Published in final edited form as:
J Am Stat Assoc. 2010 September 1; 105(491): 1215–1227. doi:10.1198/jasa.2010.tm09329.

False Discovery Rate Control With Groups

James X. Hu[Ph.D. Candidate],
Department of Statistics, Yale University, New Haven, CT 06511

Hongyu Zhao[Professor of Biostatistics Division], and
Department of Epidemiology and Public Health, Yale University, New Haven, CT 06511

Harrison H. Zhou[Assistant Professor]
Department of Statistics, Yale University, New Haven, CT 06511

James X. Hu: xing.hu@yale.edu; Hongyu Zhao: hongyu.zhao@yale.edu; Harrison H. Zhou: huibin.zhou@yale.edu

Abstract

In the context of large-scale multiple hypothesis testing, the hypotheses often possess certain group structures based on additional information such as Gene Ontology in gene expression data and phenotypes in genome-wide association studies. It is hence desirable to incorporate such information when dealing with multiplicity problems to increase statistical power. In this article, we demonstrate the benefit of considering group structure by presenting a *p*-value weighting procedure which utilizes the relative importance of each group while controlling the false discovery rate under weak conditions. The procedure is easy to implement and shown to be more powerful than the classical Benjamini–Hochberg procedure in both theoretical and simulation studies. By estimating the proportion of true null hypotheses, the data-driven procedure controls the false discovery rate asymptotically. Our analysis on one breast cancer dataset confirms that the procedure performs favorably compared with the classical method.

Keywords

Adaptive procedure; Benjamini–Hochberg procedure; Group structure; Positive regression dependence

1. INTRODUCTION

Ever since the seminal work of Benjamini and Hochberg (1995), the concept of false discovery rate (FDR) and the FDR controlling Benjamini–Hochberg (BH) procedure have been widely adopted to replace traditional methods, like family-wise error rate (FWER), in fields such as bioinformatics where a large number of hypotheses are tested. For example, in gene expression microarray experiments or brain image studies, each gene or brain location is associated with one hypothesis. Usually there are tens of thousands of them. The more conservative family-wise error rate controlling procedures often have extremely low power as the number of hypotheses gets large. Under the FDR framework, the power can be increased.

In many cases, there is prior information that a natural group structure exists among the hypotheses, or the hypotheses can be divided into subgroups based on the characteristics of the problem. For example, for gene expression data, Gene Ontology (*The Gene Ontology Consortium* 2000) provides a natural stratification among genes based on three ontologies. In genome-wide association study, each marker might be tested for association with several phenotypes of interest; or tests might be conducted assuming different genetic models (Sun et al. 2006). In clinical trials, hypotheses are commonly divided into primary and secondary

based on the relative importance of the features of the disease (Dmitrienko, Offen, and Westfall 2003). Ignorance of such group structures in data analysis can be dangerous. Efron (2008) pointed out that applying multiple comparison treatments such as FDR to the entire set of hypotheses may lead to overly conservative or overly liberal conclusions within any particular subgroup of the cases.

In multiple hypothesis testing, utilizing group structure can be achieved by assigning weights for the hypotheses (or *p*-values) in each group. Such an idea of using group information and weights has been adopted by several authors. Efron (2008) considered the Separate-Class model where the hypotheses are divided into distinct groups, and showed the legitimacy of such separate analysis for FDR methods. Benjamini and Hochberg (1997) analyzed both the *p*-value weighting and the error weighting methods and evaluated different procedures. Genovese, Roeder, and Wasserman (2006) investigated the merit of multiple testing procedures using weighted *p*-values and claimed that their weighted Benjamini–Hochberg procedure controls the FWER and FDR while improving power. Wasserman and Roeder (2006) further explored their *p*-value weighting procedure by introducing an optimal weighting scheme for FWER control. Roeder et al. (2006) considered linkage study to weight the *p*-values and showed their procedure improved power considerably when the linkage study is informative. Although in clinical trials, Finner and Roter (2001) pointed out that FDR control is hardly used, it is still potentially interesting to explore possible applications of FDR with group structures in clinical trials settings. Other notable publications include Storey, Taylor, and Siegmund (2004) and Rubin, van der Laan, and Dudoit (2005).

Very few results, however, have been published so far on proper *p*-value weighting schemes for procedures that control the FDR. In this paper, we will present the Group Benjamini–Hochberg (GBH) procedure, which offers a weighting scheme based on a simple Bayesian argument and utilizes the prior information within each group through the proportion of true nulls among the hypotheses. Our procedure controls the FDR not only for independent hypotheses but also for *p*-values with certain dependence structures. When the proportion of true null hypotheses is unknown, we show that by estimating it in each group, the data-driven GBH procedure offers asymptotic FDR control for *p*-values under weak dependence. This extends the results of both Genovese, Roeder, and Wasserman (2006) and Storey, Taylor, and Siegmund (2004).

When the information on group structure is less apparent, an alternative is to apply techniques such as clustering to assign groups. It can be a good strategy when we have spatially clustered hypotheses, that is, if one hypothesis is false, the nearby hypotheses are more likely to be false. For example, Quackenbush (2001) pointed out that in microarray studies, genes that are contained in a particular pathway or respond to a common environmental challenge, should show similar patterns of expression. Clustering methods are useful for identifying such gene expression patterns in time or space.

Our simulation results indicate that when the proportions of true nulls in each group are different, the GBH procedure is more powerful than the BH procedure while keeping the FDR controlled at the desired level. The GBH procedure also works well for situations where the number of signals is small among the hypotheses. Therefore, the procedure could be applied to microarray or genome-wide association studies where a large number of genes are monitored but only a few among them are actually differentially expressed or associated with disease. We apply our procedure to the analysis of a well-known breast cancer microarray dataset using two different grouping methods. The results indicate that the GBH procedure is able to identify more genes than the BH procedure by putting more focus on the potentially important groups. Figure 1 shows the advantage of the GBH procedure over the

BH procedure under *k*-means clustering for two methods of estimating the true null hypotheses in each group.

The rest of the paper is organized as follows. After a brief review of the FDR framework and the classical BH procedure, we present our GBH procedure in Section 2.2 and investigate our weighting scheme from both practical and Bayesian perspectives. Comparison of the classical BH and the GBH procedures in terms of expected number of rejections is discussed in Section 2.4. After discussing the data-driven GBH procedure in Section 2.3, we prove its asymptotic FDR control property in Section 3. Simulation studies of the BH and GBH procedures for normal random variables are reported in Section 4, including both independent and positive regression dependent cases. In Section 5, we show an application of the GBH procedure on a breast cancer dataset, using both the Gene Ontology grouping and *k*-means clustering strategies. The proofs for the main theorems are included in the Appendix.

2. THE GBH PROCEDURE

In this section, we introduce the Group Benjamini–Hochberg (GBH) procedure. It takes advantage of the proportion of true null hypotheses, which represents the relative importance of each group. We first examine the case where the proportions are known and then discuss data-driven procedures where the proportions are estimated based on the data.

2.1 Preliminaries

We first review the FDR framework and the classical BH procedure. Consider the problem of testing N hypotheses H_i vs H_{Ai}, $i \in I_N = \{1, \ldots, N\}$ among which n_0 are null hypotheses and $n_1 = N - n_0$ are alternatives (signals). Let V be the number of null hypotheses that are falsely rejected (*false discoveries*) and R be the total number of rejected hypotheses (*discoveries*). Benjamini and Hochberg (1995) introduced the FDR, which is defined as the expected ratio of V and R when R is positive, that is,

$$\mathrm{FDR} = E\left[\frac{V}{R \vee 1}\right], \tag{2.1}$$

where $R \vee 1 \equiv \max(R, 1)$. They also proposed the BH procedure which focuses on the ordered p-values $P_{(1)} \leq \cdots \leq P_{(N)}$ from N hypothesis tests. Given a level $\alpha \in (0, 1)$, the BH procedure rejects all hypotheses of which $P_{(i)} \leq P_{(k)}$, where

$$k = \max\left\{i \in \{1, \ldots, N\} : P_{(i)} \leq \frac{i\alpha}{N}\right\}. \tag{2.2}$$

Benjamini and Hochberg (1995) proved that for independent hypotheses, the BH procedure controls the FDR at level $\pi_0\alpha$ where $\pi_0 = n_0/N$ is the proportion of true null hypotheses. Hence, the BH procedure actually controls the FDR at a more stringent level. One can therefore increase the power by first estimating the unknown parameter π_0 using, say, $\hat{\pi}_0$, and then applying the BH procedure on the weighted p-values $\hat{\pi}_0 P_i$, $i = 1, \ldots, N$ at level α. Such a data-driven method is referred to as an *adaptive procedure*.

2.2 The GBH Procedure for the Oracle Case

When group information is taken into consideration, we assume that the N hypotheses can be divided into K disjoint groups with group sizes n_g, $g = 1, \ldots, K$. Let I_g be the index set of the gth group. The index set I_N of all hypotheses satisfies

$$I_N = \bigcup_{g=1}^{K} I_g = \bigcup_{g=1}^{K} (I_{g,0} \cup I_{g,1}), \tag{2.3}$$

where $I_{g,0} = \{i \in I_g : H_i \text{ is true}\}$ consists of indices for null hypotheses and $I_{g,1} = \{i \in I_g : H_i \text{ is false}\}$ is for the alternatives. Let $n_{g,0} = |I_{g,0}|$ and $n_{g,1} = n_g - n_{g,0}$ be the number of null and alternative hypotheses in group g, respectively. Then $\pi_{g,0} = n_{g,0}/n_g$ and $\pi_{g,1} = n_{g,1}/n_g$ are the corresponding proportions of null and alternative hypotheses in group g. Let

$$\pi_0 = \frac{1}{N} \sum_{g=1}^{K} n_g \pi_{g,0} \tag{2.4}$$

be the overall proportion of null hypotheses. In this section, we consider the so-called "oracle case," where $\pi_{g,0} \in [0, 1]$ is assumed to be given for each group. The case for unknown $\pi_{g,0}$ is discussed in Section 2.3.

Definition 1 (The GBH procedure for the oracle case)

1. For each p-value in group g, calculate the weighted p-values $P^w_{g,i} = \frac{\pi_{g,0}}{\pi_{g,1}} P_{g,i}$. Let $P^w_{g,i} = \infty$ if $\pi_{g,0} = 1$. If $\pi_{g,0} = 1$ for all g, accept all the hypotheses and stop. Otherwise go to the next step.
2. Pool all the weighted p-values together and let $P^w_{(1)} \le \cdots \le P^w_{(N)}$ be the corresponding order statistics.
3. Compute

$$k = \max\left\{ i : P^w_{(i)} \le \frac{i\alpha^w}{N} \right\}, \quad \text{where } \alpha^w = \frac{\alpha}{1 - \pi_0}.$$

   If such a k exists, reject the k hypotheses associated with $P^w_{(1)}, \ldots, P^w_{(k)}$; otherwise do not reject any of the hypotheses.

The GBH procedure weights the p-values for each group depending on the corresponding proportion of true null hypotheses in the group, that is, $\pi_{g,0}$. This idea is intuitively appealing because for any group with a small $\pi_{g,0}$, more rejections are expected and vice versa. The weight $\pi_{g,0}/\pi_{g,1}$ differentiates groups by (relatively) enlarging p-values in groups with larger $\pi_{g,0}$, therefore larger power is expected after applying the BH procedure on the pooled weighted p-values.

