal intensities, even for adjacent spots of identical probes on a single slide (Deyholos and Galbraith, unpublished data). Following hybridization, we found an average 10% coefficient of variation (CV%) of median signal intensities for 900 genes printed in quadruplicate on a single DNA microarray. Less than one-tenth of these sets of quadruplicate spots had a CV% of 20% or more. Others have reported that the expression ratios derived from these signal intensities are also variable; in duplicate independent experiments with 6,000 yeast genes on DNA microarrays, 95% of the expression ratios varied by two-fold or less, with a correlation coefficient of 0.87 (44).

Furthermore, Lee and colleagues (55) reported that in control experiments with 288 genes on DNA microarrays, when data from only one hybridization was considered, up to 9.0% of the genes were misclassified as being up- or down- regulated. However, when data from three replicate hybridizations were considered together, the misclassification rate dropped to 0.7%. Therefore, although we expect that further refinements in printing and hybridization will increase precision in the future, it is clear that several replicate microarray experiments should be pooled in order to provide the most accurate estimates of expression data.

IDENTIFYING DIFFERENTIAL EXPRESSION

The most basic approach to the analysis of microarray expression data is the identification of genes that are up- or down-regulated in response to experimental conditions. For example, in a 1,443-member sample of an *Arabidopsis* EST collection, 490 genes showed differential regulation (a 2- to 15-fold change) in root compared to leaf (36). Because genes that are highly regulated in response to a particular stimulus are likely to have important roles in the corresponding cellular response, this approach can be useful in correlating functions with genes. However, one limitation of this method is that it requires an *a priori* determination of a threshold level of significance for expression ratios. If this threshold is set too high, some relevant genes will be ignored. Conversely, a threshold that is set too low will lead to false correlations and the misclassification of genes. In practice, changes in expression of at least 2.0 fold have been considered significant, although only rarely has any statistical or biological justification been offered for the selection of this threshold (44). However, general methods for

determining statistically significant threshold levels of expression, based on analysis of variance models, have also recently been described (56, 57).

IDENTIFYING PATTERNS OF CO-EXPRESSION

A potentially powerful means of analyzing microarray expression data is the identification of groups of genes that are expressed in similar patterns. This technique is valuable because co-expressed genes are likely to participate in the same biological function. This assumption facilitates the characterization of novel genes. Furthermore, the genomic regions flanking co-expressed genes can be compared for similarities, to identify the specific *cis*-acting elements that regulate each pattern of expression (58). Finally, the comparison of expression patterns for various groups of co-expressed genes can reveal interactions between the groups, leading to global descriptions of genetic regulatory programs.

Several computational methods have been developed for identifying patterns of co-expression. These include hierarchical clustering (59), Bayesian clustering, *k*-mean clustering, two-way clustering (60), Pearson rank clustering (52), self-organizing maps (SOMs) (61), support vector machines (62). Here we will discuss only two of the most common methods, namely hierarchical clustering and SOMs, which have also been compared elsewhere (63-65). In hierarchical clustering, sets of expression data derived from a variety of time points or experimental conditions are grouped together according to similarities in expression patterns. Eisen and colleagues applied a form of hierarchical clustering (pairwise average linkage clustering), to expression data from yeast and human fibroblasts (59). They reported that redundant representations of the same gene on an

array repeatedly clustered very close together, and when larger clusters of genes were examined, the genes within each cluster tended to have common roles in cellular processes. To demonstrate that the patterns that were generated had biological significance, and were not merely artifacts of the clustering process, these researchers also clustered randomized permutations of their initial data sets. The failure of the randomized data sets to be organized into meaningful clusters implies that pair-wise average linkage clustering may be a useful tool for analyzing large sets of expression data.

The use of hierarchical clustering for gene expression data has been criticized for at least three reasons: (i) because of its deterministic nature, (ii) because gene expression data are not inherently hierarchical, and (iii) because adjacent branches are not always directly related (61, 66). SOMs have been developed as an alternative means of analyzing large sets of expression data. In constructing an SOM, the experimenter first estimates the number of expected nodes, then through a process of iterative geometric calculations, genes with similar expression patterns aggregate at each node (61). The ability of SOMs to discover biologically relevant patterns of gene expression was demonstrated by the analysis of a 160 minute time-course data set from yeast. The self-organizing map automatically extracted cell-cycle periodicity as one of the most prominent patterns in the data. Software for both self-organizing maps and hierarchical clustering is freely available on the internet (<http://www-genome.wi.mit.edu/MPR/software.html> ; <http://rana.stanford.edu/software/>).

