



# HHS Public Access

Author manuscript

*Biometrika*. Author manuscript; available in PMC 2017 May 11.

Published in final edited form as:

*Biometrika*. 2015 June ; 102(2): 247–266. doi:10.1093/biomet/asu074.

## Testing Differential Networks with Applications to Detecting Gene-by-Gene Interactions

**Yin Xia,**

Department of Statistics & Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA

**Tianxi Cai,** and

Department of Biostatistics, Harvard School of Public Health, Harvard University, Boston, Massachusetts 02115, USA

**T. Tony Cai**

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

### Summary

Model organisms and human studies have led to increasing empirical evidence that interactions among genes contribute broadly to genetic variation of complex traits. In the presence of gene-by-gene interactions, the dimensionality of the feature space becomes extremely high relative to the sample size. This imposes a significant methodological challenge in identifying gene-by-gene interactions. In the present paper, through a Gaussian graphical model framework, we translate the problem of identifying gene-by-gene interactions associated with a binary trait  $D$  into an inference problem on the difference of two high-dimensional precision matrices, which summarize the conditional dependence network structures of the genes. We propose a procedure for testing the differential network globally that is particularly powerful against sparse alternatives. In addition, a multiple testing procedure with false discovery rate control is developed to infer the specific structure of the differential network. Theoretical justification is provided to ensure the validity of the proposed tests and optimality results are derived under sparsity assumptions. A simulation study demonstrates that the proposed tests maintain the desired error rates under the null and have good power under the alternative. The methods are applied to a breast cancer gene expression study.

### Keywords

Differential network; false discovery rate; Gaussian graphical model; gene-by-gene interaction; highdimensional precision matrix; large scale multiple testing

---

#### Supplementary Material

Supplementary material available at *Biometrika* online includes more extensive simulation results comparing the numerical performance of the proposed global test with that of other tests, the proofs of Lemmas 2, 3 and 4, and the Matlab code for numerical implementation.

## 1. INTRODUCTION

High throughput technologies, enabling comprehensive monitoring of a biological system, have fundamentally transformed biomedical research. Studies using such technologies have led to successful molecular classifications of diseases into clinically relevant subtypes and genetic signatures predictive of disease progression and treatment response (van't Veer et al., 2002; Gregg et al., 2008; Hu et al., 2009, e.g.). Irrespective of the technology used, analysis of high-throughput data typically considers one marker at a time and yields a list of differentially expressed genes or proteins. On the other hand, epistasis, or interactions between genes, has long been recognized as crucial to understanding the genetic architecture of disease phenotypes (Phillips, 2008; Eichler et al., 2010). Increasing empirical evidence from model organisms and human studies suggests that gene-by-gene interactions may make an important contribution to total genetic variation of complex traits (Zerba et al., 2000; Marchini et al., 2005). In this paper, we are specifically interested in gene-by-gene interactions with respect to the interactive effects of two genes on a binary disease trait  $D$ .

In the presence of gene-by-gene interactions, the dimensionality of the feature space becomes extremely high relative to the sample size. This, together with the variability of the data, imposes a significant methodological challenge in identifying gene-by-gene interactions using currently available studies, which typically have limited sample sizes and power. Recent development in interaction modeling has led to several useful methods including multi-factor dimensionality reduction (Ritchie et al., 2001; Moore, 2004), polymorphism interaction analysis (Mechanic et al., 2008), random forests (Breiman, 2001), various variations of logistic regression with interactive effects (Chatterjee et al., 2006; Chapman & Clayton, 2007; Kooperberg & Ruczinski, 2005; Kooperberg & LeBlanc, 2008) and sure independence screening (Fan & Lv, 2008). However, to overcome the high dimensionality, a majority of these methods use multistage procedures and marginal assessments of the effects of a gene pair without simultaneously accounting for the effects of other genes. Multistage procedures may have limited power in detecting genes that affect the outcome through interactions with other genes without strong main effects. The interactive effects detected through models that only consider one pair of genes at a time without conditioning on other genes may also result in false identification of interactions due to the discrepancy between conditional and unconditional effects. Furthermore, none of the existing methods provide false discovery rate control in the presence of interactions. Due to the large number of tests, the power of multiple testing procedures using the standard Bonferroni or naive false discovery rate corrections can dissipate quickly.

