Open Access Bioinformatics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Balancing false discovery and false negative rates in selection of differentially expressed genes in microarrays

This article was published in the following Dove Press journal:
Open Access Bioinformatics
18 February 2010
Number of times this article has been viewed

Byung S Park[1,2,3]
Motomi Mori[1,2,3]

[1]Division of Biostatistics, Department of Public Health and Preventive Medicine, [2]Biostatistics Shared Resource, Knight Cancer Institute, [3]Biostatistics and Design Program, Oregon Clinical and Translational Science Institute, Oregon Health and Science University, Portland, OR, USA

Abstract: Genome-wide mRNA expression profiling using microarrays is widely available today, yet analysis and interpretation of the resulting high dimensional data continue to be a challenge for biomedical scientists. In a typical microarray experiment, the number of biological samples is quite modest compared with the number of genes on a microarray, and a probability of falsely declaring differential expression is unacceptably high without any adjustment for multiple comparisons. However, a stringent multiple comparison procedure can lead to an unacceptably high false negative rate, potentially missing a large fraction of truly differentially expressed genes. In this paper we propose a new "balancing factor score" (*BFS*) method for identifying a set of differentially expressed genes. The *BFS* method combines a traditional *P* value criterion with any other informative factors (referred to as balancing factors) that may help to identify differentially expressed genes. We evaluate the performance of the *BFS* method when the observed fold change is used as a balancing factor in a simulation study and show that the *BFS* method can substantially reduce the false negative rate while maintaining a reasonable false discovery rate.

Keywords: balancing factor score method, microarrays, multiple comparisons, false discovery rate, false negative rate

Introduction

High-throughput genomic technologies such as genome-wide mRNA expression arrays (microarrays) and single nucleotide polymorphisms (SNPs) produce thousands of measurements for each sample. A common goal in a microarray experiment is to identify genes that are differentially expressed between two or more experimental conditions.[1] A typical statistical approach is to perform an appropriate statistical significance test for each gene (eg, gene-by-gene unpaired *t*-test or two-sample Wilcoxon test) followed by a multiple comparison procedure (MCP), controlling either for the overall rate of false positives or of false discoveries. Many different MCP methods have been proposed for high dimensional data sets, and a comprehensive review of these methods can be found in Dudoit and colleagues.[2] The false discovery rate (*FDR*) method[3] is perhaps the most popularly used MCP method in microarrays.[4] Storey and colleagues proposed a *Q* value as an extension of *FDR*.[5] The *Q* value is similar to the well known *P* value. The *Q* value is a measure of significance in terms of the *FDR*, while the *P* value is a measure of significance in terms of the false positive rate (FPR). The *Q* value is implemented in the Significance Analysis of Microarrays (SAM), a popularly used free software for the microarray analysis (http://www-stat.stanford.edu/~tibs/SAM/) developed by Tibshirani's group at Stanford University.

Correspondence: Motomi Mori
Knight Cancer Institute, CR145, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, OR 97239, USA
Tel +1 503 418 1555
Fax +1 503 418 0125
Email morim@ohsu.edu

Until recently MCP methods focused mostly on controlling the number of false positives, but when a difference between two groups is of a modest magnitude, the existing methods can result in a poor power to detect genes that are truly differentially expressed.[6] To improve a power of selecting truly differentially expressed genes, some authors proposed a hybrid approach such as evaluating *FDR* in conjunction with secondary measures, for example, the *FNR* or fold change. Norris and colleagues proposed a "balance probability analysis"[6] which provides a biologist with a method for interpreting results in the context of the total number of truly differentially expressed genes, *FDRs* and *FNRs* for the list of genes reaching any significance threshold. Rosenfeld suggested a modification of the classical *t*-test,[7] which is specifically designed to enhance the sensitivity of detection of differentially expressed genes. Rosenfeld proposed a test statistic called a "bio-weight", which is defined as the product of the absolute fold change and the negative decimal logarithm of the *t*-test *P* value. In this paper, we propose a "balancing factor score" (*BFS*) method which combines the *P* values from the conventional significance test and any other informative factors that may increase the sensitivity of identifying differentially expressed genes (eg, fold change or *P* value from another microarray experiment). Under certain circumstances, the proposed *BFS* method can substantially improve the detection of truly differentially expressed genes by reducing the *FNR* while keeping a reasonable level of *FDR*.

