Submitted to the Annals of Applied Statistics
2010 1
arXiv: math.PR/0000000

FALSE DISCOVERY RATE CONTROL FOR HIGH DIMENSIONAL DEPENDENT DATA WITH AN APPLICATION TO LARGE-SCALE GENETIC ASSOCIATION STUDIES

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

University of Pennsylvania and Children's Hospital of Philadelphia

Large-scale genetic association studies are increasingly utilized for identifying novel susceptible genetic variants for complex traits, but there is little consensus on analysis methods for such data. Most commonly used methods include single SNP analysis or haplotype analysis with Bonferroni correction for multiple comparisons. Since the SNPs in typical GWAS are often in linkage disequilibrium (LD), at least locally, Bonferonni correction of multiple comparisons often leads to conservative error control and therefore lower statistical power. Motivated by an application for analysis of data from the genetic association studies, we consider the problem of false discovery rate (FDR) control under the high dimensional multivariate normal model. Using the compound decision rule framework, we develop an optimal joint oracle procedure and propose to use a marginal procedure to approximate the optimal joint optimal procedure. We show that the marginal plug-in procedure is asymptotically optimal under mild conditions. Our results indicate that the multiple testing procedure developed under the independent model is not only valid but also asymptotically optimal for the high dimensional multivariate normal data under some weak dependency. We evaluate various procedures using simulation studies and demonstrate its application to a genome-wide association study of neuroblastoma (NB). The proposed procedure identified a few more genetic variants that are potentially associated with NB than the standard p-value-based FDR controlling procedure.

1. Introduction. Large-scale genetic association studies such as the genome-wide association studies (GWAS) are designed to scan the entire genome for the identification of genetic variations associated with phenotypic traits, such as a disease condition, blood pressure, or body mass index. These studies usually examine several hundred thousand single nucleotide polymorphisms (SNPs) that explain most of the genetic variation across the

*Supported by NIH grants ES009911, CA127334 and CA124709.
†Supported by NSF FRG Grant DMS-0854973
Keywords and phrases: Neuroblastoma, large scale multiple testing, weighted classification, optimal oracle rule, marginal rule, genome-wide association

genome as well as SNPs in a large array of candidate gene regions. Most GWAS involve some form of multistage design, which includes an initial scan of hundreds of thousands of SNPs on a sample of cases and controls, followed by testing a subset of the most promising markers on independent samples. Additional markers may also be included at later stages to better characterize the full spectrum of genetic variation in the targeted regions. GWAS have been demonstrated to be a powerful approach for the detection of genetic variants related to complex traits, such as age-related macular degenerative diseases (Klein *et al.*, 2005), prostate and breast cancers (Hunter *et al.*, 2007), and type 2 diabetes (Scott *et al.*, 2007). The Welcome Trust Case-Control Consortium has recently published a GWAS of seven diseases using $14,000$ cases and 3000 shared controls (WTCCC, 2007). The success of these studies has provided solid evidence that GWAS represent a powerful approach to the identification of genes involved in common human diseases.

To date, the analytical methods of GWAS have largely been limited to the single SNP or SNP-SNP pair analysis, coupled with statistical techniques such as the Bonferroni procedure for controlling multiple comparisons. However, since SNPs in typical GWAS are in linkage disequilibrium (LD) locally, simple Bonferonni correction can potentially be conservative and therefore lead to a loss of statistical power. In addition, if multiple SNPs are all in LD with the true disease variants, effectively utilizing the information from multiple SNPs in LD can potentially increase the power of detecting the SNPs associated with the disease. Although Bonferroni correction has been widely applied in GWAS, analytical and simulation studies by Sabatti *et al.* (2003) have shown that the false discovery rate (FDR) procedure of Benjamini and Hochberg (1995) can effectively control the FDR for the dependent tests encountered in case-control association studies and increase power over more traditional methods. However, the direct application of the FDR procedure can lead to loss of power due to the dependency among tests, although the FDR can still be controlled (Benjamini and Yekutieli, 2001). The most effective way of correcting this problem relies on developing a precise model for the dependency among the SNPs and incorporating it in the definition of a FDR controlling procedure (Sabatti *et al.*, 2003). For the hidden Markov model (HMM), Sun and Cai (2009) proved the optimal power of a posterior probability-based FDR procedure while controlling the FDR. However, it is not realistic to apply simple homogeneous HMM for modeling the SNP dependency. Li, Wei and Maris (2010) proposed to apply a hidden Markov random field model for GWAS analysis, but provided no theoretical guarantee on the FDR control.

In this paper, we consider the problem of FDR control for high dimen-

sional dependent data as encountered in GWAS. In such large scale multiple testing problems, we usually have thousands or even millions of hypotheses,

$$\mathrm{H}_i^0 : \theta_i = 0,\ i = 1, \ldots, m.$$

To test these hypotheses, we have a sequence of observations $\boldsymbol{x} = (x_1, \ldots, x_m)$ (usually test statistics), which comes from the random variable $\boldsymbol{X}$,

$$\boldsymbol{X} \,|\, \boldsymbol{\theta} \sim f(\boldsymbol{x} \,|\, \boldsymbol{\theta}).$$

The dependency structure of the test statistics can be specified in many ways. We consider in particular the multivariate normal distribution for $\boldsymbol{X}$ because of its wide applications. We assume that the θ_i's are independent, but the observation $\boldsymbol{x}$ given $\boldsymbol{\theta}$ have correlated measurement errors. We assume

$$\boldsymbol{X} \,|\, \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \Sigma), \tag{1}$$

where $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the mean vector and Σ is the covariance matrix. We show in Section 5 that multiple testing problem in GWAS can be formulated as this model.

The goal is to test these hypotheses simultaneously and to separate the non-null from the null based on the observations. The outcomes of multiple testing problem can be summarized as in Table 1. The false discovery rate is defined as $\mathrm{mFDR} = \mathbb{E}\,(N_{10}/R \,|\, R > 0)\,\mathbb{P}\,(R > 0)$ and the false non-discovery rate (mFNR) is defined as $\mathrm{FNR} = \mathbb{E}\,(N_{01}/S \,|\, S > 0)\,\mathbb{P}\,(S > 0)$. mFDR and mFNR are two measures similar to FDR and FNR, which are defined as $\mathbb{E}\,N_{10}/\,\mathbb{E}\,R$ and $\mathbb{E}\,N_{01}/\,\mathbb{E}\,S$, respectively. Genovese and Wasserman (2002) proved that $\mathrm{mFDR} = \mathrm{FDR} + O(m^{-1/2})$ under some weak conditions.

TABLE 1
Classification of tested hypotheses

Hypothesis	Claimed non-significant	Claimed significant	Total
Null	N_{00}	N_{10}	m_0
Non-null	N_{01}	N_{11}	m_1
Total	S	R	m

For such dependent cases, most of the papers discussing the FDR control procedure focus more on the validity of the procedure. Benjamini and Yekutili (2001), Farcomeni (2007) and Wu (2008) show that one can control FDR under the nominal level by the p value-based procedure under some dependence assumptions, arguing the p-value based procedure can be adaptive to the dependency structure. However, the efficiency issue has not been

discussed in detail in these papers. Sun & Cai (2009) considered the case that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ is a homogenous irreducible hidden Markov chain, and given θ_i, $X_i \,|\, \theta_i$ are *i.i.d.* They obtained an asymptotically optimal rule under their model, but the assumption they make for the dependency structure is strong and does not hold for data from GWAS.

In this article, we focus on the asymptotically optimal procedure under the dependency structure specified by model (1). Our theoretical development follows that of Sun & Cai (2007) by showing that the large-scale multiple testing problem has a corresponding equivalent weighted classification problem in the sense that the optimal solution to the multiple testing problem is also the optimal decision rule for the weighted classification problem. However, our development does not need the monotone likelihood ratio (MLR) condition for the equivalence between the multiple testing and weighted classification. Based on the compound decision rule, we obtain the optimal oracle rule under the assumed dependent mode and further show that one can use the marginal oracle rule to approximate the optimal joint oracle rule in order to reduce the computational complexity. It is shown that the marginal oracle statistics are uniformly consistent approximation to the joint oracle statistics. We also develop a procedure based on a marginal plug-in and obtain the asymptotic theory for the marginal plug-in estimator which, procedure-wise, is the same as the adaptive compound decision rules.

The paper is organized as follows. We first present the oracle rule for multivariate model specified by (1) and then present the marginal approximation to the joint oracle procedure and show its asymptotic optimality. We provide simulation studies and real data analysis in Section 4 and 5 to demonstrate the validity and efficiency of the marginal plug-in procedure. A brief discussion is given in Section 6. The proofs of all the theorems are presented in the Appendix.

2. Oracle Decision Rule for Multivariate Distribution. Suppose $\theta_1, \ldots, \theta_m \overset{i.i.d}{\sim} \text{Bernoulli}(p)$ and $\boldsymbol{X} = (X_1, \ldots, X_m)$ be a multivariate random variable with the distribution

$$\boldsymbol{X} \sim f(\boldsymbol{x}|\,\boldsymbol{\theta}), i = 1, \cdots, m. \tag{2}$$

There are m hypotheses:

$$\mathrm{H}_i^0 : \theta_i = 0.$$

Based on one observation of $\boldsymbol{X}$: $\boldsymbol{x} = (x_1, \ldots, x_m)$, we need to find a procedure $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ which can control FDR at level α and achieve small FNR. Usually, $\boldsymbol{X}$ can be a sequence of statistics, like T-statistics or score

statistics. Due to dependency among the $\boldsymbol{X}$, the distribution of X_i now depends not only on θ_i but the whole sequence of $\boldsymbol{\theta}$.

