

Research article

Open Access

## An improved procedure for gene selection from microarray experiments using false discovery rate criterion

James J Yang\*<sup>1</sup> and Mark CK Yang\*<sup>2</sup>

Address: <sup>1</sup>Biostatistics and Research Epidemiology, Henry Ford Health Sciences Center, Detroit, Michigan, USA and <sup>2</sup>Department of Statistics, University of Florida, Gainesville, Florida, USA

Email: James J Yang\* - [jjyang2@hfhs.org](mailto:jjyang2@hfhs.org); Mark CK Yang\* - [yang@stat.ufl.edu](mailto:yang@stat.ufl.edu)

\* Corresponding authors

Published: 11 January 2006

Received: 11 July 2005

BMC Bioinformatics 2006, 7:15 doi:10.1186/1471-2105-7-15

Accepted: 11 January 2006

This article is available from: <http://www.biomedcentral.com/1471-2105/7/15>

© 2006 Yang and Yang; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** A large number of genes usually show differential expressions in a microarray experiment with two types of tissues, and the  $p$ -values of a proper statistical test are often used to quantify the significance of these differences. The genes with small  $p$ -values are then picked as the genes responsible for the differences in the tissue RNA expressions. One key question is what should be the threshold to consider the  $p$ -values small. There is always a trade off between this threshold and the rate of false claims. Recent statistical literature shows that the false discovery rate (FDR) criterion is a powerful and reasonable criterion to pick those genes with differential expression. Moreover, the power of detection can be increased by knowing the number of non-differential expression genes. While this number is unknown in practice, there are methods to estimate it from data. The purpose of this paper is to present a new method of estimating this number and use it for the FDR procedure construction.

**Results:** A combination of test functions is used to estimate the number of differentially expressed genes. Simulation study shows that the proposed method has a higher power to detect these genes than other existing methods, while still keeping the FDR under control. The improvement can be substantial if the proportion of true differentially expressed genes is large. This procedure has also been tested with good results using a real dataset.

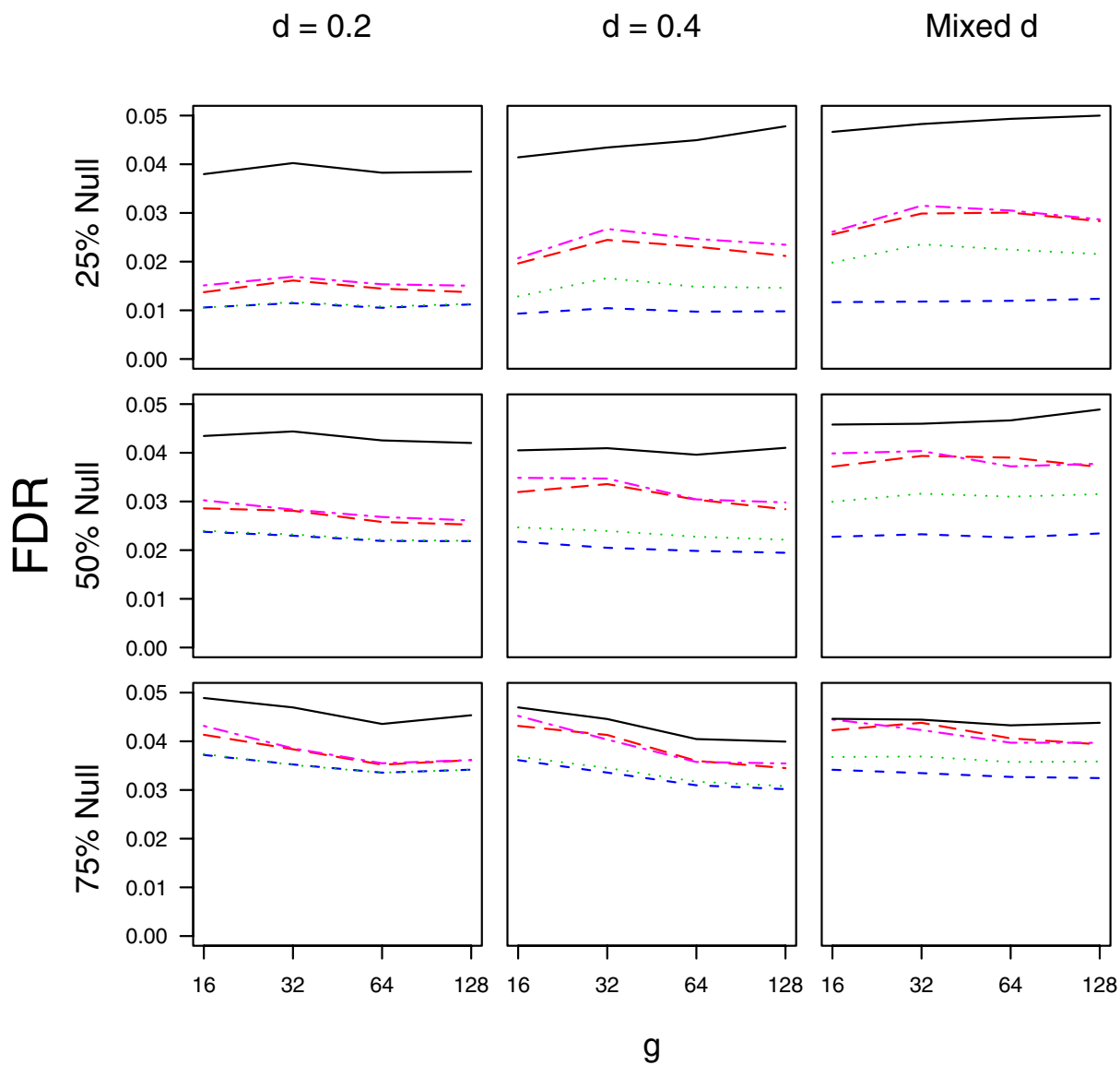
**Conclusion:** For a given expected FDR, the method proposed in this paper has better power to pick genes that show differentiation in their expression than two other well known methods.

### Background

The development of microarray technologies has created unparalleled opportunities to study the mechanism of disease, monitor disease expression and evaluate effective therapies. Because tens of thousands of genes are investigated simultaneously with a technology that has not yet been perfected, assessing uncertainty in the decision process relies on statistical modelling and theory. One key function of any statistical procedure is to control the rate

of erroneous decisions or, in the microarray case, rate of false discovery of responsible genes.

The above concern can be illustrated as a multiple comparisons problem. Suppose we are interested in testing  $g$  parameters  $(\mu_1, \dots, \mu_g) = \mu$ . For each individual parameter  $\mu_j$ , the null hypothesis is  $H_{0j}: \mu_j = 0$  and the alternative hypothesis is  $H_{1j}: \mu_j \neq 0$ . This  $\mu_j$  can be thought as the difference in mean expressions of the  $j$ th gene under two dif-



**Figure 1**

A simulation study to estimate FDR (in y-axis) for various numbers of hypotheses Tested (x-axis), proportion of true null hypotheses (left labels), and various configurations of true alternative hypotheses (top labels). All the examined methods, BH procedure (- - -); aBH procedure (.....); Fisher's combining function (-- --); Liptak's combining function (-. -.); proportion of null known (----), satisfied the required  $FDR \leq 0.05$ .

ferent conditions in a microarray experiment. A conventional method to test each hypothesis is to take a sample and then calculate its  $p$ -value from a proper statistical test. If the calculated  $p$ -value is less than a threshold determined by a testing significance level, then  $H_{0j}$  is rejected. However, when many hypotheses are simultane-

ously performed, a multiple comparisons procedure (MCP) has to be used to control the error rate [1].

The traditional MCP controls the probability of making any error in multiple selections, i.e., controls the family-wise error rate (FWER). It has been shown, however, that

this type of procedure is extremely conservative when the number of hypotheses is large. Alternatively, Benjamini and Hochberg [2] proposed the measure of *false discovery rate* (FDR) for which the expected proportion of false discoveries is controlled. This procedure is based on the idea that one can tolerate more false discoveries if the number of tests is large. For example, 5 false discoveries out of 10 selections is probably too many while 5 false discoveries out of 100 selections should be acceptable. This is particularly useful in microarray analysis since a very large number of genes usually show differential expressions. Therefore, controlling the FDR can greatly increase the power of discovery.

