EURASIP Journal on Applied Signal Processing 2004:1, 43–52

Multicriteria Gene Screening for Analysis of Differential Expression with DNA Microarrays

Alfred O. Hero

Departments of Electrical Engineering and Computer Science, Biomedical Engineering, and Statistics, University of Michigan, Ann Arbor, MI 48109, USA
Email: hero@eecs.umich.edu

Gilles Fleury

Service des Mesures, Ecole Supérieure d'Electricité, 91192 Gif-sur-Yvette, France
Email: fleury@supelec.fr

Alan J. Mears

Departments of Ophthalmology and Visual Sciences, and Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA

University of Ottawa Eye Institute, Ottawa Health Research Institute, Ottawa, ON Canada, K1H 8L6
Email: amears@ohri.ca

Anand Swaroop

Departments of Ophthalmology and Visual Sciences, and Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA
Email: swaroop@med.umich.edu

Received 10 May 2003; Revised 30 August 2003

This paper introduces a statistical methodology for the identification of differentially expressed genes in DNA microarray experiments based on multiple criteria. These criteria are false discovery rate (FDR), variance-normalized differential expression levels (paired t statistics), and minimum acceptable difference (MAD). The methodology also provides a set of simultaneous FDR confidence intervals on the true expression differences. The analysis can be implemented as a two-stage algorithm in which there is an initial screen that controls only FDR, which is then followed by a second screen which controls both FDR and MAD. It can also be implemented by computing and thresholding the set of FDR P values for each gene that satisfies the MAD criterion. We illustrate the procedure to identify differentially expressed genes from a wild type versus knockout comparison of microarray data.

Keywords and phrases: bioinformatics, gene filtering, gene profiling multiple comparisons, familywise error rates.

1. INTRODUCTION

Since Watson and Crick discovered DNA more than fifty years ago, the field of genomics has progressed from a speculative science to one of the most thriving areas of current research and development [1]. After successful completion (99%) of the Human Genome project [2], attention is turning to "functional genomics" and "proteomics," thanks principally to remarkable advances in computations and technology. These disciplines encompass the greater challenge of understanding the complex functional behavior and interaction of genes and their encoded proteins at the cellular level. This task has been significantly aided by the advent of DNA microarray technology and associated algorithms that enable researchers to filter through daunting amounts of data and genetic information. In this paper, we describe a new approach to extracting a subset of differentially expressed genes from DNA microarray data.

A DNA microarray consists of a large number of DNA probe sequences that are put at defined positions on a solid support such as a glass slide or a silicon wafer [3, 4]. After hybridization of a fluorescently labelled sample (gene transcripts) to DNA microarrays, the abundance of each probe present (called probe response) in the sample can be estimated from the measured levels of hybridization (i.e., the intensity of fluorescent signal). Two main types of DNA microarrays are in wide use for gene expression profiling: Affymetrix GeneChips [5], which are generated by photolithography; and spotted cDNA (or oligonucleotide) arrays on glass slides [6].

DNA microarrays enable biologists to study global gene expression profiles in tissues of interest over time periods and under specific conditions or treatments. For these cases, a large set of samples, consisting of several biological replicates, are hybridized to a set of microarrays. The objective is to identify subsets of genes whose expression profile over time exhibit salient behavior(s), for example, differ in response to different treatments. A crucial aspect of selecting the genes of interest is the specification of a preference ordering for ranking the probe responses. Many gene selection and ranking methods are based on testing fitness criteria such as the eigenvalue spread in a principal components analysis (PCA) of all pairs of gene expression profiles, the ratio of between-population-variation to within-population-variation, or the cross correlation between profiles [7, 8, 9].

These methods have deficiencies which have impeded their use for practical experiments. First, is the need for improved relevance of the fitness criterion to the scientific objectives of the experiment. It is often difficult for an experimenter to choose quantitative criteria that characterize the aspects of a gene expression profile of interest. Second, is the need for simultaneous control of the biological significance (minimum acceptable difference (MAD)) and the statistical significance (false discovery rate (FDR)) of differential responses discovered in the selected gene probes. A probe response difference which is too small is not of much use to the experimenter even if the difference is statistically significant. This is because the microarray experiment is usually only the first step in gene discovery; each microarray probe difference that is discovered must be validated by painstaking-followup analysis that may have limited sensitivity to small differences. Third, is the need for tight confidence intervals (CIs) on these differences. The size of a CI provides useful information on the statistical precision of an estimate of differential response.

The method we present in this paper adopts a statistical multicriteria framework for gene microarray analysis with MAD constraints on differential expression. The framework allows the experimenter to adopt multiple fitness criteria, explicitly incorporate control on biological significance in addition to statistical significance, and generate confidence intervals on discovered gene expression differences. Our method is strongly influenced by the FDR-adjusted confidence interval (FDR-CI) approach recently introduced by Benjamini and Yekutieli [10]. We illustrate our methods for a differential expression experiment designed to probe the genetic basis of retinal development. This experiment involves two populations, wild type and knockout, and the objective is to find genes that exhibit biologically and statistically significant differences between these populations. The purpose of this article is to illustrate methodology and not to report scientific findings, which will be reported elsewhere.

It is worthwhile to compare the framework developed in this paper to related work. Liu and Iba have proposed an interesting multicriteria evolutionary approach to gene selection and classification in gene microarray experiments [11]. Similarly, Fleury and Hero have proposed Pareto optimality for selecting subsets of genes using a combination of bootstrap resampling and Bayes decision theory [12, 13, 14]. Single stage [15] and multistage [16, 17, 18] screening methods which control familywise error rate (FWER) or FDR have been proposed by several authors for similar problems to ours. However, none of the above approaches account for a MAD constraint or provide CIs on the differential expression levels of the discovered genes. In contrast, our approach accounts for both FDR and MAD constraints and generates such confidence intervals using the FDR-CI framework [10]. Furthermore, we specify an algorithm for computing FDR P values for all genes at any prescribed MAD level.

Table 1: The knockout versus wild-type experiment is equivalent to a two-way layout of treatment (W or K) and time (t = Pn2, Pn10, M2).

Gene g	Pn2	Pn10	M2
W	4 samples	4 samples	4 samples
K	4 samples	4 samples	4 samples

The outline of the paper is as follows. In Section 2, we give a general description of the type of differential gene microarray experiment that will be illustrated in Section 4. In Section 3, we describe the proposed two-stage multicriteria approach. Finally, in Section 4, we illustrate these techniques for experimental data.

2. DIFFERENTIAL EXPRESSION PROFILE EXPERIMENTS

This type of experiment is very common in genetics research [19, 20] and involves comparing gene expression profiles of a set of G genes expressed in two or more populations. The data from this experiment fall into the category of a two-way layout [21], where each cell in the layout corresponds to a set of replicates of samples from one of the two populations (row) and one of T-time points (column) (see Table 1).

