



# A Comparison of Procedures for Controlling the False Discovery Rate in the Presence of Small Variance Genes: A Simulation Study

Dan Lin, Ziv Shkedy, Tomasz Burzykowski, Willem Talloen, Luc Bijnens

## ► To cite this version:

Dan Lin, Ziv Shkedy, Tomasz Burzykowski, Willem Talloen, Luc Bijnens. A Comparison of Procedures for Controlling the False Discovery Rate in the Presence of Small Variance Genes: A Simulation Study. Communications in Statistics - Simulation and Computation, 2009, 38 (10), pp.2111-2122. 10.1080/03610910903249510 . hal-00524862

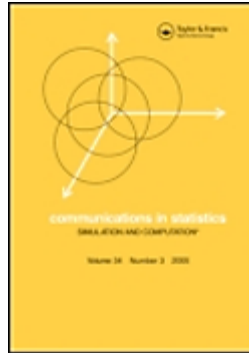
**HAL Id: hal-00524862**

**<https://hal.science/hal-00524862>**

Submitted on 9 Oct 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



### A Comparison of Procedures for Controlling the False Discovery Rate in the Presence of Small Variance Genes: A Simulation Study

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2009-0049.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	05-Aug-2009
Complete List of Authors:	Lin, Dan; Hasselt University Shkedy, Ziv; Hasselt University Burzykowski, Tomasz; Hasselt University Talloen, Willem; Janssen Pharmaceutical Bijnsens, Luc; Janssen Pharmaceuticals
Keywords:	False Discovery Rate, Fudge Factor, SAM, Power
Abstract:	The Significance Analysis of Microarrays (Tusher et al. 2001) is widely used in analyzing microarray data. It introduces a fudge factor to adjust test statistic by deflating the large value of test statistics due to the small variances. Lin et al. (2008) pointed out that the fudge factor does not effectively improve the power and control the FDR in the presence of small variance genes. Motivated by those results, we extend our study to compare several methods for choosing the fudge factor and use simulation studies to investigate the power and the control of the FDR of the considered methods.
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
fudge3.tex	



For Peer Review Only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# A Comparison of Procedures for Controlling the False Discovery Rate in the Presence of Small Variance Genes: A Simulation Study

Dan Lin<sup>1</sup>, Ziv Shkedy<sup>1</sup>, Tomasz Burzykowski<sup>1</sup>, Willem Talloen<sup>2</sup>, and Luc Bijnen<sup>2</sup>

1. I-Biostat, Hasselt University, Universitaire Campus, Building D,  
B 3590 Diepenbeek, Belgium

dan.lin@uhasselt.be

2. J&JPRD, Biometrics and Clinical Informatics,  
Beerse, Belgium

Key Words: Fudge factor; SAM; control of the FDR; and power.

## ABSTRACT

The Significance Analysis of Microarrays (SAM, Tusher *et al.*, 2001) method is widely used in analyzing gene expression data while controlling the FDR by using resampling-based procedure in the microarray setting. One of the main components of the SAM procedure is the adjustment of the test statistic. The introduction of the fudge factor to the test statistic aims at deflating the large value of test statistics due to the small standard error of gene-expression. Lin *et al.* (2008) pointed out that the fudge factor does not effectively improve the power and the control of the FDR as compared to the SAM procedure without the fudge factor in the presence of small variance genes. Motivated by the simulation results presented in Lin *et al.* (2008), in this paper, we extend our study to compare several methods for choosing the fudge factor in the modified  $t$ -type test statistics and use simulation studies to investigate the power and the control of the FDR of the considered methods.

## 1. INTRODUCTION AND NOTATION

The Significance Analysis of Microarrays (SAM) is widely used in analyzing gene expression data while empirically controlling the FDR based on a resampling-based procedure. One of the main components of the SAM procedure is the adjustment of the test statistic. The introduction of the fudge factor to the test statistic aims at deflating the large value

of test statistics due to the small standard error of gene-expression. However, the values of the modified test statistic decrease simultaneously for all the genes, and in particular, the fudge factor affects the truly differentially expressed genes with a small standard error. We noticed that the fudge factor does not effectively improve the power and the control of the FDR as compared to the SAM procedure without the fudge factor in the presence of the small variance genes (Lin *et al.*, 2008). To be specific, the part of the FDR (denoted as the  $FDR^1$  related to the non-differentially expressed genes with small variances) is inflated, while the proportion of the FDR (denoted as the  $FDR^0$  for the non-differentially expressed genes with moderate and large variances) is controlled at the desired significance level.

In this paper, we extend our study to include several methods, which adjust the test statistic in a similar way as the SAM procedure. We consider the penalized  $t$ -test statistic (Wu, 2005), selection of the fudge factor by using receiver operating characteristics curve (ROC, Broberg, 2003) and Empirical Bayes (EB, Efron and Tibshirani, 2002). We compare the performance of these methods together with the original SAM procedure in terms of power and the control of the FDR by using a simulation study.

The contents of the paper is organized as follows. In Section 2 we summarize the SAM procedure and briefly discuss results of the case study and simulation studies presented by Lin *et al.* (2008). Motivated by those results, in Section 3, we describe another three methods for the selection of the fudge factor. In Section 4, we discuss the use of simulation data to study these methods and compare the results in terms of the power and control of the FDR. The paper ends with a discussion in Section 5.

2. SELECTION OF THE FUDGE FACTOR IN THE SAM

2.1 THE SETTING

We focus on finding differentially expressed genes by using microarray data from two groups/conditions, namely the control and the treatment group. Let  $X_{ijk}$  be the gene-expression of  $i$ the gene for array  $j$  in treatment group  $k$  ( $i = 1, \dots, m; j = 1, \dots, n; k = 0$  (control), 1 (treatment)). Let  $z_{ijk}$  be an indicator variable which take the value of 1 if expression level of gene  $i$  in the microarray is obtained under the treatment and zero otherwise,

i.e.,

$$\begin{cases} z_{ijk} = 0 & k = 0 \quad \text{control,} \\ z_{ijk} = 1 & k = 1 \quad \text{treatment.} \end{cases}$$

The gene-specific linear model is given by

$$X_{ijk} = \beta_{0i} + \beta_{1i}z_{ijk} + \varepsilon_{ijk}; \quad \varepsilon_{ijk} \sim N(0, \sigma_i^2). \quad (1)$$

Here,  $\beta_{1i}$  is the effect of treatment on gene  $i$ , while  $\beta_{0i}$  corresponds to the mean of the control group for gene  $i$ . The difference between the two groups for gene  $i$  can be tested using the two sample  $t$ -test

$$t_i = \frac{\bar{X}_{i0} - \bar{X}_{i1}}{s_i \sqrt{\frac{1}{n_0} + \frac{1}{n_1}}}, \quad (2)$$

where  $\bar{X}_{i0} = \frac{\sum_{j=1}^{n_0} X_{ij0}}{n_0}$ ,  $\bar{X}_{i1} = \frac{\sum_{j=1}^{n_1} X_{ij1}}{n_1}$ , and  $n_0$  and  $n_1$  are the number of arrays in the control and treatment group, respectively. The pooled variance estimator,  $s_i$  is defined as:

$$s_i^2 = \frac{1}{n_0 + n_1 - 2} \left\{ \sum_{j=1}^{n_0} (X_{ij0} - \bar{X}_{i0})^2 + \sum_{j=1}^{n_1} (X_{ij1} - \bar{X}_{i1})^2 \right\}.$$

## 2.2 The SAM PROCEDURE AND THE FUDGE FACTOR

The SAM (Tusher *et al.*, 2001, Storey and Tibshirani, 2001) is a testing procedure for microarray analysis, which estimates the FDR by using permutations under the assumption that all null hypotheses are true. The procedure consists of three components: (1) the adjusted test statistic, (2) the approximation of the distribution of the test statistic based on permutations, and (3) the control of the FDR.