Benjamini and Yekutieli (2001) introduced the concept of positive regression dependence on subsets (PRDS) and proved that the BH procedure controls the FDR for p-values with such property. Finner, Dickhaus, and Rosters (2009, p. 603) argued that the PRDS property implies

$$\Pr\{R \geq j | P_i \leq t\} \text{ is nonincreasing in } t \tag{2.5}$$

for any $j \in I_N$, $i \in \cup_g I_{g,0}$ and $t \in (0, j\alpha/N]$. Examples of distribution satisfying the PRDS property include multivariate normal with nonnegative correlations and (absolute) multivariate *t*-distribution. It is worth pointing out that independence is a special case of PRDS; see Benjamini and Yekutieli (2001) and Finner, Dickhaus, and Rosters (2007 and Finner, Dickhaus, and Rosters (2009) for details.

For the oracle case, the following theorem guarantees that the GBH procedure controls the FDR rigorously for *p*-values with the PRDS property (hence provides FDR control for independent *p*-values as well).

Theorem 1: Assume the hypotheses satisfy (2.3) and the proportion of trull null hypotheses, $\pi_{g,0} \in [0, 1]$, is known for each group, then the GBH procedure controls the FDR at level α for *p*-values with the PRDS property.

Genovese, Roeder, and Wasserman (2006) analyzed the method of *p*-value weighting for independent *p*-values and proved FDR control of their procedure with a general set of weights. Some of the arguments in the proof of the above theorem can be implied by theorem 1 in Genovese, Roeder, and Wasserman (2006, p. 513). Nevertheless, we not only extend the result to *p*-values with the PRDS property, but also make up a small gap in their proof of FDR control (Genovese, Roeder, and Wasserman 2006, p. 514, first equation). Furthermore, the GBH procedure makes use of the information (i.e., $\pi_{g,0}$) embedded within each group, and provides a quasi-optimal way of assigning weights. Its advantage can be understood in two perspectives.

The GBH Procedure Works Well for Data With Sparse Signals: In many cases of multiple hypothesis testing, there tends to be a strong assumption that there are few signals, that is, most of the *N* hypotheses are true nulls. In microarray studies, for instance, majority of the genes are not related to certain disease, therefore we have the situation in which the $\pi_{g,0}$ of each group will be close to 1. Our weighting strategy performs well in such settings. For example, suppose we have two groups of *p*-values with $\pi_{1,0} = 0.9$ and $\pi_{2,0} = 0.99$. According to the GBH procedure, we are going to multiply 0.9/0.1 = 9 to the first group and 0.99/0.01 = 99 to the second group. Performing multiple comparison procedure on the combined weighted *p*-values means we put more attention on the *p*-values from the first group rather than the second one. As a result, more signals are expected. In the extreme case where one of the proportions is 1, say, $\pi_{1,0} = 1$ and $\pi_{2,0} \in (0, 1)$, according to the GBH procedure, all the *p*-values in the first group are rescaled to ∞, therefore no rejection (signal) would be reported in that group and our full attention would be focused on the second group. This is consistent with the fact that the first group contains no signal.

The GBH Procedure Has a Bayesian Interpretation: From the Bayesian point of view, the weighting scheme, $\pi_{g,0}/\pi_{g,1}$, can be interpreted as follows. Let $H_{g,i}$ be a hypothesis in group *g* such that $H_{g,i} = 0$ with probability $\pi_{g,0}$ and $H_{g,i} = 1$ with probability $\pi_{g,1} = 1 - \pi_{g,0}$. Let $P_{g,i}$ be the corresponding *p*-value and has a conditional distribution

$$P_{g,i}|H_{g,i}=0 \sim U_g; \quad P_{g,i}|H_{g,i}=1 \sim F_g.$$

The "Bayesian FDR" (Efron and Tibshirani 2002) of $H_{g,i}$ for $P_{g,i} \leq p$ is

$$\Pr(H_{g,i}=0|P_i \le p)=\frac{\pi_{g,0}U_g(p)}{\pi_{g,0}U_g(p)+\pi_{g,1}F_g(p)}. \tag{2.6}$$

If U_g follows a uniform distribution, the above equation becomes

$$\begin{aligned}\Pr(H_{g,i}=0|P_{gi} \le p)&=\frac{\pi_{g,0}p}{\pi_{g,0}p+\pi_{g,1}F_g(p)}\\&=\frac{(\pi_{g,0}/\pi_{g,1})p}{(\pi_{g,0}/\pi_{g,1})p+F_g(p)}\\&=\frac{[F_g(p)]^{-1}(\pi_{g,0}/\pi_{g,1})p}{[F_g(p)]^{-1}(\pi_{g,0}/\pi_{g,1})p+1}.\end{aligned}$$

Note that the above equation is an increasing function of $[F_g(p)]^{-1}(\pi_{g,0}/\pi_{g,1})p$, therefore ranking the Bayesian FDR is equivalent to focusing on the quantity

$$P_g^*=\frac{1}{F_g(p)}\frac{\pi_{g,0}}{\pi_{g,1}}p. \tag{2.7}$$

Then the ideal weight for the p-values in group g should be $[F_g(p)]^{-1}(\pi_{g,0}/\pi_{g,1})$, which can be viewed as two sources of influence on the p-values. If $F_g = F$ for all g, the first influence is through $[F(p)]^{-1}$, which can be regarded as the p-value effect. The other influence is the relative importance of the groups, that is, $\pi_{g,0}/\pi_{g,1}$. In practice, F_g is usually unknown and hard to estimate, especially when the number of alternatives is small. Hence, we just focus on the group effect in the ideal weight. Note that the weight we choose, that is, $\pi_{g,0}/\pi_{g,1}$ is not an aggressive one, since the cut-off point for the original p-values is big for important groups with small $\pi_{g,0}/\pi_{g,1}$, which implies that the ideal weight for groups with small $\pi_{g,0}/\pi_{g,1}$ is relatively smaller.

2.3 The Adaptive GBH Procedure

As mentioned in the previous sections, knowledge of the proportion of true null hypotheses, that is, π_0, can be useful in improving the power of FDR-controlling procedures. Such information, however, is not available in practice. Estimating the unknown quantity using observed data is then a natural idea, which brings us to the adaptive procedure.

Definition 2 (The adaptive GBH procedure)

1. For each group, estimate $\pi_{g,0}$ by $\hat{\pi}_{g,0}$.
2. Apply the GBH procedure in Definition 1, with $\pi_{g,0}$ replaced by $\hat{\pi}_{g,0}$.

Various estimators of π_0 were proposed by Schweder and Spjøtvoll (1982) and Storey (2002) and Storey, Taylor, and Siegmund (2004) based on the tail proportion of p-values, and by Efron et al. (2001) based on the mixture densities of null and alternative distribution of hypotheses. Jin and Cai (2007) estimated π_0 based on the empirical characteristic function and Fourier analysis. Meinshausen and Rice (2006) and Genovese and Wasserman (2004) provided consistent estimators of π_0 under certain conditions.

The adaptive GBH procedure does not require a specific estimator of $\pi_{g,0}$, therefore people may choose their favorite estimator in practice. We take the following two examples to illustrate the practical use of the adaptive GBH procedure.

Example 1 [Least-Slope (LSL) method]: The least-slope (LSL) estimator proposed by Benjamini and Hochberg (2000) performs well in situations where signals are sparse. Hsueh, Chen, and Kodell (2003) compared several methods including Schweder and Spjøtvoll (1982), Storey (2002), and the LSL estimator, and found that the LSL estimator gives the most satisfactory empirical results.

Definition 3 (Adaptive LSL GBH procedure)

1. For p-values in each group g, starting from $i = 1$, compute $l_{g,i} = (n_g + 1 - i)/(1 - P_{g,(i)})$, where $P_{g,(i)}$ is the ith order statistics in group g. As i increases, stop when $l_{g,j} > l_{g,j-1}$ for the first time.
2. For each group, compute the LSL estimator of $\pi_{g,0}$

$$\gamma_g^{\mathrm{LSL}} = \min\left(\frac{\lfloor l_{g,j}\rfloor + 1}{n_g}, 1\right). \tag{2.8}$$

3. Apply the GBH procedure at level α with $\pi_{g,0}$ replaced by γ_g^{LSL}.

The LSL estimator is asymptotically related to the estimator proposed by Schweder and Spjøtvoll (1982). It is also conservative in the sense that it overestimates $\pi_{g,0}$ in each group.

Example 2 [The Two-Stage (TST) method]: Benjamini, Krieger, and Yekutieli (2006) proposed the TST adaptive BH procedure and showed that it offers finite-sample FDR control for independent p-values.

Definition 4 (Adaptive TST GBH procedure)

1. For p-values in each group g, apply the BH procedure at level $\alpha' = \alpha/(1 + \alpha)$. Let $r_{g,1}$ be the number of rejections.
2. For each group, compute the TST estimator of $\pi_{g,0}$

$$\gamma_g^{\mathrm{TST}} = \frac{n_g - r_{g,1}}{n_g}. \tag{2.9}$$

3. Apply the GBH procedure at level α' with $\pi_{g,0}$ replaced by γ_g^{TST}.

The TST method applies the BH procedure in the first step and uses the number of rejected hypotheses as an estimator of the number of alternatives.

Both the LSL and TST methods are straightforward to implement in practice and in the next section we show both of them have good asymptotic properties. Our simulation and real data analysis show that they outperform the adaptive BH procedure, in which the group structure of the data is not considered.

Remark 1: We should point out that in applications, the adaptive GBH procedure *does not* rely on which estimator people choose. The performance, however, does depend on the distribution of signal among groups. If there is no significant difference in the proportions of signals among hypotheses for different groups, the adaptive GBH procedure degenerates to uni-group case. As long as the groups are dissimilar in terms of true null proportion and the estimator of $\pi_{g,0}$ can detect (not necessarily fully detect) the proportion of true null hypotheses for each group, the adaptive GBH procedure is expected to outperform the adaptive BH procedure.

2.4 Comparison of the GBH and BH Procedures

In previous sections, we show that the GBH procedure controls the FDR for the finite sample case when the $\pi_{g,0}$'s are known. It is of interest to compare the performance of GBH with that of the BH procedure. In this section, we are going to compare the expected number of rejections for the two procedures.

Benjamini and Hochberg (1995) showed that the BH procedure controls the FDR at level $\pi_0\alpha$. In order to compare the BH and GBH procedures at the same α level, we consider the following rescaled p-values:

$$\text{BH: } \pi_0\mathbf{P} \quad \text{vs} \quad \text{GBH: } \frac{\pi_{g,0}}{\pi_{g,1}}(1-\pi_0)\mathbf{P}_g, \tag{2.10}$$

where $\pi_{g,0} \in (0, 1)$. Note that $\pi_0 = \pi_{g,0}(1 - \pi_0)/\pi_{g,1}$ when $\pi_{g,0} = \pi_0$ for all g.

