Published in final edited form as:

J Appl Stat. 2013 February 1; 40(2): 347–357. doi:10.1080/02664763.2012.743976.

Empirical null distribution based modeling of multi-class differential gene expression detection

Xiting Cao^{1,†}, Baolin Wu^{1,*}, and Marshall I. Hertz²

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

²Department of Medicine, University of Minnesota, Minneapolis, MN, USA

Abstract

In this paper, we study the multi-class differential gene expression detection for microarray data. We propose a likelihood based approach to estimating an empirical null distribution to incorporate gene interactions and provide more accurate false positive control than the commonly used permutation or theoretical null distribution based approach. We propose to rank important genes by p-values or local false discovery rate based on the estimated empirical null distribution. Through simulations and application to a lung transplant microarray data, we illustrate the competitive performance of the proposed method.

Keywords

Empirical Bayes modeling; Empirical null distribution; False discovery rate; Gene expression data; Differential expression detection

1. Introduction

When detecting differentially expressed genes, interaction among genes could change the distribution of detection statistics across genes and compromise the accuracy and efficiency of significant gene detection. For the two-class microarray data, Efron [1, 2, 3] proposed the empirical null distribution based modeling of the two-sample t-statistic (or its variants, e.g., normally transformed) to incorporate gene interactions and accurately model the distribution of null genes, and showed its competitive performance compared to the permutation and theoretical null distribution based approach.

In this paper, we extend the empirical null modeling approach to differential gene expression detection for the multi-class microarray data. We develop the empirical null distribution based p-value and local false discovery rate to rank important genes and control false positives. The rest of the paper is organized as following. We introduce the proposed method in Section 2, and analyze a lung transplant microarray data to illustrate the proposed method in Section 3. Section 4 is devoted to simulation studies and we end with a discussion in Section 5.

2. Empirical null distribution based modeling of multi-class differential gene expression detection

Given a multi-class microarray data, denote x_{ijg} as the observed expression values for gene $i = 1, \dots, m$ from sample $j = 1, \dots, n_g$ of class $g = 1, \dots, G$. Let $n = \sum_{g=1}^G n_g$. Most commonly used methods for testing differential expressions are based on variants of the F-statistic. For gene i , it is defined as

$$t_i = \frac{\sum_{g=1}^G n_g (\bar{x}_{ig} - \bar{x}_i)^2 / (G-1)}{\hat{\sigma}_i^2},$$

where

$$\bar{x}_{ig} = \frac{\sum_{j=1}^{n_g} x_{ijg}}{n_g}, \quad g=1, \dots, G, \quad \bar{x}_i = \frac{\sum_{g=1}^G \sum_{j=1}^{n_g} x_{ijg}}{n},$$

and $\hat{\sigma}_i^2$ is the estimated expression variance. A commonly used estimate is the sample variance, $s_i^2 = \sum_{g=1}^G \sum_{j=1}^{n_g} (x_{ijg} - \bar{x}_{ig})^2 / (n - G)$. It has been widely accepted that for typical microarray data, using some moderated/stabilized variance estimate instead of the sample variance can often improve the differential gene expression detection power due to the relative small sample size compared to the large number of genes. Following Smyth [4], we estimate the expression variance with an empirical Bayes approach, which models the individual gene variances with an inverse chi-square prior distribution with scale parameter s_0^2 and degrees of freedom d_0

$$\frac{1}{\sigma_i^2} \sim \frac{1}{d_0 s_0^2} \chi_{d_0}^2.$$

The moderated variance is calculated as a weighted average of the prior information and the sample variance, $\hat{\sigma}_i^2 = ((n - G)s_i^2 + d_0 s_0^2) / (n - G + d_0)$. Here the prior parameters d_0 and s_0^2 are empirically estimated based on all genes. In the following discussion, we use the moderated F-statistic to rank important genes and compare its performance to our proposed method.