APPLICATIONS OF MICROARRAYS

Expression profiling in model organisms. A variety of model organisms, including *D. melanogaster*, *C. elegans*, yeast, mouse, and *A. thaliana*, have been subjected to expression profiling with microarrays. The earliest experiments with each of these organisms focussed on screening EST populations for differentially-regulated genes (22, 31, 36). More recently, the analysis has expanded to include the identification of genome-level patterns of expression related to complex processes. One such process is metamorphosis in *D. melanogaster* (63), for which DNA microarray analysis identified many genes of previously unknown function, assigned novel functions to previously-characterized genes, and identified potential points of co-ordination between major developmental and metabolic pathways. Microarrays have also been used to study development in *C. elegans* (67), including the identification of genetic cascades activated by transforming growth-factor β (68). In mice, researchers were able to conclude that aging is not due to widespread changes in gene expression, by using oligonucleotide arrays to compare expression profiles of muscles from young and old animals (69). Many complex processes in *S. cerevisiae* have been successfully analyzed at the whole-genome expression level, using microarrays. These processes include the diauxic shift from respiration to fermentation (70), the mitogen associated protein-kinase signaling pathway (71, 72), the cellular response to alkylating agents (73), and ploidy-dependent regulation of gene expression (74). Currently, Arabidopsis microarrays are being used to study a variety of developmental and metabolic processes including responses to biotic and abiotic stress (75, 76) ; Deyholos and Galbraith, unpublished results).

Expression profiling in medical sciences. Expression profiling with microarrays has found wide application in the analysis of human disease. Cancer biologists have used

DNA microarrays to compare gene expression in neoplastic and normal ovarian tissues (54), tumorigenic and non-tumorigenic melanoma cell lines (44), and invasive and non-invasive melanoma biopsies from a single patient (77). Similarly, oligonucleotide microarrays have been used in the expression analysis of genes related to colon cancer (60), breast cancer (78), and Wilms tumor (69). Together, these reports have correlated aspects of cancer pathology with changes in expression of hundreds of genes, the majority of which had not been previously associated with this disease. Expression profiles from both DNA and oligonucleotide arrays are also being developed as diagnostic tools for cancer classification (79, 80). Other medical applications of microarrays include the identification of human genes related to inflammatory disease (24), pulmonary fibrosis (81) and responses to ionizing radiation (37), heat shock (22), growth factors (45), and interferons (82). Finally, DNA microarray expression profiling has been used to analyze virulence, drug resistance, and host responses in pathogens such as *Mycobacterium tuberculosis* (83), and human cytomegalovirus (84).

Although this review has focused on expression profiling, we note that microarrays have also been adapted to several other high-throughput methods of nucleic acid analysis. For example, DNA microarrays have been used for linkage-disequilibrium mapping (85), and the detection of gene amplification (86). Oligonucleotide arrays are particularly well-suited to the detection of single-nucleotide polymorphisms, which are especially relevant to the treatment of human disease (87, 88).

FUTURE DIRECTIONS

The next major advances in microarray technology are likely to involve further automation, especially of those processes that are upstream of array printing, or which would improve sample tracking from start to finish. Pin and printer designs are also constantly evolving to improve spot uniformity, spot density, yield of arrays per print, and printing speed. The process of hybridization is another likely target for technical improvement, both through automation, and through the development of devices to improve uniformity of hybridization across the array surface. Finally, the widespread use of scanners with the ability to read four or more channels will further expand the information obtained from each microarray experiment.

We expect that DNA microarrays will continue to be applied to expression analysis of an increasing number of economically important species. For example, the National Science Foundation, which is one of the major funding agencies in the United States, is currently funding DNA microarray research on species as diverse as maize, rice, soybean, ice plant, potato, loblolly pine, and Arabidopsis (72, 89).