For group g, let D_g be the distribution of p-values such that

$$D_g(t)=\pi_{g,0}U_g(t)+\pi_{g,1}F_g(t), \tag{2.11}$$

where U_g and F_g are the distribution functions for p-values under the null and alternative hypotheses. Let $\tilde{D}_g(t)$ be the empirical cumulative distribution function of p-values in group g, that is,

$$\tilde{D}_g(t)=\frac{1}{n_g}\left(\sum_{i\in I_{g,0}}\{P_i \le t\}+\sum_{i\in I_{g,1}}\{P_i \le t\}\right). \tag{2.12}$$

It is proved in Lemma 2 that under weak conditions $\tilde{D}_g(t)$ converges uniformly to $D_g(t)$.

For the uni-group case, in the framework of (2.10), it has been proved by several authors (Benjamini and Hochberg 1995; Storey 2002; Genovese and Wasserman 2002; Genovese, Roeder, and Wasserman 2006) that the threshold of the BH procedure can be written as

$$T_{\text{BH}}=\sup_{t\in[0,\pi_0]}\left\{t:\frac{t}{C_N(t/\pi_0)}\le\alpha\right\},$$

where $C_N(t)=\frac{1}{N}\sum_{i\in I_N}\{P_i \le t\}$ is the empirical cumulative distribution function of the p-values, and the procedure rejects any hypothesis with a p-value less than or equal to T_{BH}.

We can extend this result to the framework of GBH. For notation purpose define $\mathbf{a}=\{a_g\}_{g=1}^K$ where $ag = \pi_{g,0}(1 - \pi_0)/(1 - \pi_{g,0})$. Let $G_N(\mathbf{a}, t)$ be the empirical distribution of the weighted p-values, that is,

$$G_N(\mathbf{a}, t) = \frac{1}{N}\sum_{g=1}^{K} n_g \tilde{D}_g\left(\frac{t}{a_g}\right) = \frac{1}{N}\sum_{g=1}^{K}\left(\sum_{i\in I_{g,0}}\left\{P_i \le \frac{t}{a_g}\right\} + \sum_{i\in I_{g,1}}\left\{P_i \le \frac{t}{a_g}\right\}\right). \tag{2.13}$$

Note that $N \cdot G_N(\mathbf{a}, t)$ is the number of rejections for the (oracle) GBH procedure with respect to the threshold t on the weighted p-values. When $\pi_0 < 1$, where π_0 defined in (2.4) is the overall proportion of null hypotheses, it can be shown that the threshold of the GBH procedure is equivalent to

$$T_{\mathrm{GBH}} = \sup_{t\in c(\mathbf{a})}\left\{t : \frac{t}{G_N(\mathbf{a}, t)} \le \alpha\right\},$$

where $c(\mathbf{a}) = \{t : 0 \le t \le \max_g a_g\}$.

For any *fixed* threshold $t \in c(\mathbf{a})$, let $\mathbb{E}[R_{\mathrm{BH}}(t)]$ and $\mathbb{E}[R_{\mathrm{GBH}}(t)]$ be the expected number of rejections of the BH and GBH procedure, respectively. The following theorem provides a sufficient condition for $\mathbb{E}[R_{\mathrm{BH}}(t)] \le \mathbb{E}[R_{\mathrm{GBH}}(t)]$.

Lemma 1—Let U_g and F_g be the distributions of p-values under the null and alternative hypotheses in group g. Assume $U_g = U$ and $F_g = F$ for all g. If $U \sim \mathrm{Unif}[0, 1]$ and $x \mapsto F(t/x)$ is convex for $x \ge \tilde{t}$, where $\tilde{t} = (1 - \pi_0)\min_g \frac{\pi_{g,0}}{1-\pi_{g0}}$, then $\mathbb{E}[R_{\mathrm{BH}}(t)] \le \mathbb{E}[R_{\mathrm{GBH}}(t)]$.

Take the classical normal mean model for an example. Suppose we observe $X_i = \theta + Z_i$, where $Z_i \overset{\mathrm{iid}}{\sim} \mathrm{N}(0, 1)$. Consider the multiple testing problem

$$H_i : \theta = 0 \quad \text{vs} \quad H_{Ai} : \theta = \theta_A > 0, \quad i = 1, \ldots, N.$$

The distribution of p-values under alternative is $F(u) = 1 - \Phi[\Phi^{-1}(1 - u) - \theta_A]$, where Φ is the standard Normal distribution function. It can be shown that $x \mapsto F(t/x)$ is convex if

$$\theta_A \le \frac{2\varphi(\Phi^{-1}(1 - t/\tilde{t}))}{t/\tilde{t}}, \tag{2.14}$$

where φ is the standard Normal density function. Note that $t/\tilde{t}$ is the threshold of the unscaled p-values for rejecting the corresponding hypotheses in one group, therefore $t/\tilde{t}$ is small. Since the right-hand side of (2.14) is a decreasing function of $t/\tilde{t}$, (2.14) becomes $\theta_A \le 4.12$ when $t/\tilde{t} \le 0.05$ and $\theta_A \le 5.33$ when $t/\tilde{t} \le 0.01$. This suggests the convexity is true for most of the cases.

For adaptive procedures, $\pi_{g,0}$ is replaced by its estimator $\hat{\pi}_{g,0}$. Let $\widehat{\mathbf{a}} = \{\widehat{a}_g\}_{g=1}^{K}$ where $\hat{a}_g = \hat{\pi}_{g,0}(1 - \hat{\pi}_0)/(1 - \hat{\pi}_{g,0})$ and $\hat{\pi}_0 = \Sigma_g n_g \hat{\pi}_0/N$. To conduct the BH procedure adaptively, we first estimate π_0 by $\hat{\pi}_0$ and then perform the BH procedure at level $\alpha/\hat{\pi}_0$. The corresponding threshold of the adaptive BH procedure is

$$\widehat{T}_{\mathrm{BH}} = \sup_{t\in[0,\widehat{\pi}_0]} \left\{ t: \frac{t\pi_0/\widehat{\pi}_0}{C_N(t/\widehat{\pi}_0)} \le \alpha \right\}, \tag{2.15}$$

and the threshold of the adaptive GBH procedure is

$$\widehat{T}_{\mathrm{GBH}} = \sup_{t\in c(\widehat{\mathbf{a}})} \left\{ t: \frac{\sum_{g=1}^{K} \pi_g \pi_{g,0} t/\widehat{a}_g}{G_N(\widehat{\mathbf{a}}, t)} \le \alpha \right\}, \tag{2.16}$$

where

$$G_N(\widehat{\mathbf{a}}, t) = \frac{1}{N} \sum_{g=1}^{K} \left(\sum_{i\in I_{g,0}} \{P_i \le t/\widehat{a}_g\} + \sum_{i\in I_{g,1}} \{P_i \le t/\widehat{a}_g\} \right). \tag{2.17}$$

Remark 2—Both (2.15) and (2.16) depend on the data, hence they are no longer fixed. In the next section we are going to prove that $\hat{T}_{\mathrm{BH}}$ and $\hat{T}_{\mathrm{GBH}}$ converges in probability to some fixed t^*_{BH} and t^*_{GBH}, respectively. Theorem 4 in Section 3.1 demonstrates that $t^*_{\mathrm{BH}} \le t^*_{\mathrm{GBH}}$ and therefore the adaptive GBH procedure rejects more than the adaptive BH procedure asymptotically.

3. GBH ASYMPTOTICS

In many applications of multiple hypothesis testing, not only are the proportions of true null hypotheses unknown, but the number of hypotheses is also very large. It is hence applicable to analyze the behavior of the GBH procedure for large N. In this section, we focus on the asymptotic property of the adaptive GBH procedure.

Genovese, Roeder, and Wasserman (2006) and Storey, Taylor, and Siegmund (2004) proved some useful results for asymptotic FDR control using empirical process argument for the BH procedure. We extend them further in the setting of the GBH procedure. We first discuss the case where we have consistent estimator of the proportion of true null hypotheses, then move on to a more general case.

3.1 Adaptive GBH With Consistent Estimator of $\pi_{g,0}$

When $N \to \infty$ and the number of groups K is finite, we assume the following condition is satisfied in every group

$$\frac{n_g}{N} \to \pi_g, \quad \frac{n_{g,0}}{n_g} \to \pi_{g,0}, \quad \frac{n_{g,1}}{n_g} \to \pi_{g,1}, \quad \text{where } \pi_g, \pi_{g,0}, \pi_{g,1} \in (0,1). \tag{3.1}$$

By the construction, $\Sigma_g \pi_g = 1$ and $\pi_{g,0} + \pi_{g,1} = 1$. The following lemma shows that (2.12) converges uniformly to (2.11) under the above condition.

Lemma 2—Under (3.1), let $U_g(t)$ and $F_g(t)$ be continuous functions. For any $t \ge 0$, if the p-values satisfy

$$\frac{1}{n_{g,0}}\sum_{i\in I_{g,0}}\{P_i \le t\} \xrightarrow{\text{a.s.}} U_g(t), \tag{3.2}$$

$$\frac{1}{n_{g,1}}\sum_{i\in I_{g,1}}\{P_i \le t\} \xrightarrow{\text{a.s.}} F_g(t). \tag{3.3}$$

Then $\sup_t|\tilde{D}_g(t) - D_g(t)| \xrightarrow{\text{a.s.}} 0$.

Storey, Taylor, and Siegmund (2004) described *weak dependence* as any type of dependence in which Conditions (3.2) and (3.3) are satisfied. Weak dependence contains the case of independent *p*-values, but for *p*-values with the PRDS property, these conditions are not necessarily true. An example is given in Section 4.

In this section, we focus on the case when we have consistent estimator of $\pi_{g,0}$ in every group, that is,

$$\widehat{\pi}_{g,0} \xrightarrow{P} \pi_{g,0} \in (0, 1) \quad \text{for all } g. \tag{3.4}$$

Recall that $\widehat{\mathbf{a}}=\{\widehat{a}_g\}_{g=1}^K$, where $\hat{a}_g = \hat{\pi}_{g,0}\,(1 - \hat{\pi}_0)/(1 - \hat{\pi}_{g,0})$. Under the above condition, we have $\widehat{\mathbf{a}} \xrightarrow{P} \mathbf{a}$. Let $G(\mathbf{a}, t)=\sum_{g=1}^K \pi_g(\pi_{g,0}U_g(t/a_g)+\pi_{g,1}F_g(t/a_g))$ be the limiting distribution of the weighted *p*-values for all groups and let $B(\mathbf{a}, t) = t/G(\mathbf{a}, t)$. Then define

$$t^*_{\text{GBH}} = \sup_{t\in C(\mathbf{a})}\{t: B(\mathbf{a}, t) \le \alpha\}.$$

The following theorem establishes the asymptotic equivalence of (2.16) and t^*_{GBH}, and thus implies asymptotic FDR control of the adaptive GBH procedure.

Theorem 2—Suppose Conditions (3.1) through (3.4) are satisfied for all groups. Suppose further that $U_g(t) = t$ for $0 \le t \le 1$ in every group. If $t \mapsto B(\mathbf{a}, t)$ has a nonzero derivative at t^*_{GBH} and $\lim_{t\downarrow 0} B(\mathbf{a}, t) \ne \alpha$, then $\widehat{T}_{\text{GBH}} \xrightarrow{P} t^*_{\text{GBH}}$ and $\text{FDR}(\hat{T}_{\text{GBH}}) \le \alpha + o(1)$.

Note that the statement of the theorem has a similar flavor as theorem 2 in Genovese, Roeder, and Wasserman (2006, p. 515). But our assumption is weaker and more importantly, the $\hat{\pi}_{g,0}$'s are estimated based on the data.