It has been shown that the multiple testing problem has a close relationship with the weighted classification problem under independent case (Sun & Cai, 2007) and under the dependency specified by the hidden Markov model (Sun & Cai, 2009). We first consider the related weighted classification problem under the model (2) and have the following theorem.

THEOREM 1. *Define the loss function:*

$$L_\lambda(\boldsymbol{\delta}, \boldsymbol{\theta}) = \frac{1}{m} \sum_i \left[\lambda \delta_i \, \mathrm{I}\left(\theta_i = 0\right) + (1 - \delta_i) \, \mathrm{I}\left(\theta_i = 1\right)\right]. \tag{3}$$

Consider the model defined in (2). Suppose that p and f are known. Then the classification risk $\mathbb{E}\left[L_\lambda(\boldsymbol{\theta}, \boldsymbol{\delta})\right]$ is minimized by the Bayes rule $\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda) = (\delta_1, \ldots, \delta_m)$, where

$$\delta_i = \mathrm{I}\left\{\Lambda_i(\boldsymbol{x}) = \frac{(1-p) f(\boldsymbol{x} |\, \theta_i = 0)}{p f(\boldsymbol{x} |\, \theta_i = 1)} \leq \frac{1}{\lambda}\right\}, \ i = 1, \ldots, m. \tag{4}$$

The minimum classification risk is

$$\mathcal{R}_\lambda(\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda)) = p + \int_K [\lambda(1-p) f(\boldsymbol{x} |\, \theta_i = 0) - p f(\boldsymbol{x} |\, \theta_i = 1)] \, \mathrm{d}\boldsymbol{x},$$

where $K = \{\boldsymbol{x} \in \Omega : \lambda(1-p) f(\boldsymbol{x} |\, \theta_i = 0) \leq p f(\boldsymbol{x} |\, \theta_i = 1)\}$.

The proof of this theorem follows that of Sun and Cai (2007) and is omitted here. The rule given in Theorem 1 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 $\boldsymbol{\delta}(\Lambda, \lambda)$ as defined in (4). Let $G_i^s(t) = \mathbb{P}\left(\Lambda_i \leq t \mid \theta_i = s\right)$, $s = 0, 1$, be the conditional cdf's of $\Lambda_i(\boldsymbol{x})$. The marginal cdf of $\Lambda_i(\boldsymbol{x})$ is then given by $G_i(t) = \mathbb{P}\left(\Lambda_i \leq t\right) = (1-p) G_i^0(t) + p G_i^1(t)$. Define the average conditional cdf's of $\boldsymbol{\Lambda}$, $G^s(t) = (1/m) \sum_{i=1}^m G_i^s(t)$ and average conditional pdf's of $\boldsymbol{\Lambda}$, $g^s(t) = (\,\mathrm{d}/\,\mathrm{d}t) G^s(t)$, $s = 0, 1$. We have the following theorem indicating that Λ is also optimal in the multiple testing problem.

THEOREM 2. *Let $\mathcal{D}_s = \{\boldsymbol{\delta} : \ \delta_i = \mathrm{I}\{\Lambda_i < \lambda, i = 1, \cdots, m\}$, where Λ_i's are defined in (4). Given a* mFDR *level α and a decision rule $\boldsymbol{\delta}(\boldsymbol{S}, R) = \left[\mathrm{I}\left(S_i(\boldsymbol{x}) \in R_i\right)\right]_{i=1,\ldots,m}$. Suppose* $\mathrm{mFDR}(\boldsymbol{\delta}(\boldsymbol{S}, R)) \leq \alpha$*, there exists a λ determined by $\boldsymbol{\delta}(\boldsymbol{S}, R)$, such that $\boldsymbol{\delta}(\Lambda, \lambda) \in \mathcal{D}_s$ outperforms $\boldsymbol{\delta}(\boldsymbol{S}, R)$ in the sense that*

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

and

$$\text{mFNR}(\boldsymbol{\delta}(\Lambda, \lambda)) \geq \text{mFNR}(\boldsymbol{\delta}(\boldsymbol{S}, R)).$$

Theorem 2 reveals that the optimal solution for the multiple testing problem comes from the set $\mathcal{D}_s$. Instead of searching for all the decision rules, we only need to search $\mathcal{D}_s$ for the optimal rule. We show next that for a given $mFDR$ value α, the optimal rule for the multiple testing problem is unique.

THEOREM 3. *Consider the optimal decision rule $\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda)$ in (4) for the weighted classification problem with the loss function (3). Then there exists a unique $\lambda(\alpha)$, such that $\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda(\alpha))$ controls $mFDR$ at level α and minimize* mFNR *among all the decision rules.*

Theorem 3 shows that there exists a one-to-one mapping between the mFDR value α and the thresholding λ, which determines the optimal rule. However, usually it is not easy to get the corresponding $\lambda(\alpha)$ for a given α. In this case, we need to develop an oracle rule that depends on α directly instead of $\lambda(\alpha)$. Define

$$T_{\text{OR},i} = \frac{(1-p)f(\boldsymbol{x}|\,\theta_i = 0)}{f(\boldsymbol{x})}. \tag{5}$$

Obviously, $T_{\text{OR},i} = \Lambda_i/(1+\Lambda_i)$ increases with Λ_i. Thus, for a given mFDR value α, we can rewrite the optimal oracle rule (4) as

$$\delta_{OR,i} = \text{I}\left(T_{\text{OR},i} < \tilde{\lambda}(\alpha) = \frac{\lambda(\alpha)}{1+\lambda(\alpha)}\right). \tag{6}$$

Let $R_{\tilde{\lambda}} = \frac{1}{m}\sum_{i=1}^{m} \text{I}\,(T_{\text{OR},i} < \tilde{\lambda})$, $V_{\tilde{\lambda}} = \frac{1}{m}\sum_{i=1}^{m} \text{I}\,(T_{\text{OR},i} < \tilde{\lambda}, \theta_i = 0)$ and $Q_{\tilde{\lambda}} = V_{\tilde{\lambda}}/R_{\tilde{\lambda}}$ be the number of rejections, number of false rejections and false discovery proportion at the realization point $\boldsymbol{x}$ for the oracle rule (6). Note that

$$\begin{aligned}\text{FDR} = \mathbb{E}\,[Q_{\tilde{\lambda}}] &= \mathbb{E}\,[\,\mathbb{E}\,(V_{\tilde{\lambda}}/R_{\tilde{\lambda}})|\,\boldsymbol{x}] \\ &= \mathbb{E}\left[\frac{\sum_{i=1}^{m}\text{I}\,(T_{OR,i} < \tilde{\lambda})T_{OR,i}}{\sum_{i=1}^{m}\text{I}\,(T_{OR,i} < \tilde{\lambda})}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{mR_{\tilde{\lambda}}} T_{OR,(i)}}{mR_{\tilde{\lambda}}}\right],\end{aligned}$$

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

$$\frac{1}{k}\sum_{i=1}^{k} T_{\text{OR},(i)} \leq \alpha. \tag{7}$$

If for every k, (7) is satisfied, then $\text{FDR} = \mathbb{E}\left[\sum_{i=1}^{R} T_{OR,(i)}/R\right] \leq \alpha$.

Based on the argument presented above, we have the following oracle procedure for multivariate data.

THEOREM 4. *Consider the model defined in (2). Suppose that p and f are known. Define $T_{OR,i}$ as in (5). Then the following procedure controls FDR at level α:*

(8)

$$\textit{Let } k = \max\left\{ l : \frac{1}{l}\sum_{i=1}^{l} T_{\text{OR},(i)}(\boldsymbol{x}) \leq \alpha \right\}, \textit{ then reject all } \text{H}_{(i)},\ i = 1, \ldots, k.$$

The final oracle rule (8) consists of two steps: the first step is to calculate the oracle statistics T_{OR_i}, and the second step is to rank the statistics and calculate the running means from the smallest to the largest in order to determine the cutoff. We then reject all the hypotheses that corresponding to the $T_{\text{OR},(i)}$ which are below the cutoff.

3. Marginal Approximation to the Joint Oracle Procedure. Theorem 2, Theorem 3, together with Theorem 4 show nice properties of oracle rule. To sum up, the oracle rule is a good choice for the multiple testing problem since it controls mFDR at level α and minimize mFNR asymptotically at the same time. However, it is often hard to get the oracle rules or equivalently the oracle statistics because of two reasons. First, while deriving the oracle statistics, we assume that the non-null proportion p, and the distributions $f(\boldsymbol{x}|\,\boldsymbol{\theta})$ are all known, which is usually not true in practice. This consequently leads to estimation difficulty to the problem. Second, even if we know $f(\boldsymbol{x}|\,\boldsymbol{\theta})$, it is still computationally hard to calculate $f(\boldsymbol{x}|\,\theta_i = 0)$ and the mixed distribution $f(\boldsymbol{x})$. Note that $f(x) = \sum_{\theta_1,\ldots,\theta_m} f(\boldsymbol{x}\,|\,\boldsymbol{\theta})$. The computational complexity to get the whole sequence of $\boldsymbol{T}_{\text{OR}}$ is $O(m \cdot 2^m)$. Therefore, we need some technique to make the problem computationally feasible.

We show in Section 3.1 that under some additional assumptions on the covariance structure of model (2), we can use the marginal oracle statistics to approximate the joint oracle statistic, which also lead to an optimal FDR controlling procedure. In section 3.2, we discuss how to estimate the marginal oracle statistics, and show the asymptotic property of the plug-in rule.

3.1. *Marginal Oracle Statistics.* We made the following additional assumptions on the model (2):

(A). The non-null proportion $p^{(m)} \to 0$, as $m \to \infty$.