In this paper, through a Gaussian graphical model framework, we translate the problem of identifying gene-by-gene interactions associated with a binary trait  $D$  into the comparison of two high-dimensional precision matrices. Let  $G$  denote a  $p \times 1$  vector of genomic markers and assume that, conditional on  $D = d$ ,  $G \sim \mathcal{N}(\mu_d, \Sigma_d)$ , for  $d = 1, 2$ . Then the posterior risk given  $G$  is

$$\text{pr}(D=1|G) = g \left\{ \text{constant} - \frac{1}{2} G^T (\Omega_1 - \Omega_2) G + G^T (\Omega_1 \mu_1 - \Omega_2 \mu_2) \right\}$$

where  $g(x) = e^x/(1 + e^x)$  and  $\Omega_d = (\omega_{i,j,d}) = \sum_d^{-1}$  is the precision matrix for  $G$  conditional on  $D = d$ . Hence, an interaction between the gene pair  $(i, j)$  affects the disease risk if and only if  $\delta_{i,j} = \omega_{i,j,1} - \omega_{i,j,2} = 0$ . The difference between the two precision matrices, denoted by  $\Delta = (\delta_{i,j}) = \Omega_1 - \Omega_2$ , is called the differential network. This type of model for a differential network has been used in Li et al. (2007) and Danaher et al. (2014). We thus propose to test for gene-by-gene interactions both by testing the global hypotheses

$$H_0: \Delta = 0 \text{ versus } H_1: \Delta \neq 0, \quad (1)$$

and by simultaneously testing the hypotheses

$$H_{0,i,j}: \delta_{i,j} = 0 \text{ versus } H_{1,i,j}: \delta_{i,j} \neq 0, \quad 1 \leq i < j \leq p,$$

while controlling for the overall false discovery rate at a pre-specified level.

Few authors have considered testing the equality of two precision matrices in the high-dimensional setting. The global null hypothesis  $\Delta = 0$ , or equivalently  $\Omega_1 = \Omega_2$ , corresponds to the hypothesis that none of the gene pairs have interactive effects on  $D$ . The equality of two precision matrices is equivalent to the equality of two covariance matrices, and the latter has been studied under various alternatives. Under the dense alternative, where  $\Sigma_1$  and  $\Sigma_2$  differ in a large number of entries, various sum-of-square type testing procedures have been proposed (Schott, 2007; Srivastava & Yanagihara, 2010; Li & Chen, 2012). Under the sparse alternative with  $\Sigma_1$  and  $\Sigma_2$  differing only in a small number of entries, Cai et al. (2013) introduced a particularly powerful test. However, in the gene-by-gene interaction setting, the goal is to identify the structure of the differential network. In such cases, it is often reasonable to assume that  $\Delta$  is sparse, while  $\Sigma_1 - \Sigma_2$  is not. Hence, testing procedures that can leverage information on the sparsity of  $\Delta$  may improve power. Furthermore, due to the fundamental difference between conditional and unconditional dependences, the various procedures for testing the covariance matrices may not be well adapted to testing specific entries of the precision matrices.

The first goal of this paper is to develop a global test for  $H_0: \Delta = 0$  that is powerful against sparse alternatives. We then develop a multiple testing procedure for simultaneously testing the hypotheses  $\{H_{0,i,j}: 1 \leq i < j \leq p\}$  with false discovery rate control to infer the structure of the differential network. In the high-dimensional setting, there is no sample precision matrix that one can use to approximate  $\Omega_d$ . We propose to infer  $\Omega_d$  by relating its elements to the coefficients of a set of regression models for  $G$  conditional on  $D = d$ . We then construct test statistics based on the covariances between the residuals from the fitted regression models. The testing procedures are easy to implement. A Matlab implementation is available in the Supplementary Material.

## 2. Global Testing of Differential Networks

### 2.1. Notation and Definitions

In this section we consider testing the global hypothesis (1). We begin with notation and definitions that will be used in the rest of the paper. Let  $X_k \in \mathbb{R}^p$  and  $Y_k \in \mathbb{R}^p$  denote  $G$  given  $D=1$  and  $D=2$ , respectively,  $X_k \sim \mathcal{N}(\mu_1, \Sigma_1)$  for  $k=1, \dots, n_1$ ,  $Y_k \sim \mathcal{N}(\mu_1, \Sigma_2)$  for  $k=1, \dots, n_2$ , where  $\Sigma_d = (\sigma_{i,j,d})$  for  $d=1, 2$ , and  $\{X_k : k=1, \dots, n_1\}$  and  $\{Y_k : k=1, \dots, n_2\}$  are independent observations from the two populations. Let  $X = (X_1, \dots, X_{n_1})^T$  and  $Y = (Y_1, \dots, Y_{n_2})^T$  denote the data matrices. Let  $\Omega_d = (\omega_{i,j,d}) = \sum_d^{-1}$ , for  $d=1, 2$ .

