Available online at www.sciencedirect.com
ScienceDirect
Genomics 88 (2006) 846–854
GENOMICS
www.elsevier.com/locate/ygeno

Method

Ranking analysis of microarray data: A powerful method for identifying differentially expressed genes

Yuan-De Tan [a], Myriam Fornage [a], Yun-Xin Fu [b,c,*]

[a] *Institute of Molecular Medicine, School of Public Health, University of Texas at Houston, Houston, TX 77030, USA*
[b] *Laboratory for Conservation and Utilization of Bioresources, Yunnan University, Kunming, Yunnan 650, China*
[c] *Human Genetics Center, School of Public Health, University of Texas at Houston, Houston, TX 77030, USA*

Received 24 March 2006; accepted 2 August 2006
Available online 18 September 2006

Abstract

Microarray technology provides a powerful tool for the expression profile of thousands of genes simultaneously, which makes it possible to explore the molecular and metabolic etiology of the development of a complex disease under study. However, classical statistical methods and technologies fail to be applicable to microarray data. Therefore, it is necessary and motivating to develop powerful methods for large-scale statistical analyses. In this paper, we described a novel method, called Ranking Analysis of Microarray Data (RAM). RAM, which is a large-scale two-sample *t*-test method, is based on comparisons between a set of ranked *T* statistics and a set of ranked *Z* values (a set of ranked estimated null scores) yielded by a "randomly splitting" approach instead of a "permutation" approach and a two-simulation strategy for estimating the proportion of genes identified by chance, i.e., the false discovery rate (FDR). The results obtained from the simulated and observed microarray data show that RAM is more efficient in identification of genes differentially expressed and estimation of FDR under undesirable conditions such as a large fudge factor, small sample size, or mixture distribution of noises than **Significance Analysis of Microarrays**.

Keywords: Microarray; *t* test; Ranking analysis; False discovery rate

Microarray technology provides a powerful tool for measuring the expression levels of large numbers of genes simultaneously and creates unparalleled opportunities to study complex physiological or pathological processes, including the development of disease, that are mediated by the coordinated action of multiple genes [1]. Detection of genes differentially expressed across experimental, biological, and/or clinical conditions is a major objective of microarray experiments. Methods for finding genes significantly differentially expressed in the context of microarray data analysis can be classified into three major groups [2,3]: marginal filters, wrappers [4], and embedded approaches [5,6]. The wrapper and embedded methods are a type of search algorithm by which candidate gene subsets that are useful to build a good predictor are constructed and selected and then evaluated by using a classification algorithm [3,7]. The filter approaches are a type of simple and fast method including *t* tests and nonparametric scoring [8,9] and analysis of variance [1,10] for searching for the features (genes) or feature (gene) subsets that are irrelevant and independent of each other [3,7]. For the microarray data, the filter approaches encounter a challenging simultaneous inference problem, as the probability of committing a type I error increases with the number of tests performed [11]. To resolve the statistical problem in testing a large family of null hypotheses, several multiple procedures have been developed. The Bonferroni procedure, the Holm procedure [12], the Hochberg procedure [13], and the Westfall and Young procedure [14] address the multiple test problem by controlling the family-wise error rate, which is the probability that at least one false positive occurs over the collective tests [15]. However, these methods are based on the assumption that different tests are independent of each other; they are, thus, not well suited to microarray data, often being too stringent, and may yield no

* Corresponding author. Human Genetics Center, University of Texas at Houston, Houston, TX 77030, USA. Fax: +1 713 500 0900.
E-mail address: Yunxin.fu@uth.tmc.edu (Y.-X. Fu).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.08.003

or few positive genes [16] and may result in unnecessary loss of power. Benjamini and Hochberg [17] have proposed an alternative measure, the false discovery rate (FDR), to control erroneous rejection of a number of true null hypotheses. FDR is an expected proportion of the false positives among all the positives detected. The FDR-based multiple testing approaches, such as the Benjamini and Hochberg (BH) procedure [17,18] and the Benjamini and Liu procedure [19], have been developed for testing for a large family of hypotheses. These procedures are generally suited to larger sample sizes because small sample sizes lead FDR to be too "granular" [16]. Most recently, Storey [20] and Storey and Tibshirani [21] developed a new measure, i.e., positive FDR (pFDR), that is an arguably more appropriate variation. It multiplies the FDR by a factor of π_0, which is the estimated proportion of non-differentially expressed genes to all genes on the arrays [22]. The estimate of pFDR is smaller than the estimate of FDR [22]. Tsai et al. [23] suggested the use of the conditional FDR (cFDR) on the most significant findings. Pounds and Cheng [15,24] proposed the spacing LOESS histogram approach to estimate of cFDR. Tusher et al. [16] developed a new FDR-based method, called Significance Analysis of Microarrays (SAM). SAM is very popular because it can identify genes with significant change of the level of expression and can estimate FDR based on permutations. However, the conventional permutation approach is not the most appropriate method for estimating the null distribution for most microarray data because sample sizes in such experiments are commonly small and yield a relatively small number of permutations leading to inaccurate ranking of scores. Although SAM has the advantage of being distribution-free, its use of a fudge factor (S_0) makes it mostly applicable to normal distributions because S_0 is in general smaller than or equal to 1 in normal distributions. Nonnormal distributions or small sample sizes can produce a larger S_0, which often makes SAM lose its power or become not applicable.

These problems in SAM led us to develop a new statistical method called Ranking Analysis of Microarray (RAM) Data. The overall approach of RAM is somewhat similar to that of SAM, which is to identify genes with significant changes in expression through the use of gene-specific t tests, but RAM evaluates its significance based on an improved empirical distribution generated by a "randomly splitting" approach instead of the "permutation" approach and implementation of a simulation-based interval method for estimation of FDR. As a result, RAM has all the major advantages of SAM, plus it performs very well for small sample sizes, which are typical in microarray experiments.

Methods

T statistic

For simplicity, we will focus our discussion on the analysis of expression data from experiments of two different classes (designated as 1 and 2), which is very common in practice. The two classes may correspond to two different genotypes of individuals, treatments, cell types, tissues, etc. Let N be the number of genes examined and m_{ik} be the number of replicate observations for the expression of gene k (k=1, …, N) in class g (g=1, 2). We will refer to the collection of all the observations for a given gene in class g as sample g. Therefore, m_{gk} is the size of sample g for gene k. Typically $m_{11}=m_{12}=\ldots=m_{1N}=m_1$ and $m_{21}=m_{22}=\ldots=m_{2N}=m_2$, otherwise the experiment is said to have some missing observations.

