

## Research Article

# An Empirical Bayes Optimal Discovery Procedure Based on Semiparametric Hierarchical Mixture Models

**Hisashi Noma and Shigeyuki Matsui**

*Department of Data Science, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan*

Correspondence should be addressed to Hisashi Noma; [noma@ism.ac.jp](mailto:noma@ism.ac.jp)

Received 20 December 2012; Revised 23 February 2013; Accepted 4 March 2013

Academic Editor: Su-Yun Huang

Copyright © 2013 H. Noma and S. Matsui. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple testing has been widely adopted for genome-wide studies such as microarray experiments. For effective gene selection in these genome-wide studies, the optimal discovery procedure (ODP), which maximizes the number of expected true positives for each fixed number of expected false positives, was developed as a multiple testing extension of the most powerful test for a single hypothesis by Storey (*Journal of the Royal Statistical Society, Series B*, vol. 69, no. 3, pp. 347–368, 2007). In this paper, we develop an empirical Bayes method for implementing the ODP based on a semiparametric hierarchical mixture model using the “smoothing-by-roughening” approach. Under the semiparametric hierarchical mixture model, (i) the prior distribution can be modeled flexibly, (ii) the ODP test statistic and the posterior distribution are analytically tractable, and (iii) computations are easy to implement. In addition, we provide a significance rule based on the false discovery rate (FDR) in the empirical Bayes framework. Applications to two clinical studies are presented.

## 1. Introduction

The comprehensive, gene expression microarrays are a powerful tool for screening the differentially expressed genes among different phenotypes such as clinical subtypes and prognostic classes of disease from a large pool of candidate genes. Gene screening studies have the potential to be useful for elucidating disease biology and aggressiveness, identifying new therapeutic targets, and developing new molecular diagnostics for optimized medicine for individual patients [1–4]. The high dimensionality of microarray data, however, has posed special challenges in extracting a small number of relevant genes from a large quantity of noise variables by gene screening analysis. In addition to the control of false positives [5–7], improvement of the efficiency of gene screening analysis is important.

For efficient screening of differentially expressed genes, Storey [8] developed the optimal discovery procedure (ODP), which can be interpreted as an extension of the most powerful test for a single hypothesis testing [9] to multiple testing. Storey defined an optimality criterion for multiple testing that maximizes the expected number of true positives (ETP) for

each fixed level of expected false positives (EFP) [8]. The ODP was developed as a testing procedure that achieves this optimality; it improves the overall performance of multiple tests by “borrowing strength” across tests. However, in applying the ODP, the following two components must be estimated: (a) the true status of each significance test (null or alternative) and (b) the true probability distribution corresponding to each test [8]. To address these estimation problems, Storey et al. [10] constructed the generalized likelihood ratio statistic [11], using maximum likelihood estimates. Cao et al. [12] also proposed a Bayesian approach, and Woo et al. [13] developed a computationally efficient method called “modular ODP.”

More recently, Noma and Matsui [14] developed an empirical Bayes approach for the ODP that can effectively circumvent the estimation problems of (a) and (b). Under a hierarchical, random effects model, they showed that the ODP was derived as a testing rule based on the marginal likelihood ratio statistic. In their numerical evaluations based on simulations, the empirical Bayes method nearly achieved the theoretical bound of ETP (i.e., average power) and performed well, compared with the existing methods under a broad range of conditions [14].

A critical aspect of this empirical Bayes approach is that a fully parametric form in the hierarchical modelling must be specified. Although most of the empirical Bayes methods previously discussed in microarray studies [15–21] assume fully parametric natural conjugate models for the prior distribution, this assumption might be restrictive because of possible substantial diversity of effect size among informative genes caused by complex molecular pathways, whose distribution is generally unknown in microarray studies. In this article, we develop an empirical Bayes method for the ODP based on a semiparametric hierarchical mixture model, in which the prior distribution for effect sizes is not specified parametrically. We estimate the nonparametric component of the prior distribution by applying the “smoothing-by-roughening” approach [22, 23] and provide an effective computational method for the ODP statistic.

This article is organized as follows. In Section 2, we provide a brief overview of the theoretical results regarding the ODP methods by Storey [8] and Noma and Matsui [14]. In Section 3, we describe the semiparametric hierarchical mixture model and how to implement the ODP method. We present applications to prostate cancer and lymphoma clinical studies in Section 4. Finally, we provide a discussion in Section 5.

## 2. The Optimal Discovery Procedure in Multiple Significance Testing

**2.1. Storey’s Optimal Discovery Procedure.** In this section, we briefly review the theoretical results regarding the ODP methods. We denote the datasets from  $m$  genes as  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ , where  $\mathbf{x}_k = (x_{k1}, x_{k2}, \dots, x_{kn})^T$  ( $k = 1, \dots, m$ ). In microarray studies,  $\mathbf{x}_k$  corresponds to a set of log-transformed, normalized gene expressions from  $n$  samples for the  $k$ th gene (see Section 4). Also, suppose that the datasets are random vectors defined in a common probability space. The optimal goal when identifying differentially expressed genes is to find as many true positives as possible, without incurring too many false positives [10]. Storey [8] defined an optimal criterion of multiple testing as a rule of statistical significance that maximizes ETP for a certain fixed EFP level.

He derived the following lemma that gives the multiple testing procedure that achieves the optimal criterion.

**Lemma 1** (Storey [8]). *Suppose the  $k$ th test has a null density  $f_k(\mathbf{x})$  and alternative density  $g_k(\mathbf{x})$  ( $k = 1, \dots, m$ ). Also, assume that the null hypothesis is true for the  $k$ th test ( $k = 1, \dots, m_0$ ) and the alternative is true for  $k = m_0 + 1, \dots, m$ , without loss of generality. The following significance threshold statistic achieves the ODP criterion:*

$$S_{\text{ODP}}(\mathbf{x}) = \frac{g_{m_0+1}(\mathbf{x}) + g_{m_0+2}(\mathbf{x}) + \dots + g_m(\mathbf{x})}{f_1(\mathbf{x}) + f_2(\mathbf{x}) + \dots + f_{m_0}(\mathbf{x})}. \quad (1)$$

Given a fixed cut-off  $\lambda$  ( $0 \leq \lambda < \infty$ ) chosen to attain an acceptable EFP level, the null hypothesis for the  $k$ th test is rejected if and only if  $S_{\text{ODP}}(\mathbf{x}_k) \geq \lambda$ .