Any gene whose temporal profile differs from wild-type to knockout populations is called "differentially expressed" in the experiment. One variant of this experiment is called the wild-type versus knockout experiment. In such an experiment, one has a control population (wild type) of subjects and a treated population (knockout) of subjects whose DNA has been altered in some way. Each population is comprised of T different age groups arranged in T subpopulations. M independent samples are taken from each subpopulation and are hybridized to a different microarray, yielding G pairs of expression profiles (see Figure 1 for profiles of the gene having probe set number 101996_*at*). This generates a total of $2MT$ microarrays. It is common to express the differential response between wild-type and knockout responses in terms of *foldchange* expressed as the ratio of these responses. For example, a foldchange of 2.0, or 1.0 in log base 2 at a given time corresponds to a wild-type response which is twice as large as the knockout response. We denote by $\{\mu_t(g)\}_{t=1}^{T}$ and $\{\eta_t(g)\}_{t=1}^{T}$ the true log wild-type and log knockout expression profiles, respectively, expressed as log base 2 of the true hybridization abundances.

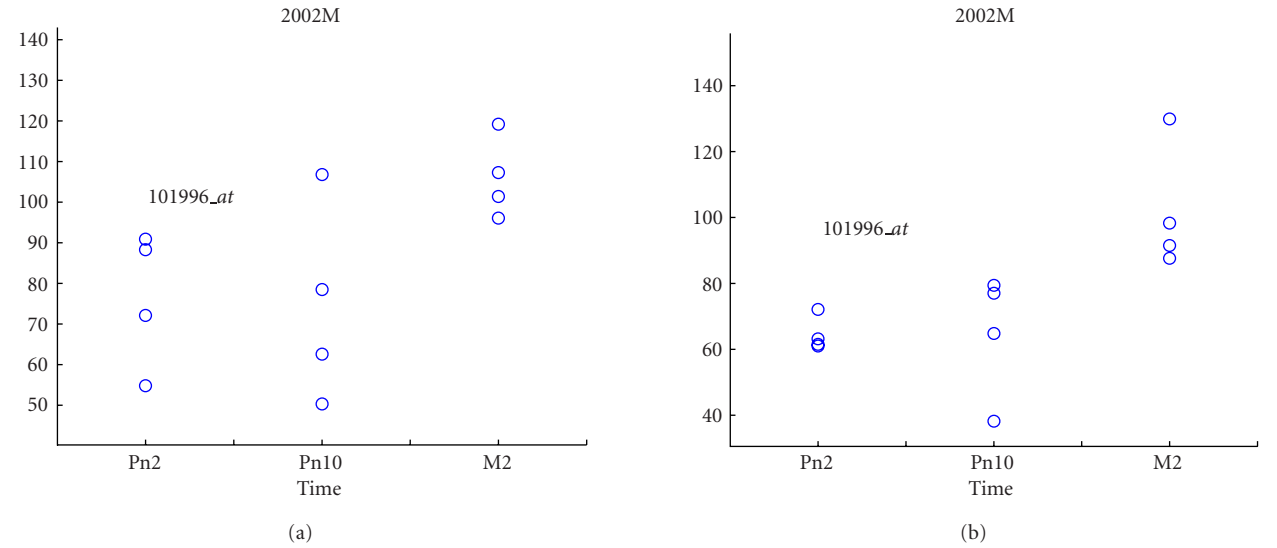

Figure 1: Responses for a particular gene (probe set number 101996_*at*) in (a) knockout mouse versus (b) wild-type mouse for the differential expression study discussed in Section 4. There are three-time points (labeled Pn2, Pn10, and M2) and at each time point, there are four replicates. The y-axis denotes log base 2 hybridization level extracted by RMA from Affymetrix GeneChips.

Figure 2 illustrates the three-dimensional multicriteria space of mean differential responses $\{\mu_t(g) - \eta_t(g)\}_{t=1}^{3}$ for the three-time point experiment described in Section 4. A "MAD box" which defines unacceptably small (inside box) versus acceptably large (outside box) differential responses, and a scatter of a small subset of all the sample mean differential responses (dots) from the experiment are also indicated. Our objective is to discover which genes are likely to have a "positive differential response" falling outside of the box in Figure 2. A very commonly used method is to simply apply a threshold to the sample means to detect those who fall outside of the box in Figure 2 as positive responses. However, as will be shown, this method does not account for statistical sampling uncertainty and can lead to many false positives.

The objective can be stated mathematically as follows: find a set of gene probes which satisfy the MAD constraint: $|\mu_t(g) - \eta_t(g)| > \text{fcmin}$ for at least one $t \in \{1, \ldots, T\}$. Here, the MAD constraint is quantified by the user-specified minimum magnitude foldchange fcmin (expressed in log base 2). Thus, we need to simultaneously test the G pairs of the two-sided hypotheses

$$\begin{aligned} H_0(g) &: |\mu_1(g) - \eta_1(g)| \\ &\quad \le \text{fcmin and, } \cdots, \text{and } |\mu_T(g) - \eta_T(g)| \\ &\quad \le \text{fcmin}, \\ H_1(g) &: |\mu_1(g) - \eta_1(g)| \\ &\quad > \text{fcmin or, } \cdots, \text{or } |\mu_T(g) - \eta_T(g)| \\ &\quad > \text{fcmin}, \end{aligned} \tag{1}$$

where $g = 1, \ldots, G$. Of course, when we must decide between $H_0(g)$ and $H_1(g)$ based on a random sample, there will generally be decision errors in the form of false positives (decide $H_1(g)$ when $H_0(g)$ is true) and false negatives (decide $H_0(g)$ when $H_1(g)$ is true). For any test, the experimenter needs to be able to control both its statistical and biological level of significance. The *statistical level of significance* of the test is specified by the false positive rate. In contrast, the *biological level of significance* of the test is specified by fcmin.

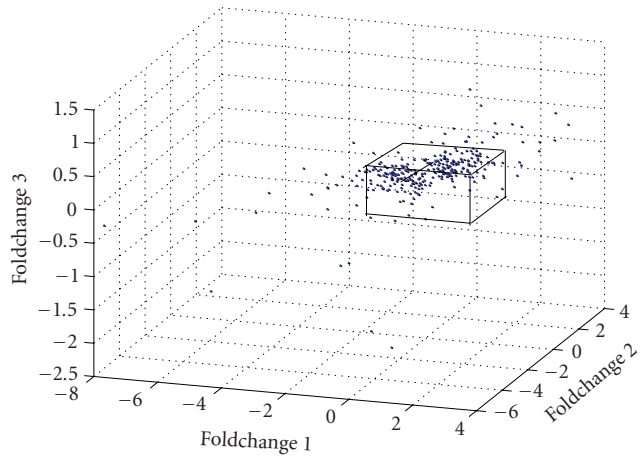

Figure 2: Three-dimensional multicriteria space for knockout and wild-type profiles over three-time points shown in Figure 1. The three criteria are the differential probe responses at each time point. A scatter plot of sample means of the differential responses along with a box of edge length 2fcmin distinguishing biologically significant responses (outside box) from biologically insignificant responses (inside box) is shown.