Let $\bar{x}_{gk}$ and σ^2_{gk} represent the mean and variance, respectively, of the expression of gene k in sample g. Define for gene k

$$td_k = \bar{x}_{1k} - \bar{x}_{2k}.$$

The traditional t-test statistic for testing if there is a significant difference between two sample means is equal to

$$t_k = d_k/\sigma_k,$$

where in the current context

$$\sigma_k = \sqrt{\sigma^2_{1k}/m_{1k} + \sigma^2_{2k}/m_{2k}}\ ,$$

for unequal variances for the two class experiments, or

$$\sigma_k = \sqrt{\{[(m_{1k}-1)\sigma^2_{1k} + (m_{2k}-1)\sigma^2_{2k}]/(m_{1k}+m_{2k}-2)\}(1/m_{1k}+1/m_{2k})}\ ,$$

for equal variances. Although the traditional t statistic is a reasonable choice for some expression data sets, its applicability is often questionable because a small sampling variance ($<<1$), which can often arise due to randomness from a large number of genes and small sample size, and relatively large value of d_k may lead to an erroneous conclusion. Such an effect is generally known as the fudging effect. To reduce the fudging effect, Tusher et al. [16] proposed a modified t statistic defined as

$$T_k = d_k/(S_0 + \sigma_k),$$

where S_0 is a constant representing the minimal coefficient of variation of t_k computed as a function of σ_k in the moving windows across the data. However, in our own studies, we noted that the fudging effect using the modified t statistic is still quite strong when the sample size is small. In particular, small sample size often leads to an unreasonably large value of S_0 that dominates the test statistic and consequently reduces the power of the analysis. To circumvent the problem, we propose a simple alternative correction δ_k for the variance of expression for gene k as

$$\delta_k = \sqrt{A_k + \sigma^2_{1k}/m_{1k} + \sigma^2_{2k}/m_{2k}} \tag{1a}$$

for the case of unequal variances, and

$$\delta_k = \sqrt{A_k + \{[(m_{1k}-1)\sigma^2_{1k} + (m_{2k}-1)\sigma^2_{2k}]/(m_{1k}+m_{2k}-2)\}(1/m_{1k}+1/m_{2k})}, \tag{1b}$$

for the case of equal variances, where

$$A_k = \begin{cases} 1 & \text{if} \quad d_k > \sigma_k < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Thus, the t statistic for the difference in expression levels of gene k is redefined as

$$T_k = d_k/\delta_k. \tag{3}$$

Since $T_k=t_k$ unless $d_k>\sigma_k<1$, the new test statistic is a simpler extension of the traditional t statistic than that proposed by Tusher et al. [16].

Ranking analysis

To identify genes whose expression levels are significantly different under two experimental conditions, a common practice is to rank the genes according to their values of the chosen statistics, which in our situation is T. Suppose T_{k^*} is the k^*-th largest T value, then its corresponding gene k is said to have significantly different expression between the two experimental conditions for a given threshold value Δ if

$$|T_k* - Z_k*| > \Delta, \tag{4}$$

where $Z_{k*}=E(T_{k*})$ is the expectation of T_{k*} under the null hypothesis that there is no gene having a significant difference in expression. This type of test is known as the ranking test.

To enable the ranking test, it is critical to obtain a good estimate of Z_{k*}. Tusher et al. [16] proposed a permutation approach for this purpose, which uses a standard permutation procedure for each gene. This process works well if the sample size is large. When the sample size is small, however, the number of permutated samples for each gene is rather small, which leads to a biased ranking test and even renders the test not applicable. This appears to be caused by the randomness introduced by permutations that lead to biased tail distributions for ranked values. The observations from analyzing both real and simulated data led us to develop a randomly splitting (RS) approach to estimate Z as follows.

First each sample is randomly split into two subsamples with size difference not larger than a given value C. We found that it is best to set $C=4$. For the J-th split, let $\bar{x}^J_{hgk}$ be the mean of subsample h of sample g for gene k. Define $\bar{e}^J_{1k}=\bar{x}^J_{11k}-\bar{x}^J_{12k}$ and $\bar{e}^J_{2k}=\bar{x}^J_{21k}-\bar{x}^J_{22k}$, and hence,

$$\bar{e}^J_k=\frac{1}{2}\left(\bar{e}^J_{1k}+\bar{e}^J_{2k}\right). \tag{5}$$

The splitting process is carried out for every gene. Define $Z^J_k=\bar{e}^J_k/\delta_k$. The set of Z^J_k values is then ranked. Let Z^J_{k*} be the k*-th largest value for the J-th split. Then we estimate Z_{k*} by the mean of Z^J_{k*} over all the splits, i.e.,

$$\bar{Z}_{k*}=\frac{1}{M}\sum_{J=1}^{M} Z^J_{k*}. \tag{6}$$

Fig. 1A shows the use of Z_{k*} in the identification of the genes that are differentially expressed in a set of 3000 genes in a stroke-response experiment. In this figure the solid line represents $T=Z$ and the two dashed lines represent the lower and upper boundaries corresponding to a threshold Δ. The dots below the lower boundary and over the upper boundary represent genes that are significantly expressed at the given threshold Δ.

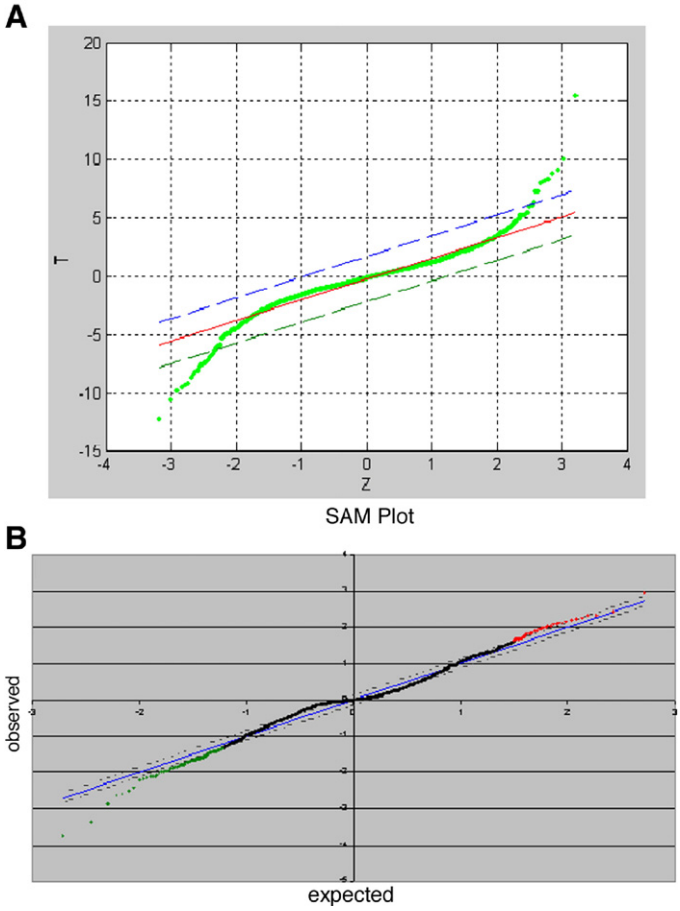

Fig. 1. Identification of the genes significantly differentially expressed. (A) A plot of T values vs Z values based on the observed data of 3000 genes in two samples, each consisting of 12 rats in response to stroke, where estimates of Z values were obtained by use of the RS approach. (B) A plot of observed T vs expected T (Z) in SAM. The simulated data set comprised 30% expression noises following gamma distribution and 70% following normal distribution where expression levels of 3000 genes in two samples, each consisting of 12 replicates, were simulated using one set of the observed sample means and two sets of the observed sample variances and treatment effect values of $G=10R$ (for 30% of the genes), where R is a random uniform variable over (0, 1].

Estimate of FDR

Consider a series of threshold values $\Delta_i(i=1,..,L)$. Let $N(i)$ be the number of genes that are significant at the threshold Δ_i by the ranking analysis. $N(i)$ then comprises two parts: the number of true positives $N_t(i)$ and the number of false positives $N_f(i)$. Therefore $N(i)=N_t(i)+N_f(i)$. FDR at threshold Δ_i can be written as $R_{FD}(i)=N_f(i)/N(i)$, which must be estimated since $N_f(i)$ is unknown. To improve the accuracy of estimating FDR, we propose a new strategy to obtain FDR as an average of two estimates each derived from simulation under a specific condition. The first estimate is carried out as follows.

