Statistics and Its Interface Volume 4 (2011) 417–430

Optimal false discovery rate control for dependent data

Jichun Xie, T. Tony Cai, John Maris and Hongzhe Li*

This paper considers the problem of optimal false discovery rate control when the test statistics are dependent. An optimal joint oracle procedure, which minimizes the false non-discovery rate subject to a constraint on the false discovery rate is developed. A data-driven marginal plug-in procedure is then proposed to approximate the optimal joint procedure for multivariate normal data. It is shown that the marginal procedure is asymptotically optimal for multivariate normal data with a short-range dependent covariance structure. Numerical results show that the marginal procedure controls false discovery rate and leads to a smaller false non-discovery rate than several commonly used p-value based false discovery rate controlling methods. The procedure is illustrated by an application to a genome-wide association study of neuroblastoma and it identifies a few more genetic variants that are potentially associated with neuroblastoma than several p-value-based false discovery rate controlling procedures.

Keywords and phrases: Large scale multiple testing, Marginal rule, Optimal oracle rule, Weighted classification.

1. INTRODUCTION

False discovery rate control, introduced in the seminal paper by [1], is one of the most important methodological developments in multiple hypothesis testing. Although the original false discovery rate controlling procedure was developed for independent p-values, Benjamini and Yekutieli [3] showed that one can control the false discovery rate by the p value-based procedure under a certain positive dependency assumption, and thus demonstrated that the p-value based procedure can be adaptive to certain dependency structure. The result has recently been generalized to a family of step-up procedures that still control the false discovery rate under more general dependency among the p-values [4, 5, 20]. When the proportion of the true nulls is relatively small, these procedures are often conservative. To overcome this drawback, adaptive plug-in procedures have been developed by incorporating an estimator of the unknown proportion of the nulls in the threshold of the previous procedures [2, 5, 21].

*Corresponding author.

The focus of these new procedures is mainly on controlling the false discovery rate when the test statistics or the p-values are dependent. However, efficiency issue has not been studied in these papers. Efron [8] used a local false discovery rate to carry out both size and power calculations on large-scale testing problems. Efron [9] further investigated the issue of correlated z-values and accuracy of large-scale statistical estimates under dependency. Sun and Cai [23] developed an adaptive multiple testing rule for false discovery rate control and showed that such a procedure is optimal in the sense that it minimizes the false non-discovery rate while controlling the false discovery rate. In particular, the marginal false discovery rate and marginal false non-discovery rate are used as the criteria for multiple testing, where the marginal false discovery rate is defined as mFDR $= E(N_{10})/E(R)$, the proportion of the expected number of nulls (N_{10}) among the expected number of rejections (R), and marginal false non-discovery rate is defined as mFNR $= E(N_{01})/E(S)$, the proportion of the expected number of non-nulls among the expected number of non-rejections (S). Genovese and Wasserman [10] showed that under the independence assumption, mFDR and FDR (mFNR and FNR) are asymptotically the same in the sense that mFDR $=$ FDR $+ O(m^{-1/2})$, where m is the number of hypotheses being tested. It will be shown in Section 2 that such asymptotic equivalence holds in a more general setting.

Sun and Cai [24] considered the case where the underlying latent indicator variable of being the null follows a homogenous irreducible hidden Markov chain and obtained an asymptotically optimal rule under this hidden Markov model. In this paper, we consider the problem of optimal FDR control for general dependent test statistics. Consider m null hypotheses and let θ_i take value 0 if the ith null hypothesis is true and 1 otherwise. The null hypotheses can then be written as $\mathrm{H}_i^0 : \theta_i = 0$ $(i = 1, \dots, m)$. To test these hypotheses, one has a sequence of test statistics $x = (x_1, \dots, x_m)$, which represents a realization of a random vector X where

$$(1)\quad X \mid \theta \sim g(x \mid \theta), \quad \theta_i \overset{i.i.d.}{\sim} \mathrm{Bernoulli}(p), \quad i = 1, \dots, m.$$

We first present an oracle decision rule under this general model for dependent statistics. Our development follows that of [23] by showing that the large-scale multiple testing problem has a corresponding equivalent weighted classification problem in the sense that the optimal solution to multiple testing is also the optimal decision rule for weighted classification. However, our development does not require the monotone likelihood ratio condition for such an equivalence. In particular, let $I\{.\}$ or $I(.)$ be the indicator function, Sun and Cai [23] focused on a class of decision rules

$$\mathcal{S} = \{\delta(x) : \delta_i(x) = \mathrm{I}\{T(x_i) < c\}\},$$

where $T(x_i)$ is a function of x_i that can depend on unknown quantities such as proportion of nonnulls and/or the distribution of X_i. Sun and Cai [23] further assumes that $T(x_i)$ has monotone likelihood ratio. It was shown that the optimal solution to the weighted classification problem is optimal in $\mathcal{S}$ for the multiple testing problem. Our results show that the optimal solution to the weighted classification problem is optimal among all decision rules for multiple testing. Based on the classification rule, the optimal oracle rule for multiple testing under the dependency model (1) is obtained.

We further consider the case when X follows a multivariate normal distribution, where the observation x has the following distribution,

(2)
$$X \mid \theta \sim \mathrm{N}(\mu \mid \theta, \Sigma), \quad \theta_i \overset{i.i.d.}{\sim} \mathrm{Bernoulli}(p), \quad i = 1, \dots, m,$$

where $\mu \mid \theta$ is the conditional mean vector and Σ is the covariance matrix. Under this model and the short-range dependency of Σ, the marginal oracle statistics are shown to be a uniformly consistent approximation to the joint oracle statistics. The marginal oracle statistics are much easier to compute than the joint oracle statistics. We also develop a data-driven marginal plug-in procedure and establish its optimality for the FDR control. Extensive simulations and an application to a genome-wide association study of neuroblastoma are presented to demonstrate the numerical properties of this marginal plug-in procedure. The numerical results show that the marginal procedure controls false discovery rate and leads to a smaller false non-discovery rate than several commonly used p-value based false discovery rate controlling methods.

2. ORACLE DECISION RULE FOR MULTIVARIATE TEST STATISTICS

In this section, we develop an oracle decision rule for multiple testing under dependence using a compound decision-theoretic framework. Consider m null hypotheses H_i^0, $i = 1, \dots, m$ and let θ_i take value 0 if the ith null hypothesis is true and 1 otherwise. In this paper we shall assume that $\theta_1, \dots, \theta_m$ are independent and identically distributed Bernoulli variables with success probability p. Let $X = (X_1, \dots, X_m)$ be a sequence of test statistics for the m null hypotheses $\mathrm{H}_i^0 : \theta_i = 0$ with the following density,

$$\text{(3)} \quad x \mid \theta \sim g(x \mid \theta), \quad \theta_i \overset{i.i.d.}{\sim} \mathrm{Bernoulli}(p), \quad i = 1, \dots, m.$$

Based on an observation of X: $x = (x_1, \dots, x_m)$, the goal is to construct a multiple testing procedure $\delta = (\delta_1, \dots, \delta_m)$ which achieves the minimum mFNR while controlling mFDR at a pre-specified level α. From the model (3), we note that the marginal distribution of X_i now depends not just on θ_i but the whole sequence of θ. Therefore, the methods of [23] and [24] cannot be applied to the model we consider here.

As in [23] for the independent case, our goal is to develop in the dependent case a multiple testing procedure which minimizes the marginal false non-discovery rate while controlling the marginal false discovery rate. As mentioned in the introduction, Genovese and Wasserman [10] showed that under the independence assumption, mFDR and FDR (mFNR and FNR) are asymptotically the same in the sense that $\text{mFDR} = \text{FDR} + O(m^{-1/2})$. The following theorem shows that the asymptotic equivalence between the FDR and the mFDR holds in a more general short-range dependence setting.

Theorem 1. *Suppose that $X = (X_1, \dots, X_m)$ is a sequence of random variables with the same marginal density f and the short range dependency, so that X_i and X_j are independent if $|i - j| > m^\tau$, for $0 \le \tau < 1$. Let $\hat{\delta}_i = \mathrm{I}(S_i \in \mathcal{R})$ be a short-ranged rule to test H_i^0, in the sense that S_i only depends on the variables that are dependent with X_i,*

$$S_i = S(X_{i-\lfloor m^\tau \rfloor}, \dots, X_{i+\lfloor m^\tau \rfloor}) \quad (0 \le \tau < 1).$$

Further suppose that

$$\mathrm{pr}\,(S_i \in \mathcal{R}, \theta_i = 1) \ge \mathrm{pr}\,(S_i \in \mathcal{R}, \theta_i = 0),$$

and

$$\mathrm{pr}\,(S_i \in \mathcal{R}, \theta_i = 1) > 0 \quad (i = 1, \dots, m).$$

Then the FDR (FNR) *of the rule $\hat{\delta}$ can be approximated by the* mFDR (mFNR)*,*

$$\lim_{m\to\infty} (\text{FDR} - \text{mFDR}) = 0, \qquad \lim_{m\to\infty} (\text{FNR} - \text{mFNR}) = 0.$$

It has been shown that the multiple testing problem is equivalent to a weighted classification problem in the independent case [23] and under the dependency specified by the hidden Markov model [24]. We consider the corresponding weighted classification problem under the model (3) and have the following theorem.

Theorem 2. *Define the loss function:*

$$\text{(4)} \quad L_\lambda(\delta, \theta) = \frac{1}{m} \sum_i \{\delta_i \,\mathrm{I}\,(\theta_i = 0) + \lambda(1 - \delta_i)\,\mathrm{I}\,(\theta_i = 1)\}.$$

Consider the model defined in (3). Suppose that p and g are known. Then the classification risk $E\{L_\lambda(\theta, \delta)\}$ is minimized by the Bayes rule $\delta(\Lambda, \lambda) = (\delta_1, \dots, \delta_m)$, where

$$\text{(5)} \qquad \delta_i = \mathrm{I}\left\{\Lambda_i(x) = \frac{(1-p)g(x \mid \theta_i = 0)}{pg(x \mid \theta_i = 1)} < \lambda\right\},$$

for $i = 1, \ldots, m$. The minimum classification risk is

$$\mathcal{R}_\lambda(\delta(\Lambda, \lambda)) = p + \int_K \{(1-p)g(x \mid \theta_i = 0) - \lambda p g(x \mid \theta_i = 1)\}\, \mathrm{d}x,$$

where $K = \{x \in \Omega : (1-p)g(x \mid \theta_i = 0) < \lambda p g(x \mid \theta_i = 1)\}$.