The  $t$ -test statistic for each gene  $i$  is modified in the SAM procedure as follows:

$$t_i^{SAM} = \frac{\bar{X}_{i1} - \bar{X}_{i0}}{s_i + s_0} = \frac{\Delta_i}{s_i + s_0}, \quad (3)$$

The constant  $s_0$  in (3) is called the fudge factor. It is calculated as the percentile of the gene-wise standard errors that minimizes the coefficient of variation (CV) of the median absolute deviance (MAD) of the SAM test statistics using (3). This modification is used to overcome bias for genes with expression difference  $\Delta_i$  close to zero, which have a large value

of the test statistic due to a small sample variance. By using an inflated standard error of the test statistics, the SAM addresses the problem of the dependence of the value of the test statistic on the variance of expression levels for a particular gene.

The SAM is a resampling-based procedure and uses permutations to approximate the null distribution of the test statistic. Thus, the choice of the test statistic does not have an effect on the SAM procedure. This distribution-free property in the SAM procedure allows to include in it any form of the test statistic. The permutation matrix  $\mathbf{T}$  consists of element  $t_{ib}$  which is the test statistic for the  $i$ th gene in the  $b$ th permutation (Ge *et al.*, 2003; Lin *et al.*, 2008). The permutation-based raw  $p$ -values can be calculated as

$$P_i = \frac{\#(b : |t_{ib}| \geq |t_i|)}{B}, \quad (4)$$

where  $t_i$  is the observed value of the test statistic for gene  $i$ . This is the the basis for the permutation approach, which adjusts  $p$ -values obtained above by using the Benjamini and Hochberg procedure for controlling the FDR (BH-FDR).

The control of the FDR is performed, once the permutation matrix  $\mathbf{T}$  is obtained. The SAM procedure requires the test statistics of each permutation to be sorted for all the genes such that the first row of the sorted matrix  $\mathbf{T}^{SAM}$  is the minimum test statistic across permutations and the last row is the maximum. The element  $t_{(i)b}$  in  $\mathbf{T}^{SAM}$  is the ordered test statistic in permutation  $b$ , i.e.,  $t_{(1)b} \leq t_{(2)b} \leq \dots \leq t_{(m)b}$ . The expected values of the observed test statistics are approximated by the means of the rows of  $\mathbf{T}^{SAM}$ , i.e.,  $\bar{t}_{(i)} = \frac{1}{B} \sum_{b=1}^B t_{(i)b}$ .

To call a gene significant, the difference between the observed and expected values of the test statistic needs to be larger than a certain cut-off value  $\lambda$  (Parmigiani *et al.*, 2003). For a grid of  $\lambda$  values, the corresponding number of significant genes can be listed; at the same time, the number of false positives, arising from any permutation matrix  $\mathbf{T}^{SAM}$ , is estimated. Under the null hypotheses, we expect that no differentially expressed genes are present for each permutation. Consequently, the median or 90 percentile number of false positives, corresponding to  $\lambda$ , can be obtained from the permutation matrix. In this way, the FDR can be calculated for each value of  $\lambda$  and an acceptable value of  $\lambda$  can be chosen to control the FDR at the desired level. Note that the SAM procedure estimates the proportion

of true null hypotheses and obtain the FDR ( $\pi_0 E(V/R) = (m_0/m)E(V/R)$ , where  $m_0$  is the number of non-differentially expressed genes and estimated from permutations),  $V$  is the number of false positives,  $R$  is the number of discoveries, and  $\pi_0 = m_0/m$  is the proportion of true null hypotheses.

In order to investigate the effect of the SAM fudge factor on the protection against genes with small variance, we decompose the number of true null hypotheses ( $m_0$ ) into two types (see Figure 1):  $m_0^0$  truly non-differentially expressed genes with a moderate to large variance and  $m_0^1$  truly non-differentially expressed genes with a relatively small variance (i.e.,  $\leq 5\%$  percentile of variance in the data). Accordingly, the falsely rejected true hypotheses  $V$  are divided into sets of  $V^0$  and  $V^1$  for these two types of genes, respectively. Consequently, the FDR can be decomposed as  $\text{FDR} = \text{FDR}^0 + \text{FDR}^1 = E(V^0/R) + E(V^1/R)$ , where  $R$  is the total number of genes declared significant (Lin *et al.*, 2008).

Hypothesis	Test result		Total
	Accept $H_0$	Reject $H_0$	
$H_0$ True	$U$	$V$ $\swarrow V^0$ $\searrow V^1$	$m_0$ $\nearrow m_0^0$ $\searrow m_0^1$
$H_0$ False	$T$	$S$	$m_1$
Total	$W$	$R$	$m$

Figure 1: Decision in multiple testing (Benjamini and Hochberg 1995).

As mentioned above, a fudge factor  $s_0$  in the denominator of the test statistic is used to adjust the test statistic. How the fudge factor is calculated is described in SAM users guide and technical document (Chu *et al.*, 2001).

### 2.3 RESULTS OF THE SIMULATION (PRESENTED IN LIN *ET AL.* 2008)

Lin *et al.* (2008) conducted the simulation study to investigate the performance of the SAM procedure in terms of power and control for the FDR in the presence of small variance genes. The simulation study was set up in the way mimicking the real life data, in which the gene-specific variances and artificial treatment effects (assuming to be large enough after



multiplicity adjustment) were used to generate gene expression levels under three treatment groups and one control.

The fudge factor in the SAM procedure aims to protect against small variance genes to result in false positives. However, the simulation study showed that by decomposing the FDR into the  $FDR^0$  (the part of false positives among the truly differentially expressed genes) and the  $FDR^1$  (the part of false positives caused by the non-differentially expressed genes but with small variances), the  $FDR^0$  was controlled at the nominal level (say 0.05) in irrespective of the proportion of small variance genes; while the  $FDR^1$  was not controlled but inflated as the number of small variance genes increased. Figure 2 shows each compositions of the FDR under two settings (no small variance genes and 10% small variance genes present in the data).

