

Fundamentals of DNA Hybridization Arrays for Gene Expression Analysis

BioTechniques 29:1042-1055 (November 2000)

Willard M. Freeman, Daniel J. Robertson and Kent E. Vrana
Wake Forest University School of Medicine, Winston-Salem, NC, USA

ABSTRACT

DNA hybridization arrays [also known as macroarrays, microarrays and/or high-density oligonucleotide arrays (Gene Chips™)] bring gene expression analysis to a genomic scale by permitting investigators to simultaneously examine changes in the expression of literally thousands of genes. For hybridization arrays, the general approach is to immobilize gene-specific sequences (probes) on a solid state matrix (nylon membranes, glass microscope slides, silicon/ceramic chips). These sequences are then queried with labeled copies of nucleic acids from biological samples (targets). The underlying theory is that the greater the expression of a gene, the greater the amount of labeled target, and hence, the greater output signal. In spite of the simplicity of the experimental design, there are at least four different platforms and several different approaches to processing and labeling the biological samples. Moreover, investigators must also determine whether they will utilize commercially available arrays or generate their own. This review will cover the status of the hybridization array field with an eye toward underlying principles and available technologies. Future developments and technological trends will also be evaluated.

INTRODUCTION

As investigators work on more complete and annotated copies of the human genome project, their attention turns to methods for using this wealth of information. Two questions generated by sequencing the entire complement of human genetic material are (i) how differential expression of that information is associated with health and disease and (ii) how mutations or small natural variations in that sequence produce genetic disorders and/or increased risk for disease. The former of these, gene expression, is the cornerstone of functional genomics. In the present context, functional genomics is defined as the study of all the genes expressed by a specific cell or group of cells and the changes in their expression pattern during development, disease or environmental exposure (Figure 1). While sequence polymorphisms are sometimes included as part of the functional genomics field, this review will place such work under genomics because it represents variations in DNA sequence. Specifically, this review will look at the use of hybridization arrays to study gene expression.

Hybridization arrays have created a wave of interest and skepticism in the past five years. While many scientists view the use of arrays to monitor gene expression for thousands of genes as the dawn of functional genomics (13,19,27), others see the technology as expensive nonhypothesis-driven descriptive research—the ultimate “fishing experiment” (9,47). Both views have valid points, and this controversy can be typical of any new field of study. Although gene expression studies using multiplex hybridization arrays have been performed on a wide range of research topics including cell biology, aging, cancer, environmental toxicity and

drug abuse (20,29,37,42,54,75), performing these experiments in a manner that yields accurate results presents a unique technical challenge.

This review will elaborate the technical underpinnings of hybridization arrays and describe potential problems that must be addressed for reliable determination of gene expression changes. The basics of multiplex hybridization arrays will be presented first, followed by the descriptions of the four different types of array platforms: macroarrays, microarrays, high-density oligonucleotide arrays (Gene Chips™; Affymetrix, Santa Clara, CA, USA) and microelectronic arrays. Next, probe selection and design will be discussed. Subsequently, because all arrays employ the same four basic components: target labeling, target-probe hybridization, detection and data analysis (Figure 2), these steps will be individually discussed. Finally, central aspects of experimental design will be reviewed. This presentation is not intended to advocate a specific experimental approach or any single version of the technology. Neither will this commentary address uses of arrays for multiplex sequencing or polymorphism detection (21,32) (important topics worthy of independent discussion in their own right; “genomics” in Figure 1). Rather, this discussion focuses on the “good practice” use of arrays for monitoring differential gene expression with the realization that the best choice of experimental options depends on the specific application.

ARRAY BASICS

In the past, analysis of gene expression (through measurement of steady-state levels of mRNA) was conducted one gene at a time. Northern blotting, dot blots and quantitative RT-PCR were

the common methods for investigating such changes in gene expression. Northern blot analysis (3) works with sample RNA that has been resolved by agarose electrophoresis and bound to a membrane. A gene-specific probe is then labeled and hybridized to the immobilized RNA, but large amounts of RNA are frequently required, only one gene is analyzed at a time and the approach requires the production of an individual and specific probe for each gene of interest. Dot blots (35,59) are an attempt to increase the throughput of northern blots by eliminating the need for electrophoretic resolution, but they suffer from the same problems of single-gene analysis. Quantitative RT-PCR was subsequently developed in the hope that it would increase the throughput and reduce the sample size needed to measure gene expression. Quantitative RT-PCR certainly is more sensitive and therefore requires less RNA than blotting methods, but this procedure presents unique problems in the form of

designing appropriate amplification standards and characterizing reaction kinetics for each gene of interest (30).

Differential display (43,44), serial analysis of gene expression (SAGE) (71) and total gene expression analysis (TOGA) (63) offer great promise because they are multiplex technologies, but they have limited development and acceptance. The very advantage of these approaches (unbiased screening) is the source of their greatest limitation in that identified genes must then be sequenced, identified and analyzed in a serial fashion. These procedures can be time and labor intensive and prone to false positives. However, these techniques are still the best for gene discovery—that is, finding unknown genes. Hybridization array technology, on the other hand, offers to bypass many of the limitations of these techniques by simultaneously creating labeled copies of multiple sample RNAs and then hybridizing them to many different gene-specific fixed DNA molecules (Figure

2). The nomenclature has developed whereby the labeled sample RNA is termed the target and the individual gene sequences placed on the array are termed probes (Figure 2) (50).

It is noteworthy that, while arrays are increasingly used for gene expression analysis, they only measure relative and not absolute levels of gene transcription. That is, the relative levels of RNAs can be described (e.g., Sample A has 50% more of the specific RNA than Sample B), but absolute amounts (e.g., Sample A has 1000 copies of the RNA and Sample B has 500 copies of the transcript) cannot be determined. Most hybridization arrays are not designed to differentiate between alternatively spliced transcripts of the same gene and, in some cases, between highly homologous members of a gene family. Finally, a change in mRNA does not necessarily correlate with a change in protein (4), and the translated protein often requires further modification to realize its full activity. These latter two points are a common and legitimate criticism of the technology. However, until proteomic technologies (49) (Figure 1) become universally accessible to the research community, hybridization arrays are the best opportunity for studying gene expression on a genomic scale.