Although DNA microarrays can be used to answer almost any gene-dependent biological question in any organism, the relative cost of equipment and materials has so far limited their application, with the most expensive consumables being fluorescent dyes, polymerases, and coated slides. Currently, the cost of a single microarray experiment is described in terms of hundreds of dollars. As with other developing technologies, we expect that technical advances and economies of scale will reduce the cost of producing and using DNA microarrays. Within a few years, if the cost of each hybridization experiment is reduced to just a few dollars, microarrays could be used to address population-scale questions in evolutionary biology and ecology. Inexpensive

microarrays could also be widely applied in toxicology and the environmental sciences for hazard assessment, and in pharmacology for large-scale screening of chemical libraries for novel and beneficial effects. Finally, as cost decreases, and accuracy and precision become standardized, microarrays will be routinely used in medical and agricultural diagnostics, and likely even the forensic sciences.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant # 9813360). We thank Georgina Lambert and Harish Rekapally for proofreading this manuscript.

LITERATURE CITED

1. Alwine J, Kemp D, Stark G. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci USA* 1977; 74: 5350-5354.
2. Liang P, Pardee A. Differential display of eukaryotic messenger RNA by means of polymerase chain reaction. *Science* 1992; 257: 967-971.
3. Zinn K, DiMaio D, Maniatis T. Identification of two distinct regulatory regions adjacent to the human beta-interferon gene. *Cell* 1983; 12: 721-732.
4. McCulloch R, Choong C, Hurley D. An evaluation of competitor type and size for use in the determination of mRNA by competitive PCR. *PCR Methods Appl* 1995; 4: 219-226.
5. Schena M, Shalon D, Davis R, Brown, P. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270: 467-470.
6. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* 1999; 21: 33-37.
7. Baxevanis A. The molecular biology database collection: an online compilation of relevant database resources. *Nuc Acids Res* 2000; 28: 1-7.
8. Goffeau A, Barrell B, Bussey H, Davis R, Dujon B, Feldmann H, Galibert F, Hoheisel J, Jacq C, Johnston M, Louis E, Mewes H, Murakami Y, Philippsen P, Tettelin H, Oliver S. Life with 6000 genes. *Science* 1996; 274: 546-567.
9. *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998; 282: 2012-2018.
10. Adams, M. *et al.*, The genome sequence of *Drosophila melanogaster*. *Science* 2000; 287: 2185-2195.
11. Science Scope. Genomaniacs. *Science* 2000; 287: 27.

12. Marra M, Hillier L, Waterston R, Expressed sequence tags- ESTablishing bridges between genomes. *Trends Genet* 1998; 14: 4-6.
13. Velculescu V, Zhang L, Vogelstein B, Kinzler K. Serial analysis of gene expression. *Science* 1995; 270: 484-487.
14. Phimister B, A note on nomenclature. *Nat Genet* 1999; 21: 1.
15. Lipshutz R, Fodor S, Gingeras T, Lockhart D, High density synthetic oligonucleotide arrays. *Nat Genet* 1999; 21 Suppl.: 20-24.
16. Warrington J, Dee S, Tulson M. Large-scale genomic analysis of gene expression microarrays. in *Microarray Biochip Technology*, Schena, M., ed., Eaton Publishing, 2000. Sunnyvale, CA, pp. 119-148.
17. McGall G, Labadie J, Brock P, Wallraff G, Nguyen, T. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci USA* 1996; 93: 13555-13560.
18. McGall G, Barone A, Diggelmann M, Fodor S, Gentalen E, Ngo N. The efficiency of light-directed synthesis of DNA arrays on glass substrates. *J Am Chem Soc* 1997; 119: 5081-5090.
19. Pirrung M, Fallon L, McGall G. Proofing of photolithographic DNA synthesis with 3',5'-dimethoxybenzoinyloxycarbonyl-protected deoxynucleoside phosphoramidites. *J Org Chem* 1998; 63: 241-246.
20. Loftus SK, Chen Y, Gooden G, Ryan JF, Birznieks G, Hilliard M, Baxenvanis A D, Bittner M, Meltzer P, Trent J, Pavan W. Informatic selection of a neural crest-melanocyte cDNA set for microarray analysis. *Proc. Natl. Acad. Sci. USA* 1999; 96: 9277-9280.
21. Singer V. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal Biochem* 1997; 247.
22. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996; 93: 10614-10619.
23. Welford SM, Gregg J, Chen E, Garrison D, Sorenson PH, Denny CT, Nelson SF. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization *Nucleic Acids Research* 1998; 26: 3059-3065.
24. Heller RA, Schena M, Chai A, Shalon D, Bedilion T, Gilmore J, Woolley DE, Davis RW. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays *Proc. Natl. Acad. Sci. USA* 1997; 94: 2150-2155.
25. Yang GP, Ross DT, Kuang WW, Brown PO, Weigel RJ. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nuc Acids Research* 1999; 27: 1517-1523.
26. Bowtell, D. Options available - from start to finish - for obtaining expression data by microarray. *Nature Genetics* 1999; 21: 25-32.
27. Giege P, Konthur Z, Walter G, Brennicke A. An ordered *Arabidopsis thaliana* mitochondrial cDNA library on high-density filters allows rapid systematic analysis of plant gene expression: a pilot study. *Plant J* 1998; 15: 721-726.
28. Khan J, Saal LH, Bittner ML, Chen Y, Trent JM, Meltzer PS. Expression profiling in cancer using cDNA microarrays. *Electrophoresis* 1999; 20: 223-229.

29. Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. *Nat Genet* 1999; 21: 5-14.
30. Nicola NA, Kristjansson KR, Fasman GD, Interaction of poly(L-Lysine) and copolymers of lysine with immobilized DNA. *Archives of Biochemistry and Biophysics* 1979; 193: 204-212.
31. Lasharki 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.
32. Mace M, Montagu J, Rose S, McGuinness G. Novel microarray printing and detection technologies. in: *Microarray Biochip Technology*, Schena, M., ed., Eaton Publishing, 2000. Sunnyvale, CA, pp. 39-6.
33. Rose D. Microfluidic technologies and instrumentation for printing DNA microarrays. in: *Microarray Biochip Technology.*, Schena, M., ed., Eaton Publishing, 2000. Sunnyvale, CA, pp. 19-38.
34. Cheung V, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. *Nat Genet* 1999; 21: 15-19.
35. de Saizieu, A., Certa, U., Warrington, J., Gray, C., Keck, W., and Mous, J., Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat Biotech* 1998; 16: 45-48.
36. Ruan Y, Gilmore J, Conner T. Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J* 1998; 15: 821-833.
37. Amundson SA, Bittner M, Chen Y, Trent J, Meltzer P, Fornace AJ. Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene* 1999; 18: 3666-3672.
38. Diehn M, Eisen M, Botstein D, Brown P. Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nat Genet* 2000; 25: 58-62.
39. Iyer V, Eisen M, Ross D, Schuler G, Moore T, Lee J, Trent J, Staudt L, Hudson J, Boguski M, Lashkari D, Shalon D, Botstein D, Brown P. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; 83-87.
40. Luo L, Salunga R, Guo H, Bittner A, Joy K, Galindo J, Xiao H, Rogers K, Wan J, Jackson M, Erlander M. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 1999; 5: 117-121.
41. Van Gelder R, von Zastrow M, Yool A, Dement W, Barchas J, Eberwine J. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990; 87:1663-1667.
42. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Methods* 1996; 6: 301-306.
43. Chee M, Yang R, Hubbell E, Berno A, Huang X, Stern D, Winkler J, Lockhart D, Morris M, Fodor S. Accessing genetic information with high density DNA arrays. *Science* 1996; 274: 610-614.
44. DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su Y A, Trent, JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 1996; 14: 457-460.