Material and methods

Motivating example

The *BFS* method was motivated by an actual microarray experiment conducted by a biomedical scientist from the Oregon National Primate Research Center at the Oregon Health and Science University. The experiment examined the effect of age and gender on gene expression of hippocampus among rhesus macaques. The experiment was performed as a 2×3 factorial design with a total of six groups defined by two gender levels (male, female) and three age levels (adult, middle-aged, old). There were approximately four animals per group with a total of 24 animals. The microarray experiment was carried out by the OHSU Gene Microarray Shared Resource Affymetrix Microarray Core using the Rhesus Macaque GeneChip. After appropriate pre-processing and normalization, a two-way analysis of variance was fitted to each gene separately, and the *FDR* adjustment was applied to each factor (gender, age, gender by age interaction) independently. After the *FDR* adjustment, there were no genes that met a *FDR P* value < 0.05 for the interaction effect and only a handful of genes that met the *FDR P* value < 0.05 for gender and age main effects. Because the cost of obtaining additional samples was prohibitive, we explored alternative methods of identifying a set of differentially expressed genes and for assessing the *FNR* of the *FDR*-adjusted *P* value criterion. Although our motivation came from one particular study, we expect that similar situations arise elsewhere. In some experiments, the cost of obtaining biologic samples is very high, and it is not always possible to increase the sample size. In addition, many microarray experiments are performed primarily as a screening procedure with identified genes being evaluated by another experimental platform (such as quantitative real time transcription – polymerase chain reaction [qRT-PCR]). In those circumstances, it may be acceptable to have a slightly higher *FDR* in exchange for a larger pool of potentially differentially expressed genes. These considerations prompted us to investigate a hybrid approach which is based on both *P* values and other factors that may help identify a set of differentially expressed genes.

Notations and definitions

To simplify, we will focus on a problem of identifying a set of differentially expressed genes between two groups (eg, normal vs disease), where one may apply a simple unpaired *t*-test or Wilcoxon rank sum test to each gene individually. Table 1 summarizes the resulting test outcomes and the definition of key concepts used in this paper. Specifically we focus on the *FNR* and *FDR*.

Illustration using simulated data

We simulated a microarray data set with the following conditions: (1) 52,865 genes; (2) 30 samples (15 samples per group) and (3) 10% of genes ($N_1 = 5{,}287$) truly differentially expressed with the mean difference of 0.748 standard deviation unit [ie, *t*-statistics of 2.05 ($t_{0.025,\ 28}$)]. In each gene, the *FDR P* value threshold of 5% is used to identify differentially expressed genes. As expected, the *FDR* is approximately 5%,

Table 1 Possible outcomes from N significant tests and definitions of FNR, FPR, FDR, and FNDR

	Declare *DE*	**Declare *non-DE***	**Total**
True DE	*TP*	*FN*	N_1
Truly not DE	*FP*	*TN*	N_2
Total	S_1	S_2	*N*

Notes: Sensitivity is 1-FNR, while specificity is 1-FPR.
Abbreviations: TP, true positive; FN, false negative; FP, false positive; TN, true negative; FNR, false negative rate = FN/N_1; FPR, false positive rate = FP/N_2; FDR, false discovery rate = FP/S_1; FNDR, false non-discovery rate = FN/S_2.

that is, 69 genes are identified as differentially expressed with only four being false positives. Surprisingly, however, 5,222 genes are false negative, ie, falsely declared as not differentially expressed, and the *FNR* is 0.99. The *FDR*-adjusted *P* value criterion sacrifices too many false negatives in order to control the number of false positives. If the raw *P* values without any MCP adjustments are used, the *FNR* is 0.49, but the *FDR* is 0.47, and thus too high to be acceptable, leading to costly follow-up experiments with very few conformed genes. Figure 1 displays the volcano plot, ie, a scatter plot of observed mean differences in the x axis vs -$\log_{10}$ (*FDR P* values) in the y axis. The red dots represent true differentially expressed genes, and the vertical line represents the *FDR*-adjusted *P* value threshold value of 0.05. In a typical microarray study, genes above the vertical line are declared as differentially expressed genes. As seen in Figure 1, most of the false negative genes are associated with a larger fold change value. An alternative method that incorporates the fold change information may lead to a better detection of differentially expressed genes, and may balance between the *FDR* and *FNR*.