For each gene, two samples of m replicates are simulated from a normal distribution, one with a mean randomly set to be $\bar{y}^J_{1k}=\frac{1}{2}\left(\bar{x}^J_{11k}+\bar{x}^J_{12k}\right)$ or $\frac{1}{2}\left(\bar{x}^J_{11k}+\bar{x}^J_{22k}\right)$ and variance σ^2_{1k}, another with a mean randomly set to be $\bar{y}^J_{2k}=\frac{1}{2}\left(\bar{x}^J_{21k}+\bar{x}^J_{12k}\right)$ or $\frac{1}{2}\left(\bar{x}^J_{21k}+\bar{x}^J_{22k}\right)$ and variance σ^2_{2k}, where $\bar{x}^J_{hgk}$ is the mean of subsample h of the sample g for gene k produced by the RS procedure in the observed data.

The process will produce M sets of simulated data, each subjecting to the ranking analysis described in the previous section. For each simulated data set, every ranked position has thus a corresponding T value that is denoted by T^J_{k*1}. Since we are concerned about false positives, we consider only those genes that are not significant in the original ranking analysis. Comparing T^J_{k*1} to $\bar{Z}_{k*}$ for every ranking position will allow one to identify genes that are becoming significant. The number of such genes in the J-th set of simulation data at the threshold Δ_i is denoted by $N(1,J,i)$.

Let $N(1,i)=\sum_{J=1}^{M} N(1,J,i)/M$, which is the mean number of $N(1,J,i)$. For an ascending series of threshold values, $N(1,i)$ rises initially and declines when the threshold value exceeds a certain value Δ. Define

$$f(1,i)=\frac{2N(1,i)}{N(\Delta)+N(1,i)} \tag{7}$$

as the first estimate of FDR where $N(\Delta)=\max_{i=1}^{L} N(1,i)$ and $N(1,i)=N(\Delta)$ when $\Delta_i<\Delta$. $f(1,i)$ is thus a decreasing function and bounded between 1 and 0 (see Fig. 2).

The second estimate of FDR is obtained also from simulation. The simulation of the two samples for each gene is done in the same way as the first simulation, except that the two means are set to be equal, i.e., $\bar{y}^J_{1k}=\bar{y}^J_{2k}=\frac{1}{2}\left(\bar{x}^J_{11k}+\bar{x}^J_{12k}\right)$ or $\frac{1}{2}\left(\bar{x}^J_{21k}+\bar{x}^J_{22k}\right)$. Also correspondingly for the J-th simulation data set, ranking analysis of the T values leads to T^J_{k*2}, where "2" represents the second simulation. T^J_{k*2} is compared to its average $\bar{T}_{k*2}$, and the significances across all the ranking positions at threshold Δ_i are counted as $N(2,J,i)$. Let $N(2,i)=\max_{J=1}^{M} N(2,J,i)$. Since the noise distribution produced by the RS approach from the simulated data agrees well with that produced by the RS approach from the observed data (see Figs. 3B, 4C, and 4D), $N(2,i)$ is a reasonable estimate of $N_f(i)$ for a given threshold Δ_i. However, to avoid the possibility that $R_{FD}(i)=N(2,i)/N(i)=\infty$ occurs when $N(i)=0$, in particular, in the extreme cases of which there is no or small expression difference between two samples, we define

$$f(2,i)=\frac{N(2,i)}{N(i)+N(2,i)} \tag{8}$$

as the second estimate of FDR. Eq. (8) shows that $f(2,i)=1$ when $N(i)=0$ and $N(2,i)\geq 1$, $f(2,i)=0.5$ when $N(i)=N(2,i)$, $f(2,i)<0.5$ when $N(i)>N(2,i)$, and $f(2,i)=0$ when $N(i)\geq 1$ and $N(2,i)=0$.

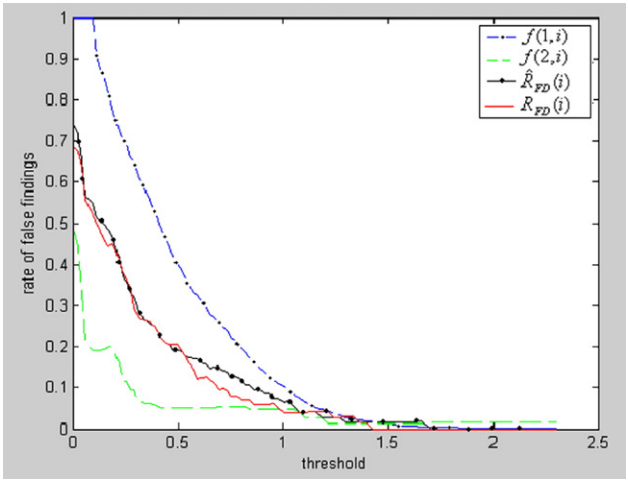

Fig. 2. Estimation of FDR. $f(1,i)$ and $f(2,i)$ are two threshold functions and are used to construct an estimation interval for estimate of FDR at threshold Δ_i. $R_{FD}(i)$ and $\hat{R}_{FD}(i)$ are the true and the estimated FDR, respectively, at threshold Δ_i, where $R_{FD}(i)$ was calculated by comparing genes identified by RAM with those given treatment effect ($G=10R$).

Although we intended to find lower and upper bounds for FDR, it can be seen from Fig. 2 that although the two estimates of FDR provide two bounds for FDR, $f(1,i)$ does not remain as the lower bound nor the upper bound, same as $f(2,i)$. The role of the two in bounding FDR switches after a certain threshold value. For this reason, we explore a single estimate of FDR, which value lies between the two bounds. One conservative estimate is to give more weight to the larger of the two bounds, which results in the third estimate of FDR,

$$f(3,i) = a_i f(1,i) + b_i f(2,i), \tag{9}$$

where $a_i = f(1,i)/[f(1,i)+f(2,i)]$ and $b_i = 1-a_i$. We found that at threshold level Δ_i, a better estimate of FDR is obtained by

$$f_i = \frac{1}{3}[f(1,i) + f(2,i) + f(3,i)]. \tag{10}$$

To smooth the estimates of FDR further, consider the difference between the numbers of genes found to be significant at adjacent thresholds Δ_i and Δ_{i+1} and define a recursive formula modifying the probability f_i as

$$f_i = f_i p_i + f_{i+1} q_i, \tag{11}$$

where $p_i = [N(i)-N(i+1)]/[1+N(i)-N(i+1)]$ and $q_i = 1-p_i$. Eq. (11) suggests that $f_{i+1} = f_i$ if $N(i) = N(i+1)$. Thus, the number of the false discoveries among those found to be significant at threshold Δ_i in the observed data is estimated by

$$\hat{N}_f(i) = f_i N(i) \tag{12}$$

and an estimate of FDR at threshold Δ_i is given by

$$\hat{R}_{FD}(i) = \hat{N}f(i)/N(i) = f_i. \tag{13}$$

It can be seen from Fig. 2 that the line for the true value $R_{FD}(i)$ agrees well with that for $\hat{R}_{FD}(i)$, indicating that $\hat{R}_{FD}(i)$ is a good estimate of FDR. We also found that, if no gene in the simulation was found to be significant, $\hat{R}_{FD}(i)$ would be more than 0.5 at threshold Δ_i of $f(1,i) < f(2,i)$ (the result is not shown).

