

Seeing the Unseen: Microarray-Based Gene Expression Profiling in Vision

Sepideh Zarepari,¹ Alfred Hero,^{2,3,4} Donald J. Zack,^{5,6,7} Robert W. Williams,^{8,9} and Anand Swaroop^{1,10}

Completion of genome sequencing efforts of humans and model organisms¹⁻³ have provided exciting opportunities for vision research. Questions that were hard to envisage, let alone resolve, suddenly appear within reach. We are on the verge of untangling gene networks that regulate development and cell fate and cellular processes that determine adaptation or apoptosis. Patterns of gene expression specific for a particular cell at a given time can now be determined at the genome level. Microarray and related technologies are revolutionizing vision research by making possible simultaneous analysis of thousands of genes, and hopefully soon, the entire transcriptome.⁴ The challenge is to understand how relatively simple variations in DNA sequence and gene expression contribute to complex biological phenomena, including normal development, aging, disease pathogenesis, and differences among individuals. We will briefly review strategies for designing microarray experiments for expression profiling, image processing including normalization, and data analysis including statistics, clustering, and pathway construction.

TYPES OF MICROARRAYS

Microarray experiments include two components: (1) immobilized DNA on solid surfaces called “probes” and (2) labeled cDNAs (or cRNAs) derived from RNA samples that are being analyzed called “targets.” We will discuss commonly used cDNA arrays on glass-slides⁵ and short oligonucleotide arrays on quartz chips (Affymetrix, Santa Clara, CA), with their unique advantages and shortcomings.

From the Departments of ¹Ophthalmology and Visual Sciences, ²Electrical Engineering and Computer Science, ³Statistics, ⁴Biomedical Engineering, and ¹⁰Human Genetics, University of Michigan, Ann Arbor, Michigan; the Departments of ⁵Ophthalmology, ⁶Molecular Biology and Genetics, and ⁷Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland; and the Departments of ⁸Anatomy and Neurobiology and ⁹Pediatrics, University of Tennessee Health Science Center, Memphis, Tennessee.

Supported by grants from the National Eye Institute, The Foundation Fighting Blindness (Owings Mills, Maryland), the Macula Vision Research Foundation (West Conshohocken, Pennsylvania), the Elmer and Sylvia Sramek Charitable Foundation (Chicago, Illinois), and Research to Prevent Blindness (RPB; New York, New York). AS is the Harold F. Falls Collegiate Professor and DJZ the Guerrieri Professor of Genetic Engineering and Molecular Ophthalmology. AS and DJZ are RPB Senior Scientific Investigators.

Submitted for publication February 20, 2004; accepted March 12, 2004.

Disclosure: **S. Zarepari**, None; **A. Hero**, None; **D.J. Zack**, None; **R.W. Williams**, None; **A. Swaroop**, None.

Corresponding author: Anand Swaroop, Department of Ophthalmology and Visual Sciences, W. K. Kellogg Eye Center, 1000 Wall Street, Ann Arbor, MI 48105; swaroop@umich.edu.

STUDY DESIGN

The goal of microarray experiments is to measure the concentration of a given mRNA species in cells or tissues of interest. The measurements are typically *relative*, although we would often prefer to measure *absolute* concentrations. Figure 1 illustrates a generic microarray design. In general, mRNA is reverse transcribed into cDNA, labeled with appropriate dye, and hybridized against the array. Most microarray gene-profiling studies involve single-factor comparisons and/or time-series analysis. In a single-factor comparison, expression profiles from two or more groups are compared to each other, such as wild-type versus knockout, or normal versus diseased tissue. In time-series analysis, expression changes are compared across a group of samples at different time points—for example, cultures exposed to trophins or toxins for different periods, or samples from different age groups. In cDNA arrays, two experimental options are available. In *direct comparisons*, RNA from one sample is labeled with Cy-3, and RNA from a different sample with Cy-5. In *indirect comparisons*, a common reference RNA is labeled with Cy-3, and RNA from each of the experimental samples is labeled with Cy-5 and hybridized separately against the labeled reference RNA. The choice of the approach depends on the biological question and available resources. *Direct comparison* can provide a more precise measurement of the difference in expression between two samples in single-factor experiments. In contrast, *indirect comparison* allows assessment across several targets and makes possible comparison across multiple groups and different experiments. Overall, this method requires less RNA and fewer slides when comparing numerous samples, and though it has slightly higher noise, it provides equivalent results.⁶⁻⁸ There are more complex experimental paradigms, such as 2 × 2 factorial and loop designs, but their discussion are covered elsewhere.^{6,9,10}