There are three aspects to the hypothesis-testing problem (1) which make it nonstandard:

(i) standard tests on differences in means, such as the paired t test, treat any nonzero difference as significant,

whereas (1) specifies that only differences exceeding the specified MAD level of fcmin are significant;

(ii) a positive response ($H_1(g)$) is described by multiple criteria, here equal to the T magnitude log response ratios at each point in time;

(iii) the G pairs of hypotheses must be tested simultaneously.

For the case $G = T = 1$, the first aspect can be treated by applying methods for composite hypothesis testing such as generalized likelihood ratio tests, unbiased tests, and CI test procedures [22, 23]. When fcmin $= 0$, (ii) and (iii) can be handled by applying a standard method, like paired t-test, to (1) for each gene probe g, implemented with a multiplicity error-correction factor, for example, Bonferroni, FWDR, or FDR, [24]. However, such a repeated test of significance will result in excessive false positives corresponding to small log response ratios that are biologically insignificant (do not satisfy the MAD constraint) but are statistically significant.

3. MULTICRITERIA GENE SCREENING METHOD

Define $\underline{\xi}(g) = [\xi_1(g), \ldots, \xi_T(g)]$ the true differential response vector associated with gene probe g, where $\xi_t(g) = \mu_t(g) - \eta_t(g)$. Given the DNA microarray data, our objective is to test the G hypotheses (1) involving a total of $P = GT$ unknown parameters $\{\underline{\xi}(g)\}_{g=1}^{G}$.

Any test of (1) must test over multiple criteria $\{\underline{\xi}_t(g)\}_t$ and multiple genes at a given level of biological significance MAD $=$ fcmin and a given level of statistical significance max FDR $= \alpha$. Unless fcmin $= 0$, this is a doubly composite hypothesis-testing problem since the parameter values ξ_t are not specified under H_0 or H_1. Due to the presence of multiple criteria and multiple genes, this problem falls into the area of multiple testing, simultaneous inference, and repeated tests of significance [25, 26]. Two standard measures of statistical significance of a test of (1) are its FWER and its FDR [25]. A mathematically convenient notation for a test of (1) is $\phi(g)$, which is called a *test function*, taking on values 0 or 1 depending on whether the test declares H_0 or H_1 for probe g, respectively. With $\mathcal{G}_0$ denoting the probes not having positive responses, the FWER and FDR of a test ϕ can be mathematically defined as

$$\begin{aligned} \text{FWER}\,(\mathcal{G}_0) &= 1 - E\left[\Pi_{g=1}^{G}(1-\phi(g))\psi_{\mathcal{G}_0}(g)\right], \\ \text{FDR}\,(\mathcal{G}_0) &= E\left[\frac{\sum_{g=1}^{G}\phi(g)\psi_{\mathcal{G}_0}(g)}{\sum_{g=1}^{G}\phi(g)}\right], \end{aligned} \tag{2}$$

where $E[Z]$ denotes statistical expectation of a random variable Z and $\psi_{\mathcal{G}_0}(g)$ is the indicator function of the set $\mathcal{G}_0$. In words, the FWER is the probability that the test of all G pairs of hypotheses (1) yields at least one false positive in the set of declared positive responses. In contrast, the FDR is the average proportion of false positives in the set of declared positive responses. The FDR is dominated by the FWER and is therefore a less stringent measure of significance. Both FWER and FDR have been widely used for gene microarray analysis [16, 17, 24, 27].

It is useful to contrast the FWER and FDR to the per-comparison error rate (PCER). The PCER refers to the false positive error rate incurred in testing a single pair of hypothesis $H_0(g)$ versus $H_1(g)$ for a single gene, say, gene $g = g_o$, and does not account for multiplicity of the hypotheses (1). The PCER is the probability that random sampling errors would have caused g_o to be erroneously selected, generating a false positive, based on observing microarray responses for gene g_o only. If an experimenter were only interested in deciding on the biological significance of a single gene g_o, based only on observing probes for that gene, then reporting PCER(g_o) would be sufficient for another biologist to assess the statistical significance of the experimenter's statement that g_o exhibits a positive response. In contrast to the PCER, FWER and FDR communicate statistical significance of an experimenter's finding of biological significance after observing all gene responses. The FWER is the probability that there are any false positives among the set of genes selected. On the other hand, the FDR refers to the expected proportion of false positives among the selected genes. The FDR is a less stringent criterion than the FWER [25, 27, 28].

The FWER can be upper bounded as a function of $\{\text{PCER}(g)\}_{g=1}^{G}$ using Bonferroni-type methods [26] or it can be computed empirically from the sample by resampling methods [29]. The FDR can be computed by applying the step-down procedure of Benjamini and Hochberg [25] to the list of PCER P values over all genes. For a given g, the PCER P value, denoted $p(g)$, of a test ϕ is a function of the microarray measurements and is defined as the minimum value of PCER for which $H_0(g)$ would be falsely rejected by the test. The set of gene responses which pass the test ϕ at a specified FDR can be simply determined after ordering the genes indices according to increasing PCER P value $p(g_{(1)}) \le \cdots \le p(g_{(G)})$. Specifically, for a fixed value $\alpha \in [0, 1]$ of maximum acceptable FDR, the FDR-constrained test will declare the following set $\mathcal{G}_1$ of genes as positive responses [28]:

$$\begin{aligned} \mathcal{G}_1 &= \{g_{(1)}, \ldots, g_{(K)}\}, \\ K &= \max\left\{k : p(g_{(k)}) \le \frac{k\alpha}{G\nu}\right\}. \end{aligned} \tag{3}$$

In this expression, $\nu = 1$ if the decisions $\phi(g)$ can be assumed statistically independent over $g = 1, \ldots, G$, while $\nu = 1/\sum_{k=1}^{G} k^{-1}$ without the independence assumption.

A test which controls a maximum level α of acceptable FDR is said to be an FDR test of level α. We propose a test ϕ of (1) at FDR level α and MAD level fcmin based on intersecting simultaneous CIs on the T differences $\underline{\xi}(g)$ with the unacceptable difference region $[-\text{fcmin}, \text{fcmin}]$. We will specify a two-stage direct implementation and a single-stage inverse implementation in the following subsections. First, however, we recall some facts about simultaneous CIs.