Benjamini and Hochberg [3] proved that Simes' procedure [4] can be used to control expected FDR at  $\alpha$  ( $0 < \alpha < 1$ ) when tests of true null hypotheses are either independent or positively dependent. More specifically, let  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(g)}$  be the ordered  $p$ -values and  $J + 1$  be the smallest integer satisfying

$$p_{(J+1)} > (J + 1) \times \frac{\alpha}{g} \tag{1}$$

If  $J \geq 1$ , then rejecting  $H_{(0j)}$ ,  $j = 1, \dots, J$  ensures the expected FDR at  $\alpha$ . We call this method the *BH procedure*.

Suppose the number of true nulls is  $g_0$ , or the proportion is  $\pi_0 = g_0/g$ . Benjamini and Hochberg [3] proved that the expected FDR of BH procedure is less than or equal to

$$\frac{g_0}{g} \alpha.$$

When  $g_0$  is small compared to  $g$ , using the original  $\alpha$  in (1) loses of power because it can be replaced by the larger value  $\alpha/\pi_0$  and still control the FDP at  $\alpha$ . Here, the power of a selection procedure is defined as the proportion of the alternative hypotheses that are correctly identified. Obviously, if we knew  $\pi_0$  in advance, we could replace  $\alpha$  in (1) by  $\alpha/\pi_0$  to increase power. Since  $\pi_0$  is unknown, the key question here is how to estimate the number of true null hypotheses  $g_0$ .

In an earlier published paper, Benjamini and Hochberg [5] proposed an *adaptive procedure* (aBH) which, by simulation, showed a higher power than BH procedure. The idea of the aBH procedure is to estimate  $g_0$  based on the fact that the  $p$ -value under the alternative hypothesis is stochastically smaller than that under the null, which is uniformly distributed on  $(0, 1)$ . On a quantile plot of  $p$ -values ( $p_{(j)}$  versus  $j$ ), the slope passing through  $(j, p_{(j)})$  and  $(g + 1, 1)$  increases just as  $j$  increases if the  $P_{(j)}$ s are corresponding to the subset of true alternative hypotheses. The first time the slope decreases indicates a change point.

Using this stopping rule in conjunction with the Lowest Slope (LSL) estimator,  $(1 - p_{(j)})/(g + 1 - j)$ , Benjamini and Hochberg proposed their aBH method to estimate  $g_0$  as follows:

**Alg. 1 aBH method in  $g_0$  estimation**

1. If there is no hypothesis rejected by the BH procedure, then stop and declare no discovery.
2. Calculate  $m_{(j)} = (1 - p_{(j)})/(g + 1 - j)$ ,  $\forall j = 1, \dots, g$ .
3. Let  $J^*$  be the largest value such that  $m_{(J^*)} < m_{(J^* - 1)}$ . Define  $b$  as the smallest integer larger than  $1/m_{(J^*)}$ .
4. The number of true null hypotheses is estimated as  $\hat{g}_0 = \min(b, g)$ .

Recently, Storey and Tibshirani [6] observed and demonstrated in their Figure 1 that the density histogram of  $p$ -values based on  $p_1, \dots, p_g$  from a microarray experiment looked flat in the region of  $(0.5, 1)$ . Based on the fact that the null  $p$ -values of genes are uniformly distributed, most of genes in the region of  $(0.5, 1)$  should be from the null. Therefore, they used a smoothing method to estimate  $g_0$  based on the flat region of the observed  $p$ -values. However, the implicit requirements of their method are: 1)  $g$  should be large, and 2) the proportion of true null hypotheses should also be large, such as 0.5, so that  $g_0$  can be estimated accurately. The method developed in this paper tries to bypass these two requirements so that a broad range of multiple testing problems can be applied. It is conceivable that when the number of relevant genes are known for certain target disease, a chip with a small number of genes will be widely used for diagnosis. The FDR methods, including both the BH and the aBH versions, have now been widely accepted for microarray analyses. They have been implemented in the R program [7], which can be incorporated into the Bioconductor package [8]. Several microarray analysis computer programs, such as SAM (Significance Analysis of Microarrays) [9] and GeneSpring [10] also use the FDR criterion to identify differentially expressed genes. Yang *et al.* [11] have used the FDR criterion to determine the sample size in microarray experiments.

**Methods**

Without loss of generality, we describe the method in a one-sample testing problem. Its extension to commonly used paired- $t$  test, two-sample  $t$  test, or nonparametric rank tests in microarray analysis is very straightforward. Let  $Y_i = (Y_{i1}, \dots, Y_{ig})'$ ,  $i = 1, \dots, n$  be  $g$ -variate vector samples from populations with means  $\mu = (\mu_1, \dots, \mu_g)'$ . Define  $Y^{(j)} = (Y_{1j}, \dots, Y_{nj})'$ ,  $j = 1, \dots, g$  as a row vector for the  $j$ th component. The null hypothesis of the  $j$ th component is  $H_{0j}: \mu_j = 0$  and the alternative hypothesis is  $H_{1j}: \mu_j \neq 0$ . Let the  $p$ -

value for rejecting the null  $H_{0j}$  be  $p_j$  derived from a properly chosen test statistic.

Once the  $p$ -values are derived, we first summarize the information content of  $p$ -values using a differentiable real-valued decreasing function  $h$ , i.e.,  $h(p_1, \dots, p_g)$  is decreasing to all its arguments. We call  $h$  the combining function and its value,  $\eta = h(p_1, \dots, p_g)$ , the global statistic. Next, define  $\eta^{(0)}$  to be the observed global statistic of  $\eta$  derived directly from the observed data. Note that a small value of  $p_j$  indicates  $H_{0j}$  is less likely to be true. Since  $h$  is a decreasing function of its argument  $p_j$ , a large value of  $\eta^{(0)}$  indicates a subset of the null hypotheses is likely to be true alternatives. To determine whether  $\eta^{(0)}$  is large enough to make such a claim, we need to know the distribution of  $\eta$  when all the nulls are true. The distribution of  $\eta$  depends on the distribution of its arguments, the correlation between its arguments, and the combining function  $h$ . In a microarray experiment, there seems no way to model or estimate so many correlations (see, for example, Chapter 6 of Pesarin [12]). Hence, a reasonable approach which can tackle the correlation within each experimental unit is to determine the critical value by a permutation test. More specifically, we calculate all  $B = 2^n$  values of  $\eta$  based on  $(\omega_1 Y_1, \dots, \omega_n Y_n)$ , where  $\omega_i = \pm 1$  (The multivariate two-sample permutation method can be found in Pesarin [12]). Hence, we have a set of reference values  $\{\eta_1^{(0)}, \dots, \eta_B^{(0)}\}$  from the  $B$  permutations, and can now define the  $p$ -value of the global  $\eta^{(0)}$ -statistic as

$$p^{(0)} = \frac{\sum_{b=1}^B I(\eta_b^{(0)} \geq \eta^{(0)})}{B} \tag{2}$$

where  $I(x)$  is the indicator function that takes value 1 if the statement  $x$  is true and 0 if it is not. To determine whether the global null hypothesis of all  $\mu_{(j)}$ ,  $j = 1, \dots, g$ , is zero is to evaluate if the global  $p$ -value,  $p^{(0)}$ , is less than or equal to a significance level. However, this level is part of the estimation process. It will be determined by the data (see later Alg. 2).

When the global null hypothesis is rejected, it indicates that not all null hypotheses  $H_{0j}$ s are true. Immediate question is, which subset of hypotheses is the true alternative. To determine whether to reject the hypothesis  $H_{(0j)}$  is a multiple testing problem. We can, however, determine the size of true alternatives using the following iterated process. We regard the gene with the smallest  $p$ -value as the major contributor for the rejection and claim that the null

hypothesis is not true with this gene. When this gene is removed from the data set, we continue the same process with the rest  $g - 1$  genes and the  $p^{(0)}$  computed by (2) is now denoted by  $p^{(1)}$ . The whole process is then repeated to produce a sequence of pseudo-global  $p$ -values,  $p^{(s)}$ ,  $s = 1, \dots, g - 1$ . We call the later step global  $p$ -value pseudo because it is only based on the subset of the original data. The detail of how to generate pseudo-global  $p$ -values is given in the **Algorithm** section. First, we will use these pseudo-global  $p$ -values to estimate the number of true null genes  $g_0$ .