Ideally when a gene is not differentially expressed, under normal distribution assumption the moderated F statistic theoretically will follow a F-distribution with $(G - 1, n - G + d_0)$ degrees of freedom, which can be used to compute differential expression significance p-values [4]. The permutation approach can also be used to derive p-values. Let $p_{(1)} \ p_{(2)} \ \dots \ p_{(m)}$ be the ordered p-values. Benjamini and Hochberg [5] proposed the following false discovery rate (FDR) procedure to control the proportion of false positives. For a given significance level $0 < \alpha < 1$, define

$$k_\alpha = \arg \max_i \left\{ p_{(i)} \leq \frac{i}{m} \alpha \right\}.$$

When declaring all genes with $p_i \leq p_{(k_\alpha)}$ as significant, FDR will be controlled at level α . An alternative approach is to directly estimate FDR (see, e.g., Efron [6]). For example, when calling all genes with p-values less than $p_{(i)}$ as significant, the estimated FDR is

$$\widehat{\text{FDR}} = \widehat{m\theta_0} \frac{p_{(i)}}{i},$$

where the proportion of true null genes, θ_0 , needs to be estimated. Given the p-values of all genes, we follow the approach of Langaas *et al.* [7] that estimated the null gene proportion based on the p-value density fitted with a convex decreasing density model. Comparing the FDR control and estimation approaches, the main difference is the use of null gene proportion θ_0 . In general the estimation approach is more powerful than the control approach.

The commonly observed interactions among genes could change the distribution of test statistics across genes, and the theoretical null distribution in general does not provide an adequate fit for null genes. While as shown in Efron [1], the choice of null distribution is crucial in the estimation and control of FDR. With the large number of genes available, we can empirically estimate the null distribution to obtain a more accurate fit and provide appropriate FDR control.

2.1 Empirical modeling of null gene distribution

Figure 1 compares the theoretical and empirical null fitting of the moderated F-statistics for the gene expression data from a lung transplant study [8]. The study measured the gene expressions of bronchoalveolar lavage cell samples from lung transplant recipients and the goal is to find the genes related to acute rejection. Biopsies were graded from 0 to 4 for 'A' (perivascular inflammation) and 'B' (lymphocytic bronchiolitis) scores, indicating an increase in the severeness of acute rejection. The subjects are divided into three groups based on the sum of A and B scores.

For this data, the theoretical null distribution of the moderated F-statistic is F-distribution with (2, 22) degrees of freedom (estimated prior degree of freedom is $\hat{d}_0 = 1.0$). We assume that those moderated F-statistics less than the sample median are coming from null genes, and we plotted their histogram, quantile plot and empirical cumulative distribution function. In the left panel of Figure 1, the theoretical null distribution shown in dotted line deviates largely from the actual data, and provides a very poor fit. The solid line is the estimated null distribution based on a likelihood approach which will be discussed later, and clearly fits the data much better. In the middle panel, the dotted line is the quantile of theoretical null distribution against the sample quantiles of moderated F-statistics. The solid line is from the empirical null distribution, and agrees with the diagonal line very well, indicating a very good fit. In the right panel, we compare the empirical cumulative distribution function of the moderated F-statistics against the estimated theoretical null (dotted line) and empirical null (solid line) distributions. Overall the empirical null distribution provides a much better fit to the data than the theoretical null.

The maximum likelihood estimation of the empirical null distribution is based on a truncated distribution of the significant genes and assumes that all moderated F-statistics in a region

near 0 come from null genes. The assumption provides a reasonable approximation in most microarray data analysis, where researchers are only interested in detecting a relatively small number of differentially expressed genes, e.g., 10% among hundreds of thousands of candidate genes. Even if some truly significant genes have relatively small moderated F-statistic value, they would not substantially affect the estimate of the null distribution.