Similarly, for the adaptive BH procedure, we define the distribution of all *p*-values as $C(t) = \pi_0 U(t) + (1 - \pi_0)F(t)$, where $U(t)$ and $F(t)$ are continuous functions. Let t^*_{BH} be such that

$$t^*_{\text{BH}} = \sup_{t\in[0,\pi_0]}\left\{t: \frac{t}{C(t/\pi_0)} \le \alpha\right\}.$$

The following theorem illustrates that asymptotically the adaptive GBH procedure has more expected number of rejections than the adaptive BH procedure. Note that $R_{\mathrm{BH}}(\cdot)$ and $R_{\mathrm{GBH}}(\cdot)$ denote the number of rejections of the BH and GBH procedures, respectively.

Theorem 3—Under Conditions (3.1) through (3.4). Assume in each group $U_g(t) = U(t) = t$, $0 \le t \le 1$ and $F_g(t) = F(t)$, where $x \mapsto F(t/x)$ is convex for $x \ge \min_g a_g$. Assume further that both $B(\mathbf{a}, t)$ and $t/C(t/\pi_0)$ are increasing in t. If $\pi_0 \ge \alpha$ and $\lim_{t\downarrow 0} t/C(t/\pi) \le \alpha$, then $t^*_{\mathrm{BH}} \le t^*_{\mathrm{GBH}}$, and therefore

$$\mathbb{E}[R_{\mathrm{BH}}(\widehat{T}_{\mathrm{BH}})]/\mathbb{E}[R_{\mathrm{GBH}}(\widehat{T}_{\mathrm{BH}})] \le 1+o(1).$$

Remark 3—Sometimes the assumption that all the alternative hypotheses across different groups follow the same distribution may not be appropriate. The condition $F_g(t) = F(t)$ in the above theorem is necessary to establish Theorem 3. However, that assumption is not a necessity in establishing Theorem 2 and Theorem 4, where we show FDR control for the adaptive GBH procedure.

3.2 Discussion for Inconsistent Estimator of $\pi_{g,0}$

For a general estimator of $\pi_{g,0}$, let $\hat{\pi}_{g,0} \in (0, 1]$ be an estimator of $\pi_{g,0}$ such that

$$\widehat{\pi}_{g,0} \xrightarrow{P} \zeta_g \in (0, 1] \quad \text{and} \quad \bar{\zeta}=\sum_g \pi_g \zeta_g<1, \tag{3.5}$$

where the latter condition means at least one ζ_g is less than 1 among all groups. Let $\rho=\{\rho_g\}_{g=1}^K$ where $\rho_g = \zeta_g/(1 - \zeta_g)$ and $\rho_g = \infty$ when $\zeta_g = 1$. Then, we have $\widehat{\mathbf{a}} \xrightarrow{P} \rho$. Let $G(\rho, t)=\sum_{g=1}^K \pi_g(\pi_{g,0}U_g(t/\rho_g)+\pi_{g,1}F_g(t/\rho_g))$ be the limiting distribution of the weighted p-values for all groups. Denote $B(\rho, t)=\frac{\sum_{g=1}^K \pi_g \pi_{g,0} t/\rho_g}{G(\rho,t)}$ and define

$$t^*_{\mathrm{GBH}} = \sup_{t\in c(\rho)} \{t: B(\rho, t) \le \alpha\}. \tag{3.6}$$

Theorem 4—Suppose Conditions (3.1) through (3.3) and (3.5) are satisfied for all groups. Suppose further that $U_g(t) = t$ for $0 \le t \le 1$ and $\zeta_g \ge b_g\pi_{g,0}$ for some $b_g > 0$ in every group. If $t \mapsto B(\boldsymbol{\rho}, t)$ has a nonzero derivative at t^*_{GBH} and $\lim_{t\downarrow 0} B(\rho, t) \ne \alpha$, then $\widehat{T}_{\mathrm{GBH}} \xrightarrow{P} t^*_{\mathrm{GBH}}$ and $\mathrm{FDR}(\hat{T}_{\mathrm{GBH}}) \le \alpha/\min_g\{b_g\} + o(1)$. In particular, $\mathrm{FDR}(\hat{T}_{\mathrm{GBH}}) \le \alpha + o(1)$ when $b_g \ge 1$ for all groups.

The theorem generalizes the result in Theorem 2 and indicates that the adaptive GBH procedure controls the FDR at level α not only for consistent estimators of $\pi_{g,0}$'s, but also for asymptotically conservative estimators.

Remark 4—For the TST estimator γ_g^{TST} in (2.9), note that

$$\gamma_g^{\mathrm{TST}}=1-\frac{1}{n_g}\sum_{i\in I_g}\{P_i\le \widehat{T}_0\},$$

where $\hat{T}_0$ is the threshold for the BH procedure in the first step. Following from Theorem 4, $\widehat{T}_0 \xrightarrow{P} t_0$, where t_0 satisfies $t_0/(\pi_{g,0}t_0 + \pi_{g,1}F_g(t_0)) = \alpha'$. Since $F_g(t_0) \le 1$, it can be shown that $t_0 \le (1 - \pi_{g,0})\alpha'/(1 - \alpha'\pi_{g,0})$. Then,

$$\gamma_g^{\mathrm{TST}} \xrightarrow{P} \pi_{g,0}(1-t_0)+\pi_{g,1}(1-F_g(t_0))=1-\frac{t_0}{\alpha}\ge(1-\alpha')\pi_{g,0}.$$

Therefore, by Theorem 4 the adaptive TST GBH procedure controls FDR at level $\alpha'/(1 - \alpha')$ = α asymptotically.

Remark 5—As $n_g \to \infty$, the LSL estimator γ_g^{LSL} defined in (2.8) can be viewed as a special case of the estimator $\hat{\pi}_{g,0}(\lambda)$ proposed by Schweder and Spjøtvoll (1982). For fixed λ, $\hat{\pi}_{g,0}(\lambda)$ satisfies

$$\widehat{\pi}_{g,0}(\lambda)=\frac{n_g-\sum_{i\in I_g}\{P_i\le\lambda\}}{n_g(1-\lambda)}\xrightarrow{\text{a.s}}\pi_{g,0}+\pi_{g,1}\frac{1-F_g(\lambda)}{1-\lambda}\ge\pi_{g,0},$$

under Conditions (3.2) and (3.3). Therefore, $\hat{\pi}_{g,0}(\lambda)$ is asymptotically conservative and by Theorem 4 the FDR is controlled asymptotically at α for $\hat{\pi}_{g,0}(\lambda)$.

4. SIMULATION STUDIES

For simplicity, assume the hypotheses are divided into two groups. Without loss of generality, assume there are n observations in each group. Consider the following model, let

$$S_{gi}=\theta_i+\sqrt{1-\xi_g}\cdot Z_{gi}-\sqrt{\xi_g}Z_0,\quad i=1,\ldots,n;g=1,2 \tag{4.1}$$

be the ith test statistic in group g, where Z_{gi} and Z_0 are independent standard Normal random variables. Note that $\mathrm{Cov}(T_{gu}, T_{gv}) = \xi_g$, for $u, v \in \{1, \ldots, n\}$, $u \ne v$ and the model satisfies the PRDS property discussed in Section 2.2 when $0 \le \xi_g \le 1$. Similar dependence structures were considered in Finner, Dickhaus, and Roters (2007) and Benjamini, Krieger, and Yekutieli (2006). Note that when $\xi_g > 0$, Conditions (3.2) and (3.3) are not satisfied for large N due to the extra Z_0 term.

Consider the hypothesis testing problem with two groups $H_0 : \theta_j = 0$ vs $H_a : \theta_j > 0$, for $j = 1, \ldots, 2n$. In this section, we compare the performances of the BH and GBH procedures for both oracle and adaptive cases. For the adaptive BH procedure, we compute the (either LSL or TST) estimator $\hat{\pi}_0$ for all p-values and then apply the BH procedure at level $\alpha/\hat{\pi}_0$.

Four combinations of $\pi_{g,0}$'s are considered: (1) $\pi_{1,0} = 0.9$ vs $\pi_{2,0} = 0.2$; (2) $\pi_{1,0} = 0.8$ vs $\pi_{2,0} = 0.4$; (3) $\pi_{1,0} = 0.99$ vs $\pi_{2,0} = 0.9$; (4) $\pi_{1,0} = 0.999$ vs $\pi_{2,0} = 0.9$. In each case, we generate $n_g = 10{,}000$ test statistics for each of the two groups based on (4.1). In every group, $n_g\pi_{g,0}$ of

the hypotheses are null and the rest are alternatives with corresponding $\theta = 3$ in one group and $\theta = 5$ in the other group. Other combinations of n and θ's are also considered and the results are similar. Since we have the information about which hypothesis is from the alternative, the power for the two procedures can be obtained, which is the proportion of true rejections among the false null hypotheses. The power of the BH and GBH procedures is evaluated *in pairs* based on 200 iterations for each of the 20 FDR levels between 0.01 and 0.2. The results for the oracle and adaptive cases are as follows.

For the oracle case with independent *p*-values, Figure 2 indicates that the GBH procedure outperforms the BH procedure in all four cases, especially when $\pi_{g,0}$'s are close to 1 (the last two panels). The more the groups differ in $\pi_{g,0}$, the larger the difference is obtained in the power of the two procedures. This is also true for *p*-values with the PRDS property. Figure 3 shows the power difference between the GBH and BH procedures for *p*-values under model (4.1) with $\xi_1 = \xi_2 = 0.5$. All points being above zero indicates the GBH procedure outperforms the BH procedure for all four cases.

For the adaptive case with independent *p*-values, we estimate the unknown $\pi_{g,0}$'s using either the TST or LSL method introduced in Section 2.3. Figure 4 indicates that the average of the false discovery proportion (FDP) is controlled at a prespecified FDR level for both the BH and GBH procedures with either the TST or LSL method. The power improvement of the adaptive GBH over the adaptive BH procedure is shown in Figure 5. Both the TST GBH and the LSL GBH procedures are more powerful than the corresponding adaptive BH procedures.

We also analyze the performance of the adaptive GBH procedure for weighting scheme other than $\pi_{g,0}/\pi_{g,1}$. According to (2.6), when U_g is uniform, the Bayesian FDR is $[\pi_{g,0}/D_g(p)]p$, where $D_g(p)$ is the distribution function of *p*-values in group g. It's therefore natural to consider the weight $\hat{\pi}_{g,0}/\tilde{D}_g(p)$, where $\tilde{D}_g$ is the empirical distribution, as pointed out by a referee. Although this weight takes into consideration of the distribution of *p*-values in each group, the power of the adaptive procedure using this weight is often low in the situation where we have sparse signals and estimating the alternative distribution is difficult.

5. APPLICATIONS

van't Veer et al. (2002) used microarrays to study the primary breast tumors of 78 young patients, of which 44 developed cancer in less than 5 years and the other 34 were cancer free during that period. In total 24,184 genes were monitored and *p*-values were obtained for each gene by comparing the mean ratio of $\log_{10}$ intensities. A fraction of the data is listed in Table 1.

In order to apply the GBH procedure which makes use of the group structure, we need to stratify the genes first. Here we consider two grouping strategies.