For subscripts, we use the convention that  $i$  stands for the  $i^{\text{th}}$  entry of a vector and  $(i,j)$  for the entry in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of a matrix,  $k$  represents the  $k^{\text{th}}$  sample and  $d$  indexes the binary trait. Let  $\beta_{i,1} = (\beta_{1,i,1}, \dots, \beta_{p-1,i,1})^T$  denote the regression coefficients of  $X_{k,i}$  regressed on the rest of the entries of  $X_k$  and let  $\beta_{i,2} = (\beta_{1,i,2}, \dots, \beta_{p-1,i,2})^T$  denote the regression coefficients of  $Y_{k,i}$  regressed on the rest of the entries of  $Y_k$ .

For any vector  $\mu_d$  with dimension  $p \times 1$ , let  $\mu_{-i,d}$  denote the  $(p-1) \times 1$  vector by removing the  $i^{\text{th}}$  entry from  $\mu_d$ . For a symmetric matrix  $A$ , let  $\lambda_{\max}(A)$  and  $\lambda_{\min}(A)$  denote the largest and smallest eigenvalues of  $A$ . For any  $p \times q$  matrix  $A$ ,  $A_{i,-j}$  denotes the  $i^{\text{th}}$  row of  $A$  with its  $j^{\text{th}}$  entry removed and  $A_{-i,j}$  denotes the  $j^{\text{th}}$  column of  $A$  with its  $i^{\text{th}}$  entry removed. The matrix  $A_{-i,-j}$  denotes a  $(p-1) \times (q-1)$  matrix obtained by removing the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $A$ . For an  $n \times p$  data matrix  $U = (U_1, \dots, U_n)^T$ , let  $U_{\cdot,-i} = (U_{1,-i}^T, \dots, U_{n,-i}^T)^T$  with dimension  $n \times (p-1)$ ,  $\bar{U}_{\cdot,-i} = n^{-1} \sum_{k=1}^n U_{k,-i}$  with dimension  $1 \times (p-1)$ ,  $U_{(i)} = (U_{1,i}, \dots, U_{n,i})^T$  with dimension  $n \times 1$ ,  $\bar{U}_{(i)} = (\bar{U}_i, \dots, \bar{U}_i)^T$  with dimension  $n \times 1$ , where  $\bar{U}_i = n^{-1} \sum_{k=1}^n U_{k,i}$ , and  $\bar{U}_{(\cdot,-i)} = (\bar{U}_{\cdot,-i}^T, \dots, \bar{U}_{\cdot,-i}^T)^T$  with dimension  $n \times (p-1)$ . For tuning parameters  $\lambda$ , let  $\lambda_{n_d,i,d}$  represent the  $i^{\text{th}}$  tuning parameter for binary trait  $d$ , which depends on the sample size  $n_d$ .

For a vector  $\beta = (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p$ , define the  $\ell_q$  norm by  $|\beta|_q = (\sum_{i=1}^p |\beta_i|^q)^{1/q}$  for  $1 \leq q < \infty$ . A vector  $\beta$  is called  $k$ -sparse if it has at most  $k$  nonzero entries. For a matrix  $\Omega = (\omega_{i,j})_{p \times p}$ , the matrix 1-norm is the maximum absolute column sum,  $\|\Omega\|_{L_1} = \max_{1 \leq j \leq p} \sum_{i=1}^p |\omega_{i,j}|$ , the matrix elementwise infinity norm is defined to be  $\|\Omega\|_{\infty} = \max_{1 \leq i,j \leq p} |\omega_{i,j}|$  and the elementwise  $\ell_1$  norm is  $\|\Omega\|_1 = \sum_{i=1}^p \sum_{j=1}^p |\omega_{i,j}|$ . For a matrix  $\Omega$ , we say  $\Omega$  is  $k$ -sparse if each row/column has at most  $k$  nonzero entries. For a set  $\mathcal{H}$ , denote by  $|\mathcal{H}|$  the cardinality of  $\mathcal{H}$ . For two sequences of real numbers  $\{a_n\}$  and  $\{b_n\}$ , write  $a_n = O(b_n)$  if there exists a constant  $C$  such that  $|a_n| \leq C|b_n|$  holds for all  $n$ , write  $a_n = o(b_n)$  if  $\lim_{n \rightarrow \infty} a_n/b_n = 0$ , and write  $a_n \asymp b_n$  if there are positive constants  $c$  and  $C$  such that  $c|a_n/b_n| \leq C$  for all  $n$ .