Let θ be an unknown parameter, for example, a gene's foldchange $\xi_1(g)$ at time $t = 1$. A PCER $(1-\alpha)\times 100\%$ CI on θ is an interval $I(\alpha) = [a, b]$ with random data-dependent endpoints that covers the true θ value, say θ_o, with probability at least $1 - \alpha$:

$$P(a \le \theta_o \le b \mid \theta = \theta_o) \ge 1 - \alpha. \quad (4)$$

There is always a trade-off between confidence level $1-\alpha$ and precision (CI length) since the length $b-a$ of $I(\alpha)$ generally increases as α decreases. Let $\mathcal{A}$ be any subset of $\mathbb{R}$. A PCER CI on θ can be converted to a PCER level-α test of the hypotheses $H_0(g) : \theta \in \mathcal{A}$ versus $H_1(g) : \theta \notin \mathcal{A}$ by the simple procedure: "reject H_0 if the $(1-\alpha)\times 100\%$ CI on θ does not intersect $\mathcal{A}$" [22].

Multiple parameters, $\theta_1, \ldots, \theta_P$, can be simultaneously covered by FWER $(1-\alpha)\times 100\%$ CIs $\{I^p(1-(1-\alpha)^{1/P})\}_{p=1}^{P}$, where $I^p(\alpha)$ is a PCER $(1-\alpha)\times 100\%$ CI on θ_p. Under the assumption that each of the P PCER CIs are statistically independent, the FWER intervals cover all the parameters with probability at least $1-\alpha$ [26]. A less stringent set of CIs $\{I^p(\alpha/P)\}_{p=1}^{P}$, which can be applied to dependent sets of PCER CIs, is guaranteed to cover at least $(1-\alpha)P$ of the unknown parameters [26, 30]. When the number of P of parameters is random, as occurs when the number of parameters results from some initial screening, the above methods cannot be applied. It was for this situation that the FDR-CI approach was developed [10]. If P is the result of initial screening at an FDR level α of Q parameters having PCER-CIs $\{I^p(\alpha)\}_{p=1}^{Q}$, then the FDR-CIs on the P parameters are defined as $\{I^p(P\alpha/Q)\}_{p=1}^{P}$. The FDR-CIs are guaranteed to cover at least $(1-\alpha)\times 100\%$ of the P unknown parameters.

Below, we give two equivalent FDR-CI procedures for screening differentially expressed genes with FDR and MAD constraints.

3.1. Direct two-stage screening procedure

Stage 1. Gene screening at MAD level 0 extracts a set of G_1 genes $\mathcal{G}_1$ by testing (1) under the relaxed MAD constraint fcmin $= 0$ using an FDR level-α test via the step-down procedure (3).

Stage 2. Gene screening at MAD level fcmin > 0 extracts a set $\mathcal{G}_2$ of positive genes from those in $\mathcal{G}_1$ as follows. For each gene $g \in \mathcal{G}_1$, construct T simultaneous CIs, denoted as $\{I_t^g(\alpha)\}_{t=1}^{T}$, of FWER level $(1-\alpha)\times 100\%$ on the true fold-changes $\{\mu_t(g)-\eta_t(g)\}_{t=1}$. Convert these into $(1-\alpha)\times 100\%$ FDR-CIs by the method of Benjamini and Yekutieli [10]: $I_t^g(\alpha) \to I_t^g(G_1\alpha/G)$, $t = 1, \ldots, T$, $g = 1, \ldots, G$. Finally, define the set of indices $\mathcal{G}_2$ of gene profiles having at least one-time point, where the FDR-CI does not intersect $[-\text{fcmin}, \text{fcmin}]$:

$$\mathcal{G}_2 = \{g \in \mathcal{G}_1 : (\cup_{t=1,2,3} I_t^g(G_1\alpha/G) \cap [-\text{fcmin}, \text{fcmin}]) = \varnothing\}, \quad (5)$$

where $\varnothing$ denotes the empty set. It follows from [10, Section 3.1] that the set $\mathcal{G}_2$ has FDR less than or equal to α at MAD level fcmin.

3.2. Inverse screening procedure: FDR P values

In many practical situations, the experimenter may not be comfortable in specifying a MAD or FDR criterion in advance. In these situations, it is more useful to solve the following "inverse problem:" what is the most stringent pair of criteria (α, fcmin) that would lead to including a particular gene among the positives $\mathcal{G}_2$? For fixed fcmin, the most stringent (minimum) value α for which a gene would fall into $\mathcal{G}_2$ is called the FDR P value. The FDR P value for a gene g_o can be computed by (1) computing the PCER P value sequence $\{p(g)\}_{g=1}^{G}$; (2) arranging the PCER P value sequence in an increasing order $p(g_{(1)}) \le \cdots \le p(g_{(G)})$; (3) finding the minimum value $\alpha = \alpha(g_o)$ for which at least one of the PCER CIs $\{I_t^{g_o}(\alpha)\}_{t=1}^{T}$ does not intersect $[-\text{fcmin}, \text{fcmin}]$; and (4) computing the integer index

$$N(\alpha(g_o)) = \sum_{k=1}^{G} I\left(p(g_{(k)})\frac{k}{G} \le 1 - (1-\alpha(g_o))^T\right), \quad (6)$$

where $I(A) = 1$ if statement A is true and $I(A) = 0$ otherwise; the FDR P value of g_o is then simply $p(g_i)$, where $i = N(\alpha(g_o))$. Repeating this as g_o ranges over $1, \ldots, G$ gives a sequence of FDR P values at MAD level fcmin that can be thresholded to determine the set of positive genes $\mathcal{G}_2$ at any desired FDR level of significance.

4. APPLICATION TO A WILD-TYPE VERSUS KNOCKOUT EXPERIMENT

These experiments were performed to investigate the role of a specific retinal transcription factor Nrl [31] in the development of mouse retina. The retinal samples were taken from four pairs ("biological replicates") of wild-type and knockout (Nrl deficient) mice [32] at three different time points: postnatal day 2 (Pn2), postnatal day 10 (Pn10), and 2 months of age (M2). The samples were then hybridized to a total of twenty-four MGU74Av2 Affymetrix GeneChips. The log base 2 probe responses were extracted from Affymetrix GeneChips using the robust microarray analysis (RMA) package [33]. We denote the measured wild-type and knockout responses by $W_{t,m}(g)$ and $K_{t,m}(g)$, where $m = 1, \ldots, M$, $t = 1, \ldots, T$, and $g = 1, \ldots, G$ are microarray replicate, time, and gene probe location on the microarray, respectively. For this experiment, $G = 12421$, $M = 4$, and $T = 3$. To construct CIs on foldchanges, we define the vector of paired t-test statistics:

$$\underline{\hat{\xi}}(g) = \left[\frac{|\overline{W_1}(g) - \overline{K_1}(g)|}{s_1(g)/\sqrt{M/2}}, \frac{|\overline{W_2}(g) - \overline{K_2}(g)|}{s_2(g)/\sqrt{M/2}}, \frac{|\overline{W_3}(g) - \overline{K_3}(g)|}{s_3(g)/\sqrt{M/2}}\right], \quad (7)$$

where $g = 1, \ldots, G$. Here, $\overline{W_t}(g) = M^{-1}\sum_{m=1}^{M} W_{t,m}(g)$ and $\overline{K_t}(g) = M^{-1}\sum_{m=1}^{M} K_{t,m}(g)$ denote the sample mean of the M replicates at time t for wild-type and knockout treatments, respectively, and

$$s_t^2(g) = (2(M-1))^{-1}\left(\sum_{m=1}^{M}(W_{t,m}(g) - \overline{W_t}(g))^2 + \sum_{m=1}^{M}(K_{t,m}(g) - \overline{K_t}(g))^2\right) \quad (8)$$

denotes the pooled sample variance at time t.