It should be noted that the ODP statistic is composed of all the density functions on  $m$  tests. In other words, through borrowing strengths across  $m$  tests, the overall power is improved more than the conventional most powerful test for single hypothesis tests [9, 11]. Also, the optimality result is held regardless of correlation structure among the datasets  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ .

As mentioned in Section 1, for applying  $S_{\text{ODP}}(\mathbf{x})$ , there are two practical issues: (a) estimation of the true status of each hypothesis (null or alternative), corresponding to the numerator and denominator of  $S_{\text{ODP}}(\mathbf{x})$ , and (b) estimation of the true probability density functions  $f_k(\mathbf{x})$ ’s and  $g_k(\mathbf{x})$ ’s at which  $\mathbf{x}_k$ ’s are evaluated. Storey et al. [10] provided an practical estimating method, motivated by the generalized likelihood ratio test for single significance tests [11]. They noticed that the significance rule based on

$$S_{\text{ODP}}(\mathbf{x}) + 1 = \frac{f_1(\mathbf{x}) + \dots + f_{m_0}(\mathbf{x}) + g_{m_0+1}(\mathbf{x}) + \dots + g_m(\mathbf{x})}{f_1(\mathbf{x}) + f_2(\mathbf{x}) + \dots + f_{m_0}(\mathbf{x})} \quad (2)$$

is equivalent to that of  $S_{\text{ODP}}(\mathbf{x})$  in Lemma 1. For each test  $k = 1, \dots, m$ , let  $\hat{f}_k(\mathbf{x})$  be an estimate of the null density function with all the unknown parameters replaced by their maximum likelihood estimates under the constraints of the null hypothesis, and let  $\hat{g}_k(\mathbf{x})$  be the analogous estimate for the alternative hypothesis using the unconstrained maximum likelihood estimates. Then, Storey et al. [10] proposed to use the canonical plug-in estimate:

$$\hat{S}_{\text{ODP}}(\mathbf{x}) = \frac{\hat{g}_1(\mathbf{x}) + \hat{g}_2(\mathbf{x}) + \dots + \hat{g}_m(\mathbf{x})}{\hat{w}_1 \hat{f}_1(\mathbf{x}) + \hat{w}_2 \hat{f}_2(\mathbf{x}) + \dots + \hat{w}_m \hat{f}_m(\mathbf{x})}, \quad (3)$$

where  $\hat{w}_k$  ( $k = 1, 2, \dots, m$ ) is an estimate of the status of the  $k$ th hypothesis. For estimating  $\hat{w}_k$ , Storey et al. [10] ranked the tests based on a univariate statistic (e.g., the  $t$ -statistic for two-class comparison) and estimated it as  $\hat{w}_k = 1$  when the  $k$ th test fell below the estimated proportion of true nulls by Storey [6], and  $\hat{w}_k = 0$  otherwise. Other estimating procedures for  $S_{\text{ODP}}(\mathbf{x})$  were also proposed by Cao et al. [12] and Woo et al. [13].

**2.2. Empirical Bayes Approach by Noma and Matsui [14].** As Storey [8] pointed out, the ODP is analogous to the well-known shrinkage estimation for multiple-point estimation [24, 25]. Because the shrinkage estimation is interpreted as the empirical Bayes estimation [26–28], hierarchical modelling would be a natural formulation for information sharing across tests, both in deriving the ODP and in developing estimation methods for implementing it [14].

Suppose that the probability density functions for  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  have the same parametric form  $f(\mathbf{x} \mid \theta_k, \psi_k)$ , where  $\theta_k$  is the parameter of interest and  $\psi_k$  is the nuisance parameter for the  $k$ th test. We assume the following prior distribution for the parameters  $(\theta_k, \psi_k)$ :

$$\begin{aligned} \text{null: } (\theta_k, \psi_k) &\sim G_0(\theta, \psi \mid \xi_0), \\ \text{alternative: } (\theta_k, \psi_k) &\sim G_1(\theta, \psi \mid \xi_1), \end{aligned} \quad (4)$$

where  $\xi_0$  and  $\xi_1$  are the hyperparameters of the prior distribution.

**Lemma 2** (Noma and Matsui [14]). *Under the empirical Bayes framework with the prior distribution (4), the following significance threshold function achieves the ODP criterion:*

$$R_{\text{ODP}}(\mathbf{x}) = \frac{E_{G_1}[f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi})]}{E_{G_0}[f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi})]} \quad (5)$$

$$= \frac{\int f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi}) dG_1(\boldsymbol{\theta}, \boldsymbol{\psi} | \xi_1)}{\int f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi}) dG_0(\boldsymbol{\theta}, \boldsymbol{\psi} | \xi_0)}.$$

For a fixed cut-off  $\lambda$  ( $0 \leq \lambda < \infty$ ) chosen to reach an acceptable EFP level, the null hypothesis for the  $k$ th test is rejected if and only if  $R_{\text{ODP}}(\mathbf{x}_k) \geq \lambda$ .

See Noma and Matsui [14] for interpretations of  $R_{\text{ODP}}(\mathbf{x})$ .

For the specification of the prior distributions  $G_0(\boldsymbol{\theta}, \boldsymbol{\psi} | \xi_0)$  and  $G_1(\boldsymbol{\theta}, \boldsymbol{\psi} | \xi_1)$ , Noma and Matsui [14] adopted the empirical Bayes method [26, 27]. The empirical Bayes implementation is used to obtain estimates for  $\xi_0$  and  $\xi_1$  from the data, which can then be substituted into the ODP statistic,  $R_{\text{ODP}}(\mathbf{x})$ ,

$$\hat{R}_{\text{ODP}}(\mathbf{x}) = \frac{\int f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi}) dG_1(\boldsymbol{\theta}, \boldsymbol{\psi} | \hat{\xi}_1)}{\int f(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\psi}) dG_0(\boldsymbol{\theta}, \boldsymbol{\psi} | \hat{\xi}_0)}, \quad (6)$$

where  $\hat{\xi}_0$  and  $\hat{\xi}_1$  are the estimators of  $\xi_0$  and  $\xi_1$ , respectively.