BFS method

A balancing factor is any informative factor that may help identify differentially expressed genes in addition to the standard p values from statistical significance testing. A *BFS* is defined as a quantitative measure of a balancing factor ranging from 0 (strong differential expression) to 1 (no differential expression), analogous to the p value. A *BFS* can be created by first transforming it to the standardized *z* score and computing the standard normal probability. For example, consider a fold change (FC) defined as: sign(Δ)*$2^{|\Delta|}$, where Δ is a mean difference of normalized log signal values between the experimental and control groups. A fold change of −2 implies twofold downregulation, while a fold change of 2 means twofold upregulation in the experimental group compared with the control group. A fold change can be used as a balancing factor, and the *BFS* can be created by using the standardized normal probability as:

$$BFS = P\left(Z > \left| \frac{|FC| - \overline{|FC|}}{SD_{|FC|}} \right| \right).$$

For each gene, a *BFS* statistic (T_{BFS}) can be computed as:

$$T_{BFS} = \lambda_0 p + \sum_{i=1}^{k} \lambda_i BFS_i$$

where *p* is the unadjusted *P* value for the conventional test, λ_0 is the weight for the unadjusted *P* value, and λ_i is a weight for the *ith* balancing factor so that $\lambda_0 + \sum_{i=1}^{k} \lambda_i = 1$. T_{BFS} is always between 0 and 1, and one can consider it as a modified *P* value. We will then apply the *FDR* method to T_{BFS} to calculate the *FDR*-adjusted T_{BFS} value, denoted by $FDR(T_{BFS})$. A gene is then declared as differentially expressed

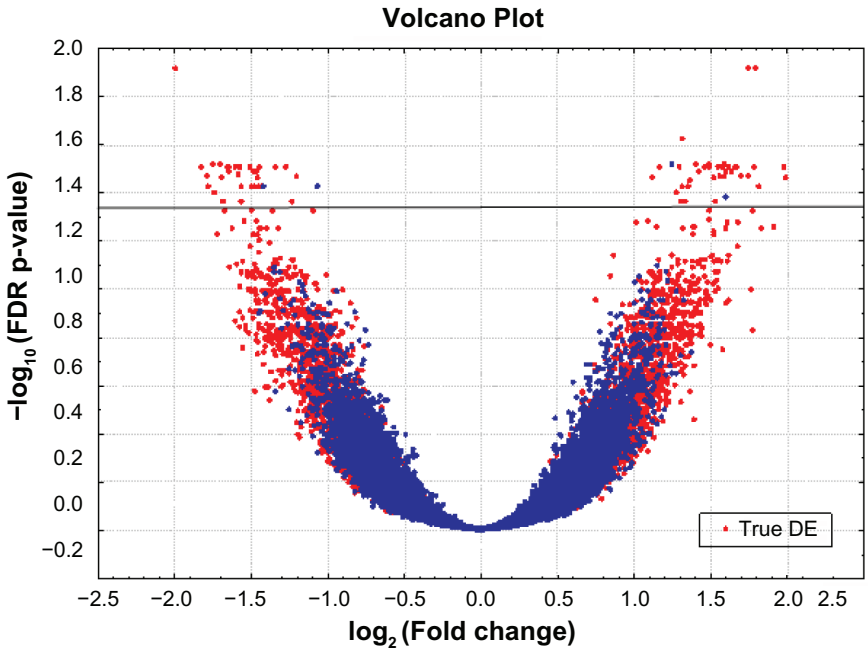

Figure 1 Volcano plot: scatter plot of log (fold change) in the x axis vs $-\log_{10}$ (FDR *P* value) in the y axis. The red dots represent true differentially expressed genes, and the vertical line represents the FDR-adjusted *P* value threshold value of 0.05.

if $FDR(T_{BFS})$ is less than α, specified cut-off value (usually 0.05), and otherwise declared not differentially expressed. Note that if $\lambda_0 = 1$ and $\lambda_1 = 0$ for all $i > 0$, then this is equivalent to the *FDR*-adjusted *P* value criterion.

A choice of balancing factors is not unique, but there are obvious candidates including a fold change, *P* value from the previous microarray experiment, *P* value from the published study or publicly available data, or *P* value representing biologic relevance (eg, a likelihood of a particular gene belonging to a pathway of interest).

