



Review

The microarray: potential applications for ophthalmic research

Ann S. Wilson,¹ Bridget G. Hobbs,² Terence P. Speed,² P. Elizabeth Rakoczy³

¹Department of Molecular Ophthalmology, Lions Eye Institute and ³the Centre for Ophthalmology and Vision Science, University of Western Australia, Nedlands, WA, Australia; ²Division of Genetics and Bioinformatics, Walter and Eliza Hall Institute of Medical Research, Melbourne, VIC, Australia

The microarray is a revolutionary technology combining molecular biology and computer technology in the high throughput, simultaneous analysis of global gene expression. It is emerging as a powerful and valuable research tool that holds great promise in elucidating the molecular mechanisms involved in complex diseases. The information gained may provide direction toward identifying appropriate targets for therapeutic intervention. Despite the enormous potential of this technology, however, a number of issues exist that complicate gene expression analysis and require further resolution. This paper reviews these issues as well as the conceptual, practical and statistical aspects of microarray technology, including its current use in research and clinical applications. Furthermore, the advantages and potential benefits of this technology in ophthalmic research are discussed, with particular attention to retinal diseases, and its possible application in the identification of genes involved in ocular disease progression that may serve as clinical markers or potential therapeutic targets.

Accumulative evidence suggests that increasing numbers of human diseases, both acquired and genetic, result fundamentally from alterations in DNA sequence, or altered gene expression patterns. With the completion of the Human Genome Project, the possibility of linking specific genes to the susceptibility of particular diseases is improving. However, most diseases are complex and polygenic, resulting from the cooperation of a variety of genes in the initiation and progression of the diseased state. Therefore, gene expression studies or functional genetics are needed to elucidate the underlying molecular mechanisms that determine the physiological state of the cell, particularly in disease.

Currently, Northern [1] and dotblot analysis [2,3], reverse transcriptase-coupled polymerase chain reaction (RT-PCR), differential display [4] and serial analysis of gene expression (SAGE) [5] are techniques used for studying gene expression at the mRNA level. However, these techniques are time consuming and are associated with a number of shortfalls. Traditionally, information on specificity and relative abundance of expression products are derived by semi-quantitative techniques such as the Northern and dotblot analyses and ribonuclease protection assays, which often require large amounts of RNA for the study of a limited number of genes. The more sophisticated methods of differential display and SAGE have been used to screen a larger subset of complementary DNA (cDNA) clones. However, these too suffer from limitations such as cost and technical issues (i.e., are labor intensive, can require complex sample preparation, and particularly with SAGE, dependence on the availability of high-throughput se-

quencing and the complete annotation of the genome) rendering these techniques non-conducive for rapid, comprehensive genomic investigation.

A new and advanced technology has emerged allowing the study of gene expression on a global level: the microarray. The microarray enables the parallel screening of a plethora of genes [6,7] and facilitates the comparative analysis of a large number of samples [8,9]. This technology has countless advantages over conventional techniques including economy of size (i.e., miniaturized arrays), high sensitivity, the use of non-toxic chemicals and the capacity for simultaneously analyzing thousands of genes with the minimal amount of sample. The microarray is proving to be a powerful research tool in gaining insight into the molecular biological study of genes in the tissues of interest. The main large-scale application of the microarray is comparative expression analysis, however, interest in the analysis of DNA variation is increasing. Other potential and current applications for this technology include diagnostics and genotyping, and pharmacogenetics.

This technology, however, is still in its infancy and a number of issues and drawbacks attenuate the power of the microarray. These issues include: (i) Conceptual, to design experiments that fully exploit the power of genome-wide expression analysis and maximize the meaningful information that can be extracted; (ii) Technical, to control for environmental (e.g., temperature) and biological (e.g., heterogeneity of population examined) factors that may compromise gene expression patterns; (iii) Statistical, to validate the data obtained and enable comparisons between arrays; (iv) Logistical, to effectively and efficiently deal with the vast amounts of data generated; (v) Confirmational, to perform essential follow-up studies for the authentication of the microarray data

Correspondence to: Dr. A. S. Wilson, Department of Molecular Ophthalmology, Lions Eye Institute, 2 Verdun Street, Nedlands, WA 6009, Australia; email: awilson@eye.uwa.edu.au

and finally the (vi) Cost of the microarray project itself. Still, these problems are likely to be resolved with experience, economies of scale, collaborations and cooperation between laboratories, free-market competition and time.

THE MICROARRAY TECHNOLOGY

The principle of the microarray experiment is the reverse of the Northern blot analysis, where mRNA extracted from the tissue of interest is used to generate a labeled sample, often referred to as the “target”. This target is hybridized onto a high density, ordered array of tens of thousands of transcript sequences (“probes”) immobilized onto an impermeable, rigid support, such as glass, thereby allowing the detection and quantification of a large number of gene transcripts simultaneously. Many different microarray systems have been developed both by academic institutions and commercial companies. However, two conceptually different systems have predominated

categorized primarily by the material arrayed: complementary DNA (cDNA) and high-density oligonucleotide arrays. When referring to oligonucleotide arrays information will largely pertain to the Affymetrix GeneChip® unless otherwise specified. Figure 1 demonstrates the fundamental differences between the two types of microarray platforms in regard to sample preparation, design and production of the arrays, and the data images generated which are discussed in detail below.

The cDNA Array: Schena and colleagues [10] pioneered the first cDNA array in 1995, which was generated by the robotic printing of double stranded cDNAs (up to 2 kb in length) as spots at defined locations on glass microscope slides. Nylon filters are also used as platforms for the cDNA microarray. The spots range from 100-300 µm in size and are set an equal distance apart. The arrayed cDNAs are usually amplified PCR products derived from cDNA clone sets, con-

TABLE 1. LIST OF AVAILABLE MICROARRAYS

Company	Name	Type of Array	Arrays available and Comments
Affymetrix, Inc.	GeneChip® Expression Arrays	High density oligonucleotide genome arrays (can contain about 12,000 genes per chip)	Human, rat, mouse, drosophila, Arabidopsis, yeast, E.coli. Rat toxicology/neurobiology arrays available. Customized array service available.
SuperArray, Inc.	GEArray and GEArray Q series; Pathway-specific Gene Expression Arrays	Nylon membrane cDNA arrays (GEArray has 20 cDNA fragments spotted in duplicate. GEArray Q series has 96 genes)	Cell cycle, apoptosis, cytokine, chemokine and receptor, cancer, PathwayFinder, tumor & metastasis, angiogenesis, adhesion and ECM. Both array types are available as kits containing 2, 4, and 12 membranes.
Clontech	Atlas Arrays	Nylon membrane cDNA macroarrays, glass slide arrays and plastic films (Plastic films and glass slides have long oligos)	Human: 3.8K & 8K, cancer, apoptosis, cardiovascular, cell cycle, cytokine/receptor, oncogene/tumor suppressor, hematology/immunology, neurobiology, stress, toxicology. Rat: 1K, stress, toxicology. Mouse: 1K, cancer, stress, toxicology
Agilent	Agilent Catalog Microarray kits, Agilent Custom Oligonucleotide Microarrays	cDNA microarrays and in situ oligonucleotide arrays (inkjet printed)	Human gene expression arrays, cDNA arrays spotted using pre-defined cDNA clone sets from Incyte Genomics. Oligonucleotide arrays feature either 25-mer or 60-mer in situ oligonucleotide synthesized probes on glass slides in either 8.4K or 22K format. Customized service available.
Sequenom	SpectroCHIP	Silicon chip containing high density array of mass spectrometry analysis sites	Designed for genotyping.
Incyte Genomics	GEM series of microarrays	cDNA microarray	Human, rat, arabidopsis, mouse. cDNA arrays using sequence-verified cDNA clones
Nanogen	NanoChip Microelectronic Array	Electronic chip for RNA and DNA analysis, is 99-site array	Ideal for SNP (single nucleotide polymorphism), STR (short tandem repeat) and point mutations research for clinical diagnostics
NEN Life Sciences Products, Inc.	MICROMAX	cDNA slide microarray	Cuts down on starting material, utilizes 1-4 µg total RNA