An essential aspect of good microarray study design is replication, both to increase the likelihood of identifying true positives and decrease the risk of false positives. There are two types of replication: (1) technical replication in which the same RNA sample (or aliquots from the same source) are applied to different arrays and (2) biological replication in which RNA samples from independent sources are used. Technical replicates are essential in quantifying the experimental variation due to “technical” aspects of the experiment, such as those arising from differences in RNA preparation, degradation, labeling, array irregularities, hybridization, washing, and image analysis. Biological replicates help quantify the variation from the actual biological system being studied, such as differences between experimental samples caused by litter effects, pathogen load, or diet. Both types of replicates are essential to allow appropriate statistical tests to deduce meaningful conclusions about the sampled population. In cDNA arrays, technical replicates often include a dye swap in which the same RNA sample (labeled once with one dye and then with the other) is hybridized against two separate arrays. These exper-

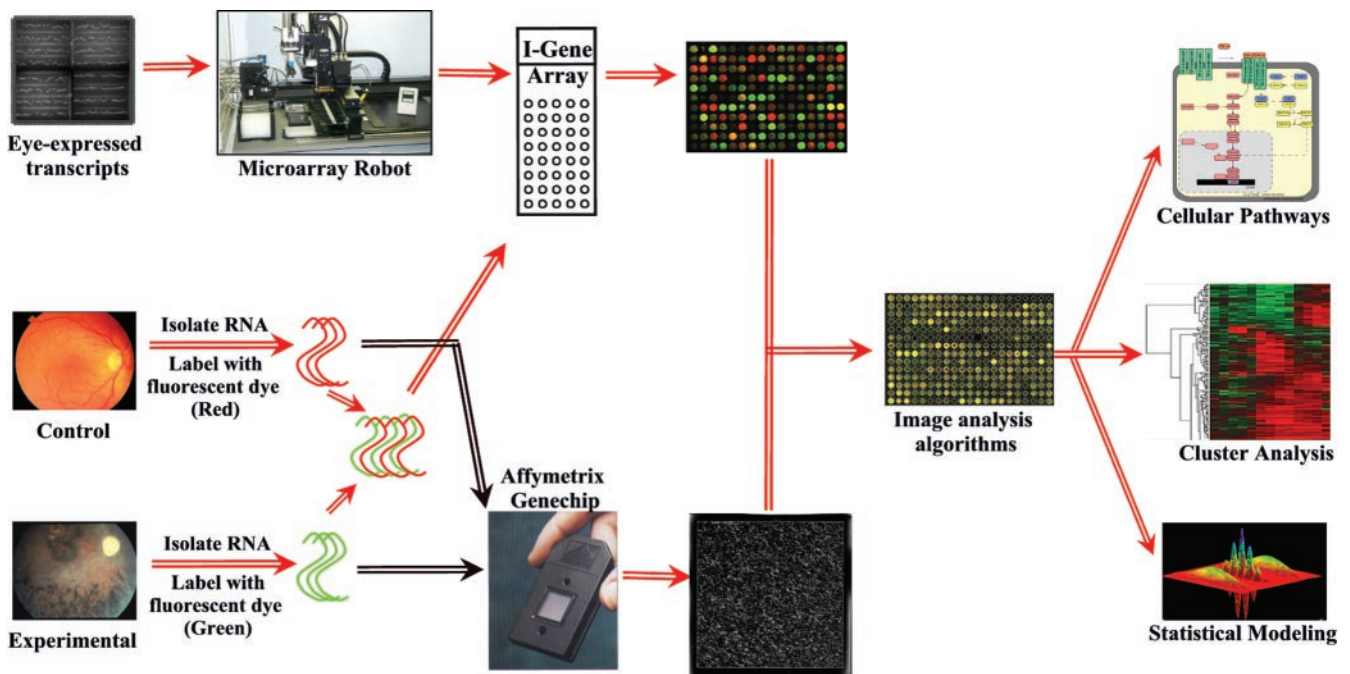


FIGURE 1. A schematic representation of microarray experiments. cDNA arrays use a two-color hybridization system, in which cDNAs from one sample are labeled with one dye, often Cy-3, and cDNAs from another sample are labeled with a different dye, often Cy-5. By measurement of the amount of each dye bound to every spot, the relative abundance of mRNA species is estimated. With Affymetrix (Santa Clara, CA) oligonucleotide arrays, mRNAs from a single sample are reverse transcribed into cDNAs, which serve as a template in an *in vitro* transcription reaction to produce biotin-labeled cRNAs that are fragmented and hybridized to the array.