We observe that, intuitively,  $p^{(0)} \leq p^{(1)} \leq \dots \leq p^{(g-1)}$ , and this monotone increasing property will be proved in (J.1) of the **Justification** section. In addition, we will prove in (J.2) of the Justification section that if  $r$  pseudo-global  $p$ -values are less than a given value  $\beta$  ( $0 < \beta < 1$ ), the estimator of the number of null genes is

$$\hat{g}_0 = g - r + \beta / (1 - \beta)^2 \tag{3}$$

and this estimator ensures that  $E[\hat{g}_0] \geq g_0$ . (The conservativeness of this estimate under other conditions using other techniques has also been mentioned by Storey and Tibshirani [13] and Efron et al. [14].) Since the inequality  $E[\hat{g}_0] \geq g_0$  holds for any value of  $\beta$ , the best choice of  $\beta$  is the one which makes  $\hat{g}_0$  closest to  $g_0$ . By defining  $\Delta = \beta / (1 - \beta)^2$  and  $\rho(\beta) = r - \Delta$ , we observe that  $\rho(\beta)$  is an increasing function and then a decreasing function of  $\beta$  for the following reasons. When  $\beta$  is small,  $\Delta$  increases slowly. However,  $r$  is an increasing function in  $\beta$  and  $r$  is always greater than 1. Therefore,  $\rho(\beta)$  is an increasing function for small value of  $\beta$ . On the other hand, when  $\beta \rightarrow 1$ ,  $\Delta$  reaches  $\infty$ . Since  $r$  is finite,  $\rho(\beta)$  becomes a decreasing function in this range. Since (3) is equivalent to  $\hat{g}_0 = g - \rho(\beta)$ , the optimal value of  $\beta$  should be determined as

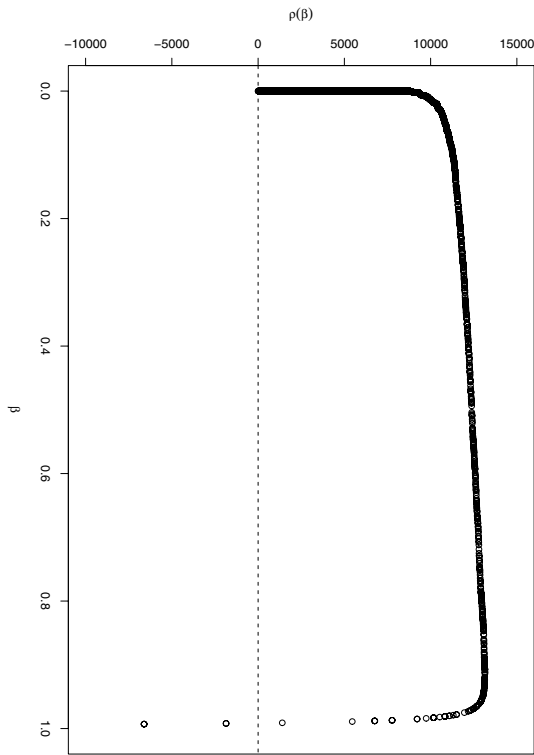
$$\hat{\beta} = \arg \max_{\beta} \rho(\beta)$$

and the method of estimating  $g_0$  is equivalent to finding the optimal  $\beta$  value.

The estimation of  $g_0$  can thus be summarized as

**Alg. 2 global-p method in  $g_0$  estimation**

1. List  $r_\beta$  as the integer such that  $p^{(s-1)} \leq \beta \forall s = 1, \dots, r_\beta$ , and  $p^{(r_\beta)} > \beta$  for a large number of  $\beta$ s for  $0 < \beta < 1$ .



**Figure 5**  
The plot of function  $\rho(\beta)$  in a real data analysis to determine the  $\arg \max \rho(\beta)$  for the tumor data.

2. Find  $\beta$  such that  $\rho(\beta) = r_{\beta} - \beta / (1 - \beta)^2$  is maximized (see Figure 5). Let this  $\beta$  be  $\beta^*$ .
3. Let  $r_{\beta^*}$  be the integer such that  $p^{(s-1)} \leq \beta^* \forall s = 1, \dots, r_{\beta^*}$ , and  $p^{(r_{\beta^*})} > \beta^*$
4. Let  $\hat{g}_0 = \min[g - r_{\beta^*} + \beta^* / (1 - \beta^*)^2, g]$ .

We called this *global-p* method and denoted it by  $\wp$ .

**Alg. 3 Differentially expressed gene selection with given FDR**

With  $g_0$  estimated, to control the FDR at  $\alpha$  level for identifying differentially expressed genes in microarray data analysis with the conservative estimate  $\hat{g}_0$ , we simply test each sub-hypothesis based on (1) with the new  $\alpha \equiv \alpha / \hat{g}_0$ , i.e., rejects all  $H_{(j)}, j = 1, \dots, J$ , if  $p_{(j)} \leq j \times \frac{\alpha}{\hat{g}_0}$  and  $p_{(j+1)} > (J$

$+ 1) \times \frac{\alpha}{\hat{g}_0}$ . We also need to address the choices for the combining function  $h$ . In this paper, we consider only two commonly used ones. 1) Fisher's sum of logarithm

$$\eta = h(p_1, \dots, p_g) = -2 \sum_{j=1}^g \log(p_j)$$

and 2) Liptak's sum of inverse standard Gaussian distribution functions

$$\eta = h(p_1, \dots, p_g) = \sum_{j=1}^g \Phi^{-1}(1 - p_j).$$

More discussion on the choices of combining functions can be found in Birnbaum [15] and Folks [16]. The computer program for the proposed method, *global-p*, has been implemented in R script and is publicly available [17].

**Results**  
**Simulation**

Two simulation studies were conducted. First, when the number of hypotheses are small and second, when the total number of hypotheses is extremely large. For small number of hypotheses, we used the following configurations: numbers of hypotheses are 16, 32, 64, and 128 with sample size  $n = 10$  and the proportions of  $g_0/g$  being 0, 0.25, 0.5 and 0.75. The true expressions under the alternative hypotheses are assumed to be variance 1 Gaussian random variables with non-zero mean values  $\mu_j = d_j$ . In one set of experiments, all  $d_j$ s are set as 0.2 or 0.4; in the other set the  $d_j$ s are divided into 4 equal size blocks with values 0.2, 0.4, 0.6 or 0.8 in each block. The number of permutations is  $B = 1,000$  and the number of simulations is 20,000. The FDR is set at  $\alpha = 0.05$ . Four procedures are used for testing: BH, aBH, our *global-p* test one with Fisher's combining function, and one with Liptak's combining function. For the purpose of comparison, we also plug in the true value of  $g_0$  into the aBH method. This is the ideal situation to reach the highest power.