The list comprises some of the types of microarrays currently available. Information provided includes the companies that distribute these arrays, the type of array, number of elements per array, and other available services or kits.

taining sequence-verified genes and expressed sequence tags (ESTs) [11], or custom cDNA libraries [12] (see application notes from ABgene). Theoretically, greater than 30,000 genes can be spotted onto a standard glass microscope slide, although 10,000 to 20,000 spots are usual. Commercially available cDNA arrays are also available. Table 1 is a sample listing of the vendors of commercially available arrays, however, a comprehensive list of vendors (and industry links) of microarray instrumentation and accessories such as spotters, scanners, and analysis software is maintained by Leming Shi, PhD.

The spotted cDNA array enables the flexibility in tailoring arrays, particularly smaller customized arrays, for specific research needs. However, this technology can suffer from spotting errors and inconsistencies. The filtered cDNA arrays have the advantage of being relatively affordable, requiring no specific equipment for use. These types of arrays have also proven useful when examining tissues of limiting amounts of RNA. These features combined make the filtered cDNA arrays attractive for the researcher and are widely accessible to a greater number of laboratories.

The Oligonucleotide Array: The oligonucleotide array [13] is generated by the in situ synthesis of short 25mers by either photolithography onto a silicon chip (or GeneChip® developed by Affymetrix) or by ink-jet technology (developed by Rosetta Inpharmatics). An alternative is to print presynthesized oligonucleotides ranging from 60-100mers onto glass slides. The GeneChip® contains between approximately 7,000 and 20,000 functionally known genes and ESTs per array (Table 1 lists some of the available GeneChips®). Each gene or EST on the array is represented by between 11 and 20 different oligonucleotide sequences designed to be exactly complementary to different parts of the coding region for that gene. These are known as the perfect match (PM) probes. Each of these PM oligonucleotide probes is paired with a mismatch (MM) probe identical to the PM probe except for the alteration of the central base, a feature designed to enable correction for non-specific hybridization.

The oligonucleotide array offers the advantage of speed in generating an array as no time is expended in the handling of cDNA resources. However, the designing of the oligonucleotide set can be time consuming and is dependent upon public databases that are, at present, not complete for all genomes. The oligonucleotide probe is often designed to represent the unique sequence of the transcript reducing cross-hybridization and allowing the detection of splice variants and homologous genes. Use of multiple, short probes for measuring expression in a single gene in the GeneChip® improves the hybridization specificity and reduces the rate of false positives through averaging over the available probes and the GeneChip® requires much less starting material in comparison to the cDNA slide array.

Ideally, using both types of arrays could maximize the information gained. The GeneChip® could be used for gene hunting directing the researcher towards gene families and co-regulated genes, or signaling pathways that may be modified as a result of the stimulus or condition. These groups of genes could then be spotted onto a customized cDNA array allow-

ing the further interrogation and examination of the genes of interest. Although this would be idyllic, it is nonetheless appreciated that such endeavors are not so simple and the high cost of the commercially available microarrays, the sophisticated equipment and software and the availability and cost of clone libraries make such undertakings unapproachable for a majority of academic laboratories. However, shared microarray facilities and commercial services are surfacing making the technology more affordable and widely accessible.

THE MICROARRAY EXPERIMENT

Controls: One challenge facing the scientist performing microarray experiments is the designation of suitable controls. Controls are needed to ensure that significant experimental biases are not introduced at any step during the experiment. Spiked controls used in sample preparations and sophisticated internal controls on the microarray are adopted for quality control measures, ensuring that the samples and microarray hybridizations are optimal.

Good experimental design is essential and it is important to ensure that samples for comparison are as closely matched as possible, such that the only difference is the stimulus, drug or factor being examined. However, a number of issues arise including the definition of normal and overcoming the heterogeneity of the cell population [14]. Normal is not always easy to define. Variations can exist in both the patient (e.g., ethnicity, sex, age, genetic background, disease states) and the sample (e.g., anatomical location, developmental stage, degree of inflammation), which are likely to affect the gene expression profiles. For tissue samples and biopsies comprising a variety of different cells, the solution may be in laser capture microdissection (LCM) or by culturing a pure population of the cells of interest. However, LCM and in vitro culture are associated with a number of problems. These include the damage of RNA as a result of formalin fixing (frozen sections are better for RNA but are often difficult to prepare and histology can be compromised), the samples generated by LCM can be small and in vitro culture of cells may not reflect that which occurs in vivo producing different gene expression profiles. However, in situ hybridization following microarray analysis may reveal the specificity of the transcript to a particular cell type in the tissue examined.

For experiments where no appropriate control exists (e.g., across the board comparisons of tumors between patients), a reference sample is often employed for cDNA arrays. The reference is total RNA harvested from a variety of tissue sources and is designed to hybridize to as many genes on the microarray as possible. This enables the quantitative measurement of expression for any given gene as well as allowing comparisons of microarray data between experiments.

Sample Preparation, Labeling, and Hybridization: The targets for microarrays are usually labeled products prepared from mRNA samples, although genomic DNA may also be analyzed [15]. Frequently, total RNA pools are extracted and utilized to maximize the yields of mRNA that can be obtained from any given amount of tissue. The purity of the RNA is critical to ensure the greatest hybridization performance as

contaminating cellular proteins and DNA can result in significant levels of non-specific binding of fluorescent labeled probes.

Samples for hybridization to either microarray platform are generated by reverse transcribing the extracted and purified mRNA or total RNA using a modified oligonucleotide d(T) primer (such as the T7-d(T)₂₄ primer, see the Affymetrix technical manuals) to produce a double stranded cDNA (ds

cDNA) (Figure 1). The ds cDNA can then be amplified by in vitro transcription (IVT) using a T7 RNA polymerase promoter [15] (see also the RNA amplification protocol and the Affymetrix technical manuals) to produce an amplified RNA product or cRNA. The IVT reaction linearly amplifies the target making possible the profiling of transcripts from the minutest amount of sample. Current protocols for sample preparation for microarray hybridization utilize between 5-20 µg of