If BFS_i follows beta distributions for the *ith* test, λ_0 is the probability of there being no difference, and λ_1 is the *P* value of being truly differentially expressed in *ith* test, then the *BFS* method has the same structural form as the mixture model proposed by Allison and colleagues.[8] The *BFS* method can be extended to identify a set of differentially expressed genes when the microarray experiment is performed at multiple times or using different platforms such as DNA methylation. In a special case when a balancing factor is a *P* value from another, independently conducted experiment, the *BFS* test statistic can be expressed as: $T_{BFS} = \lambda_0 p_0 + \lambda_1 p_1$, where p_0 is the *P* value from the current experiment and p_1 is the *P* value from another study. In this case, we can determine the actual distribution and cut-off value for the *BFS* method that provides a specified error rate (see Appendix 1).

Results

A simulation study is conducted to evaluate the performance of the *BFS* method when a fold change is considered as a balancing factor. A total of 500 simulated data sets are generated under the following conditions: (1) Two groups; (2) 15 samples per group; (3) 10,000 genes (*N*) on each microarray; (4) the proportion of truly differentially expressed genes (N_1) is either 5% or 10%; (5) among truly differentially expressed genes, the fold change varies from 1.5, 2.0 and 3.0, and (6) the expression values are normally distributed. The following procedure is applied to each simulated data set:

1. Define $BFS = P\left(Z > \left|\frac{|FC| - \overline{|FC|}}{SD_{|FC|}}\right|\right)$, where $SD_{|FC|}$ is the standard deviation of the absolute fold change values
2. Calculate the *BFS* using $T_{BFS} = \lambda_0 p + \lambda_1 BFS$, where *p* is a *P* value from two-sample *t*-test
3. Calculate Benjamini Hochberg *FDR*-adjusted T_{BFS}, denoted as $FDR(T_{BFS})$
4. If $FDR(T_{BFS})$ is smaller than the pre-determined threshold value, typically 0.05, then declare the gene as differentially expressed.

Simulation results are summarized in terms of empirical *FDR* and *FNR*. We evaluate *FDR* (T_{BFS}) for various values of λ_1 between 0 to 1. If $\lambda_1 = 0$, this is equivalent to the standard *FDR*-adjusted *P* value criterion. The performance of the *BFS* method is shown as a function of the weight λ_1 and is compared with the *FDR*-adjusted *P* value criterion as well as the criterion based on unadjusted *P* values. Table 2 shows the performance of the *BFS* method when the proportion of truly differentially expressed genes is 10% with a relatively large fold change (FC = 3.0). The *FNR* of the *FDR*-adjusted *P* value criterion is 0.12, whereas the *FNR* of the *BFS* method is consistently smaller for all λ_1 (0.07–0.11) while maintaining a comparable *FDR* level. In this simulation scenario, regardless of a choice of λ_1, the *BFS* method not only identifies more true positive genes but also identifies fewer false positives than typically used selection methods. Table 3 shows the simulation results for the other extreme when the proportion of truly differentially expressed

Table 2 Total true DE = 1000 10%, true difference FC = 3.0, cutoff α = 0.05

		Number of genes identified as DE		*False discovery rate FDR*		*False negative rate FNR*	
		Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**
Raw *P* value		1438.60	0.9303	0.31	0.0004	0.01	0.0002
FDR *P* value		925.81	0.6382	0.05	0.0003	0.12	0.0005
BFS λ_1	0.1	927.92	0.6129	0.04	0.0003	0.11	0.0005
	0.2	930.02	0.5859	0.04	0.0003	0.11	0.0005
	0.3	932.58	0.5572	0.04	0.0003	0.10	0.0005
	0.4	935.30	0.5278	0.04	0.0003	0.10	0.0004
	0.5	938.78	0.5052	0.04	0.0003	0.09	0.0004
	0.6	943.02	0.4935	0.03	0.0003	0.09	0.0004
	0.7	947.87	0.4764	0.03	0.0003	0.08	0.0004
	0.8	953.73	0.4533	0.03	0.0003	0.08	0.0004
	0.9	961.66	0.4683	0.04	0.0003	0.07	0.0004