**Table 1: False discovery rate when all null hypotheses are true.**

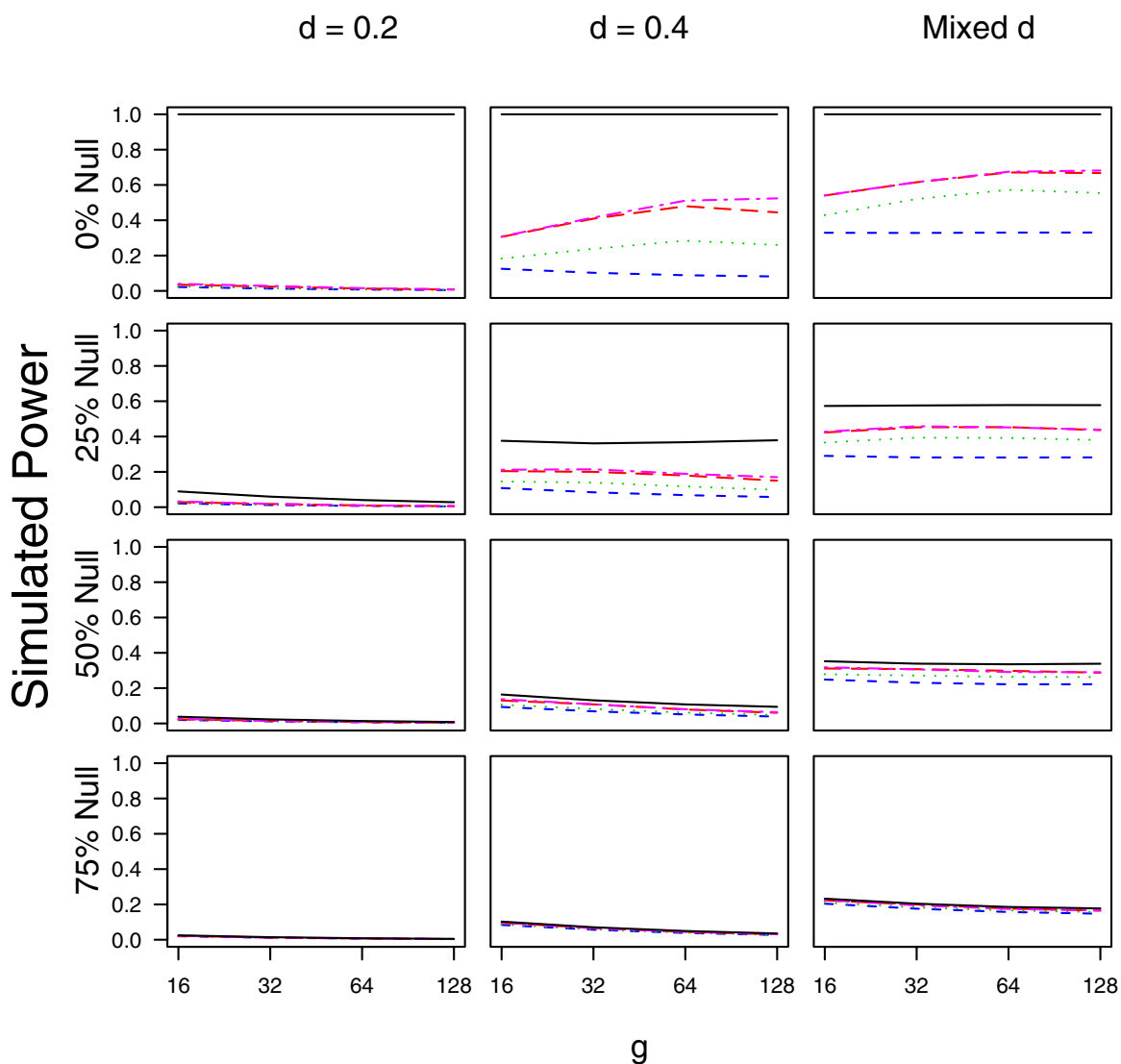
g	BH	aBH	Fisher	Liptak
16	0.047	0.047	0.048	0.048
32	0.049	0.049	0.049	0.050
64	0.049	0.049	0.049	0.049
128	0.048	0.048	0.049	0.049

Table 1 lists the simulated expected FDR when all null hypotheses are true. Figures 1 and 2 show the estimated FDR and power under various gene expression compositions.

Based on this simulation results, all the four methods have their expected FDR below the nominal FDR  $\alpha = 0.05$ . Both combining functions in our proposed method have a higher power than the aBH procedure in most situations, while there is little difference between them. The

improvement over the aBH methods is substantial when the proportion of the true null hypotheses is small.

Since most microarray data consist of tens of thousands of genes simultaneously and many of their expressions are correlated, our second simulation tried to reflect these facts. Storey and Tibshirani [13] proposed a "clumpy" dependence model in which genes are partitioned into blocks. The gene expressions are assumed independent between blocks but dependent within the same block.



**Figure 2**

The same simulation study with estimated average powers, using the same legends as Figure 1. The powers of the new method are clearly higher than the existing two, especially when the null proportion is small.

Following the clumpy model, 10 test and 10 control samples were generated from normal random variables with mean 0 and standard deviation 1 where each sample contained 10,000 genes. For test samples, genes with expression difference were added by 3 to represent the expression differences. The number of differentially expressed genes  $g_0$  were 100, 2,000, 5,000, 8,000 which corresponded to the proportion of nulls  $\pi_0$  at 0.99, 0.8, 0.5, 0.2. To simulate intra-block dependency, we generated one vector of normal random variables with mean 0 and standard deviation 0.2 in each block of size 50 and add it to every gene in that block. This process creates correlations between genes. Since the true expression differences for non-null genes are moderate large in this simulation, we expect that any good method should accurately estimate  $\pi_0$ .

In a recent article, Broberg [18] did a comparative review of various newly developed methods for estimating  $\pi_0$ . To make a comparison, we presented the means and standard errors of estimates of  $\pi_0$  for all the methods describe in Broberg's paper in addition to our proposed method ( $\varphi$ ) over 1000 simulations. The details of  $\pi_0$  estimation methods (BUM, SPLOSH, QVALUE, Bootstrap LSE, SEP, LSL, mgf, PRE) can be found in Broberg's paper and the programs for these methods can be found in the add-on packages of the freely available R software [19]. The simulation results are shown in Table 2.

As expected, all of them performed very well for large  $\pi_0$ . If both the mean (bias) and standard error (stability) over the whole range of  $\pi_0$  are considered, LSL and Global-p stand out and LSL is better in stability. However, as we will see from the next study, LSL seems to be over-conservative in real data analysis. The means of both combining functions of our proposed method performed well but Liptak's function has a higher standard errors. Therefore, Fisher's function will be used for the real data analysis. During the simulation, we also noticed that current implementation of SPLOSH, SEP, mgf, and PRE methods in R programs failed to work when the number of hypotheses was small.

### Real data analysis

We use a publically available experimental data set from the Stanford microarray database [20] first to compare the differences between BH, aBH and our method with Fisher's combining function. The purpose of this experiment was to identify genes that have different expressions between prostate tumor tissue and matched normal tissue. It consists of a total of 82 microarrays: 41 arrays were from primary prostate tumors and the other half were from matched normal prostate tissues. Each array contains 32,152 different human genes. We chose this data

set because of the large number of replicates. With this amount of replicates, we had a better idea of the ground truth.

To compare performances of various methods, we split the whole data set into two groups: a test data set and a confirmation data set. Eleven pairs of tumor and normal arrays were chosen for the test data set and the remaining arrays were used for the confirmation. The number of eleven test pairs were taken from a systematic sampling to avoid bias. Patients 1, 5, 9, ..., 41 in the original order from the database were chosen as the test set. Expression information is missing if it is labelled as failed, contaminated, or flagged. Genes were removed from our study if more than half of their expressions were missing in the original data set. A total of 24,865 genes was used to identify differences in expressions using BH, aBH and our proposed procedures at 0.01 expected FDR level. Since the experiment was designed with paired tumor-normal arrays, the paired  $t$ -test was used to derive the  $p$ -value of each gene in the test data set. The BH procedure identified 1,254 genes, the aBH procedure identified 1,523 genes, and our proposed method identified 2,119 genes. Our test is apparently more powerful if it can maintain the required FDR 0.01. Since we did not have the biological information, another approach was to estimate FDR based on our rejection set. Specifically, suppose the rejection set contains the  $p$ -values that are less than  $\xi$  and the estimated proportion of null hypotheses is  $\hat{\pi}_0$ . An intuitive estimate for FDR is [21,22]

$$\widehat{\text{FDR}}(\xi) = \frac{\pi_0 \xi}{\widehat{\text{Pr}}[P \leq \xi]}, \quad (4)$$

where  $\widehat{\text{Pr}}[P \leq \xi] = \sum_{j=1}^g I(p_j \leq \xi) / g$ . Based on the confirm data set, our proposed method reported an estimator of  $\pi_0$ ,  $\hat{\pi}_0 = 0.5326$ . Using equation (4), the estimated FDRs for BH, aBH and our method were 0.0054, 0.0067, 0.0099, respectively.