We empirically model the null genes with a scaled F-distribution with $(G-1, n-G+d_0)$ degrees of freedom, non-centrality parameter λ_0 and scale parameter σ_0

$$t \sim \frac{1}{\sigma_0} f\left(\frac{t}{\sigma_0}; G-1, n-G+d_0, \lambda_0\right).$$

Given a selected cutoff value C_0 , define $\mathcal{I}_0 = \{i : t_i \leq C_0\}$ and let $m_0 = \sum_{i=1}^m I(t_i \leq C_0)$ be the corresponding number of genes. Define the null gene probability

$$H_0(\lambda_0, \sigma_0) = \Pr(t \leq C_0 | \text{null}) = F\left(\frac{C_0}{\sigma_0}; G-1, n-G, \lambda_0\right).$$

Here $f(t; \nu_1, \nu_2, \lambda)$ and $F(t; \nu_1, \nu_2, \lambda)$ are the probability density and distribution functions of F-distribution with (ν_1, ν_2) degrees of freedom and non-centrality parameter λ . When assuming all moderated F-statistics less than C_0 are coming from null genes, we can compute the marginal probability

$$\theta = \Pr(t \leq C_0) = \theta_0 H_0(\lambda_0, \sigma_0).$$

The likelihood can be written as (all non-null genes are only counted as larger than C_0 and treated equally in a Binomial likelihood)

$$L(\lambda_0, \sigma_0, \theta_0) = \theta_0^{m_0} (1-\theta_0)^{m-m_0} \left[\prod_{i \in \mathcal{J}_0} \frac{f(t_i/\sigma_0; G-1, n-G+d_0, \lambda_0)}{\sigma_0 H_0(\lambda_0, \sigma_0)} \right]$$

For convenience of maximization, we did the following transformation

$$\lambda_0 = \exp(\eta), \quad \sigma_0 = \exp(\tau), \quad \theta_0 = \frac{1}{1 + \exp(-\varphi)}.$$

With no constraint on (η, τ, φ) , we can easily maximize the log likelihood numerically.

For the lung transplant microarray data, we choose C_0 as the median of moderated F-statistics, and obtain $\hat{\sigma}_0 = 3.40$, $\hat{\sigma}_0 = 0.39$ and $\hat{\theta}_0 = 0.999$. The estimated empirical null distribution fits the data very well as shown in Figure 1. The theoretical null distribution provides a very poor fit. When comparing the moderated F-statistics versus the theoretical null and estimated empirical null distributions using the Kolmogorov-Smirnov test, the reported significance p-values are 2.2×10^{-16} and 0.34 respectively.

For large scale significant gene selection, a popular approach is using p-values to rank genes. We propose to compute the empirical null distribution based p-values as follows

$$1 - F(t_i / \widehat{\sigma}_0; G-1, n-G+d_0, \widehat{\lambda}_0),$$

which can then be used to estimate and control FDR. We can similarly rank genes using p-values computed from the theoretical null distribution. In Figure 2 we compare the estimated FDR based on p-values computed from the theoretical and empirical null distributions for the lung transplant microarray data. Controlling FDR at 0.1, no gene is detected as significant using the theoretical null distribution, and we identified 19 significant genes using the empirical null distribution.

2.2 Empirical Bayes ranking with local false discovery rate

Alternatively we can use the empirical Bayes modeling approach to rank genes based on the local FDR [3]. Genes are either null or non-null with prior probability θ_0 and $\theta_1 = 1 - \theta_0$. Define h_0 as the density of the moderated F-statistics for null genes and h_1 for non-null genes. The marginal density of the moderated F-statistic is therefore $h = \theta_0 h_0 + \theta_1 h_1$. The local FDR is defined as [2, 3]

$$fdr(t) = \Pr(\text{gene is null} | t) = \theta_0 \frac{h_0(t)}{h(t)},$$

which is the posterior probability of a gene being null given its moderated F-statistic according to the Bayes rule. We estimate h_0 using the empirical null distribution as described previously. Given the large amount of genes, we could get a good estimate of h non-parametrically. Following Efron [3], we propose a Poisson regression model to estimate h based on $\{t_i\}_{i=1}^m$.