Table 2: Two stage FDR-CI algorithm for screening genes from the knockout versus wild-type experiment.

Stage 1	Compute and sort PCER P values according to (9) Select gene indices $\mathcal{G}_1$ according to (3)
Stage 2	Construct simultaneous PCER CIs using (10) Select gene indices $\mathcal{G}_2$ according to (5)

For Stage 1 of the screening procedure, we consider the simple and standard (see [26]) simultaneous test of (1) at MAD level fcmin $= 0$: "decide $H_1(g)$ if $\max_{t=1,2,3}(|\overline{W_t}(g) - \overline{K_t}(g)|/s_t(g)/\sqrt{M/2}) >$ fcmin." Under the large M approximation that the paired t test statistic has a Student t distribution [34], and assuming time independence of cells in the two-way layout of Table 1, we can easily compute both the PCER P value for this test:

$$p(g) = 1 - \left[2\mathcal{T}_{2(M-1)}(\hat{\xi}(g)) - 1\right]^3, \tag{9}$$

and simultaneous $(1-\alpha)\times 100\%$ CIs, $I_1^g(\alpha)$, $I_2^g(\alpha)$, $I_3^g(\alpha)$, for the temporal foldchanges $\{\mu_t(g) - \eta_t(g)\}_{t=1,2,3}$ of gene g:

$$\begin{aligned} &\overline{W_t}(g) - \overline{K_t}(g) - \frac{s_t(g)}{\sqrt{M/2}\mathcal{T}^{-1}_{2(M-1)}}\left(1-\frac{\alpha}{2}\right) \\ &\quad\le \mu_t(g) - \eta_t(g) \\ &\quad\le \overline{W_t}(g) - \overline{K_t}(g) + \frac{s_t(g)}{\sqrt{M/2}\mathcal{T}^{-1}_{2(M-1)}}\left(1-\frac{\alpha}{2}\right), \end{aligned} \tag{10}$$

$t = 1, 2, 3$. In the above inequality, $\mathcal{T}_\nu : \mathbb{R} \mapsto [0,1]$ denotes the Student t cumulative distribution function with ν degrees of freedom and $\mathcal{T}_\nu^{-1}$ denotes its functional inverse, that is, the Student t quantile function.

With the above expressions, we can find the set $\mathcal{G}_1$ of gene indices which pass Stage 1 FDR screening by substituting the sorted PCER P values (9) into the step-down algorithm (3). Stage 2 of screening selects gene indices according to the FDR-CIs from (5). This direct two-stage screening stage procedure is summarized in Table 2. Alternatively, the inverse procedure of Section 3.2 can be implemented using (9) and the explicit expression for the $\alpha(g)$ sequence

$$\alpha(g) = 2\left[1 - \mathcal{T}_{2(M-1)}\left(\frac{\max|\overline{W_t}(g) - \overline{K_t}(g)| - \text{fcmin}}{s_t(g)/\sqrt{M/2}}\right)\right], \tag{11}$$

where $g = 1, \ldots, G$.

4.1. Experimental results

Figures 3 and 4 illustrate the direct and inverse implementations of the FDR-CI screening procedure. In Figure 3, the direct screening procedure is constrained by MAD and FDR criteria fcmin $= 2.0$ and $\alpha = 0.2$, respectively. As there are ($T = 3$)-time points and $G = 12\,421$ genes, there are $GT = 37\,263$ parameters for which FDR-CIs are constructed. A gene passes the screening if at least one of the three time instants has an FDR-CI that does not intersect the interval $[-\text{fcmin}, \text{fcmin}]$. The test is implemented by defining two rank orderings of the FDR-CIs of the genes according to (1) the FDR-CI with minimum upper boundary over the three time points; and (2) the FDR-CI with maximum lower boundary over the time points. Figures 3a and 3b show relevant segments of these two ordered sequences of CIs. Screening all genes with maximum lower endpoints > fcmin and minimum upper endpoints < −fcmin generates the set of declared positive genes $\mathcal{G}_2$.

Figure 4 illustrates the inverse procedure specified in Section 3.2 for screening differentially expressed genes. First, the FDR P values are computed for each gene at several MAD levels of interest. For each MAD level fcmin, we plot the ordered FDR P values. These can be plotted on the same gene index axis since the induced gene ordering is independent of MAD level. FDR P value curves for four different levels of fcmin are illustrated in Figure 4. The figure also illustrates how for FDR and MAD constraints $\alpha = 0.2$ and fcmin $= 0.32$, respectively, the G_2 positive responses $\mathcal{G}_2$ can be extracted from the FDR P value curve by thresholding. Notice that for fixed α, the size G_2 decreases rapidly as the MAD criterion becomes more stringent, that is, as fcmin increases.

Figure 5 shows nine of the top ranked (in FDR P value) differentially expressed gene profiles in (log base 2 scale) among the 59 genes selected by either the direct or inverse implementations of the FDR-CI screening procedure. In the figure, the level of significance constraint is FDR $\le \alpha = 0.2$ and the minimum foldchange constraint is MAD > fcmin $= 1.0$.

In Table 3, we compare the performance of the proposed screening algorithm, labeled "Two-stage FDR-CI," to two other algorithms, called "Thresholded FDR" and "Thresholded RMA." All three algorithms aim to control MAD at a level of fcmin $= 1.0$ (log base 2). The "Two-stage FDR-CI" and "Thresholded FDR" algorithms aim to control FDR at a level of $\alpha = 0.2$ in addition to MAD. Both of these latter algorithms were implemented as two-stage algorithms with common Stage 1, which is to select the gene responses $g \in \mathcal{G}_1$ that pass the paired-t test of hypotheses (1) with fcmin $= 0$ at a FDR level of 20%. The second stage of the "Two-stage FDR-CI" algorithm selects $\mathcal{G}_2$ as a subset of $\mathcal{G}_1$ at the prescribed FDR-CI level of 20%. Stage 2 of the "Thresholded FDR" algorithm simply selects the subset of genes $g \in \mathcal{G}_1$ having at least one sample mean foldchange exceeding fcmin $= 1.0$, that is, it implements the following filter:

$$\max_{t=1,2,3} |\overline{W_t}(g) - \overline{K_t}(g)| > 1.0 \tag{12}$$

on probes $g \in \mathcal{G}_1$. The single-stage "Thresholded RMA" algorithm, a nonstatistical method commonly used in many microarray studies, implements the filter (12) on the responses of each g in the original set of 12 421 genes as indicated in Figure 2.