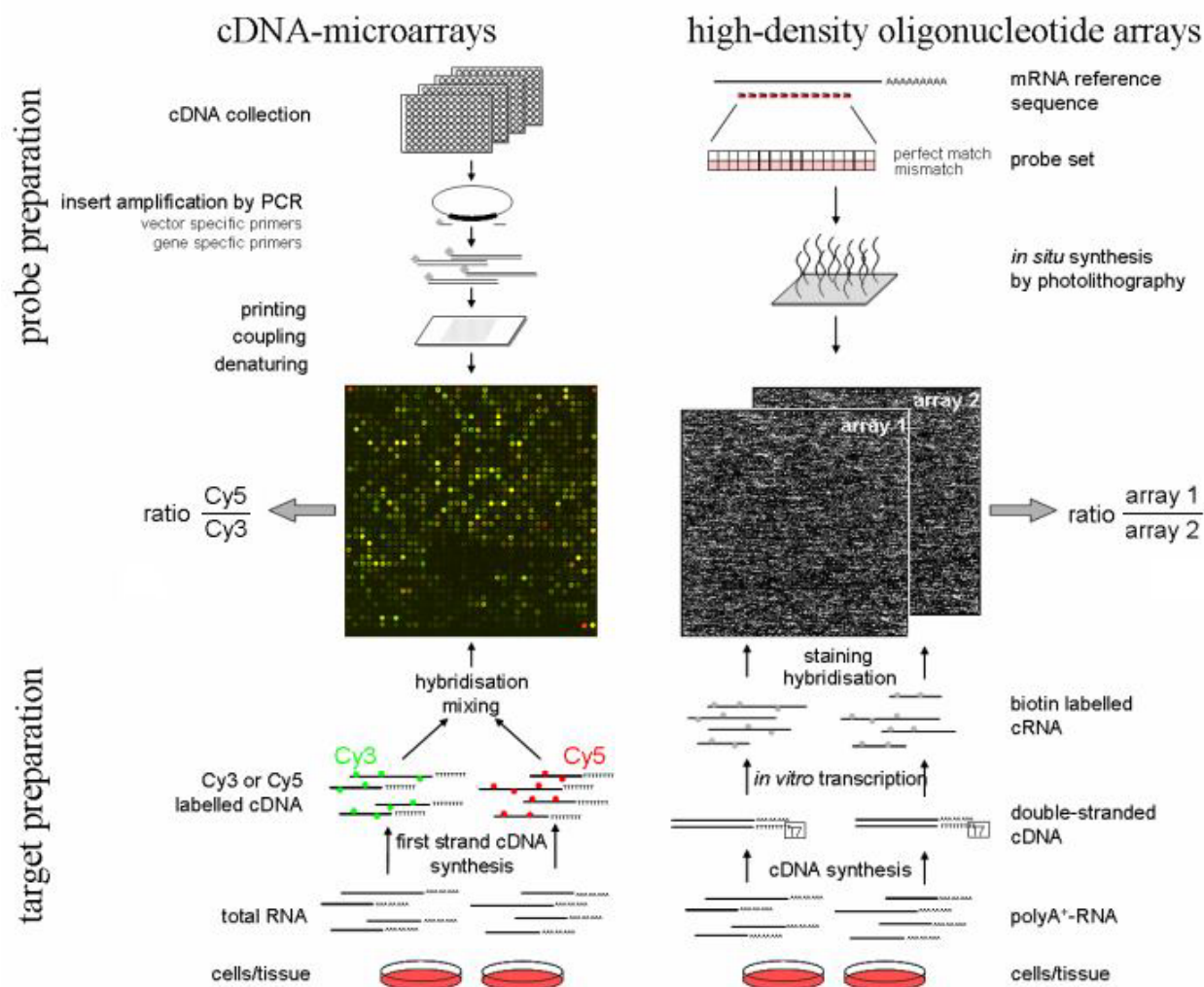


Figure 1. Schematic overview of spotted cDNA microarrays and high-density oligonucleotide arrays. **cDNA microarrays:** Array preparation: inserts from cDNA collections or libraries are amplified and the PCR products printed at specified sites on glass slides using high-precision arraying robots. These probes are attached by chemical linkers. Target preparation: RNA from 2 different tissues or cell populations is used to synthesize cDNA in the presence of nucleotides labeled with 2 different fluorescent dyes (eg: Cy3 and Cy5). Both samples are mixed in a small volume of hybridization buffer and hybridized to the array, resulting in competitive binding of differentially labeled cDNAs to the corresponding array elements. High resolution confocal fluorescence scanning of the array with two different wavelengths corresponding to the dyes used provides relative signal intensities and ratios of mRNA abundance for the genes represented on the array. **High-density oligonucleotide microarrays:** Array preparation: sequences of 16-20 short oligonucleotides (typically 25mer) are chosen from the mRNA reference sequence of each gene, often representing the unique part of the transcript. Light-directed, in situ oligonucleotide synthesis is used to generate high-density probe arrays containing over 300,000 individual elements. Target preparation: Total RNA from different tissues or cell populations is used to generate cDNA carrying a transcriptional start site for T7 DNA polymerase. During IVT, biotin-labeled nucleotides are incorporated into the synthesized cRNA molecules which is then fragmented. Each target sample is hybridized to a separate probe array and target binding is detected by staining with a fluorescent dye coupled to streptavidin. Signal intensities of probe array element sets on different arrays are used to calculate relative mRNA abundance for the genes represented on the array. Modified and reprinted with permission from Nature Cell Biology (Vol. 3, No. 8, pp. E190-E195) Copyright ©2001 Macmillan Publishers Limited.

total RNA depending upon the labeling method used. For tissues with limiting amounts of RNA (about 1 µg) or cells obtained from laser-capture microdissection, or even to the level of a single cell [14,16] a second round of target amplification by IVT or post-hybridization amplification may be adopted.

For cDNA arrays, fluorescent dyes can be incorporated directly into the cDNA during the reverse transcription of the RNA or amplified RNA in cases of limiting starting material. The fluorescent dyes most commonly used because of cost and availability are the cyanine dyes: Cy3 and Cy5; which have good photostability, and are widely separated in terms of their excitation and emission spectra. However, these dyes are large molecules that do not integrate easily into the reverse transcription reaction, and the incorporation of these dyes during sample preparation is not equal thereby making it necessary to perform reciprocal labeling of duplicate samples with swapped dyes. Some of the problems associated with the Cy-dyes are circumvented by improved dye technology now available. Examples include the Fairplay™ Microarray Labeling Kit (Stratagene, La Jolla, CA; the kit enzymatically modifies the cDNA prior to chemically linking the Cy-dye) and the Alexa Fluor Dyes (Molecular Probes, Eugene, OR), which are designed such that they do not interfere with enzymatic interactions or target binding sites once incorporated into the cDNA. Indirect labeling post reverse transcription techniques, such as the use of DNA dendrimer probes (Genisphere, Hatfield, PA) are also an option. These techniques have shown improved uniformity of dye incorporation, sensitivity and consistency [17].

For detection of gene expression using the Affymetrix GeneChip®, the sample is labeled during the IVT process which occurs without introducing significant bias. This results in a biotinylated cRNA product that is then fragmented to produce short targets. A secondary detection method is used where a streptavidin-conjugated fluorescent marker is bound to the biotinylated, fragmented cRNA, prior to hybridization to the GeneChip®. The secondary labeling strategy amplifies the signal enabling the detection of low abundance transcripts.