Motivated by this results, we extend the simulation study and consider three more methods in selecting the fudge factor in the  $t$ -type test statistic and compare their performance in terms of control of the  $FDR^0$ ,  $FDR^1$ , and the power.

### 3. OTHER METHODS FOR SELECTING THE FUDGE FACTOR

In this section, we consider other methods for the selection of the fudge factor: the penalized linear regression, the operating characteristic curve (ROC), and the empirical Bayes.

#### 3.1 PENALIZED LINEAR REGRESSION

Wu (2005) proposed the penalized linear regression as a method for the detection of differentially expressed genes. The penalized  $t$ -test statistic can be written as

$$t_i^* = \text{sign}(\bar{X}_{i0} - \bar{X}_{i1}) \frac{(|\bar{X}_{i0} - \bar{X}_{i1}| - \varphi)_+}{\sqrt{(1/n_0 + 1/n_1)s_i^2 + \varphi^2/(n-2)}}, \quad (5)$$

The shrinkage parameter  $\varphi$  is chosen to minimize the correlation between the penalized  $t$ -test statistics in (5) and  $s_i$  for all genes. It is easy to note that  $\varphi$  is applied to both the nominator and denominator and shrinks the absolute mean difference in the numerator by  $\varphi$ . The denominator is shrunken by  $\varphi^2/(n-2)$ . Thus, the advantage of the penalized  $t$ -statistic

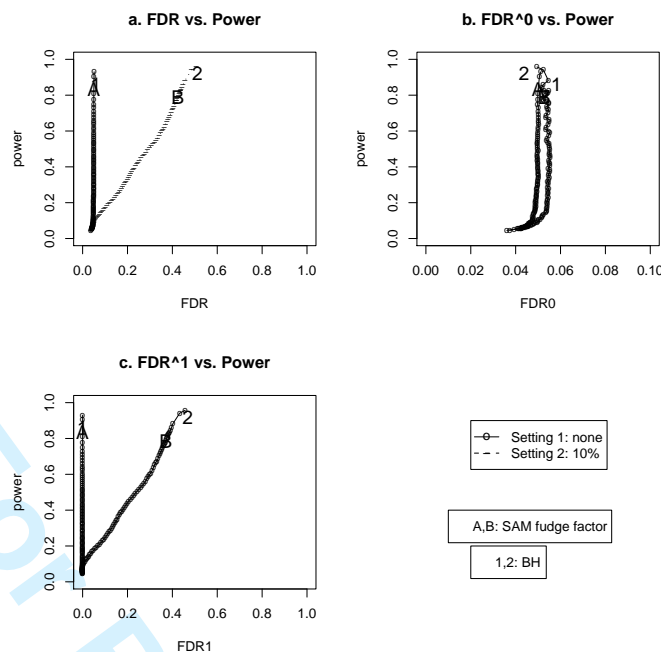


Figure 2: *a. FDR vs. power, b.  $FDR^0$  vs. power, c.  $FDR^1$  vs. power under the two settings. Black line: setting 1 ( $m_0^1=0$ ); green line: setting 2 ( $m_0^1=5100$ ). The power and FDR were obtained for a grid of 100 fudge factors. A and B represent the power and FDR by using the fudge factor selected by the SAM (which minimizes the CV of the SAM test statistics), while 1 and 2 represent the power and FDR obtained by using the BH-FDR procedure.*

is that it shrinks the absolute mean differences to zero, deflating large values of test statistics due to the small standard error.

Wu (2005) argued that, for a given value of  $\varphi$ , the penalized  $t$ -test statistic has the same property as the SAM  $t$ -test statistic. Wu (2005) proposed to replace the SAM test statistic with the penalized  $t$ -test statistic and to control the FDR by using the SAM procedure.

### 3.2 ROC METHOD

Broberg (2003) proposed a method to jointly optimize the number of genes that are falsely declared positive and the number of genes that are falsely declared negative by optimizing over a range of the FDR and fudge factor values. For this purpose, he utilized the receiver

operating characteristic curve (ROC curve), which plots the false positive rate (FPR) against the false negative rate (FNR). The FPR is defined as the number of false positives among the significant genes divided by the total number of genes. Similarly, the FNR is defined as the number of false negatives divided by the total number of genes.

For each combination of the FDR and of the fudge factor  $s_0$ , the procedure calculates the number of false positives and false negatives. The criterion to be used for the choice of the fudge factor is to minimize  $C = \sqrt{\text{FPR}^2 + \text{FNR}^2}$ . Due to the large amount of calculations, the optimization of  $C$  is performed over a lattice of  $(\text{FDR}, s_0)$  values. Lovell *et al.* (1996) and Genovese and Wasserman (2001) present other versions of the minimizing criterion. For example, one of the criteria incorporates weight  $\omega$  in order to express the relative importance of the FPR compared to the FNR, what results in  $C = \sqrt{\omega^2 \text{FPR}^2 + \text{FNR}^2}$ .

According to Storey (2001), the false positive probability estimator can be defined as  $\widehat{\text{FPR}} = \hat{p}_0 \alpha$ , which represents the proportion of non-differentially expressed genes  $\hat{p}_0$  multiplied by the probability  $\alpha$  that such a gene produces a significant result. The estimator of the false negative probability equals  $\widehat{\text{FNR}} = 1 - \hat{p}_0(1 - \alpha) - \hat{p}(\alpha)$ , where  $\hat{p}(\alpha)$  is the relative frequency of genes that have a  $p$ -value less than  $\alpha$ . By using the estimates  $\widehat{\text{FNR}}$  and  $\widehat{\text{FPR}}$ , corresponding to the significance level  $\alpha$ , we can obtain the fudge factor  $s_0$  that minimizes  $C$ . Then, the SAM procedure is carried out with the corresponding value of  $s_0$ .

### 3.3 THE EMPIRICAL BAYES APPROACH

Efron *et al.* (2001) and Efron and Tibshirani (2002) modeled the distribution of the adjusted test statistics  $t_i$  in (3), as a mixture of two components. One component corresponds to the differentially expressed genes, and the other to the non-differentially expressed genes. Denoting the density of the former by  $f_1$  and the latter by  $f_0$ , the mixture density of the test statistics is given by

$$f(t) = \pi_0 f_0(t) + (1 - \pi_0) f_1(t), \quad (6)$$

where  $\pi_0$  is the prior probability that a gene is differentially expressed. Application of Bayes' rule to (6) results in the posterior probability that a gene with test statistic  $t$  is differentially

expressed, i.e.,

$$p_i(t) = 1 - \pi_0 \frac{f_0(t)}{f(t)}. \quad (7)$$

Following Efron *et al.* (2001) and Efron and Tibshirani (2002), a gene is called differentially expressed if its posterior probability (7) is larger than or equal to 0.9. The FDR for the resulting rejection region  $\Gamma = \{t : p_i(t) \leq 0.9\}$  is then estimated by using a resampling approach as follows,

$$\widehat{\text{FDR}}(\Gamma) = \hat{\pi}_0 \frac{\#\{t_{ib} \in \Gamma\}/B}{\max\{\#\{t_i \in \Gamma\}, 1\}},$$

where  $B$  is total number of permutations.