The rule given in Theorem 2 is optimal for the weighted classification problem. We next show that the optimality property can be extended to the multiple testing problem. Consider the optimal rule $\delta(\Lambda, \lambda)$ as defined in (5). Let $G_i^s(t) = \mathrm{pr}(\Lambda_i \le t \mid \theta_i = s)$, $s = 0, 1$, be the conditional cumulative density functions (CDF) of $\Lambda_i(x)$. The CDF of $\Lambda_i(x)$ is then given by $G_i(t) = \mathrm{pr}(\Lambda_i \le t) = (1-p)G_i^0(t) + pG_i^1(t)$. Define the average conditional CDF's of Λ, $G^s(t) = (1/m)\sum_{i=1}^m G_i^s(t)$ and average conditional probability density functions of Λ, $g^s(t) = (\mathrm{d}/\mathrm{d}t)G^s(t)$, $s = 0, 1$. The following theorem shows that Λ is also optimal for multiple testing.

Theorem 3. *Let $\mathcal{D}_s = \{\delta : \delta_i = \mathrm{I}(\Lambda_i < \lambda), i = 1, \ldots, m\}$, where Λ_i's are defined in (5). Given an mFDR level α and a decision rule*

$$\delta(S, R) = \{\mathrm{I}(S_1(x) \in R_1), \ldots, \mathrm{I}(S_m(x) \in R_m)\}$$

with $\mathrm{mFDR}\{\delta(S,R)\} \le \alpha$, then there exists a λ determined by $\delta(S, R)$, such that $\delta(\Lambda, \lambda) \in \mathcal{D}_s$ outperforms $\delta(S, R)$ in the sense that

$$\mathrm{mFDR}\{\delta(\Lambda, \lambda)\} \le \mathrm{mFDR}\{\delta(S, R)\} \le \alpha,$$

and

$$\mathrm{mFNR}\{\delta(\Lambda, \lambda)\} \le \mathrm{mFNR}\{\delta(S, R)\}.$$

Theorem 3 reveals that the optimal solution for the multiple testing problem belongs to the set $\mathcal{D}_s$. Instead of searching for all decision rules, one only needs to search in the collection $\mathcal{D}_s$ for the optimal rule. The following result shows that for a given mFDR value α, the optimal rule for the multiple testing problem is unique.

Theorem 4. *Consider the optimal decision rule $\delta(\Lambda, \lambda)$ in (5) for the weighted classification problem with the loss function (4). There exists a unique $\lambda(\alpha)$, such that $\delta\{\Lambda, \lambda(\alpha)\}$ controls mFDR at level α and minimizes mFNR among all decision rules.*

Theorem 4 shows that there exists a one-to-one mapping between the mFDR value α and the thresholding λ, which determines the optimal rule. However, it is often hard to obtain the corresponding $\lambda(\alpha)$ for a given α, in which case we need to develop an oracle rule that depends on α directly instead of $\lambda(\alpha)$. Define

$$T_{\mathrm{OR},i} = \frac{(1-p)g(x \mid \theta_i = 0)}{g(x)}, \tag{6}$$

and, clearly $T_{\mathrm{OR},i} = \Lambda_i/(1+\Lambda_i)$ and increases with Λ_i. Thus, for a given mFDR value α, one can rewrite the optimal oracle rule (5) as

$$\delta_{\mathrm{OR},i} = \mathrm{I}\left\{T_{\mathrm{OR},i} < \tilde{\lambda}(\alpha) = \frac{\lambda(\alpha)}{1+\lambda(\alpha)}\right\}. \tag{7}$$

Let $R_{\tilde{\lambda}} = \sum_{i=1}^m \mathrm{I}(T_{\mathrm{OR},i} < \tilde{\lambda})/m$, $V_{\tilde{\lambda}} = \sum_{i=1}^m \mathrm{I}(T_{\mathrm{OR},i} < \tilde{\lambda}, \theta_i = 0)/m$ and $Q_{\tilde{\lambda}} = V_{\tilde{\lambda}}/R_{\tilde{\lambda}}$. Then,

$$\begin{aligned}\mathrm{mFDR} = E(Q_{\tilde{\lambda}}) &= E\{E(V_{\tilde{\lambda}}/R_{\tilde{\lambda}} \mid x)\} \\ &= E\left\{\frac{\sum_{i=1}^m \mathrm{I}(T_{\mathrm{OR},i} < \tilde{\lambda})T_{\mathrm{OR},i}}{\sum_{i=1}^m \mathrm{I}(T_{\mathrm{OR},i} < \tilde{\lambda})}\right\} = E\left(\frac{\sum_{i=1}^{mR_{\tilde{\lambda}}} T_{\mathrm{OR},(i)}}{mR_{\tilde{\lambda}}}\right),\end{aligned}$$

where $T_{\mathrm{OR},(i)}$ is the ith order statistic of $T_{\mathrm{OR},i}$ ($i = 1, \ldots, m$). Suppose the number of rejections is $k = mR$, then the false discovery proportion is controlled at α if

$$\frac{1}{k}\sum_{i=1}^k T_{\mathrm{OR},(i)} \le \alpha. \tag{8}$$

If for every k, (8) is satisfied, then

$$\mathrm{mFDR} = E\left(\sum_{i=1}^R T_{\mathrm{OR},(i)}/R\right) \le \alpha.$$

Based on the argument presented above, we have the following joint oracle procedure for multiple testing with dependent statistics.

Theorem 5. *Consider the model defined in (3). Suppose that p and g are known. Define $T_{\mathrm{OR},i}$ as in (6). Then the following procedure controls mFDR at level α:*

$$\text{Reject all } \mathrm{H}_{(i)}\ (i = 1, \ldots, k), \quad \text{where } k = \max\left\{l : \frac{1}{l}\sum_{i=1}^l T_{\mathrm{OR},(i)} \le \alpha\right\}. \tag{9}$$

The final oracle rule (9) consists of two steps: the first step is to calculate the oracle statistics $T_{\mathrm{OR}} = (T_{\mathrm{OR},1}, \ldots, T_{\mathrm{OR},m})$, and the second step is to rank the statistics and calculate the running averages from the smallest to the largest in order to determine the cutoff. We then reject all the hypotheses with $T_{\mathrm{OR},(i)}$ below the cutoff.

3. MARGINAL APPROXIMATION TO THE JOINT ORACLE PROCEDURE FOR MULTIVARIATE NORMAL DATA

3.1 Marginal oracle rule

Theorems 3–5 show the optimality of the joint oracle rule for multivariate test statistics. However, the oracle rule assumes that the non-null proportion p and the distribution

$g(x|\theta)$ are both known. Even if $g(x|\theta)$ is known, it is still computationally challenging to calculate $g(x|\theta_i = 0)$ and the mixture distribution

$$g(x) = \sum_{\theta_1,\ldots,\theta_m} \left\{ g(x\,|\,\theta) \prod_{i=1}^{m} p^{\theta_i}(1-p)^{1-\theta_i} \right\}.$$

The computational complexity to obtain T_{OR} is $O(m2^m)$.

To resolve the computational difficulty associated with the oracle rule, we show in the following section that under the multivariate normal model (3) with short-range dependence covariance structure, certain marginal oracle statistics can be used to approximate the joint oracle statistics, which leads to a computationally feasible optimal mFDR controlling procedure.

We make the following additional assumptions on the model (3):

(A) The non-null proportion goes to zero: $\lim_{m\to\infty} p^{(m)} = 0$.

(B) The data $x^{(m)} = (x_1, \ldots, x_m)$ is an observation of the random variable $X^{(m)} = (X_1, \ldots, X_m)$, which follows a multivariate normal distribution given the latent variable $\theta^{(m)} = (\theta_1, \ldots, \theta_m)$,

$$X^{(m)}|\,\theta^{(m)} \sim \text{N}(\mu^{(m)} \mid \theta^{(m)}, \Sigma^{(m)}), \tag{10}$$

where $(\mu^{(m)} \mid \theta^{(m)}) = (\mu_1 \mid \theta_1, \ldots, \mu_m \mid \theta_m)$. The variable $\mu_i \mid (\theta_i = 0)$ follows a point mass distribution at point 0 and $\mu_i \mid (\theta_i = 1)$ follows the distribution with CDF $F^1(\mu)$. Without loss of generality, assume X is rescaled so that $\Sigma_{ii}^{(m)} = 1$. Under this model, X_i has the same unknown marginal probability density

$$f(x) = \int \varphi(x-\mu)\,\mathrm{d}F(\mu), \quad \text{where } F(\mu) = (1-p)I(\mu \geq 0) + pF^1(\mu). \tag{11}$$

(C) The minimum eigenvalue of $\Sigma^{(m)}$ is bounded away from 0:

$$\liminf_{m\to\infty} \lambda_{\min}(\Sigma^{(m)}) = \kappa > 0.$$

(D) The correlation structure of $X^{(m)}$ is short ranged with $\Sigma_{ik}^{(m)} = 0$ whenever $|i-k| \geq m^\tau$ for some constant $\tau \in (0,1)$.

Define

$$T_{\text{MG},i} = (1-p)f(x_i|\,\theta_i = 0)/f(x_i) \tag{12}$$

to be the marginal oracle rule, which only involves the marginal distributions of x_i. The following theorem shows that as $m \to \infty$, $T_{\text{MG},i}$ can approximate $T_{\text{OR},i}$ well.

Theorem 6. *Under the assumptions* (A)–(C)*, let* $T_{\text{OR},i}$ *and* $T_{\text{MG},i}$ *be defined as in equations* (6) *and* (12)*. Then for all* $\epsilon > 0$ *and for all* $i = 1, \ldots, m$,

$$\lim_{m\to\infty} \text{pr}\,(|\,T_{\text{MG},i} - T_{\text{OR},i}\,| > \epsilon) = 0.$$

Theorem 6 reveals that the marginal oracle statistics $T_{\text{MG}} = (T_{\text{MG},i}, \ldots, T_{\text{MG},m})$ are a uniformly consistent approximation to the joint oracle rule determined by T_{OR}. Note that T_{MG} in Theorem 6 is a separable rule with a computation complexity of $O(m)$, much smaller than the complexity of the joint oracle rule.

3.2 Estimating the marginal oracle statistics

In model (10), the marginal densities $f(x_i)$'s are the same for all $i = 1, \ldots, m$, as well as $f(x_i\,|\,\theta_i)$. From now on, we use f to denote the marginal density and let f, f_0 and f_1 be the marginal density, marginal density under the null and marginal density under the alternative for X_i. Denote estimator of f, f_0, and the non-null proportion p as $\hat{f}$, $\hat{f}_0$ and $\hat{p}$. Let $\hat{T}_{\text{MG},i} = \{(1-\hat{p})\hat{f}_0(x_i)/\hat{f}(x_i)\} \wedge 1$.