In comparison with the existing ODP estimation methods based on  $S_{\text{ODP}}(\mathbf{x})$ , the empirical Bayes approach has two advantages. First, it estimates only the hyperparameters  $\xi_0$  and  $\xi_1$  for the random effects models, and not the parameters in the density function  $\mathbf{x}_k$  ( $k = 1, \dots, m$ ), for which a large number of parameters (thousands or more in microarray studies) must be estimated as in the procedure of Storey et al. [10]. Second, in the empirical Bayes approach, the estimation of the true status, null or alternative, for each test is not needed.

### 3. Semiparametric Hierarchical Mixture Model and the ODP

**3.1. Semiparametric Hierarchical Mixture Model.** In this section, we consider applying the ODP to concrete microarray analyses of two-class comparisons (with classes denoted as “0” and “1”), for example, good prognosis and poor prognosis, on the basis of the expression levels of  $m$  genes from  $n$  samples. The gene expression data considered here comprise normalized log ratios from two-color complementary DNA arrays or normalized log signals from oligonucleotide arrays (e.g., Affymetrix GeneChip). Let  $n_0$  and  $n_1$  be the sample sizes of classes 0 and 1, respectively. Let  $\mu_{0k}$  and  $\mu_{1k}$  ( $k = 1, 2, \dots, m$ ) be the means of the gene expression levels for classes 0 and 1, respectively. We consider the detection of

differentially expressed genes to be when  $\mu_{0k} \neq \mu_{1k}$ . That is, testing problem for

$$H_{0k} : \mu_{0k} = \mu_{1k} \text{ versus } H_{1k} : \mu_{0k} \neq \mu_{1k} \quad (7)$$

$$(k = 1, 2, \dots, m).$$

We assume that gene expressions of the  $k$ th gene are normally distributed within each class,

$$x_{k1}, \dots, x_{kn_0} \sim N(\mu_{0k}, \sigma_k^2), \quad (8)$$

$$y_{k1}, \dots, y_{kn_1} \sim N(\mu_{1k}, \sigma_k^2),$$

where  $\sigma_k^2$  is a common within-class variance ( $k = 1, 2, \dots, m$ ), as in Storey et al. [10]. Let  $d_k$  be the difference of the mean of the two classes for the  $k$ th gene ( $k = 1, 2, \dots, m$ ),  $\mu_{1k} - \mu_{0k}$ , which is the parameter of interest. Let  $\bar{x}_k, \bar{y}_k$  denote the sample means of the gene expression levels for classes 0 and 1, respectively. We consider the following translation:

$$w_{ki} = \sqrt{\frac{n_0}{n_0 - 1}} (x_{ki} - \bar{x}_k), \quad (9)$$

$$z_{ki} = \sqrt{\frac{n_0}{n_0 + 1}} (y_{ki} - \bar{x}_k).$$

Using this translation, the nuisance parameters  $\mu_{0k}$ 's are eliminated in the models of  $w_{ki}$  and  $z_{ki}$ . Specifically,

$$w_{k1}, \dots, w_{kn_0} \sim N(0, \sigma_k^2), \quad (10)$$

$$z_{k1}, \dots, z_{kn_1} \sim N(\theta_k, \sigma_k^2),$$

where  $\theta_k = \sqrt{n_0/(n_0 + 1)}d_k$ . Note that the testing problems in (7) are equivalent to testing

$$H'_{0k} : \theta_k = 0 \text{ versus } H'_{1k} : \theta_k \neq 0 \quad (11)$$

$$(k = 1, 2, \dots, m).$$

Let  $\mathbf{w}_k = (w_{k1}, \dots, w_{kn_0})$  and  $\mathbf{z}_k = (z_{k1}, \dots, z_{kn_1})$ .

For applying the ODP to the testing problems  $H'_{0k} : \theta_k = 0$  ( $k = 1, 2, \dots, m$ ) under model (10), we assume the following random effects distributions for the model parameters  $(\theta_k, \sigma_k^2)$ :

$$\theta_k \sim \pi_0 \delta(\theta) + \pi_1 G(\theta),$$

$$\sigma_k^2 \sim \text{Gamma}^{-1}(\alpha, \beta), \quad (12)$$

$$\left[ P(\sigma_k^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma_k^{-2(\alpha+1)} \exp\left(-\frac{\beta}{\sigma_k^2}\right) \right],$$

where  $\delta(\theta)$  in the first component of  $\theta_k$  is the Dirac delta function, representing nondifferential expression between two classes. The second component of  $\theta_k$  represents a component of the differential expression, and  $\pi_0$  and  $\pi_1$  are the mixing proportions for the components ( $\pi_0 + \pi_1 = 1$ ). The random effects distribution of the gene-specific variances  $\sigma_k^2$  is specified as the parametric conjugate model for analytical tractability. For modeling  $G(\theta)$ , instead of adopting a fully parametric conjugate normal distribution model as done by Noma and Matsui [14], we apply the “smoothing-by-roughening” approach of Laird and Louis [22] and Shen and Louis [23]. This approach was proposed to obtain the nonparametric maximum likelihood estimate of the prior distribution in empirical Bayes analysis [29, 30]. This approach starts with a smooth estimate of the prior distribution and uses the EM algorithm to “roughen” it towards the nonparametric maximum likelihood estimate. For modeling  $G(\theta)$ , we place supports at equally spaced discrete mass points at a series of nonzero points  $\mathbf{a} = (a_1, a_2, \dots, a_L)$ , for sufficiently large values of  $L$ . We assume that

$$\Pr(\theta = a_j) = p_j \quad (j = 1, 2, \dots, L), \quad (13)$$

where  $p_1 + p_2 + \dots + p_L = 1$ . We denote the hyperparameters of the distribution function for the differential component as  $\mathbf{p} = (p_1, p_2, \dots, p_L)$ . Shen and Louis [23] provided a detailed guideline for implementing the smoothing-by-roughening approach based on their numerical evaluations. Shen and Louis [23] recommended that the number of grid points should be at least 200. For simplicity, they suggested the grid points can be equally spaced on the support of the distribution. Increasing the number of grid points and spacing them according to the prior density will provide a better approximation of  $G(\theta)$  and therefore will improve the estimation, particularly for the tail areas. However, too large numbers might not necessarily be beneficial. For more details involving analytical and numerical evaluations, see the guideline in Shen and Louis [23]. In this article, we refer to this hierarchical model as the “semiparametric hierarchical mixture model.” The important property of this model is that the posterior distributions of  $\theta_k$  and  $\sigma_k^2$  ( $k = 1, 2, \dots, m$ ) can be obtained analytically. We provide the analytical representation of the posterior distributions in Section 3.4.