The cDNA slide array utilizes a system of competitive hybridization (Figure 1). Two samples, each labeled with distinct fluorescent dyes (for example Cy3 (colored green) and Cy5 (colored red)) are simultaneously hybridized to the same microarray. The normalized Cy3/Cy5 ratio of each gene on the array reflects the relative abundance of the relevant transcripts in relation to each other. For two samples representing experimental and control, where, for example, the experimental sample is labeled with Cy3 (green) dye and the control sample is labeled with Cy5 (red) dye, green spots on the microarray represent those genes upregulated compared to control, red represents those genes that are downregulated in the experimental sample, and yellow encodes those genes of equal abundance in both experimental and control samples. Unlike cDNA microarrays, the hybridization of sample to the Affymetrix GeneChip® is not competitive; control and experimental samples are hybridized to separate chips (Figure 1) and comparisons of the two are performed in subsequent analysis. The differential fluorescent intensities are represented as alterations

in transcriptional profiles between the compared samples.

Signal Detection and Data Extraction: Detection of the hybridized targets is achieved by fluorescence scanning following laser excitation of the fluorophores (cDNA slide and oligonucleotide arrays) or phosphor imaging (cDNA filter arrays). The cDNA slide and oligonucleotide arrays can produce well-defined, clearly delineated signals resulting in images amenable to data extraction by highly developed, digital image processing techniques.

The image processing steps in the analysis of a microarray are similar for both types of microarrays. The first step is gridding or addressing in which the exact locations of the individual spots or transcript signals are determined. The next step is to classify the microarray pixels as either foreground (corresponding to a spot/transcript signal) or background. Finally the foreground is extracted, contributions due to background intensity are subtracted, and calculations of spot or transcript intensity are made. Oligonucleotide microarrays require an additional calculation prior to obtaining transcript intensity values. In this final step the transcript intensity values represented by the PM sequences are corrected for non-specific hybridization as measured by the MM sequences.

In practice each of the steps of addressing, classification and signal extraction can be highly problematic. Artifacts on the arrays can complicate the addressing and classification stages, while irregularly shaped spots can cause problems in the foreground extraction process. A variety of software tools utilizing different mathematical algorithms to perform microarray image analysis and circumvent some of these problems are available [18], both commercially and from public sources. A comprehensive list is maintained by Yuk Fai Leung.

Data extractions from film or phosphor images of radioactive hybridizations present a number of additional problems for image analysis. Nylon filters frequently undergo non-linear warping of the matrix resulting in a lack of geometric regularity of the array. This makes it difficult to overlay grids to accurately specify target locations. The disintegrating nuclide spreads detectable particles distorting the signal morphology making difficult the application of automatic detection and signal extraction procedures. Due to the radiation spread, local background for individual signals is difficult to determine, as the transition from signal to background is not readily discernible [19]. Other potential problems associated with the nylon membrane include the regular stripping and re-use of the membrane, which may result in the removal of some of the original probe present on the membrane making comparisons between samples using the same nylon filter unreliable and undesirable.

A vast amount of data is generated from the microarray experiment. Data management becomes a critical issue important particularly when performing further downstream data analyses. Databases for the storing and retrieval of gene expression data from any given experiment and array type, and for any organism or artificial source serve as important resources particularly for the dissemination of the gene expression data for public forum and exchange of accurate information, and for subsequent independent verification of the data

at a later stage if required. A number of gene expression databases have been generated and are accessible to the public, including the Gene Expression Omnibus (provided by the National Center for Biotechnology Information), which is a repository for gene expression and hybridization array data. Another database to consider is MGED provided by the Microarray Gene Expression Data Group, which also provides standards for the representations of microarray expression data (MIAME and MAGE), information and discussion on ontology and normalization methods.

DATA ANALYSIS OF MICROARRAYS

The list of microarray software includes references to software which can be used to perform analysis of the intensity data derived from the image analysis for a given microarray experiment consisting of one or more arrays. The general microarray experiment compares experimental samples against one another or against a common reference sample, where the aim is to discriminate between the various experimental samples based on their differential expression. Examples of experiments include measuring gene expression in a number of different tissues, or in the same tissue at several different points in time. For both cDNA and oligonucleotide arrays a necessary first step in the analysis of the microarray intensity data, is within-array and between-array normalization. This process aims to remove the effects of factors other than gene expression from the data. These factors can arise from various sources, including different dye hybridization efficiencies in the case of cDNA microarrays, and different scanner settings in the case of both types of arrays. Such noise factors can mask transiently expressed genes and reduce comparability of the array data [20]. A simple and common normalization method for microarrays is global normalization where it is assumed that the array to array bias, and/or the green to red bias in cDNA arrays, is a constant factor across the whole

array. Under such an assumption the data are normalized by subtracting a constant on the log scale to get normalized values. However, the bias is often seen to be intensity dependent, or dependent on the location of the spots on the arrays. As such normalization methods which apply an intensity and/or location dependent correction are preferable [21,22]. After normalization, differences between the intensity of a pair of arrays in the case of the oligonucleotide arrays, or between the red and green intensities in the case of cDNA arrays, can be related directly to differential gene expression.

In a simple pairwise comparison of gene expression between two samples, an experimental (T) and control (C), the results can be shown in a plot of the log intensity ratio $M = \log_2(T/C)$ versus mean log intensity $A = (1/2)\log_2(TC)$, for each gene on the array [22]. Figure 2 shows an example of such a plot. An M vs A plot is equivalent to a 45° rotation of a conventional scatter plot, along with appropriate scaling of the co-ordinates. Scatter plots are dominated by the high degree of correlation between the two samples. For the purpose of detecting differential expression, the true interest in a scatter plot is not this correlation but the points that deviate from the 45° line. An M vs A plot increases the visibility of these points, thereby making it easier to observe the differential expression between the samples. The log scale is used due to the large range of intensity values, spanning several orders of magnitude. In general the expression levels of most of the genes in the two samples do not change, thereby resulting in the majority of the points being scattered symmetrically around the horizontal line $M = 0$. Differentially expressed genes will appear away from this line, with those having the greatest intensities appearing to the right of the plot.

Given such a plot the next task is to determine the boundary between significantly differentially expressed and unchanged genes, given the natural and experimental variability of the microarray experiment. Major sources of variability of the data include normal physiological gene expression variations [23] (where normal variance for many tightly regulated tissue-specific genes can be within 20-30%) and noise that may be introduced by slide heterogeneities, printing irregularities, spotting fluctuations and poor sample hybridization [24,25]. Many studies apply an arbitrary fold-change threshold to define significant differential expression, usually 2-fold. This is essentially a ranking method, which while intuitive, suffers from the limitation of not detecting truly differentially expressed genes, which have changed by less than the arbitrary cut-off. It is preferable that genes are determined to be differentially expressed on the basis of a statistical test, with an associated confidence level. For a sufficient number of replicates for both treatment and control samples there are methods for estimating a measure of significance of the observed differences between samples. Examples include estimating the false discovery rate [26], and calculating adjusted p-values [27]. These methods and others like them simulate the underlying distribution, against which it is then possible to test the measured gene expression differences and assign a significance value to them.