Efron *et al.* (2001) suggested to specify the optimal choice of the fudge factor by running the EB procedure for several values of  $s_0$ , and by selecting the value of  $s_0$  that leads to the largest number of differentially expressed genes.

#### 4. SIMULATION STUDY

In this section we investigate the performance of the SAM procedure with variants of the procedure resulting from using the fudge factor, obtained by using the penalized  $t$ -test (T), the ROC (R), and the SAM procedure (S). Additionally, we consider the use of the EB (E) and based on permutation  $p$ -values controlling the FDR by using the BH-FDR procedure (B).

##### 4.1 SIMULATION SETTING

We generate 10,000 genes on each array. In total, 100 data sets are generated with 16 arrays under two conditions (control and treatment) with eight arrays for each condition. We assume that 10% of the 10,000 genes are truly differentially expressed. The mean gene-expression for the two groups is defined as follows:

$$\mu_{ik} = \begin{cases} \delta_{ik} \times \sigma_i & i \leq 1000 \text{ and } k = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

It is assumed that  $\delta_{ik} \sim U(2.8, 4.5)$ , while  $\sigma_i$  is randomly sampled from the estimated standard error distribution observed in the case study analyzed by Lin *et al.* (2008). The means

of treatment groups for genes with small variance are generated from normal distribution  $N(0, 0.3)$  for the four treatment groups. We consider four settings, in which the number of small variance genes is equal to zero, 500, 1000, and 2000. The variance of these genes is chosen to be equal to 0.0036, which is the 5% percentile of the variances in the observed case study data and is expected to be small to inflate the test statistics.

## 4.2 SIMULATION RESULTS

### 4.2.1 THE FPR VS. FNR

First, we study the relationship between the estimated mean FPR and FNR of the five methods under the four settings across 100 simulation datasets, where the FPR and FNR are defined in Section 3.2. As we observe in Figure 3, the optimal results which minimizing  $C = \sqrt{FPR^2 + FNR^2}$  is obtained for the SAM procedure, except that in the first setting (as with no small variance genes) the SAM and permutation approach show similar results. As the proportion of small variance genes increases (i.e.,  $m_0^1 = 500, 1000, \text{ and } 2000$  genes corresponding to 5%, 10%, and 20% of the total number of genes, respectively), the FPR and FNR increase. The permutation approach yields a higher FPR as compared to the other procedures, because the procedure does not adjust the test statistic to protect against the small variance genes. On the other hand, the penalized  $t$  approach tends to yield a higher FNR, because the penalized  $t$ -test statistic downplays the truly differentially expressed genes as well as non-differentially expressed genes with small variance. In-between these methods, coincidentally, the ROC and EB methods show similar results, but the difference between them seems to increase as the number of the small variance genes increases.

### 4.2 THE FDR VS. POWER

To make a distinction between genes that are falsely declared differentially expressed due to a small variance from the rest of truly non-differentially expressed genes, we decomposed the FDR as  $FDR = FDR^0 + FDR^1 = E(V^0/R) + E(V^1/R)$ , where  $V^0$  is the number of truly non-differentially expressed genes with moderate to big variances and  $V^1$  is the number of truly non-differentially expressed genes with small variances ( $\leq 0.0036$  in our case). In what

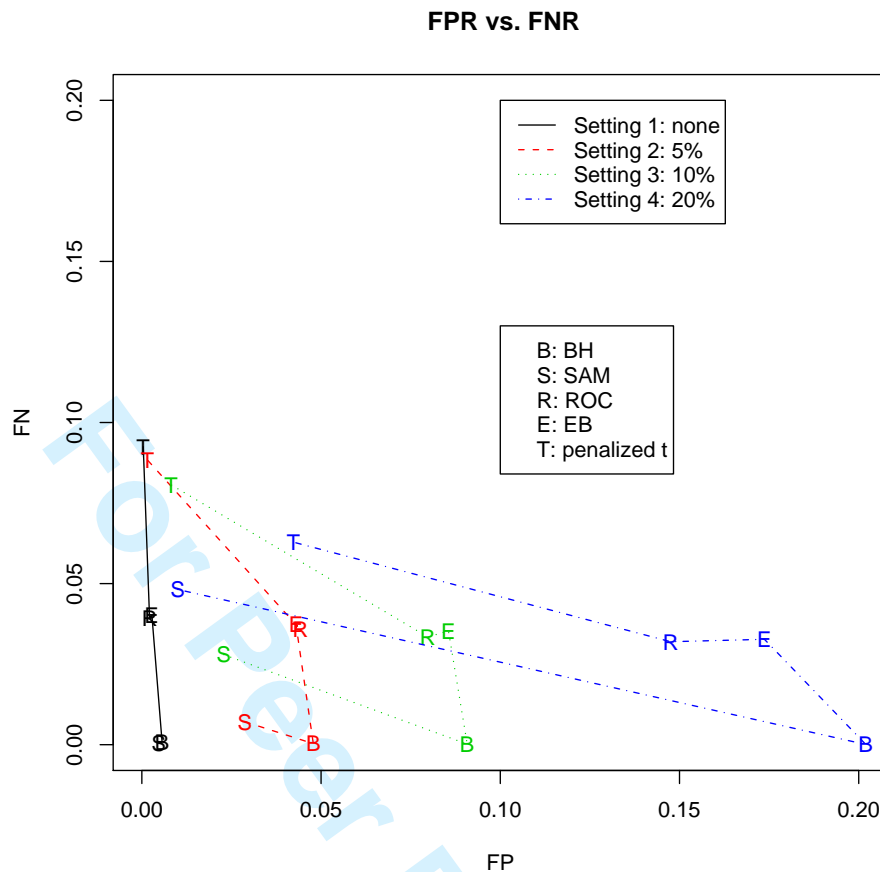


Figure 3: *FPR vs. FNR* using the five methods under the four settings. Black line: setting 1 ( $m_0^1=0$ ); red line: setting 2 ( $m_0^1=500$ ); green line: setting 3 ( $m_0^1=1000$ ); blue line: setting 4 ( $m_0^1=2000$ ). B: permutation  $p$ -values adjusted by the BH-FDR procedure, S: SAM, R: ROC, E: EB, and T: penalized  $t$ .

follows, we investigate the control of the FDR (including  $FDR^0$  and  $FDR^1$ ) and power for the considered methods.