ARRAY TYPES

The hybridization array, an ingenious inversion of the northern blot (frequently considered the reverse northern), has spawned a number of different formats. Current array formats can be categorized into four groups: macroarrays, microarrays, high-density oligonucleotide arrays (Gene Chips) and microelectronic arrays. While the nomenclature in the field is vague and variable, we use these terms to describe specific formats. In many places in the literature, all of these formats are referred to as microarrays; however, we will use the term hybridization array to describe the technology as a whole, while the term microarray will describe a specific subset of hybridization arrays. These varying platforms differ according to matrix, probe number/density, array size and type of label. An un-

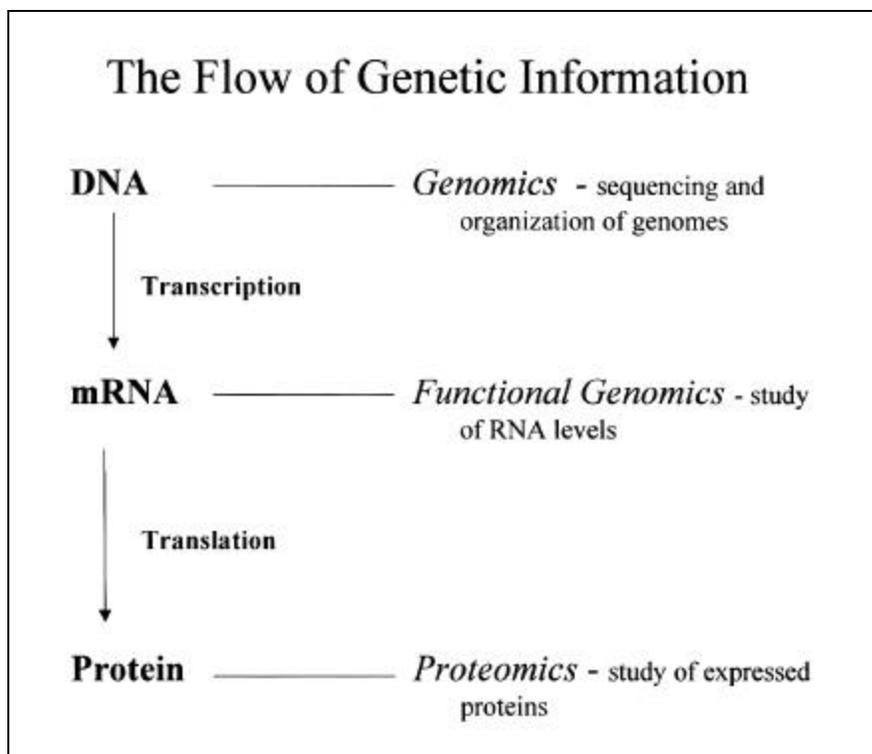


Figure 1. The flow of genetic information. Genetic information flows from DNA into mRNA through transcription and then from mRNA to protein through translation. The different technologies for each level of analysis are indicated. It should be noted that there is some controversy over whether polymorphism analysis should be included in functional genomics. For the present discussion, we chose to include this under genomics because it represents structural variations in DNA sequence, albeit with the potential to represent functional changes.

Review

Understanding of the strengths and weaknesses of each platform is necessary to decide which is appropriate for an individual investigator's research aims.

Macroarrays

Macroarrays (Figure 3A) are generally defined as those arrays that rely on robotically spotted probes that have been immobilized on a membrane-based matrix. The term macroarray as opposed to microarray (discussed next) refers to the generally lower probe density on these arrays. While density varies among arrays, the term macroarray is useful because of other inherent differences of membrane-based arrays. The idea of macroarrays grew from early hybridization experiments against cDNA libraries (31,38). These experiments were combined with improvements in molecular biology and the use of robotic workstations to spot individual cDNA clones onto membranes (48,66,80). Currently, DNA clones,

PCR products or oligonucleotides are all spotted onto membranes as probes. Macroarrays are unique among hybridization arrays in that they use radioactive target labeling, although chemiluminescent labeling has also been described (52). After radioactively labeling the target, different samples are hybridized to individual separate arrays. Phosphorimagers (or less frequently X-ray film) are then used to detect the bound target. These arrays, typically containing between 200 and 5000 genes, are commercially available for a wide variety of organisms and genes and can be obtained from a number of companies. The number of commercially available macroarrays is rapidly expanding, and these companies' offerings are listed on the Internet at www.gene-chips.com/, www.wfubmc.edu/physpharm/genetech/ and www.people.cornell.edu/pages/alm13/chips.html. Custom macroarrays can also be constructed in-house by hand or by robot and can contain as few as a dozen or as

many as thousands of genes. The complexity of generating these arrays is a topic unto itself (15,16,24).

Microarrays

Microarrays (Figure 3B) can be differentiated from macroarrays in three ways. First, microarrays use a glass or plastic slide as a matrix. Second, they use fluorescent dye-labeling detection. Third, they tend to have a larger number and higher density of probes than macroarrays. As with macroarrays, probes are made from either clones, PCR products or oligonucleotides spotted robotically. A leader in this area has been P.O. Brown at Stanford University, who introduced this form of microarray in 1995 (20,55,58). This approach has advantages in that hybridization takes place in a flow cell or small hybridization chamber, which uses a much smaller hybridization solution volume as compared to macroarrays, thereby increasing the relative target concentration. A competitive fluorescent scheme allows both sample groups (control and treated) to be hybridized to the same array and is discussed further in the section on detection. Like macroarrays, an ever-expanding number of microarrays are commercially available and listed on the Internet (www.api.com, www.gene-chips.com/, www.wfubmc.edu/physpharm/genetech/ and www.people.cornell.edu/pages/alm13/chips.html). Many research institutions are currently investing heavily in the equipment to produce custom microarrays in-house. This process has been reviewed (15,16,24) and allows researchers to focus on specific genes of particular interest to their research. One caveat with custom production, however, is that it is an involved effort (requiring intensive bioinformatics and molecular biology), and many groups find that scale-up times are about a year.