The maximum likelihood estimates of the hyperparameters  $\boldsymbol{\eta} = (\pi_0, \pi_1, \mathbf{p}, \alpha, \beta)$  are obtained by the EM algorithm [31]. Because the posterior distributions of  $\theta_k$  and  $\sigma_k^2$  are obtained analytically, the E-step of the EM algorithm can be implemented easily. The details of the EM algorithm are presented in the appendix.

**3.2. The Optimal Discovery Procedure.** The ODP statistic  $R_{\text{ODP}}(\mathbf{w}, \mathbf{z})$  for the testing problems (11) is obtained as the marginal likelihood ratio. The marginal likelihood of the null hypothesis corresponds to the marginal likelihood of the

nondifferential component:

$$\begin{aligned} h_0(\mathbf{w}_k, \mathbf{z}_k) &= \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = 0, \sigma_k^2) p(\sigma_k^2 \mid \alpha, \beta) d\sigma_k^2 \\ &= \frac{\Gamma(n/2 + \alpha)}{\Gamma(\alpha)} \beta^\alpha (2\pi)^{-n/2} \\ &\quad \times \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} z_{ki}^2}{2} + \beta \right)^{-(n/2 + \alpha)}. \end{aligned} \quad (14)$$

Similarly, the marginal likelihood of the alternative hypothesis in (11) corresponds to that of the differential component:

$$\begin{aligned} h_1(\mathbf{w}_k, \mathbf{z}_k) &= \sum_{j=1}^L \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) \\ &\quad \times p(\sigma_k^2 \mid \alpha, \beta) d\sigma_k^2 \Pr(\theta_k = a_j) \\ &= \frac{\Gamma(n/2 + \alpha)}{\Gamma(\alpha)} \beta^\alpha (2\pi)^{-n/2} \\ &\quad \times \sum_{j=1}^L \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki} - a_j)^2}{2} + \beta \right)^{-(n/2 + \alpha)} p_j \end{aligned} \quad (15)$$

under the semiparametric hierarchical mixture model.

Thus, the ODP statistic (the marginal likelihood ratio) is given as

$$\begin{aligned} R_{\text{ODP}}(\mathbf{w}_k, \mathbf{z}_k) &= \left( \sum_{j=1}^L \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) \right. \\ &\quad \times p(\sigma_k^2 \mid \alpha, \beta) d\sigma_k^2 \Pr(\theta_k = a_j) \Big) \\ &\quad \times \left( \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = 0, \sigma_k^2) p(\sigma_k^2 \mid \alpha, \beta) d\sigma_k^2 \right)^{-1} \\ &= \left( \sum_{j=1}^L \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki} - a_j)^2}{2} + \beta \right)^{-(n/2 + \alpha)} p_j \right) \\ &\quad \times \left( \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} z_{ki}^2}{2} + \beta \right)^{-(n/2 + \alpha)} \right)^{-1}. \end{aligned} \quad (16)$$

By plugging-in the maximum likelihood estimate of  $\boldsymbol{\eta}$ , the empirical Bayes testing procedure is implemented based on

$$\begin{aligned} \widehat{R}_{\text{ODP}}(\mathbf{w}_k, \mathbf{z}_k) \\ = \frac{\sum_{j=1}^L \left( \left( \sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki} - a_j)^2 \right) / 2 + \widehat{\beta} \right)^{-(n/2+\widehat{\alpha})} \widehat{p}_j}{\left( \left( \sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} z_{ki}^2 \right) / 2 + \widehat{\beta} \right)^{-(n/2+\widehat{\alpha})}}. \end{aligned} \quad (17)$$

Note that the ODP statistic is obtained in closed form using the maximum likelihood estimate of  $\boldsymbol{\eta}$ .

**3.3. Assessing Significance.** For assessing significance, we can obtain an empirical Bayes estimator of FDR [5–7] under the semiparametric hierarchical mixture model as follows. For any given threshold  $\lambda$ , we denote the significant region  $\Lambda(\lambda) = \{(\mathbf{w}, \mathbf{z}) \mid R_{\text{ODP}}(\mathbf{w}, \mathbf{z}) \geq \lambda\}$  for each test. Let  $\Xi$  denote the set of indices for the significant genes. In the framework of Bayesian selection rules, the FDR accords to the misclassification error rate of differential/non-differential classification [6, 7, 32], and one of the well-known Bayesian estimator of the FDR [33, 34] is

$$\widehat{\text{FDR}}(\lambda) = \frac{1}{|\Xi|} \sum_{k \in \Xi} \Pr(H_{0k} \mid \mathbf{w}_k, \mathbf{z}_k), \quad (18)$$

where

$$\begin{aligned} \Pr(H_{0k} \mid \mathbf{w}_k, \mathbf{z}_k) \\ = \frac{\pi_0 h_0(\mathbf{w}_k, \mathbf{z}_k)}{\pi_0 h_0(\mathbf{w}_k, \mathbf{z}_k) + \pi_1 h_1(\mathbf{w}_k, \mathbf{z}_k)} \\ = \left( \pi_0 \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} z_{ki}^2}{2} + \beta \right)^{-(n/2+\alpha)} \right) \\ \times \left( \pi_0 \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} z_{ki}^2}{2} + \beta \right)^{-(n/2+\alpha)} \right. \\ \left. + \pi_1 \sum_{j=1}^L \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki} - a_j)^2}{2} + \beta \right)^{-(n/2+\alpha)} p_j \right)^{-1}. \end{aligned} \quad (19)$$

Accordingly, an empirical Bayes estimator of  $\widehat{\text{FDR}}(\lambda)$  for the significant region  $\Lambda(\lambda) = \{(\mathbf{w}, \mathbf{z}) \mid R_{\text{ODP}}(\mathbf{w}, \mathbf{z}) \geq \lambda\}$  can be obtained by plugging-in the maximum likelihood estimate of  $\boldsymbol{\eta}$ .

Note that we conducted small simulations for checking correctness of the proposed ODP procedure and these results were provided in the Supplementary Materials available online at <http://dx.doi.org/10.1155/2013/568480>.