### 2.2. Testing Procedure

It is well known (e.g., Anderson, 2003, Section 2.5), that in the Gaussian setting the precision matrix can be described in terms of regression models. Specifically, we may write

$$X_{k,i} = \alpha_{i,1} + X_{k,-i}\beta_{i,1} + \varepsilon_{k,i,1}, \quad (i=1, \dots, p; \quad k=1, \dots, n_1), \quad (2)$$

$$Y_{k,i} = \alpha_{i,2} + Y_{k,-i}\beta_{i,2} + \varepsilon_{k,i,2}, \quad (i=1, \dots, p; \quad k=1, \dots, n_2), \quad (3)$$

where  $\varepsilon_{k,i,d} \sim N(0, \sigma_{i,i,d} - \sum_{i,-i,d} \sum_{-i,-i,d}^{-1} \sum_{-i,i,d}) (d=1, 2)$ , are independent of  $X_{k,-i}$  and  $Y_{k,-i}$  respectively, and  $\alpha_{i,d} = \mu_{i,d} - \sum_{i,-i,d} \sum_{-i,-i,d}^{-1} \mu_{-i,d}$ . The regression coefficient vectors  $\beta_{i,d}$  and the error terms  $\varepsilon_{k,i,d}$  satisfy

$$\beta_{i,d} = -\omega_{i,i,d}^{-1} \Omega_{-i,i,d}, \quad r_{i,j,d} = \text{cov}(\varepsilon_{k,i,d}, \varepsilon_{k,j,d}) = \frac{\omega_{i,j,d}}{\omega_{i,i,d} \omega_{j,j,d}},$$

where  $\text{cov}(\cdot, \cdot)$  denotes the population covariance. Since the null hypothesis  $H_0: \beta_{i,d} = 0$  is equivalent to the hypothesis

$$H_0: \max_{1 \leq i \leq j \leq p} |\omega_{i,j,1} - \omega_{i,j,2}| = 0,$$

a natural approach to test  $H_0$  is to first construct estimators of  $\omega_{i,j,d}$  and then base the test on the maximum standardized differences. We first construct estimators of  $r_{i,j,d}$

Let  $\hat{\beta}_{i,d} = (\hat{\beta}_{1,i,d}, \dots, \hat{\beta}_{p-1,i,d})^T$  be estimators of  $\beta_{i,d}$  satisfying

$$\max_{1 \leq i \leq p} |\hat{\beta}_{i,d} - \beta_{i,d}|_1 = o_p\{(\log p)^{-1}\}, \quad (4)$$

$$\max_{1 \leq i \leq p} |\hat{\beta}_{i,d} - \beta_{i,d}|_2 = o_p\{(n_d \log p)^{-1/4}\}. \quad (5)$$

Estimators  $\hat{\beta}_{i,d}$  that satisfy (4) and (5) can be obtained easily via methods such as the lasso and Dantzig selector. See Section 2.3 for details. Define the residuals by

$$\hat{\varepsilon}_{k,i,1} = X_{k,i} - \bar{X}_i - (X_{k,-i} - \bar{X}_{\cdot,-i}) \hat{\beta}_{i,1}, \quad \hat{\varepsilon}_{k,i,2} = Y_{k,i} - \bar{Y}_i - (Y_{k,-i} - \bar{Y}_{\cdot,-i}) \hat{\beta}_{i,2}.$$

A natural estimator of  $r_{i,j,d}$  is the sample covariance between the residuals,

$$\tilde{r}_{i,j,d} = \frac{1}{n_d} \sum_{k=1}^{n_d} \hat{\varepsilon}_{k,i,d} \hat{\varepsilon}_{k,j,d}. \quad (6)$$