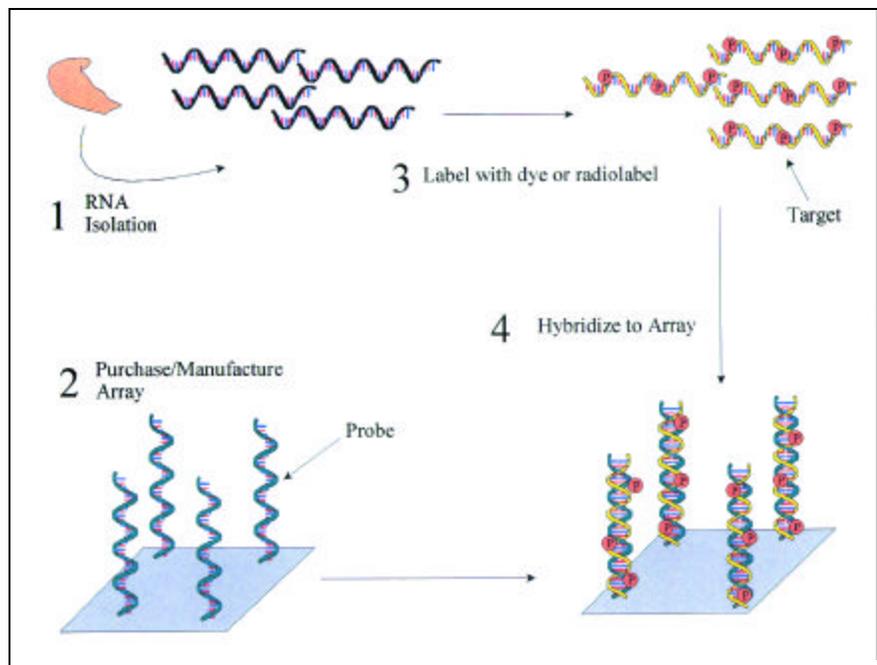


Figure 2. Hybridization array experimental scheme. All hybridization arrays are based on the same four steps. (i) RNA is isolated from samples. (ii) An array with many gene-specific probes for the organism being studied is purchased or constructed. (iii) Labeled targets are created from the sample RNA. (iv) The targets are hybridized to the probes, and the relative signals are measured. In the case of microarrays, fluorescent dyes are used to label the RNA targets (one color for each condition). These are then mixed and used to query the array in a competitive manner. The detection system discriminates the relative signal outputs. High-density oligonucleotide arrays and microelectronic arrays use the same fluorescent dye with a separate array for each sample. In the case of macroarrays, the RNA is converted to radioactively labeled cDNAs, and the control and experimental targets are used in separate hybridization and washing experiments.

High-Density Oligonucleotide Arrays

High-density oligonucleotide arrays (Figure 3C) differ from other formats in that the probe is generated in situ on the surface of the matrix. The leader in these arrays is Affymetrix and their combinatorial synthesis method (40).

Review

This method makes use of a process called photolithography to construct probes on the array surface by making oligonucleotides one base at a time. This synthetic scheme has been described in detail elsewhere (28,46). Because the combinatorial synthesis scheme has a 95% efficiency at each step, synthesis of oligonucleotides longer than 25 bases is problematic. As a result of using these 25-mer oligonucleotides for gene expression analysis, mismatches and spurious target-probe binding can take place because of the limited specificity and binding affinity for a 25 residue oligonucleotide. To

overcome this problem, a series of oligonucleotides that differ by a one-base mismatch from the gene-specific probe are also included on the array and can be used to determine the amount of mismatch hybridization, which can then be subtracted from the signal (41, 78). These arrays, which are available only from Affymetrix, contain between 40 000 and 60 000 probes (including multiple mismatch controls for each gene) and provide the highest density of probes of any array. The greatest hindrance to widespread use of this technology has historically been the high cost and limited organisms available.

Microelectronic Arrays

Microelectronic chips (Figure 3D) are one of the newer formats for hybridization arrays. They are the result of a combination of advances in molecular biology and semiconductor microfabrication techniques. Instead of a membrane or a glass slide platform, these arrays consist of sets of electrodes covered by a thin layer of agarose coupled with an affinity moiety (permitting biotin-avidin immobilization of probes). Each microelectrode is 80 μm in diameter and is capable of generating a current (22,60). The incorporation of

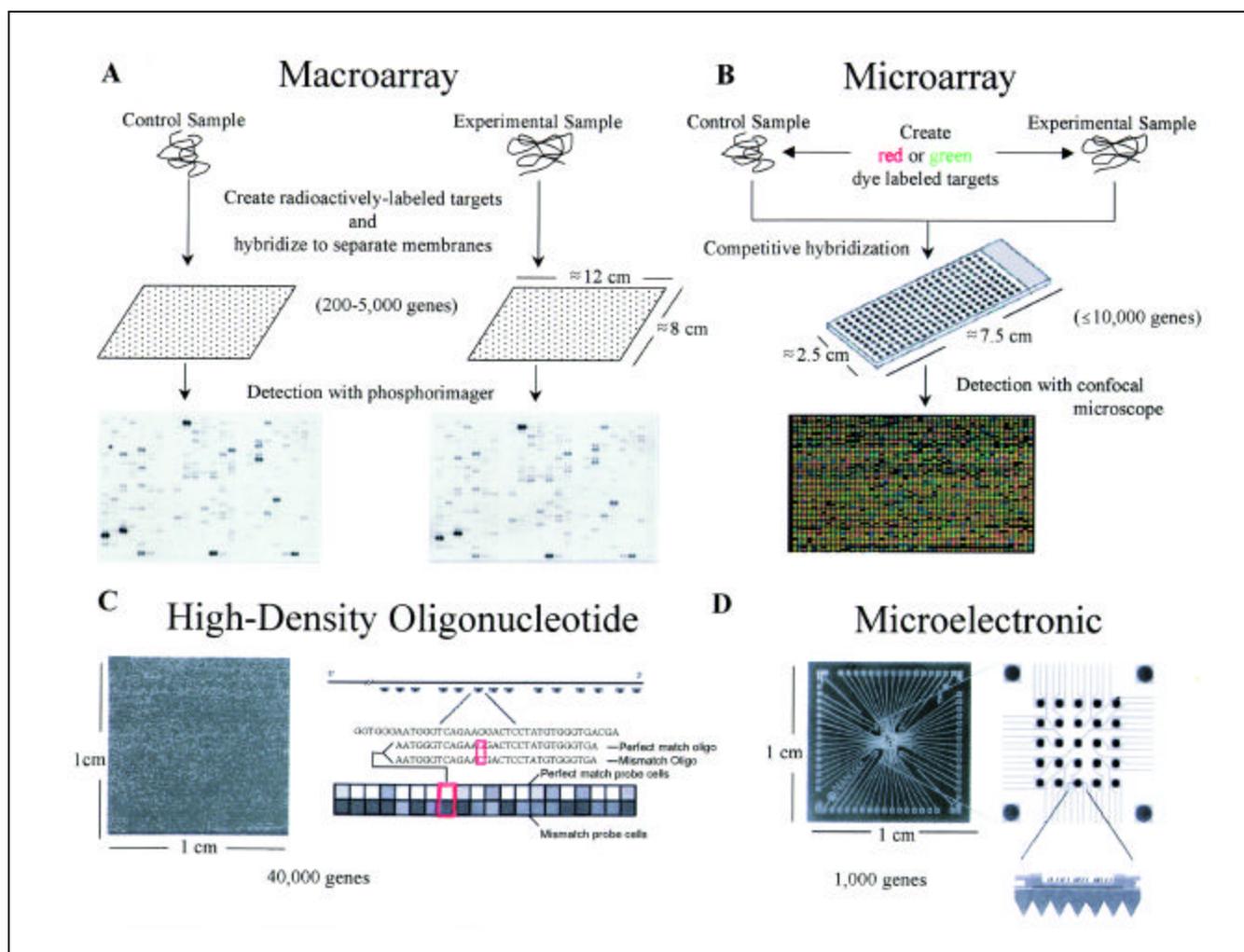


Figure 3. Array types. (A) Macroarrays use a nylon or nitrocellulose membrane matrix onto which nucleic acid probes have been deposited. Samples are radioactively labeled and hybridized in parallel. Detection is best accomplished through phosphorimaging (although film autoradiography can be employed). (B) Microarrays use a silicon or plastic matrix and a fluorescence-based labeling scheme in which both control and experimental samples are hybridized to the same array in a competitive manner. Fluorescent scanners are used for detection. (C) High-density oligonucleotide arrays are constructed by in situ synthesis of probes on the silicon matrix. The probes are limited in length and therefore specificity. A mismatch detection scheme is used to determine specific hybridization. (D) Microelectronic arrays use an electric field generated by individually controllable electrodes to immobilize probes and to control target hybridization. Reversal of the electric field is used for washing the chips.

controllable electric fields gives a new degree of control over probe deposition and target hybridization (see hybridization section) (14,33). This technology is still under development by Nanogen (San Diego, CA, USA) but has great potential because of its ability to move nucleic acids (and proteins) around on the surface of the chip.