**3.4. The Posterior Distribution.** Under the semiparametric hierarchical mixture model, the posterior distributions of  $\theta_k$

and  $\sigma_k^2$  ( $k = 1, 2, \dots, m$ ) can be obtained analytically as noted above; this gives a computational advantage for implementing Bayesian inference. The posterior distribution will be used to derive estimates and confidence intervals of  $d_k$ 's, in addition to the statistical significance measures such as the  $q$ -value.

Let  $\gamma_{0k}$  and  $\gamma_{1k}$  denote indicator variables to which the  $k$ th gene component belongs ( $k = 1, 2, \dots, m$ ). The joint posterior distribution of  $(\theta_k, \sigma_k^2)$  is expressed as a mixture of differential and non-differential components:

$$\begin{aligned} p(\theta_k, \sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k) \\ = p(\theta_k, \sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k, \gamma_{0k} = 1) \Pr(\gamma_{0k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k) \\ + p(\theta_k, \sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k, \gamma_{1k} = 1) \Pr(\gamma_{1k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k), \end{aligned} \quad (20)$$

where  $p(\theta_k, \sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k, \gamma_{0k} = 1) = \delta(\theta) p(\sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k)$ . The weight  $\Pr(\gamma_{0k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k) = \Pr(H_{0k} \mid \mathbf{w}_k, \mathbf{z}_k)$  is given in Section 3.3, and  $\Pr(\gamma_{0k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k) = 1 - \Pr(\gamma_{1k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k)$ . The posterior distribution of the differential component is expressed as

$$\begin{aligned} p(\theta_k = a_j, \sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k, \gamma_{1k} = 1) \\ = \Pr(\theta_k = a_j \mid \sigma_k^2, \mathbf{w}_k, \mathbf{z}_k) p(\sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k), \end{aligned} \quad (21)$$

where

$$\Pr(\theta_k = a_j \mid \sigma_k^2, \mathbf{w}_k, \mathbf{z}_k) = \frac{p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) p_j}{\sum_{l=1}^L p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_l, \sigma_k^2) p_l}, \quad (22)$$

$$\begin{aligned} p(\sigma_k^2 \mid \mathbf{w}_k, \mathbf{z}_k) \\ = \frac{p(\mathbf{w}_k, \mathbf{z}_k \mid \sigma_k^2) p(\sigma_k^2)}{\int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \sigma_k^2) p(\sigma_k^2) d\sigma_k^2} \\ = \frac{\sum_{j=1}^L p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) p(\sigma_k^2) p_j}{\sum_{j=1}^L \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) p(\sigma_k^2) d\sigma_k^2 p_j}, \end{aligned} \quad (23)$$

$$\begin{aligned} \int_0^\infty p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k = a_j, \sigma_k^2) p(\sigma_k^2) d\sigma_k^2 \\ = \frac{\Gamma(n/2 + \alpha)}{\Gamma(\alpha)} \beta^\alpha (2\pi)^{-n/2} \\ \times \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki} - a_j)^2}{2} + \beta \right)^{-(n/2+\alpha)}. \end{aligned} \quad (24)$$



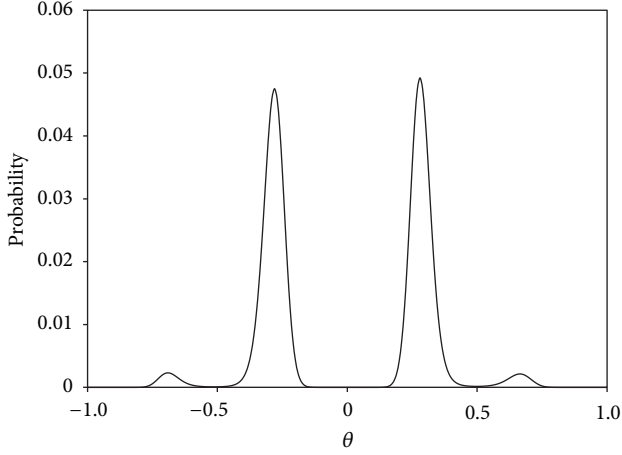


FIGURE 1: The nonparametric estimate for the differential component of the prior distribution of the prostate cancer data [38].

The marginal posterior distribution of  $\theta_k$  can be obtained as follows:

$$\begin{aligned}
 & \Pr(\theta_k = a_j | \mathbf{w}_k, \mathbf{z}_k) \\
 &= \frac{p(\mathbf{w}_k, \mathbf{z}_k | \theta_k = a_j) \Pr(\theta_k = a_j)}{\sum_{l=1}^L p(\mathbf{w}_k, \mathbf{z}_k | \theta_k = a_l) \Pr(\theta_k = a_l)} \\
 &= \frac{\left( \left( \sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki}^2 - a_j)^2 \right) / 2 + \beta \right)^{-(n/2+\alpha)} p_j}{\sum_{l=1}^L \left\{ \left( \left( \sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki}^2 - a_l)^2 \right) / 2 + \beta \right)^{-(n/2+\alpha)} p_l \right\}}. \quad (25)
 \end{aligned}$$

#### 4. Applications

We illustrate our method using real data sets from two clinical studies with microarrays. Because previous simulation studies have indicated that the Storey's method [10] is more powerful than many other multiple testing methods [10, 12, 35], we chose to use this method as a reference.

**4.1. Prostate Cancer Example.** Efron [36, 37] analyzed a modified gene expression data set from Singh et al. [38] with  $m = 6033$  genes for  $n = 102$  samples;  $n_1 = 52$  prostate cancer patients and  $n_0 = 50$  healthy controls. For the proposed empirical Bayes approach, we placed the grid points for  $G(\theta)$  equally spaced on  $[-1, 1]$  by 0.01 (except for 0;  $L = 200$ ). The hyperparameters were estimated as  $\hat{\pi}_0 = 0.716$ ,  $\hat{\pi}_1 = 0.284$ ,  $\hat{\alpha} = 8.624$ , and  $\hat{\beta} = 7.373$ , and the estimated prior distribution for  $G(\theta)$  is given in Figure 1. The estimated prior distribution was skewed and multimodal. This indicates that analytically tractable parametric models (for example, the normal distribution) might be inadequate in this case.