Table 3 Total true DE = 500 5%, true difference FC = 1.5, cutoff $\alpha = 0.05$

		Number of genes identified as DE		*Number of false positives*		*Number of false negatives*	
		Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**
Raw *P* value		646.81	1.0700	476.81	0.9495	329.99	0.4820
FDR *P* value		0.54	0.0454	0.08	0.0148	499.55	0.0384
BFS λ_1	0.1	0.61	0.0504	0.09	0.0152	499.48	0.0433
	0.2	0.73	0.0559	0.11	0.0164	499.37	0.0483
	0.3	0.89	0.0641	0.13	0.0197	499.25	0.0559
	0.4	1.08	0.0763	0.19	0.0250	499.11	0.0641
	0.5	1.43	0.0949	0.27	0.0348	498.84	0.0754
	0.6	1.96	0.1159	0.39	0.0413	498.43	0.0910
	0.7	3.74	0.1937	0.88	0.0713	497.14	0.1445
	0.8	10.32	0.4203	3.12	0.1654	492.80	0.2816
	0.9	84.71	1.8028	42.30	1.0927	457.59	0.7604

genes is small (5%) with a moderate change (FC = 1.5). In this setting, the *FDR*-adjusted *P* value criterion fails to identify true differentially expressed genes. It only detects 0.54 genes on average. If the unadjusted *P* value criterion is used, on average, 647 genes are declared to be differentially expressed, while 476 genes are false positives. Even in this extreme case, the *BFS* method provides a reasonable number of genes (0–85 genes depending on λ_1) that can be pursued further for confirmation. Figures 2a and 2b illustrate the *FDR* (2a) and *FNR* (2b), respectively, as a function of λ_1, when

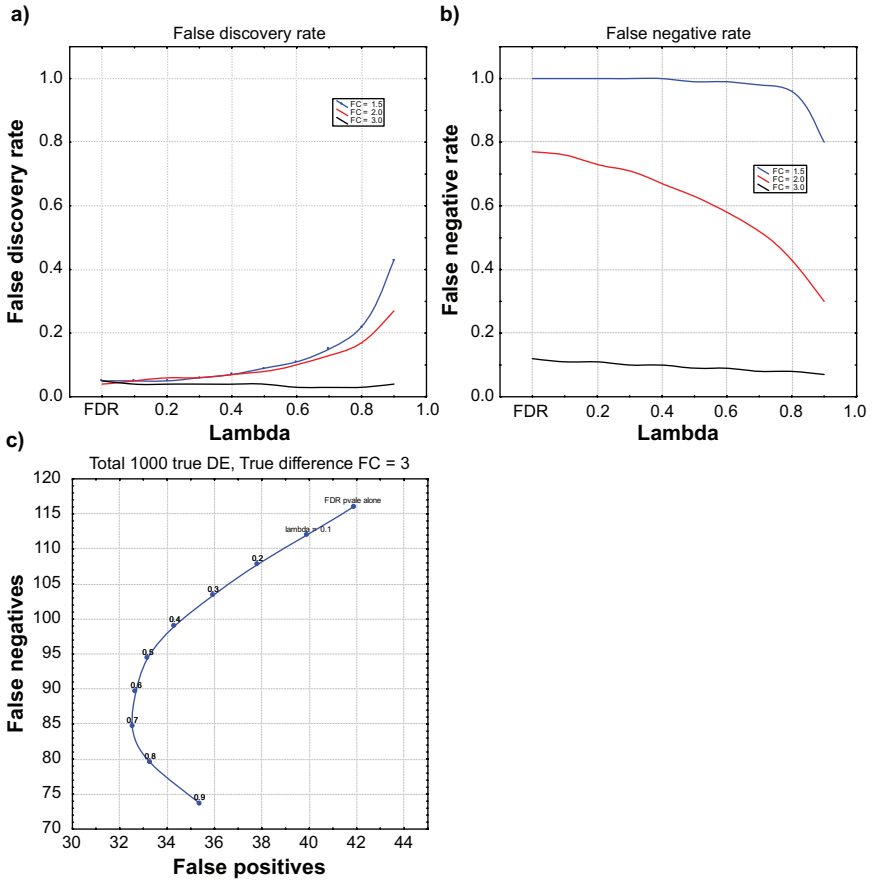

Figure 2 Operating characteristics of the BFS analysis in comparison with FDR alone analysis in terms of FDR and FNR. Figure 2a displays the FDR of various choice of λ_1; 2b) the FNR of various choice of λ_1. 2c) displays the performs of balancing factor score analysis when there are 10% of truly DE genes in the sample and the true change between groups was assumed to be threefold changes. y axis represents number false negatives, and x axis represents number of false positives from the simulation.