Simulation results

Estimate of the null distribution

To determine if the empirical distributions obtained by the permutation approach and the RS approach are appropriate for the analysis of expression data, we simulated three sets of microarray data each consisting of 3000 genes and two samples of 12 replicates each. The means and variances for the two sample of each gene are set to be one of observed means and two of the observed, respectively. In our real microarray data sets, the expression levels of 3000 genes were measured in two different strains (the spontaneously hypertensive rat and the stroke-prone spontaneously hypertensive rat) each consisting of 12 rats. In the first simulation data set, all 3000 genes were set to have no treatment effect. In the second and third simulation data sets, treatment effects of $G=10R$ and $G=30R$, respectively, where R is a random variable in the uniform distribution (0,1], were randomly assigned to 30% of the genes.

In the ranking analysis, a set of Z_{k*} values for each simulated data set was computed from 100 permutations or 100 random splits. As Z_{k*} is an estimate of T_{k*} under the null hypothesis, a desirable property is that Z_{k*} has a linear relationship with T_{k*}.

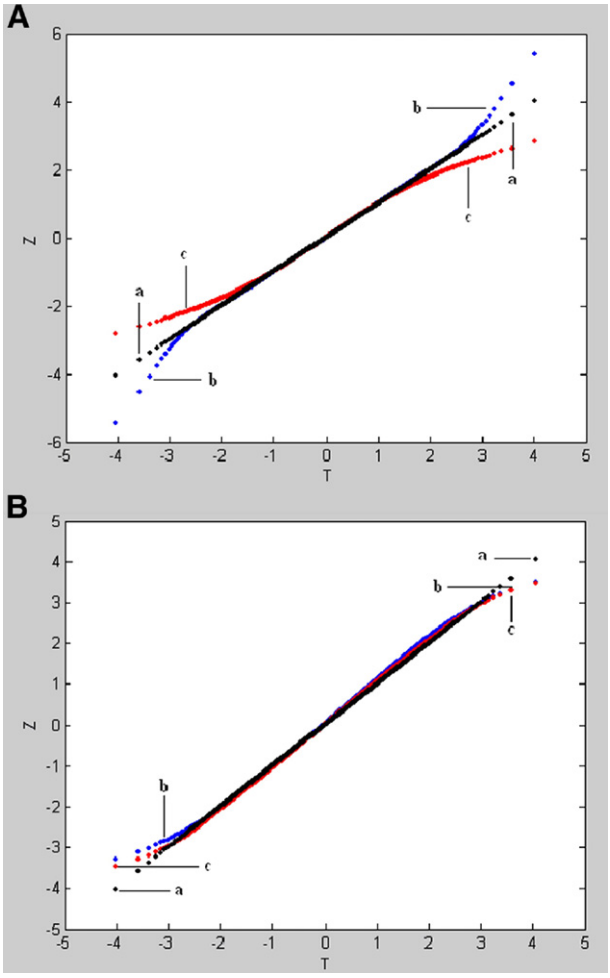

Fig. 3. Plots of Z values vs T values. The observed Z value (curve b) and simulated Z value (curve c) were obtained by (A) the permutation approach and (B) the RS approach. The observed microarray data of 3000 genes were obtained in two samples, each consisting of 12 rats. The first set of the simulated microarray data was produced using the pseudorandom generator and one set of 3000 observed means and two sets of 3000 observed variances in which no gene was not given a treatment effect value. The simulated T values (curve a) were a set of 3000 null scores produced from 100 repeated simulations of the first set of the simulated data (see text).

This property can be seen by plotting Z_{k^*} versus T_{k^*}. Fig. 3 shows the plot of Z obtained by the permutation (Fig. 3A) and RS (Fig. 3B) approaches. It can be seen from Fig. 3A that the Z distribution obtained by the permutation approach from either the observed or the first simulated data sets remarkably deviates from the null distribution when $|T|$ is large. More specifically, in the tails of T, the observed Z values remarkably overestimate the null scores, whereas the simulated Z values underestimate the null scores. These patterns were also seen from simulation incorporating different treatment effects on gene expression. In Fig. 4A, the Z^* values obtained by the permutation approach from the second simulation data set in which 30% of the genes were given treatment effect values of $10R$ are in between the Z values obtained by the permutation approach from the first simulation data set, in which no gene was given treatment effect, and the null scores (simulated T values) when $T>1.5$ or < -1.5, whereas in Fig. 4B, Z^* values from the third simulation data set, in which 30% of the genes were given treatment effect values of $30R$, are much larger than the null scores at $T>3$ or much smaller than the null scores at $T<-3$. These results indicate that when the treatment effect contributing to expression variations of genes is weak or lacking, the Z distribution yielded by the permutation approach would negatively deviate from the null distribution, i.e., $Z_{k^*} \le T_{k^*} > 0$ or $Z_{k^*} \ge T_{k^*} < 0$, so that type I errors observed in the ranking test would be more than those expected. However, when a large treatment effect to a different extent contributes to expression variations of a part of the genes, the Z distribution would remarkably positively deviate from the null distribution, i.e., $Z_{k^*} \ge T_{k^*} > 0$ or $Z_{k^*} \le T_{k^*} < 0$. In this case type II errors observed in the ranking test would be much more than those expected. These observations in the case of small samples are in fact a general feature of the permutation approach (see Appendix A).

It can be seen from Fig. 3B, however, that the Z distributions obtained by the RS approach from the observed and the first simulated data sets and the simulated T distribution (the null distribution) are almost overlapped with each other. This is also shown in Figs. 4C and 4D, in which the Z^* values were obtained by the RS approach from the second and third simulation data sets and the Z values from the first simulation data set. The results similar to those shown in Figs. 4C and 4D were obtained in the case of a sample size of 6. These results strongly suggest