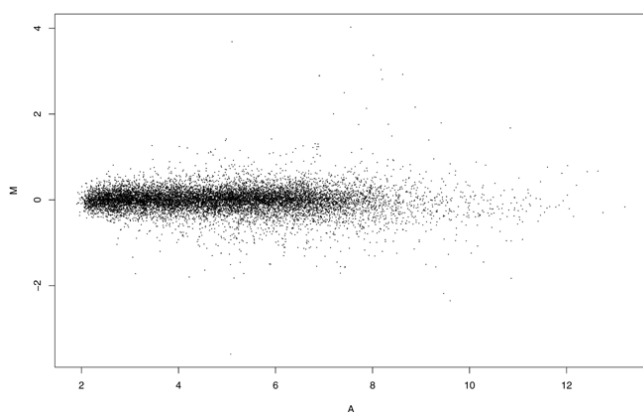


Figure 2. M versus A scatter plot. M vs A for microarray intensity data from treatment (T) and control (C) samples, where M is the log intensity ratio $M = \log_2(T/C)$, and A is the mean log intensity. In this case average $\log_2(PM)$ for each probe set was used as the measure of expression for each gene. Other summary measures of expression would show similar behavior.

For experiments more complicated than the two sample experiment described here, more sophisticated analysis techniques are needed. Such experiments may involve the inter-comparison of data from many microarrays, for example, experiments comparing gene expression in several different tissues or cell types, or time-course experiments where gene expression in the same sample is measured at several different time points. The aim here is to examine the patterns of co-regulated gene expression, which characterize the different tissue types or time points in the experiment.

A common approach in these types of experiments is hierarchical clustering [28-30] although other techniques such as K-means clustering and self-organizing maps have also been used [31,32]. Clustering methods use mathematical operations and similarity measures akin to correlation, to group together genes with similar patterns of expression. Such groups may indicate co-regulated genes, genes encoding proteins that interact, genes with related functions or gene families, and signatures of individual signaling pathways within the data sets providing information on the networks operating within a given cell type or tissue. In an experiment involving different diseased tissue samples, cluster analysis can be utilized to characterize the different tissues by their gene expression patterns, as in the case of classifying tumors [33], thereby providing a molecular basis for diagnosis in patients. When investigations involve ESTs (expressed sequence tags of unknown function), methods utilizing sequence homology to detect functional linkages among predicted proteins [34] and comparing these to proteins of known function [35] are used. Databases used to extract sequence information include BLAST, GenBank and dbEST provided in the public domain by the National Center for Biotechnology Information (NCBI). Cluster analysis can be used in this context to aid in the assignment of a function or potential role in a given pathway to unknown genes based on the set of known genes that fall into known cluster sets.

Another method of analyzing gene expression data across multiple arrays is to derive a linear model for the expression of a given gene, with the elements of the model corresponding to the data for the gene across all the arrays, including various error components. The model can then be fitted to the data using least squares estimation. The coefficients of the fitted model for a given gene are the estimates of the gene expression for the conditions of each given microarray [36].

Due to the uncertainties inherent in the microarray experiment, replication is always advisable, as well as verification by alternative techniques, preferably quantitative, such as Northern hybridizations, RNase protection assays and real-time RT-PCR. Such confirmational studies, while costly in terms of sample and finance, can help to establish any variability within the experimental system studied and validate the change in gene expression between control and experimental samples.

APPLICATIONS OF THE MICROARRAY

Experimental Approaches: elucidating gene function: The microarray provides a variety of ways for investigating scientific questions arising in cell biology with the different ex-

perimental approaches falling between two extremes; local and global. The local approach aims to identify the single gene responsible for the mediation of the cascade of genes expressed that eventually results in the establishment of a particular phenotype (e.g., transformation and gene knockouts [37] or protection from apoptosis). This approach has been associated with little success, as a single phenotype could potentially result from the altered expression of hundreds of genes. The microarray offers few clues in distinguishing which of these genes is important in establishing a given phenotype let alone identifying the single gene from which the deluge of gene events originate. The microarray, however, has successfully been adapted for the large-scale identification and genotyping of genetic mutations such as single nucleotide polymorphisms (SNPs) in the human [38] and rice [39] genomes.

Results to date have shown that the power of the microarray in analyzing cell function is in the provision of a global presentation of gene expression, rather than as a means to identify a single critical gene. For the global approach, expression profiles of samples, of varying phenotypes, impacted by a variety of conditions, stimuli or experimental manipulations can be compared. Examples include the expression profiling of yeast [40] and mammalian cells [41] during the cell cycle and the response of fibroblasts to serum [42] or PDGF [43] stimulation.

Disease Diagnosis and Therapy: The challenge facing researchers today is the development of new approaches for the identification of disease causing agents, the understanding at a fundamental level of the pathogenic processes and, from this information, creating an efficient means of treatment.

In the past, drug discovery involved the collaboration of biochemistry and medical chemistry in the identification and optimization of the therapeutic effects of small molecules in targeting specific enzymes implicated in a pathophysiological process. After intensive laboratory investigation and animal trials, clinical trials ensued to identify those molecules with therapeutic efficacy in humans before being marketed. This process is long, tedious and very costly. Pharmacologists, toxicologists and members of the pharmaceutical industry now utilize the microarray as a means of monitoring modified gene expression in cells or tissues in response to medication (see application notes from ABgene), to identify and validate the targets and to identify any toxic effects as a result of drug metabolism [44,45]. The benefits of such studies speak for themselves. Importantly, the use of the microarray minimizes the time and cost of identifying potential drugs for therapy too. Possibly in the near future, the microarray could facilitate the identification of those individuals that display adverse reactions to specific drug treatments as well.

To date, the microarray has largely been adopted in cancer research and diagnosis. Before the microarray, cancer classification techniques relied heavily on subjective judgments made based on tumor histology. It has also been revealed that tumors presenting with similar histology may not necessarily be of the same molecular classification. Therefore, the ability to distinguish morphologically similar tumors in order to ac-

curately identify the tumor and hence provide the appropriate treatment is important. Recently the use of the microarray in accurately diagnosing and classifying tumors [33,46], enabling the prediction of outcomes and facilitating the recommendation of the best-matched treatment has been demonstrated. The potential value of the microarray in understanding the mechanics of cancer at a molecular level are great. The information gained would lead to clinical benefits such as maximizing therapeutic efficacy and reducing any toxic effects as well as possibly inspiring new tactics in treating cancer.

For infectious diseases, major contributors to population mortality and increased health costs, the progress in understanding the pathogenic processes and the discovery and development of new therapies has been limited. The microarray would prove to be a valuable tool in disease research with diagnostic and prognostic indications. The feasibility of this approach for disease research has been suggested in expression profile studies of human macrophages infected with a variety of viral and bacterial pathogens [47,48], allowing pathogen classification.

Gene expression profiling has also shown promise in the unveiling of the molecular biological consequences of cytokine induced b-cell damage and repair in type 1 diabetes, an autoimmune disease [49]. The anticipated effect of these studies is the invaluable insight into the pathogenesis and the potential for initiating new approaches in combating these diseases.

Current Applications in Ophthalmic Research: Research into the underlying molecular mechanisms that occur as a result of ocular diseases have been gaining some ground, however the causes of such diseases and their pathogenesis remain relatively unknown. Many grouped ocular diseases, such as inherited retinal degenerations, including retinitis pigmentosa (RP), are believed to be linked by a common mechanism that results in photoreceptor cell death. Studies using animal and transgenic models have tried to determine this common link between the diseases but progress is slow. Microarray technology has gradually permeated its way into ophthalmic research, the benefits of which are great. The microarray shows potential as a useful tool in the elucidation of the genes involved in ocular disease progression and may therefore aid in the characterization and classification of ocular diseases and possibly identify a common mechanism linking grouped diseases.