5.1 Grouping Using Gene Ontology (GO)

The GO project (*The Gene Ontology Consortium* 2000) provides detailed annotations for a gene product's biology. It consists of three ontologies, namely Biological Process, Molecular Function, and Cellular Component, each representing a key concept in Molecular Biology. The GO terms are classified into one of the three ontologies. Based on the GO terms, one can construct a top-down tree diagram, in which the higher nodes represent more general biological concepts.

The tree structure provides the idea of GO grouping which can be summarized as follows. After choosing one of the three ontologies, say Biological Process, some higher nodes are

selected as ancestors according to the generic GO slim file, which contains the broad overview of each ontology without the detail of each GO terms (available at http://www.geneontology.org/GO.slims.shtml). Next, for those genes with GO IDs, we trace them upward to the nodes we have chosen. Genes that share common ancestors are then grouped together. The biggest concern for GO grouping in our case is that the mapping rate is low. Even though the GO consortium updates their data base on a daily basis, not every gene in our data has a GO ID. For our case, 9492 of the 24,184 genes have the annotation information for Biological Process, therefore the mapping rate for our data is 9492/24,184 ≈ 39%.

However, we may still use the remaining 9492 genes to see the difference of using group information in multiple hypothesis testings. We first divide the genes into four groups with respect to Biological Process, that is, (1) Cell communication; (2) Cell growth and/or maintenance; (3) Development; and (4) Multifunction. The results for the adaptive BH and GBH procedures are listed in Table 2. For simplicity, we just report the results for the LSL method.

At FDR level 0.15, Table 2 indicates that the adaptive GBH procedure focuses more on groups with smaller estimated $\pi_{g,0}$'s, that is, groups (2) and (4), and is able to discover genes that are not detectable using the adaptive BH procedure. In fact, as shown in Figure 6, using either the LSL or TST method, the adaptive BH procedure cannot detect any signals when the FDR level is less than 0.15.

Even though the mapping rate for this dataset is low, the idea of GO grouping could be a good choice if the data were collected in terms of GO identities; or the mapping between the GO ID and other gene IDs (e.g., GenBank Accession Number) was more complete. Then each group may correspond to different biological processes or genetic functions within the tumor and the GBH method can help us to find more signals among desired groups.

5.2 Grouping Using *k*-Means Clustering

Another grouping idea is to apply clustering. Here we choose *k*-means clustering with initial points satisfying maximum separation rule based on all the 78 samples. Note that we are *not* just clustering the *p*-values. Unlike GO grouping, *k*-means clustering makes use of the whole dataset and we do not have to worry about the mapping rate. Although we do have the difficulties regarding cluster analysis, for example, the choice of initial points, number of clusters, and the interpretation of each cluster, we use it as an illustrative example to compare the performances of the adaptive BH and GBH procedures.

In order to have a reliable estimator for each group, six clusters are selected such that within each cluster there are at least 200 genes. Table 3 shows the results for the two procedures using the LSL method at FDR level 0.1. Most of the additional discoveries found by the adaptive GBH procedure come from the first cluster, which is expected to contain more signals because the estimated $\pi_{g,0}$ is relatively smaller than the others. Gene-annotation enrichment analysis confirms that those 109 genes selected by the GBH procedure in the first cluster are closely associated with cell cycle, mitosis, chromosome segregation, and phosphoprotein, which are common factors related to breast cancer. Similar analyses on the four and five-cluster cases indicate that the number of genes detected by the adaptive GBH procedure is 145 and 226, respectively. Out of those genes, 94 of them are overlapping with the six-cluster case. Comparing with an average of eight genes discovered by random grouping, which assigns groups randomly with the same group sizes as the above three cases, clustering and using the GBH procedure is advantageous in our case.

For comparison of the two procedures over a range of FDR levels, Figure 1 shows the increment in the number of signals detected by the adaptive GBH over BH procedure for both the LSL and TST methods. This indeed shows that by applying the GBH procedure, more signals can be detected.

6. SUMMARY

We have presented a new approach of *p*-value weighting procedure GBH for controlling the FDR when the hypotheses are believed to have some group structure. We prove that it controls the FDR for hypotheses with the positive regression dependence property when the proportions of true null hypotheses $\pi_{g,0}$'s are known in each group. The weighting scheme $(\pi_{g,0}/\pi_{g,1})$ for the *p*-values in each group makes it possible to focus on groups that are expected to have more signals.

By estimating $\pi_{g,0}$ for each group, we propose the adaptive GBH procedure and show that it controls the FDR asymptotically under weak dependence. We demonstrate the benefit of the adaptive GBH over BH by two methods of estimating $\pi_{g,0}$, namely the LSL (Benjamini and Hochberg 2000) and the TST (Benjamini, Krieger, and Yekutieli 2006) estimators. As we have pointed out, the choice of the estimator for $\pi_{g,0}$ in general does not affect the performance of the adaptive GBH procedure. In practice, people may choose the estimator based on their own preference.

Acknowledgments

James Hu and Harrison Zhou's research is supported in part by NSF grant DMS-0645676. Hongyu Zhao's research is supported in part by NSF grant DMS-0714817 and NIH grant GM59507.

References

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Ser B. 1995; 57:289–300.

Benjamini Y, Hochberg Y. Multiple Hypothesis Testing With Weights. Scandinavian Journal of Statistics. 1997; 24:407–418.

Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing With Independent Statistics. Journal of Educational and Behavioral Statistics. 2000; 25:60–83.

Benjamini Y, Yekutieli D. The Control of the False Discovery Rate in Multiple Testing Under Dependency. The Annals of Statistics. 2001; 29:1165–1188.

Benjamini Y, Krieger MA, Yekutieli D. Adaptive Linear Step-Up Procedures That Control the False Discovery Rate. Biometrika. 2006; 93(3):491–507.

Dmitrienko A, Offen WW, Westfall HP. Gatekeeping Strategies for Clinical Trials That Do Not Require All Primary Effects to Be Significant. Statistics in Medicine. 2003; 22:2387–2400. [PubMed: 12872297]

Efron B. Simultaneous Inference: When Should Hypothesis Testing Problems be Combined? The Annals of Applied Statistics. 2008; 2 (1):197–223.

Efron B, Tibshirani R. Empirical Bayes Methods and False Discovery Rates for Microarrays. Genetic Epidemiology. 2002; 23:70–86. [PubMed: 12112249]

Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes Analysis of a Microarray Experiment. Journal of the American Statistical Association. 2001; 96:1151–1160.

Finner H, Roters M. On the False Discovery Rate and Expected Type 1 Errors. Biometrical Journal. 2001; 43(8):985–1005.

Finner H, Dickhaus T, Roters M. Dependency and False Discovery Rate: Asymptotics. The Annals of Statistics. 2007; 35:1432–1455.

Finner H, Dickhaus T, Roters M. On the False Discovery Rate and an Asymptotically Optimal Rejection Curve. The Annals of Statistics. 2009; 37(2):596–618.

Genovese CR, Wasserman L. A Stochastic Process Approach to False Discovery Control. The Annals of Statistics. 2002; 32:1035–1061.

Genovese CR, Wasserman L. A Stochastic Process Approach to False Discovery Control. The Annals of Statistics. 2004; 32:1035–1061.

Genovese CR, Roeder K, Wasserman L. False Discovery Control With P-Value Weighting. Biometrika. 2006; 93:509–524.

Hsueh H, Chen J, Kodell R. Comparison of Methods for Estimating the Number of True Null Hypotheses in Multiplicity Testing. Journal of Biopharmaceutical Statistics. 2003; 13(4):675–689. [PubMed: 14584715]

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

Meinshausen N, Rice J. Estimating the Proportion of False Null Hypotheses Among a Large Number of Independently Tested Hypotheses. The Annals of Statistics. 2006; 34(1):373–393.

Quackenbush J. Computational Analysis of Microarray Data. Nature Review Genetics. 2001; 2:418–427.

Roeder K, Bacanu SA, Wasserman L, Devlin B. Using Linkage Genome Scans to Improve Power of Association in Genome Scans. The American Journal of Human Genetics. 2006; 78(2):243–252.

Rubin, D.; van der Laan, M.; Dudiot, S. Technical report. University of California, Berkeley, Division of Biostatistics; 2005. Multiple Testing Procedures Which Are Optimal at a Simple Alternative; p. 171

Schweder T, Spjøtvoll E. Plots of P-Values to Evaluate Many Tests Simultaneously. Biometrika. 1982; 69:493–502.

Storey JD. A Direct Approach to False Discovery Rates. Journal of the Royal Statistical Society, Ser B. 2002; 64:479–498.

Storey JD, Taylor JE, Siegmund D. Strong Control, Conservative Point Estimation, and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach. Journal of the Royal Statistical Society, Ser B. 2004; 66:187–205.

Sun L, Craiu RV, Paterson AD, Bull SB. Stratified False Discovery Control for Large-Scale Hypothesis Testing With Application to Genome-Wide Association Studies. Genetic Epidemiology. 2006; 30(6):519–530. [PubMed: 16800000]

The Gene Ontology Consortium. Gene Ontology: Tool for the Unification of Biology. Nature Genetics. 2000; 25:25–29. [PubMed: 10802651]

van't Veer LJ, et al. Gene Expression Profiling Predicts Clinical Outcome of Breast Cancer. Nature. 2002; 415:530–536. [PubMed: 11823860]

Wasserman, L.; Roeder, K. "Weighted Hypothesis Testing," technical report. Carnegie Mellon University, Dept. Statistics; 2006.

APPENDIX: PROOFS

Proof of Theorem 1

The proof is based on the proof of theorem 4.1 in Finner, Dickhaus, and Roters (2009). Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$ be the multiple testing procedure. $\phi_i = 0$ means retaining H_i and $\phi_i = 1$ means rejecting H_i. The FDR for oracle GBH is

$$\begin{aligned}\mathrm{FDR}(\phi_{\mathrm{GBH}}) &= E\left[\frac{V}{R\vee 1}\right] = E\left[\frac{V}{R\vee 1}\sum_{j\in I_N}\{R=j\}\right]\\ &= \sum_{g=1}^{K}\sum_{i\in I_{g,0}}\sum_{j\in I_N}\frac{1}{j}\Pr(R=j,\phi_{g,i}=1)\\ &= \sum_{g=1}^{K}\sum_{i\in I_{g,0}}\sum_{j\in I_N}\frac{1}{j}\Pr\left(R=j,\frac{\pi_{g,0}}{\pi_{g,1}}P_{g,i}\le\frac{j}{N}\alpha^{w}\right).\end{aligned}$$

Note that if $\pi_{g,0} = 0$ or $\pi_{g,0} = 1$ for some g, that group doesn't contribute to the FDR because $I_{g,0} = \varnothing$ if $\pi_{g,0} = 0$ and $\Pr(R=j, \frac{\pi_{g,0}}{\pi_{g,1}} P_{g,i} \le \frac{j}{N}\alpha^w)=0$ if $\pi_{g,0} = 1$ (we treat $\pi_{g,0}/\pi_{g,1}$ as ∞). Let $\eta = \{g : \pi_{g,0} \in (0, 1)\}$. Then

$$\mathrm{FDR}(\phi_{\mathrm{GBH}})=\sum_{g\in\eta}\sum_{i\in I_{g,0}}\sum_{j\in I_N}\frac{1}{j}\Pr\left(R=j, \frac{\pi_{g,0}}{\pi_{g,1}}P_{g,i} \le \frac{j}{N}\alpha^w\right).$$