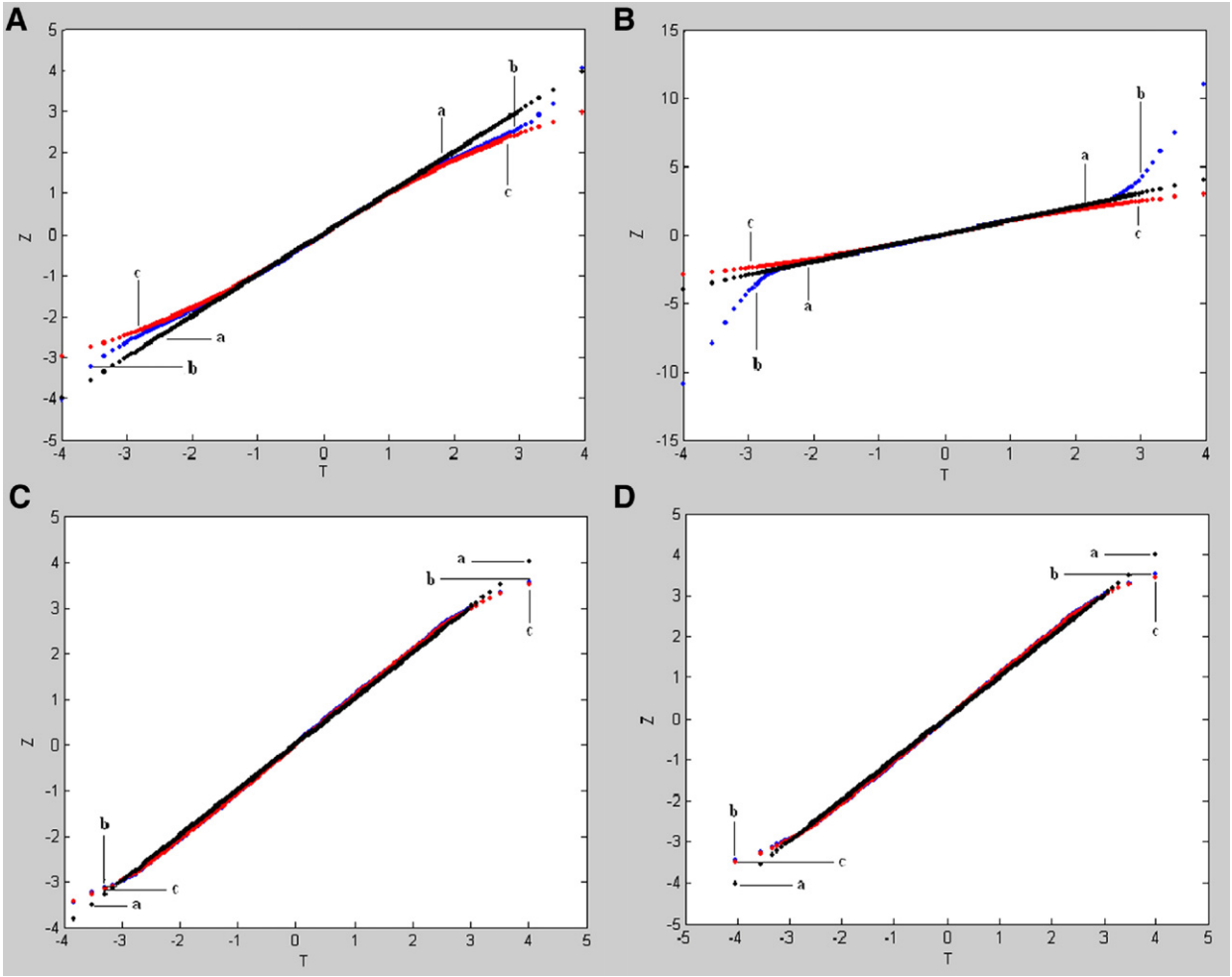

Fig. 4. Plots of Z values vs T values. The simulated Z values were obtained from the first (curve c), second (curve b in A and C), and third (curve b in B and D) sets of the simulated data of 3000 genes. In the first, second, and third simulation sets, treatment effect values of $G=0R$, $10R$, and $30R$ were randomly assigned to 30% of the genes, where R is a random variate in the uniform distribution (>0,1] (see text for simulation). The simulated T values (curve a) were a set of 3000 null scores (see text). The results shown in (A) and (B) were obtained by the permutation approach and those shown in (C) and (D) were based on the RS approach.

that the Z distribution, as an empirical distribution, produced by the RS approach is a desirable approximation of the null distribution and in particular it is independent of treatment effect or sample size, which is essential for the rank test.

Estimate of FDR

Since it is generally unknown if a given gene is expressed differently under two different conditions, it is not necessarily best to use real data of gene expression to evaluate an FDR estimator. Therefore, we also conducted a computer simulation for comparing expression status (significance or insignificance) of a gene identified by a method with its real status. In this simulation study, we also generated two data sets of 3000 genes in which treatment effect values of 10R were randomly assigned to 10 and 30% of the genes, respectively, and sample size was set to be 6 replicates. This simulation procedure was iterated 20 times. Four criteria, i.e., absolute average, maximum and minimum, and variance of differences between the estimated and the true numbers of the false discoveries across all $\hat{R}_{FD}(i)\% \le \lambda$ obtained from these 20 two-sample simulated data sets were used to assess an estimator. We set $\lambda = 40$, 30, 20, 10, and 5%. Table 1 summarizes the results obtained by applying RAM and SAM (the software comes from http://www-stat.stanford.edu/~tibs/SAM/) to these simulated data sets in the situations of 10 and 30% of the genes given effect values of 10R. The results shown in Table 1 clearly indicate that the RAM estimator has a much better accuracy in estimating FDR than the SAM estimator. In particular, for FDR of 5%, which is an important threshold value in practice, the RAM estimate is, on average, 0.65 false discoveries with variance <1 and variation interval of 1–3 false discoveries, whereas the SAM estimate is, on average, about 2 false discoveries with variance larger than 6 and variation interval of 7 false discoveries. Fig. 2 shows the whole profile of the RAM estimates of FDRs over all given thresholds based on the second simulation data set. In this profile, the estimated and true curves are well agreed, suggesting that the RAM estimate is reliable.

Table 1
Difference between the estimated and the true false discoveries at $\hat{R}_{FD}\% \le \lambda$ obtained by SAM and RAM from the simulated microarray data of 3000 genes

Method	λ	Absolute average	Variance	Maximum	Minimum
30% of genes received treatment effect values of G = 10R					
RAM	40	3.021	18.787	16	−17
	30	2.398	9.659	7	−8
	20	2.119	7.677	6	−8
	10	1.363	3.554	4	−7
	5	0.649	0.739	2	−1
SAM	40	5.309	55.240	11	−25
	30	3.406	18.915	11	−9
	20	3.044	16.582	11	−9
	10	2.209	9.214	6	−4
	5	1.850	8.684	6	−1
10% of genes received treatment effect values of G = 10R					
RAM	40	1.961	7.219	8	−5
	30	1.471	3.963	6	−3
	20	1.046	1.835	3	−3
	10	0.641	0.763	2	−2
	5	0.300	0.333	0	−1
SAM	40	3.182	18.129	12	−11
	30	2.468	9.873	8	−4
	20	2.048	7.268	7	−3
	10	1.826	6.909	7	0
	5	1.667	6.705	7	0

Table 2
The results obtained by SAM and RAM from the simulated microarray data sets of 3000 genes of which 30% were given treatment effect values of 8R and 30% of the expression noise followed a gamma distribution and the rest followed a normal distribution

SAM				RAM					
Δ_i	$N(i)$	$\hat{N}_f(i)$	$\hat{R}_{FD}(i)$ %	Δ_i	$N(i)$	$\hat{N}_f(i)$	$N_f(i)$	$\hat{R}_{FD}(i)$ %	$R_{FD}(i)$ %
0.00050	1127	1160	102.9	0.0676	1821	1296	1279	71.2	70.2
0.01035	351	292.5	83.3	0.0851	1715	1199	1199	69.9	69.9
0.01217	350	286	81.7	0.1025	1660	1147	1157	69.1	69.7
0.01943	346	282	81.5	0.1374	1505	753	1040	50.0	69.1
0.02073	345	277	80.2	0.1724	1372	462	938	33.7	68.4
0.02932	315	248	78.7	0.1900	1288	369	875	28.6	67.9
0.03368	311	239.5	77.0	0.2251	1137	295	764	25.9	67.2
0.04193	308	230	74.6	0.2428	984	232	657	23.6	66.8
0.05636	301	218.5	72.5	0.2782	802	172	523	21.4	65.2
0.06288	289	206.5	71.4	0.3138	401	82	230	20.4	57.4
0.06851	256	178	69.5	0.3677	102	21	21	20.6	20.6
0.08908	153	99	64.7	0.3858	99	20	20	20.2	20.2
0.11274	135	83	61.4	0.4039	97	19	19	19.6	19.6
0.12576	127	75.5	59.4	0.4405	95	18	18	18.9	18.9
0.13374	124	73	58.8	0.4589	92	17	17	18.5	18.5
0.14284	123	70	56.9	0.4775	89	16	16	18.0	18.0
0.14912	120	68	56.6	0.4961	87	15	15	17.2	17.2
0.15884	88	45.5	51.7	0.5337	83	14	15	16.9	18.1
0.16771	86	43	50.0	0.5909	79	12	13	15.2	16.5
0.18042	85	42	49.4	0.6103	77	11	13	14.3	16.9
0.18894	80	38	47.5	0.6494	74	10	11	13.5	14.9
0.19440	74	35	47.2	0.6692	72	10	10	13.9	13.9
0.19750	73	35	47.9	0.6892	71	9	9	12.7	12.7
0.20497	71	33	46.4	0.7502	68	8	9	11.8	13.2
0.20650	48	22	45.8	0.7918	65	6	8	9.2	12.3
0.21360	44	20	45.4	0.8344	64	6	7	9.4	10.9
0.21474	39	18	46.1	0.8560	61	6	5	9.8	8.2
0.21516	38	17	44.7	0.8779	60	5	5	8.3	8.3
0.21815	26	11	42.3	0.9226	57	5	4	8.8	7.0
0.22433	19	7	36.8	1.0156	47	3	4	6.4	8.5
0.23141	19	7	36.8	1.0893	42	3	2	7.1	4.8
0.23953	15	6	40.0	1.1147	41	3	2	7.3	4.9
0.27464	14	4	28.5	1.1939	40	2	2	5.0	5.0
0.45645	5	1	20.0	1.3691	36	2	2	5.6	5.6
1.02531	1	1	100.0	1.4011	34	2	1	5.9	2.9
1.04895	0	1	NA	1.4340	32	1	1	3.1	3.1
				2.2410	24	0	0	0	0