Other Formats

A number of alternative formats to hybridization arrays are under development. Xanthon (Research Triangle Park, NC, USA) uses a microplate and electrochemical detection. Lynx Therapeutics (Hayward, CA, USA) has developed a cell sorter-based scheme (10). Interactiva (Ulm, Germany) employs a gold matrix for arrays. Finally, Gene Logic (Berkeley, CA, USA) has produced fiber optic bundle, bead-based arrays. All of these technologies have interesting capabilities, but few published reports are available on their use in gene expression experiments. Their future is dependent on being easier, cheaper and more reliable than current hybridization arrays.

PROBES

Probe design and production are the key developments that have made hybridization arrays possible. Without the wealth of sequence data available from various genome projects and sequence databases, the generation of probes would not be possible. Using these genetic databases, specific probes can be constructed for each gene to be placed on the array. Probes should be both organism- and gene-specific; otherwise, nonspecific hybridization can compromise the experiment. Some probes can be used across species, among humans and higher primates, for example, where there is a high degree of sequence identity (29). Probes for homologous gene families should also be constructed to hybridize to regions with the lowest sequence conservation. Physically, probes take three different forms: complex cDNA clones, PCR products and oligonucleotides. Clones, PCR products and oligonucleotides are used with macroarrays, microarrays

and microelectronic arrays, whereas high-density oligonucleotide arrays use only short oligonucleotide probes. Clone-based probes are spotted onto arrays as whole genes or fragments of genes and are usually obtained from clone libraries. Frequently, the sequences of these clone-based probes are not specifically designed for regions of low homology to other genes. This can result in nonspecific hybridization. Clone probes can be anywhere from a few hundred base pairs to several kilobases. PCR Product probes, when properly designed, represent non-homologous regions of a gene generated by PCR from a cDNA clone, library or RNA and tend to be 200–500 bp in length. Oligonucleotide probes differ from the other probes in that they can be deposited by printing or synthesized in situ on the array matrix by photolithography (28,45). In situ synthesized oligonucleotide probes generally do not exceed 25 bases in length because of reaction efficiency limits. Because 25 bases are not often enough to be gene specific, a number of different oligonucleotide probes are used per gene on high-density oligonucleotide arrays with deliberate mismatches incorporated to discriminate mispriming (41). One limitation of this process is that high-density oligonucleotide arrays generally cannot be used across species, and custom arrays cannot be readily produced by researchers as with other formats.

TARGET LABELING

To measure the amount of probe-target hybridization, a labeled target must be created. This is usually accomplished by a reverse transcription reaction where one of the nucleotides in the reaction mixture contains either a radioactive or fluorescent label. A label as such is any chemical or radiochemical that can later be used for detection.

The most basic method for creating a labeled cDNA is reverse transcription with the retroviral enzyme reverse transcriptase. Reverse transcription requires a primer for initiation, which can be oligo(dT) (to target 3' ends of messenger RNAs), random hexamers (to indiscriminately copy all RNA) or gene-spe-

cific primers. While gene-specific primers reduce noise by copying only the genes for which there are probes on the hybridization array, their use also requires the production of hundreds or thousands of primers. Other labeling techniques include amplified antisense RNA (70,73), tyramide signal amplification (TSA) (36), strand displacement amplification (SDA) (76) and PCR (62,69). These methods have the advantage of signal amplification during labeling by creating multiple copies or labels. This helps overcome problems associated with low expression of some genes or small sample sizes. The kinetics of the amplification step, however, must be exactly reproducible in these approaches; otherwise, changes observed on the array could be the result of differential amplification artifacts.

HYBRIDIZATION

Hybridization and washing are critical to the generation of high-quality hybridization array gene expression data. Regardless of the format, the even distribution of both hybridization and wash solutions is important for optimization of target-probe interactions and minimization of background, respectively. Recall that the purpose of hybridization is to enable complementary target and probe sequences to specifically bind. To accomplish this, the hybridization solution must be evenly mixed to permit targets (in the hybridization solution) to be evenly distributed over the probes (on the solid matrix). Similarly, wash solutions must be evenly distributed over the array to remove nonhybridized probes, reduce nonspecific hybridization and minimize background. In that there are several physical types of arrays, there are also different ways in which they can be hybridized and washed. In this section, three different hybridization techniques will be reviewed: (i) traditional nylon membrane methods, (ii) glass microscope slides in flow cells and (iii) electric field-enhanced methods. Generally, macroarrays use traditional membrane methods, microarrays and high-density oligonucleotide arrays use flow cells, and microelectronic chips use a directed electric field in solution.

Traditional Methods

Since Edwin Southern's introduction of nucleic acid blotting (61), most hybridization schemes have relied on hybridization taking place in sealed bags or solution-filled bottles. Hybridization bottles are designed to roll or agitate in a temperature-controlled environment. These bottles are also used for the sequentially more stringent washes (lower salt concentration and higher temperatures) that serve to minimize background. Conveniently, most nylon membrane-based arrays are hybridized in this way. Therefore, molecular biology researchers familiar with membrane hybridization techniques can immediately use existing hybridization equipment for macroarray experiments.

A technical challenge exists, however, in that arrays are hybridized and washed in different bottles and yet are compared to each other. Differences in hybridization or washing between the two membranes can result in inconsistent, poor-quality data. Therefore, when using this technique, great care must be taken to ensure identical conditions during hybridization and washing of both macroarrays in an experiment.

Flow Cells

Microarrays and high-density oligonucleotide arrays use a glass slide or silicon chip matrix that is anchored to the bottom of a sealed chamber through which solutions can be directed. Temperature is controlled by a metal plate beneath the slide. Alternatively, small hybridization cassettes, into which the solutions can be added, can be used in water baths. In either format, temperature and volume control are optimized, and consistent hybridization and washing steps are attainable. Though most laboratories are not equipped with this hardware, many researchers believe that the increased throughput and hybridization efficiency make the technology worth the investment.

Electric Field

The previous two hybridization techniques rely on the random, passive interaction between matched targets and probes. The passive approach has inherent difficulties. Different target-

probe pairs have different melting temperatures and hybridization kinetics. The temperature and salt concentrations of hybridization solutions represent the best conditions for the probe population mean and therefore will not be optimal for some pairs. In addition, the kinetics of hybridization are dependent on the target concentration in solution, and low concentration targets will require a longer time to hybridize, if they hybridize at all. Using an electric field to control hybridization can potentially eliminate many of the problems of passive hybridization.