From Figure 4a we can see that the (overall) FDR is not well controlled at the desired level of 0.05 for all the five methods. The SAM procedure seems to yield the lowest FDR, while the ROC and EB approaches yield a similar, high value of the FDR. More importantly, taking SAM as the reference procedure, we can make two sets of comparisons: (1) between the SAM, penalized  $t$ , and ROC methods that utilize the same approach to control the FDR,

but differ in the choice of the fudge factor, and (2) between the SAM, EB, and permutation approaches, which differ not only in the choice of the fudge factor, but also regarding the approach used to control the FDR.

For the first set of results, as  $m_0^1$  increases, the FDR (Figure 4a) increases for the penalized  $t$  (from 0.1, to 0.22 and to 0.52) and ROC methods (from 0.4, to 0.57, and to 0.7), and the power increases for the penalized  $t$  approach (from 0.1, to 0.21, and to 0.4) and it is maintained for the ROC approach (about 0.6). The SAM procedure behaves differently: it maintains the FDR at around 0.2 at the price of losing the power (from 0.9, to 0.7, and to 0.5). Using the decomposition of FDR into the  $FDR^0$  (shown in Figure 4b) and  $FDR^1$  (presented in Figure 4c), we observe different patterns for the  $FDR^0$  for the three methods. For the small variance genes, the  $FDR^0$  is controlled at 0.05 for the SAM and ROC methods under the four settings, but not for the penalized  $t$ , for which it increases as the number of small variance genes increases. The penalized  $t$ -test statistic is modified by adding a constant to both the denominator and numerator, which results in declaring more non-differentially expressed genes with a moderate or large variance as significant. On the other hand, the penalized  $t$  approach reduces the  $FDR^1$ , in particular when 5% or 10% non-differentially genes with small variance are present. However, it does not do so for a larger proportion of such genes (i.e., 20%). In all, we can conclude that the SAM procedure is most successful in controlling the overall FDR at the level about 0.2, as compared to other approaches, but it does not control the FDR at the desired level of 0.05. The main component of the FDR for the SAM and ROC approaches is the  $FDR^1$ . The penalized  $t$  method has a tendency to control the  $FDR^1$ , but it loses the control of the  $FDR^0$ . We also note that the variability of the FDR and  $FDR^0$  for the penalized  $t$  approach is large, as shown in Figures 5a and 5b.

For the second set of methods (the SAM, EB, and permutation approach), from Figure 4a we can observe that the FDR for the three methods is not controlled at the desired level of 0.05. The EB approach yields the largest FDR among the three methods and a constant power (around 60%). The SAM approach shows a decreasing FDR (0.25, 0.24, and 0.18, respectively) and decreasing power (90%, 70%, and 50%, respectively) as the proportion of

non-differentially expressed genes with a small variance increases (from 5% to 10%, and to 20%). For the permutation approach, the FDR (0.35, 0.5, and 0.7) increases, but the power is consistently close to 1. In Figure 4b, the  $FDR^0$  for the three methods is around 0.05. Hence, the  $FDR^1$  shows the same pattern as the overall FDR.

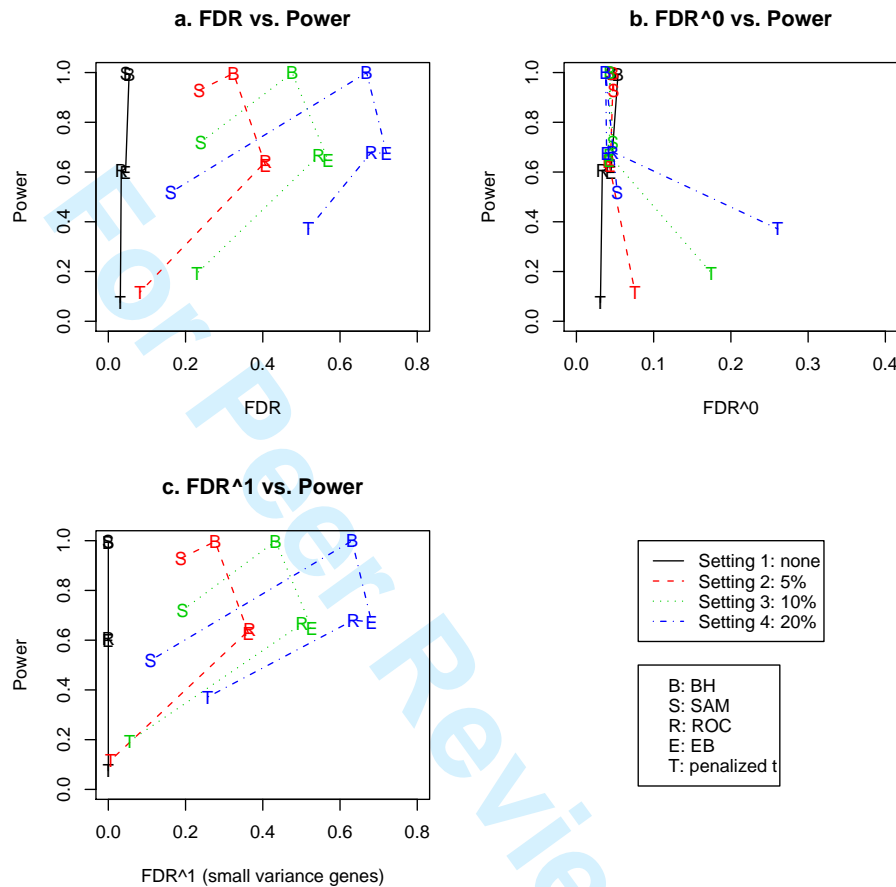


Figure 4: Panel a. FDR vs. power, panel b.  $FDR^0$  vs. power, panel c.  $FDR^1$  vs. power under the four settings. Black line: setting 1 ( $m_0^1=0$ ); red line: setting 2 ( $m_0^1=500$ ); green line: setting 3 ( $m_0^1=1000$ ); blue line: setting 4 ( $m_0^1=2000$ ). S: SAM, B: permutation p-values adjusted BH-FDR, E: empirical Bayes, R: ROC, and T: penalized t approach.

Figure 5 shows the power of the five methods. We can conclude that the EB and ROC methods yield not only a similar power, but also a similar FDR. Both methods produce rather consistent values of the FDR and power regardless of the proportion of non-differentially



expressed genes with small variance. The permutation approach performs slightly better than the EB and ROC methods, yielding a lower FDR and a higher power. The worst performance with respect to the power is observed for the penalized  $t$  approach. This method tends to control the  $FDR^1$ , but to lose control of the  $FDR^0$ . It can be regarded as the best method to choose the fudge factor for controlling the FDR, but at the cost of losing power.