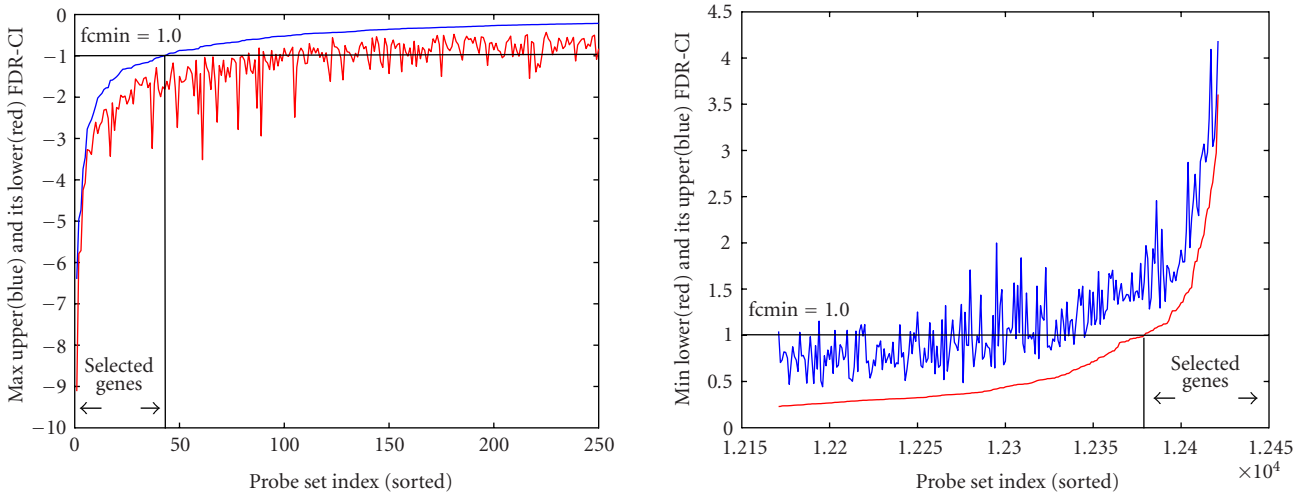

FIGURE 3: Segments of upper and lower curves specifying the 80% FDR-CI on the foldchanges $\{\mu_t(g) - \eta_t(g)\}_{t=1,2,3}$ for the knockout versus wild-type study. Upper and lower curves in each figure sweep out FDR-CI upper and lower boundaries on foldchange for all genes (indexed by probe set number). In (a) the curves sweep out the sequence of FDR-CIs indexed in an increasing order of the (maximum) lower CI boundary and in (b) the ordering is in an increasing order of the (minimum) upper CI boundary. Only those genes whose three FDR-CIs do not intersect $[-\text{fcmin}, \text{fcmin}]$ are selected by the second stage of screening. When the MAD foldchange criterion is fcmin = 2.0 (1.0 in log base 2), these genes are obtained by thresholding the curves as indicated.

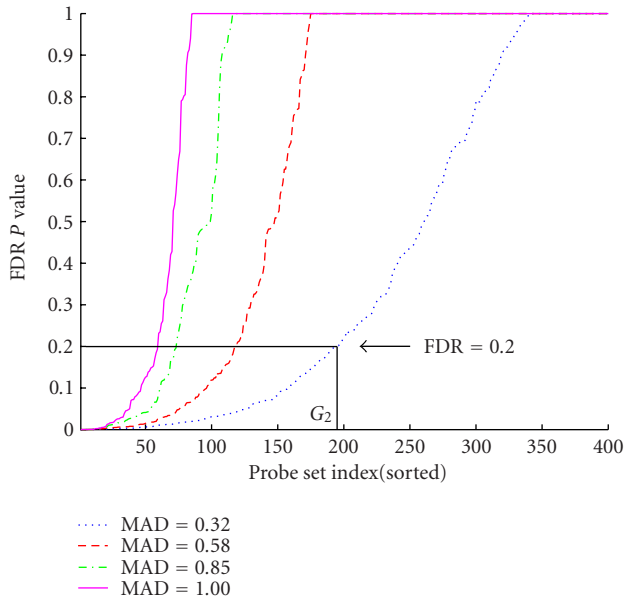

FIGURE 4: Plots of FDR P value curves over sorted list of gene indices for four values of the MAD criterion: fcmin = 0.32, 0.58, 0.85, 1.0 (log base 2) corresponding to wild-type/knockout MAD ratios of 1.25, 1.5, 1.8, and 2.0, respectively. Constraints FDR $\leq$ 0.2 and foldchange > 0.32 determine a set $\mathcal{G}_2$ of G_2 differentially expressed genes by thresholding the corresponding curve as indicated.

The number of screened and discovered genes for the three algorithms is indicated in the first two columns of Table 3. The maximum and median of the FDR P values of the discovered genes is indicated in the third and fourth columns for each algorithm. The last column indicates the maximum length of the FDR-CIs on foldchanges of the discovered genes. We conclude from Table 3 that the proposed "Two-stage FDR-CI" algorithm outperforms the other algorithms in terms of (1) maintaining the FDR requirement that false positives do not exceed 20% (column 4); (2) ensuring a substantially lower median FDR P value than the others (column 5); (3) discovering genes that have tighter (on the average) CIs on biologically significant (> 1.0) foldchange (column 6).

5. CONCLUSION

Signal processing for analysis of DNA microarrays for gene expression profiling is a rapidly growing area and there are enough challenges to keep the community busy for years. It is essential that signal processing methods be relevant and capture the biological aims of the experimenter. To this aim, in this paper, we developed a flexible multicriteria approach to gene selection and ranking for screening differentially expressed gene profiles. The proposed criteria capture the gene expression differences at multiple time points, account for minimum acceptable foldchange constraints, and control false discovery rate. In many cases, biological significance requires minimum hybridization levels, for example, as implemented by Affymetrix in their "absent calls" for weakly expressed genes. This can be easily captured by incorporating an addition criterion, the minimum acceptable mean expression level, into our multicriteria approach.

Table 3: Performance comparison of three algorithms for selecting genes with magnitude (log base 2) foldchange > 1.0. Thresholded RMA and Thresholded FDR are significantly worse in terms of statistical significance (P value) than the proposed Two-stage FDR-CI algorithm (columns 4 and 5). Furthermore, the average length of the CIs on foldchanges of the discovered genes are shorter for the Two-Stage FDR-CI algorithm than for the other algorithms (column 6).