During electric field-driven hybridization (or active hybridization), a current is generated at a single test site, drawing all targets in the hybridization solution toward the site because of DNA's inherent negative charge. Active attraction increases the effective concentration of target at a probe site and thus speeds the process to such an extent that only minutes are required for hybridization, compared to hours for passive hybridization. Next, by reversing the current, the site can be washed of any mismatches and spurious binding because any incompletely hybridized targets will not possess the hydrogen bonding strength required to overcome the repulsive forces of the current. A further advantage of this technique is that multiple samples can be assayed by the sequential hybridization of individual samples to different sets of probes on the same array. This technology is currently under development (Nanogen) and, if it can be made inexpensive enough, could represent the future of hybridization arrays (14,33).

DETECTION

Whether elucidating developmental gene expression changes or characterizing a tumor cell line, array experiments all share a common thread in their design—creating a clear picture. Thus, sensitive and accurate signal output detection is imperative. In this section, two methods of detection will be discussed: (i) radioisotopic and (ii) fluorescence.

Radioisotopic Detection

Detection of nucleic acids on macro-

Review

arrays is most commonly accomplished through the use of a radioactive isotope in conjunction with an imaging plate/phosphorimaging system. After hybridization and washing, the array is apposed to an image plate, and the signal intensity is detected when the plate emulsion is excited by the high-energy radioactive particles (recall that the labeling step yielded a radiolabeled cDNA) (53). Following an appropriate exposure time, the image plate is developed by scanning with a laser of an appropriate wavelength, thus releasing the energy of the excited electrons. This release is detected by the phosphorimager yielding a digital image of the radioactivity (both location and intensity). After image capture, analysis is performed using computer software that ascribes signal intensities to all of the pixel within each spot. In this way, the intensity of a gene-specific spot on the control array can be compared to the same spot on the exper-

imental array. Classically, X-ray film has been used to image radioactivity; however, phosphorimaging's superior dynamic range, shorter exposure times and direct digital output have made it the method of choice.

Fluorescence Detection

Fluorescence detection is used with microarrays, high-density oligonucleotide arrays and microelectronic chips. The main reason for using fluorescence is that gene expression experiments can use a competitive hybridization scheme in which both experimental and control samples are hybridized to the same array. It is possible to differentiate the signals from the control and experimental samples because of a fluorescence phenomenon known as the Stokes shift. The Stokes shift of a fluorescent tag is the difference between its excitation and emission wavelengths. Fluorescent

tags are compounds that absorb light of a certain wavelength, exciting electrons within the compound to a higher energy state. From this higher energy state, photons are emitted at a specific wavelength for each compound. Therefore, if two samples (control and experimental) are labeled with different fluorescent tags having different emission wavelengths, their respective signals can be detected separately within the same experiment. Because the emission of energy is generally linear, the amount of signal detected is relative to the amount of fluorescently labeled cDNA hybridized to the probe. For example, one sample is labeled with a red dye and the other is labeled with a green dye. When the samples are mixed and used to query an array, if there is more mRNA for a given gene in one sample, there will be more dye of that particular color and, following hybridization, a predominance of that dye signal over the other.

The advantage of a competitive hybridization scheme such as this is that comparison of inter-array hybridization, as seen with the macroarray format, is eliminated.

Fluorescence is generally detected using a confocal scanning microscope, which uses lasers for generating light of the correct excitation wavelength and photomultiplier tubes for detection (57). The confocal nature of these devices is such that having a pair of focal points restricts the detection of light to one direction and reduces light from spurious sources above, below or to the side of the part of the array being scanned. A drawback of confocal scanners, however, is that each excitation wavelength must have its own expensive laser, and they therefore tend to be costly. Charge-coupled device (CCD) detectors have been used in an attempt to provide more flexibility because they can use a single light source with different filters for different excitation wavelengths. The CCD detectors are hampered by hardware issues, however, and time will tell if they can become a competitor of the confocal microscope.

DATA ANALYSIS

The creation, hybridization and detection of hybridization arrays may seem to be a daunting task. It would appear that once an image of the array, with relative intensities for each sample, has been generated, the experiment would nearly be finished. Unfortunately, this is not the case as scientists are now learning that the massive amounts of data generated by arrays pose a new challenge (6,11,26,65,72). In this section, basics in array normalization and data management will be presented, along with a look at the mathematical models being developed to help create a relevant biological story from expression data.

Basics

The first steps in data analysis are background subtraction and normalization. The principals of both are similar to the techniques used with conventional nucleic acid or protein blotting. Background subtraction pulls the non-

specific background noise out of the signal detected for each spot and allows comparison of specific signals. For illustration, if the signal intensities for the control and experimental spots are 4 and 6, respectively, it would appear that the experimental is 50% higher. However, if a background of 2 is subtracted from both signal intensities, the experimental value is actually 100% higher than the control. Background is often taken from the blank areas on the array. A complication to background subtraction is that variations in background across the array can affect some spots more than others. An alternative is to use either a local background for the area around each spot or spots with the lowest signal intensities for background determination. The latter may be a more accurate determination of non-specific background because it represents the nonspecific binding of targets to probe. Background intensities from blank areas (no nucleic acids) do not contain probe and are arguably a different form of background.

Normalization is the process of accounting for differences between separate arrays. All macroarray and high-density oligonucleotide experiments and any other multiple array experiments may require the use of normalization for consistent comparisons. For example, when a pair of macroarrays representing control and treated samples shows a difference in overall or total signal intensity, such differences can arise from unequal starting amounts of RNA or cDNAs, from labeling reactions of different efficiencies or from subtle differences in hybridization efficiencies. Any of these factors can skew the results. One common method of normalization is to use a housekeeping gene, thought to be invariant under experimental conditions, for comparison (74). If the signal for this gene is higher on one array than the other, it can be used to normalize the data. Housekeeping genes are problematic for many experiments because housekeeping genes do in fact vary under some experimental conditions (64). To overcome the variability of these genes, some researchers have turned to a "basket" or "sum" approach for normalization. This strategy is based on the precept that the overall or global levels of gene expression are

unaffected by the experimental conditions, and changes are equally distributed between inductions and reductions. Therefore, arrays can be normalized by taking the sum of the intensities for all control and experimental spots and equilibrating them. Even this approach is invalid for some experiments, however, and exogenous synthetic RNA standards have been used to normalize (23).

Clustering Analysis

Because the sheer quantity of data generated by arrays exceeds the ability of manual human assessment, new modes of analysis are needed to effectively and exhaustively explore hybridization array results. To this end, there have been advances in computational biology and bioinformatics that have helped to create a biological story out of the databases generated by hybridization array data (6,18,79). As this topic warrants an in-depth review of its own, the current discussion will provide only a basic overview of histograms and clustering algorithms.