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

$$\boldsymbol{X}^{(m)} | \, \boldsymbol{\theta}^{(m)} \sim \mathrm{N}(\boldsymbol{\mu}(\boldsymbol{\theta}^{(m)}), \Sigma^{(m)}), \tag{9}$$

where $\boldsymbol{\mu}(\boldsymbol{\theta}^{(m)}) = (\mu(\theta_1), \ldots, \mu(\theta_m))$, $\mu(\theta_i = 0) = \beta_0$ and $\mu(\theta_i = 1)$ follows the discrete distribution taking μ^d with probability p_d, $d = 1, \ldots, D$. Note that $p = \sum_{d=1}^{D} p_d$. Let $\vartheta_i = d$, if $\mu_i = \beta_d$. Then $\mu_i(\vartheta_i = d) = \beta_d$ and $\mathbb{P}(\vartheta_i = d) = p_d$, $d = 0, \ldots, m$. Without loss of generality, assume $\boldsymbol{X}$ is re-scaled so that $\Sigma_{ii}^{(m)} = 1$.

(C). As $m \to \infty$, 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 $\boldsymbol{X}^{(m)}$ is short ranged, so that $\forall i \ \Sigma_{ik}^{(m)} = 0$ when $|i - k| \geq m^{\tau}$ for some constant $\tau \in (0, 1)$.

Let

$$T_{\mathrm{MG}}(x_i) = (1 - p) f(x_i | \, \theta_i = 0) / f(x_i), \tag{10}$$

which only involves the marginal distribution for x_i. The following theorem shows that as $m \to \infty$, $T_{\mathrm{MG},i}$ can approximates $T_{\mathrm{OR},i}$ well.

THEOREM 5. *Under the assumptions (A) - (C), let* $T_{\mathrm{OR},i}(\boldsymbol{x})$ *and* $T_{\mathrm{MG},i} = T_{\mathrm{MG}}(x_i)$ *be defined as in equation (5) and equation (10). Then* $\forall \ \epsilon > 0$*, for all* $i = 1, \ldots, m$*,*

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

Theorem 5 reveals that the marginal oracle statistic, though not as optimal as, is a uniformly consistent approximation to the joint oracle rule determined by $\boldsymbol{T}_{\mathrm{OR}}$. Note that $\boldsymbol{T}_{\mathrm{MG}}$ in Theorem 5 is a separable rule with the computation complexity $O(m)$, much smaller than the complexity of the joint oracle rule. In other words, we pay a little price in optimality but get great improvement in computational efficiency.

3.2. *Estimating the Marginal Oracle Statistics.* In model (9), 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 $\hat{T}_{\mathrm{MG},i}(x_i) = [(1-\hat{p})\hat{f}_0(x_i)/\hat{f}(x_i)] \wedge 1$.

For many multiple testing problem, we usually know the theoretical null distribution for X_i, *i.e.* we know f_0. In short-ranged correlated cases, Jin & Cai (2007) provided a consistent estimator for p. However, estimating f under the assumptions (A) - (D) is not straightforward. Although we assume f is a normal mixture in assumption (B), we do not really know how many components the mixture has. Therefore, parametric methods might not be a good choice to estimate f. As for non-parametric methods, there are many methods available for density estimation under the independent case. However, few of them can be applied to the dependent case.

We consider the kernel estimator of the density functions. It is not very different to obtain the non-parametric density estimation under the short-ranged dependent case from the independent case. Suppose the range of correlation is $B = m^{\tau},\ \tau \in (0,1)$ as in assumption (B). Define $K = m/B = m^{1-\tau}$. Let's rank each coordinate of $\boldsymbol{X}$ in the following way so that it becomes a matrix:

$$
\begin{array}{cccccc}
 & k=1 & 2 & \dots & K \\
b=1 & X_1 & X_{B+1} & \dots & X_{(K-1)B+1} \\
2 & X_2 & X_{B+2} & \dots & X_{(K-1)B+2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
B & X_B & X_{2B} & \dots & X_{KB}
\end{array} \tag{11}
$$

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 line of (11) is an independent subsequence of $\boldsymbol{X}$ with length $K = m^{1-\tau}$. Therefore, we can get kernel estimators $\hat{f}_b(x)$ based on $\boldsymbol{X}_{b,\cdot} = (X_{b,1}, \dots, X_{b,K})$, $b = 1, \dots, B$. Define $\hat{f}(x) = \sum_{b=1}^{B} \hat{f}_b(x)/B$. Let $\tilde{f}_b(x) = \mathbb{E}\,\hat{f}_b(x)$. Silverman (1978) obtained the convergence rate at which $\sup_x|\hat{f}(x) - \tilde{f}(x)|$ and $\sup_x|\tilde{f}(x) - f(x)|$ converges to zero. Note that $X_{b,\cdot}$ are identically distributed with correlation. Clearly, $\tilde{f}_b(x) = \tilde{f}(x)$ for all b and the convergence rate at which $\sup_x|\hat{f}_b(x) - \tilde{f}(x)|$ converges to zero doesn't depend on b. Thus $\sup_b \sup_x|\hat{f}_b(x) - \tilde{f}(x)| \to 0$, and

$$
\sup_x|\hat{f}(x) - \tilde{f}(x)| \le \frac{1}{B}\sum_{b=1}^{m} \sup_x|\hat{f}_b(x) - \tilde{f}(x)| \le \sup_b \sup_x|\hat{f}_b(x) - \tilde{f}(x)| \to 0.
$$

Together with $\sup_x|\tilde{f}(x) - f(x)| \to 0$, we have $\sup_x|\hat{f}(x) - f(x)| \to 0$.

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

$\sum_{b=1}^{B}\sum_{k=1}^{K} w(x, X_{b,k})/(KB)$ is the same as the kernel estimator viewing $X_1, \ldots, X_m$ are *i.i.d.* However, because of the correlation between the samples, the optimal convergence rate and the corresponding window width might be changed. Detailed analysis of the convergence rate is out of the scope of this paper.

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 plug-in procedure:

(12)
$$\text{Let } k = \max\{i : \frac{1}{i}\sum_{j=1}^{i} \hat{T}_{MG,(j)} \le \alpha\};\ \text{then reject all } \mathrm{H}_{(i)},\ i = 1, \ldots, k.$$

The next theorem show that the plug-in procedure (12) is asymptotically valid and optimal in the sense that it asymptotically controls mFDR under the given level α and minimizes mFNR.

THEOREM 6. *Assume* $\theta_i \overset{i.i.d}{\sim} \text{Bernoulli}(p^{(m)})$, $i = 1, \ldots, m$. *Let* $x_i \,|\, \theta_i$ *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* $\sup_x|\hat{f}(x) - f(x)| \to 0$ *and* $\sup_x|\hat{f}_0(x) - f_0(x)| \to 0$. *Define* $\hat{T}_{\mathrm{MG},i} = \hat{T}_{\mathrm{MG}}(x_i) = (1-p)\hat{f}_0(x_i)/\hat{f}(x_i)$. *Then asymptotically, the plug-in procedure (12) controls the mFDR at the given level* α *and simultaneously minimizes the mFNR.*

4. Simulation Studies. We conducted simulation studies to demonstrate that the marginal procedure works well, in the sense that it can not only control FDR but also has great efficiency. In the simulation studies, we assume the multivariate mixture normal model:

$$\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \Sigma), \tag{13}$$

where $\theta_i \overset{i.i.d}{\sim} \text{Bernoulli}(p)$. The model we used also satisfies the Assumptions (A) - (D).

4.1. *Simulation 1: The Performance of the Marginal Oracle Rule..* In the following simulations, we set the number of the random variables m=6000 and assume that the precision matrix Σ^{-1} is a blocked matrix with block size equal to 60. Within each block, the sub-precision matrix is a banded matrix with bandwidth 1, diagonal entries 1 and off-diagonal entries 0.2. We set $\mu(\theta_i = 0) = 0$. We developed an efficient computational algorithm

to compute the joint oracle statistics under the banded precision matrix assumption and the computation complexity is $O(2^b m)$, 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.

Figure 1 shows that the marginal oracle rule is close to the joint oracle rule and better than the BH procedure. Panel (a) plots mFNR as a function of mFDR, setting $\mu(\theta_i = 1) = 2.5$, the non-null proportion $p = 0.2$. Note that the mFDR level is the level we would like to control, and does not mean all the procedures shown here can achieve the level. Actually, BH procedure is overconservative and the actual mFDR for BH procedure can never achieve the specified mFDR level. In panel (b) of this figure, we plot the mFDR versus the alternative mean value, setting $p = 0.2$ and mFDR $= 0.1$. Here, we aim to demonstrate that the marginal oracle procedure can control mFDR. Panel (c) shows the mFNR *vs.* the alternative mean at the same setting as (b). Panels (b) and (c) show that the marginal oracle procedure can control mFDR and result in almost as small mFNR as the joint oracle procedure, which is optimal. Panel (d) plots mFDR versus the non-null proportion p, setting $\mu(\theta_i = 1) = 3.5$ and mFDR $= 0.1$. We noticed that when p increases, the BH procedure becomes more and more conservative, but the marginal oracle procedure still performs almost as efficiently as the joint oracle procedure.

Panels (e) - (g) of Figure 1 show the same comparisons when the non-null proportion $p = 0.02$. We still observed better performances in terms of mFNR for the joint oracle and the marginal oracle procedures over the BH procedure. The differences in mFNR increase as p increases (panel (h) of Figure 1).

4.2. *Simulation 2: The Performance of the Plug-in Marginal Rule.* We next compare the plug-in marginal procedure, the oracle marginal procedure and the BH procedure in the multivariate settings. In all simulation studies, we have $m = 6000$, and the covariance matrix Σ is a blocked matrix with block size equal to 10. Within each block, the sub-matrix is a banded matrix with bandwidth 4. We set the diagonal entries of Σ equal to 1 and other nonzero entries equal to 0.2. We set $\mu(\theta_i = 0) = 0$, with probability $1 - p$ and $\mu(\theta_i = 1) = \beta_1$ or β_2, with probability p_1 and p_2 respectively. Note that $p = p_1 + p_2$.