For many multiple testing problems, the theoretical null distribution f_0 for X_i is typically known. In short-range dependency cases, Cai and Jin [6] provided an estimator for p with the minimax convergence rate. Estimating f under the assumptions (A)–(D) is not straightforward even when f is a normal mixture under the assumption (B). However, it is easy to obtain a non-parametric density estimate under the short-range dependency assumption. Suppose that the range of correlation is $B = m^\tau$ $(0 \leq \tau < 1)$ as in assumption (B). Define $K = m/B = m^{1-\tau}$. We rank each coordinate of x in the following way so that it becomes a matrix:

$$\begin{array}{cccccc} & k=1 & 2 & \cdots & K \\ b=1 & x_1 & x_{B+1} & \cdots & x_{(K-1)B+1} \\ 2 & x_2 & x_{B+2} & \cdots & x_{(K-1)B+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B & x_B & x_{2B} & \cdots & x_{KB} \end{array} \tag{13}$$

To facilitate the discussion, in the remaining part of this subsection, we use double subindex to denote x_i, i.e. $x_{b,k}$ is $x_{(k-1)B+b}$ in the original vector notation. Note that each row of (13) is an independent subsequence of x with length $K = m^{1-\tau}$. Therefore, we can obtain a kernel estimator $\hat{f}_b(x)$ based on $x_{b,\cdot} = (x_{b,1}, \ldots, x_{b,K})$, $b = 1, \ldots, B$. Define $\hat{f}(x_0) = \sum_{b=1}^{B} \hat{f}_b(x_0)/B$. Note that f is a mixture of normal densities and therefore, it is infinitely differentiable. Van der Vaart [26] showed that for any bounded a differentiable function f such that $\int|f^{(a)}(z)|\,\mathrm{d}z \leq M$, with appropriately chosen bandwidth $h = n^{-a/(1+2a)}$, where n is the sample size, for any x_0, the kernel estimate $\hat{f}(x_0)$ can achieve the bound

$$E\{\hat{f}(x_0) - f(x_0)\}^2 \leq Cn^{-2a/(1+2a)}, \tag{14}$$

where C only depends on the true marginal density f, the Kernel function and the bandwidth h. In our setting, set the sample size $n = m/B = K$. For each row of (13), the kernel estimator $\hat{f}_b(x)$ can attain the estimation upper bound (14). Therefore,

$$E\{\hat{f}_b(x_0) - f(x_0)\}^2 \leq CK^{-2a/(1+2a)} = Cm^{-2a(1-\tau)/(1+2a)},$$

for $b = 1, \ldots, B$. Define $\hat{f}(x_0) = \sum_{b=1}^{B} \hat{f}_b(x_0)/B$. When $0 \leq \tau < 1$,

$$E\{\hat{f}(x_0) - f(x_0)\}^2 = \frac{1}{B^2} \sum_{b=1}^{B} E\{\hat{f}_b(x_0) - f(x_0)\}^2 \to 0.$$

The kernel estimators of $f(x_0)$ can be written as $\hat{f}_b(x_0) = \sum_{k=1}^{K} w(x_0, X_{b,k})/K$, where w is the kernel function. Thus

$$\hat{f}(x_0) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}_b(x_0) = \frac{1}{KB} \sum_{b=1}^{B} \sum_{k=1}^{K} w(x_0, X_{b,k})$$

is the same as the kernel estimator viewing $X_1, \ldots, X_m$ as independent and identically distributed.

3.3 Asymptotic validity and optimality of the marginal plug-in procedure

After obtaining the estimators for p, f and f_0, we define the marginal plug-in procedure as

$$\text{(15)} \quad \text{Reject all } \mathrm{H}_{(i)} \ (i = 1, \ldots, k), \quad \text{where } k = \max\left\{ i : \frac{1}{i} \sum_{j=1}^{i} \hat{T}_{\mathrm{MG},(j)} \leq \alpha \right\},$$

where $\hat{T}_{\mathrm{MG},(j)}$ is the jth order statistic of $\hat{T}_{\mathrm{MG},j}$ $(j = 1, \ldots, m)$.

The next theorem shows that the plug-in procedure (15) is asymptotically valid and optimal in the sense that it asymptotically controls mFDR under the given level α and minimizes mFNR.

Theorem 7. *Assume θ_i follows* Bernoulli(p) *independently $(i = 1, \ldots, m)$. Let $x = (x_1, \ldots, x_m)$ be dependent observations satisfying the Assumptions* (A), (B) *and* (C). *Let $\hat{p}$ be the consistent estimator of p, and $\hat{f}$, $\hat{f}_0$ be estimators of f and f_0 satisfying for all x_0, $E\{\hat{f}(x_0) - f(x_0)\}^2 \to 0$ and $E\{\hat{f}_0(x_0) - f_0(x_0)\}^2 \to 0$. Define $\hat{T}_{\mathrm{MG},i} = (1-\hat{p})\hat{f}_0(x_i)/\hat{f}(x_i)$. The plug-in procedure* (15) *asymptotically controls the m*FDR *at the given level α and simultaneously minimizes the m*FNR.

4. SIMULATION STUDIES

4.1 Simulation 1: The performance of the marginal oracle rule

In this section we evaluate the numerical performance of the marginal procedure. In our simulations, we assumed a multivariate mixture normal model:

$$\text{(16)} \qquad X \mid \theta \sim \mathrm{N}(\mu \mid \theta, \Sigma),$$

where θ_i follows Bernoulli(p). We chose the parameters so that the Assumptions (A)–(D) hold. All simulation results were based on 1,000 replications.

In the following simulations, we set the number of hypotheses as $m = 6,000$ and assume that the covariance matrix Σ is a block-diagonal matrix with block size of 30. Within each block, the sub-precision matrix is a banded matrix with bandwidth 1, diagonal elements of 1 and off-diagonal elements of 0.2. We set $\mu(\theta_i = 0) = 0$. We have developed an efficient computational algorithm to compute the joint oracle statistics under the banded precision matrix assumption with computation complexity of $O(2^B m)$, where B is bandwidth for the precision matrix. Our aim is to compare the performance of the optimal oracle rule with the marginal oracle rule when the true parameters are known.

We considered different settings of the parameters and present the results in Figure 1. Overall, we observe that the results from the marginal oracle rule are similar to those from the optimal joint oracle rule. Both procedures control the FDR at the desired levels and have smaller mFNRs than the Benjamini and Hochberg procedure. The upper left and upper right panels show the observed FDR versus the alternative mean value for mFDR $= 0.05$, and $p = 0.20$ and $p = 0.02$, respectively, indicating that the marginal oracle rule can indeed control FDR at the desired level. As expected, the Benjamini and Hochberg procedure is quite conservative. The upper middle left and right plots show the FNR as a function of the mFDR for $\mu(\theta_i = 1) = 2.50$, and $p = 0.20$ and $p = 0.02$, respectively. The lower middle left and right plots show the FNR versus the alternative mean value with mFDR $= 0.05$, and $p = 0.20$ and $p = 0.02$, respectively. These plots indicate that the joint oracle and the marginal oracle rules perform similarly, and both have smaller FNR than the BH procedure. Finally, the bottom plot shows the FNR as a function of the non-null proportion p for mFDR $= 0.05$ and $\mu_1 = 3.5$. We observe that as p increases, the Benjamini and Hochberg procedure becomes increasingly more conservative, but the marginal oracle rule still performs almost as efficiently as the joint oracle rule.

4.2 Simulation 2: The performance of the plug-in marginal rule

In this simulation, we evaluated the performance of the marginal plug-in procedure and compared this with other p-value-based procedures, including Storey's α procedure [21], two two-stage adaptive procedures of [5], the adaptive procedure of [2], the q-value procedure [22] and the Benjamini and Hochberg procedure [1], in different settings, to evaluate their empirical FDR and FNR. We chose the Benjamini and Hochberg procedure since it is the first and still one of the most widely used procedures for FDR control. The Storey's α procedure was chosen since [5] compared it with several other procedures and concluded that it is the best procedure among them in terms of controlling FDR and increased power under both independent and dependent settings.The q-value