If we further reject more sub-hypotheses beyond the number provided by our method, the FDR will exceed the assigned level. For example, if we reject the next 500 null sub-hypotheses that have the smallest  $p$ -values among those not previously rejected, the estimated FDR is 0.0137 which is larger than the pre-assigned 0.01 level.

Moreover, we calculated the standard difference (paired  $t$ -statistic) of expressions using the confirmation data set. The confirmation data set contains 30 pairs of arrays so

**Table 2: The means (first row) and standard errors (second row) of estimates  $\hat{\pi}_0$  based on various methods when  $\pi_0 = 0.2, 0.5, 0.8, 0.99$ .**

True $\pi_0$	0.200	0.5000	0.8000	0.9900
BUM	0.1124 0.0021	0.3962 0.0024	0.7288 0.0021	1.0000 0.0000
SPLOSH	0.8770 0.0737	0.9030 0.0592	0.9343 0.0482	0.9594 0.0271
LSL	0.2011 0.0005	0.5015 0.0007	0.8016 0.0006	0.9904 0.0003
Bootstrap LSE	0.1938 0.0079	0.4905 0.0135	0.7863 0.0184	0.9787 0.0159
QVALUE	0.1975 0.0107	0.4979 0.0176	0.7980 0.0242	0.9871 0.0154
SEP	0.9982 0.0004	0.4978 0.0078	0.7982 0.0089	0.9883 0.0090
mgf	-	0.4944 0.0044	0.7973 0.0055	0.9905 0.0062
PRE	-	0.1712 0.0084	0.6568 0.0107	0.9815 0.0066
Global-p (Fisher)	0.2005 0.0014	0.5002 0.0012	0.8002 0.0018	0.9899 0.0019
Global-p (Liptak)	0.1998 0.0017	0.4992 0.0038	0.7978 0.0037	0.9884 0.0027

that the statistic is a good estimate for the unknown standard difference. Figure 3 shows histograms of absolute standard differences based on genes identified by the BH procedure, by the aBH procedure, by our method, and the extra 500 genes beyond our method. From the histograms, several hundred differentially expressed genes that were not identified by BH or aBH in the test arrays were identified by our method and most of them have standard expression differences greater than 2. However, the next group of 500 genes beyond our method may contain too many false discoveries to make the FDR acceptable.

Next, we use this data set to compare different estimates of proportion of null genes,  $\pi_0$ . The whole data set was partitioned into four sub-data sets. We labelled the arrays from 1 to 41 based on the original order in their database. Sub-data set 1 contained 11 arrays with labels 1, 5, 9, ..., 41; sub-data set 2 contains 10 arrays with labels 2, 6, ..., 38; sub-data set 3 contains 10 arrays with labels 3, 7, ..., 39; sub-data set 4 contained the remaining arrays. Although we did not know the true value of  $\pi_0$  in real data analysis, we used the four sub-data sets to compare the variation of different estimation methods. In addition, we used the whole data set to check if the estimates are reliable. For global- $p$  method, we also plot the function of  $\rho(\beta)$  in Figure 5 using the whole data set. The maximum of  $\rho(\beta)$  occurs at  $\beta = 0.91$ .

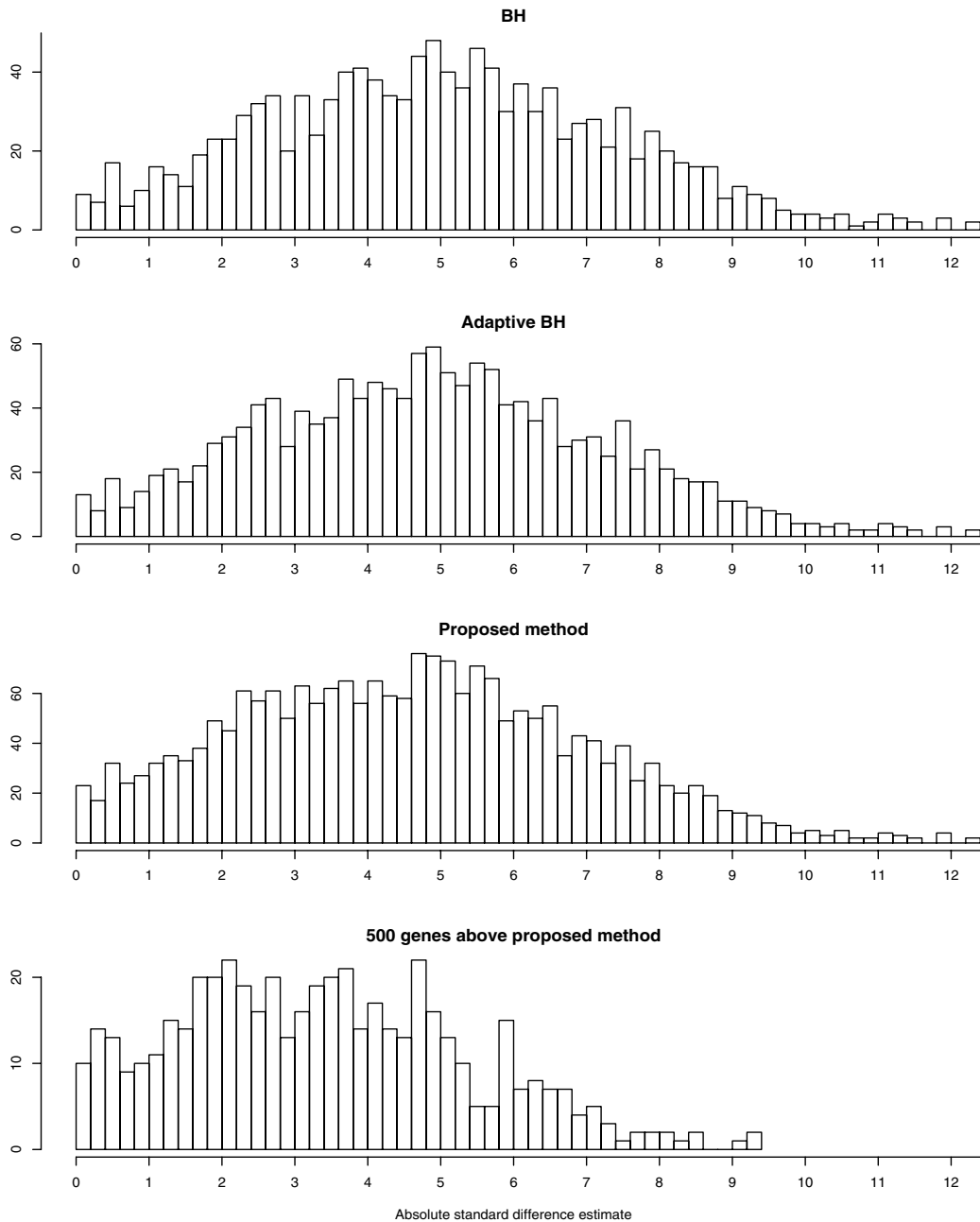
The estimates  $\hat{\pi}_0$  were listed in Table 3. All  $\hat{\pi}_0$  estimates in the subsets were larger than the ones in the whole data set. This is expected because genes with small expressional

differences are more likely to be classified as null genes in a small sample. From the  $\pi_0$  estimates from the whole data set, we may say the LSL gave a large  $\pi_0$  value, PRE gave a small value and all the others gave similar estimates. We used LSL, PRE and Global-p to draw the  $p$ -value density histogram using the whole data set in Figure 4. The upper panel is the overall view and the lower panel is the closer view. The green dash-dot line is the height using LSL estimate; the red dash line PRE estimate; the blue dot line global- $p$  estimate. The LSL method, which is also shown in Broberg's study, is too conservative producing the largest value. The PRE method underestimated  $\pi_0$ . If we used  $\hat{\pi}_0$  by PRE method to improve FDR procedure, it is likely to have a higher FDR than the nominal level. We think our  $\pi_0$  estimate is reliable because it is close but higher than the height of observed  $p$ -values in the region of (0.6, 1) if we assume genes with expressional differences have rare chance to have large observed  $p$ -values.