$N_f(i)$ and $R_{FD}(i)$ are the true number and the rate of false discoveries according to the comparison between identified and true genes differentially expressed in the simulated data.

Identification of differentially expressed genes

The exact distribution of the expression level of a gene is unknown in microarray experiments. For some genes, normal distributions may be appropriate, while for others gamma distribution may be more accurate, and for some none of the

standard distributions may be adequate. When many thousands of genes are examined simultaneously, a variety of distributions is likely present. Therefore, it is appropriate to evaluate a method using data generated from a mixture of distributions. For simplicity, we limited ourselves in the simulation to using gamma and normal distributions to yield data sets consisting of 3000 genes in two samples each having six replicates. Then at random we mixed them together at a given proportion (for example, 30% gamma distribution and 70% normal distribution) to construct a new set of microarray data. We applied SAM and RAM to the simulation data set. The results are summarized in Fig. 1B and Table 2, in which the exchangeability (fudging) factor $S_0 = 10.75$ at percentile 33%. One can find in Fig. 1B that all dots on the plots are close to the expected lines, suggesting that SAM fails to work on such data, whereas the other result in Table 2 shows that RAM works very well for identifying genes that are significantly differentially expressed and for the estimation of FDR.

Application to the real microarray data

Both SAM and RAM were applied to the two-sample real microarray data of 7129 genes obtained from two small samples (four replicates for each sample) provided in the SAM software package. The results shown in Table 3 are helpful for explaining the observation in Table 1 of Tusher et al. [16]. A larger S_0 ($S_0 = 3.3$) is the primary cause for SAM's poor performance: 12% FDR in the 48 genes identified to be significant at threshold $\Delta = 1.2$. It can be seen from Table 3 that RAM found 61 genes having significant expressional change at an acceptable FDR level of 3.3%, whereas SAM identified only 21 genes at an acceptable FDR level of 4.7%. The difference of 40 genes between both is because of an unnecessarily larger fudging factor ($S_0 = 3.4$) used in SAM. Indeed, some of 40 genes have $d > \sigma < 1$, suggesting that a large value of S_0 indeed led some truly differentially expressed genes to be missed by SAM.

Table 3
Numbers of genes called significant and of the false discoveries estimated by SAM and RAM from the observed microarray data sets of 7129 genes in four replicate experiments provided in the SAM software

SAM ($S_0 = 3.46$ at percentile = 0.01)				RAM			
Δ_i	$N(i)$	$\hat{N}_f(i)$	$\hat{R}_{FD}(i)$ %	Δ_i	$N(i)$	$\hat{N}_f(i)$	$\hat{R}_{FD}(i)$ %
0.00676	4046	3736.4	92.3	0.04641	6834	5060	74.0
0.02311	4011	3682.4	91.8	0.10520	6392	4261	66.7
0.03355	3952	3621.4	91.6	0.16402	5956	3964	66.6
0.04874	3933	3591.4	91.3	0.22289	5539	1993	36.0
0.07252	3893	3551.5	91.2	0.28185	5144	1656	32.2
0.08402	3882	3536.5	91.1	0.34092	4736	1389	29.3
0.08731	3855	3499.5	90.7	0.40013	4188	1430	34.1
0.08885	3305	2955.7	89.4	0.45952	3752	1043	27.8
0.08977	3211	2879.7	89.6	0.51911	3245	553	17.0
0.09132	1936	1716.2	88.6	0.57893	2660	617	23.2
0.09278	1751	1529.3	87.3	0.63901	1795	100	5.6
0.09538	1739	1510.3	86.8	0.69939	1480	110	7.4
0.09691	1718	1487.3	86.5	0.76010	1220	71	5.8
0.09886	1703	1464.3	85.9	0.82118	783	61	7.8
0.10159	752	568.2	75.5	0.88266	310	17	5.5
0.10943	739	550.2	74.4	**0.94457**	**61**	**2**	**3.3**
0.11610	599	436.8	72.9	1.00697	58	2	3.4
0.12068	531	383.3	72.1	1.06988	57	1	1.8
0.13922	410	268.3	65.4	1.13336	57	1	1.8
0.15164	352	218.4	62.0				
0.18234	268	155.9	58.1				
0.19807	261	147.9	56.6				
0.20716	213	116.9	54.9				
0.33398	167	64.9	38.9				
0.43301	124	39.9	32.2				
0.57814	88	19.4	22.1				
0.65578	74	12.9	17.5				
0.76837	62	9.9	16.1				
0.86358	46	5.9	13.0				
1.24876	36	2.9	8.3				
1.38245	26	1.9	7.6				
1.60219	**21**	**0.9**	**4.7**				
2.03175	12	0.9	8.3				
2.43241	11	0.9	9.0				
2.69035	3	0.9	33.3				
4.19555	0	0.9	NA				

Discussion

In conventional statistical resampling, permutation is a popular approach to estimate a null distribution. However, as seen from our analysis and as indicted in Appendix A, the distribution-free method based on permutations would be generally biased because for microarray data analysis small sample sizes limit the number of distinct permutation samples and ranking the T statistics at each permutation does not completely remove the treatment effect contributing to gene-expression variations. The RS approach is developed in this paper to circumvent the aforementioned problems of SAM. The resulting RAM has the advantage of being insensitive to the treatment effect that often presents in real data and having a better estimate of FDR. Another important advantage of RAM is that it works well for small sample sizes which is particularly useful for analyzing microarray data that often have small sample sizes. In addition, the RS approach can be easily extended to the pair data set (see Appendix B).

FDR is often used to control error rate in the BH procedure [18] and in SAM [16,22]. In practice, for a multiple-test method based on t statistics, it is important to obtain an accurate estimate of FDR. In SAM, the FDR estimate is realized through the permutation approach in which fluctuations around expectation occur among permutated samples. The fluctuations would be impacted on by the data themselves, i.e., sample size, treatment effect, and data noise. The RAM estimator of FDR is based on a two-simulation strategy so that it avoids these impacts on the estimate of FDR. Our simulation results indicate that the RAM estimator of FDR is generally accurate at a given threshold of interest.