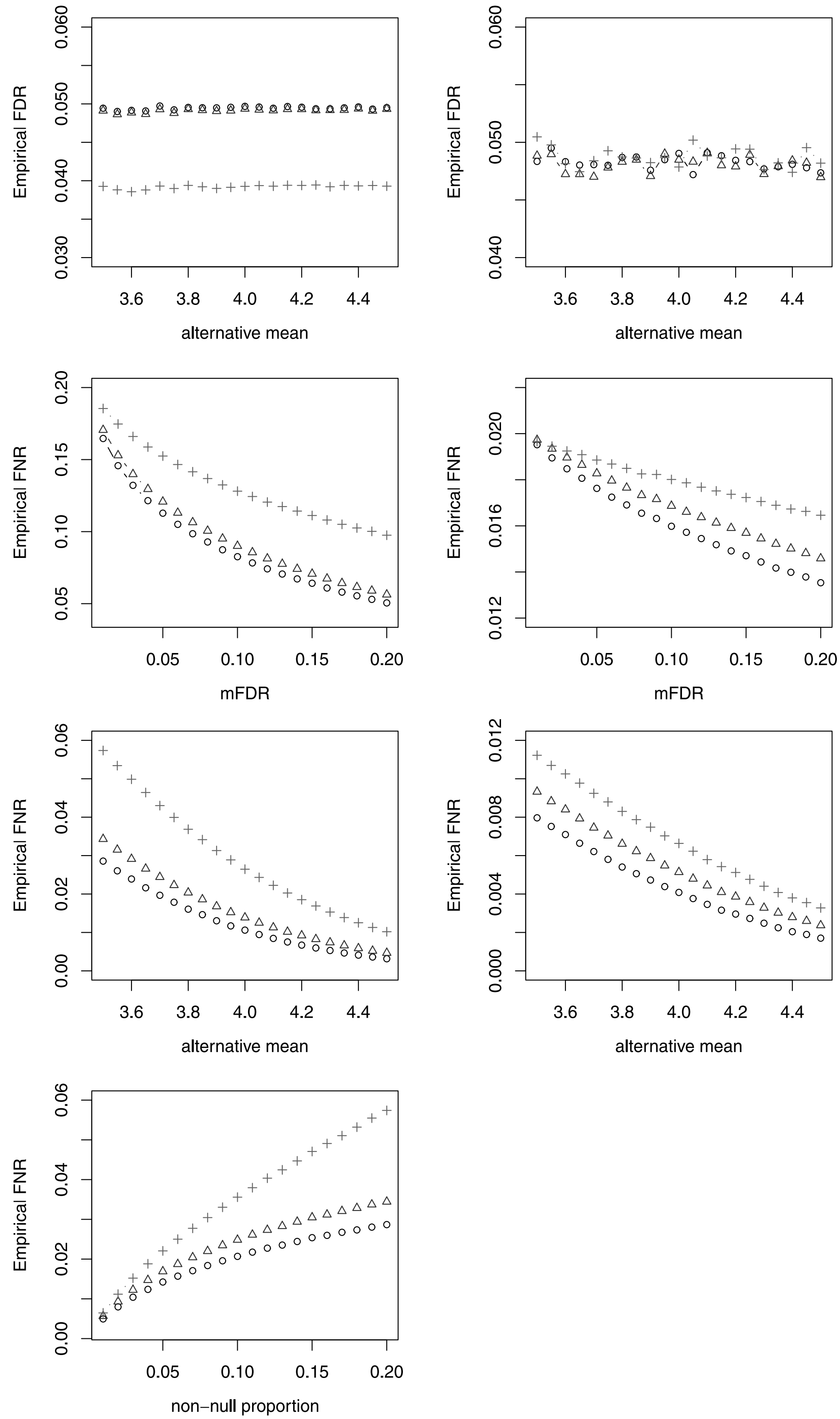

Figure 1. Comparison of the joint oracle procedure (black circle), the marginal oracle procedure (dark grey triangle) and the Benjamini and Hochberg procedure (grey plus sign). The upper left: $\mathrm{mFDR} = 0.05$, $p = 0.2$; *the upper right:* $\mathrm{mFDR} = 0.05$, $p = 0.02$; *the upper middle left:* $\mu_1 = 2.5$, $p = 0.2$; *the upper middle right:* $\mu_1 = 2.5$, $p = 0.02$; *the lower middle left:* $\mathrm{mFDR} = 0.05$, $p = 0.2$; *the lower middle right:* $\mathrm{mFDR} = 0.05$, $p = 0.02$; *the lower left:* $\mu_1 = 3.5$, $\mathrm{mFDR} = 0.05$.

Table 1. Comparison of empirical FDR*(10^{-2} unit) vs.* mFDR*(10^{-2} unit) of six different procedures, including the marginal plug-in procedure (MG), the original Benjamini & Hochberg's procedure (BH), the Storey's α procedure (Storey), the Blanchard & Roquain's two stage procedure with Holm's step-down for the first step (BR I), and the Blanchard & Roquain's two stage procedure controlling the family-wise error rate for the first step (BR II), the procedure of [2] (BK), and the q-value procedure (QV)*

		mFDR				
Model	Method	1	3	5	7	9
A1	MG	1.00(0.31)	3.07(0.50)	5.26(0.63)	7.45(0.75)	9.61(0.80)
	BH	0.79(0.28)	2.34(0.47)	3.94(0.57)	5.52(0.70)	7.10(0.76)
	Storey	0.98(0.30)	2.89(0.52)	4.90(0.65)	6.87(0.81)	8.85(0.86)
	BR I	0.39(0.20)	1.17(0.34)	1.95(0.42)	2.74(0.51)	3.54(0.56)
	BR II	0.39(0.20)	1.17(0.34)	1.95(0.42)	2.74(0.51)	3.54(0.56)
	BK	0.97(0.30)	2.89(0.52)	4.90(0.65)	6.86(0.81)	8.83(0.86)
	QV	1.00(0.31)	2.93(0.53)	4.96(0.68)	6.94(0.85)	8.88(0.91)
A2	MG	1.03(1.07)	3.21(1.72)	5.44(2.14)	7.87(2.35)	10.33(2.71)
	BH	0.96(1.09)	2.93(1.73)	4.86(2.22)	6.88(2.48)	8.79(2.93)
	Storey	0.98(1.11)	2.97(1.74)	4.93(2.26)	6.97(2.51)	8.92(2.94)
	BR I	0.50(0.82)	1.49(1.27)	2.38(1.64)	3.42(1.83)	4.40(2.10)
	BR II	0.50(0.82)	1.49(1.27)	2.38(1.63)	3.42(1.83)	4.40(2.10)
	BK	1.08(1.14)	3.25(1.80)	5.29(2.29)	7.42(2.51)	9.38(2.92)
	QV	1.00(1.10)	3.00(1.75)	4.99(2.27)	7.04(2.53)	9.01(2.94)
A3	MG	0.98(0.31)	3.09(0.56)	5.27(0.73)	7.45(0.82)	9.72(0.97)
	BH	0.77(0.27)	2.32(0.50)	3.93(0.64)	5.49(0.71)	7.12(0.84)
	Storey	0.94(0.30)	2.88(0.56)	4.90(0.73)	6.86(0.81)	8.89(0.95)
	BR I	0.38(0.20)	1.14(0.35)	1.95(0.46)	2.74(0.51)	3.55(0.60)
	BR II	0.38(0.20)	1.14(0.35)	1.95(0.46)	2.74(0.51)	3.55(0.60)
	BK	0.94(0.30)	2.87(0.56)	4.89(0.72)	6.85(0.80)	8.88(0.94)
	QV	0.96(0.31)	2.91(0.58)	4.94(0.76)	6.89(0.84)	8.96(1.03)
A4	MG	1.06(2.33)	3.88(4.74)	6.20(5.44)	8.79(6.47)	11.63(7.02)
	BH	0.87(1.88)	3.04(4.00)	4.68(4.59)	6.67(5.68)	8.48(6.09)
	Storey	0.89(1.90)	3.10(4.06)	4.78(4.66)	6.80(5.78)	8.65(6.18)
	BR I	0.45(1.28)	1.55(2.90)	2.33(3.21)	3.42(4.17)	4.21(4.31)
	BR II	0.45(1.28)	1.55(2.90)	2.33(3.21)	3.42(4.17)	4.20(4.30)
	BK	0.98(1.97)	3.35(4.13)	5.10(4.69)	7.18(5.75)	9.07(6.13)
	QV	0.91(1.93)	3.18(4.11)	4.90(4.72)	6.93(5.79)	8.86(6.29)

procedure is chosen because it is also widely used and least conservative. The procedures of [5] and [2] were developed for dependent test statistics and chosen for comparisons.

In all simulations, we set $m = 6,000$, where two thousand variables were generated from a block-structured covariance with block size of 40, two thousand variables were generated from a block-structured covariance of block size of 20, and two thousand variables were generated independently. We generated $\mu(\theta_i = 1)$ from distribution $0.8\,\text{Unif}(-6,-3) + 0.2\,\text{Unif}(3,6)$ and considered four different scenarios:

A1. $p = 0.2$ and within-block covariance matrix is randomly generated and is sparse. For blocks of size of 40, the proportion of non-zero off-diagonal elements is about 0.03 and for blocks of size 20, the proportion of non-zero off-diagonal elements is about 0.08. All the within-block covariance matrices are regularized so that the condition number is equal to the block size.

A2. $p = 0.02$ and within-block covariance matrix structure is the same as Model A1.

A3. $p = 0.2$ and within-block covariance matrix is exchangeable with non-diagonal elements of 0.2.

A4. $p = 0.2$ and within-block covariance matrix is exchangeable with non-diagonal elements of 0.8.

The empirical FDR comparison results are shown in Table 1. We observe that the marginal plug-in procedure controls the FDR well when within-block correlation is low or moderate. However, when the within-block correlation becomes high as in Model A4, the marginal plug-in approach results in inflated FDR. This is due to the use of kernel estimate of the marginal density when the correlations among the observations are high. Although other procedures can control FDR at the desired levels, the procedures of [5] are too conservative in the settings we considered.

We compared the FNR result in the moderate correlation setting where the marginal plug-in procedure can control FDR at the desired level and present the results in Table 2. Within each block, the sub-covariance matrix was randomly generated and was sparse. We generated $\mu(\theta_i = 1)$ from two

Table 2. Comparison of empirical FNR(10^{-2} unit) vs. mFDR, β_2, p_1 and p. See Table 1 for details of these seven methods compared