Using the proof of theorem 4.1 in Finner, Dickhaus, and Rosters (2009), we have

$$\begin{aligned}\mathrm{FDR}(\phi_{\mathrm{GBH}}) &\le \sum_{g\in\eta}\frac{\pi_{g,1}\alpha^w}{\pi_{g,0}N}\sum_{i\in I_{g,0}}\Pr\left(R\ge 1 | P^w_{g,i} \le \tfrac{1}{N}\alpha^w\right)\\ &=\sum_{g\in\eta}\frac{\pi_{g,1}\alpha^w}{\pi_{g,0}N}\sum_{i\in I_{g,0}}1=\sum_{g\in\eta}\frac{\pi_{g,1}\alpha^w}{\pi_{g,0}N}\cdot n_g\pi_{g,0}\\ &=\frac{\sum_{g\in\eta}n_g\pi_{g,1}}{\sum_{g=1}^K n_g\pi_{g,1}}\alpha \le \alpha.\end{aligned}$$

Proof of Lemma 1

For the unweighted case, the expected number of rejections of BH procedure for a given threshold t, where $t \le \tilde{t} = (1 - \pi_0) \max_g \pi_{g,0}/\pi_{g,1}$ is

$$\begin{aligned}\mathbb{E}[R_{\mathrm{BH}}(t)]&=\mathbb{E}\sum_{g=1}^K\sum_{j\in I_g}\{\pi_0 P_j \le t\}\\ &=\mathbb{E}\sum_{g=1}^K\left(\sum_{j\in I_{g,0}}\left\{P_j \le \tfrac{t}{\pi_0}\right\}+\sum_{j\in I_{g,1}}\left\{P_j \le \tfrac{t}{\pi_0}\right\}\right)\\ &\le Nt+\sum_{g=1}^K n_g\pi_{g,1}F\left(\tfrac{t}{\pi_0}\right).\end{aligned}$$

Similarly, the expected number of rejections of GBH procedure for $t \le \tilde{t}$ is

$$\begin{aligned}\mathbb{E}[R_{\mathrm{GBH}}(t)]&=\mathbb{E}\sum_{g=1}^K\sum_{j\in I_g}\left\{\frac{\pi_{g,0}}{\pi_{g,1}}(1-\pi_0)P_j \le t\right\}\\ &=Nt+\sum_{g=1}^K n_g\pi_{g,1}F\left(\frac{\pi_{g,1}t}{\pi_{g,0}(1-\pi_0)}\right).\end{aligned}$$

Let $\varepsilon_g = n_g\pi_{g,1}/\Sigma_g\, n_g\pi_{g,1}$ and $x_g = \pi_{g,0}\,(1 - \pi_0)/\pi_{g,1}$. Now that $x \mapsto F(t/x)$ is convex for all $x \ge \tilde{t}$. We have $F(t/\Sigma_g\varepsilon_g x_g) \le \Sigma_g\, \varepsilon_g F(t/x_g)$, that is,

$$F\left(\frac{t}{\pi_0}\right) \le \frac{1}{\sum_g n_g\pi_{g,1}}\sum_{g=1}^K n_g\pi_{g,1}F\left(\frac{\pi_{g,1}t}{\pi_{g,0}(1-\pi_0)}\right).$$

Therefore, $\mathbb{E}[R_{\mathrm{BH}}(t)] \le \mathbb{E}[R_{\mathrm{GBH}}(t)]$ for $t \le \tilde{t}$.

Proof of Lemma 2

Consider the estimator of $D_g(t)$ defined in (2.12). Under (3.1), for any $t \geq 0$,

$$\sup_t \left|\tilde{D}_g(t) - D_g(t)\right| \leq \sup_t \left|\frac{n_{g,0}}{n_g}\frac{1}{n_{g,0}}\sum_{i\in I_{g,0}}\{P_i \leq t\} - \pi_{g,0}U_g(t)\right| + \sup_t \left|\frac{n_{g,1}}{n_g}\frac{1}{n_{g,1}}\sum_{i\in I_{g,1}}\{P_i \leq t\} - \pi_{g,1}F_g(t)\right|$$
$$\leq \left|\frac{n_{g,0}}{n_g} - \pi_{g,0}\right| + \pi_{g,0}\sup_t \left|\frac{1}{n_{g,0}}\sum_{i\in I_{g,0}}\{P_i \leq t\} - U_g(t)\right| + \left|\frac{n_{g,1}}{n_g} - \pi_{g,1}\right| + \pi_{g,1}\sup_t \left|\frac{1}{n_{g,1}}\sum_{i\in I_{g,1}}\{P_i \leq t\} - F_g(t)\right|,$$

where $\sup_t |n_{g,0}^{-1}\sum_{i\in I_{g,0}}\{P_i \leq t\} - U_g(t)| \xrightarrow{\text{a.s.}} 0$ and $\sup_t |\{n_{g,1}^{-1} \times \sum_{i\in I_{g,1}}\{P_i \leq t\} - F_g(t)| \xrightarrow{\text{a.s.}} 0$ by Glivenko–Cantelli Theorem. Therefore, $\sup_t |\tilde{D}_g(t) - D_g(t)| \xrightarrow{\text{a.s.}} 0$.

Proof of Theorem 2

Theorem 4 generalizes this theorem. See the proof of Theorem 4.

Proof of Theorem 3

Under the conditions that $U_g(t) = U(t) = t$ and $F_g(t) = F(t)$ for all g, in the proof of Lemma 2, we show that $G(\mathbf{a}, t) \geq C(t/\pi_0)$ for all $0 \leq t \leq \min_g a_g$. Since $G(\mathbf{a}, \pi_0) \leq \pi_0 = t/C(t/\pi_0)|_{t=\pi_0}$ and both $G(\mathbf{a}, t)$ and $t/C(t/\pi_0)$ are increasing, we have $G(\mathbf{a}, t) \geq C(t/\pi_0)$ for all $0 \leq t \leq \max_g a_g$. Deduce $B(\mathbf{a}, t) \leq t/C(t/\pi_0)$ for $t \in c(\mathbf{a})$. Therefore $t^*_{\text{BH}} \leq t^*_{\text{GBH}}$. Conditions $\lim_{t\downarrow 0} t/C(t/\pi) \leq \alpha$ and $\pi_0 \geq \alpha$ guarantee that $t^*_{\text{BH}} > 0$.

Note that both $G(\mathbf{a}, t)$ and $t/C(t/\pi_0)$ are continuous, we have

$$\frac{t^*_{\text{BH}}}{C(t^*_{BH}/\pi_0)} = \frac{t^*_{\text{GBH}}}{G(\mathbf{a}, t^*_{\text{GBH}})}.$$

Since $t^*_{\text{BH}} \leq t^*_{\text{GBH}}$, deduce $C(t^*_{\text{BH}}/\pi_0) \leq G(\mathbf{a}, t^*_{\text{GBH}})$. For the adaptive BH procedure, the number of rejections $R_{\text{BH}}(\hat{T}_{\text{BH}}) = \Sigma_{i\in I_N}\{P_i \leq \hat{T}_{\text{BH}}/\hat{\pi}_0\} = N \cdot C_N(\hat{T}_{\text{BH}}/\hat{\pi}_0)$, and

$$|C_N(\widehat{T}_{\text{BH}}/\widehat{\pi}_0) - C(t^*_{\text{BH}}/\pi_0)| \leq \sup_{t\geq 0}|C_N(t/\widehat{\pi}_0) - C(t/\widehat{\pi}_0)| + |C(\widehat{T}_{\text{BH}}/\widehat{\pi}_0) - C(t^*_{\text{BH}}/\pi_0)|,$$

where $\sup_{t\geq 0}|C_N(t/\widehat{\pi}_0) - C(t/\widehat{\pi}_0)| \xrightarrow{\text{a.s.}} 0$ by Glivenko–Cantelli Theorem and $|C(\widehat{T}_{\text{BH}}/\widehat{\pi}_0) - C(t^*_{\text{BH}}/\pi_0)| \xrightarrow{P} 0$ by continuous mapping theorem. Therefore $C_N(\widehat{T}_{\text{BH}}/\widehat{\pi}_0) \xrightarrow{P} C(t^*_{\text{BH}}/\pi_0)$. Analogously one can show that $G_N(\widehat{\mathbf{a}}, \widehat{T}_{\text{GBH}}) \xrightarrow{P} G(\mathbf{a}, t^*_{\text{GBH}})$. A more generalized argument is shown in the proof of Theorem 4. By dominant convergence theorem we have $N^{-1}\mathbb{E}[R_{\text{BH}}(\widehat{T}_{\text{BH}})] \to C(t^*_{\text{BH}}/\pi_0)$ and $N^{-1}\mathbb{E}[R_{\text{GBH}}(\widehat{T}_{\text{GBH}})] \to G(\mathbf{a}, t^*_{\text{GBH}})$. Therefore $\mathbb{E}[R_{\text{BH}}(\hat{T}_{\text{BH}})]/\mathbb{E}[R_{\text{GBH}}(\hat{T}_{\text{GBH}})] \leq 1 + o(1)$.

Proof of Theorem 4

The proof applies Glivenko–Cantelli Theorem as in Storey, Taylor, and Siegmund (2004) and Genovese, Roeder, and Wasserman (2006). Let $S = c(\hat{\mathbf{a}}) \cup c(\boldsymbol{\rho})$. For any $t \in S$, we have

$$\sup_{t\in S}|G_N(\widehat{\mathbf{a}},t) - G(\rho,t)| \le \sup_{t\ge 0}|G_N(\widehat{\mathbf{a}},t) - G(\widehat{\mathbf{a}},t)| + \sup_{t\ge 0}|G(\widehat{\mathbf{a}},t) - G(\rho,t)|.$$

Note that for $t \ge 0$,

$$\begin{aligned}\sup_t|G_N(\widehat{\mathbf{a}},t) - G(\widehat{\mathbf{a}},t)| &= \frac{1}{N}\sum_{g=1}^{K} n_g \sup_t|\tilde{D}_g(t/\widehat{a}_g) - D_g(t/\widehat{a}_g)| \\ &\le \frac{1}{N}\sum_{g=1}^{K} n_g \sup_t|\tilde{D}_g(t) - D_g(t)| \xrightarrow{\text{a.s.}} 0,\end{aligned} \tag{A.1}$$

where the last step is implied by Lemma 2. On the other hand,

$$\sup_{t\ge 0}|G(\widehat{\mathbf{a}},t) - G(\rho,t)| = \frac{1}{N}\sum_{g=1}^{K} n_g \sup_{t\ge 0}|D_g(t/\widehat{a}_g) - D_g(t/\rho_g)|.$$

Since D_g is continuous on $[0, +\infty)$ and $\lim_{t\to\infty} D_g(t) = 1$ is finite, D_g is uniform continuous. By continuous mapping theorem, $|t/\widehat{a}_g - t/\rho_g| \xrightarrow{P} 0$. Therefore $\sup_{t\in S}|D_g(t/\widehat{a}_g) - D_g(t/\rho_g)| \xrightarrow{P} 0$. So we have $\sup_{t\in S}|G_N(\widehat{\mathbf{a}},t) - G(\rho,t)| \xrightarrow{P} 0$.