Divide the range of observations for t_i into K intervals with equal length (for moderated F-statistics, we typically choose the range as $[0, \max_j t_j]$). Denote the sample count for each interval by $T_k = \#\{t_i \text{ in interval } k\}$, $k = 1, \dots, K$. We fit T_k with a Poisson regression model,

$$T_k \sim \text{Poisson}(\lambda_k), \quad \log(\lambda_k) = \sum_{i=0}^p \beta_i b_i(x_k),$$

where x_k is the midpoint of interval k , $b_i(x_k)$ is the generated B-spline basis matrix for a natural cubic spline with p degrees of freedom, and λ_k is the expectation of T_k . The density at x_k can be estimated as

$$\widehat{h}(x_k) = \frac{\lambda_k}{\sum_{j=1}^K \lambda_j} \frac{1}{\ell_k}, \quad \ell_k = \text{length of interval } k.$$

For general x , $\widehat{h}(x)$ is estimated based on linear interpolation. The choice of K and p is not critical and the default value is $K = 120$ and $p = 7$ as implemented in the R package *locfdr* [2].

The local FDR is estimated as

$$\widehat{fdr}(t) = \widehat{\theta}_0 \frac{\widehat{h}_0(t)}{\widehat{h}(t)},$$

$\widehat{\theta}_0$ and $\widehat{h}_0$ are derived from the empirical null distribution maximum likelihood estimation and $\widehat{h}$ is the non-parametric estimate based on the Poisson regression model. Using local FDR as the ranking statistic, we estimate the overall FDR using the Monte Carlo approximation. Specifically we simulate a large set of null data based on the estimated empirical null distribution and then calculate the corresponding proportion of false positives to approximate the false positive rate in the observed data.

Given a cut off value c_0 for the local FDR, the overall FDR can be estimated as

$$\widehat{\text{FDR}} = \frac{m\widehat{\theta}_0 \sum_{b=1}^B I(\widehat{fdr}_b^0 < c_0) / B}{\sum_{i=1}^m I(\widehat{fdr}_i < c_0)}.$$

Here $\widehat{fdr}_b^0$ is the estimated local FDR for simulated null gene $b = 1, \dots, B$. We fix $B = 10^6$ in the simulation study and lung transplant data analysis.

3. Application to lung transplant microarray data

We apply the three methods to detect differentially expressed genes for the lung transplant microarray data. Figure 2 compares the estimated FDR based on the empirical Bayes (EB) with the empirical null distribution, and p-values computed from the theoretical and empirical null distributions. Overall the empirical null distribution based p-value ranking performs the best. When controlling FDR at 0.1, both the EB ranking and p-value based on the empirical null distribution identified the same set of 19 significant genes.

Table 1 shows the 19 genes sorted by their p-values computed from the empirical null distribution. The last column is their ranking by local FDR. Some of the selected genes have been shown associated with the human immune process, which is crucial in organ transplant allograft rejection. For example, EZR was found significantly associated with lung adenocarcinoma [9]. STAT6 encodes a protein that plays a central role in exerting interleukin-4 (IL-4) responses and IL-4 is known associated with transplant allograft rejection [see 10–12, e.g.]. PSMC3 regulates the class II transactivator, which is critical for initiation of adaptive immune responses [13]. GSTA4 encodes an enzyme involved in cellular defense against toxic, carcinogenic, and pharmacologically active electrophilic compounds and was found significantly associated with regeneration of graft tissue after transplant [14]. IGKC encodes a protein that could be combined with other proteins to produce a novel immunosuppressant for clinical heart transplantation [15].

Overall we can see that the empirical null distribution based ranking approaches perform very well. And the p-value ranking based on the empirical null distribution has the best overall performance. In the following section we conduct simulation studies to investigate the performance of the proposed methods.

4. Simulation Study

We compare the performance of p-value ranking based on the theoretical and empirical null distributions, and local FDR based EB ranking approach.

We simulate the F-statistics for 10^4 genes, assuming 95% of them are from null genes, following a scaled non-central F-distribution $\sigma_0^{-1} f(\sigma_0^{-1} t; 2, 40, \lambda_0)$, and the rest are from differentially expressed genes, following $\sigma_0^{-1} f(\sigma_0^{-1} t; 2, 40, \lambda)$. We conducted simulations for $\sigma_0 = (1, 2, 0.5)$, $\lambda_0 = (0, 0.5)$ and $\lambda = (5, 10)$.