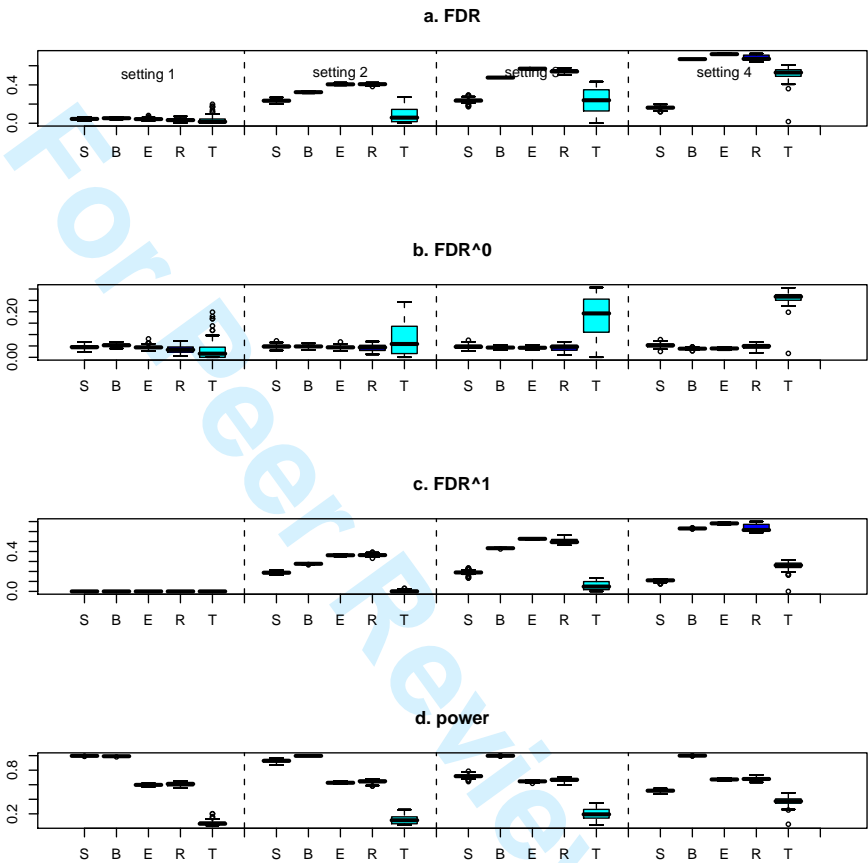


Figure 5: Panel a. Boxplot of FDR, panel b. Boxplot of  $FDR^0$ , panel c. Boxplot of  $FDR^1$ , panel d. Boxplot of power under the four settings. Black line: setting 1 ( $m_0^1=0$ ); red line: setting 2 ( $m_0^1=500$ ); green line: setting 3 ( $m_0^1=1000$ ); blue line: setting 4 ( $m_0^1=2000$ ). S: SAM, B: permutation p-values adjusted BH-FDR, E: empirical Bayes, R: ROC, and T: penalized  $t$  approach.

## 5. CONCLUSION AND DISCUSSION

The question of interest in this paper is the capability of the SAM procedure to control the FDR for truly non-differentially expressed genes with a small variance. We have shown in the simulation study for the SAM procedure (Tusher *et al.* 2001) that the overall FDR cannot be controlled even when the proportion of truly non-differentially expressed genes with a small variance is relatively small (i.e., 5% of the total number of genes in the array). Moreover, we have shown that, when the FDR is decomposed into the  $FDR^0$  and  $FDR^1$ , there is no problem to control the  $FDR^0$  (regardless of the proportion of non-differentially expressed genes with a small variance). However, the  $FDR^1$ , associated with the small variance genes, is not well controlled. When the proportion of non-differentially expressed genes with a small variance increases, the SAM procedure with the fudge factor either loses power or does not control the FDR at the desired level. Thus, the automatic selection of the fudge factor does not guarantee the power and the control of the FDR in the SAM approach.

We have also investigated several other methods for selecting the fudge factor. We have shown that none of these methods controls the FDR at the desired level. The SAM procedure is most successful in controlling the FDR but at the price of losing power. The penalized  $t$  approach shows poor performance in the control of both the FDR and power, with a poor control over truly non-differentially expressed genes with a moderate or large variance. Coincidentally, the EB and ROC approaches, which differ in the methods used to control the FDR and to choose the fudge factor, yield similar and consistent results with respect to the power and FDR, regardless of the proportion of non-differentially expressed genes with a small variance.

Note that the proportion of non-differentially expressed genes with a small variance ( $m_0^1$ ) is unknown in the case study. Mixture modeling approach (Delmar *et al.*, 2003; Manda *et al.*, 2007) have been proposed to estimate this proportion. None of the methods considered in the simulation study takes this proportion into consideration. The possibility of choosing the fudge factor with the information of the proportion of small variance genes and of modifying the test statistics for genes with different variances should be explored further. The aim of

such an approach would be to achieve a good balance between the control of the FDR and power.

ACKNOWLEDGEMENT

Financial support from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

BIBLIOGRAPHY

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2002) A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. *Bioinformatics* **19**, 185–193.

Broberg, P. (2003) Statistical methods for ranking differentially expressed genes. *Genome Biology*, **4**, R41.

Chu, G., Narasimhan, B., Tibshirani, R., Tusher, V. (2001), SAM “Significance Analysis of Microarrays, Users guide and technical document.

Delmar, P., Robin, S., Tronik-LeRoux, D., Daudin, J-J. (2005) Mixture Model on the variance for the analysis of gene expression data, *J. R. Stat. Soc. C.*, **54**, 31–50.

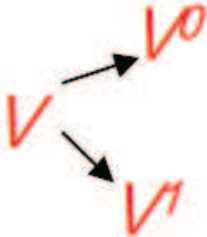
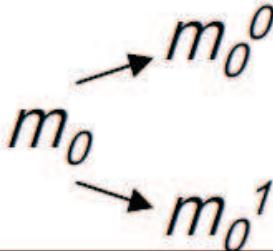
Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. *Ann. Stat.*, **32**, 407–499.

Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V. (2001) Empirical Bayes Analysis of a Microarray Experiment, *Journal of the American Statistical Association*, **96**, 1151–1160.

Efron, B., and Tibshirani, R. (2002) Empirical Bayes Methods and False Discovery Rates for Microarrays, *Genetic Epidemiology*, **23**, 70–86.