However, when  $i \neq j$ ,  $\tilde{r}_{i,j,d}$  tends to be biased due to the correlation induced by the estimated parameters and it is desirable to construct a bias-corrected estimator. Lemma 2 shows that

$$\tilde{r}_{i,j,d} = \tilde{R}_{i,j,d} - \tilde{r}_{i,i,d}(\hat{\beta}_{i,j,d} - \beta_{i,j,d}) - \tilde{r}_{j,j,d}(\hat{\beta}_{j-1,i,d} - \beta_{j-1,i,d}) + o_p\{(n_d \log p)^{-\frac{1}{2}}\}$$

where  $\tilde{R}_{i,j,d}$  is the empirical covariance between  $\{\varepsilon_{k,i,d}: k = 1, \dots, n_d\}$  and  $\{\varepsilon_{k,j,d}: k = 1, \dots, n_d\}$ . For  $1 \leq i < j \leq p$ ,  $\beta_{i,j,d} = -\omega_{i,j,d}/\omega_{j,j,d}$  and  $\beta_{j-1,i,d} = -\omega_{i,j,d}/\omega_{i,i,d}$ . Thus, we propose a bias-corrected estimator of  $r_{i,j,d}$  as

$$\hat{r}_{i,j,d} = -(\tilde{r}_{i,j,d} + \tilde{r}_{i,i,d}\hat{\beta}_{i,j,d} + \tilde{r}_{j,j,d}\hat{\beta}_{j-1,i,d}), \quad 1 \leq i < j \leq p. \quad (7)$$

The bias of  $\hat{r}_{i,j,d}$  is of order  $\max\{r_{i,j,d}(\log p/n_d)^{1/2}, (n_d \log p)^{-1/2}\}$ .

For  $i = j$ , note that  $r_{i,i,d} = 1/u_{i,i,d}$ . We show in Lemma 2 that

$$\max_{1 \leq i \leq p} |\tilde{r}_{i,i,d} - r_{i,i,d}| = O_p\{(\log p/n_d)^{1/2}\},$$

which implies that  $\hat{r}_{i,i,d} = \tilde{r}_{i,i,d}$  is a nearly unbiased estimator of  $r_{i,i,d}$ . A natural estimator of  $\omega_{i,j,d}$  can then be defined by

$$T_{i,j,d} = \frac{\hat{r}_{i,j,d}}{\hat{r}_{i,i,d}\hat{r}_{j,j,d}}, \quad 1 \leq i \leq j \leq p \quad (8)$$

We test  $H_0: \theta_{i,j,d} = 0$  based on the estimators  $\mathcal{T} = \{T_{i,j,d}: 1 \leq i \leq j \leq p\}$

The estimators  $T_{i,j,d}$  in  $\mathcal{T}$  are heteroscedastic and possibly have a wide range of variability. We first standardize  $T_{i,j,d}$  before combining information from all entries in

$\mathcal{T}$ . Let  $U_{i,j,d} = (1/n_d) \sum_{k=1}^{n_d} \{\varepsilon_{k,i,d}\varepsilon_{k,j,d} - E(\varepsilon_{k,i,d}\varepsilon_{k,j,d})\}$  and

$\tilde{U}_{i,j,d} = (r_{i,j,d} - U_{i,j,d}) / (r_{i,i,d}r_{j,j,d})$ . It will be shown in Lemma 2 that, uniformly in  $1 \leq i < j \leq p$ ,

$$|T_{i,j,d} - \tilde{U}_{i,j,d}| = O_p\{(\log p/n_d)^{\frac{1}{2}}\} r_{i,j,d} + o_p\{(n_d \log p)^{-\frac{1}{2}}\}.$$

Let  $\theta_{i,j,d} = \text{var}(\tilde{U}_{i,j,d})$ . Note that

$$\theta_{i,j,d} = \text{var}\{\varepsilon_{k,i,d}\varepsilon_{k,j,d} / (r_{i,i,d}r_{j,j,d})\} / n_d = (1 + \rho_{i,j,d}^2) / (n_d r_{i,i,d}r_{j,j,d}),$$

where  $\rho_{i,j,d}^2 = \beta_{i,j,d}^2 r_{i,i,d} / r_{j,j,d}$ . We then estimate  $\theta_{i,j,d}$  by

$$\hat{\theta}_{i,j,d} = (1 + \hat{\beta}_{i,j,d}^2 \hat{r}_{i,i,d} / \hat{r}_{j,j,d}) / (n_d \hat{r}_{i,i,d} \hat{r}_{j,j,d}).$$

Define the standardized statistics

$$W_{i,j} = \frac{T_{i,j,1} - T_{i,j,2}}{(\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})^{1/2}}, \quad 1 \leq i \leq j \leq p. \tag{9}$$