Model	Method			mFDR		
		0.01	0.03	0.05	0.07	0.09
B1	MG	12.79(0.35)	9.33(0.35)	7.60(0.35)	6.42(0.34)	5.57(0.32)
	BH	14.00(0.36)	10.73(0.39)	9.01(0.39)	7.84(0.38)	6.96(0.36)
	Storey	13.62(0.37)	10.15(0.39)	8.38(0.39)	7.17(0.38)	6.28(0.35)
	BR I	15.68(0.33)	12.88(0.37)	11.31(0.38)	10.19(0.39)	9.37(0.38)
	BR II	15.68(0.33)	12.88(0.37)	11.31(0.38)	10.19(0.39)	9.37(0.38)
	BK	13.78(0.37)	10.33(0.40)	8.54(0.41)	7.32(0.39)	6.41(0.36)
	QV	13.38(0.39)	9.98(0.40)	8.24(0.39)	7.07(0.38)	6.19(0.37)
B2	MG	1.75(0.09)	1.54(0.10)	1.41(0.11)	1.30(0.12)	1.23(0.11)
	BH	1.79(0.09)	1.60(0.10)	1.48(0.12)	1.38(0.12)	1.31(0.12)
	Storey	1.79(0.09)	1.60(0.11)	1.48(0.12)	1.38(0.12)	1.31(0.12)
	BR I	1.86(0.07)	1.73(0.09)	1.64(0.11)	1.56(0.11)	1.51(0.11)
	BR II	1.87(0.07)	1.73(0.09)	1.64(0.11)	1.56(0.11)	1.51(0.11)
	BK	1.78(0.09)	1.59(0.10)	1.47(0.12)	1.37(0.12)	1.30(0.12)
	QV	1.79(0.09)	1.59(0.10)	1.48(0.12)	1.38(0.12)	1.31(0.12)
				β_2		
		3	3.75	4.5	5.25	6
C1	MG	7.61(0.35)	3.60(0.26)	1.93(0.18)	1.37(0.16)	1.18(0.15)
	BH	9.00(0.38)	4.24(0.28)	2.26(0.20)	1.78(0.18)	1.71(0.18)
	Storey	8.37(0.38)	3.79(0.28)	2.02(0.19)	1.62(0.18)	1.56(0.18)
	BR I	11.31(0.38)	5.86(0.32)	3.11(0.23)	2.32(0.19)	2.17(0.20)
	BR II	11.31(0.38)	5.86(0.32)	3.11(0.23)	2.32(0.19)	2.17(0.20)
	BK	8.53(0.40)	3.83(0.28)	2.03(0.19)	1.62(0.18)	1.56(0.18)
	QV	8.23(0.39)	6.08(6.08)	4.43(0.31)	3.23(0.26)	2.46(0.23)
C2	MG	1.41(0.11)	1.16(0.11)	0.92(0.10)	0.72(0.10)	0.56(0.09)
	BH	1.48(0.12)	1.23(0.11)	0.99(0.11)	0.78(0.10)	0.60(0.09)
	Storey	1.48(0.12)	1.23(0.11)	0.99(0.11)	0.78(0.10)	0.59(0.09)
	BR I	1.63(0.11)	1.42(0.11)	1.17(0.12)	0.95(0.11)	0.74(0.10)
	BR II	1.64(0.11)	1.42(0.11)	1.17(0.12)	0.95(0.11)	0.74(0.10)
	BK	1.47(0.12)	1.22(0.12)	0.97(0.11)	0.76(0.10)	0.58(0.09)
	QV	1.47(0.12)	1.23(0.12)	0.98(0.11)	0.77(0.10)	0.59(0.09)
				p_1		
		0	0.25	0.5	0.75	1
D1	MG	5.96(0.32)	7.80(0.34)	8.22(0.36)	7.79(0.35)	5.96(0.31)
	BH	9.00(0.38)	8.99(0.38)	9.00(0.38)	9.00(0.39)	8.97(0.37)
	Storey	8.37(0.38)	8.36(0.38)	8.37(0.38)	8.36(0.38)	8.34(0.37)
	BR I	11.30(0.38)	11.31(0.38)	11.31(0.4)	11.30(0.38)	11.28(0.38)
	BR II	11.30(0.38)	11.31(0.38)	11.31(0.4)	11.30(0.38)	11.28(0.38)
	BK	8.53(0.39)	8.52(0.39)	8.53(0.39)	8.53(0.39)	8.50(0.38)
	QV	8.23(0.39)	8.22(0.39)	8.24(0.39)	8.22(0.39)	8.24(0.40)
D2	MG	1.29(0.12)	1.36(0.12)	1.40(0.11)	1.44(0.11)	1.46(0.11)
	BH	1.48(0.12)	1.48(0.12)	1.47(0.11)	1.48(0.12)	1.48(0.12)
	Storey	1.48(0.12)	1.48(0.12)	1.47(0.11)	1.48(0.12)	1.48(0.12)
	BR I	1.64(0.11)	1.63(0.11)	1.64(0.11)	1.64(0.11)	1.64(0.11)
	BR II	1.64(0.11)	1.63(0.11)	1.64(0.11)	1.64(0.11)	1.64(0.11)
	BK	1.47(0.12)	1.47(0.12)	1.46(0.11)	1.47(0.12)	1.47(0.12)
	QV	1.48(0.12)	1.47(0.12)	1.47(0.11)	1.48(0.12)	1.47(0.12)
				p		
		0.01	0.07	0.13	0.19	0.25
E	MG	0.80(0.07)	3.95(0.21)	6.15(0.27)	7.93(0.33)	9.49(0.38)
	BH	0.81(0.07)	4.04(0.22)	6.48(0.29)	8.63(0.35)	10.74(0.41)
	Storey	0.82(0.08)	3.98(0.22)	6.22(0.29)	8.07(0.35)	9.71(0.40)
	BR I	0.87(0.06)	4.78(0.22)	7.94(0.30)	10.83(0.35)	13.66(0.41)
	BR II	0.88(0.06)	4.78(0.22)	7.94(0.30)	10.83(0.35)	13.66(0.41)
	BK	0.80(0.08)	4.00(0.22)	6.30(0.30)	8.22(0.36)	9.93(0.42)
	QV	0.81(0.07)	3.96(0.22)	6.17(0.30)	7.94(0.37)	9.49(0.43)

point mass distribution $(p_1/p)\delta(\beta_1)+(p_2/p)\delta(\beta_2)$. The non-null proportion was $p = p_1 + p_2$. We considered four sets of models in order to compare the empirical FNR obtained from different methods.

In Model B, we considered the empirical FNR versus mFDR level under two different non-null proportions $p = 0.2$ (Model B1) and $p = 0.02$ (Model B2), respectively. We set $p_1 = 0.8p$, $\beta_1 = -3$, $p_2 = 0.2p$ and $\beta_2 = 3$. Under such settings, the two-stage procedures of [5] have lower FNR than the Benjamini and Hochberg procedure because they are quite conservative. Other procedures have smaller FNR than the Benjamini and Hochberg procedure especially when p is large. The marginal plug-in method has the smallest FNR values.

Model C considered the FNR as a function of β_2 under $p = 0.2$ (Model C1) and $p = 0.02$ (Model C2), respectively, setting $\alpha = 0.05$, $\beta_1 = -3$, $p_1 = 0.1$ and $p_2 = 0.1$. Decrease in the empirical FNR of the marginal plug-in procedure becomes larger as the alternative distribution becomes more asymmetric. This trend is confirmed in Model D, where we set mFDR $= 0.05$, $\beta_1 = -3$, $\beta_2 = 3$ and considered FNR as a function of p_1 with $p = 0.2$ (Model D1) and $p = 0.02$ (Model D2), respectively. Compared to other procedures, the marginal plug-in procedure resulted in higher efficiency when the alternative marginal distribution was highly asymmetric. This is because all other procedures are based on the p-values, which are the probabilities under the null, ignoring the information from the alternative marginal distribution. The marginal plug-in procedure compares the probability under the marginal null and marginal alternative and therefore is more adaptive to the shape of the marginal alternative.

In Model E, we set $\beta_1 = -3$, $\beta_2 = 3$, $p_1 = p_2 = 0.5p$ and considered FNR as a function of the non-null proportion p. As p increases, the marginal plug-in procedure, the Storey's α procedure, the procedure of [2] and the q-value procedure performed more efficiently than the Benjamini and Hochberg procedure since all of them are adaptive to the proportion of the non-nulls. The marginal plug-in method still gives the smallest FNR values.

5. APPLICATION TO ANALYSIS OF CASE-CONTROL GENETIC STUDY OF NEUROBLASTOMA

We applied the proposed FDR controlling procedure to a case-control genetic study of neuroblastoma conducted at the Children's Hospital of Philadelphia. Neuroblastoma is a pediatric cancer of the developing sympathetic nervous system and is the most common solid tumors outside the central nervous system. It is a complex disease, with rare familial forms occurring due to mutations in PHOX2B or ALK [15, 16], and several common variations being enriched in sporadic neuroblastoma cases [14]. The latter genetic associations were discovered in a genome-wide association study of sporadic neuroblastoma cases, compared to children without cancer, conducted at the Children's Hospital of Philadelphia. After initial quality controls on samples and the marker genotypes, our discovery data set contained 1,627 neuroblastoma case subjects of European ancestry, each of which contained 479,804 markers. To correct the potential effects of population structure, 2,575 matching control subjects of European ancestry were selected based on their low identity-by-state estimates with case subjects.

For each marker, a score statistic was obtained by fitting a logistic regression model using an additive coding of the genotypes. Due to linkage disequilibrium among the markers, these score statistics cannot be treated as independent. However, despite existence of long-range linkage disequilibrium, the linkage disequilibrium in general decay as the distance between two markers decreases [13, 18]. We therefore can reasonably assume that the score statistics across all the markers are short-range dependent. To apply our proposed method, we used the method of [6] to estimate p and used the kernel density estimation for the marginal densities. We set different nominal mFDR values (mFDR $\in$ [0.01,0.2]) and examined the number of rejections based on the proposed marginal plug-in procedure and several p-value based procedures (see Figure 2). For a given mFDR level, the marginal plug-in procedure always identifies more significant markers than the BH and other p-value based procedures that can account for dependency of the test statistics, suggesting that it may lead to a smaller FNR and therefore better power of detecting the neuroblastoma associated markers. This is especially important for initial genome-wide scanning in order to identify the potential candidate markers for follow-up studies.

The marginal plug-in procedure identified 30 markers that are associated with neuroblastoma for mFDR $= 0.05$. In contrast, the Benjamini and Hochberg procedure identified 24 markers, the Storey's α procedure identified 24 markers, and the two stage procedures of [5] and the adaptive procedure of [2] identified 21, 21 and 25 markers, respectively. The six additional markers identified by our proposed marginal plug-in procedure, but missed by the standard Benjamini and Hochberg procedure, are presented in Table 3. The BARD1 gene provides instructions for making a protein that helps control cell growth and division. Within the nucleus of cells, the BARD1 protein interacts with the protein produced from the BRCA1 gene. Together, these two proteins mediate DNA damage response [12]. This provides some biological evidence for the association between the BARD1 gene and neuroblastoma. In fact, the recent publication of [7] identified the variants in BARD1 tumor suppressor gene influence susceptibility to high-risk neuroblastoma. Another gene, DGKI, is known to regulate Ras guanyl-releasing protein 3 and inhibits Rap1 signaling [19]. However, the association between the variant in DGKI gene and neuroblastoma is not clear and deserves further biological validation. Gene XPO4, encodes a nuclear export protein whose substrate,

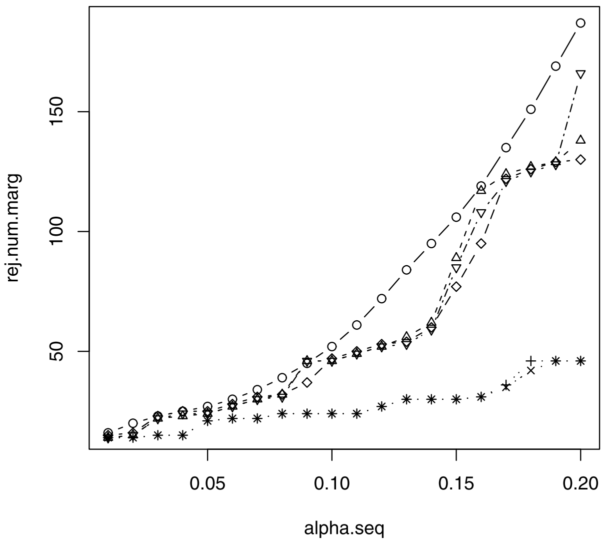

Figure 2. Comparison of number of rejections vs the mFDR *level for the marginal plug-in procedure (circle) and the p-value based procedures, including the Storey's α procedure (upward triangle), the Blanchard & Roquain procedure with Holm's step down in the first step (plus sign), the Blanchard & Roquain procedure with family-wise error control in step 1 (cross sign), the procedure of [2] (diamond) and the Benjamini and Hochberg procedure (downward triangle) for the case-control neuroblastoma genome-wide association study.*

EIF5A2, is amplified in human tumors, is required for proliferation of XPO4-deficient tumor cells, and promotes hepatocellular carcinoma in mice [27]. Another gene IKZF1 that encodes the lymphoid transcription factor IKAROS, has recently been reported to be associated with the poor outcome in acute lymphoblastic leukemia [17], and was also discovered to harbor common variations associated with susceptibility to this disease [25]. The NR5A2 gene encodes a transcription factor that has been discovered to be responsible for the reprogramming of differentiated cells into stem cells. Stem cells generated from differentiated cells are known as induced pluripotent stem cells [11].