iments provide a measure of bias in the labeling and hybridization of each dye. The spots on cDNA arrays are often duplicated on each slide. Although not independent replicates, the duplicate spots can assist in determining the quality of the hybridization. In general, multiple independent biological replicates for each condition should be included.

A common dilemma is how many replicates are appropriate to detect true biological change. To answer this, we need to know the variance in signal intensity, the magnitude of the effect to be detected (i.e., level of fold change), the rate of false positives, and the desired power of the study (i.e., the probability of detecting the specified fold change). In microarrays, it is difficult to estimate the variance in signal intensity before performing experiments, because variance for each probe differs, depending on sequence, expression level, tissue source, and complexity (cross hybridization).¹¹ In general, at least three to five replicates are needed.^{12,13} Increasing the sample size naturally reduces the error term. Assuming that the within-group error term is approximately normal, the critical values of the *t*-test may be used to estimate the probabilities of both type I and type II errors.

To maximize the extraction of meaningful data, it is critical to reduce sample variation that is not due to the effect or treatment being tested. When human tissues are used, the variables include obvious differences between samples (sex, ethnicity, age, and health status), premortem conditions (such as, drugs, cause of death, body temperature, and number of days on ventilators), postmortem interval, dissection, and tissue preservation methods. Although animal experiments are easier to standardize than human studies, even comparisons between seemingly identical samples from isogenic mice can reveal significant variability.¹⁴ Experiments involving tightly controlled populations of animals must still consider the effect of hormones, immune system status, light level, and time of day at tissue collection.

Batch effects can also introduce spurious results. For example, if one performs all studies of male retinas in the first phase of a study and female retinas in the second phase a few months later, one has unintentionally introduced a serious design confound between sex and batch. An interleaved experimental design (each group represented equally in each batch) with technical replications across batches is almost essential. Technical sources of variation are also a major source of false-positive results and batch effects. One should use the same RNA extraction protocol performed, if possible, by the same person for all samples in any given experiment. The quality of RNA used as starting point for microarrays is an essential variable, as degraded RNA samples can yield false positive results.^{15,16} Additional problematic batch effects can be introduced by reagents used to transcribe, amplify, and label mRNA. Batch effects can be detected by including a common "spiked" technical replicate in each batch. However, batch corrections are usually transcript-specific, and with sufficient biological and technical overlap between batches, one can compute batch corrections for each transcript. This is not possible with small batches, and large batches are often necessary. When using oligonucleotide arrays, it is preferable to use arrays from the same lot for all samples, since arrays from different lots can exhibit different background levels. Similarly, if using cDNA arrays, it is better to use slides from the same printing, though the use of a reference target in a two-color system offers obvious advantages. It is also important to treat control and experimental samples evenly rather than using one set of reagents/arrays for control samples and another for experimental samples.

One potential limitation of using any tissue as the source of RNA is the heterogeneity of cell types. Low-abundance transcripts present in a tissue often are not detected by microarrays. For example, the Affymetrix arrays can detect mRNAs with a relative abundance down to 5 to 10 transcripts per

million.¹⁷ This corresponds roughly to a few transcripts per cell. In some cases, an abundant transcript in a rare cell type is diluted and is not detectable in a bulk analysis. Methods, such as single cell isolation,¹⁸ laser capture microdissection,^{19,20} and flow sorting of cells,²¹ allow researchers to isolate specific cell types and generate their expression profile. RNA may have to be amplified in one or more rounds to produce sufficient quantities of cRNA or cDNA, but this may introduce changes in relative abundance of transcripts due to nonlinear amplification. However, several studies have compared expression profiles obtained from amplified RNA and total RNA and have found good correlations between the two samples.^{22,23} Despite this host of somewhat intimidating factors, reliable and comparative gene expression profiles can be obtained if the study is designed correctly, controlling for variations and using appropriate replicates.