The microarray has been useful in the determination of the cause of photoreceptor cell death. Different approaches to tackling this question have resulted in a number of possibilities. Antagonists of the N-methyl-D-aspartic acid (NMDA) receptor serve in the treatment of elevated intraocular pressure associated with glaucoma and are found to be neuroprotective in the ischemic retina, yet the signaling transduction pathways mediated by the NMDA receptor remain unclear. Nylon membrane arrays were used to examine NMDA-induced apoptosis in the retinas of 18 Sprague Dawley rats over 24 h. The data from quadruplicate replicates revealed that the neurotoxin NMDA caused an upregulation of several apoptosis-associated genes in the retina [50]. These were the receptors TNF-R1 and Nur77, FasL (a member of the TNF

family) and GADD 45 and GADD153, of unknown function, which were verified by semi-quantitative RT-PCR techniques. It is probable that these genes may be representative of an intracellular signaling cascade resulting in photoreceptor cell death. Furthermore, antagonists of NMDA were shown to inhibit the expression of these genes. However, whether these genes translate to proteins involved in apoptosis remains to be seen. The array was also spotted with only 588 known genes therefore it is impossible to determine whether other genes, not present on the array, may be involved.

Jones and colleagues [51] focused their study on identifying genes associated with inherited retinal degeneration by comparing gene expression profiles of retinas dissected from postmortem human eyes from donors with retinitis pigmentosa (RP; n=4) and those with no history of ocular disease (n=5). Using cDNA filter arrays spotted with 205 apoptosis-related partial cDNAs, increased expression of SFRP2 (secreted Frizzled-related protein-2) was observed in RP affected human retinal samples. This protein has been linked to the Wnt signaling pathway responsible for cell fate. The findings suggest that altered Wnt signal transduction may be a vital step in the degenerative process leading to eventual photoreceptor cell death. Due to the limited amount of RNA in this study, the microarray hybridization was not performed in replicate although confirmed by Northern blot, in situ hybridization and immunocytochemical analyses.

Gene expression profiles of current retinal degenerative disease animal models have been generated with the aim of identifying the early alterations in gene expression that may contribute to photoreceptor cell death. In a time course examination of light-induced apoptosis in mice deficient in arrestin and rho kinase [52], hierarchical clustering and self-organizing maps of the microarray data revealed a decrease in a cluster of photoreceptor-specific genes well before morphological damage could be observed. Differential changes in gene expression were noted over time. Despite the increase in stress proteins, damage-inducible genes and genes involved in DNA repair, little to no change was observed in the expression of those genes involved in the initiation of apoptosis, which may be contributed to the short time period studied. In the case of the rd/rd model for inherited retinal degeneration, microarray analysis using cDNA arrays containing 588 elements on post-natal mice identified increased expression in nm23-M2, a factor involved in metastasis suppression and transcriptional regulation, during retinal degeneration [53]. It is believed that this increase in nm23-M2 may be stress-related and a response to photoreceptor loss and altered retinal structure, providing a possible marker for retinitis pigmentosa.

Another approach to tackling the investigation of photoreceptor degeneration has been to understand the regulation of the processes involved in the terminal maturation of photoreceptor cells. A study utilizing the microarray identified the photoreceptor homeobox gene Crx as the modulator of a network of genes, which may be responsible for the differentiation of the photoreceptor cell [54]. Mutations in any of these genes (identified as Crx targets) resulted in photoreceptor degeneration. For this study, a mouse retinal cDNA array was

constructed from 960 random clones from an adult mouse retina cDNA library where the individual inserts were PCR amplified and sequenced prior to spotting with a piezo-electric array printing system onto glass slides. Verification of the data was performed using Northern and in situ hybridization of the 17 genes of interest. This project is an example of a very large undertaking that may have been quite costly, however the payoffs proved worthwhile. By identifying mutations in the mediators of this particular pathway, it then becomes possible to identify potential therapeutic targets and the approach to therapy becomes focused; appropriately matched for the specific disease.

The above study is an example of the need to produce an array due to the serious under-representation of eye-specific gene arrays available commercially. The need for eye specific gene arrays is further exemplified by the study on gene expression in the retina of developing mice [55], which found genes specific for eye development and only expressed in the retina. This study also demonstrated the lack of overlap in the genes and ESTs present in the library generated to currently available mouse cDNA clone sets such as the NIA 15 K cDNA clone set emphasizing the need for retinal specific databases and retinal specific libraries and arrays. Recently, Swaroop and colleagues [56] have managed to isolate 7500 clones from mouse eye/retina cDNA libraries (of which >3,100 have been sequenced) to produce slide arrays (I-gene arrays) containing over 2000 clones. Currently, the production of slides with >6000 clones are underway. These I-gene microarrays will be useful in the generation of expression profiles of mouse eye genes that will serve to elucidate molecular pathways involved in development, aging and disease.

The global approach to gene expression profiling achieved with microarrays allows the characterization of a pathological process that can be applied to any disease. Such an approach would be valuable for polygenic diseases such as diabetic retinopathy. In the streptozotocin (STZ)-induced diabetic rat model, alteration in retinal gene expression was examined over 3 weeks using high-density cDNA filter arrays (containing 5147 rat genes) [57]. The data revealed increases in the number of upregulated genes over time followed by a decrease in the number by 21 days post-induction. These gene expression changes are likely to be diabetes-related as no such alterations were evident in non-diabetic controls and STZ non-converters. These genes included insulin/glucose metabolism related genes, signal transducers, proliferation and differentiation related genes, apoptotic genes, inflammation related genes and secreted growth factors and regulators. The information gained from such investigations would piece together the detailed knowledge of the pathophysiology of diabetic retinopathy, which is necessary for the development of a rational therapy.

The microarray has also been used to identify potential markers for disease. Glaucoma is commonly associated with elevated intraocular pressure. From organ perfusion studies of anterior segments obtained from human postmortem eyes (2 pairs), microarray revealed the induction of several genes in the trabecular meshwork in response to increased pressure [58]. The products of these genes (e.g., interleukin 6, cathep-

sin-L, metallothionein, preprotachykinin-1 to name a few) are involved in blood vessel permeability homeostatic mechanisms such as vascular permeability, secretion, extracellular matrix remodeling and reactive oxygen species scavenging. These could serve as predictors of the development of glaucoma, based on their known activities.

Many ocular disorders have no appropriate animal model leaving investigators to rely on surgical specimens that often contain very limited amounts of RNA. Amplification of samples derived from autopsy corneas of normal individuals and those with pseudophakic bullous keratopathy (PBK) [59] for microarray has enabled the expression profiling of this corneal disease. Results from commercial filter array have implicated the altered expression of certain adhesion molecules, namely β_6 -integrin and β -catenin, in the formation of epithelial bullae and the microcystic changes in PBK.

In vitro studies involving the culture of corneal epithelial cells is one means of circumventing the problem of limiting RNA for gene analysis. The expression profiles of human corneal epithelial cells HCEC grown in cultures supplemented with or without TGF β 1 have been examined [60] to study its role in the eye. Using cDNA filter arrays, TGF β 1 was found to modify the expression of numerous genes. Those genes shown to be downregulated the most (α_3 -integrin, transferrin receptor, cyclin D1 and a serine protease inhibitor) displayed a variety of functions including cell adhesion, regulation of cell cycle and proliferation, demonstrating the multifunctional role of TGF β in the maintenance of HCECs.