Let $B_N(\widehat{\mathbf{a}},t) = \frac{\sum_g n_{g,0} t/\widehat{a}_g}{N \cdot G_N(\widehat{\mathbf{a}},t)}$. According to (2.16) and (3.6), $\hat{T}_{\text{GBH}} = \sup_{t\in c(\hat{\mathbf{a}})}\{t : B_N(\hat{\mathbf{a}}, t) \le \alpha\}$ and $t^*_{\text{GBH}} = \sup_{t\in c(\rho)}\{t : B(\rho,t) \le \alpha\}$, where $B(\rho,t) = \frac{\sum_g \pi_g \pi_{g,0} t/\rho_g}{G(\rho,t)}$. Note that the assumption $\lim_{t\downarrow 0} B(\pi, t) \ne \alpha$ implies $t^*_{\text{GBH}} > 0$.

We first show $\widehat{T}_{\text{GBH}} \xrightarrow{P} t^*_{\text{GBH}}$. For any $\xi > 0$, note that $B(\boldsymbol{\rho}, t)$ is increasing for $t \ge \max_g \rho_g$, therefore $B(\rho, t^*_{\text{GBH}} + \xi) > \alpha$, otherwise it contradicts with t^*_{GBH} being the supremum. Fix $\delta > 0$, for any $\delta' \ge \delta$, let $t' = t^*_{\text{GBH}} + \delta'$. Then

$$\begin{aligned}\inf_{\delta'\ge\delta} B_N(\widehat{\mathbf{a}},t') &= \inf_{\delta'\ge\delta} \frac{(1/N)\sum_g n_{g,0} t'/\widehat{a}_g}{G_N(\widehat{\mathbf{a}},t')} \\ &\ge \inf_{\delta'\ge\delta} \frac{\sum_g \pi_g \pi_{g,0} t'/\rho_g - \sum_g |n_{g,0}/N\widehat{a}_g - \pi_g\pi_{g,0}/\rho_g| t'}{G(\rho,t') + \sup_{t\in S}|G_N(\widehat{\mathbf{a}},t) - G(\rho,t)|} \\ &\ge \frac{1-\varepsilon_1}{[1/\inf_{\delta'\ge\delta} B(\rho,t')] + \varepsilon_2},\end{aligned}$$

where $\varepsilon_1 = \Sigma_g |n_{g,0}/N\hat{a}_g - \pi_g\pi_{g,0}/\rho_g| / (\Sigma_g \pi_g \pi_{g,0}/\rho_g)$, and $\varepsilon_2 = \sup_{t\in S}|G_N(\hat{\mathbf{a}}, t) - G(\boldsymbol{\rho}, t)| / (\Sigma_g \pi_g \pi_{g,0} t'/\rho_g)$. Since $\varepsilon_2 \xrightarrow{P} 0$ and $\inf_{\delta'\ge\delta} B(\rho, t') > \alpha$, it can be derived that $\Pr(\cap_{\delta'\ge\delta}\{B_N(\hat{\mathbf{a}}, t') > \alpha\}) \to 1$ which implies $\Pr(\widehat{T}_{\text{GBH}} < t^*_{\text{GBH}} + \delta) \to 1$.

On the other hand, since $B(\boldsymbol{\rho}, t)$ has a nonzero derivative at t^*_{GBH}, it must be positive, otherwise t^*_{GBH} cannot be the supremum of all t such that $B(\boldsymbol{\rho}, t) \le \alpha$. Thus, $t \mapsto B(\boldsymbol{\rho}, t)$ is an increasing function and for any $\xi > 0$, $B(\rho, t^*_{\mathrm{GBH}} - \xi) < \alpha$. For any $\delta > 0$, let $t^\circ = t^*_{\mathrm{GBH}} - \delta$,

$$\begin{aligned} B_N(\widehat{\mathbf{a}}, t^\circ) &= \frac{(1/N)\sum_g n_{g,0} t^\circ / \widehat{a}_g}{G_N(\widehat{\mathbf{a}}, t^\circ)} \\ &\le \frac{\sum_g \pi_g \pi_{g,0} t^\circ / \rho_g + \sum_g |n_{g,0}/N \widehat{a}_g - \pi_g \pi_{g,0}/\rho_g| t^\circ}{G(\rho, t^\circ) - \sup_{t \in S} |G_N(\widehat{\mathbf{a}}, t) - G(\rho, t)|} \\ &\le \frac{1+\varepsilon_1}{[1/B(\rho, t^\circ)] - \varepsilon_2}, \end{aligned}$$

where $\varepsilon_2 \xrightarrow{P} 0$ and $B(\rho, t^\circ) < \alpha$. Then $\Pr(B_N(\hat{\mathbf{a}}, t^\circ) < \alpha) \to 1$. Deduce $\Pr(\widehat{T}_{\mathrm{GBH}} > t^*_{\mathrm{GBH}} - \delta) \to 1$. Combine this and previous result we get $\widehat{T}_{\mathrm{GBH}} \xrightarrow{P} t^*_{\mathrm{GBH}}$.

Next, we prove $\mathrm{FDR}(\hat{T}_{\mathrm{GBH}}) \le \alpha / \min_g\{b_g\} + o(1)$. Let

$$H_N(\widehat{\mathbf{a}}, t) = \frac{1}{N} \sum_{g=1}^{K} \sum_{i \in I_{g,0}} \{P_i \le t / \widehat{a}_g\}$$

be the empirical distribution of p-values under null hypothesis for adaptive GBH procedure. Note that $\widehat{T}_{\mathrm{GBH}} \xrightarrow{P} t^*_{\mathrm{GBH}}$ implies $\Pr(\widehat{T}_{\mathrm{GBH}} > t^*_{\mathrm{GBH}} - \delta) \to 1$ for any $\delta > 0$. Since $t^*_{\mathrm{GBH}} > 0$, deduce $\Pr(\hat{T}_{\mathrm{GBH}} > 0) \to 1$. On the other hand, the assumption $\hat{\zeta} < 1$ rules out the situation where $\hat{T}_{\mathrm{GBH}} / \hat{a}_g \to 0$ for all groups. Therefore $\Pr(\Sigma_g \Sigma_{i \in I_g} \{P_i \le \hat{T}_{\mathrm{GBH}} / \hat{a}_g\} \ge 1) \to 1$. Then the false discovery proportion (FDP) is

$$\mathrm{FDP}(\widehat{T}_{\mathrm{GBH}}) = \frac{N^{-1} \sum_{g=1}^{K} \sum_{i \in I_{g0}} \{P_i \le \widehat{T}_{\mathrm{GBH}} / \widehat{a}_g\}}{N^{-1} \sum_{g=1}^{K} \sum_{i \in I_g} \{P_i \le \widehat{T}_{\mathrm{GBH}} / \widehat{a}_g\}} = \frac{H_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}})}{G_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}})},$$

where $H_N(\hat{\mathbf{a}}, \hat{T}_{\mathrm{GBH}})$ satisfies

$$\begin{aligned} &\left| H_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - \tfrac{1}{N} \sum_{g=1}^{K} n_{g,0} U_g(\widehat{T}_{\mathrm{GBH}} / \widehat{a}_g) \right| \\ &\le \tfrac{1}{N} \sum_{g=1}^{K} n_{g,0} \sup_{t \in c(\widehat{\mathbf{a}})} \left| \tfrac{1}{n_{g,0}} \sum_{i \in I_{g,0}} \{P_i \le t / \widehat{a}_g\} - U_g(t / \widehat{a}_g) \right| \\ &\le \tfrac{1}{N} \sum_{g=1}^{K} n_{g,0} \sup_{t \ge 0} \left| \tfrac{1}{n_{g,0}} \sum_{i \in I_{g,0}} \{P_i \le t\} - U_g(t) \right|. \end{aligned}$$

By Condition (3.2), Glivenko–Cantelli Theorem implies $\sup_{t \ge 0} |\frac{1}{n_{g,0}} \times \sum_{i \in I_{g,0}} \{P_i \le t\} - U_g(t)| \xrightarrow{\text{a.s.}} 0$. Therefore,

$$\left| H_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - \frac{1}{N}\sum_{g=1}^{K} n_{g,0} U_g(\widehat{T}_{\mathrm{GBH}}/\widehat{a}_g) \right| \xrightarrow{\text{a.s.}} 0. \tag{A.2}$$

Now that $\widehat{T}_{\mathrm{GBH}} \xrightarrow{P} t^*_{\mathrm{GBH}}$ and by (3.1),

$$\frac{1}{N}\sum_{g=1}^{K} n_{g,0} U_g(\widehat{T}_{\mathrm{GBH}}/\widehat{a}_g) \xrightarrow{P} \sum_{g=1}^{K} \pi_g \pi_{g,0} U_g(t^*_{\mathrm{GBH}}/\rho_g). \tag{A.3}$$

Combine (A.2) and (A.3) we have

$$H_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) \xrightarrow{P} \sum_{g=1}^{K} \pi_g \pi_{g,0} U_g(t^*_{\mathrm{GBH}}/\rho_g). \tag{A.4}$$

On the other hand,

$$\begin{aligned} &|G_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - G(\rho, t^*_{\mathrm{GBH}})| \\ \le\ &|G_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - G(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}})| + |G(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - G(\rho, t^*_{\mathrm{GBH}})| \\ \le\ &\sup_{t\ge 0} |G_N(\widehat{\mathbf{a}}, t) - G(\widehat{\mathbf{a}}, t)| + |G(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - G(\rho, t^*_{\mathrm{GBH}})|, \end{aligned}$$

where $\sup_{t\ge 0}|G_N(\widehat{\mathbf{a}}, t) - G(\widehat{\mathbf{a}}, t)| \xrightarrow{\text{a.s.}} 0$ by (A.1) and $|G(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) - G(\rho, t^*_{\mathrm{GBH}})| \xrightarrow{P} 0$ by continuous mapping theorem. Therefore,

$$G_N(\widehat{\mathbf{a}}, \widehat{T}_{\mathrm{GBH}}) \xrightarrow{P} G(\rho, t^*_{\mathrm{GBH}}). \tag{A.5}$$

Since $t^*_{\mathrm{GBH}} > 0$ and $\bar{\zeta} < 1$, we have $G(\rho, t^*_{\mathrm{GBH}}) > 0$. By (A.4) and (A.5),

$$\mathrm{FDP}(\widehat{T}_{\mathrm{GBH}}) \xrightarrow{P} \frac{\sum_{g=1}^{K} \pi_g \pi_{g,0} U_g(t^*_{\mathrm{GBH}}/\rho_g)}{G(\rho, t^*_{\mathrm{GBH}})}.$$

By dominated convergence theorem,

$$\mathrm{FDR}(\widehat{T}_{\mathrm{GBH}}) = E[\mathrm{FDP}(\widehat{T}_{\mathrm{GBH}})] \to \frac{\sum_{g=1}^{K} \pi_g \pi_{g,0} U_g(t^*_{\mathrm{GBH}}/\rho_g)}{G(\rho, t^*_{\mathrm{GBH}})}. \tag{A.6}$$

Note that $\zeta_g \ge b_g \pi_{g,0}$ for some $b_g > 0$. Deduce $\rho_g \ge b_g \pi_{g,0}/(1 - \zeta_g)$. Since $U_g(t) \le t$ for all $t \ge 0$, we have

$$\frac{\sum_{g=1}^{K}\pi_g\pi_{g,0}U(t^*_{\mathrm{GBH}}/\rho_g)}{G(\rho,t^*_{\mathrm{GBH}})} \le \frac{1}{\min_g\{b_g\}}\frac{(1-\bar{\zeta})t^*_{\mathrm{GBH}}}{G(\rho,t^*_{\mathrm{GBH}})} \le \frac{\alpha}{\min_g\{b_g\}}.$$

Hence, $\mathrm{FDR}(\hat{T}_{\mathrm{GBH}}) \le \alpha/\min_g\{b_g\} + o(1)$.