10% of genes are truly differentially expressed. The *BFS* method appears to make the most impact on *FNR* when the fold change is moderate (FC = 2.0) as shown in Figure 2b. Figure 2c demonstrates that the *BFS* method is superior not only for reducing false negatives but also false positives when compared with the *FDR*-adjusted *P* value criterion in some conditions. Tables 4 and 5 summarize additional simulations results on *FDR* and *FNR*.

Discussion

Genome-wide mRNA expression profiling using microarrays is widely available in biomedical research. However, the cost of a microarray experiment including sample generation, preparation, assays and chips can be prohibitive, and many microarray experiments are performed using modest number of biologic samples and are often inadequately powered. Therefore it is not surprising that we have to balance our risk of false positives and false negatives when identifying a set of differentially expressed genes. We have shown that the *BFS* method incorporating additional informative factors such as a fold change in addition to *P* values from statistical significance testing can improve the overall detection rate by reducing the number of false negatives while keeping the number of false positives at a reasonable level. The *BFS* method can be extended easily to incorporate other balancing factors, such as *P* values from the previous experiment or another study.

In this paper, we focused on the *FNR* and *FDR*. Genovese and Wasserman introduced the false nondiscovery rate (FNDR)[9] and proposed methods that incorporate both *FDR* and FNDR. The FNDR may be a conceptually better measure because it complements the *FDR*. Genovese and Wasserman suggested using 1-π as a risk function to compare the performance of various multiple testing procedures,[5] where $\pi = 1 - FDR - FNDR$. This is equivalent to the concept of power introduced by Sarkar,[10] reflecting how well a multiple comparison procedure performs in terms of the overall correct decisions. However in most studies, the number of genes that are declared as differentially expressed is much smaller, and the FNDR contributes very little relative to the *FDR*. Hence, the FNDR may not contribute significantly to the risk function.

Table 4 5% True differentially expressed genes

		Cutoff 0.05						Cutoff 0.10					
		FC = 1.5		FC = 2.0		FC = 3.0		FC = 1.5		FC = 2.0		FC = 3.0	
		FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE
Raw		0.74	0.66	0.56	0.25	0.49	0.01	0.80	0.53	0.69	0.15	0.66	0.01
P value		0.0007	0.0010	0.0006	0.0009	0.0005	0.0002	0.0004	0.0010	0.0004	0.0007	0.0003	0.0001
FDR		0.04	1.00	0.05	0.88	0.05	0.18	0.09	1.00	0.09	0.76	0.10	0.11
P value		0.0075	0.0001	0.0013	0.0012	0.0005	0.0010	0.0098	0.0002	0.0013	0.0015	0.0006	0.0007
λ_1	0.1	0.04	1.00	0.05	0.86	0.05	0.17	0.09	1.00	0.11	0.75	0.10	0.10
		0.0076	0.0001	0.0012	0.0013	0.0005	0.0009	0.0096	0.0002	0.0013	0.0015	0.0007	0.0007
	0.2	0.05	1.00	0.06	0.85	0.05	0.16	0.10	1.00	0.12	0.73	0.11	0.09
		0.0079	0.0001	0.0013	0.0013	0.0005	0.0009	0.0096	0.0002	0.0013	0.0015	0.0007	0.0006
	0.3	0.06	1.00	0.07	0.83	0.06	0.15	0.12	1.00	0.14	0.70	0.12	0.08
		0.0089	0.0001	0.0013	0.0013	0.0005	0.0008	0.0098	0.0002	0.0014	0.0015	0.0007	0.0006
	0.4	0.07	1.00	0.08	0.81	0.06	0.13	0.14	0.99	0.17	0.67	0.14	0.07
		0.0098	0.0001	0.0013	0.0013	0.0006	0.0008	0.0097	0.0003	0.0014	0.0015	0.0007	0.0006
	0.5	0.09	1.00	0.10	0.79	0.07	0.12	0.18	0.99	0.21	0.64	0.15	0.06
		0.0098	0.0002	0.0014	0.0013	0.0006	0.0007	0.0097	0.0004	0.0014	0.0015	0.0007	0.0005
	0.6	0.10	1.00	0.13	0.75	0.08	0.10	0.24	0.99	0.26	0.60	0.18	0.05
		0.0096	0.0002	0.0015	0.0013	0.0006	0.0007	0.0092	0.0006	0.0015	0.0014	0.0008	0.0005
	0.7	0.14	0.99	0.19	0.71	0.10	0.08	0.30	0.97	0.36	0.55	0.21	0.04
		0.0097	0.0003	0.0016	0.0013	0.0006	0.0006	0.0070	0.0009	0.0015	0.0013	0.0008	0.0004
	0.8	0.24	0.99	0.29	0.66	0.12	0.06	0.48	0.92	0.54	0.48	0.27	0.03
		0.0092	0.0006	0.0017	0.0012	0.0007	0.0005	0.0039	0.0015	0.0014	0.0012	0.0009	0.0003
	0.9	0.47	0.92	0.58	0.58	0.17	0.04	0.81	0.49	0.80	0.40	0.36	0.02
		0.0039	0.0015	0.0018	0.0012	0.0008	0.0004	0.0004	0.0011	0.0014	0.0012	0.0008	0.0002