The results are shown in Figure (2). Panel (a) plots mFNR versus mFDR level we would like to control, setting $\beta_1 = -3$, $p_1 = 0.16$, $\beta_2 = 3$, $p_2 = 0.04$. From the plot, we observe that the BH procedure has larger mFNR and

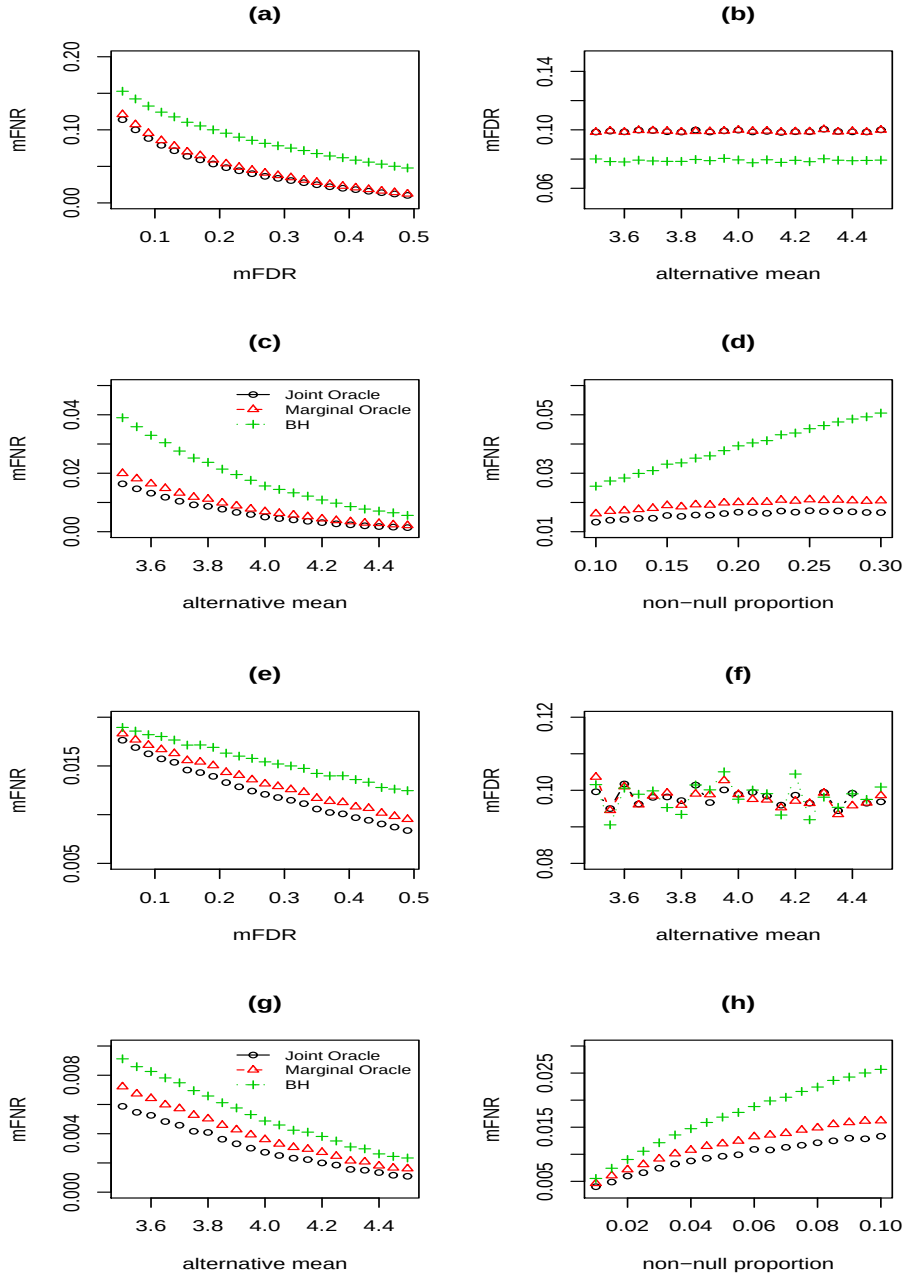


FIG 1. *Comparison of the joint oracle procedure, the marginal oracle procedure and the BH procedure. (a)* $\mu_1 = 2.5$, $p = 0.2$*; (b)* mFDR $= 0.1$, $p = 0.2$*; (c)* mFDR $= 0.1$, $p = 0.2$*; (d)* $\mu_1 = 3.5$, mFDR $= 0.1$*; (e)* $\mu_1 = 2.5$, $p = 0.02$*; (f)* mFDR $= 0.1$, $p = 0.02$*; (g)* mFDR $= 0.1$, $p = 0.02$*; (h)* $\mu_1 = 3.5$, mFDR $= 0.02$.

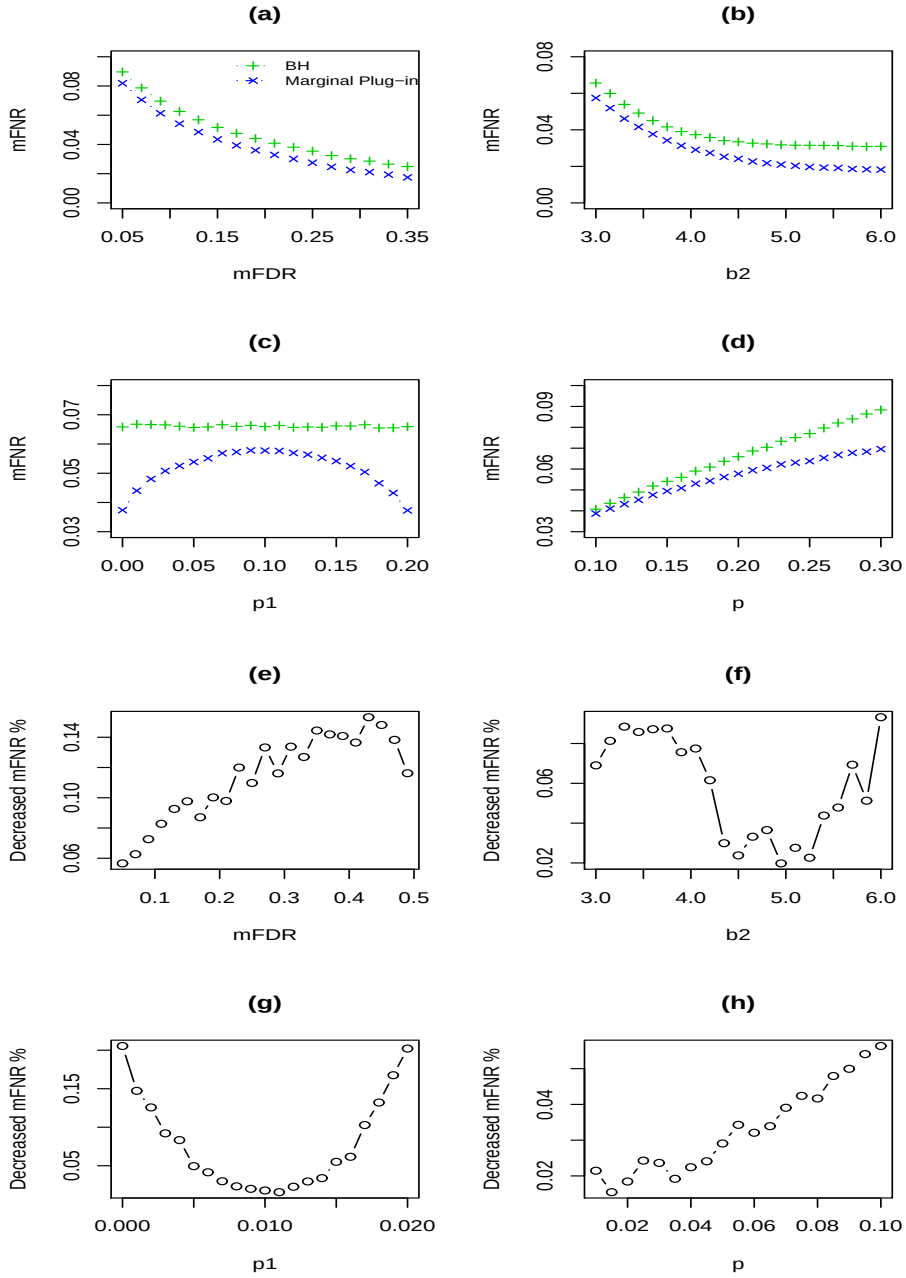


FIG 2. *Comparison of the joint oracle procedure, the marginal oracle procedure and the BH procedure. (a)* $\beta = (-3, 3)$, $p = (0.16, 0.04)$*; (b)* $\mathrm{mFDR} = 0.1$, $\beta_1 = -3$, $p = (0.04, 0.16)$*; (c)* $\mathrm{mFDR} = 0.1$, $\beta = (-3, 3)$, $p_1 + p_2 = 0.2$*; (d)* $\mathrm{mFDR} = 0.1$, $\beta = (-3, 3)$, $p_1 = p_2 = p/2$*; (e)* $\beta = (-3, 3)$, $p = (0.016, 0.004)$*; (f)* $\mathrm{mFDR} = 0.1$, $\beta_1 = -3$, $p = (0.004, 0.016)$*; (g)* $\mathrm{mFDR} = 0.1$, $\beta = (-3, 3)$, $p_1 + p_2 = 0.02$*; (h)* $\mathrm{mFDR} = 0.1$, $\beta = (-3, 3)$, $p_1 = p_2 = p/2$.

is more conservative. Panel (b) plots mFNR as a function of β_2, setting $\alpha = 0.1$, $\beta_1 = -3$, $p_1 = 0.1$ and $p_2 = 0.1$. The difference between mFNR increases as the alternative distribution becomes more asymmetric. This trend is confirmed in panel (c), where we set mFDR $= 0.1$, $\beta_1 = -3$, $\beta_2 = 3$, $p = 0.2$, and plot mFNR as a function of p_1. Compared to BH procedure, the marginal plug-in procedure can gain more in efficiency when the alternative marginal distribution is highly asymmetric. This is mainly because BH procedure only considers the p-value, which is the probability under the null. It does not use 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 panel (d), $\beta_1 = -3$, $\beta_2 = 3$, $p_1 = p_2 = 0.5p$. We plot mFNR versus the non-null proportion p. As p increases, BH procedure performs more conservatively and loses more efficiency compared with the marginal plug-in procedure.