In an idealized setting in which all expression levels are normally distributed, SAM and RAM both work well for identifying differentially expressed genes. However, in the case in which most of the expression levels follow a normal distribution and a small fraction, for example, 30% of the genes, possibly follow a gamma distribution, SAM performs poorly or

even fails to work due to a larger fudge factor S_0 whereas RAM continues to performs well. In addition, small sample size makes it possible to produce sample variances far smaller than 1 in a large-scale gene-expression profile. This situation, as seen in Tusher et al. [16], also produces a larger fudging factor for SAM, but in RAM this fudging impact can effectively be excluded.

Acknowledgments

This research was supported by grants from the U.S. National Institutes of Health, R01 NS41466 (M.F.) R01 HL69126 (M.F.), and R01 GM50428 (Y.F.) and by funds from Yunnan University. We thank the High Performance Computer Center of Yunnan University for computational support and Sara Barton for editorial assistance.

Appendix A

Suppose we have two classes $X_k=\{x_{k1},\ldots, x_{km})$ and $Y_k=\{y_{k1},\ldots,y_{km})$ of m replicates for gene k. A permutation produces two resampling classes $X_k'=\{x_{k1},\ldots, x_{km-r},y_{k1},\ldots,y_{kr}\}$ and $Y_k'=\{x_{k1},\ldots,x_{kr}, y_{k1},\ldots,y_{km-r}$. From these resampling class data, we have two resampling means,

$$\bar{X}_k{}' = \frac{1}{m}\left(\sum_{j=1}^{m-r} x_{kj} + \sum_{j=1}^{r} y_{kj}\right) = \frac{1}{m}\sum_{j=1}^{m-r} x_{kj} + \frac{1}{m}\sum_{j=1}^{r} y_{kj}, \quad \text{(A1a)}$$

$$\bar{Y}_k{}' = \frac{1}{m}\left(\sum_{j=1}^{r} x_{kj} + \sum_{j=1}^{m-r} y_{kj}\right) = \frac{1}{m}\sum_{j=1}^{r} x_{kj} + \frac{1}{m}\sum_{j=1}^{m-r} y_{kj}. \quad \text{(A1b)}$$

Let $x_{kj}=\mu_k+\tau_{xk}+e_{xkj}$ and $y_{kj}=\mu_k+\tau_{yk}+e_{ykj}$, where μ_k is overall mean (expectation) for expression levels of gene k, τ_{xk} and τ_{yk} are assumed to be treatment effects contributing to the expression variation of gene k, and e_{xkj} and e_{ykj} are expression noises. Thus, these two means can also be expressed as

$$\bar{x}_k{}' = \mu_k + \frac{1}{m}\left[(m-r)\tau_{xk} + r\tau_{yk}\right] + \frac{1}{m}\sum_{j=1}^{m-r} e_{xkj} + \frac{1}{m}\sum_{j=1}^{r} e_{ykj}, \quad \text{(A2a)}$$

$$\bar{y}_k{}' = \mu_k + \frac{1}{m}\left[r\tau_{xk} + (m-r)\tau_{yk}\right] + \frac{1}{m}\sum_{j=1}^{r} e_{xkj} + \frac{1}{m}\sum_{j=1}^{m-r} e_{ykj}, \quad \text{(A2b)}$$

where r is the number of exchanged members between two classes. It is clear that with the difference between $\bar{X}_k'$ and $\bar{Y}_k'$, the treatment effect difference is

$$d(\tau_k) = \frac{1}{m}\left[(m-r)\tau_{xk} + r\tau_{yk}\right] - \frac{1}{m}\left[r\tau_{xk} + (m-r)\tau_{yk}\right] = 0$$

if $r=m/2$, otherwise, $d(\tau_k)\neq 0$. In addition, the rank of Z values across all positions at each permutation changes the Z values in position k^* in the rank space, so that the component dealing with $d(\tau_{k^*})$ in the Z value in position k^* in the rank space, that is, $\frac{1}{M}\sum_{J=1}^{M} /d\left(\tau_{k^*}^J\right)/\sigma_{k^*}^J \neq 0$, *where* $d\left(\tau_{k^*}^1\right)/\sigma_{k^*}^1 > 0,\ldots,$ $d\left(\tau_{k^*}^M\right)/\sigma_{k^*}^M > 0$ *or* $d\left(\tau_{k^*}^1\right)/\sigma_{k^*}^1 < 0,\ldots,d\left(\tau_{k^*}^M\right)/\sigma_{k^*}^M < 0, \sigma_{k^*}^j$ is a pooled standard deviation of two samples in position k^* at permutation J. This indicates that the Z distribution obtained by the permutation approach contains treatment effect differences for the microarray experiments if $r\neq m/2$. This is why a large treatment effect on expression levels of a part of the genes leads to an obviously "positive deviation" of the Z distribution obtained by the permutation approach from the null distribution as seen in Figs. 3A, 4A, and 4B, say, $Z_k^* \geq T_{k^*} \geq 0$ or $Z_k^* \leq T_{k^*} \leq 0$, where $T_{k^*}=d(e_{k^*})/\sigma_{k^*}$ is a null score of the T statistic.

For no treatment effect, i.e., $\tau_{xk}=\tau_{yk}=0$, and for small sample size for gene k, $\sum e_k \geq 0$ or $\sum e_k \leq 0$, and hence, Eqs. (A1a) and (A1b) are changed to

$$\begin{aligned}\bar{x}' &= \frac{1}{m}\sum_{j=1}^{m-r} e_{xkj} + \frac{1}{m}\sum_{j=1}^{r} e_{ykj} \\ &= \frac{1}{m}\sum_{j=1}^{m} e_{xkj} - \frac{1}{m}\sum_{j=1}^{r} e_{xkj} + \frac{1}{m}\sum_{j=1}^{r} e_{ykj} \\ &= \bar{e}_{xk} - \bar{e}_{xk}(r) + \bar{e}_{yk}(r), \end{aligned} \quad \text{(A3a)}$$

$$\begin{aligned}\bar{y}_k{}' &= \frac{1}{m}\sum_{j=1}^{m-r} e_{ykj} + \frac{1}{m}\sum_{j=1}^{r} e_{xky} \\ &= \frac{1}{m}\sum_{j=1}^{m} e_{ykj} - \frac{1}{m}\sum_{j=1}^{r} e_{ykj} + \frac{1}{m}\sum_{j=1}^{r} e_{xky} \\ &= \bar{e}_{yk} - \bar{e}_{yk}(r) + \bar{e}_{xk}(r). \end{aligned} \quad \text{(A3b)}$$

In the difference between $\bar{X}_k'$ and $\bar{Y}_k'$, there is an error difference,

$$\begin{aligned} d(\varepsilon_k) &= (\bar{e}_{xk} - \bar{e}_{yk}) - [\bar{e}_{xk}(r) + \bar{e}_{xk}(r)] + [\bar{e}_{yk}(r) + \bar{e}_{yk}(r)] \\ &= d(e_k) - 2\bar{e}_{xk}(r) + 2\bar{e}_{yk}(r) \\ &= d(e_k) + 2d[e_k(r)], \end{aligned} \quad \text{(A4)}$$