DATA ANALYSIS

A major challenge in obtaining biologically relevant results with microarray studies is the analysis of large, complex, and sometimes noisy data sets. We will discuss three levels of data analysis. Low-level analysis consists primarily of normalization and filtering data sets (generating lists of interesting transcripts), whereas mid- and high-level analyses consist of more advanced statistical methods, including clustering, network analysis, and integration of known biological patterns.

Low-Level Analysis

Microarray data (i.e., signal intensities) can be transformed from the original scale on which hybridization signal is measured. The original distribution usually has a right skew due to a small number of genes that exhibit high expression levels. Taking the logarithm (usually base-2) typically produces a distribution of values that is close to normal. More sophisticated data-transformation methods²⁴ exploit an accurate statistical model for the raw responses and account for background intensity variations that can cause biases in log-transformed data. To compare expression levels effectively across multiple samples and arrays, data must be adjusted in several additional steps. For example, there is often large variation within and between slides (especially problematic with cDNA arrays). Factors including differences in dye incorporation, hybridization efficiencies, and scanner settings can contribute to technical variation and idiosyncratic nonlinearities. Normalization procedures, in general, attempt to soften the impact of these extrinsic sources of variation while preserving the underlying biological variation that is the focus of the research. There are two levels of normalization: within-array and across arrays. Within-array normalization is often used to compensate for uneven spotting and hybridization. It can be performed locally (i.e., at pin level) or globally. Local normalization may be more appropriate when there are inconsistencies within an array, such as those caused by differences in the spotting pins, surface variation within slide, and differences in hybridization quality across the slide. Several methods of normalization are available for each type of array, and new methods are continually being developed.

We briefly mention a few widely used normalization methods (for details see Refs. 25–27). With respect to cDNA arrays, a robust, locally weighted scatterplot smoothing (*lowess*) method²⁸ has been developed for normalization. It offers the advantage of removing intensity-dependent effects on the log₂ ratio values by reducing the contribution from array spots that are on the two extremes of signal intensity. Moreover, *lowess* and other normalizations may be applied globally or locally. Another approach is data-driven *lowess* normalization, which

uses a cluster of least-altered genes on the array identified by a rank-based algorithm.⁷ In the case of short oligonucleotide arrays from Affymetrix, the key programs, MAS 5.0 and its successor, GREX, offer the option of using different methods for global normalization. A robust multiarray analysis (RMA) has been developed that offers the advantage of performing normalization at probe level for all the probes on an array.²⁹ RMA and the Affymetrix implementation called GREX, allows users to conduct a global background adjustment and across array normalization. Another normalization method is *z*-score transformation, which allows standardization of data across multiple arrays and comparison of the data independent of the hybridization intensities.³⁰ Bolstad et al.²⁹ describe quantitative comparisons between some of the normalization methods for oligonucleotide arrays.

Mid-Level Analysis

The next step is to determine whether an observed change in gene expression represents a true biological effect or a random fluctuation. Many studies have used arbitrary fold change threshold levels (often twofold) on sample mean probe response to identify differentially expressed genes, without accounting for statistical variations. Although it may seem logical to pursue transcripts that exhibit a comparatively high level of modulation, large differences in sample mean response can be statistically insignificant, whereas much smaller differences may be highly significant and biologically relevant. Fold change cutoffs are too arbitrary to be useful in general and should be accompanied by statistical tests. Conventional *t*-tests are appropriate when the error variance within each group appears to be normally distributed. If not, then nonparametric, rank-order tests are more appropriate and more robust. In either type, *P*-values must be adjusted for the number of comparisons performed. Given the large number of comparisons in microarrays, the likelihood is high that many of the putative differentially expressed genes are false positives. Several modified versions of *t*-tests have been developed for analysis of microarray data (Bioconductor Web site). Alternatively, false-discovery rate (FDR) can be used to estimate the proportion of false-positives among all the differentially expressed genes. FDR is more powerful than applying a family-wise error rate (FWER), with Bonferroni's correction to the *P*-value, for multiple testing. Fold change cutoff criteria can be incorporated into the FDR framework by implementing false discovery rate confidence interval (FDR-CI) analysis.³¹ This allows the experimenter to control statistical significance and biological significance simultaneously, in reporting positive differential responses. Finally, analysis of variance (ANOVA) can be used in more complex microarray experiments involving more than two conditions. For an excellent review, see Cui and Churchill.³²