Figure 2 presents the results of multiple testing by the method of Storey et al. [10] and the empirical Bayes method developed in this article. This plot summarizes the correspondence of the numbers of significant genes and  $q$ -values [6, 7] for varying values of the cut-off  $\lambda$ . For a fair comparison, the

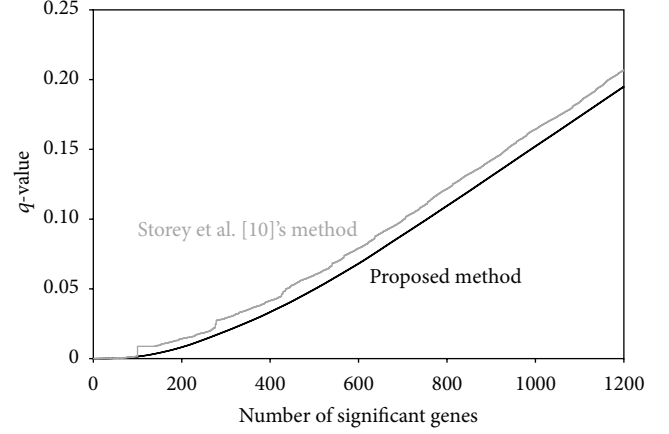


FIGURE 2: Results of the multiple testing for the prostate cancer data [38].

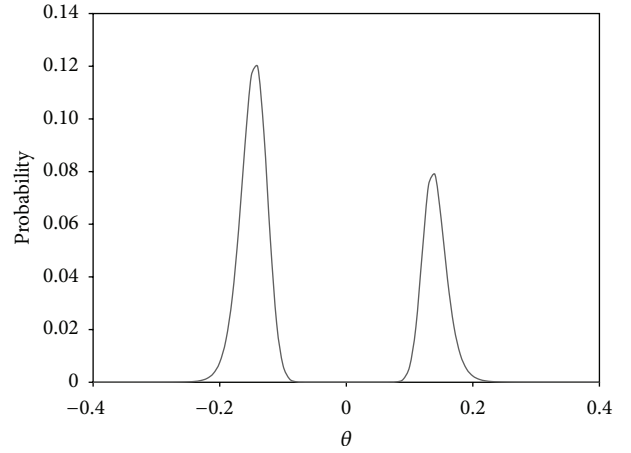


FIGURE 3: The nonparametric estimate for the differential component of the prior distribution of the lymphoma data [39].

same FDR-controlling method presented in Section 3.3 was applied to these two methods. For any levels of the  $q$ -values, the number of significant genes identified by the proposed method was greater than that of the Storey et al. [10] method.

**4.2. Lymphoma Example.** Dave et al. [39] analyzed Affymetrix HG-U133A and U133B (Affymetrix, Santa Clara, CA) microarray data for 191 biopsy specimens from patients with untreated follicular lymphoma with  $m = 44,928$ . The data are available from the BRB-ArrayTools Data Archive for Human Cancer Gene Expression [40]. We compared patients who died within 5 years (poor prognosis;  $n_1 = 51$ ) with patients who survived for more than 5 years (good prognosis;  $n_0 = 109$ ).

For the maximum likelihood estimate of the semiparametric prior distribution, the hyperparameters were estimated as  $\hat{\pi}_0 = 0.831$ ,  $\hat{\pi}_1 = 0.169$ ,  $\hat{\alpha} = 1.266$ , and  $\hat{\beta} = 0.311$ . The estimated prior distribution for  $G(\theta)$  is given in Figure 3. (The grid points were placed on  $[-1, 1]$  equally spaced by 0.01 except for 0;  $L = 200$ .) Compared with the prostate

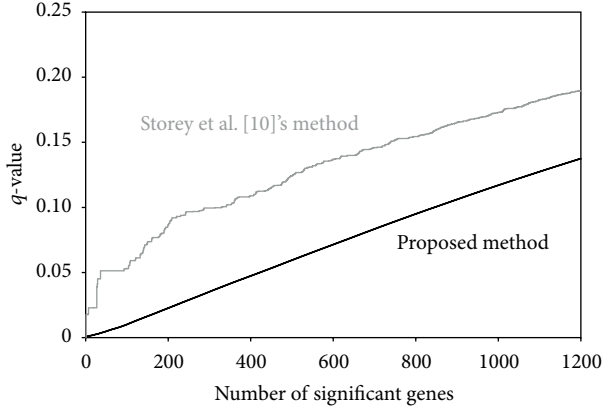


FIGURE 4: Results of the multiple testing for the lymphoma data [39].

cancer data, large masses of the estimated  $G(\theta)$  are located in the regions of small  $\theta$ , and thus the signal on differential expression contained in the data would be weaker than that in the prostate cancer data. In Figure 4, the relationships of the numbers of significant genes and their  $q$ -values are presented. The number of significant genes identified by the empirical Bayes method was greater than that of the method of Storey et al. [10].

## 5. Discussion

For efficient gene screening using high-dimensional microarray data, multiple testing methodologies are effective tools for controlling false positives findings. Although the multiple testing can provide a relevant framework to ensure control of false positives for a set of significant genes (e.g., FDR [5–7]), researchers would like to find as many true positive genes as possible for further investigations. The optimal discovery procedure (ODP), introduced by Storey [8], would provide an efficient solution for this requirement. Storey et al. [10] provided desirable results in many numerical evaluations [10, 12, 35].

In this article, we proposed an empirical Bayes ODP method based on a semiparametric hierarchical model, through relaxing the parametric prior assumption in the empirical Bayes ODP method by Noma and Matsui [14]. Hierarchical mixture models and empirical Bayes methods form a basis for information sharing across genes and provide a framework for efficient multiple testing [15–21]. Although empirical Bayes methods are more robust to prior misspecifications than eliciting a single prior, parametric empirical Bayes methods can suffer from a lack of robustness when the true prior is not from the assumed parametric family, as pointed out by Morris [28], such as a failure to incorporate the mixture structure or specify the form of the effect-size distribution ( $G$ ). Actually, the estimated prior distribution in the prostate cancer example (Section 4.1) had a multimodal and skewed shape. In such examples, natural conjugate families might be inadequate for modeling an unknown random effects distribution. The “smoothing-by-roughening” approach [22, 23] is one of the flexible

modeling methods in empirical Bayes inference [41–43] and could be a basis for effective empirical Bayes inference for a broad range of situations. An extension of our method is to specify a nonparametric distribution for the random effect distribution of  $\sigma_k^2$  through applying the smoothing-by-roughening approach, although it requires substantial computation for estimating a large number of parameters and can suffer from unstable parameter estimates.