Finally, we propose the following test statistic for testing the global null hypothesis  $H_0$ ,

$$M_n = \max_{1 \leq i \leq j \leq p} W_{i,j}^2 = \max_{1 \leq i \leq j \leq p} \frac{(T_{i,j,1} - T_{i,j,2})^2}{\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}}. \tag{10}$$

The asymptotic properties of  $M_n$  will be studied in detail in Section 3. Intuitively,  $\{W_{i,j}\}$  are approximately standard normal variables under the null  $H_0$  and they are only weakly dependent under suitable conditions. Thus  $M_n$  is the maximum of the squares of  $p(p+1)/2$  such random variables, so its value should be close to  $2 \log\{p(p+1)/2\} \approx 4 \log p$  under  $H_0$ . We show in Section 3 that, under certain regularity conditions,  $M_n - 4 \log p - \log \log p$  converges to a type I extreme value distribution under  $H_0$ :  $\tau = 0$ .

Based on the limiting null distribution of  $M_n$ , which will be developed in Section 3.1, we define the test  $\psi_\alpha$  by

$$\Psi_\alpha = I(M_n \geq q_\alpha + 4 \log p - \log \log p) \tag{11}$$

where  $q_\alpha$  is the  $1 - \alpha$  quantile of the type I extreme value distribution with the cumulative distribution function  $\exp\{-(8\pi)^{-1/2} e^{-t/2}\}$ , i.e.,

$$q_\alpha = -\log(8\pi) - 2 \log \log(1 - \alpha)^{-1}. \tag{12}$$

The hypothesis  $H_0$  is rejected whenever  $\Psi_\alpha = 1$ .

### 2.3. Data-driven estimation of regression coefficients

The testing procedure requires the estimation of regression coefficients  $\beta_{i,d}$  for  $i = 1, \dots, p$  and  $d = 1, 2$ . Various estimators have been studied in the literature, including the lasso and Dantzig selector. Here, we use the lasso by solving the optimization problem,

$$\hat{\beta}_{i,1} = D_{i,1}^{-1/2} \arg \min_{u \in \mathbb{R}^{p-1}} \{(2n_1)^{-1} |(X_{\cdot, -i} - \bar{X}_{(\cdot, -i)}) D_{i,1}^{-1/2} u - (X_{(i)} - \bar{X}_{(i)})|_2^2 + \lambda_{n_1, i, 1} |u|_1\}, \tag{13}$$

$$\hat{\beta}_{i,2} = D_{i,2}^{-1/2} \arg \min_{v \in \mathbb{R}^{p-1}} \{ (2n_2)^{-1} |(Y_{\cdot,-i} - \bar{Y}_{(\cdot,-i)}) D_{i,2}^{-1/2} v - (Y_{(i)} - \bar{Y}_{(i)})|_2^2 + \lambda_{n_2,i,2} |v|_1 \}, \quad (14)$$

where  $D_{i,d} = \text{diag}(\sum_{-i,-i,d} \hat{\sigma}_{i,i,d})$  and  $\lambda_{n_d,i,d} = \kappa_d (\hat{\sigma}_{i,i,d} \log p / n_d)^{1/2}$ ,  $d = 1, 2$ . Then by Proposition 4.2 of Liu (2013), under Condition (C1) given in Section 3 and a mild condition on the sparsity of  $\beta_{i,d}$  ( $i = 1, \dots, p$ ,  $d = 1, 2$ ), the convergence rates in (4) and (5) can be guaranteed by using any  $\kappa_d > 2$ . The result is formally stated in Corollary 1. In practice,  $\kappa_d = 2$  works well for global testing of  $H_0 : \omega = 0$ , and for the multiple testing procedure with false discovery rate control, a data-driven algorithm is proposed in Section 5 to select  $\kappa_d$  adaptively.

### 2.4. Discussion

The global test  $\psi_a$  given in (11) is based on estimators of  $\omega_{i,j,1} - \omega_{i,j,2}$ . Here we estimate  $\omega_{i,j,d}$  by first constructing estimators for  $r_{i,j,d} = \omega_{i,j,d} / (\omega_{i,i,d} \omega_{j,j,d})$ , and then estimating  $r_{i,j,d}$  through bias correction of the residuals  $\hat{r}_{i,j,d}$  defined in (7).