For binary experiments, where there is a control and a single experimental condition, a common first step in data analysis is to create a histogram. Histograms serve to rank genes in order of magnitude of change from greatest induction to greatest reduction (29,51). Moreover, the histogram provides a visual sense of the distribution (between induction and reduction) and any skew to the results. Such a comparison highlights those genes that show the greatest fold (+ or -) changes. These genes can subsequently be made into a simple list and examined further at the level of protein or activity.

For more complex experimental designs, where there are two or more experimental conditions (typically, an experiment looking at multiple time-points, doses or groups), the computational requirements are much greater. The question then becomes not one of a simple change under one condition, but of how one gene (out of thousands) changes over multiple conditions. The first step in understanding such complex relationships is to plot the expression of each gene over the various conditions. With large experiments analyzing thousands of genes, these

data increase dramatically, and as a result, it can be difficult to find patterns in the data. To this end, clustering algorithms have been employed. Stated simply, these approaches seek to find groups of genes, or clusters, that behave similarly across the experimental conditions. Clusters and the genes within them can subsequently be examined for commonalities in function or sequence to better understand how and why they behave similarly (17). A number of different methods: k-means, self-organizing maps, hierarchical clustering, vector machines and Bayesian statistics are employed for clustering analysis (2,7,12,25,34,67,68,75,77). In contrast to an expression histogram that focuses only on the genes whose expression changes, an advantage of clustering analysis is that it uses all of the data generated. In a multi-endpoint experiment, clustering analysis can also identify groups of genes that do not change their expression pattern under some, or all, of the conditions. The fact that a group of genes is unperturbed or unresponsive in a specific experimental condition can lead to valuable biological insights. Clustering analyses will be critical for the mining of public expression databases that are being generated (References 8, 18 and Web sites: www.cmgm.stanford.edu/pbrown/, www.nhgri.nih.gov/DIR/LCG/15K/HTML/index.html and www.arraydata.org).

Clustering techniques are also being used to combine gene expression data generated from hybridization arrays with other experimental data. This is the likely direction of functional genomic research in the future. The goal is being able to make correlations between seemingly disparate measures such as biochemical, electrophysiological or behavioral observations. This type of clustering, which combines gene expression with other data, has begun to be demonstrated in the cancer field (54,56).

EXPERIMENTAL DESIGN

All successful science is based on sound experimental design. From a practical standpoint, this is especially true of hybridization array experiments because the time and resources that can be wasted on poorly designed function-

al genomic research is staggering. For both the beginning researcher and those already conducting experiments using hybridization arrays, it is worth examining the concerns of sample collection, fold changes, sensitivity and post-hoc confirmation.

Sample Collection

Sample collection is a basic element of experimental design for many molecular biological experiments, but it is worth reiterating. Depending on the cells or tissue being examined, it is often unavoidable that a sample will contain multiple cell types. In complex samples such as brain tissue, there is routinely a heterogeneous cell population. Therefore, observed changes may represent a change in one cell type or all cell types. Similarly, smaller changes occurring in only one cell type may be hidden. Thus, researchers must be mindful of this heterogeneous population when drawing conclusions. Similarly, in comparing normal and cancer samples, there will be obvious differences in the proportion of the cell types (i.e., cancer cells will be over-represented). Therefore, interpretations of differences in gene expression may be complicated by the sheer mass of one cell over another. A promising technological solution is laser capture microdissection, which allows very small and identified cellular populations to be dissected (42).

The timing of tissue collection goes hand-in-hand with the nature of the collected tissue, and, therefore, sample collection times will be important. For example, in an experiment in which cells undergo programmed cell death, the collection time point will determine if causative changes or endpoint changes are observed. If a late timepoint is chosen, it becomes increasingly difficult to distinguish changes due to the general breakdown of cellular processes from those that have triggered cell death.

Detection Sensitivity

Detection sensitivity in array research takes two very distinct forms. The first, termed threshold sensitivity, is the ability to detect one RNA species out of a population and is a concern for rarely expressed messages, for small

sample sizes and is the traditional issue of sensitivity common to other techniques. The second sensitivity parameter is "fold-change sensitivity", the ability of hybridization arrays to reliably determine a certain magnitude difference in expression. The claimed fold-change sensitivity of different platforms varies from 2- to 0.75-fold for fluorescence-based protocols and 1.5- to 0.66-fold for radioactivity-based methods. Determination of this parameter is crucial to characterizing the technology and ensuring that researchers choose the technology most appropriate to their goals (5). For research involving systems that undergo large gene expression changes (e.g., yeast cell-cycle regulation or organ developmental processes where tenfold changes are expected), one can detect such changes with fluorescent protocols. Other research efforts, for example in neuroscience, where gene changes are less dramatic, may find radioactivity-based methods more applicable.

Post-Hoc Confirmation

One of the most common criticisms of hybridization arrays is that when hundreds or thousands of genes are examined at once, some will appear changed by random chance. This is a statistical reality and highlights the requirement for post-hoc confirmation of changes seen with arrays. This is because a single array experiment, representing an n of one, lacks the sample size needed for statistical analysis. Therefore, tests on individual samples themselves are necessary to produce statistical significance. Such corroborating experiments can examine the gene changes at the level of mRNA (Northern blot, quantitative RT-PCR), protein (immunoblot) or activity (enzymatic activity, DNA binding or other measures). The protein and activity tests are recommended because they assess the gene of interest at a level closer to function or actually address the function itself. Protein analysis is important because increased levels of transcription do not always translate into increased levels of protein (4). In addition, protein assessment is achieved with fundamentally different experimental techniques and may not be subject to the same sources

of error as the array. Unfortunately, immunoblotting and activity assays would appear to return researchers to the single-gene assay that hybridization arrays were intended to avoid. This is not true in practice because large amounts of genes have already been screened by the array. The optimal solution to ascribing relevance to the data is to develop techniques by which confidence intervals for individual genes can be generated from arrays and these results can be combined with proteomic techniques under development (49). Alternatively, as costs are decreased, experiments will ultimately permit hybridization arrays for each sample. Many researchers are exploring the use of small (in terms of the number of genes) arrays that focus on a specific gene family or pathway. Regardless of the confirmation strategy used, the ultimate goal of hybridization array-based experiments is not technical showmanship but the creation of biological narratives to illuminate the issue at hand.

CONCLUSION

Potential applications for hybridization arrays exist anywhere the levels of gene expression are of interest. As delineated in this discussion, the technology will have to be optimized and tested for each area of application. Foremost is the need for probes specific to many different animal and plant models. This will be made possible by the growing number of genome projects for model organisms [e.g., *Arabidopsis thaliana* (39) and *Drosophila melanogaster* (1)]. Similarly, targeted (custom) arrays will need to be generated for areas such as cancer biology, neuroscience and developmental biology.