Table 5 10% True differentially expressed genes

		Cutoff 0.05						Cutoff 0.10					
		FC = 1.5		FC = 2.0		FC = 3.0		FC = 1.5		FC = 2.0		FC = 3.0	
		FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE	FDR SE	FNR SE
Raw		0.57	0.66	0.37	0.25	0.31	0.01	0.66	0.53	0.51	0.15	0.48	0.00
P value		0.0007	0.0007	0.0006	0.0006	0.0004	0.0002	0.0005	0.0007	0.0004	0.0006	0.0003	0.0001
FDR		0.05	1.00	0.04	0.77	0.05	0.12	0.09	0.99	0.09	0.63	0.09	0.06
P value		0.0074	0.0001	0.0006	0.0011	0.0003	0.0005	0.0072	0.0002	0.0007	0.0012	0.0004	0.0004
λ_1	0.1	0.05	1.00	0.05	0.76	0.04	0.11	0.09	0.99	0.10	0.61	0.09	0.06
		0.0075	0.0001	0.0006	0.0011	0.0003	0.0005	0.0066	0.0003	0.0007	0.0011	0.0004	0.0004
	0.2	0.05	1.00	0.06	0.73	0.04	0.11	0.11	0.99	0.11	0.58	0.09	0.06
		0.0071	0.0001	0.0006	0.0011	0.0003	0.0005	0.0065	0.0003	0.0007	0.0011	0.0004	0.0004
	0.3	0.06	1.00	0.06	0.71	0.04	0.10	0.12	0.99	0.12	0.55	0.09	0.05
		0.0070	0.0001	0.0007	0.0011	0.0003	0.0005	0.0059	0.0004	0.0007	0.0011	0.0004	0.0003
	0.4	0.07	1.00	0.07	0.67	0.04	0.10	0.15	0.98	0.14	0.51	0.09	0.05
		0.0070	0.0002	0.0007	0.0012	0.0003	0.0005	0.0054	0.0005	0.0008	0.0011	0.0004	0.0003
	0.5	0.09	0.99	0.08	0.63	0.04	0.09	0.18	0.98	0.17	0.46	0.10	0.05
		0.0072	0.0002	0.0007	0.0012	0.0003	0.0004	0.0047	0.0006	0.0008	0.0010	0.0004	0.0003
	0.6	0.11	0.99	0.10	0.58	0.03	0.09	0.22	0.96	0.20	0.41	0.10	0.04
		0.0065	0.0003	0.0007	0.0011	0.0003	0.0004	0.0034	0.0009	0.0008	0.0010	0.0004	0.0003
	0.7	0.15	0.98	0.13	0.52	0.03	0.08	0.30	0.92	0.25	0.34	0.10	0.04
		0.0054	0.0005	0.0007	0.0011	0.0003	0.0004	0.0024	0.0011	0.0009	0.0009	0.0004	0.0003
	0.8	0.22	0.96	0.17	0.43	0.03	0.08	0.44	0.80	0.33	0.26	0.11	0.03
		0.0034	0.0008	0.0007	0.0009	0.0003	0.0004	0.0015	0.0015	0.0009	0.0008	0.0003	0.0003
	0.9	0.43	0.80	0.27	0.30	0.04	0.07	0.68	0.48	0.44	0.17	0.12	0.03
		0.0015	0.0014	0.0007	0.0007	0.0003	0.0004	0.0005	0.0008	0.0008	0.0006	0.0003	0.0002

We applied the *BFS* method to the microarray experiment described previously as the motivating example. The *BFS* method did not offer a larger set of differentially expressed genes than the set identified by the *FDR P* value < 0.05 criterion. For example, there were 24 genes that met *FDR P* value < 0.05 for gender difference. When the *BFS* method is applied, the number of genes actually decreased to 12. This was due to the fact that the average fold change was 1.05 with only 18 genes showing more than a twofold change. Based on these analyses, we concluded that age and gender effects on gene expression were very subtle in hippocampus.