6. DISCUSSION

We have studied in this paper the multiple testing problem under the setting where the test statistics are dependent. It is shown that for any multiple testing problem under the dependent model (3), there exists a corresponding weighted classification problem, such that the optimal decision rule for the classification problem is also an optimal solution to the multiple testing problem. Although the oracle rule obtained is optimal under the dependent model (3), it is not easy to implement. A marginal oracle rule is shown to approximate the optimal joint oracle rule when the test statistics follow a multivariate normal distribution with short range dependence. A data-driven marginal plug-in procedure is developed and is shown to be asymptotically valid and optimal in such a setting.

It should be emphasized that the marginal plug-in procedure is not necessarily the same as the adaptive compound decision rules. We showed that the marginal density of X_i can be estimated by treating the data as independent. The procedure we adopted here produces a consistent estimator for the marginal density. However, for dependent data, this might not be the optimal choice. It is an interesting future research topic to develop alternative estimation methods that can estimate the marginal densities with a faster convergence rate. In addition, in order to show the asymptotic optimality of the marginal plug-in procedure, we require Assumptions (A)–(D). Among them, Assumptions (A) and (C) are necessary and cannot be weakened. For the dependency structure that does not satisfy either Assumption (A) or (C), the marginal plug-in procedure does not have the asymptotic optimality. How to obtain a practical and asymptotically optimal FDR controlling procedure under more general dependency structure for test statistics is still a challenging open problem.

Table 3. Six single nucleotide polymorphism markers identified by the marginal plug-in procedure, but missed by the Benjamini and Hochberg procedure for the neuroblastoma data. Z score^2 is the square of the score statistic from a logistic regression with genotypes coded as 0, 1, and 2. Chrom stands for Chromosome

Marker	Gene	Chrom	Z score^2	p-value
rs4770073	XPO4	13	21.74	3.13×10^{-6}
rs7557557	BARD1	2	21.79	3.05×10^{-6}
rs1714518	RSRC1	3	21.86	2.93×10^{-6}
rs10248903	IKZF1	7	2.25	4.03×10^{-6}
rs3828112	NR5A2	1	20.59	5.67×10^{-6}
rs2059320	DGKI	7	20.42	6.21×10^{-6}

ACKNOWLEDGEMENTS

This research was partially supported by NIH grants CA127334 and CA124709.

APPENDIX A. PROOFS

We collect the proofs of the main theorems in this Appendix.

Proof of Theorem 1. Define $U_0 = \sum_{i=1}^m \mathrm{I}\,(S_i \in \mathcal{R}, \theta_i = 0)$, $U_1 = \sum_{i=1}^m \mathrm{I}\,(S_i \in \mathcal{R}, \theta_i = 1)$. Let $V_0 = E(U_0)/m = \mathrm{pr}\,(S_i \in \mathcal{R}, \theta_i = 0)$ and $V_1 = E(U_1)/m = \mathrm{pr}\,(S_i \in \mathcal{R}, \theta_i = 1)$. The mFDR and FDR can be written as

$$\mathrm{mFDR} = \frac{E(U_0)}{E(U_0 + U_1)} = \frac{V_0}{V_0 + V_1},$$

$$\text{FDR} = E\{\Psi(U_0,U_1)\} = mE\left(\frac{U_0}{U_0+U_1}\right).$$

Let $\Psi(y_1,y_2)=y_1/(y_1+y_2)$. By Taylor expansion, FDR $= E\{\Psi(U_0,U_1)\} = \text{mFDR} + (A) + (B)$, where

$$\begin{aligned}(A) &= \frac{1}{2}\frac{\partial^2\Psi}{\partial^2 y_1}(V_0,V_1)E\{(U_0-mV_0)^2\}\\ &\quad+\frac{\partial^2\Psi}{\partial y_1\partial y_2}(V_0,V_1)E(U_0-mV_0)E(U_1-mV_1)\\ &\quad+\frac{1}{2}\frac{\partial^2\Psi}{\partial^2 y_2}(V_0,V_1)E\{(U_1-mV_1)^2\},\\ (B) &= \sum_{k_1+k_2=3}\frac{1}{k_1!k_2!}E\{\zeta\cdot(U_0-V_0)^{k_1}(U_1-V_1)^{k_2}\},\end{aligned}$$

with $\zeta(U_0,U_1,V_0,V_1) = \int_0^1(1-t)^3\frac{\partial^3\Psi}{\partial y_1^{k_1}\partial y_2^{k_2}}(tU_0+(1-t)mV_0, tU_1+(1-t)mV_1)\,dt$.

We need to bound (A) and (B) respectively. For any $k_1+k_2=k$,

$$\begin{aligned}\left|\frac{\partial^k\Psi(y_1,y_2)}{\partial y_1^{k_1}\partial y_2^{k_2}}(V_0,V_1)\right| &\le Ck!\frac{\max(mV_0,mV_1)}{(mV_0+mV_1)^{k+1}}\\ &\le \frac{Ck!}{(V_0+V_1)^k m^k},\end{aligned}$$

$$\begin{aligned}&E\{\,(U_0-mV_0)^2\}\\ &\quad= mE\{(I\{S_i\in\mathcal{R},\theta_i=0\}-V_0)^2\}\\ &\qquad+\sum_{i=1}^m\sum_{j\neq i}E\{(I\{S_i\in\mathcal{R},\theta_i=0\}-V_0)\\ &\qquad\times(I\{S_j\in\mathcal{R},\theta_j=0\}-V_0)\}\\ &\quad\le mV_0(1-V_0)+4m^{1+\tau}.\end{aligned}$$

The inequality is due to the fact that S_i $(i=1,\dots,m)$ is short-ranged and the dependency structure of X is short-ranged, so that S_i and S_j are independent if $|i-j|>2m^\tau$. Similarly, we can show that $E\{(U_1-mV_1)^2\}\le Cm^{1+\tau}$. Therefore, by Cauchy–Schwarz inequality,

$$(A)\le\frac{C}{(V_0+V_1)^2m^2}\{mV_0(1-V_0)+2m^{1+\tau}\} = C_1m^{-1+\tau},$$

where C_1 depends on V_0 and V_1.

For any $k_1+k_2=k$,

$$\begin{aligned}&\left|(1-t)^k\frac{\partial^k\Psi(y_1,y_2)}{\partial y_1^{k_1}\partial y_2^{k_2}}(tU_0+(1-t)mV_0, tU_1+(1-t)mV_1)\right|\\ &\quad\le\frac{Ck!(1-t)^k\max\{tU_0+(1-t)V_0, tU_1+(1-t)V_1\}}{\{t(U_0+U_1)+(1-t)m(V_0+V_1)\}^{k+1}}\\ &\quad\le\frac{Ck!}{\{\frac{t}{1-t}(U_0+U_1)+m(V_0+V_1)\}^k}\le\frac{Ck!}{(V_0+V_1)^k m^k}.\end{aligned}$$

Thus, $|\zeta(U_0,U_1,V_0,V_1)|\le C_1/m^k$.

Let $W_{0,i}=\mathrm{I}(S_i\in\mathcal{R},\theta_i=0)-V_0$ and $W_{1,i}=\mathrm{I}(S_i\in\mathcal{R},\theta_i=1)-V_1$. We have

$$\begin{aligned}E\{(U_0-mV_0)^3\} &= mE(W_{0,i}^3)+\sum_{i=1}^m\sum_{j\neq i}E(W_{0,i}W_{0,j}^2)\\ &\quad+\sum_{i=1}^m\sum_{j\neq i}\sum_{l\neq i,j}E(W_{0,i}W_{0,j}W_{0,l})\\ &\le m+m(4m^\tau)+m(4m^\tau)(8m^\tau)\\ &\le Cm^{1+2\tau},\end{aligned}$$

and

$$\begin{aligned}&E\{(U_0-mV_0)(U_1-mV_1)^2\}\\ &\quad= m^2E(W_{0,i}W_{1,j}^2)+\sum_{i=1}^m\sum_{j=1}^m\sum_{l\neq j}E(W_{0,i}W_{1,j}W_{1,l})\\ &\quad\le m^2+m^2(4m^\tau)\le Cm^{2+\tau}.\end{aligned}$$

Similarly, $E\{(U_1-mV_1)^3\}\le Cm^{1+2\tau}$ and $E\{(U_0-mV_0)(U_1-mV_1)^2\}\le Cm^{2+\tau}$. Therefore,

$$(B)=\sum_{k_1+k_2=3}\frac{C_1}{m^3}E\{(U_0-mV_0)^{k_1}(U_1-mV_1)^{k_2}\}\le C_1m^{-1+\tau}.$$

This leads to FDR $=$ mFDR $+2C_1m^{-1+\tau}$.

Following a similar argument, we can show that FNR $=$ mFNR $+2C_1m^{-1+\tau}$. □

The proof of Theorem 2 follows that of [23] and ia omitted here.

Proof of Theorem 3. Given the FDR level α, consider a decision rule $\delta(S,R)$, with $\text{mFDR}(\delta(S,R))\le\alpha$. Suppose that the expected rejection number for $\delta(S,R)$ is r. For $\delta(\Lambda,\lambda)$ defined in (5), the expected number of rejections is

$$\begin{aligned}r(\lambda) &= E\left\{\sum_{i=1}^m\mathrm{I}(\Lambda_i<1/\lambda)\right\}\\ &= mpG^1(1/\lambda)+m(1-p)G^0(1/\lambda).\end{aligned}$$

Thus, $r(\lambda)$ monotonically decreases with λ. In addition, it is easy to see that

$$\lim_{\lambda\to0}\{r(\lambda)/m\}=1,\qquad\lim_{\lambda\to\infty}r(\lambda)=0.$$

Therefore, for a given rejection number r determined by $\delta(S,R)$, there exists a unique $\lambda(r)$ such that the rule $\delta(\Lambda,\lambda(r))$ has the same rejection number.