We now consider the cases when the non-null proportion p=0.02, in which case the maximum possible mFNR is very small, and mFNR must be less than 0.02. Instead of showing the actual mFNRs as in previous simulations, we plot in panels (e) - (h) of Figure (2) the percentages of the mFNR decreases over the BH procedure when the marginal plug-in procedure was used. Similarly as before, we observed substantial decreases of the mFNRs using the marginal plug-in procedure over the BH procedure.

5. Application to analysis of case-control genetic study of neuroblastoma.

5.1. *Problem formulation and justification of the model assumptions.* In large-scale genetic association studies, we can formulate the problem as a multiple testing problem under the model (1). In such studies, for each SNP, usually we have n_1 cases and n_2 controls. For simplicity of demonstration, let $Y_{ik} = 1$ denote whether the kth patient in cases has at least one minor allele at the SNP i; otherwise, $Y_{ik} = 0$, $k = 1, \ldots, n_1$. Similarly, define Z_{il} as an indicator of whether the lth patient in controls has at least one minor allele at the SNP i, $l = 1, \ldots, n_2$. The linkage disequilibrium (LD) between the ith and the jth SNPs determines that Y_{ik} and Y_{jk} are not independent from each other, and so are Z_{il} and Z_{jl}. We assume the covariance structure due to LD among the cases is D_1 and the LD structure among the controls is D_2 and let $D_{1,ij} = \mathrm{Cov}(Y_{ik}, Y_{jk})$, and $D_{2,ij} = \mathrm{Cov}(Z_{il}, Zjl)$. We also have $\mathrm{Cov}(Y_{ik}, Z_{jl}) = 0$, $\forall\ i, j, k, l$, and $\mathrm{Cov}(Y_{ik}, Y_{jl}) = \mathrm{Cov}(Z_{ik}, Z_{jl}) = 0$, for $k \neq l$. Let $X_i = \sqrt{n_1}\bar{Y}_i - \sqrt{n_2}\bar{Z}_i$, where $\bar{Y}_i = \sum_{k=1}^{n_1} Y_{ik}/n_1$ and $\bar{Z}_i = \sum_{l=1}^{n_2} Y_{il}/n_2$. Define the latent variable θ_i as the indicator whether SNP i

is truly associated with the disease or not. It is reasonable to assume θ_i's are independent from each other because they are determined by biological nature. We have $\mathbb{E}\, X_i = 0$ if $\theta_i = 0$ and $\mathbb{E}\, X_i = \beta_i$ if $\theta_i = 1$. The covariance of X_i and X_j is given by

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}(\sqrt{n_1}\bar{Y}_i - \sqrt{n_2}\bar{Z}_i, \sqrt{n_1}\bar{Y}_j - \sqrt{n_2}\bar{Z}_j) \\ &= \frac{1}{n_1}\sum_{k=1}^{n_1}\sum_{l=1}^{n_1}\mathrm{Cov}(Y_{ik}, Y_{jl}) + \frac{1}{n_2}\sum_{k=1}^{n_2}\sum_{l=1}^{n_2}\mathrm{Cov}(Z_{ik}, Z_{jl}) \\ &= D_{1,ij} + D_{2,ij}\end{aligned}$$

which depends on the LD structure between X_i and X_j. Similarly, we can show that $\mathrm{Cov}(\boldsymbol{X}) = D_1 + D_2$. If we assume that $D_1 = D_2 = D$, then $\mathrm{Cov}(\boldsymbol{X}) = 2D$, which follows the model (1) with $\Sigma = 2D$.

We note that these assumptions (A) - (D) hold in the setting of large-scale genetic association studies. Assumption A is usually true since there is only a small set of markers that are related to a given disease, and the proportion of these markers among the total markers for testing is usually extremely low. Even if we increase the markers for testing, the number of genes that are associated with the disease should not expand. For assumption (B), recall that the data vector $\boldsymbol{X}$ are the score statistics or z-statistics summarized over all the SNP markers. When the sample size is large, it is reasonable to assume that they follow a joint normal distribution with marginal variance of 1 after re-scaling. Assumption (C) bounds the minimum eigenvalue of $\Sigma^{(m)}$ away from zero to make sure that the new variable added into the system is not close to any linear combination of the existing variables such that $\Sigma^{(m)}$ is becoming nearly singular. Assumption (D) imposes special structures on $\Sigma^{(m)}$, requiring that the correlation between two scores is zero if the two markers are distant enough.

5.2. *Results from a case-control study of neuroblastoma.* We applied the proposed FDR controlling procedure to a case-control genetic study of neuroblastoma (NB) 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 (Mosse *et al.*, 2005; Mosse *et al.*, 2008), and several common variations being enriched in sporadic neuroblastoma cases (Maris *et al.*, 2008). The latter genetic associations were discovered in a GWAS of sporadic NB cases, compared to children without cancer, conducted at the Children's Hospital of Philadelphia. After initial quality controls on samples

and the SNP genotypes, our discovery data set contained 1627 neuroblastoma case subjects of European ancestry, each of which contained 479,804 SNP markers. To correct the potential effects of population structure, 2575 matching control subjects of European ancestry were selected based on their low identity-by-state estimates with case subjects. Beside the original discovery set of cases and controls, we also have a replication data set that contains 398 case subjects and 1507 control subjects.

For each SNP, a score statistic was obtained by fitting a logistic regression model using an additive coding of the genotypes. We used the method of Jin & Cai (2007) to estimate p and used the kernel density estimation for the marginal densities. We set different nominal FDR values ($\text{mFDR} \in [0.01, 0.2]$) and examined the number of rejections based on the proposed marginal plug-in procedure and the BH procedure (see Figure 4). For a given mFDR level, the marginal plug-in procedure always identify more significant SNPs than the BH procedure, suggesting that it may lead to a smaller FNR and therefore better power of detecting the NB associated SNPs. This is especially important for initial genome-wide scanning in order to identify the potential candidate SNPs for a given FDR for follow-up studies.

Figure 4 shows at $\text{mFDR} = 0.05$ the SNPs claimed to be significant for both procedures. The marginal plug-in procedure identified 30 SNPs that are associated with NB. In contrast, BH procedure identified 24 SNPs. The six additional SNPs identified by our proposed marginal plug-in procedure, but missed by the standard BH procedure, are presented in Table 2. The p-values from the replication set are also shown in this table. It is interesting to see that two of these SNPs were replicated in the independent replication set, including the SNP rs7557557 in the BARD1 gene and the SNP rs2059320 in the DGKI gene. 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 (Irminger-Finger and Jefford, 2006). This provides some biological evidence for the association between the BARD1 gene and NB. Another gene, DGKI, is known to regulate Ras guanyl-releasing protein 3 and inhibits Rap1 signaling (Regier et al., 2005). However, the association between the variant in DGKI gene and NB is not clear and deserves further biological validation.

Other genes, although not replicated in our relatively small replication set, may also be associated with NB. For example, gene XPO4, encodes a nuclear export protein whose substrate, EIF5A2, is amplified in human tumors, is required for proliferation of XPO4-deficient tumor cells, and promotes

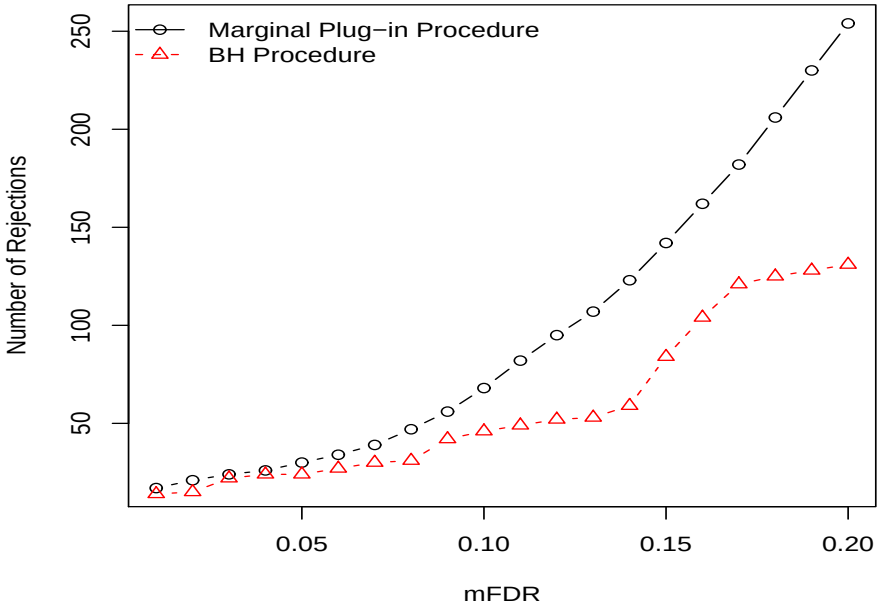


FIG 3. *Comparison of number of rejections for the marginal plug-in procedure and BH procedure for the case-control NB genetic association study.*

hepatocellular carcinoma in mice (Zender *et al.*, 2008). Another gene IKZF1 that encodes the lymphoid transcription factor IKAROS, has recently be reported to be associated with the poor outcome in acute lymphoblastic leukemia (Mullighan *et al.*, 2009), and was also discovered to harbor common variations associated with susceptibility to this disease (Trevino *et al.*, 2009). 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 (Heng *et al.*, 2010).