Liu (2013) considered multiple testing of entries of a single precision matrix  $\Omega = (\omega_{i,j})$ . In the one-sample case,  $\omega_{i,j} = 0$  is equivalent to  $r_{i,j} = \omega_{i,j} / (\omega_{i,i} \omega_{j,j}) = 0$  under the null and  $r_{i,j}$  is easier to estimate. The procedure in Liu (2013) is based on the estimation of  $r_{i,j}$  instead of  $\omega_{i,j}$ . However, in Section 4 we will also consider multiple testing between two groups, and  $\omega_{i,j,1} - \omega_{i,j,2}$  is not equivalent to  $r_{i,j,1} - r_{i,j,2}$ . Thus, it is necessary to construct testing procedures based directly on estimators of  $\omega_{i,j,1} - \omega_{i,j,2}$ .

Testing the global hypothesis  $H_0 : \Omega_1 = \Omega_2$  is equivalent to testing  $H_0 : \Sigma_1 = \Sigma_2$ , which has been well studied (Schott, 2007; Srivastava & Yanagihara, 2010; Li & Chen, 2012; Cai et al., 2013). In particular, Cai et al. (2013) constructed a global test for  $H_0 : \Sigma_1 = \Sigma_2$  that is powerful against the alternative where  $\Sigma_1 - \Sigma_2$  is sparse. However, in many applications, the goal is to learn the structure of the differential network, and we are interested in both testing the global hypothesis  $H_0 : \Omega_1 = \Omega_2$  and multiple testing of the entrywise hypotheses  $H_{0,i,j} : \omega_{i,j,1} - \omega_{i,j,2}$ . In such cases, it is often reasonable to assume that  $\Omega_1 - \Omega_2$  is sparse, but  $\Sigma_1 - \Sigma_2$  is not. Hence, testing procedures for  $H_0 : \Sigma_1 = \Sigma_2$  cannot leverage information on the sparsity of  $\Omega_1 - \Omega_2$  and more importantly do not naturally lead to a multiple testing procedure for simultaneously testing the entrywise hypotheses  $H_{0,i,j} : \omega_{i,j,1} - \omega_{i,j,2}$ .

## 3. Theoretical Results for the Global Test

### 3.1. Asymptotic Null Distribution of $M_n$

In this section, we analyze the properties of the new test for testing the global null hypothesis  $H_0 : \omega = 0$ , including the null distribution of the test statistic  $M_n$ , the asymptotic size and power. We are particularly interested in the power of the new test under the alternative with  $\omega$  sparse. We further show that the power is minimax rate optimal.



Under assumptions (C1) and (C2), Theorem 1 indicates that under  $H_0$ ,  $M_n - 4 \log p + \log \log p$  converges weakly to a Gumbel random variable with distribution function  $\exp\{-(8\pi)^{-1/2} e^{-t/2}\}$ .

- (C1) Assume that  $\log p = \alpha n^{1/5}$ ,  $n_1 \asymp n_2$ , and for some constant  $C_0 > 0$ ,  $C_0^{-1} \leq \lambda_{\min}(\Omega_d) \leq \lambda_{\max}(\Omega_d) \leq C_0$ , for  $d = 1, 2$ . There exists some  $\tau > 0$  such that  $|A_d| = \alpha p^{1/16}$  where  $A_\tau = \{(i, j) : |\omega_{i,j,d}| \leq (\log p)^{-2-\tau}, 1 \leq i < j \leq p, \text{ for } d = 1 \text{ or } 2\}$ .
- (C2) Let  $D_d$  be the diagonal of  $\Omega_d$  and let  $(\eta_{i,j,d}) = R_d = D_d^{-1/2} \Omega_d D_d^{-1/2}$ , for  $d = 1, 2$ . Assume that  $\max_{1 \leq i < j \leq p} |\eta_{i,j,d}| = \eta_d \leq 1$  for some constant  $0 < \eta_d < 1$ .

Condition (C1) on the eigenvalues is a common assumption in the high-dimensional setting and implies that most of the variables are not highly correlated with each other. Condition (C2) is also mild. For example, if  $\max_{1 \leq i < j \leq p} |\eta_{i,j,d}| = 1$ , then  $\Omega_d$  is singular. The following theorem states the asymptotic null distribution for  $M_n$ .

**Theorem 1**—Suppose that (C1), (C2), (4) and (5) hold. Then under  $H_0$ , for any  $t \in \mathbb{R}$ ,

$$\text{pr}(M_n - 4 \log p + \log \log p \leq t) \rightarrow \exp\{-(8\pi)^{-1/2} \exp(-t/2)\}, \text{ as } n_1, n_2, p \rightarrow \infty, \quad (15)$$

where  $M_n$  is defined in equation (10). Under  $H_0$ , the convergence in (15) is uniform for all  $\{X_k : k = 1, \dots, n_1\}$  and  $\{Y_k : k = 1, \dots, n_2\}$  satisfying (C1), (C2), (4) and (5).