**Conclusion**

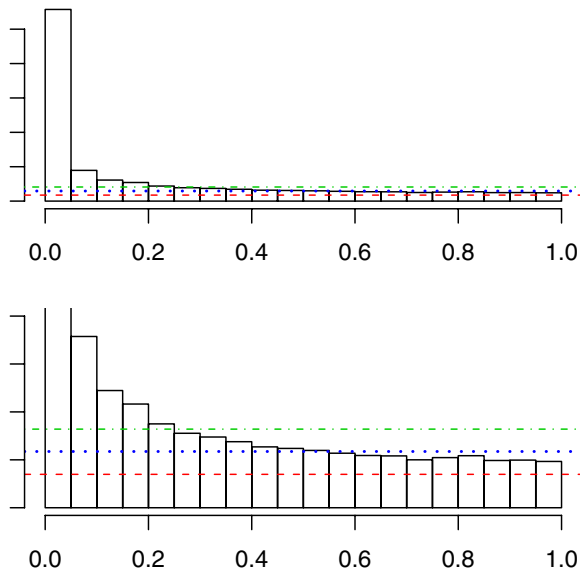
Benjamini and Hochberg's FDR procedure [2] opens an useful approach for multiple comparisons in microarray analysis. Due to the complexity in microarray experiments, it seems unlikely that one method can dominate all the other methods in all the cases. In this paper, we have demonstrated a new method that is intuitive appearing because:





**Figure 3**

The histograms of the absolute standard difference estimates (*x*-axis) in the confirmation data set for the selected genes from the three procedures based on the test data set for the prostate tumor data. The top is the selected gene histogram by the BH procedure; the second is that by the aBH procedure; and the third is the one by our method. The similarity of the three histograms means that the additional 596 genes picked by our method should maintain the same FDR (0.01) as the other two. The bottom panel contains the next 500 genes beyond our proposed method. The left shift of this histogram means that the additional genes may contain too many false discoveries.



**Figure 4**  
The histogram of  $p$ -values in a real data analysis. About half of the large  $p$ -values to the right look uniformly distributed.

1. It uses permutation test which can take care the complex correlation structures in gene expressions.
2. Its global test based on sequentially eliminated significant genes should provide a good stopping rule because all the remaining genes are always considered together.

With the support of simulation and real data studies, the new method should be a viable alternative to find the differentially expressed genes in microarray experiments.

**Algorithm**

We illustrate the algorithm for calculating pseudo-global  $p$ -values based on marginal  $p$ -values,  $p_1, \dots, p_{g_0}$ . Please note that the procedure described here is the same regardless of  $g_0$ . We consider  $h(p_1, \dots, p_{g_0})$  in the summation form:

$$\eta = h(p_1, \dots, p_{g_0}) = \sum_{i=1}^{g_0} \tilde{h}(p_i), \tag{5}$$

where  $h$  is a differentiable decreasing function in which  $\tilde{h}'(p) = \frac{d\tilde{h}(t)}{dt} \Big|_{t=p}$  exists and  $\tilde{h}(p) \rightarrow \infty$  as  $p \rightarrow 0$ . Note that equation (5) totally meets the requirement that  $h$  should be a differentiable real-valued decreasing function. The realization of  $h$  can be, for instance,  $h(p) = -2\log(p)$  if we use

Fisher's sum of logarithm or  $h(p) = \Phi^{-1}(1 - p)$  if we use Liptak's sum of inverse standard Gaussian distribution function. Recall that  $p_j$  is the  $p$ -value obtained from  $j$ th row vector  $(Y_{1j}, \dots, Y_{nj})$  and the ordered marginal  $p$ -values are  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(g_0)}$ . The  $p$ -values for individual genes are denoted by subscripts while the pseudo-global  $p$ -values in (2) are denoted by superscripts. Parentheses are used to reveal the order property.

In one-sample testing problem, we permute data  $B$  times, where each time we assign  $w_i = \pm 1$  with equal probability to  $(w_1 Y_{1j}, \dots, w_n Y_{nj})$ . For each permuted data, we use  $p_{b(j)}$  to be the  $p$ -value from the  $b$ th permuted data for the gene that produced  $p_{(j)}$ ,  $j = 1, \dots, g_0$ ,  $b = 1, \dots, B$ .

We summarize  $p_{(i)}$  and  $p_{b(i)}$  using the following table.

$p_{(1)}$	$p_{1(1)}$	$\dots$	$p_{b(1)}$	$\dots$	$p_{B(1)}$
$p_{(2)}$	$p_{1(2)}$	$\dots$	$p_{b(2)}$	$\dots$	$p_{B(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$p_{(g_0)}$	$p_{1(g_0)}$	$\dots$	$p_{b(g_0)}$	$\dots$	$p_{B(g_0)}$

(6)

Note that the first column is  $p$ -values of genes from the original data while the remaining columns are from the permuted data.

To simplify the notation, let  $(p_{(i)}) = (p_{i})$ ,  $(p_{b(i)}) = (p_{b(i)})$  and  $\eta$  is just the sum of  $(p_{i})$ . Therefore, we can transform the table (6) into the following table with the addition of the last row where its value  $\eta$  is the sum of the  $(p_{i})$  in its corresponding column.

$\tilde{h}_{(1)}$	$\tilde{h}_{1(1)}$	$\dots$	$\tilde{h}_{b(1)}$	$\dots$	$\tilde{h}_{B(1)}$
$\tilde{h}_{(2)}$	$\tilde{h}_{1(2)}$	$\dots$	$\tilde{h}_{b(2)}$	$\dots$	$\tilde{h}_{B(2)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\tilde{h}_{(g_0)}$	$\tilde{h}_{1(g_0)}$	$\dots$	$\tilde{h}_{b(g_0)}$	$\dots$	$\tilde{h}_{B(g_0)}$
$\Downarrow$	$\Downarrow$	$\dots$	$\Downarrow$	$\dots$	$\Downarrow$
$\eta^{(0)}$	$\eta_1^{(0)}$	$\dots$	$\eta_b^{(0)}$	$\dots$	$\eta_B^{(0)}$

(7)

Once  $\eta^{(0)}$  and  $\eta^{(1)}, \dots, \eta^{(B)}$  are available, the global  $p$ -value, by definition, is

$$p^{(0)} = \frac{\sum_{b=1}^B I(\eta_b^{(0)} \geq \eta^{(0)})}{B}$$

The same process is directly used to calculate the pseudo-global  $p$ -values after removing the genes which have the smallest  $p$ -values. Actually, we can use the available data illustrated in table (8). By induction, suppose  $p_{(1)}, \dots, p_{(s)}$

**Table 3: Estimates of  $\hat{\pi}_0$  based on the prostate tumor data set from the Stanford microarray database. The first four rows are estimated based on the four partitions of the whole data set while the last row the is based on the data set. The means and standard errors (5th and 6th rows) are calculated based on the estimates of the four sub-data sets.**

	BUM	SPLOSH	LSL	Bootstrap LSE	QVALUE	SEP	mgf	PRE	Global-p
Sub-data 1	0.4011	0.4844	0.8071	0.4802	0.4662	0.5792	0.5677	0.4536	0.5583
Sub-data 2	0.5211	0.5710	0.9106	0.5797	0.5679	0.5865	0.6866	0.5727	0.6722
Sub-data 3	0.5529	0.6104	0.9302	0.6235	0.6070	0.6232	0.7120	0.6097	0.7082
Sub-data 4	0.6056	0.6627	0.9628	0.6709	0.6618	0.6760	0.7529	0.6646	0.7493
Mean	0.5202	0.5821	0.9027	0.5886	0.5758	0.6162	0.6798	0.5751	0.6702
Sd	0.0867	0.0752	0.0672	0.0813	0.0826	0.0443	0.0796	0.0894	0.0821
Whole	0.3940	0.4706	0.6598	0.3949	0.3882	0.5979	0.4706	0.2803	0.4721