Previously, many efficient gene selection methods have been proposed for microarray studies, not only the multiple testing methodologies, for example, Bayesian ranking methods [20, 21, 43]. For evaluating practical values of these methods, comprehensive numerical investigations, including large-scale simulations and applications to many real datasets, would be a worthwhile subject.

## Appendix

### A. The EM Algorithm

The maximum likelihood estimate of the hyperparameters  $\boldsymbol{\eta}$  is obtained using the EM algorithm [31]. Regarding  $\theta_k$ s,  $\sigma_k^2$ s,  $\gamma_{0k}$ s, and  $\gamma_{1k}$ s as missing variables, the log complete data likelihood is expressed as

$$\begin{aligned}
 \log L(\boldsymbol{\eta}) &= \sum_{k=1}^m \gamma_{0k} \log p(\mathbf{w}_k, \mathbf{z}_k, \sigma_k^2 \mid \boldsymbol{\eta}, \gamma_{0k} = 1) \\
 &\quad + \gamma_{1k} \log p(\mathbf{w}_k, \mathbf{z}_k, \theta_k, \sigma_k^2 \mid \boldsymbol{\eta}, \gamma_{1k} = 1) \\
 &\quad + \gamma_{0k} \log \pi_0 + \gamma_{1k} \log \pi_1 \\
 &= \sum_{k=1}^m \gamma_{0k} \log p(\mathbf{w}_k, \mathbf{z}_k \mid \sigma_k^2, \gamma_{0k} = 1) \\
 &\quad + \gamma_{1k} \log p(\mathbf{w}_k, \mathbf{z}_k \mid \theta_k, \sigma_k^2, \gamma_{1k} = 1) \\
 &\quad + \gamma_{1k} \log \Pr(\theta_k = a_j \mid \boldsymbol{\eta}, \gamma_{1k} = 1) \\
 &\quad + \log p(\sigma_k^2 \mid \boldsymbol{\eta}) + \gamma_{0k} \log \pi_0 + \gamma_{1k} \log \pi_1.
 \end{aligned} \tag{A.1}$$

The first and second terms in the sum do not depend on the hyperparameters. Thus the objective function in each  $M$ -step is given by

$$\begin{aligned}
 Q(\boldsymbol{\eta} \mid \boldsymbol{\eta}^{(t)}) &= \sum_{k=1}^m E[\gamma_{1k} \log \Pr(\theta_k = a_j \mid \boldsymbol{\eta}, \gamma_{1k} = 1) \\
 &\quad + \log p(\sigma_k^2 \mid \boldsymbol{\eta}) \\
 &\quad + \gamma_{0k} \log \pi_0 + \gamma_{1k} \log \pi_1 \mid \mathbf{w}_k, \mathbf{z}_k, \boldsymbol{\eta}^{(t)}] \\
 &= \sum_{k=1}^m \omega_{1k}^{(t)} E[\log \Pr(\theta_k = a_j \mid \boldsymbol{\eta}, \gamma_{1k} = 1) \mid \mathbf{w}_k, \mathbf{z}_k, \boldsymbol{\eta}^{(t)}] \\
 &\quad + E[\log p(\sigma_k^2 \mid \boldsymbol{\eta}) \mid \mathbf{w}_k, \mathbf{z}_k, \boldsymbol{\eta}^{(t)}] \\
 &\quad + \omega_{0k}^{(t)} \log \pi_0 + \omega_{1k}^{(t)} \log \pi_1,
 \end{aligned} \tag{A.2}$$

where  $\omega_{0k}^{(t)} = \Pr(\gamma_{0k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k)$ ,  $\omega_{1k}^{(t)} = \Pr(\gamma_{1k} = 1 \mid \mathbf{w}_k, \mathbf{z}_k)$ . The superscripts indicate the number of the iterations. Solving the optimization problem, the  $M$ -step is expressed as

$$\begin{aligned} \pi_0^{(t+1)} &= \frac{1}{m} \sum_{k=1}^m \omega_{0k}^{(t)}, \quad \pi_1^{(t+1)} = 1 - \pi_0^{(t+1)}, \\ p_j^{(t+1)} &= \left( \sum_{k=1}^m \omega_{1k}^{(t)} p_j^{(t)} \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki}^2 - a_j)^2}{2} + \beta^{(t)} \right)^{-(n/2+\alpha^{(t)})} \right) \\ &\quad \times \left( \left\{ \sum_{k=1}^m \omega_{1k}^{(t)} \right\} \sum_{l=1}^L \left\{ p_l^{(t)} \left( \frac{\sum_{i=1}^{n_0} w_{ki}^2 + \sum_{i=1}^{n_1} (z_{ki}^2 - a_l)^2}{2} + \beta^{(t)} \right)^{-(n/2+\alpha^{(t)})} \right\} \right)^{-1}. \end{aligned} \quad (\text{A.3})$$

Also, the optimization problem for  $\alpha$  and  $\beta$  is equivalent to the maximum likelihood estimation of the shape and scale parameters for the inverse gamma distribution. It can be solved simply by any numerical optimization methods provided in standard software (e.g., nlm or optim in R).

## Acknowledgment

This work was supported by a Grant-in-Aid from Japan Society for the Promotion of Science (24800081).