For $\delta(S,R)$, suppose the expected false discovery number is v_S and the true discovery number is k_S. Similarly, suppose v_Λ and k_Λ is the expected false and true discovery number for $\delta(\Lambda,\lambda)$. Then $r=v_S+k_S=v_\Lambda+k_\Lambda$. Consider the loss function

$$L_{\lambda(r)}=\frac{1}{m}\sum_i\{\lambda(r)\delta_i\,\mathrm{I}(\theta_i=0)+(1-\delta_i)\,\mathrm{I}(\theta_i=1)\},$$

then the risk for $\delta(S,R)$ and $\delta(\Lambda,\lambda)$ is $\mathcal{R}_{\lambda(r)} = p + (1/m)\{\lambda(r)v_L - k_L\}$, $L = S,\Lambda$. Since $\mathcal{R}_{\lambda(r)}(\delta(\Lambda,\lambda)) \le \mathcal{R}_{\lambda(r)}(\delta(S,R))$, it implies that $v_\Lambda \le v_S$ and $k_\Lambda \ge k_S$.

Let mFDR_L and mFNR_L be the mFDR and mFNR of $\delta(S,R)$ and $\delta(\Lambda,\lambda)$, $L = S,\Lambda$. Note that

$$\text{mFDR}_L = v_L/r \quad \text{and} \quad \text{mFNR}_L = (m_1 - k_L)/(m-r).$$

The fact that $v_\Lambda \le v_S$ and $k_\Lambda \ge k_S$ leads to $\text{mFDR}_\Lambda \le \text{mFDR}_S$ and $\text{mFNR}_\Lambda \le \text{mFNR}_S$. □

The proof of Theorem 4 requires the following Lemma.

Lemma 1. *Consider the oracle classification statistic Λ defined in (5). Let $g^s(t)$, $s = 0,1$ be the average conditional pdf's of Λ. Then $g^1(t)/g^0(t) = (1-p)/(pt)$. That is, $g^1(t)/g^0(t)$ is monotonically decreasing in t.*

The proof of Lemma 1 is similar to the proof of *Corollary* 1 in [24] and is omitted.

Proof of Theorem 4. Note that Lemma 1 implies that

$$\int_0^c g^0(t)\,\mathrm{d}t \Big/ \int_0^c g^1(t)\,\mathrm{d}t < \int_0^c g^0(t)\,\mathrm{d}t \Big/ \int_0^c \frac{g^1(c)}{g^0(c)} g^0(t)\,\mathrm{d}t = g^0(c)/g^1(c),$$

which is equivalent to $g^1(c)G^0(c) < g^0(c)G^1(c)$. Similarly, we can get $g^1(c)(1-G^0(c)) > g^0(c)(1-G^1(c))$.

Let $c = 1/\lambda$. We show that mFDR strictly increases with c.

$$\text{mFDR} = \frac{E\{\sum_{i=1}^n \mathrm{I}(T_i \le c, \theta_i = 0)\}}{E\{\sum_{i=1}^m \mathrm{I}(T_i \le c)\}} = \frac{(1-p)\sum_{i=1}^m G_i^0(c)}{\sum_{i=1}^m G_i(c)} = \frac{(1-p)G^0(c)}{G(c)}.$$

The derivative

$$\frac{\mathrm{d}}{\mathrm{d}c}\text{mFDR} = \frac{p(1-p)\{g^0(c)G^1(c) - g^1(c)G^0(c)\}}{(G(c))^2} > 0.$$

Therefore, the mFDR strictly increases with c and therefore decreases with λ.

We can show that $\text{mFNR} = p\{1-G^1(c)\}/\{1-G(c)\}$ and the derivative of mFNR is $\{g^0(c)(1-G^1(c)) - g^1(c)(1-G^0(c))\}/\{1-G(c)\}^2 < 0$. Therefore, mFNR strictly decreases with c and increases with λ. □

Proof of Theorem 6. First, we have

$$\begin{aligned}
&|\,T_{\text{MG},i} - T_{\text{OR},i}\,| \\
&= \left|\frac{(1-p)f(x_i|\theta_i=0)}{f(x_i)}\right|\left|\frac{g(x_{-i}|x_i,\theta_i=0)}{g(x_{-i}|x_i)} - 1\right| \\
&\le \left|\frac{\{g(x_{-i}|x_i,\theta_i=0) - g(x_{-i}|x_i,\theta_i=1)\}\,\mathrm{pr}\,(\theta_i=1|x_i)}{g(x_{-i}|x_i)}\right| \\
&= \left|\frac{pf(x_i|\theta_i=1)\{g(x_{-i}|x_i,\theta_i=0) - g(x_{-i}|x_i,\theta_i=1)\}}{g(x)}\right| \\
&\le \left|\frac{2p|\Omega_{ii}^{(m)}|^{1/2}}{\sum_\theta \exp\{-\frac{1}{2}(x-\mu(\theta))^{\mathrm{T}}\Sigma^{-1}(x-\mu(\theta))\}\,\mathrm{pr}\,(\theta)}\right|.
\end{aligned}$$

For m sufficiently large, $\lambda_{\min}(\Sigma^{(m)}) > \kappa/2$ and $|\,\Omega_{ii}^{(m)}\,| \le \lambda_{\max}(\Omega^{(m)}) < 2/\kappa$. For all $\eta > 0$, there exists an l_0, such that $\mathrm{pr}\,(|\sum_\theta \exp\{-\frac{1}{2}(x-\mu(\theta))^{\mathrm{T}}\Sigma^{-1}(x-\mu(\theta))\}\,\mathrm{pr}\,(\theta)| < l_0)) < \eta$. For all $\varepsilon > 0$, there exists an m sufficiently large, such that $2\sqrt{2}p/(\sqrt{\kappa}\varepsilon) < l_0$. Then for all $i = 1,\ldots,m$,

$$\begin{aligned}
&\mathrm{pr}\,(|\,T_{\text{MG},i} - T_{\text{OR},i}\,| > \epsilon) \\
&= \mathrm{pr}\left(\left|\sum_\theta \exp[-\{x-\mu(\theta)\}^{\mathrm{T}}\Sigma^{-1}\{x-\mu(\theta)\}/2] \right.\right. \\
&\qquad \left.\left. \times\ \mathrm{pr}\,(\theta)\right| < l_0\right) < \eta.
\end{aligned}$$

□

The proof of Theorem 7 requires the following two lemmas.

Lemma 2. *Assume Assumptions* (A), (B) *and* (C) *hold. Let $\hat{p}$, $\hat{f}$, and $\hat{f}_0$ be estimates such that $\hat{p}$ converges to p in probability, and for all x, $E\{\hat{f}(x) - f(x)\}^2 \to 0$ and $E\{\hat{f}_0(x) - f_0(x)\}^2 \to 0$, then $\hat{T}_{\text{MG},i}$ converges to $T_{\text{OR},i}$ in probability.*

The proof of Lemma 2 is similar to the proof of Lemmas A.1 and A.2 in [23] and is omitted.

Lemma 3. *Let $\hat{R}_\lambda = \frac{1}{m}\sum_{i=1}^m \mathrm{I}\,(\hat{T}_{\text{MG},i} \le \lambda)$ and $\hat{V}_t = \frac{1}{m}\sum_{i=1}^m \mathrm{I}\,(\hat{T}_{\text{MG},i} \le t)\hat{T}_{\text{MG},i}$. Define $\hat{Q}_\lambda = \hat{V}_\lambda/\hat{R}_\lambda$. Then for $\alpha < t < 1$, $\hat{Q}_\lambda$ converges to Q_λ in probability.*

Proof of Lemma 3. Let $\nu_i = E\{\mathrm{I}\,(T_{\text{OR},i} < \lambda)\}$ and $Z_i = \mathrm{I}\,(T_{\text{OR},i} < \lambda) - \nu_i$. On $\{x : |T_{\text{OR},i} - T_{\text{MG},i}| \le \varepsilon\}$, $\mathrm{I}\,(T_{\text{OR},i} < \lambda) = \mathrm{I}\,(T_{\text{MG},i} < \lambda)$ holds unless $\lambda - \varepsilon \le T_{\text{MG},i} \le \lambda + \varepsilon$. Therefore, for any $\varepsilon > 0$,

$$\begin{aligned}
&|E\{I(T_{\text{OR},i} < \lambda)\} - E\{\mathrm{I}\,(T_{\text{MG},i} < \lambda)\}| \\
&\quad \le \mathrm{pr}\,(|T_{\text{OR},i} - T_{\text{MG},i}| > \varepsilon) + \mathrm{pr}\,(\lambda - \varepsilon < T_{\text{MG}} < \lambda + \varepsilon).
\end{aligned}$$

The first term converges to zero uniformly by Theorem 6, and the second term converges to zero uniformly by the continuity of T_{MG}. Consequently, $\nu_i \to \mathrm{pr}\,(T_{\text{MG}} < \lambda)$ uniformly. The short range dependency structure of X leads to the short range dependency structure of Z, and consequently

$$\begin{aligned}
\mathrm{var}(\bar{Z}) &= \mathrm{var}(Z_i)/m + \sum_{i=1}^m \sum_{j\ne i} \mathrm{cov}(Z_i,Z_j)/(m^2) \\
&\le 1/m + m^{-1+\tau} \to 0.
\end{aligned}$$

By weak law of large numbers of triangle arrays, we have $\bar{Z}$ converges to 0 in probability. Thus,

$$|R_\lambda - \mathrm{pr}\,(T_{\text{MG}} < \lambda)| \le \frac{1}{m}\sum_{i=1}^m |\nu_i - \mathrm{pr}\,(T_{\text{MG}} < \lambda)| + |\bar{Z}| \to 0.$$

Similarly, we write $\hat{R}_\lambda$ as $\hat{R}_\lambda = \sum_{i=1}^m \{\mathrm{I}(\hat{T}_{\mathrm{MG},i} < \lambda) - \mathrm{pr}(\hat{T}_{\mathrm{MG}} < \lambda)\} + \mathrm{pr}(\hat{T}_{\mathrm{MG}} < \lambda)$. The first part goes to 0 by the weak law of large numbers for triangle arrays, and the second part goes to $\mathrm{pr}(T_{\mathrm{MG}} < \lambda)$. Then we have $\hat{R}_\lambda$ converges to R_λ in probability.