High-Level Analysis

It is often helpful to identify transcripts with similar patterns of change, because they may provide insights into functional pathways or common regulatory mechanisms. Hierarchical clustering, K-means clustering, self-organizing maps, and mixture-model clustering have been used for grouping and discrimination among the many "signatures" implicit in array data sets. Hierarchical clustering is an unsupervised method that uses similarity or distance measures to distinguish among samples. Using a list of differentially expressed genes, genes with similar expression covariance across a set of conditions/samples are clustered together on the vertical axis. If a large number of samples are available, a subset could be used for supervised clustering by cross validation. Here, a subset of data is used as a training set, and then this learned information is used to

analyze the remaining data to predict how well remaining data fit into predicated classes. This has been used successfully in cancer research to predict cancer subtypes and to generate prognostic and diagnostic biomarkers.³³ Self-organizing maps (SOMs) and K-means clustering are also used to visualize similar gene expression patterns across multiple samples or conditions or time-points. K-means clustering can optimally partition genes into a fixed number of clusters by minimizing within-group dissimilarity for a specified number of groups. The quality of clusters obtained may depend on the number of clusters specified. Hence, the number of clusters must be chosen carefully and should be repeated several times with different numbers to find the best clusters. Similarly in SOM, the user must define the number of clusters. The clusters are arranged so that neighboring clusters are similar to each other. In general, as the number of clusters increases, the more similar expression patterns are obtained among genes in any given cluster; however, this is accompanied by higher statistical variation of clusters, leading to unstable clustering results. The tradeoff between high specificity (many clusters) and stability (few clusters) makes a priori selection of the number of clusters difficult. Unsupervised mixture-model-based clustering algorithms³⁴ that automatically select the number of clusters from the data have been adapted recently to clustering gene time courses in cDNA microarrays.⁸ These algorithms characterize the clusters as locally Gaussian and use a complexity penalized expectation-minimization (EM) maximum-likelihood strategy to fit cluster groupings to the data. A very different grouping strategy, called Pareto analysis, was recently introduced for detecting trends in gene expression from a time-course microarray experiment.³⁵ In this method, genes are filtered through multiobjective optimization to identify the genes that maximize a set of user-specified fitness criteria yet are statistically stable.³⁶ Various other clustering methods for gene microarray experiments are described in detail in a book.³⁷

After generating a list of differentially expressed genes, one can ask how these genes interact with each other and how they contribute to the phenotype of interest. In yeast³⁸ and *Drosophila*,³⁹ extensive networks of transcriptional regulators have been constructed using microarray data to identify interacting transcripts. In mammalian systems, the progress is hampered by our lack of knowledge about function of many known genes and the organization of pathways. One available approach involves exploiting gene classification schemes formalized as gene ontologies. A gene ontology is a structure that can help highlight pathway components and categories of molecules affected by a treatment or disease process. Because a gene may be involved in more than one pathway, many different pathways may be represented. Additional publicly available resources are Gene MicroArray Pathway Profiler, the Kyoto Encyclopedia of Genes and Genomes, and Biocarta. One of the most powerful tools for exploring networks of interacting genes is WebQTL, which allows one to explore arbitrary complex associative networks built around essentially any transcript, including poorly annotated expressed sequence tags (ESTs).⁴⁰ In the next few years we can expect many novel programs to be developed to explore and dissect gene networks and pathways.

VALIDATION

Microarray studies often yield a large number of differentially expressed genes that may require validation depending on study design and statistical methods used. In simple comparisons involving a few groups and samples, the expression changes in a subset of genes (defined based on functional

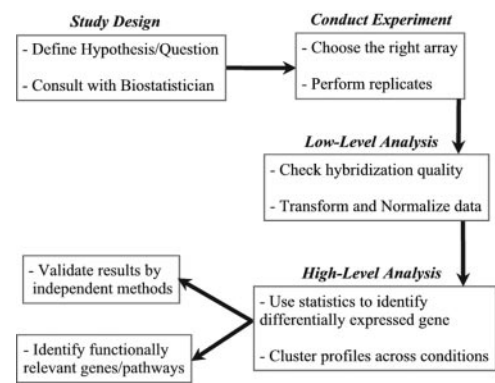


FIGURE 2. A guide for designing microarray experiments highlighting the important components.

interest, fold change levels, or FDRs) should be confirmed by independent methods (such as real-time RT-PCR). In more complex studies involving large sample sizes, validations may be impractical. Thus, considerable attention must be paid to proper study design and statistical methods so that results can be interpreted with confidence.