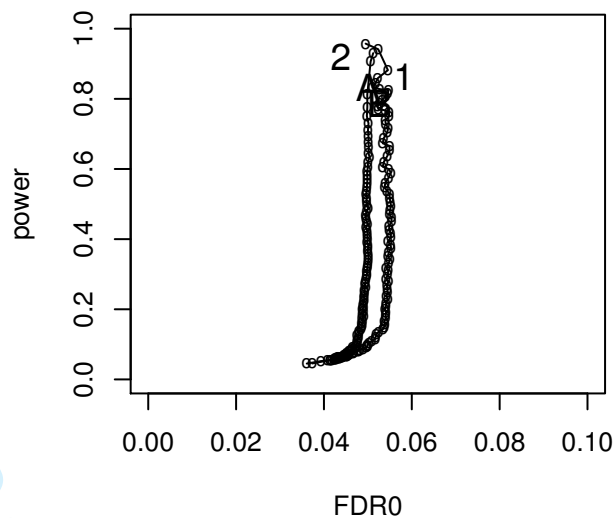
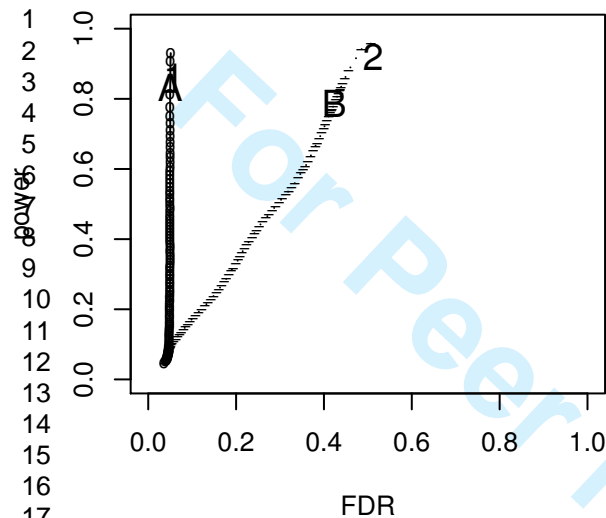
Genovese, C. and Wasserman, L. (2001) Operating characteristics of the FDR procedure. *Technical Report*. New York, Carnegie Mellon University 2001.

- Lin D., Shkedy Z., Burzykowski T., Göhlmann H. W.H., De Bondt A., Perera T., Geerts T., and Bijmans L. (2008) Significance analysis of microarray (SAM) for comparisons of several treatments with one control. *Biometric Journal, MCP*, **50**(5), 801–823.
- Lovell, D.R., Dance, C.R., Niranjana, M., Prager, R.W., Dalton, K.J. (1996) Ranking the effect of different features on the classification of discrete valued data. In *Engineering Applications of Neural Networks*. Kingston on Thames, London, 487–494.
- Manda, S.O., Walls, R.E., and Gilthorpe, M.S. (2007) A full Bayesian hierarchical mixture model for the variance of gene differential expression. *BMC Bioinformatics*, **8**, 124.
- Storey, J.D. (2001) A direct approach to false discovery rates. *Technical Report*. Stanford, CA: Stanford University.
- Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal statistical society, series B*, 58.
- Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response, *Proceedings of the National Academy of Sciences*, **98**, 5116–5121.
- Wu, T.D. (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.* **195**, 53–65.

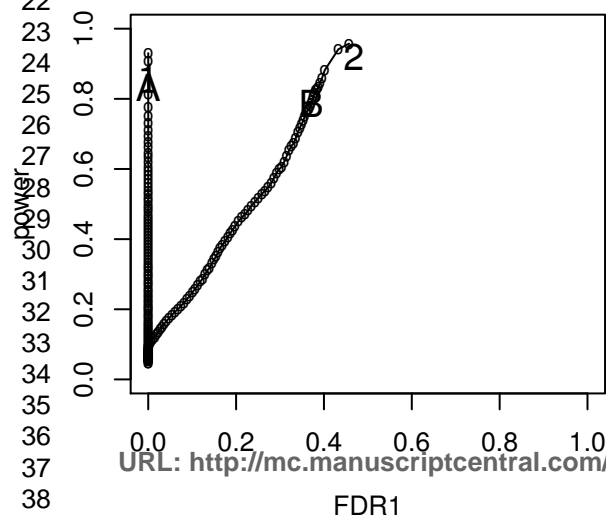
Hypothesis	Test result		Total
	Accept $H_o$	Reject $H_o$	
$H_o$ True	$U$		
$H_o$ False	$T$	$S$	$m_1$
Total	$W$	$R$	$m$