6. Discussion. We have reexamined the relationship between the multiple testing problem and weighted classification problem using the compound decision rule framework proposed under the dependent model instead of the independent model. Our theoretical argument does not require the monotone likelihood condition (MLR), which is a necessary condition for Sun & Cai (2007). Our analysis builds upon the connection between the

TABLE 2
Seven SNPs identified by the proposed procedure, but missed by the BH procedure for the NB case-control GWAS data.

SNP	Gene	Chr	$Zscore^2$	p-value	Rep. p-value
rs4770073	XPO4	13	21.74	3.13e-06	0.56
rs7557557	BARD1	2	21.79	3.05e-06	0.027
rs1714518	RSRC1	3	21.86	2.93e-06	0.23
rs10248903	IKZF1	7	21.25	4.03e-06	0.45
rs3828112	NR5A2	1	20.59	5.67e-06	0.39
rs2415603	Unknown	14	20.68	5.43e-06	0.39
rs2059320	DGKI	7	20.42	6.21e-06	0.052

Note: Fig 4 is too large to put together with the main file. Please see the pdf file on the last page.

FIG 4. *Analysis of case-control NB genetic association data: score statistics for 479804 SNPs, where those labeled in red were identified by both the marginal plug-in procedure and the BH procedure, and those labeled in blue were identified only by the marginal plug-in procedure for FDR=5%.*

multiple-testing problem and the weighted classification in the sense that we show for any multiple testing problem under the dependent model (2),

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. In other words, after we find the optimal decision rule for the weighted classification problem, we can claim that it is optimal among all the testing procedures for the multiple testing problems, including all the p-value based procedures. Though the oracle rule obtained is optimal, under the dependent model (2), it is hard to estimate and compute. We have developed a marginal oracle rule to approximate the optimal joint oracle rule and then discussed how to estimate the marginal rule. We have shown that the marginal plug-in procedure is asymptotically valid and optimal. Procedure-wise, the marginal plug-in rule is the same as the adaptive compound decision rules. This indirectly proved that although the adaptive compound decision rule is developed under the independent model, it is asymptotically valid and optimal under the short-range dependency case.

The theory and methods developed in this paper can also be applied to the models where the dependency not only lies in the measurement error, but also in the mean level. Suppose the mean of X_i is determined by the whole sequence of $\boldsymbol{\theta}$; in other words, $\mathbb{E}(X_i) = A_{i,\cdot}\boldsymbol{\mu}(\boldsymbol{\theta})$, where $A_{i,\cdot}$ is a vector. We can then assume

$$\boldsymbol{X} \mid \boldsymbol{\theta} \sim \mathrm{N}(A\boldsymbol{\mu}(\boldsymbol{\theta}), \Sigma), \tag{14}$$

where A is a full-rank matrix. Let $\tilde{X} = A^{-1}X$, $\tilde{\Sigma} = A^{-1}\Sigma(A^{-1})^{\mathrm{T}}$. Then model (14) can be reparameterized as

$$\tilde{\boldsymbol{X}} \mid \boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \tilde{\Sigma}), \tag{15}$$

If A is known, then model (15) has the same formula as model (1).

It should be emphasized that the marginal plug-in procedure does not necessarily to be the same as the adaptive compound decision rules. We showed that we can estimate the marginal density of X_i pretending that they are independent from each other. The procedure we adopted here produces consistent estimator for the marginal density. However, for correlated 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 need the assumptions (A)-(D). Among them, the assumptions (A) and (C) are necessary and cannot be weakened. For the dependency structure that does not satisfy the assumption (A) and (C), the marginal plug-in procedure does not have the asymptotic optimal property and how to obtain an asymptotically optimal test procedure is still an open problem.

APPENDIX A: PROOFS

We collect the proofs of the main theorems in this Appendix. The proofs of Theorem 1 and Theorem 2 follows that of Sun and Cai (2007) and are omitted here.

Appendix A1: Proof of Theorem 3

The proof of Theorem 3 requires the following Lemma.

LEMMA 1. *Consider the oracle classification statistic* $\mathbf{\Lambda}$ *defined in (4). Let* $g^s(t),\ s = 0, 1$ *be the average conditional pdf's of* $\mathbf{\Lambda}$*. Then*

$$\frac{g^1(t)}{g^0(t)} = \frac{1-p}{p \cdot t}.$$

That is, $g^1(t)/g^0(t)$ *is monotonically decreasing in* t*.*

Proof of Lemma 1. Suppose $t = 1/\lambda$. Consider the loss function (3), then $\delta_i(\Lambda_i, 1/t) = \mathrm{I}(\Lambda_i < t)$ is the optimal rule for the weighted classification problem. Consider a class of decision rules $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$, with $\delta_i(\Lambda_i, 1/t^*) = \mathrm{I}(\Lambda_i < t^*)$. Obviously, $\boldsymbol{\delta}(\mathbf{\Lambda}, 1/t)$ defined in (4) belongs to the class. Note that for any decision rule $\boldsymbol{\delta}(\mathbf{\Lambda}, t^*)$ belongs to this class,

$$\begin{aligned}
\mathcal{R}_\lambda(\boldsymbol{\delta}(\mathbf{\Lambda}, 1/t^*)) &= \mathbb{E}\left\{\frac{1}{m}\sum_{i=1^m}[(1/t)(1-\theta_i)\delta_i + \theta_i(1-\delta_i)]\right\} \\
&= \frac{1}{m}\sum_{i=1}^{m}[(1/t)\,\mathbb{P}(\theta_i = 0, \Lambda_i \le t^*) + \mathbb{P}(\theta_i = 1, \Lambda_i > t^*)] \\
&= p + (1/t)(1-p)G^0(t^*) - pG^1(t^*).
\end{aligned}$$

The optimal t^* that minimizes $\mathcal{R}_\lambda(\boldsymbol{\delta})$ satisfies $g^1(t^*)/g^0(t^*) = (1-p)/(pt)$. We know that $\boldsymbol{\delta}(\mathbf{\Lambda}, 1/t)$ is the optimal rule, and thus $t^* = t$. Hence $g^1(t)/g^0(t) = (1-p)/(pt)$.

Proof of Theorem 3. Note that Lemma 1 implies that

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

which is equivalent to $g^1(c)G^0(c) < g^0(c)G^1(c)$. Similarly,

$$\int_c^\infty g^0(t)\,\mathrm{d}t \Big/ \int_c^\infty g^1(t)\,\mathrm{d}t > \int_c^\infty g^0(t)\,\mathrm{d}t \Big/ \int_c^\infty \frac{g^1(c)}{g^0(c)} g^0(t)\,\mathrm{d}t = g^0(c)/g^1(c),$$

which is equivalent to $g^1(c)(1-G^0(c)) > g^0(c)(1-G^1(c))$.

Let $c = 1/\lambda$.

Now we show that mFDR is strictly increasing with c.

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

The derivative

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

Therefore, the mFDR is strictly increasing with c and therefore decreasing with λ.

Similarly, $\text{mFNR} = p(1-G^1(c))/(1-G(c))$. 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 is strictly decreasing with c and increasing with λ.

Therefore, for a given mFDR level α, we have a unique $\lambda(\alpha)$ such that $\text{mFDR}(\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda(\alpha))) = \alpha$ and $\text{mFNR}(\boldsymbol{\delta}(\boldsymbol{\Lambda}, \lambda(\alpha)))$ is the smallest among all the decision rules that can control mFDR.

Appendix A2: Proof of Theorem 5.

$$\begin{aligned}
&|\, T_{\text{MG},i}(\boldsymbol{x}) - T_{\text{OR},i}(\boldsymbol{x}) \,| \\
&= \left| \frac{(1-p)f(x_i|\,\theta_i = 0)}{f(x_i)} \right| \left| \frac{f(\boldsymbol{x}_{-i}|\,x_i, \theta_i = 0)}{f(\boldsymbol{x}_{-i}|\,x_i)} - 1 \right| \\
&\le \left| \frac{[\, f(\boldsymbol{x}_{-i}|\,x_i, \theta_i = 0) - f(\boldsymbol{x}_{-i}|\,x_i, \theta_i = 1)\,]\,\mathbb{P}(\theta_i = 1|\,x_i)}{f(\boldsymbol{x}_{-i}|\,x_i)} \right| \\
&= \left| \frac{pf(x_i|\theta_i = 1)[\, f(\boldsymbol{x}_{-i}|\,x_i, \theta_i = 0) - f(\boldsymbol{x}_{-i}|\,x_i, \theta_i = 1)\,]}{f(\boldsymbol{x})} \right| \\
&\le \left| \frac{2|\,\Omega_{ii}^{(m)}\,|^{1/2} \cdot p}{\sum_{\boldsymbol{\vartheta}} \exp[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}(\boldsymbol{\vartheta}))]\,\mathbb{P}(\boldsymbol{\vartheta})} \right|
\end{aligned}$$

For m sufficiently large, $\lambda_{\min}(\Sigma^{(m)}) > \kappa/2$. And thus $|\,\Omega_{ii}^{(m)}\,| \le \lambda_{\max}(\Omega^{(m)}) < 2/\kappa$.