## References

- [1] S. Matsui and H. Noma, "Analysis of DNA microarrays," in *Handbook of Statistics in Clinical Oncology*, J. Crowley and A. Hoering, Eds., pp. 561–581, Chapman and Hall/CRC, Boca Raton, Fla, USA, 3rd edition, 2012.
- [2] R. M. Simon, E. L. Korn, L. M. McShane et al., *Design and Analysis of DNA Microarray Investigations*, Springer, New York, NY, USA, 2003.
- [3] T. Strachan and A. Read, *Human Molecular Genetics*, Garland Science, New York, NY, USA, 4th edition, 2011.
- [4] C. Wiuf and C. L. Andersen, *Statistics and Informatics in Molecular Cancer Research*, Oxford University Press, New York, NY, USA, 2009.
- [5] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate—a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B*, vol. 57, no. 1, pp. 289–300, 1995.
- [6] J. D. Storey, "A direct approach to false discovery rates," *Journal of the Royal Statistical Society, Series B*, vol. 64, no. 3, pp. 479–498, 2002.
- [7] J. D. Storey, "The positive false discovery rate: a Bayesian interpretation and the  $q$ -value," *Annals of Statistics*, vol. 31, no. 6, pp. 2013–2035, 2003.
- [8] J. D. Storey, "The optimal discovery procedure: a new approach to simultaneous significance testing," *Journal of the Royal Statistical Society, Series B*, vol. 69, no. 3, pp. 347–368, 2007.
- [9] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London, Series A*, vol. 231, pp. 289–337, 1933.
- [10] J. D. Storey, J. Y. Dai, and J. T. Leek, "The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments," *Biostatistics*, vol. 8, no. 2, pp. 414–432, 2007.
- [11] E. L. Lehmann, *Testing Statistical Hypotheses*, Springer, Berlin, Germany, 2nd edition, 1986.
- [12] J. Cao, X. J. Xie, S. Zhang, A. Whitehurst, and M. A. White, "Bayesian optimal discovery procedure for simultaneous significance testing," *BMC Bioinformatics*, vol. 10, article 5, 2009.
- [13] S. Woo, J. T. Leek, and J. D. Storey, "A computationally efficient modular optimal discovery procedure," *Bioinformatics*, vol. 27, no. 4, pp. 509–515, 2011.
- [14] H. Noma and S. Matsui, "The optimal discovery procedure in multiple significance testing: an empirical Bayes approach," *Statistics in Medicine*, vol. 31, no. 2, pp. 165–176, 2012.
- [15] R. Gottardo, J. A. Pannucci, C. R. Kuske, and T. Brettin, "Statistical analysis of microarray data: a Bayesian approach," *Biostatistics*, vol. 4, no. 4, pp. 597–620, 2003.
- [16] C. M. Kendzierski, M. A. Newton, H. Lan, and M. N. Gould, "On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles," *Statistics in Medicine*, vol. 22, no. 24, pp. 3899–3914, 2003.
- [17] K. Lo and R. Gottardo, "Flexible empirical Bayes models for differential gene expression," *Bioinformatics*, vol. 23, no. 3, pp. 328–335, 2007.
- [18] I. Lönnstedt and T. Speed, "Replicated microarray data," *Statistica Sinica*, vol. 12, no. 1, pp. 31–46, 2002.
- [19] M. A. Newton, C. M. Kendzierski, C. S. Richmond, F. R. Blattner, and K. W. Tsui, "On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data," *Journal of Computational Biology*, vol. 8, no. 1, pp. 37–52, 2001.
- [20] H. Noma and S. Matsui, "Bayesian ranking and selection methods in microarray studies," in *Statistical Diagnostics for Cancer*, F. Emmert-Streib and M. Dehmer, Eds., pp. 57–74, Wiley-VCH, Weinheim, Germany, 2013.
- [21] H. Noma, S. Matsui, T. Omori, and T. Sato, "Bayesian ranking and selection methods using hierarchical mixture models in microarray studies," *Biostatistics*, vol. 11, no. 2, pp. 281–289, 2010.
- [22] N. M. Laird and T. A. Louis, "Smoothing the non-parametric estimate of a prior distribution by roughening. A computational study," *Computational Statistics and Data Analysis*, vol. 12, no. 1, pp. 27–37, 1991.
- [23] W. Shen and T. A. Louis, "Empirical bayes estimation via the smoothing by roughening approach," *Journal of Computational and Graphical Statistics*, vol. 8, no. 4, pp. 800–823, 1999.
- [24] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate distribution," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 197–206, University of California Press, Berkeley, Calif, USA, 1956.
- [25] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.



- [26] B. Efron and C. Morris, "Limiting the risk of Bayes and empirical Bayes estimators, part II: the empirical Bayes case," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 130–139, 1972.
- [27] B. Efron and C. Morris, "Data analysis using Stein's estimator and its generalizations," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 311–319, 1975.
- [28] C. N. Morris, "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 47–55, 1983.
- [29] N. M. Laird, "Nonparametric maximum likelihood estimation of a mixing distribution," *Journal of the American Statistical Association*, vol. 73, no. 364, pp. 805–811, 1978.
- [30] N. M. Laird, "Empirical Bayes estimate using the nonparametric maximum likelihood estimate of the prior," *Journal of Statistical Computation and Simulation*, vol. 15, no. 2-3, pp. 211–220, 1982.
- [31] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [32] H. Noma and S. Matsui, "Optimality of gene ranking based on univariate  $P$ -values for detecting differentially expressed genes," *Japanese Journal of Biometrics*, vol. 31, no. 1, pp. 13–21, 2010.
- [33] G. J. McLachlan, R. W. Bean, and L. B. T. Jones, "A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays," *Bioinformatics*, vol. 22, no. 13, pp. 1608–1615, 2006.
- [34] M. A. Newton, A. Noueiry, D. Sarkar, and P. Ahlquist, "Detecting differential gene expression with a semiparametric hierarchical mixture method," *Biostatistics*, vol. 5, no. 2, pp. 155–176, 2004.
- [35] J. T. Leek, E. Monsen, A. R. Dabney, and J. D. Storey, "EDGE: extraction and analysis of differential gene expression," *Bioinformatics*, vol. 22, no. 4, pp. 507–508, 2006.
- [36] B. Efron, "Size, power and false discovery rates," *Annals of Statistics*, vol. 35, no. 4, pp. 1351–1377, 2007.
- [37] B. Efron, "Microarrays, empirical Bayes and the two-groups model," *Statistical Science*, vol. 23, no. 1, pp. 1–22, 2008.
- [38] D. Singh, P. G. Febbo, K. Ross et al., "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 203–209, 2002.
- [39] S. S. Dave, G. Wright, B. Tan et al., "Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells," *The New England Journal of Medicine*, vol. 351, no. 21, pp. 2159–2169, 2004.
- [40] Y. Zhao and R. Simon, "BRB-ArrayTools Data Archive for human cancer gene expression: a unique and efficient data sharing resource," *Cancer Informatics*, vol. 6, pp. 9–15, 2008.
- [41] S. Matsui and H. Noma, "Estimation and selection in high-dimensional genomic studies for developing molecular diagnostics," *Biostatistics*, vol. 12, no. 2, pp. 223–233, 2011.
- [42] S. Matsui and H. Noma, "Estimating effect sizes of differentially expressed genes for power and sample-size assessments in microarray experiments," *Biometrics*, vol. 67, no. 4, pp. 1225–1235, 2011.
- [43] H. Noma and S. Matsui, "Empirical Bayes ranking and selection methods via semiparametric hierarchical mixture models in microarray studies," *Statistics in Medicine*. In press.