Cell culture is often adopted for the investigation of a number of age-related disorders, such as cataract and age-related macular degeneration (AMD), which are often associated with dysfunctional RPE layers. In vitro studies combined with cDNA filter microarrays have been employed to examine senescence [61], an arrested state in which cells remain viable, and the effects of advanced glycation end products (AGEs) on the aging of RPE [62]. Clusters of genes involved in cell maintenance of basement membrane and the deposition and maintenance of extracellular matrix were found to be affected in cell senescence and modified by AGEs, demonstrating that adverse effects to cell architecture contribute to RPE dysfunction. However, AGEs, found in human Bruch's membrane, basal deposits and drusen, was found to alter the expression of those genes involved in the regulation of cell differentiation and apoptosis. This study identified genes not previously known to be regulated by AGEs, expressed by RPE or implicated in the aging process. Further investigation of the gene sets is underway to fully comprehend their role in the development of AMD and cataract.

Future Potential: The combined information obtained from the likes of ocular research studies noted above, would lead to the delineation of common pathways that link and group the different types of inherited retinal degenerations, age-related diseases and other ocular disorders. Much can be learned from the microarray experiments performed in cancer studies. By examining the clustering of groups of genes over a series of specimen samples, it may be possible to elucidate the gene expression patterns associated with particular ocular

diseases. The establishment of the commonly featured pathways of these diseases may lead to the development of new therapeutic targets and/or therapies for treatment of the grouped ocular diseases and disorders.

Another method may be to analyze genomic DNA. Array-format comparative genomic hybridizations [15], which measure gene copy number in DNA samples (disease compared to normal), may be adopted to define genetic mutations responsible for disease, including chromosomal insertions and possibly chromosomal deletions.

As the eye comprises several cell types, expression profiling of individual tissues and cell types (achieved by combining laser capture microdissection techniques with microarray) in normal and diseased states is needed. Such examination may allow one to define the initial source of the disease such as dysfunctions in the photoreceptor, RPE cell layer, vascular tissue of retina or cornea. Alternatively, downstream localization studies (such as *in situ* hybridization) post-microarray of the genes of interest may also provide information regarding tissue specificity of the genes. The generation of comprehensive databases of generic gene expression profiles would enable the signature expression profiles for specific cell types, which can then be extracted from whole tissue profiles. Currently, a number of signature expression profiles have been generated including that for native human retinal pigment epithelium [63] and human cornea [64] both derived from human donor tissues. The libraries generated allow comprehensive examination of tissue specific genes enabling the identification of genes responsible for retinal and/or macular disorders. The clinical significance of the databases generated is the increase in understanding of normal and abnormal tissue function in relation to treating the disease and possible application to transplantation. Other gene expression signature profiles of importance include those for specific processes such as synaptic plasticity in the visual cortex [65] or those generated to understand the effects of different cytokines on the differentiation process of corneal epithelial cells *in vitro* [66] or the downstream events of Pax6 activity in the regulation of lens development in mice [67].

Additionally, it is necessary to combine the gene expression data of microarray with protein expression profiles and histological information of normal and pathologically affected tissues. Much of the mRNA expressed in cells does not translate to form protein products and whether these proteins are involved in the processes examined need to be confirmed. Other complicating factors including post-translational modifications and protein isoforms may alter their function *in vivo*. Therefore, it is important to investigate the protein profiles and the respective functions of the expressed genes. The histological link provides the phenotypic information of the effects of certain gene expression patterns. This "genetic fingerprinting" of the histological progress of disease allows the genome-wide screening and monitoring of disease and genetic mutations.

CONCLUSION

The global approach to gene expression profiling enables the elucidation of the underlying pathophysiological process that occur in disease, allowing disease characterization and classification and the possible development of rational therapies. One must be mindful, though, that the microarray experiment is currently expensive involving considerable attention, thorough experimental planning, good statistical analysis and the employment of a number of follow-up confirmational techniques. Experimental designs that maximally exploit the global gene expression profiles need to be employed and experimental manipulations need to be rigorously controlled, as small differences in the microenvironment potentially amount to large changes in the expression profile.

Although the computational tools available for the analysis of microarray data is underdeveloped, the microarray still remains an advanced technology that enables the high throughput examination of the genome of interest and its regulatory networks. It proves to be a valuable tool in gaining insight into cellular processes and, therapeutically, identifying genes that may serve as molecular markers of disease or as potential targets for gene therapy.

ACKNOWLEDGEMENTS

The authors acknowledge with thanks the financial support provided by the National Health and Medical Research Council (Australia), the Juvenile Diabetes Research Foundation (USA) and WestPac Australia. This work is part of the research effort of the Diabetic Retinopathy Consortium, Perth.

REFERENCES

1. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 1977; 74:5350-4.
2. Lennon GG, Lehrach H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 1991; 7:314-7.
3. Kafatos FC, Jones CW, Efstratiadis A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* 1979; 7:1541-52.
4. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992; 257:967-71.
5. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 1995; 270:484-7.
6. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996; 14:1675-80.
7. 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 U S A* 1996; 93:10614-9.
8. DeRisi JL, Iyer VR, Brown PO. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 1997; 278:680-6.
9. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 1997; 15:1359-67.

10. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with complementary DNA microarray. *Science* 1995; 270:467-70.
11. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252:1651-6.
12. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nat Genet* 1999; 21:10-4.
13. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999; 21:20-4.
14. King HC, Sinha AA. Gene expression profile analysis by DNA microarrays: promise and pitfalls. *JAMA* 2001; 286:2280-8.
15. 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. *Nat Genet* 1999; 23:41-6.
16. Brady G. Expression profiling of single mammalian cells—small is beautiful. *Yeast* 2000; 17:211-7.
17. Yu J, Othman MI, Farjo R, Zarepari S, MacNee SP, Yoshida S, Swaroop A. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. *Mol Vis* 2002; 8:130-7.
18. Yang YH, Buckley MJ, Dudoit S, Speed TP. Comparison of methods for image analysis on cDNA microarray. *Journal of Computational and Graphical Statistics* 2002; 11:108-36.
19. Brandle N, Bischof H, Lapp H. A generic and robust approach for the analysis of spot array images. In: Bittner ML, Chen Y, Dorsel AN, Dougherty ER, editors. *Microarrays: optical technologies and informatics*. Vol 4266. *Proceedings of SPIE*; 2001 Jan 21-22, San Jose, CA. Bellingham (WA): SPIE; 2001. p. 1-12.
20. Hill AA, Brown EL, Whitley MZ, Tucker-Kellogg G, Hunter CP, Slonim DK. Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. *Genome Biol* 2001; 2:research0055.1-0055.13.
21. Finkelstein DB, Gollub J, Ewing R, Sterky F, Somerville S, Cherry JM. Iterative linear regression by sector. In: Lin SM, Johnson KF, editors. *Methods of microarray data analysis*. Boston: Kluwer Academic; 2002. p. 57-68.
22. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002; 30:e15.
23. Pritchard CC, Hsu Li, Delrow J, Nelxon PS. Project normal: defining normal variance in mouse gene expression. *Proc Natl Acad Sci U S A* 2001; 98:13266-71.
24. Watson SJ, Meng F, Thompson RC, Akil H. The “chip” as a specific genetic tool. *Biol Psychiatry* 2000; 48:1147-56.
25. Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H. Normalization strategies for cDNA microarrays. *Nucleic Acids Res* 2000; 28:E47.
26. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001; 98:5116-21.
27. Dudoit S, Yang YH, Callow MJ, Speed TP. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* 2002; 12:111-40.
28. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998; 95:14863-8.
29. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A* 1999; 96:6745-50.
30. Stuart RO, Bush KT, Nigam SK. Changes in global gene expression patterns during development and maturation of the rat kidney. *Proc Natl Acad Sci U S A* 2001; 98:5649-54.
31. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999; 22:281-5.
32. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* 1999; 96:2907-12.
33. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; 403:503-11.
34. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. Protein function in the post-genomic era. *Nature* 2000; 405:823-6.
35. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Cherry JM, Botstein D. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 1998; 282:2022-8.
36. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. Assessing gene significance from cDNA microarray expression data via mixed models. *J Comput Biol* 2001; 8:625-37.
37. Eggleston P, Zhao Y. Gene targeting in mosquito cells: a demonstration of ‘knockout’ technology in extrachromosomal gene arrays. *BMC Genet* 2001; 2:11.
38. Wang DG, Fan JB, Siao CJ, Bero A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lander ES, et al. Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 1998; 280:1077-82.
39. Jaccoud D, Peng K, Feinstein D, Kilian A. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res* 2001; 29:E25.
40. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998; 9:3273-97.
41. Cho RJ, Huang M, Campbell MJ, Dong H, Steinmetz L, Sapinosa L, Hampton G, Elledge SJ, Davis RW, Lockhart DJ. Transcriptional regulation and function during the human cell cycle. *Nat Genet* 2001; 27:48-54.
42. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J Jr, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; 283:83-7.
43. Fambrough D, McClure K, Kazlauskas A, Lander ES. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* 1999; 97:727-41.

44. Roses AD. Pharmacogenetics and the practice of medicine. *Nature* 2000; 405:857-65.
45. Gerhold D, Lu M, Xu J, Austin C, Caskey CT, Rushmore T. Monitoring expression of genes involved in drug metabolism and toxicology using DNA microarrays. *Physiol Genomics* 2001; 5:161-70.
46. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001; 98:10869-74.
47. Manger ID, Relman DA. How the host 'sees' pathogens: global gene expression responses to infection. *Curr Opin Immunol* 2000; 12:215-8.
48. Young RA. Biomedical discovery with DNA arrays. *Cell* 2000; 102:9-15.
49. Cardozo AK, Kruhoffer M, Leeman R, Orntoft T, Eizirik DL. Identification of novel cytokine-induced genes in pancreatic beta-cells by high-density oligonucleotide arrays. *Diabetes* 2001; 50:909-20.
50. Laabich A, Li G, Cooper NG. Characterization of apoptosis-genes associated with NMDA mediated cell death in the adult rat retina. *Brain Res Mol Brain Res* 2001; 91:34-42.
51. Jones SE, Jomary C, Grist J, Stewart HJ, Neal MJ. Altered expression of secreted frizzled-related protein-2 in retinitis pigmentosa retinas. *Invest Ophthalmol Vis Sci* 2000; 41:1297-301.
52. Choi S, Hao W, Chen CK, Simon MI. Gene expression profiles of light-induced apoptosis in arrestin/rhodopsin kinase-deficient mouse retinas. *Proc Natl Acad Sci U S A* 2001; 98:13096-101.
53. Jones SE, Jomary C, Grist J, Stewart HJ, Neal MJ. Identification by array screening of altered nm23-M2/PuF mRNA expression in mouse retinal degeneration. *Mol Cell Biol Res Commun* 2000; 4:20-5.
54. Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL. Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene *Crx*. *Curr Biol* 2000; 10:301-10.
55. Mu X, Zhao S, Pershad R, Hsieh TF, Scarpa A, Wang SW, White RA, Beremand PD, Thomas TL, Gan L, Klein WH. Gene expression in the developing mouse retina by EST sequencing and microarray analysis. *Nucleic Acids Res* 2001; 29:4983-93.
56. Farjo R, Yu J, Othman MI, Yoshida S, Sheth S, Glaser T, Baehr W, Swaroop A. Mouse eye gene microarrays for investigating ocular development and disease. *Vision Res* 2002; 42:463-70.
57. Joussen AM, Huang S, Poulaki V, Camphausen K, Beecken WD, Kirchhof B, Adamis AP. In vivo retinal gene expression in early diabetes. *Invest Ophthalmol Vis Sci* 2001; 42:3047-57.
58. Gonzalez P, Epstein DL, Borras T. Genes upregulated in the human trabecular meshwork in response to elevated intraocular pressure. *Invest Ophthalmol Vis Sci* 2000; 41:352-61.
59. Spirin KS, Ljubimov AV, Castellon R, Wiedoeft O, Marano M, Sheppard D, Kenney MC, Brown DJ. Analysis of gene expression in human bullous keratopathy corneas containing limiting amounts of RNA. *Invest Ophthalmol Vis Sci* 1999; 40:3108-15.
60. Hayashida-Hibino S, Watanabe H, Nishida K, Tsujikawa M, Tanaka T, Hori Y, Saishin Y, Tano Y. The effect of TGF-beta1 on differential gene expression profiles in human corneal epithelium studied by cDNA expression array. *Invest Ophthalmol Vis Sci* 2001; 42:1691-7.
61. Shelton DN, Chang E, Whittier PS, Choi D, Funk WD. Microarray analysis of replicative senescence. *Curr Biol* 1999; 9:939-45.
62. Honda S, Farboud B, Hjelmeland LM, Handa JT. Induction of an aging mRNA retinal pigment epithelial cell phenotype by matrix-containing advanced glycation end products in vitro. *Invest Ophthalmol Vis Sci* 2001; 42:2419-25.
63. Buraczynska M, Mears AJ, Zarepari S, Farjo R, Filippova E, Yuan Y, MacNee SP, Hughes B, Swaroop A. Gene expression profile of native human retinal pigment epithelium. *Invest Ophthalmol Vis Sci* 2002; 43:603-7.
64. Jun AS, Liu SH, Koo EH, Do DV, Stark WJ, Gottsch JD. Microarray analysis of gene expression in human donor corneas. *Arch Ophthalmol* 2001; 119:1629-34.
65. Prasad SS, Kojic LZ, Li P, Mitchell DE, Hachisuka A, Sawada J, Gu Q, Cynader MS. Gene expression patterns during enhanced periods of visual cortex plasticity. *Neuroscience* 2002; 111:35-45.
66. Lee JS, Liu JJ, Hong JW, Wilson SE. Differential expression analysis by gene array of cell cycle modulators in human corneal epithelial cells stimulated with epidermal growth factor (EGF), hepatocyte growth factor (HGF), or keratinocyte growth factor (KGF). *Curr Eye Res* 2001; 23:69-76.
67. Chauhan BK, Reed NA, Zhang W, Duncan MK, Kilimann MW, Cvekl A. Identification of genes downstream of Pax6 in the mouse lens using cDNA microarrays. *J Biol Chem* 2002; 277:11539-48.