In some cases, there may exist a unique λ_1 that satisfies $FPR(\lambda_1) = FNR(\lambda_1)$; however, our simulation results show that λ_1 does not always exist, as seen in Figure 2. When the magnitude of differential expression is relatively small (eg, FC = 1.5), the *FDR*-adjusted *P* value criterion fails to identify differentially expressed genes (see Table 3). However, the *BFS* method offers at least a handful of genes that biomedical researchers can pursue and then perform a confirmation experiment. In this situation, the choice of λ_1 should depend on the resources available to the researcher for follow-up experiments and her/his willingness to take a risk for potential false positive genes. When the magnitude of differential expression is large (eg, FC = 3.0), a medium range of λ_1 reduces the number of false negatives as well as false positives. In the real experiment we have no idea regarding the proportion of truly differentially expressed genes or the magnitude of differential expression, therefore an analytical and/or empirical approach for finding an optimal choice of λ_1 would be helpful in the future.

Acknowledgments

This publication is made possible with support from the Oregon Clinical and Translational Research Institute (OCTRI), grant number UL1 RR024140 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research.

Disclosures

The authors report no conflicts of interest in this work.

References

1. Slonim DK. From patterns to pathways: Gene expression data analysis comes of age. *Nat Genet*. 2002;32 Suppl:502–508.
2. Dudoit S, Shaffer J, Boldrick J. Multiple hypothesis testing in microarray experiments. *Stat Sci*. 2003;49:71–103.
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57:289–300.
4. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98:5116–5121.
5. Storey JD, Tibshirani R. Statistical significance for genome-wide studies. *Proc Natl Acad Sci U S A*. 2003;100:9440–9445.
6. Norris AW, Kahn CR. Analysis of gene expression in pathophysiological states: Balancing false discovery and false negative rates. *Proc Natl Acad Sci U S A*. 2006;103:649–653.
7. Rosenfeld S. Detection of differentially expressed genes in small sets of cDNA microarrays. *J Data Sci.* 2007;5:JDS-341.
8. Allison DB, Gadby G, Heo M, et al. Mixture model approach for the analysis of microarray gene expression data. *Comput Stat Data Anal*. 2003;39:1–20.
9. Genovese C, Wasserman L. Operating characteristics and extensions of the false discovery rate procedure. *J R Stat Soc Series B Stat Methodol.* 2002;64:499–517.
10. Sarkar SK. *FDR*-controlling stepwise procedures and their false negatives rates. Invited talk: The 3rd International Conference on Multiple Comparisons. Bethesda, MD: Technical Report, Temple University; August, 2002.

Appendix 1

Proposition: Let X_1 and X_2 be independent random variables with continuous density functions f_1, f_2 and distribution functions F_1, F_2, respectively. For $0 < \lambda < 1$, let P denote the distribution function of $T_\lambda = \lambda X_1 + (1-\lambda) X_2$ and let $G(t; \lambda) = P\{T_\lambda \leq t\}$.

$$G(t;\lambda) = \int_{-\infty}^{\infty} F_1[\lambda^{-1}(t-x)+x] f_2(x)\, dx$$

Proof

$$G(t;\lambda) = P\{T_\lambda \leq t\} = \mathrm{E}[\mathrm{P}\{T_\lambda \leq t \mid X_2\}]$$
$$= \int_{-\infty}^{\infty} \mathrm{P}\{T_\lambda \leq t \mid X_2\} f_2(x)\, dx$$

Because of independence we can write:

$$\mathrm{P}\{T_\lambda \leq t \mid X_2 = x\} = \mathrm{P}\{\lambda X_1 + (1-\lambda)\, x \leq t \mid X_2 = x\}$$
$$= \mathrm{P}\{\lambda X_1 + (1-\lambda)\, x \leq t\} = F_1(\lambda^{-1}\,[t + (1-\lambda)x])$$
$$= F_1(\lambda^{-1}\,(t-x) + x)$$

This can be easily extended to $k > 1$.