Here we reported the results from 100 simulations. Figure 3 compares the estimated FDR, with their bias and variance summarized in Table 2 for $\lambda = 10$. When $\sigma_0 = 1$ and $\lambda_0 = 0$, the true null distribution is exactly the theoretical null distribution, and all methods perform similarly with the estimated FDR close to the true values. When $\sigma_0 < 1$, using theoretical null distribution could significantly underestimate FDR, which gives overly optimistic results, while for $\sigma_0 > 1$, the theoretical null yields very conservative estimates. The two empirical null distribution based approaches have similar performance. Similar patterns are observed for $\lambda = 5$ and the corresponding results are summarized in Table 3 and Figure 4.

5. Discussion

Through application and simulation studies, we illustrated the importance of using the empirical null instead of theoretical null distribution in large-scale significance analysis of multi-class differential gene expression detection. When estimating the empirical null distribution with the maximum likelihood, we have assumed that all moderated F-statistics smaller than a pre-selected cutoff value are coming from null genes. In the simulation studies, we have found that the choice of cutoff value has some impact on the performance of local FDR and p-value ranking based on the empirical null distribution (please see the supplementary materials for complete results). Intuitively the cutoff value selection involves the bias and variance tradeoff of the empirical null and associated FDR estimation. It will be interesting to investigate some principled way of choosing the optimal cutoff value.

In addition to the previous empirical Bayes ranking, we could also adopt the idea of McLachlan *et al.* [17] by converting the p-values (based on the empirical null distribution) to z-scores and fitting a two-component mixture model to all the z-values. We will explore this approach and report the results in the future.

Most existing local FDR procedures either assume a known standard distribution for the null density (e.g., Guedj *et al.* [16]) or estimate it using computationally intensive permutation methods (e.g., Pawitan *et al.* [18]). The proposed empirical null modeling approach is flexible without assuming some known standard null distribution and efficient without permutations. This will become increasingly relevant as larger and larger data sets are collected.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by NIH grant GM083345 and CA134848. We would like to thank two anonymous referees for their constructive comments that have dramatically improved the presentation of the paper.