45. Fambrough D, McClure K, Kazlauskas A, Lander, E. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* 1999; 97: 727-741.
46. Kane MD, Jatkoa TA, Stumpf CR, Lu J, Thomas JD, Madore SJ. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nuc Acid Res* 2000; 28: 4552-4557.
47. Basarsky T, Verdnik D, Ye Zhai J, Wellis D. Overview of a microarray scanner: design essentials for an integrated acquisition and analysis platform, in *Microarray Biochip Technology*, Schena, M., ed., Eaton Publishing, 2000. Sunnyvale, CA, pp. 265-284.
48. Zhou Y, Kalocsai P, Chen J, Shams S. Information processing issues and solutions associated with microarray technology., in *Microarray Biochip Technology*, M, S., Ed., Eaton Publising, 2000. Sunnyvale, Ca, pp. 167-200.
49. Chen Y, Dougherty ER, and Bittner ML. Ratio-Based Decisions and the Quantitative Analysis of cDNA Microarray Images *Journal of Biomedical Optics* 1997; 2: 364-374.
50. Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrac H, Herzel H. Normalization strategies for cDNA microarrays *Nucleic Acids Res* 2000; 28: i-v.
51. Richmond T, and Sommerville S. Chasing the dream: plant EST microarrays. *Curr Opin Plant Biol* 2000; 3: 108-116.
52. Voehringer D, Hirschberg D, Xiao J, Lu Q, Roederer M, Lock, C, Steinman L, Herzenberg L. Gene microarray identification of redox and mitochondrial elements that control resistance or sensitivity to apoptosis. *Proc Natl Acad Sci USA* 2000; 97: 2680-2685.
53. Kapp U, Yeh WC, Patterson B, Elia AJ, Kagi D, Ho A, Hessel A, Tipsword M, Williams A, Mirtsos C, Itie A, Moyle M, Mak TW. Interleukin 13 Is Secreted by and Stimulates the Growth of Hodgkin and Reed-Sternberg Cells *Journal of Experimental Medicine* 1999; 189: 1939-1945.
54. Wang K, Gan L, Jeffery E, Gayle M, Gown AM, Skelly M, Nelson PS, Ng WV, Schummer M, Hood L, Mulligan L. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray *Gene* 1999; 229: 101-108.
55. Lee MLT, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proc Nat Acad Sci USA* 2000; 97: 9834-9839.
56. Aharoni A, Keizer LCP, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen A, De Vos R, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, and O'Connell AP. Identification of the *SAAT* gene involved in strawberry flavor biogenesis by use of DNA microarrays *Plant Cell* 2000; 12: 647-662.
57. Greller L, and Tobin F. Detecting selective expression of genes and proteins. *Genome Res* 1999; 9: 282-296.
58. Wolfsberg T, Gabrielian A, Campbell M, Cho R, Spouge J, D, L. Candidate regulatory sequence elements for cell cycle-dependent transcription in *Saccharomyces cerevisiae*. *Genome Res* 1999; 9: 775-792.

59. Eisen MB, Spellman PT, Brown PO, and Botstein D. Cluster analysis and display of genome-wide expression patterns Proc. Natl. Acad. Sci. USA 1998; 95: 14863-14868.
60. Alon U, Barkai N, Notterman D, Gish K, Ybarra S, Mack D, Levine A. 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.
61. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, and 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.
62. Brown M, Grundy W, Lin D, Cristianini N, Sugnet C, Furey T, Manuel Ares J, and Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 2000; 97: 262-267.
63. White K, Rifkin S, Hurban P, and Hogness D. Microarray analysis of Drosophila development during metamorphosis. Science 1999; 286: 2179-2184.
64. Zhang M. Large-scale gene expression data analysis: a new challenge to computational biologists. Genome Res 1999; 9: 681-688.
65. Heyer L, Kruglyak S, and Yooseph S. Exploring expression data: identification and analysis of coexpressed genes. Genome Res 2000; 9: 1106-1115.
66. Toronen P, Kolehmainen M, Wong G, and Castren E. Analysis of gene expression data using self-organizing maps. FEBS Lett 1999; 451: 142-146.
67. Hill AA, Hunter CP, Tsung BT, Tucker-Kellogg G, and Brown EL. Genomic analysis of gene expression in *C. elegans* Science 2000; 290: 809-812.
68. Mochii M, Yoshida S, Morita K, Kohara Y. Identification of transforming growth factor-beta-regulated genes in *Caenorhabditis elegans* by differential hybridization of arrayed cDNAs Proc Natl Acad Sci USA 1999; 96: 15020-15025.
69. Lee S, Huang K, Palmer R, Truong V, Herzlinger D, Kolquist K, Wong J, Paulding C, Yoon S, Gerlad W, Oliner J, Haber D. The Wilms tumor suppressor *WT1* encodes a transcriptional activator of *amphiregulin* Cell 1999; 98: 663-673.
70. DeRisi J, Iyer VR, and Brown P. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997; 278: 680-686.
71. Madhani H, Galitski T, Lander E, Fink G. Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatures of signaling mutants. Proc Natl Acad Sci USA 1999; 96: 12530-12535.
72. Roberts C, Nelson B, Marton M, Stoughton R, Meyer M, Bennett H, He Y, Dai H, Walker W, Hughes T, Tyers M, Boone C, Friend S. Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science 2000; 287: 873-880.
73. Jelinsky S and Samson L. Global response of *Saccharomyces cerevisiae* to an alkylating agent. Proc Natl Acad Sci USA 1999; 96: 1486-1491.
74. Galitski T, Saldanha A, Styles C, Lander E, Fink G. Ploidy regulation of gene expression. Science 1999; 285: 251-254.