a. FDR vs. Power

b. FDR<sup>0</sup> vs. Power



c. FDR<sup>1</sup> vs. Power

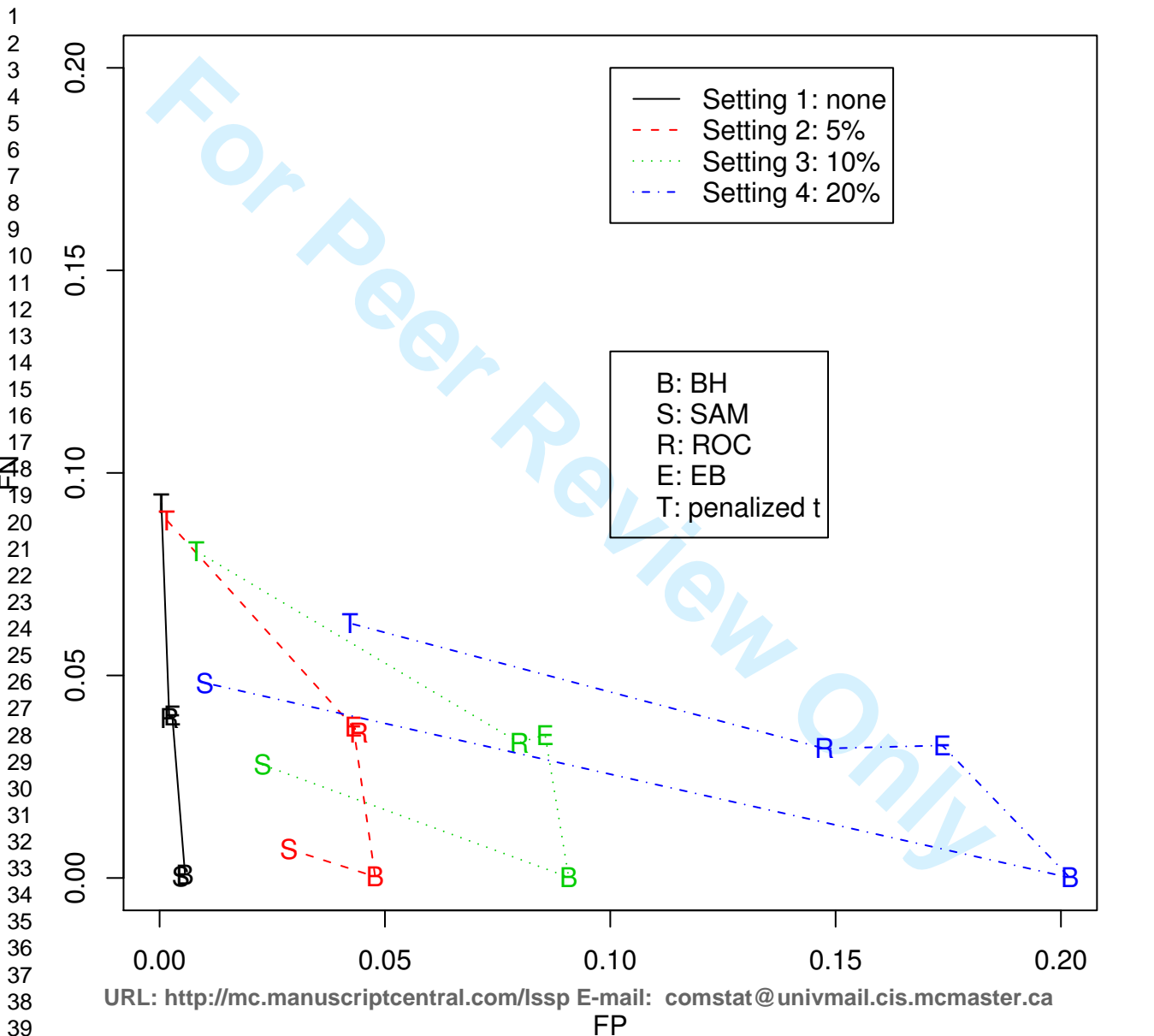


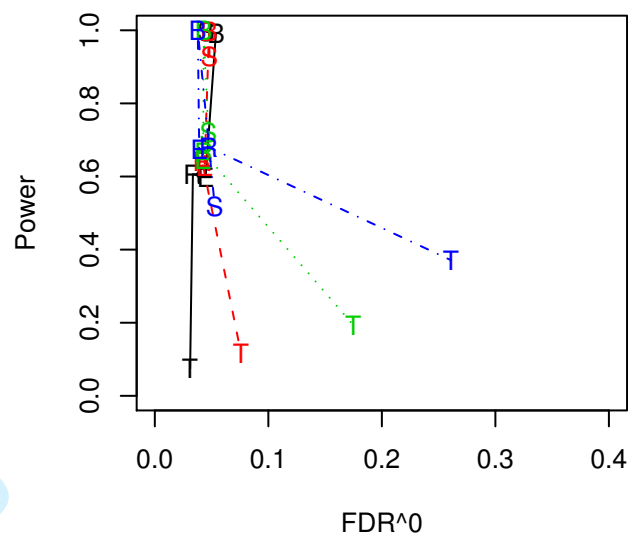
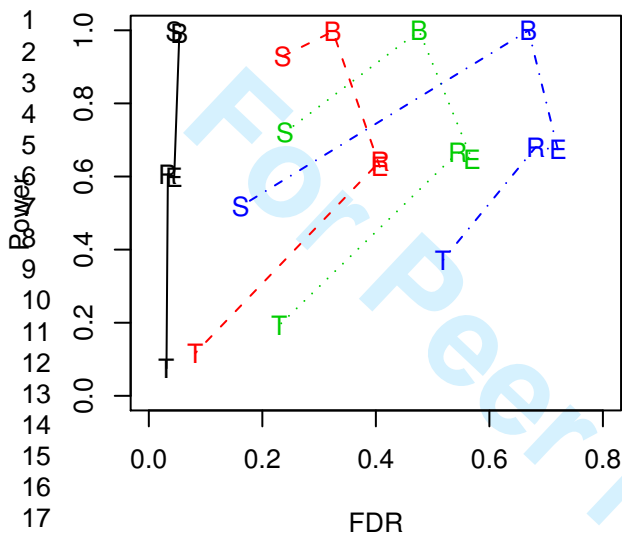
—○— Setting 1: none  
- - - Setting 2: 10%

A,B: SAM fudge factor

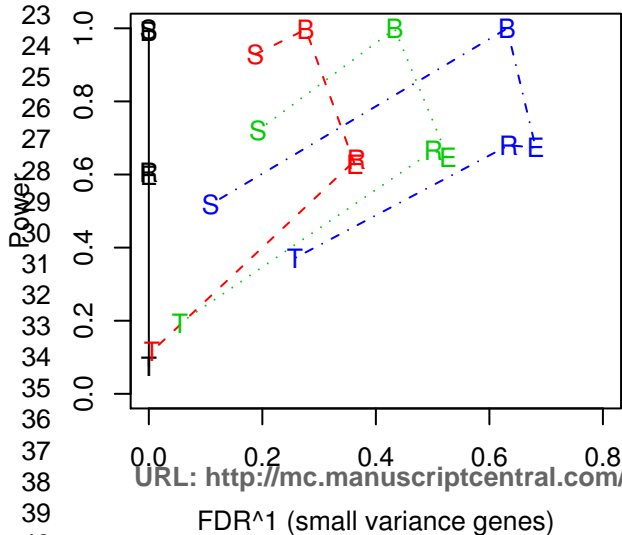
1,2: BH

FPR vs. FNR





c. FDR<sup>1</sup> vs. Power

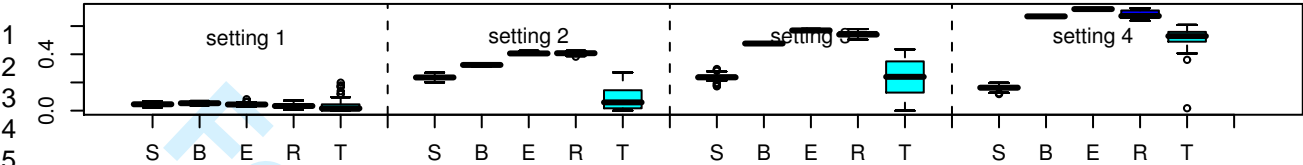


— Setting 1: none  
- - Setting 2: 5%  
... Setting 3: 10%  
- . Setting 4: 20%

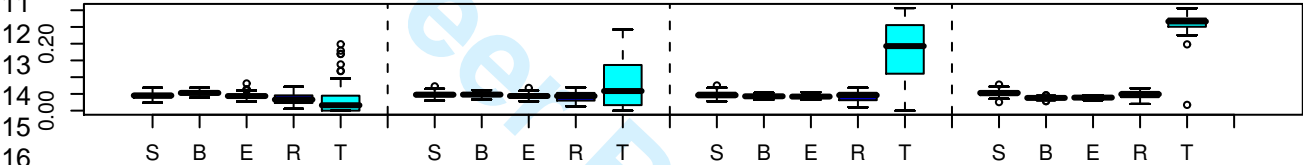
B: BH  
S: SAM  
R: ROC  
E: EB  
T: penalized t



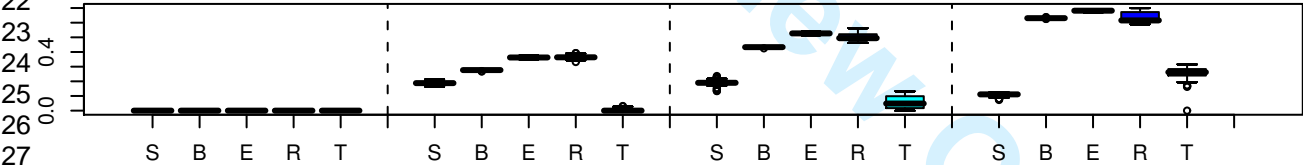
a. FDR



b.  $FDR^0$



c.  $FDR^1$



d. power