References

1. Efron B. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association*. 2004; 99:96–104.
2. Efron B. Correlation and Large-Scale Simultaneous Significance Testing. *Journal of the American Statistical Association*. 2007a; 102:93–103.
3. Efron B. Size, power, and false discovery rates. *Annals of Statistics*. 2007b; 35(4):1351–1377.
4. Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*. 2004; 3(1):Article 3.
5. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995; 57:289–300.
6. Efron B. Robbins, empirical bayes and microarrays. *The Annals of Statistics*. 2003; 31:366–378.
7. Langaas M, Ferkingstad E, Lindqvist B. Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society Series B (Methodological)*. 2005; 67:555–572.
8. Gimino VJ, Lande JD, Berryman TR, King RA, Hertz MI. Gene Expression Profiling of Bronchoalveolar Lavage Cells in Acute Lung Rejection. *American Journal of Respiratory and Critical Care Medicine*. 2003; 168:1237–1242. [PubMed: 12958056]
9. Tokunou M, Niki T, Saitoh Y, Imamura H, Sakamoto M, Hirohashi S. Altered expression of the ERM proteins in lung adenocarcinoma. *Lab Invest*. 2000; 80(11):1643–1650. [PubMed: 11092524]
10. Manchanda PK, Kumar A, Sharma RK, Goel H, Mittal RD. Association of pro/anti-inflammatory cytokine gene variants in renal transplant patients with allograft outcome and cyclosporine immunosuppressant levels. *Biologics: Targets & Therapy*. 2008; 2(4):875–884.
11. Manchanda PK, Mittal RD. Analysis of cytokine gene polymorphisms in recipient's matched with living donors on acute rejection after renal transplantation. *Molecular and Cellular Biochemistry*. 2008; 311(1–2):57–65. [PubMed: 18165865]
12. Pawlik A, Domanski L, Rozanski J, Czerny B, Juzyszyn Z, Dutkiewicz G, Myslak M, Haasa M, Sojewski M, Dabrowska-Zamojcin E. The association between cytokine gene polymorphisms and kidney allograft survival. *Annals of Transplantation: Quarterly of the Polish Transplantation Society*. 2008; 13(2):54–58. [PubMed: 18566561]
13. Truax AD, Koues OI, Mentel MK, Greer SF. The 19S ATPase S6a (S6'/TBP1) regulates the transcription initiation of class II transactivator. *J Mol Biol*. 2010; 395(2):254–269. [PubMed: 19853614]
14. Bradley SP, Pahari M, Uknis ME, Rastellini C, Cicalese L. Gene expression profiles characterize early graft response in living donor small bowel transplantation: a case report. *Transplant Proc*. 2006; 38(6):1742–1743. [PubMed: 16908269]
15. Meiser BM, Reiter C, Reichenspurner H, Uberfuhr P, Kreuzer E, Rieber EP, Riethmüller G, Reichart B. Chimeric monoclonal CD4 antibody—a novel immunosuppressant for clinical heart transplantation. *Transplantation*. 1994; 58(4):419–423. [PubMed: 7915442]
16. Guedj M, Robin S, Celisse A, Nuel G. Kerfdr: a semi-parametric kernel-based approach to local false discovery rate estimation. *BMC Bioinformatics*. 2009; 10(1):84. [PubMed: 19291295]
17. McLachlan GJ, Bean RW, Jones LBT. A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics*. 2006; 22(13):1608–1615. [PubMed: 16632494]
18. Pawitan Y, Calza S, Ploner A. Estimation of false discovery proportion under general dependence. *Bioinformatics*. 2006; 22(24):3025–3031. [PubMed: 17046978]

Figure 1. Theoretical and empirical null distribution comparison: dashed lines are the sample histogram, quantile plot, and empirical distribution function; solid lines are the estimates from empirical null distribution fitting; and dotted lines are the estimates from theoretical null distribution fitting.

Figure 2. FDR estimates based on the empirical Bayes method with the empirical null and theoretical null distributions for the lung transplant study.

Figure 3.

Estimated FDR versus total number of rejections (TNR) for local FDR ranking, p-value ranking based on theoretical and empirical null distributions: the solid black lines are true values and the dashed lines are estimated values.

Figure 4.

Estimated FDR versus total number of rejections (TNR) for local FDR ranking, p-value ranking based on theoretical and empirical null distribution: the solid black lines are true values and the dashed lines are estimated values.

Table 1

Top 19 genes ranked by p-values computed from the empirical null distribution.

Gene	p-value	fdr rank
EZR (ezrin)	2.1e-06	1
UBA52 (ubiquitin A-52 residue ribosomal protein fusion product 1)	3.8e-06	9
STAT6 (signal transducer and activator of transcription 6, interleukin-4 induced)	1.7e-05	16
NFS1 (NFS1 nitrogen fixation 1 homolog (S. cerevisiae))	2.1e-05	11
EIF3I (eukaryotic translation initiation factor 3, subunit I)	2.4e-05	8
RAD21 (RAD21 homolog (S. pombe))	2.6e-05	6
GINS1 (GINS complex subunit 1 (Psf1 homolog))	2.9e-05	2
RAE1 (RAE1 RNA export 1 homolog (S. pombe))	4.0e-05	3
PSMC3 (proteasome (prosome, macropain) 26S subunit, ATPase, 3)	4.1e-05	4
SELT (selenoprotein T)	4.2e-05	5
HCCS (holocytochrome c synthase (cytochrome c heme-lyase))	4.7e-05	7
NONO (non-POU domain containing, octamer-binding)	5.5e-05	10
UGT2B28 (UDP glucuronosyltransferase 2 family, polypeptide B28)	6.1e-05	12
CLINT1 (clathrin interactor 1)	6.7e-05	13
GSTA4 (glutathione S-transferase alpha 4)	7.4e-05	14
IGKC (immunoglobulin kappa constant)	7.6e-05	15
UBQLN2 (ubiquilin 2)	8.5e-05	17
CHRNA3 (cholinergic receptor, nicotinic, alpha 3)	9.0e-05	18
DAZAP2 (DAZ associated protein 2)	9.7e-05	19