have been removed, the values of the raw data and their reference sets are

$$\begin{matrix}
 \hat{h}_{(s+1)} & \hat{h}_{1(s+1)} & \dots & \hat{h}_{b(s+1)} & \dots & \hat{h}_{B(s+1)} \\
 \hat{h}_{(s+2)} & \hat{h}_{1(s+2)} & \dots & \hat{h}_{b(s+2)} & \dots & \hat{h}_{B(s+2)} \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 \hat{h}_{(g_0)} & \hat{h}_{1(g_0)} & \dots & \hat{h}_{b(g_0)} & \dots & \hat{h}_{B(g_0)} \\
 \downarrow & \downarrow & \dots & \downarrow & \dots & \downarrow \\
 \eta^{(s)} & \eta_1^{(s)} & \dots & \eta_b^{(s)} & \dots & \eta_B^{(s)}
 \end{matrix} \tag{8}$$

and the pseudo-global  $p$ -value after removing the  $s$  genes is

$$p^{(s)} = \frac{\sum_{b=1}^B I(\eta_b^{(s)} \geq \eta^{(s)})}{B}$$

**Justification**

Our justification is based on extreme cases to make the proofs tractable. However, this requirement is not very critical. Our simulation and real data analysis show that even with moderate  $g_0$  value, the pseudo-global  $p$ -value,  $p^{(s)}$ , starts to increase when our procedure has removed most of the significant genes. Besides, it jumps up very fast to reach the point that  $p^{(s)}$  will be larger than threshold  $\beta$  and stop removing genes. Let  $g$  be the total number of sub-hypotheses considered. Given  $g$ , let  $\mathbf{d} = (d_1, \dots, d_g)$  denote the vector containing the true values of  $\mu_j$ ,  $j = 1, \dots, g$  for each sub-hypothesis; and  $R(g, \mathbf{d})$  be the number of  $Y^{(i)}$  removed using procedure  $\wp$ . Without loss of generality, we assume  $d_{(1)} \leq \dots \leq d_{(g-g_0)} < 0$  and  $d_{(j)} = 0$  for  $j = g - g_0 + 1, \dots, g$ . We have

$$E_{\wp}[R(g, \mathbf{d})] \leq g - g_0 + E_{\wp}[R(g_0, \mathbf{0})], \tag{9}$$

and the equality holds when  $d_{(g-g_0)} \rightarrow -\infty$ . The equality affirms that the sub-sample  $Y^{(i)}$  with extreme small  $d_j < 0$  is identified and removed from the component of  $\eta$ . Actually, when the gene expression difference is far away from zero, it will be identified by any reasonable multiple testing procedure. Hence, in this extreme case, we only focus on  $E_{\wp}[R(g_0, \mathbf{0})]$  which is based on the remaining  $g_0$  samples that have null distributions. Our goal is reduced to finding an upper bond of  $E_{\wp}[R(g_0, \mathbf{0})]$ , so that the expected value of  $\hat{g}_0$  can be bounded (below) by  $g_0$ .

**J.1 Monotone in Pseudo-global p-values**

The key feature of our algorithm is that the pseudo-global  $p$ -values are monotone increasing. We can be more precise by showing that

$$Pr \left[ p^{(s)} \leq p^{(s+1)} \mid p^{(s)} \right] \rightarrow 1 \text{ as } \frac{s}{g_0} \rightarrow 0. \tag{10}$$

Romano (Section 2)[23] has proved that the distribution of  $\eta_b^{(s)}$  can be approximated by Gaussian distribution using the Central Limit theorem. Therefore, if we let the mean and standard deviation of reference samples  $\eta_b^{(s)}$  ( $b = 1, \dots, B$ ) be  $\mu^{(s)}$  and  $\sigma^{(s)}$ , respectively, and use  $Z$  to denote the standard Gaussian random variable with distribution function  $\Phi$ , then the pseudo-global  $p$ -value  $p^{(s)}$  can be expressed as

$$\begin{aligned}
 p^{(s)} &= \frac{\sum_{b=1}^B I(\eta_b^{(s)} \geq \eta^{(s)})}{B} \\
 &\doteq Pr \left[ Z > \frac{\eta^{(s)} - \mu^{(s)}}{\sigma^{(s)}} \right] \\
 &= Pr \left[ Z > C^{(s)} \right], \tag{11}
 \end{aligned}$$

where  $C^{(s)} = \Phi^{-1}(1 - p^{(s)})$  is  $p^{(s)}$ th upper percentile of the standard Gaussian distribution. Hence, the relation between  $\eta^{(s)}$  and  $C^{(s)}$  is

$$\eta^{(s)} = \mu^{(s)} + \sigma^{(s)} C^{(s)}.$$

Similarly, for the  $\mu^{(s+1)}$  and  $\sigma^{(s+1)}$  as the mean and standard deviation of reference samples  $\eta_b^{(s+1)}$  ( $b = 1, \dots, B$ ), we repeat the same approximation method to express  $p^{(s+1)}$  as

$$\begin{aligned}
 p^{(s+1)} &= \frac{\sum_{b=1}^B I(\eta_b^{(s+1)} \geq \eta^{(s+1)})}{B} \tag{12} \\
 &\doteq Pr \left[ Z > \frac{\eta^{(s+1)} - \mu^{(s+1)}}{\sigma^{(s+1)}} \right] \\
 &= Pr \left[ Z > \frac{(\eta^{(s+1)} + \hat{h}_{(s+1)}) - (\mu^{(s+1)} + \hat{h}_{(s+1)})}{\sigma^{(s+1)}} \right] \\
 &= Pr \left[ Z > \frac{\eta^{(s)} - (\mu^{(s+1)} + \hat{h}_{(s+1)})}{\sigma^{(s+1)}} \right] \\
 &= Pr \left[ Z > \frac{\mu^{(s)} + \sigma^{(s)} C^{(s)} - (\mu^{(s+1)} + \hat{h}_{(s+1)})}{\sigma^{(s+1)}} \right] \\
 &= Pr \left[ Z > \frac{\sigma^{(s)}}{\sigma^{(s+1)}} C^{(s)} + \frac{\mu^{(s)} - \mu^{(s+1)} - \hat{h}_{(s+1)}}{\sigma^{(s+1)}} \right]. \tag{13}
 \end{aligned}$$

Comparing equations (11) and (13), we observe that to prove  $p^{(s)} \leq p^{(s+1)}$  with probability 1 asymptotically when  $p^{(s)}$  is given is equivalent to prove that

$$\frac{\sigma^{(s)}}{\sigma^{(s+1)}} C^{(s)} + \frac{\mu^{(s)} - \mu^{(s+1)} - \hat{h}_{(s+1)}}{\sigma^{(s+1)}} < C^{(s)} \tag{14}$$

with probability 1 asymptotically when  $C^{(s)}$  is given.

As  $g_0 \rightarrow \infty$ , the sample deviations  $\sigma^{(s)}$  and  $\sigma^{(s+1)}$  are almost identical so that  $\frac{\sigma^{(s)}}{\sigma^{(s+1)}} C^{(s)} \rightarrow C^{(s)}$ . Next, using table (8) again, we can express  $\mu^{(s)} - \mu^{(s+1)}$  as

$$\frac{\sum_{b=1}^B \hat{h}(p_{b(s+1)})}{B}.$$

The distribution of the reference set  $p_{1(s+1)}, \dots, p_{B(s+1)}$  converges to the uniform (0, 1) distribution since the distribution of the permuted test statistics, which are corresponding to these  $p$ -values, converges to a Normal distribution using Theorem 2.1 of Romano's work. Therefore, if we define uniform (0, 1) random variable as  $U$ ,

$$\frac{\sum_{b=1}^B \hat{h}(p_{b(s+1)})}{B}$$

converges in probability to  $E[(U)]$  by the

Law of Large Number. Hence, to prove equation (14) we only need to prove that

$$Pr[(p_{(s+1)}) \geq E[(U)]] \rightarrow 1. \tag{15}$$

Assume, without loss of generality, that  $p_{1, \dots, p_{g_0}}$  correspond to the true null. To prove equation (15), we first observe that, since  $p_{1, \dots, p_{g_0}}$  are independent uniform random variables,  $P_{(s+1)}$  is the  $(s+1)/g_0$ -th quantile of the uniform (0, 1) distribution. By defining  $\zeta_{s+1, g_0} = (s+1)/g_0$  and using the distribution of sample quantile [24], we have

$$\sqrt{g_0} (p_{(s+1)} - \zeta_{s+1, g_0}) \rightarrow N(0, \zeta_{s+1, g_0} (1 - \zeta_{s+1, g_0})).$$

Since the first derivative of  $\hat{h}$  exists, we can use the delta method to have

$$\sqrt{g_0} (\hat{h}(p_{(s+1)}) - \hat{h}(\zeta_{s+1, g_0})) \rightarrow N(0, \sigma^{*2}),$$

where  $\sigma^* = \sqrt{\zeta_{s+1, g_0} (1 - \zeta_{s+1, g_0}) [\hat{h}'(\zeta_{s+1, g_0})]^2}$  represents the asymptotical standard error of  $(p_{(s+1)})$ .