We next prove $\hat{V}_\lambda$ converges to V_λ in probability. For any $\varepsilon > 0$, if m is sufficiently large, for all i, $E(T_{\mathrm{OR},i} - T_{\mathrm{MG},i}) \leq \varepsilon + \mathrm{pr}(|T_{\mathrm{OR},i} - T_{\mathrm{MG},i}| > \varepsilon) < 2\varepsilon$. Therefore, $E(T_{\mathrm{OR},i}) \to E(T_{\mathrm{MG}})$ uniformly. We then have

$$\begin{aligned} &E\{T_{\mathrm{OR},i}\,\mathrm{I}(T_{\mathrm{OR},i} < \lambda)\} - E\{T_{\mathrm{MG}}\,\mathrm{I}(T_{\mathrm{MG}} < \lambda)\} \\ &\quad \leq \mathrm{pr}(|T_{\mathrm{OR},i} - T_{\mathrm{MG}}| > \varepsilon) \\ &\qquad + E(T_{\mathrm{OR},i} - T_{\mathrm{MG}}) + \mathrm{pr}(\lambda - \varepsilon \leq T_{\mathrm{MG}} \leq \lambda + \varepsilon). \end{aligned}$$

All three parts go to zero uniformly as $m \to \infty$. Similar to the convergence of R_λ shown above, we can obtain that V_λ converges to $E\{T_{\mathrm{MG}}\,\mathrm{I}(T_{\mathrm{MG}} < \lambda)\}$ in probability. We can show similarly that $\hat{V}_\lambda$ converges to $E\{T_{\mathrm{MG}}\,\mathrm{I}(T_{\mathrm{MG}} < \lambda)\}$ in probability. Then $\hat{V}_\lambda$ converges to V_λ in probability. Consequently, we conclude that $\hat{Q}_\lambda = \hat{V}_\lambda/\hat{R}_\lambda$ converges to $V_\lambda/R_\lambda = Q_\lambda$ in probability. □

Proof of Theorem 7. Define threshold $\lambda = \sup\{t \in (0,1) : Q(t) \leq \alpha\}$ and the plug-in threshold $\hat{\lambda} = \sup\{t \in (0,1) : \hat{Q}(t) \leq \alpha\}$. Since $\hat{Q}_\lambda$ converges to Q_λ in probability, by Lemma A.5 in [23], we have $\hat{\lambda}$ converges to λ in probability. The plug-in procedure is equivalent to rejecting H_i^0 when $\hat{T}_{\mathrm{MG},i} \leq \hat{\lambda}$. In the proof for Lemma 3, we have $\sum_{i=1}^m \mathrm{pr}(T_{\mathrm{OR},i} < \lambda)/m \to \mathrm{pr}(T_{\mathrm{MG}} < \lambda)$ and $\frac{1}{m}\sum_{i=1}^m \mathrm{pr}(T_{\mathrm{OR},i} < \lambda \mid \mathrm{H}_i^0)$ converges to $\mathrm{pr}(T_{\mathrm{MG}} < \lambda \mid \mathrm{H}_i^0)$ in probability. Following the same arguments, $\hat{T}_{\mathrm{MG},i} - \hat{\lambda} \to T_{\mathrm{MG},i} - \lambda$ uniformly, and thus

$$\frac{1}{m}\sum_{i=1}^m \mathrm{pr}(\hat{T}_{\mathrm{MG},i} < \lambda) \to \mathrm{pr}(T_{\mathrm{MG}} < \lambda),$$

and

$$\frac{1}{m}\sum_{i=1}^m \mathrm{pr}(\hat{T}_{\mathrm{MG},i} < \lambda \mid \mathrm{H}_i^0) \to \mathrm{pr}(T_{\mathrm{MG}} < \lambda \mid \mathrm{H}_i^0).$$

It follows that

$$\frac{1}{m}\sum_{i=1}^m \mathrm{pr}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda} \mid \mathrm{H}_i^0) - \frac{1}{m}\sum_{i=1}^m \mathrm{pr}(T_{\mathrm{OR},i} < \lambda \mid \mathrm{H}_i^0) \to 0,$$

$$\frac{1}{m}\sum_{i=1}^m \mathrm{pr}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda}) - \frac{1}{m}\sum_{i=1}^m \mathrm{pr}(T_{\mathrm{OR},i} < \lambda) \to 0.$$

This leads to

$$\frac{E(\hat{V}_{\hat{\lambda}})}{E\hat{R}_{\hat{\lambda}}} = \frac{E(V_\lambda)}{E(R_\lambda)} = \mathrm{mFDR}_{\mathrm{OR}}.$$

Following a similar argument, we can show that

$$E(\hat{V}_{\hat{\lambda}})/E(\hat{R}_{\hat{\lambda}}) \to \mathrm{mFNR}_{\mathrm{OR}}.$$ □

Received 11 August 2011

REFERENCES

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistics Society, Series B* **57** 289–300. MR1325392

[2] Benjamini, Y., Krieger, M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438

[3] Benjamini, Y. and Yekutieli, D. (2001). The control of false discovery rate in multiple testing under dependence. *Annals of Statistics* **76** 1165–1188. MR1869245

[4] Blanchard, F. and Fleuret, F. (2006). Occam's hammer: A link between randomized learning and multiple testing FDR control Technical Report, Fraunhofer FIRST.IDA, Berlin, Germany and EPFL, CVLAB, Lausanne, Switzerland.

[5] Blanchard, G. and Roquain, E. (2009). Adaptive FDR control under independence and dependence. *Journal of Machine Learning Research* **10** 2837–2871. MR2579914

[6] Cai, T. T. and Jin, J. (2010). Optimal rates of convergence for estimating the null and proportion of non-null effects in large-scale multiple testing. *The Annals of Statistics* **38** 100–145. MR2589318

[7] Capasso, M., Hou, C., Asgharzadeh, S., Attiyeh, E. F., Mosse, Y. P., Diskin, S. J., Cole, K., Bosse, K., Diamond, M., Laudenslager, M., Winter, C., Bradfield, J. P., Scott, R. H., Jagannathan, J., Glessner, J. T., Kim, C., London, W. B., Seeger, R. C., Li, H., Rahman, N., Rappaport, E., Hakonarson, H., Devoto, M. and Maris, J. (2009). Common variations in the BARD1 tumor suppressor gene influence susceptibility to high-risk neuroblastoma. *Nature Genetics* **41** 718–723.

[8] Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of American Statistical Association* **102** 93–103. MR2293302

[9] Efron, B. (2010). Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association* **105** 1042–1055. MR2752597

[10] Genovese, C. and Wasserman, L. (2002). Operating characteristic and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society, Series B* **64** 499–517. MR1924303

[11] Heng, J. D., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J., Orlov, Y., Huss, M., Yang, L., Lufkin, T., Lim, B. and Ng, H. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. *Cell Stem Cell* **6** 167–174.

[12] Irminger-Finger, I. and Jefford, C. E. (2006). Is there more to BARD1 than BRCA1? *Nature Review Cancer* **6** 382–391.

[13] Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22** 139–144.

[14] Maris, J. M., Yael, P. M., Bradfield, J. P., Hou, C., Monni, S., Scott, R. H., Asgharzadeh, S., Attiveh, E. F., Diskin, S. J., Laudenslager, M., Winter, C., Cole, K., Glessner, J. T., Kim, C., Frackelton, E. C., Casalunovo, T., Eckert, A. W., Capasso, M., Rappaport, E. F., McConville, C., London, W. B., Seeger, R. C., Rahman, N., Devoto, M., Grant, S. F. A., Li, H. and Hakonarson, H. (2008). A genome-wide association study identifies a susceptibility locus to clinically aggressive neuroblastoma at 6p22. *New England Journal of Medicine* **358** 2585–2593.

[15] Mosse, Y. P., Greshock, J., Margolin, A., Naylor, T., Cole, K., Khazi, D., Hii, G., Winter, C., Shahzad, S., Asziz, M. U., Biegel, J. A., Weber, B. L. and Maris, J. M. (2005). High-resolution detection and mapping of genomic DNA alterations in neuroblastoma. *Genes, Chromosomes, Cancer* **43** 390–403.

[16] Mosse, Y. P., Laudenslager, M., Longo, L., Cole, K. A., Wood, A., Attiyeh, E. F., Laquaglia, M. J., Sennett, R., Lynch, J. E., Perri, P., Laureys, G., Speleman, F., Kim, C., Hou, C., Hakonarson, H., Torkamani, A., Schork, N. J., Brodeur, G. M., Tonini, G. P., Rappaport, E., Devoto, M. and Maris, J. M. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455** 930-935.

[17] Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B., Ma, J., Liu, W., Cheng, C., Schulman, B. A., Harvey, R. C., Chen, I. M., Clifford, R. J., Carroll, W. L., Reaman, G., Bowman, W. P., Devidas, M., Gerhard, D. S., Yang, W., Relling, M. V., Shurtleff, S. A., Campana, D., Borowitz, M. J., Pui, C. H., Smith, M., Hunger, S. P., Willman, C. L., Downing, J. R. and Children's Oncology Group (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *New England Journal of Medicine* **360** 470–480.

[18] Pritchard, J. K. and Przeworski, M. (2001). Linkage Disequilibrium in humans: Models and data. *American Journal of Human Genetics* **69** 1–14.

[19] Regier, D. S., Higbee, J., Lund, K. M., Sakane, F., Prescott, S. M. and Topham, M. K. (2005). Diacylglycerol kinase i regulates ras guanyl-releasing protein 3 and inhibits Rap1 signaling. *Proceedings of the National Academy of Sciences* **102** 7595–7600.

[20] Sarkar, S. (2008). Generalizing Simes' test and Hochberg's stepup procedure. *The Annals of Statistics* **36** 337–363. MR2387974

[21] Storey, J., Taylor, J. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B* **66** 187–205. MR2035766

[22] Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences* **100** 9440–9445. MR1994856

[23] Sun, W. and Cai, T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association* **102** 901–912. MR2411657

[24] Sun, W. and Cai, T. (2009). Large-scale multiple testing under dependency. *Journal of the Royal Statistical Society, Series B* **71** 393–424. MR2649603

[25] Trevino, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M., Willman, C., Neale, G., Downing, J., Raimondi, S. C., Pui, C. H., Evans, W. E. and Relling, M. V. (2009). Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nature Genetics* **41** 1001–1005.

[26] Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, London. MR1652247

[27] Zender, L., Xue, W., Zuber, J., Semighini, C. P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J. M., Schirmacher, P., Combine, W. R., Wigler, M., Hicks, J., Hannon, G. J., Powers, S. and Lowe, S. W. (2008). An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135** 852–864.

Jichun Xie
Department of Statistics
The Fox School of Business and Management
Temple University
USA
E-mail address: jichun@temple.edu

T. Tony Cai
Department of Statistics
The Wharton School
University of Pennsylvania
USA
E-mail address: tcai@wharton.upenn.edu

John Maris
Department of Pediatrics
School of Medicine
University of Pennsylvania
USA
E-mail address: maris@chop.edu

Hongzhe Li
Department of Biostatistics and Epidemiology
School of Medicine
University of Pennsylvania
USA
E-mail address: hongzhe@upenn.edu