For all $\eta > 0, \exists\, l_0$, such that $\mathbb{P}\left(|\sum_{\boldsymbol{\vartheta}} \exp[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\vartheta}))^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\vartheta}))]\,\mathbb{P}(\boldsymbol{\vartheta})| < l_0)\right) < \eta$. $\forall\ \varepsilon > 0$, $\exists\ m$ sufficiently large, such that $\frac{2\sqrt{2}p}{\sqrt{\kappa}\varepsilon} < l_0$. Then for all $i = 1, \ldots, m$,

$$\begin{aligned}&\mathbb{P}\left(|\, T_{\mathrm{MG},i}(\boldsymbol{x}) - T_{\mathrm{OR},i}(\boldsymbol{x})\,| > \epsilon\right)\\ &\quad= \mathbb{P}\left(\left|\sum_{\boldsymbol{\vartheta}} \exp[-(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\vartheta}))^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}(\boldsymbol{\vartheta}))/2]\,\mathbb{P}(\boldsymbol{\vartheta})\right| < l_0\right) =< \eta.\end{aligned}$$

Appendix A3: Proof of Theorem 6

The proof of Theorem 6 requires the following two Lemmas.

LEMMA 2. *Assume Assumption A,B and C hold. Let $\hat{p}$, $\hat{f}$, and $\hat{f}_0$ be estimates such that $\hat{p} \xrightarrow{\mathbb{P}} p$, $\sup_x|\hat{f}(x) - f(x)| \to 0$ and $\sup_x|\hat{f}_0(x) - f_0(x)| \to 0$, then $\hat{T}_{\mathrm{MG},i} \xrightarrow{\mathbb{P}} T_{\mathrm{OR},i}$.*

Proof of Lemma 2 $\forall\ \varepsilon > 0,\ \eta > 0, \exists\, l_0$, such that $\mathbb{P}(|f(x_i)| < l_0) < \eta/2$. There also exists M, such that for $m > M$, $|\hat{f}(x_i) - f(x_i)| < \varepsilon_0$ and $|\hat{f}_0(x_i) - f_0(x_i)| < \varepsilon_0$, where ε_0 satisfies $\left|\frac{3\varepsilon_0+\varepsilon_0^2}{(l_0-\varepsilon_0)l_0}\right| < \varepsilon/2$. Then with probability at least $1 - \eta/2$,

$$\begin{aligned}&|\hat{T}_{\mathrm{MG},i} - T_{\mathrm{MG},i}|\\ &\leq |\frac{(p+\varepsilon_0)(f_0(x_i)+\varepsilon_0)f(x_i) - pf_0(x_i)(f(x_i)-\varepsilon_0)}{f(x_i)(f(x_i)-\varepsilon_0)} \leq \left|\frac{3\varepsilon_0+\varepsilon_0^2}{(l_0-\varepsilon_0)l_0}\right| < \varepsilon/2.\end{aligned}$$

Combined with Theorem 5, we have

$$\mathbb{P}(|\hat{T}_{\mathrm{MG},i} - T_{\mathrm{OR},i}| > \varepsilon) \leq \mathbb{P}(|\hat{T}_{\mathrm{MG},i} - T_{\mathrm{MG},i}| > \varepsilon/2) + \mathbb{P}(|T_{\mathrm{MG},i} - T_{\mathrm{OR},i}| > \varepsilon/2) < \eta,$$

for m sufficiently large.

LEMMA 3. *Let $\hat{R}_\lambda = \frac{1}{m}\sum_{i=1}^m \mathrm{I}\{\hat{T}_{\mathrm{MG}}(x_i) \leq \lambda\}$ and $\hat{V}_t = \frac{1}{m}\sum_{i=1}^m \mathrm{I}\{\hat{T}_{\mathrm{MG}}(x_i) \leq t\}\hat{T}_{\mathrm{MG}}(x_i)$. Define $\hat{Q}_\lambda = \hat{V}_\lambda/\hat{R}_\lambda$. Then for $\alpha < t < 1$, $\hat{Q}_\lambda \xrightarrow{\mathbb{P}} Q_\lambda$.*

Proof of Lemma 3. Let $\nu_i = \mathbb{E}(\mathrm{I}\{T_{\mathrm{OR},i} < \lambda\})$ and $z_i = \mathrm{I}(T_{\mathrm{OR},i} < \lambda) - \nu_i$. Note that on $\{\boldsymbol{x} : |T_{\mathrm{OR},i} - T_{\mathrm{MG},i}| \leq \varepsilon\}$, $\mathrm{I}\{T_{\mathrm{OR},i} < \lambda\} = \mathrm{I}\{T_{\mathrm{MG},i} < \lambda\}$ holds unless $\lambda - \varepsilon \leq T_{\mathrm{MG},i} \leq \lambda + \varepsilon$. Therefore, $\forall\ \varepsilon > 0$, $|\mathbb{E}(I\{T_{\mathrm{OR},i} < \lambda\}) - \mathbb{E}(\mathrm{I}\{T_{\mathrm{MG},i} < \lambda\})| \leq \mathbb{P}(|T_{\mathrm{OR},i} - T_{\mathrm{MG},i}| > \varepsilon) + \mathbb{P}(\lambda - \varepsilon < T_{\mathrm{MG}} < \lambda + \varepsilon)$. The first term converges to zero uniformly by Theorem 5, and the second term converges to zero uniformly by the continuity of T_{MG}. Therefore, $\nu_i \to$

$\mathbb{P}(T_{\mathrm{MG}} < \lambda)$ uniformly. Note that $\mathrm{Var}(z_i) \le 1$ and $\mathrm{Cov}(z_i, z_j) \le 1$ for $|i - j| < m^\tau$ and 0 otherwise. Therefore $\mathrm{Var}(\sum_{i=1}^m z_i)/m^2 = \sum_{i=1}^m \mathrm{Var}(z_i)/m^2 + \sum_i \sum_j \mathrm{Cov}(z_i, z_j)/m^2 \le 1/m + m^{-1+\tau} \to 0$. By weak law of large numbers of triangle arrays, we have $\frac{1}{m}\sum_{i=1}^m z_i \xrightarrow{\mathbb{P}} 0$. Thus, $|R_\lambda - \mathbb{P}(T_{MG} < \lambda)| \le \frac{1}{m}\sum_{i=1}^m |\nu_i - \mathbb{P}(T_{MG} < \lambda)| + |\frac{1}{m}\sum_{i=1}^m z_i| \to 0$. Similarly, $\hat{R}_\lambda = \frac{1}{m}\sum_{i=1}^m [\mathrm{I}\{\hat{T}_{\mathrm{MG},i} < \lambda\} - \mathbb{P}(\hat{T}_{\mathrm{MG}} < \lambda)] + \mathbb{P}(\hat{T}_{\mathrm{MG}} < \lambda)$. The first part goes to 0 by the weak law of large numbers for triangle arrays, and the second part goes to $\mathbb{P}(T_{\mathrm{MG}} < \lambda)$. Then we have $\hat{R}_\lambda \xrightarrow{\mathbb{P}} R_\lambda$.

Now let's turn to prove $\hat{V}_\lambda \xrightarrow{\mathbb{P}} V_\lambda$. $\forall\, \varepsilon > 0$, if m is sufficiently large, then for all i, $\mathbb{E}(T_{\mathrm{OR},i} - T_{\mathrm{MG},i}) \le \varepsilon + \mathbb{P}(|T_{\mathrm{OR},i} - T_{\mathrm{MG},i}| > \varepsilon) < 2\varepsilon$. Therefore, $\mathbb{E}(T_{\mathrm{OR},i}) \to \mathbb{E}(T_{MG})$ uniformly. We then have $\mathbb{E}(T_{\mathrm{OR},i}\,\mathrm{I}\{T_{\mathrm{OR},i} < \lambda\}) - \mathbb{E}(T_{\mathrm{MG}}\,\mathrm{I}\{T_{\mathrm{MG}} < \lambda\}) \le \mathbb{P}(|T_{\mathrm{OR},i} - T_{\mathrm{MG}}| > \varepsilon) + \mathbb{E}(T_{\mathrm{OR},i} - T_{\mathrm{MG}}) + \mathbb{P}(\lambda - \varepsilon \le T_{MG} \le \lambda + \varepsilon)$. All three parts go to zero uniformly as $m \to \infty$. Now similarly as shown above for R_λ, we can get $V_\lambda \xrightarrow{\mathbb{P}} \mathbb{E}(T_{\mathrm{MG}}\,\mathrm{I}\{T_{\mathrm{MG}} < \lambda\})$. Also we can show similarly that $\hat{V}_\lambda \xrightarrow{\mathbb{P}} \mathbb{E}(T_{\mathrm{MG}}\,\mathrm{I}\{T_{\mathrm{MG}} < \lambda\})$. Then $\hat{V}_\lambda \xrightarrow{\mathbb{P}} V_\lambda$.

Consequently, we can get $\hat{Q}_\lambda = \hat{V}_\lambda/\hat{R}_\lambda \xrightarrow{\mathbb{P}} V_\lambda/R_\lambda = Q_\lambda$.

Proof of Theorem 6.

Define threshold $\lambda = \sup\{t \in (0,1) : Q(t) \le \alpha\}$, and the plug-in threshold $\hat{\lambda} = \sup\{t \in (0,1) : \hat{Q}(t) \le \alpha\}$. Since $\hat{Q}_\lambda \xrightarrow{\mathbb{P}} Q_\lambda$, by Lemma A.5 in Sun & Cai (2007), we have $\hat{\lambda} \xrightarrow{\mathbb{P}} \lambda$. The plug in procedure is equivalent to rejecting H_i^0 when $\hat{T}_{\mathrm{MG},i} \le \hat{\lambda}$. In the proof for Lemma 3, we have $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} < \lambda) \to \mathbb{P}(T_{\mathrm{MG}} < \lambda)$, $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} < \lambda \,|\, \mathrm{H}_i^0) \xrightarrow{\mathbb{P}} \mathbb{P}(T_{\mathrm{MG}} < \lambda \,|\, \mathrm{H}_i^0)$. Similarly, $\hat{T}_{\mathrm{MG},i} - \hat{\lambda} \to T_{\mathrm{MG},i} - \lambda$ uniformly, and thus $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \lambda) \to \mathbb{P}(T_{\mathrm{MG}} < \lambda)$, $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \lambda \,|\, \mathrm{H}_i^0) \to \mathbb{P}(T_{\mathrm{MG}} < \lambda \,|\, \mathrm{H}_i^0)$. Therefore, $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda} \,|\, \mathrm{H}_i^0) - \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} < \lambda \,|\, \mathrm{H}_i^0) \to 0$, $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda}) - \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} < \lambda) \to 0$. Then $\mathbb{E}\,\hat{V}_{\hat{\lambda}}/\mathbb{E}(\hat{R}_{\hat{\lambda}}) = (1-p) \cdot \frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda} \,|\, \mathrm{H}_i^0)/\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} < \hat{\lambda}) \to (1-p) \cdot \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} < \lambda \,|\, \mathrm{H}_i^0)/\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{OR},i} < \lambda) = \mathbb{E}\,V_\lambda/\mathbb{E}\,R_\lambda = \mathrm{mFDR}_{\mathrm{OR}} = \alpha$.