APPLICATIONS

A PubMed search revealed over 70 published studies in which microarrays were used in vision research. Although similar studies have been performed in lens^{41,42} and other eye tissues,^{43,44} a few examples from retinal tissue are given here. The microarray studies have ranged from identification of genes altered during aging of human retina,^{11,45} genes involved in organization of developing retina in embryonic mice,⁴⁶ identification of a disease-causing gene based on differential expression between wild-type and rhodopsin knockout mice,⁴⁷ and identification of novel genes that may be preferentially expressed in the retina.⁴⁸ Although all the studies have made significant contributions to ophthalmic research, greater success could be achieved if arrays that contain all the genes expressed in the eye are used. Currently, there is an underrepresentation of the eye-specific genes in the publicly available databases. Several groups have spent considerable effort and resources to produce custom eye expression arrays. cDNA arrays containing more than 3000 ESTs from three mouse eye/retina cDNA libraries are available.⁴⁹ The NEIbank has constructed cDNA libraries from many sections, including retina and RPE/choroid that are available to the vision-research community.^{50,51} Recently, annotation and analysis of 10,000 ESTs was reported from adult mouse retina and from mouse eyes at embryonic day 15.5 and postnatal day 2.⁵²

SUMMARY

Microarray technology has been a rapidly evolving field and will continue to advance at a swift pace. As the use of arrays in ophthalmic research becomes more pervasive, it is essential that this technology be used appropriately. With this much experimental power, there is substantial risk of generating a large body of intriguing experimental artifacts. There is an equally serious possibility of missing significant results by applying procrustean or superficial statistical analyses. Figure 2 provides a flow chart that should be helpful for designing appropriate microarray experiments. It is imperative to consider study designs and the objectives carefully before embarking on microarray experiments. Appropriate normalization and statistical treatment of microarray data will extract useful new

information, thereby allowing efficient generation and validation of hypotheses.

LOOKING TO THE FUTURE

Genome-wide expression profiling with microarrays is expected to yield tremendous amounts of data. Efficient mining of these data is currently a challenge for vision scientists and computational biologists. It is critical to develop acceptable standards for microarray experimentation and guidelines for data presentation, storage, and sharing.⁵⁵ We are approaching a new frontier of vast information. Translation of this technology to uncover fundamental interacting gene networks and to perform targeted drug design for treatment of blinding diseases will require a conscious and concerted effort.

ELECTRONIC DATABASE INFORMATION

A few useful web sites for obtaining additional information on microarrays are provided below.

Affymetrix: www.affymetrix.com

Biocarta: www.biocarta.com

Bioconductor: www.bioconductor.org

dChip web site: www.dchip.org

Friedlander Retinal Expression DB: www.scripps.edu/cb/friedlander/gene_expression

Gene MicroArray Pathway Profiler: www.genmapp.org

Gene Ontology Information: www.geneontology.org

Kyoto Encyclopedia of Genes and Genomes: www.genome.ad.jp/kegg

Minimum information about a microarray experiment (MIAME): <http://www.mged.org/Workgroups/MIAME/miame.html>

Microarray Profiling of Rodent Models: www.pga.tigr.org

NEIBank: <http://neibank.nei.nih.gov>

RetBase: www.hgsc.bcm.tmc.edu/retbase

RetinalExpress: <http://odin.mdacc.tmc.edu/RetinalExpress>

Stanford Microarray Database: <http://genome-www5.stanford.edu>

Swaroop Laboratory Microarray Site and I-Gene Database: www.umich.edu/~retina/microarray.html

The Mouse SAGE site: www.mouse.biomed.cas.cz/sage

WebQTL Project: www.webqtl.org/search.html

Note Added in Proof

The FDRCI method has been implemented in R package (umarray.fdrcli) and is now publicly available at <http://www.umich.edu/~retina/microarray.html>.