As detailed in this review, DNA hybridization arrays will revolutionize gene expression research. Like any new method, researchers will have to be simultaneously aware of its power and its limitations. The scientific community and marketplace have embraced this technology, as witnessed by the research funding opportunities to academics and capitalization of array-based companies. Hybridization arrays will therefore clearly be a major tool of the postgenome era.

ACKNOWLEDGMENTS

The authors wish to thank Steve Walker, Nancy Ayers, Drake Morgan and Karen Brebner for providing valuable editorial comments on the manuscript. This work was supported by National Institutes of Health grant nos. P50DA06643 and P50AA11997 (to K.E.V.) and T32DA07246 (to W.M.F.).

REFERENCES

- Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195.
- Alon, U., N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack and A.J. Levine. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* 96:6745-6750.
- Alwine, J.C., D.J. Kemp and G.R. Stark. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. USA* 74:5350-5354.
- Anderson, L. and J. Seilhamer. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533-537.
- Baldwin, D., V. Crane and D. Rice. 1999. A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. *Curr. Opin. Plant Biol.* 2:96-103.
- Bassett, D.E., M.B. Eisen and M.S. Boguski. 1999. Gene expression informatics—it's all in your mine. *Nat. Genet.* 21:51-55.
- Ben-Dor, A., R. Shamir and Z. Yakhini. 1999. Clustering gene expression patterns. *J. Computational Biol.* 6:281-297.
- Brazma, A., A. Robinson, G. Cameron and M. Ashburner. 2000. One-stop shop for microarray data—is a universal, public DNA-microarray database a realistic goal? *Nature* 403:699-700.
- Brenner, S. 1999. Sillycon valley fever. *Curr. Biol.* 9:R671.
- Brenner, S., M. Johnson, J. Bridgman, G. Golda, D.H. Lloyd, D. Johnson, S.J. Luo, S. McCurdy et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18:630-634.
- Brent, R. 1999. Functional genomics: learning to think about gene expression data. *Curr. Biol.* 9:R338-R341.
- Brown, M.P.S., W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares and D. Haussler. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* 97:262-267.
- Brown, P.O. and D. Botstein. 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21:33-37.
- Cheng, J., E.L. Sheldon, L. Wu, A. Uribe, L.O. Gerrue, J. Carrino, M.J. Heller and J.P. O'Connell. 1998. Preparation and hybridization analysis of DNA/RNA from *E. coli* on microfabricated bioelectronic chips. *Nat. Biotechnol.* 16:541-546.
- Cheung, V.G., M. Morley, F. Aguilar, A. Massimi, R. Kucherlapati and G. Childs. 1999. Making and reading microarrays. *Nat. Genet.* 21:15-19.
- Clark, M.D., G.D. Panopoulou, D.J. Cahill, K. Busow and H. Lehrach. 1999. Construction and analysis of arrayed cDNA libraries. *Methods Enzymol.* 303:205-233.
- Claverie, J.M. 1999. Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.* 8:1821-1832.
- Claverie, J.M. 2000. Do we need a huge new centre to annotate the human genome? *Nature* 403:12.
- DeRisi, J.L., V.R. Iyer and P.O. Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278:680-686.
- DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su and J.M. Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14:457-460.
- Drmanac, R. and S. Drmanac. 1999. cDNA screening by array hybridization. *Methods Enzymol.* 303:165-178.
- Edman, C.F., D.E. Raymond, D.J. Wu, E. Tu, R.G. Sosnowski, W.F. Butler, M. Nerenberg and M.J. Heller. 1997. Electric field directed nucleic acid hybridization on microchips. *Nucleic Acids Res.* 25:4907-4914.
- Eickhoff, B., B. Korn, M. Schick, A. Poustka and J. van der Bosch. 1999. Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res.* 27:e33.
- Eisen, M.B. and P.O. Brown. 1999. DNA arrays for analysis of gene expression. *Methods Enzymol.* 303:179-205.
- Eisen, M.B., P.T. Spellman, P.O. Brown and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95:14863-14868.
- Ermolaeva, O., M. Rastogi, K.D. Pruitt, G.D. Schuler, M.L. Bittner, Y. Chen, R. Simon, P. Meltzer, J.M. Trent and M.S. Boguski. 1998. Data management and analysis for gene expression arrays. *Nat. Genet.* 20:19-23.
- Fields, S., Y. Kohara and D.J. Lockhart. 1999. Functional genomics. *Proc. Natl. Acad. Sci. USA* 96:8825-8826.
- Fodor, S.P., J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu and D. Solas. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-773.
- Freeman, W.M., D.J. Robertson, K. Brebner, W.J. Lynch, S.H. Nader, M.A. Nader, L. Gioia, J.B. Daunias et al. 2000. Functional genomics of cocaine in non-human primates and rats. *Society for Neuroscience, Abstract.*
- Freeman, W.M., S.J. Walker and K.E. Vrana. 1999. Quantitative RT-PCR: pitfalls and potential. *BioTechniques* 26:112-122.
- Gress, T.M., J.D. Hoheisel, G.G. Lennon, G. Zehetner and H. Lehrach. 1992. Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm. Genome* 3:609-619.
- Hacia, J.G. 1999. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet.* 21:42-47.
- Heller, M.J., A.H. Forster and E. Tu. 2000. Active microelectronic chip devices which utilize controlled electrophoretic fields for multiplex DNA hybridization and other genomic applications. *Electrophoresis* 21:157-164.
- Hilsenbeck, S.G., W.E. Friedrichs, R. Schiff, P. O'Connell, R.K. Hansen, C.K. Osborne and S.A.W. Fuqua. 1999. Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl. Cancer Inst.* 91:453-459.
- Kafatos, F.C., C.W. Jones and A. Efstratiadis. 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.* 7:1541-1552.
- Kerstens, H.M., P.J. Poddighe and A.G. Hanselaar. 1995. A novel in situ hybridization signal amplification method based on the deposition of biotinylated tyramine. *J. Histochem. Cytochem.* 43:347-352.
- Lee, C.K., R.G. Klopp, R. Weindrich and T.A. Prolla. 1999. Gene expression profile of aging and its retardation by caloric restriction. *Science* 285:1390-1393.
- Lennon, G.G. and H. Lehrach. 1991. Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* 7:314-317.
- Lin, X.Y., S.S. Kaul, S. Rounsley, T.P. Shea, M.I. Benito, C.D. Town, C.Y. Fujii, T. Mason et al. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402:761-768.
- Lipshutz, R.J., S.P. Fodor, T.R. Gingeras and D.J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21:20-24.
- Lockhart, D.J., H. Dong, M. Bryne, M. Follettie, M. Gallo, M. Chee, M. Mittmann, C. Wang et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* 14:1675-1680.
- Luo, L., R.C. Salunga, H.Q. Guo, A. Bittner, K.C. Joy, J.E. Galindo, H.N. Xiao, K.E. Rogers et al. 1999. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat. Med.* 5:117-122.
- Martin, K.J. and A.B. Pardee. 1999. Principles of differential display. *Methods Enzymol.* 303:234-258.
- Matz, M.V. and S.A. Lukyanov. 1998. Different strategies of differential display: areas of application. *Nucleic Acids Res.* 26:5537-5543.
- McGall, G., A. Barone, M. Diggelmann, S.P. Fodor, E. Gentalen and N. Ngo. 1997. The efficiency of light-directed synthesis of DNA arrays on glass substrates. *J. Am. Chem. Soc.* 119:5081-5090.
- McGall, G., J. Labadie, P. Brock, G. Wallraff, T. Nguyen and W. Hinsberg. 1996. Light-directed synthesis of high-density