Similarly, for mFNR, we can prove that $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} > \hat{\lambda} \,|\, \mathrm{H}_i^1) - \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} > \lambda \,|\, \mathrm{H}_i^1) \to 0$, $\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} > \hat{\lambda}) - \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} > \lambda) \to 0$. Thus, $\mathbb{E}\,\hat{V}_{\hat{\lambda}}/\mathbb{E}(\hat{R}_{\hat{\lambda}}) = p \cdot \frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} > \hat{\lambda} \,|\, \mathrm{H}_i^1)/\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{MG},i} > \hat{\lambda}) \to p \cdot \frac{1}{m}\sum_{i=1}^m \mathbb{P}(T_{\mathrm{OR},i} > \lambda \,|\, \mathrm{H}_i^1)/\frac{1}{m}\sum_{i=1}^m \mathbb{P}(\hat{T}_{\mathrm{OR},i} > \lambda) = \mathbb{E}\,V_\lambda/\mathbb{E}\,R_\lambda = \mathrm{mFNR}_{\mathrm{OR}}$.

References

Bejamini, 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.

Bejamini, Y., and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25** 60–83.

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

Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scand Journal of Statistics* **34** 275–297.

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.

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.

Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., Wang, J., Yu, K., Chatterjee, N., Orr, N., Willett, W.C., Colditz, G.A., Ziegler, R.G., Berg, C.D., Buys, S.S., McCarty, C.A., Feigelson, H.S., Calle, E.E., Thun, M.J., Hayes, R.B., Tucke,r M., Gerhard, D.S., Fraumeni, J.F. Jr, Hoover, R.N., Thomas, G., Chanock, S.J. (2007). A genome−wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics* **39** 870–874.

Irminger-Finger, I. and Jefford, C.E. (2006). Is there more to BARD1 than BRCA1? *Nat Rev Cancer* **6** 382–91.

Jin, J., and Cai, T. (2007). Estimating the Null and the Proportion of non-Null Effects in Large-Scale Multiple Comparisons. *Journal of American Statistical Association* **102** 495–506.

Klein, R.J., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh. J. (2005). Complement factor H polymorphism in age−related macular degeneration. *Science* **3085** 385–389.

Li, H., Wei, Z. and Maris, J. (2010). A hidden Markov random field model for genome-wide association studies. *Biostatistics* **11** 139–150.

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.

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., Maris, J.M. (2005). High-resolution detection and mapping of genomic DNA alterations in neuroblastoma. *Genes, Chromosomes, Cancer* **43** 390–403.

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., Maris, J.M. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455** 930–035.

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., Children's Oncology Group (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *New England Journal of Medicine* **360** 470–480.

Regier, D.S., Higbee, J., Lund,K. M., Sakane, F., Prescott, S. M., 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.

Sabatti, C., Service, S. and Freimer, N. (2003). False discovery rates in linkage and association linkage genome screens for complex disorders. *Genetics* **164** 829–833.

Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren,

W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U., Prokunina-Olsson, L., Ding, C.J., Swift, A.J., Narisu, N., Hu, T., Pruim, R., Xiao, R., Li, X.Y., Conneely, K.N., Riebow, N.L., Sprau, A.G., Tong, M., White, P.P., Hetrick, K.N., Barnhart, M.W., Bark, C.W., Goldstein, J.L., Watkins, L., Xiang, F., Saramies, J., Buchanan, T.A., Watanabe, R.M., Valle, T.T., Kinnunen, L., Abecasis, G.R., Pugh, E.W., Doheny, K.F., Bergman, R.N., Tuomilehto, J., Collins, F.S., Boehnke, M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316** 1341–1345.

Silverman, B.W. (1978). Choosing the window width when estimating a density. *Biometrika* **65** 1–11.

Sun, W., and Cai, T. (2007). Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control. *Journal of American Statistical Association* **102** 901–912.

Sun, W., and Cai, T. (2009). Large-Scale Multiple Testing under Dependency. *Journal of the Royal Statistical Society* **Series B 71** 393–424.

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., Relling, M.V. (2009) Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nature Genetics* **41** 1001–1005.

Welcome Trust Case-control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447** 661–678.

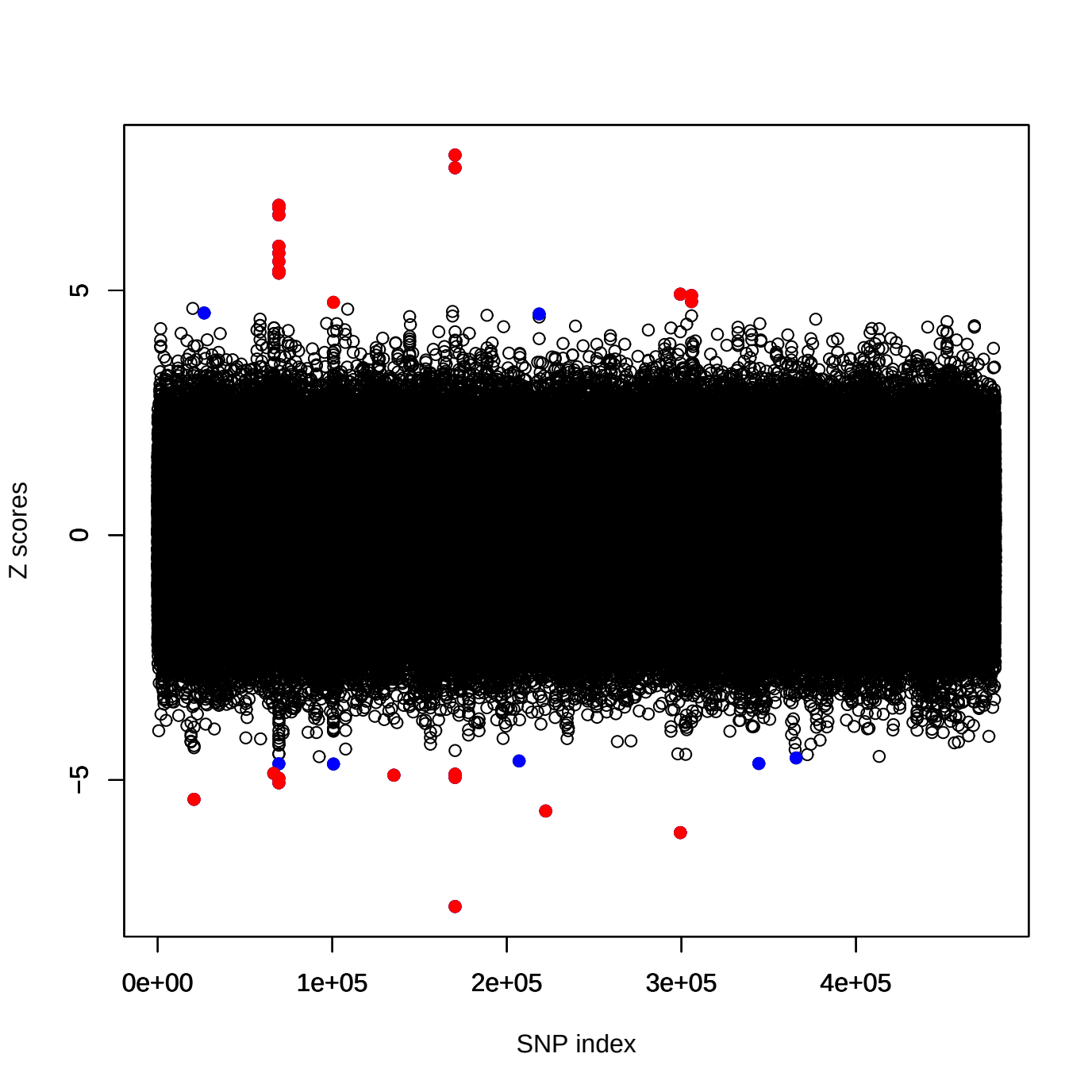
Wu, W. (2008). On False Discoverty Control under Dependence. *Annals of Statistics* **36** 364–380.

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., Lowe, S.W. (2008). An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135** 852–864.

Department of Biostatistics and Epidemiology
University of Pennsylvania School of Medicine
Philadelphia, PA 19104
E-mail: jichun@mail.med.upenn.edu
E-mail: hongzhe@upenn.edu

Department of Statistics
The Wharton School, University of Pennsylvania
Philadelphia, PA 19104.
E-mail: tcai@wharton.upenn.edu

Division of Oncology and Center for Childhood Cancer Research
Childrens Hospital of Philadelphia
Department of Pediatrics
University of Pennsylvania School of Medicine
Philadelphia, PA 19104
E-mail: Maris@email.chop.edu


Z scores
5
0
−5
0e+00
1e+05
2e+05
3e+05
4e+05
SNP index