Finally, equation (15) can be proven as,

$$\begin{aligned}
 &Pr[\hat{h}(p_{(s+1)}) < E[\hat{h}(U)]] \\
 &= Pr \left[ \frac{\hat{h}(p_{(s+1)}) - \hat{h}(\zeta_{s+1, g_0})}{\sigma^*} \right. \\
 &\quad \left. < \frac{E[\hat{h}(U)] - \hat{h}(\zeta_{s+1, g_0})}{\sigma^*} \right] \\
 &\doteq \Phi \left( \frac{E[\hat{h}(U)] - \hat{h}(\zeta_{s+1, g_0})}{\sigma^*} \right)
 \end{aligned}$$

which converges to 0 because  $E[(U)]$  is finite and  $\hbar(\zeta_{s+1, g_0}) \rightarrow \infty$  as  $\zeta_{s+1, g_0} = (s + 1)/g_0 \rightarrow 0$ .

**J.2 Determining the threshold  $\beta$  by (3)**

First, we recall that, as  $p^{(0)}$  is the global  $p$ -value based on all null data, it is uniformly distributed. That is,

$$Pr[p^{(0)} \leq \beta] = \beta.$$

We have proved that the pseudo-global  $p$ -values,  $p^{(s)}$ , are monotone increasing when  $s/g_0 \rightarrow 0$ . Therefore, there exists  $m$  such that  $p^{(0)} \leq p^{(1)} \leq \dots \leq p^{(m)}$ . Furthermore, for all  $s \leq m$ ,

$$Pr[p^{(s)} \leq \beta | p^{(j)} \leq \beta, j = 0, \dots, s - 1] \leq \beta.$$

The purpose here is to derive equation (3) for any  $\beta$  within (0,1) using the above inequality. We start with the number of genes removed by our procedure. Observe that

$$\begin{aligned} E_{\varphi}[R(g_0, \mathbf{0})] &= \sum_{k=0}^{g_0-1} k Pr[R(g_0, \mathbf{0}) = k] \\ &= \sum_{k=1}^{g_0-1} k Pr[p^{(0)} \leq \alpha, p^{(1)} \leq \beta, \dots, p^{(k-1)} \leq \beta, \\ &\quad p^{(k)} > \beta]. \end{aligned}$$

In addition,

$$\begin{aligned} Pr[p^{(0)} \leq \beta, p^{(1)} \leq \beta, \dots, p^{(k-1)} \leq \beta, p^{(k)} > \beta] &= Pr[p^{(0)} \leq \beta] \\ &\quad \left\{ \prod_{s=1}^{k-1} Pr[p^{(s)} \leq \beta | p^{(l)} \leq \beta, l = 0, \dots, s - 1] \right\} \\ &\quad Pr[p^{(k)} > \beta | p^{(i)} \leq \beta, i = 0, \dots, k - 1] \\ &\leq \beta \beta^{\min(k-1, m)} \\ &= \beta^{k_m} \end{aligned}$$

where  $k_m = \min(k, m + 1)$ . Hence, we can find an upper bound of the expected number of genes removed,  $E_{\varphi}[R(g_0, \mathbf{0})]$ , which is

$$\begin{aligned} E_{\varphi}[R(g_0, \mathbf{0})] &\leq \sum_{k=1}^{g_0-1} k \beta^{k_m} \\ &= \frac{\beta}{(1-\beta)^2} (1-\beta^{m+1}) - (m+1)\beta^{m+2} / (1-\beta) \\ &\quad + (g_0 - m - 2)\beta^{m+1} \\ &\approx \frac{\beta}{(1-\beta)^2} \text{ as } m \text{ is large.} \end{aligned} \tag{16}$$

Let  $\Delta = \Delta(\beta) = \beta / (1 - \beta)^2$ . From equations (9) and (16), we have

$$E_{\varphi}[R(g, \mathbf{d}) - \Delta] \leq g - g_0$$

so that the number of true null estimate is

$$\hat{g}_0 = g - R(g, \mathbf{d}) + \Delta$$

and this estimate ensures that

$$E_{\varphi}[\hat{g}_0] \geq g_0.$$

**References**

- Westfall PH, Young SS: *Resampling-based multiple testing* New York: John Wiley & Sons; 1993.
- Benjamini Y, Hochberg Y: **Controlling the false discovery rate: A Practical and powerful approach to multiple testing.** *JRSSB* 1995, **57**:289-300.
- Benjamini Y, Hochberg Y: **The control of the false discovery rate in multiple testing under dependency.** *Annals of Statistics* 2001:1165-1188.
- Simes RJ: **An improved Bonferroni procedure for multiple tests of significance.** *Biometrika* 1986, **73**.
- Benjamini Y, Hochberg Y: **On the adaptive control of the false discovery rate in multiple testing with independent statistics.** *Journal of Education and Behavioral Statistics* 2000, **25**:60-83.
- Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *PNAS* 2003, **100**:9440-9445.
- Reiner A, Yekutieli D, Benjamini Y: **Identifying differentially expressed genes using false discovery rate controlling procedures.** *Bioinformatics* 2003, **19**:368-375.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Li FLC, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: Open software development for computational biology and bioinformatics.** 2004 [<http://genomebiology.com/2004/5/10/R80>].
- Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *PNAS* 2001, **98**:5116-5121.
- GeneSpring 7.1. Silicon Genetics** [<http://www.silicongenetics.com>]
- Yang MCK, Yang JJ, McIndoe RA, She JX: **Microarray experimental design: power and samples size considerations.** *Physiol Genomics* 2003, **16**:24-28.
- Pesarin F: *Multivariate permutation tests with applications in Biostatistics* West Sussex, England: John Wiley & Sons; 2001.
- Storey JD, Tibshirani R: **Estimating false discovery rates under dependence, with applications to DNA microarrays.** *Technical report, Department of Statistics, Stanford University* 2001.
- Efron B, Storey JD, Tibshirani R: **Microarrays, Empirical Bayes Methods, and False Discovery Rates.** *Technical Report 217, Department of Statistics, Stanford University* 2001.
- Birnbaum A: **Combining independent tests of significance.** *JASA* 1954, **49**:559-574.

16. Folks JL: **Combination of independent tests.** In *Handbook of Statistics Volume 4*. Edited by: Krishnaiah PR, Sen PK. New York: Elsevier Science Publishers; 1984:113-121.
17. **Global-p website** [[http://www.stat.ufl.edu/~jyang/global\\_p/global.p.R](http://www.stat.ufl.edu/~jyang/global_p/global.p.R)]
18. Broberg P: **A comparative review of estimates of the proportion unchanged genes and the false discovery rate.** *BMC Bioinformatics* 2005, **6(199)**:.
19. R Development Core Team: **R: A language and environment for statistical computing.** *R Foundation for Statistical Computing, Vienna, Austria* 2005 [<http://www.R-project.org>]. ISBN 3-900051-07-0
20. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD, Pollack JR: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *PNAS* 2004, **101**:811-816.
21. Storey JD: **A direct approach to false discovery rates.** *JRSSB* 2002, **64**:479-498.
22. Genovese C, Wasserman L: **A stochastic process approach to false discovery control.** *The Annals of Statistics* 2004, **32**:1035-1061.
23. Romano JH: **On the behavior of randomization tests without a group invariance assumption.** *Journal of the American Statistical Association* 1990, **85**:.
24. Serfling RJ: *Approximation Theorems of Mathematical Statistics* New York: John Wiley & Sons; 1980.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