- oligonucleotide arrays using semiconductor photoresists. *Proc. Natl. Acad. Sci. USA* 93:13555-13560.
47. **Mir, K.U.** 2000. The hypothesis is there is no hypothesis. *Trends Genet.* 16:63-64.
 48. **Nguyen, C., D. Rocha, S. Granjeaud, M. Baldit, K. Bernard, P. Naquet and B.R. Jordan.** 1995. Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* 29:207-216.
 49. **Pandey, A. and M. Mann.** 2000. Proteomics to study genes and genomes. *Nature* 405:837-846.
 50. **Phimister B.** 2000. Going global. *Nat. Genet.* 21:1.
 51. **Pietu, G., O. Alibert, V. Guichard, B. Lamy, F. Bois, E. Leroy, R. Mariage-Sampson, R. Houlgatte, P. Soularue and C. Auffray.** 1996. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* 6:492-503.
 52. **Rajeevan, M.S., I.M. Dimulescu, E.R. Unger and S.D. Vernon.** 1999. Chemiluminescent analysis of gene expression on high-density filter arrays. *J. Histochem. Cytochem.* 47:337-342.
 53. **Robertson, D.J., W.M. Freeman and K.E. Vrana.** Phosphorimaging. *In Encyclopedia of Life Sciences.* (In Press).
 54. **Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey et al.** 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24:227-235.
 55. **Schena, M., D. Shalon, R.W. Davis and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467-470.
 56. **Scherf, U., D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold et al.** 2000. A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* 24:236-244.
 57. **Schermer, M.J.** 1999. Confocal scanning microscopy in array detection. *In* M. Schena (Ed.), *DNA Microarrays: A Practical Approach.* Oxford University Press, Oxford.
 58. **Shalon, D., S.J. Smith and P.O. Brown.** 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6:639-645.
 59. **Sim, G.K., F.C. Kafatos, C.W. Jones, M.D. Koehler, A. Efstratiadis and T. Maniatis.** 1979. Use of a cDNA library for studies on evolution and developmental expression of the chorion multigene families. *Cell* 18:1303-1316.
 60. **Sosnowski, R.G., E. Tu, W.F. Butler, J.P. O'Connell and M.J. Heller.** 1997. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc. Natl. Acad. Sci. USA* 94:1119-1123.
 61. **Southern, E.M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J.D. Mol. Biol.* 98:503-517.
 62. **Spirin, K.S., A.V. Ljubimov, R. Castellon, O. Wiedoeft, M. Marano, D. Sheppard, M.C. Kenney and D.J. Brown.** 1999. Analysis of gene expression in human bullous keratopathy corneas containing limiting amounts of RNA. *Invest. Ophthalmol. Vis. Sci.* 40:3108-3115.
 63. **Sutcliffe, J.G., P.E. Foye, M.G. Erlander, B.S. Hilbush, L.J. Bodzin, J.T. Durham and K.W. Hasel.** 2000. TOGA: an automated parsing technology for analyzing expression of nearly all genes. *Proc Natl. Acad. Sci. USA* 97:1976-1981.
 64. **Suzuki, T., P.J. Higgins and D.R. Crawford.** 2000. Control selection for RNA quantification. *BioTechniques.* 29:332-337.
 65. **Szallasi, Z.** 1999. Genetic network analysis in light of massively parallel biological data acquisition. *Pac. Symp. Biocomp.* 5-16.
 66. **Takahashi, N., H. Hashida, N. Zhao, Y. Misumi and Y. Sakaki.** 1995. High-density cDNA filter analysis of the expression profiles of the genes preferentially expressed in human brain. *Gene* 164:219-227.
 67. **Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander and T.R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96:2907-2912.
 68. **Toronen, P., M. Kolehmainen, C. Wong and E. Castren.** 1999. Analysis of gene expression data using self-organizing maps. *FEBS Lett.* 451:142-146.
 69. **Trenkle, T., J. Welsh and M. McClelland.** 1999. Differential display probes for cDNA arrays. *BioTechniques* 27:554-560.
 70. **Van Gelder, R.N., M.E. von Zastrow, A. Yool, W.C. Dement, J.D. Barchas and J.H. Eberwine.** 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* 87:1663-1667.
 71. **Velculescu, V.E., L. Zhang, B. Vogelstein and K.W. Kinzler.** 1995. Serial analysis of gene expression. *Science* 270:484-487.
 72. **Vingron, M. and J. Hoheisel.** 1999. Computational aspects of expression data. *J. Mol. Med.* 77:3-7.
 73. **Wang, E., L.D. Miller, G.A. Ohnmacht, E.T. Liu and F.M. Marincola.** 2000. High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* 18:457-459.
 74. **Warrington, J.A., A. Nair, M. Mahadevappa and M. Tsyganskaya.** 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping maintenance genes. *Physiol. Genomics* 2:143-147.
 75. **Weinstein, J.N., T.G. Myers, P.M. O'Connor, S.H. Friend, A.J. Fornace, K.W. Kohn, T. Fojo, S.E. Bates et al.** 1997. An information-intensive approach to the molecular pharmacology of cancer. *Science* 275:343-349.
 76. **Westin, L., X. Xu, C. Miller, L. Wang, C.F. Edman and M. Nerenberg.** 2000. Anchored multiplex amplification on a microelectronic chip array. *Nat. Biotechnol.* 18:199-204.
 77. **Wittes, J. and H.P. Friedman.** 1999. Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. *J. Natl. Cancer Inst.* 91:400-401.
 78. **Wodicka, L., H. Dong, M. Mittmann, M. Ho and D.J. Lockhart.** 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 15:1359-1366.
 79. **Zhang, M.Q.** 1999. Large-scale gene expression data analysis: a new challenge to computational biologists. *Genome Res.* 9:681-688.
 80. **Zhao, N., H. Hashida, N. Takahashi, Y. Misumi and Y. Sakaki.** 1995. High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene* 156:207-213.

Address correspondence to:

Dr. Kent E. Vrana
 Department of Physiology and Pharmacology
 Wake Forest University School of Medicine
 Medical Center Blvd.
 Winston-Salem, NC 27157, USA
 e-mail: kvrana@wfbmc.edu