Acknowledgments

The authors thank David Beebe, Rafal Farjo, Gilles Fleury, Leonard M. Hjelmeland, Ritu Khanna, Alan J. Mears, and Mohammad Othman for stimulating discussions and Sharyn Ferrara for administrative assistance.

References

- Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
- Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.
- Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002;420:520-562.
- Swaroop A, Zack DJ. Transcriptome analysis of the retina. *Genome Biol*. 2002;3:REVIEWS1022.
- Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270:467-470.
- Yang YH, Speed T. Design issues for cDNA microarray experiments. *Nat Rev Genet*. 2002;3:579-588.
- Yu J, He S, Friedman J, et al. Altered expression of genes of the Bmp/Smad and Wnt/Ca2+ signaling pathways in the cone-only Nrl-knockout mouse retina, revealed by gene profiling using custom cDNA microarrays. *J Biol Chem*. Submitted.
- Yu J. Transcript profiling of the cone-only Nrl-knockout retina using custom cDNA-microarrays: identification of novel targets of Nrl and of signaling pathways involved in photoreceptor function. *Biomedical Engineering*. Ann Arbor, MI: University of Michigan; 2004:148-153.
- Kerr MK, Churchill GA. Experimental design for gene expression microarrays. *Biostatistics*. 2001;2:183-201.
- Parmigiani G, Garrett E, Irizarry R, Zeger S. *The Analysis of Gene Expression Data: Methods and Software*. New York: Springer; 2003.
- Chowers I, Liu D, Farkas RH, et al. Gene expression variation in the adult human retina. *Hum Mol Genet*. 2003;12:2881-2893.
- Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA*. 2000;97:9834-9839.
- Pavlidis P, Li Q, Noble WS. The effect of replication on gene expression microarray experiments. *Bioinformatics*. 2003;19:1620-1627.
- Blackshaw S, Kuo WP, Park PJ, et al. MicroSAGE is highly representative and reproducible but reveals major differences in gene expression among samples obtained from similar tissues. *Genome Biol*. 2003;4:R17.
- Auer H, Lyianarachchi S, Newsom D, et al. Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat Genet*. 2003;35:292-293.
- Spiess AN, Mueller N, Ivell R. Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. *BMC Genomics*. 2003;4:44.
- Chudin E, Walker R, Kosaka A, et al. Assessment of the relationship between signal intensities and transcript concentration for Affymetrix GeneChip arrays. *Genome Biol*. 2002;3:RESEARCH0005.
- Wahlin KJ, Lim L, Grice EA, et al. Analysis of gene expression in isolated, microcaptured mouse photoreceptor and Müller cells. *Mol Vis*. Submitted.
- Sgroi DC, Teng S, Robinson G, et al. In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res*. 1999;59:5656-5661.
- Ohyama H, Zhang X, Kohno Y, et al. Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques*. 2000;29:530-536.
- Livesey FJ, Young TL, Cepko CL. An analysis of the gene expression program of mammalian neural progenitor cells. *Proc Natl Acad Sci USA*. 2004;101:1374-1379.
- Pabon C, Modrusan Z, Ruvolo MV, et al. Optimized T7 amplification system for microarray analysis. *Biotechniques*. 2001;31:874-879.
- Iscove NN, Barbara M, Gu M, et al. Representation is faithfully preserved in global cDNA amplified exponentially from sub-pico-gram quantities of mRNA. *Nat Biotechnol*. 2002;20:940-943.
- Yang W, Speed T. Design and analysis of comparative microarray experiments. In: Speed T, ed. *Statistical Analysis of Gene Expression Microarray Data*. Boca Raton, FL: Chapman & Hall/CRC Press; 2003;35-81.
- Quackenbush J. Microarray data normalization and transformation. *Nat Genet*. 2002;32(suppl):496-501.
- Smyth GK, Speed T. Normalization of cDNA microarray data. *Methods*. 2003;31:265-273.

27. Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. *Trends Genet.* 2003;19:649–659.
28. Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 2002;30:e15.
29. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003;19:185–193.
30. Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. *J Mol Diagn.* 2003;5:73–81.
31. Hero A, Fleury G, Mears A, Swaroop A. Multicriteria gene screening for analysis of differential expression with DNA microarrays. *EURASIP JASP.* 2004;1:43–52.
32. Cui X, Churchill GA. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* 2003;4:210.
33. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002;415:530–536.
34. McLachlan GJ, Bean RW, Peel D. A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics.* 2002;18:413–422.
35. Fleury G, Hero A, Yoshida S, et al. Pareto analysis of gene filtering in microarray experiments. *XI European Signal Processing Conference.* Toulouse, France: Telecommunications Spatiales et Aeronautiques; 2002:165–168.
36. Fleury G, Hero A, Zareparsy S, Swaroop A. Gene discovery using pareto depth sampling distributions. *J Franklin Institute.* 2004;341:55–75.
37. Chipman H, Hastie T, Tibshirani R. *Clustering Microarray Data.* In: Speed T. *Statistical Analysis of Gene Expression Microarray Data.* Chapman & Hall/CRC Press: Boca Raton; 2003:159–200.
38. Lee TI, Rinaldi NJ, Robert F, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science.* 2002;298:799–804.
39. Orian A, van Steensel B, Delrow J, et al. Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network. *Genes Dev.* 2003;17:1101–1114.
40. Wang J, Williams R, Manly K. WebQTL: Web-based complex trait analysis. *Neuroinformatics.* 2003;1:299–308.
41. Hawse JR, Hejtmancik JF, Huang Q, et al. Identification and functional clustering of global gene expression differences between human age-related cataract and clear lenses. *Mol Vis.* 2003;9:515–537.
42. Chauhan BK, Reed NA, Yang Y, et al. A comparative cDNA microarray analysis reveals a spectrum of genes regulated by Pax6 in mouse lens. *Genes Cells.* 2002;7:1267–1283.
43. Ishibashi T, Takagi Y, Mori K, et al. cDNA microarray analysis of gene expression changes induced by dexamethasone in cultured human trabecular meshwork cells. *Invest Ophthalmol Vis Sci.* 2002;43:3691–3697.
44. Weigel AL, Handa JT, Hjelmeland LM. Microarray analysis of H2O2-, HNE-, or tBH-treated ARPE-19 cells. *Free Radic Biol Med.* 2002;33:1419–1432.
45. Yoshida S, Yashar BM, Hiriyanna S, Swaroop A. Microarray analysis of gene expression in the aging human retina. *Invest Ophthalmol Vis Sci.* 2002;43:2554–2560.
46. Diaz E, Yang YH, Ferreira T, et al. Analysis of gene expression in the developing mouse retina. *Proc Natl Acad Sci USA.* 2003;100:5491–5496.
47. Kennan A, Aherne A, Palfi A, et al. Identification of an IMPDH1 mutation in autosomal dominant retinitis pigmentosa (RP10) revealed following comparative microarray analysis of transcripts derived from retinas of wild-type and Rho(–/–) mice. *Hum Mol Genet.* 2002;11:547–557.
48. Chowers I, Gunatilaka TL, Farkas RH, et al. Identification of novel genes preferentially expressed in the retina using a custom human retina cDNA microarray. *Invest Ophthalmol Vis Sci.* 2003;44:3732–3741.
49. Farjo R, Yu J, Othman MI, et al. Mouse eye gene microarrays for investigating ocular development and disease. *Vision Res.* 2002;42:463–470.
50. Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of human retina for the NEIBank Project: retbindin, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts. *Mol Vis.* 2002;8:196–204.
51. Wistow G, Bernstein SL, Wyatt MK, et al. Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: over 6000 non-redundant transcripts, novel genes and splice variants. *Mol Vis.* 2002;8:205–220.
52. Yu J, Farjo R, MacNee SP, et al. Annotation and analysis of 10,000 expressed sequence tags from developing mouse eye and adult retina. *Genome Biol.* 2003;4:R65.
53. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet.* 2001;29:365–371.

New Developments in Vision Research

Written for a broad audience, the articles in this column succinctly and provocatively review a rapidly changing area of visual science that shows progress and holds potential. Authors and topics are chosen by the Editor-in-Chief in collaboration with the Editorial Board.

To avoid bias, the Editor-in-Chief subjects these articles to the same rigorous peer review process to which all other *IOVS* articles are subjected. Space and reference limitations are imposed on the authors.

The purpose of this series is not the recognition of individual scientists, nor exhaustive review of a subject, but the stimulation of interest in a new research area.

Editor-in-Chief