where $d(e_k)=e_{xk}-e_{yk}$ and $d[e_k(r)]=\bar{e}_{yk}(r)-\bar{e}_{xk}(r)$. It is clear from Eq. (A4) that $d(\varepsilon_k)\neq d(e_k)$ if $d[e_k(r)]\neq 0$. On the other hand, due to and $\bar{e}_{xk}(r)\in\bar{e}_{xk}$ and $\bar{e}_{yk}(r)\in\bar{e}_{yk}$, $d[e_k(r)]=\bar{e}_{yk}(r)-\bar{e}_{xk}(r)$ is negatively related to $d(e_k)=\bar{e}_{xk}-\bar{e}_{yk}$, that is, if $d(e_k)>0$, then $d[e_k(r)]\leq 0$ or if $d(e_k)<0$, then $d[e_k(r)]\geq 0$. Again, the rank of the Z value across all positions leads to $d[e_{k^*}^1(r)]/\sigma_{k^*}^1 \geq 0$,...,$d(e_{k^*}^M))/\sigma_{k^*}^M \geq 0$ or $d(e_{k^*}^1(r))/\sigma_{k^*}^1 \leq 0,\ldots,$ $d[e_{k^*}^M(r)]/\sigma_{k^*}^M \leq 0$, consequently, the average of $d[e_{k^*}^J(r)]/\sigma_{k^*}^J$ in position k^* over all permutations is larger or less than or equal to 0, that is, $\frac{1}{M}\sum_{J=1}^{M} d\left[e_{k^*}^J(r)\right]/\sigma_{k^*}^J \geq 0$ *or* $\frac{1}{M}\sum_{J=1}^{M} d\left[e_{k^*}^J\right)(r)]/\sigma_{k^*}^J \leq 0$, which then results in a "negative deviation" of the Z distribution from the null distribution as seen in Figs. 3A, 4A, and 4B, i.e., $Z_{k^*} \geq T_{k^*} \leq 0$ or $Z_{k^*} \leq T_{k^*} \geq 0$.

Appendix B

For paired data, since two samples of m_k observed values $(x_{1k},\ldots, x_{m_k k})$ and $(y_{1k},\ldots,y_{m_k k})$ become a sample of m_k distant values $(d_{1k},\ldots,d_{m_k k})$, $k=1,\ldots, N$, the sample m_k of replicates for distances can also be at random cut into two subsamples. Let

$d_{ik}=x_{ik}-y_{ik}=d_k+e_{xik}-e_{yik}=d_k+e_{ik}$, $i=1,\ldots, m_k$, where d_k is the difference between treatment effects on the expression of gene k. We then have $\bar{d}_k = \sum_{i=1}^{m_k}(d_k+e_{ik})/m_k = d_k + \bar{e}_k$. In two subsamples at split J, two subsample means are expressed as $\bar{d}^J_{1k}=d_k+\bar{e}^J_{x1k}-\bar{e}^J_{y1k}$ and $\bar{d}^J_{2k}=d_k+\bar{e}^J_{x2k}-\bar{e}^J_{y2k}$, where $\bar{e}^J_{gik}$ is the average of errors in subsample i at split J for gene k in the system g ($g=x, y$). Therefore, $\bar{e}_k$ is estimated by

$$\begin{aligned}\bar{e}^J_k &= \frac{1}{2}\left(\bar{d}^J_{1k}-\bar{d}^J_{2k}\right)\\ &= \frac{1}{2}\left[d_k+\bar{e}^J_{x1k}-\bar{e}^J_{y1k}-d_k-\bar{e}^J_{x2k}+\bar{e}^J_{y2k}\right]\\ &= \frac{1}{2}\left[\left(\bar{e}^J_{x1k}-\bar{e}^J_{y1k}\right)-\left(\bar{e}^J_{x2k}-\bar{e}^J_{y2k}\right)\right]\\ &= \frac{1}{2}\left[\left(\bar{e}^J_{x1k}-\bar{e}^J_{x2k}\right)+\left(\bar{e}^J_{y1k}-\bar{e}^J_{y2k}\right)\right],\end{aligned}$$

say, $\bar{e}_k$ in the paired data is equivalent to that in the unpaired data. The null score of the T statistic is estimated by the Z -value:

$$Z^j_k = \frac{\bar{d}_k}{\sqrt{\frac{\sigma^2(d_k)}{m_k}}} = \frac{\bar{e}^J_k}{\sqrt{\frac{\sigma^2(d_k)}{m_k}}},$$

where $\sigma^2(d_k)$ is the sample variance of distances between two paired data for gene k.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.08.003.

References

[1] M.K. Kerr, M. Martin, G.A. Churchill, Analysis of variance for gene expression microarray data, J. Comput. Biol. 7 (2000) 819–839.
[2] L. Li, J.W. Jiang, X. Li, K.L. Moser, Z. Guo, L. Du, Q. Wang, E.J. Topol, Q. Wang, S. Rao, A robust hybrid between genetic algorithm and support vector machine for extracting an optimal feature gene subset, Genomics 85 (2005) 16–23.
[3] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 4 (2003) 1157–1182.
[4] E.P. Xing, M.I. Jordan, R.M. Karp, Feature selection for high-dimensional genomic microarray data, in: Machine Learning: Proceedings of the Eighteenth International Conference, San Francisco, Morgan Kaufmann, San Mateo, CA, 2001.
[5] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997) 273–324.
[6] I. Tsamardinos, C.F. Aliferis, Towards principled feature selection: relevance, filters and wrappers, in: Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, 2003.
[7] L. Wolf, A. Shashua, D. Geman, Feature selection for unsupervised and supervised inference: the emergence of sparsity in a weight-based approach, J. Mach. Learn. Res. 6 (2005) 1855–1887.
[8] P.J. Park, M. Pagano, M. Bonetti, A nonparametric scoring algorithm for identifying informative genes from microarray data, Pac. Symp. Biocomput. (2001) 52–63.
[9] L. Li, X. Li, Z. Guo, Efficiency of two filters for feature gene selection, Life Sci. Res. **7** (2003) 372–396 (in Chinese).
[10] X. Cui, J.T.G. Hwang, J. Qiu, N.J. Blades, G.A. Churchill, Improved statistical tests for differential gene expression by shrinking variance components, Biostatistics 6 (2005) 59–75.
[11] W. Pan, J. Lin, C.T. Le, How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach, Genome Biol. 3 (2002) (Research0022).
[12] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 54–70.
[13] Y. Hochberg, A sharper Bonferroni procedure for multiple tests of significance, Biometrika 75 (1988) 800–803.
[14] P. Westfall, S. Young, Resampling-Based Multiple Testing, Wiley, New York, 1993.
[15] F.E. Turkheimer, C.B. Smith, K. Schmidt, Estimation of the number of "true" null hypotheses in multivariate analysis of neuroimaging data, NeuroImage 3 (2001) 920–930.
[16] V.G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the ionizing radiation response, Proc. Natl. Acad. Sci. USA 98 (2001) 5116–5121.
[17] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J.R. Stat. Soc. Ser. B 57 (1995) 289–300.
[18] Y. Benjamini, D. Drai, G. Elmer, N. Kafkfi, I. Golani, Controlling the false discovery rate in behavior genetics research, Behav. Brain Res. 125 (2001) 279–284.
[19] Y. Benjamini, W. Liu, A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence, J. Stat. Plan. Inference 82 (1999) 163–170.
[20] J.D. Storey, A direct approach to false discovery rates, J.R. Stat. Soc. Ser. B 64 (2002) 479–498.
[21] J.D. Storey, R. Tibshirani, Statistical significance for genome wide studies, Proc. Natl. Acad. Sci. USA 100 (2003) 9440–9445.
[22] X. Cui, G.A. Churchill, Statistical tests for differential expression in cDNA microarray experiments, Genome Biol. 4 (2003) 210–219.
[23] C.-A. Tsai, H.-M. Hsueh, J.J. Chen, Estimation of false discovery rates in multiple testing: application to gene microarray data, Biometrics 59 (2003) 1071–1081.
[24] S. Pounds, C. Cheng, Improving false discovery rate estimation, Bioinformatics 20 (2004) 1737–1745.