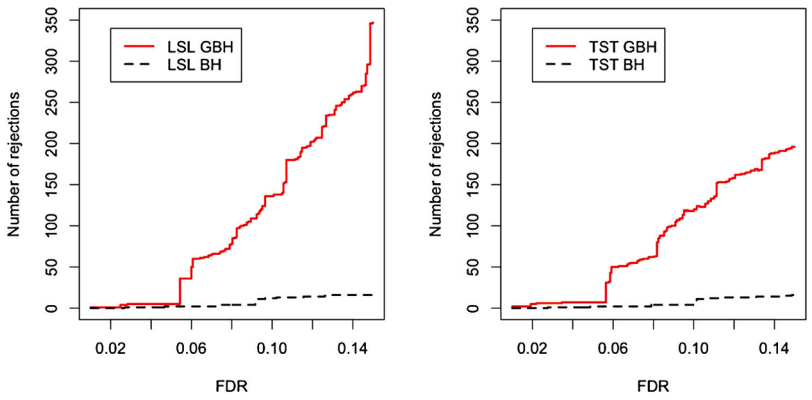

Figure 1.

Breast cancer study, 24,184 genes. Plot shows the number of signals detected by GBH and BH procedures versus prespecified FDR level. Left panel: adaptive LSL GBH procedure. Right panel: adaptive TST GBH procedure. Details for LSL and TST approaches are in Section 2.3. Genes are assigned into six groups using *k*-means clustering. Data from van't Veer et al. (2002). The online version of this figure is in color.

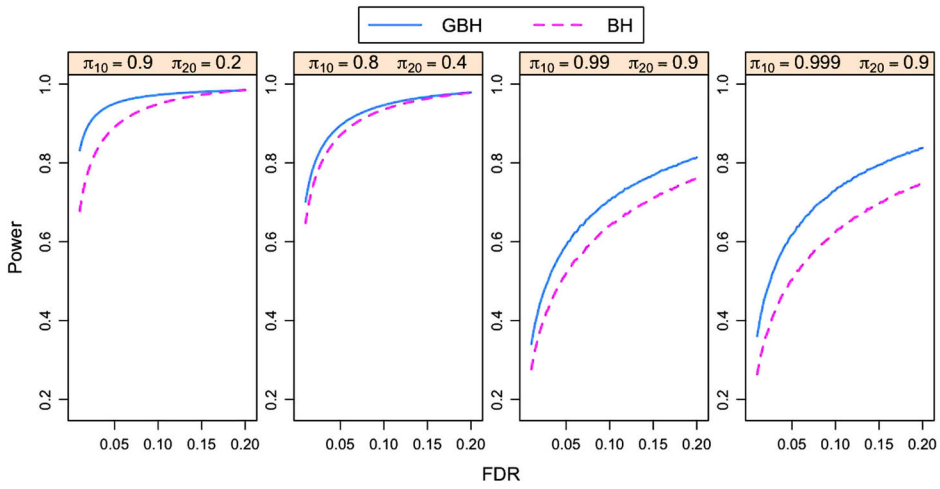

Figure 2.

Power curves of the oracle BH and GBH procedures for independent *p*-values. The *p*-values are generated based on model (4.1) with $\xi_1 = \xi_2 = 0$ and $n = 10{,}000$ for each group. Each panel corresponds to one combination of $\pi_{g,0}$'s for two groups. The online version of this figure is in color.

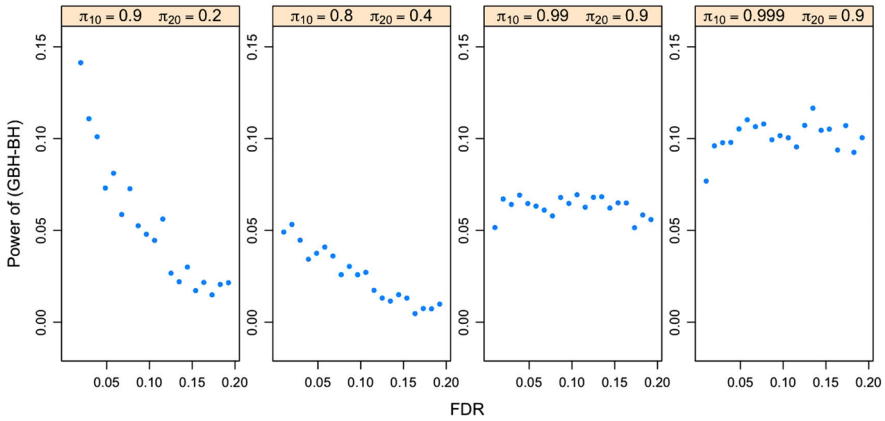

Figure 3.

Power differences of the oracle BH and GBH procedures for *p*-values with the PRDS property. The *p*-values are generated based on model (4.1) with $\xi_1 = \xi_2 = 0.5$ and $n = 10{,}000$ for each group. Each panel corresponds to one combination of $\pi_{g,0}$'s for two groups. The online version of this figure is in color.

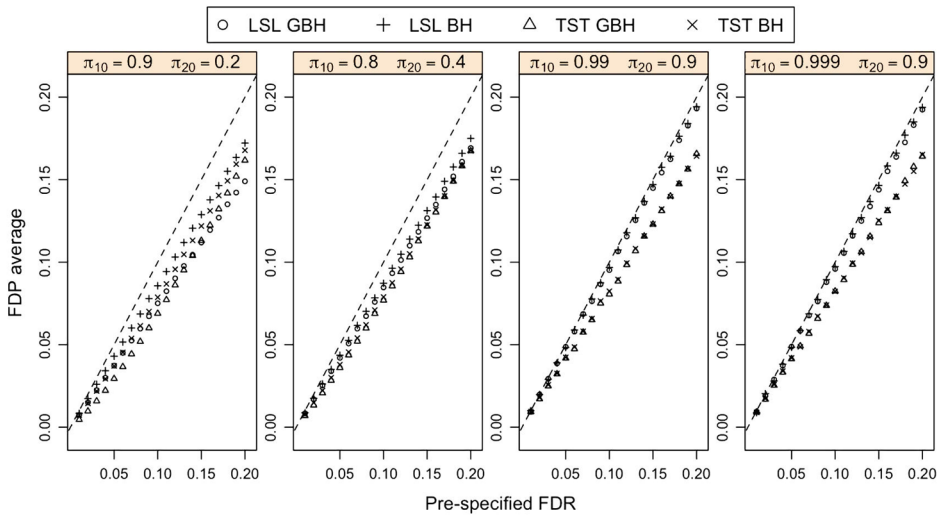

Figure 4.

Comparison of the average FDP and the prespecified FDR for the adaptive BH and GBH procedures. The dash line is the 45-degree line. The *p*-values are generated based on model (4.1) with $\xi_1 = \xi_2 = 0$ and $n = 10{,}000$ for each group. Each point of the FDP is the average of 200 iterations. The online version of this figure is in color.

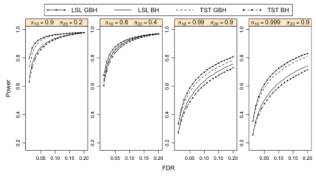

Figure 5.

Power curves of the adaptive BH and GBH procedures for independent *p*-values. The *p*-values are generated based on model (4.1) with $\xi_1 = \xi_2 = 0$ and $n = 10{,}000$ for each group. Each panel corresponds to one combination of $\pi_{g,0}$'s for two groups. The online version of this figure is in color.

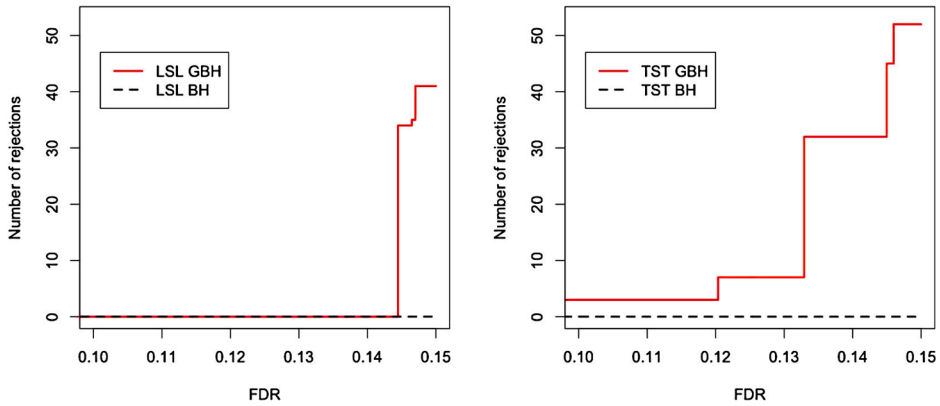

Figure 6.

Breast cancer study, 9492 genes. The plots show the number of genes detected by the adaptive BH and GBH procedures using Gene Ontology grouping. Left panel: the LSL method. Right panel: the TST method. Data from van't Veer et al. (2002). The online version of this figure is in color.

Table 1

Part of the breast cancer dataset in van't Veer et al. (2002)

	Developed cancer in 5 years				Cancer-free in 5 years				
Gene ID	**patient 1**	**patient 2**	**…**	**patient 44**	**patient 45**	**patient 46**	**…**	**patient 78**	***p*-value**
AA000990	0.080	0.130	…	0.136	−0.513	−0.098	…	−0.015	0.7937
AA001113	−0.159	−0.087	…	−0.116	0.190	−0.204	…	0.082	0.4897
AA001360	−0.018	−0.024	…	−0.255	0.114	−0.042	…	0.200	0.1224
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

NOTE: The entries are adjusted $\log_{10}$(red/green) ratios from cDNA microarrays. The *p*-values were calculated based on a two-sample *t*-test for each gene.

Table 2

Comparison of the adaptive LSL GBH and the adaptive LSL BH procedures for GO grouping. FDR level = 0.15

Group	# of genes	$\widehat{\pi}_{g,0}^{\mathrm{LSL}}$	LSL BH	LSL GBH
(1) Cell communication	593	0.995	0	3
(2) Cell growth/maintenance	4142	0.987	0	13
(3) Development	434	0.989	0	0
(4) Multifunction	4323	0.983	0	25
Total	9492		0	41

Table 3

Comparison of the adaptive LSL GBH and the adaptive LSL BH procedures for *k*-means grouping. FDR level = 0.1

Cluster	# of genes	$\widehat{\pi}_{g,0}^{\mathrm{LSL}}$	LSL BH	LSL GBH
1	1904	0.871	4	109
2	214	0.991	0	0
3	1368	0.999	0	0
4	2458	0.969	1	19
5	7058	0.999	4	2
6	11,164	0.996	3	6
Total	24,184		12	136