Bias and variance of FDR estimation: FDR is the estimated true value averaged over 100 simulations, TNR is the total number of rejections. P_t/P_e are p-values derived from the theoretical/empirical null distributions, EB is empirical Bayes ranking with local FDR based on the empirical null distribution. C_0 is based on the 75% quantile.

Table 2

$\lambda_0 = 0, \sigma_0 = 1, \lambda = 10$									
TNR	50			100			200		
	FDR	bias	stderr	FDR	bias	stderr	FDR	bias	stderr
P_t	0.019	0.003	0.006	0.049	-0.002	0.009	0.119	-0.004	0.017
P_e	0.019	0.006	0.008	0.049	0.004	0.015	0.119	0.008	0.027
EB	0.019	0.006	0.009	0.049	0.004	0.015	0.119	0.009	0.028

$\lambda_0 = 0.5, \sigma_0 = 2, \lambda = 10$									
TNR	50			100			200		
	FDR	bias	stderr	FDR	bias	stderr	FDR	bias	stderr
P_t	0.071	-0.071	0.000	0.118	-0.118	0.000	0.218	-0.217	0.0001
P_e	0.071	0.013	0.041	0.118	0.020	0.057	0.218	0.027	0.083
EB	0.071	0.013	0.041	0.118	0.021	0.058	0.218	0.028	0.084

$\lambda_0 = 0.5, \sigma_0 = 0.5, \lambda = 10$									
TNR	50			100			200		
	FDR	bias	stderr	FDR	bias	stderr	FDR	bias	stderr
P_t	0.078	0.903	0.042	0.130	0.870	0.0001	0.224	0.776	0.0001
P_e	0.078	0.001	0.047	0.130	0.005	0.068	0.224	0.013	0.092
EB	0.077	0.001	0.047	0.130	0.005	0.068	0.224	0.014	0.093

Bias and variance of FDR estimation: FDR values are averages over 100 simulations, TNR is for total number of rejections, P_lP_e are p-values derived from theoretical/empirical null distributions, EB is empirical Bayes ranking with local FDR based on empirical null distribution.

Table 3

$\lambda_0 = 0, \sigma_0 = 1, \lambda = 5$									
TNR		20		50		100			
		FDR	bias	stderr	FDR	bias	stderr	FDR	stderr
P_l		0.156	-0.011	0.044	0.234	-0.007	0.046	0.323	0.0003
P_e		0.156	0.025	0.058	0.234	0.043	0.071	0.323	0.062
EB		0.159	0.018	0.056	0.237	0.043	0.071	0.324	0.062
$\lambda_0 = 0.5, \sigma_0 = 2, \lambda = 5$									
TNR		10		20		40			
		FDR	bias	stderr	FDR	bias	stderr	FDR	stderr
P_l		0.335	-0.334	0.000	0.415	-0.415	0.0001	0.499	-0.498
P_e		0.335	0.031	0.194	0.415	0.046	0.199	0.499	0.201
EB		0.346	0.0006	0.177	0.434	0.019	0.193	0.515	0.200
$\lambda_0 = 0.5, \sigma_0 = 0.5, \lambda = 5$									
TNR		10		20		40			
		FDR	bias	stderr	FDR	bias	stderr	FDR	stderr
P_l		0.346	0.654	0.0001	0.415	0.585	0.0001	0.490	0.510
P_e		0.346	0.039	0.196	0.415	0.072	0.196	0.490	0.191
EB		0.365	0.004	0.177	0.426	0.050	0.188	0.501	0.188