	# Screened	# Discovered	Max(Pv)	Median(Pv)	Avg(FDR-CI length)
Thresholded RMA	12,421	159	1.0	0.80	1.52
Thresholded FDR	303	127	1.0	0.31	1.17
Two-stage FDR-CI	303	59	0.19	0.02	1.09

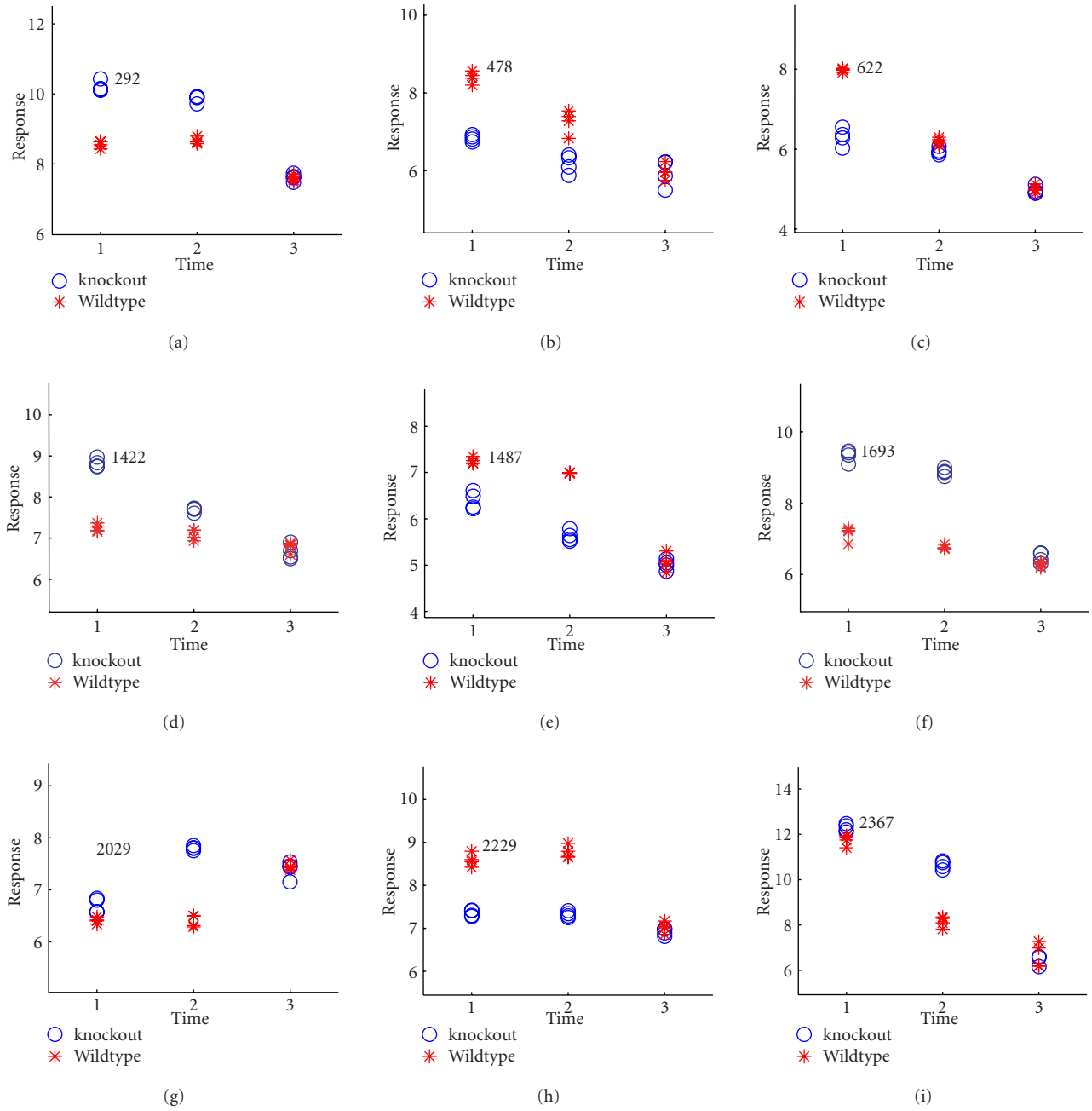

Figure 5: Gene profiles of nine of the differentially expressed genes discovered using the proposed two-stage FDR-CI procedure with constraints on level of significance $\alpha = 0.2$ and minimum foldchange fcmin = 1.0. Knockout "$\circ$" and Wildtype "$*$" are as indicated, and the numbers on each panel denote gene indices (related to the positions of the gene probes on the microarray).

ACKNOWLEDGMENT

The authors would like to thank R. Farjo for stimulating discussions and suggestions on the gene selection techniques presented in this paper. The research was supported by grants from the National Institutes of Health (EY11115 including administrative supplements), the Elmer and Sylvia Sramek Foundation, and The Foundation Fighting Blindness.

REFERENCES

[1] J. Watson and A. Berry, *DNA: The Secret of Life*, Alfred A. Knopf, NY, USA, 2003.

[2] F. C. Collins, M. Morgan, and A. Patrinos, "The Human Genome Project: lessons from large-scale biology," *Science*, vol. 300, no. 5617, pp. 286–290, 2003.

[3] P. O. Brown and D. Botstein, "Exploring the new world of the genome with DNA microarrays," *Nature Genetics*, vol. 21, suppl. 1, pp. 33–37, 1999.

[4] D. Bassett, M. B. Eisen, and M. Boguski, "Gene expression informatics—it's all in your mine," *Nature Genetics*, vol. 21, suppl. 1, pp. 51–55, 1999.

[5] Affymetrix, *NetAffx User's Guide*, 2000, http://www.netaffx.com/site/sitemap.jsp.

[6] National Human Genome Research Institute (NHGRI), *cDNA Microarrays*, 2001, http://www.nhgri.nih.gov/.

[7] T. Hastie, R. Tibshirani, M. Eisen, et al., "Gene shaving: a new class of clustering methods for expression arrays," Tech. Rep., Stanford University, Stanford, Callif, USA, 2000.

[8] A. A. Alizadeh, M. B. Eisen, R. E. Davis, et al., "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[9] M. Brown, W. N. Grundy, D. Lin, et al., "Knowledge-based analysis of microarray gene expression data by using support vector machines," *Proceedings of National Academy of Sciences*, vol. 97, no. 1, pp. 262–267, 2000.

[10] Y. Benjamini and D. Yekutieli, "False discovery rate adjusted confidence intervals for selected parameters," submitted to Journal of the American Statistical Association.

[11] J. Liu and H. Iba, "Selecting informative genes using a multiobjective evolutionary algorithm," in *Proc. Congress on Evolutionary Computation*, pp. 297–302, Honolulu, Hawaii, USA, May 2002.