75. Reymond P, Weber H, Damond M, Farmer E. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *Plant Cell* 2000; 12: 707-720.
76. Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Sommerville SC, Maners JM. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis *Proc Nat Acad Sci USA* 2000; 97: 11655-11660.
77. Maniotis AJ, Folberg R, Hess A, Seftor EA, Gardner LMG, Peer J, Trent JM, Meltzer PS, Hendrix MJC. Vascular Channel Formation by Human Melanoma Cells *in Vivo* and *in Vitro*: Vasculogenic Mimicry. *American Journal of Pathology* 1999; 155: 739-752.
78. Harkin P, Bean J, Miklos D, Song Y, Truong V, Englert C, Christians F, Ellisen L, Maheswaran S, Oliner J, Haber D. Induction of *GADD45* and JNK/SAPK-dependent apoptosis following inducible expression of *BRCA1* *Cell* 1999; 97: 575-586.
79. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesitov JP, Coller H, Loh ML, Downing JR, Caliguri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Nat Genet* 1999; 28: 531-537.
80. Alizadeh A, Eisen M, Davis R, Ma C, Lossos I, Rosenwald A, Boldrick J, Sabet H, Tran T, Yu X, Powell J, Yang L, Marti G, Moore T, Hudson J, Lu L, Lewis D, Tibshirani R, Sherlock G, Chan W, Greiner T, Weisenburger D, Armitage J, Warnke R, Levy R, Wilson W, Grever M, Byrd J, Botstein D, Brown PO. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403: 503-511.
81. Kaminski N, Allard J, Pittet J, Zuo F, Griffiths M, Morris D, Huang X, Sheppard D, Heller R. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis *Proc Natl Acad Sci USA* 2000; 97: 1778-1783.
82. Der S, Zhou A, Williams B, and Silverman R. Identification of genes differentially regulated by interferon α , β , or γ , using oligonucleotide arrays. *Proc Natl Acad Sci USA* 1998; 95: 15623-15628.
83. Wilson M, DeRisi J, Kristensen H, Imboden P, Brown P, Schoolnik G. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc Natl Acad Sci USA* 1999; 96: 12833-12838.
84. Zhu H, Cong J, Mamtora G, Gingeras T, and Shenk T. Cellular gene expression altered by human cytomegalovirus. *Proc Natl Acad Sci USA* 1998; 95: 14470-14475.
85. Cheung VG, Gregg JP, Goglin-Evans KJ, Bandong J, Stanley CA, Baker L, Higgins MJ, Nowak NJ, Shows TB, Ewens WJ, Nelson SF, Spielman RS. Linkage-disequilibrium mapping without genotyping *Nature Genetics* 1998; 18: 225-230.
86. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO. Genome-wide analysis of DNA copy-number changes using cDNA microarrays *Nature Genetics* 1999; 23: 41-46.

87. Halushka M, Fan J, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, AC. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 1999; 22: 239-247.
88. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane C, Lim E, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley G, Lander E. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; 22: 231-238.
89. Walbot V. What can plant biologists expect from the 1998 National Science Foundation plant genome research program. *Plant Physiol* 1999; 119: 1151-1156.