[12] G. Fleury, A. O. Hero, S. Yoshida, T. Carter, C. Barlow, and A. Swaroop, "Clustering gene expression signals from retinal microarray data," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 4, pp. 4024–4027, Orlando, Fla, USA, May 2002.

[13] A. Hero and G. Fleury, "Pareto-optimal methods for gene analysis," to appear in Journal of VLSI Signal Processing, Special Issue on Genomic Signal Processing.

[14] G. Fleury and A. O. Hero, "Gene discovery using Pareto depth sampling distributions," to appear in Journal of Franklin Institute.

[15] A. Reiner, D. Yekutieli, and Y. Benjamini, "Identifying differentially expressed genes using false discovery rate controlling procedures," *Bioinformatics*, vol. 19, no. 3, pp. 368–375, 2003.

[16] R. L. Miller, A. Galecki, and R. J. Shmookler-Reis, "Interpretation, design, and analysis of gene array expression experiments," *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 56, no. 2, pp. B52–B57, 2001.

[17] D. B. Allison and C. S. Coffey, "Two stage testing in microarray analysis: what is gained?," *Journal of Gerontology: Biological Sciences*, vol. 57, no. 5, pp. B189–B192, 2002.

[18] Y. Benjamini, A. Krieger, and D. Yekutieli, "Adaptive linear step-up false discovery rate controlling procedures," Tech. Rep. Research Paper 01-03, Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel, 2001.

[19] T. P. Speed, *Statistical Analysis of Gene Expression Microarray Data*, Chapman & Hall/CRC Press, Boca Raton, Fla, USA, 2003.

[20] J. Watson, M. Gilman, J. Witkowski, and M. Zoller, *Recombinant DNA*, W. H. Freeman, NY, USA, 1992.

[21] M. Hollander and D. A. Wolfe, *Nonparametric Statistical Methods*, John Wiley & Sons, NY, USA, 2nd edition, 1999.

[22] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco, Calif, USA, 1977.

[23] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I*, John Wiley & Sons, NY, USA, 1968.

[24] S. Dudoit, J. P. Shaffer, and J. C. Boldrick, "Multiple hypothesis testing in microarray experiments," Tech. Rep. Working Paper 110, Berkeley Division of Biostatistics Working Paper Series, 2002, http://www.bepress.com/ucbbiostat/paper110.

[25] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society*, vol. 57, no. 1, pp. 289–300, 1995.

[26] R. G. Miller, *Simultaneous Statistical Inference*, Springer-Verlag, NY, USA, 1981.

[27] J. D. Storey and R. Tibshirani, "Estimating the positive false discovery rates under dependence, with applications to DNA microarrays," Tech. Rep. 2001-28, Department of Statistics, Stanford University, Stanford, Callif, USA, 2001.

[28] C. R. Genovese, N. A. Lazar, and T. E. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, no. 4, pp. 870–878, 2002.

[29] P. Westfall and S. Young, *Resampling-Based Multiple Testing*, John Wiley & Sons, NY, USA, 1993.

[30] V. S. Williams, L. V. Jones, and J. W. Tukey, "Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement," *Journal of Educational and Behavioral Statistics*, vol. 24, no. 1, pp. 42–69, 1999.

[31] A. Swaroop, J. Xu, H. Pawar, A. Jackson, C. Skolnick, and N. Agarwal, "A conserved retina-specific gene encodes a basic motif/leucine zipper domain," *Proceedings of National Academy of Sciences (USA)*, vol. 89, no. 1, pp. 266–270, 1992.

[32] A. Mears, M. Kondo, P. Swain, et al., "Nrl is required for rod photoreceptor development," *Nature Genetics*, vol. 29, no. 4, pp. 447–452, 2001.

[33] R. Irizarry, B. Hobbs, F. Collin, et al., "Exploration, normalization, and summaries of high density oligonucleotide array probe level data," *Biostatistics*, vol. 4, no. 2, pp. 249–264, 2003.

[34] D. F. Morrison, *Multivariate Statistical Methods*, McGraw-Hill, NY, USA, 1967.

Alfred O. Hero received his Ph.D. degree from Princeton University in 1984. Since then, he has been a Professor with the University of Michigan, Ann Arbor, where he has appointments in the Department of Electrical Engineering and Computer Science, the Department of Biomedical Engineering, and the Department of Statistics. Alfred Hero is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE). He has received the 1998 IEEE Signal Processing Society Meritorious Service Award, the 1998 IEEE Signal Processing Society Best Paper Award, and the IEEE Third Millenium Medal. His interests are in estimation and detection, statistical communications, bioinformatics, signal processing, and image processing.

Gilles Fleury was born in Bordeaux, France in 1968. He received the M.S. degree in electrical engineering from Ecole Supérieure d'Electricite (SUPELEC) in 1990, the Ph.D. degree in signal processing from the Université de Paris-Sud, Orsay, France, in 1994, and his Habilitation à diriger la Recherche (HDR) in 2003. He is presently a Professor within the Department of Measurement of SUPELEC. He has worked in the areas of inverse problems and optimal design. His current research interests include bioinformatics, optimal nonlinear modeling, and nonuniform sampling.

Alan J. Mears received his B.S. degree (Honors) from Leeds University, U.K. in 1989 and his Ph.D. degree from the University of Alberta, Canada in 1995, both in Genetics. He was a Research Investigator at the University of Michigan from 1999 to 2003 and is currently an Assistant Professor in Ophthalmology at the University of Ottawa in Canada. His research interests include the genetics of retinal disease and the transcriptional regulation of mammalian retinal development. Alan Mears has been a member of the American Association for the Advancement of Science from 1995 to 1997, American Society of Human Genetics from 1995 to 1998, and the Association for Research in Vision and Ophthalmology from 1996 till now.

Anand Swaroop received his Ph.D. degree in biochemistry from the Indian Institute of Science in 1982 and pursued his postdoctoral research in Genetics at Yale University, initially working on Drosophila and then Human Genetics. He joined the faculty of the Department of Ophthalmology and Visual Sciences and the Department Human Genetics at the University of Michigan Medical School in July 1990. He was promoted to a Full Professor in 2000 and currently holds the appointment as Harold F. Falls Collegiate Professor. He is Director/Coordinator of the Center for Retinal and Macular Degeneration and Director of the Sensory Gene Microarray Node. His research focuses on molecular genetics of retinal and macular diseases, retinal differentiation and aging, and expression profiling. He has published over 100 manuscript. His work is supported by grants from the National Institutes of Health, The Foundation Fighting Blindness, Macula Vision Research Foundation, and Elmer and Sylvia Sramek Charitable Foundation. In 1997, Anand Swaroop received the Lew R. Wasserman Merit Award from the Research to Prevent Blindness Foundation. He is currently a member on the editorial boards of Investigative Ophthalmology and Visual Science and Molecular Vision. He reviews manuscripts and grants for several journals, international foundations, and agencies. He is also a regular member of the BDPE study section of NIH.