Equations (4) and (5) are mild conditions on the estimator of  $\beta_{i,d}$  in order to obtain the limiting distribution in Theorem 1. As discussed in Section 2.3, these conditions can be guaranteed by the lasso estimator for example.

**Corollary 1**—Suppose that (C1) and (C2) hold and  $\max_{1 \leq i \leq p} |\beta_{i,d}| = o\{n^{1/2} / (\log p)^{3/2}\}$ . Then under  $H_0$ , for any  $\kappa_d > 2$  in (13) and (14), and for any  $t \in \mathbb{R}$ ,

$$\text{pr}(M_n - 4 \log p + \log \log p \leq t) \rightarrow \exp\{-(8\pi)^{-1/2} \exp(-t/2)\}, \text{ } n_1, n_2, p \rightarrow \infty, \quad (16)$$

where  $M_n$  is defined in (10).

### 3.2. Power Analysis

We now turn to an analysis of the power of the test  $\psi_\alpha$  given in (11). We shall define the following class of precision matrices:

$$\mathcal{U}(c) = \left\{ (\Omega_1, \Omega_2) : \max_{1 \leq i < j \leq p} \frac{|\omega_{i,j,1} - \omega_{i,j,2}|}{(\theta_{i,j,1} + \theta_{i,j,2})^{1/2}} \geq c(\log p)^{1/2} \right\}. \quad (17)$$

The next theorem shows that the null parameter set in which  $\Omega_1 = \Omega_2$  is asymptotically distinguishable from  $\mathcal{U}(4)$  by the test  $\Psi_\alpha$ . That is,  $H_0$  is rejected by the test  $\Psi_\alpha$  with overwhelming probability if  $(\Omega_1, \Omega_2) \in \mathcal{U}(4)$ .

**Theorem 2**—Let the test  $\Psi_\alpha$  be given as in (11). Suppose that (C1), (4) and (5) hold. Then

$$\inf_{(\Omega_1, \Omega_2) \in \mathcal{U}(4)} \Pr(\Psi_\alpha = 1) \rightarrow 1, n, p \rightarrow \infty.$$

The following result shows that this lower bound is rate-optimal. Let  $\mathcal{T}_\alpha$  be the set of all  $\alpha$ -level tests, i.e.,  $\Pr(T_\alpha = 1) \leq \alpha$  under  $H_0$  for all  $T_\alpha \in \mathcal{T}_\alpha$ .

**Theorem 3**—Suppose that  $\log p = \alpha(n)$ . Let  $\alpha, \beta > 0$  and  $\alpha + \beta < 1$ . Then there exists a constant  $c_0 > 0$  such that for all sufficiently large  $n$  and  $p$ ,

$$\inf_{(\Omega_1, \Omega_2) \in \mathcal{U}(c_0)} \sup_{T_\alpha \in \mathcal{T}_\alpha} \Pr(T_\alpha = 1) \leq 1 - \beta.$$

Theorem 3 shows that, if  $c_0$  is sufficiently small, then any  $\alpha$  level test is unable to reject the null hypothesis correctly uniformly over  $(\Omega_1, \Omega_2) \in \mathcal{U}(c_0)$  with probability tending to one. So the order  $(\log p)^{1/2}$  in the lower bound of  $\max_{1 \leq i < j \leq p} \{|\omega_{i,j,1} - \omega_{i,j,2}| / (\theta_{i,j,1} + \theta_{i,j,2})^{1/2}\}$  in (17) cannot be improved.

#### 4. Multiple Testing with False Discovery Rate Control

If the global null hypothesis is rejected, it is often of interest to investigate the structure of the differential network. A natural approach is to carry out simultaneous testing on the elements of. In this section, we introduce a multiple testing procedure with false discovery rate control for testing  $(p^2 - p)/2$  hypotheses

$$H_{0,i,j}: \delta_{i,j} = 0 \text{ versus } H_{1,i,j}: \delta_{i,j} \neq 0, \quad 1 \leq i < j \leq p. \quad (18)$$

The standardized differences of  $T_{i,j,1}$  and  $T_{i,j,2}$  are defined by the test statistics

$W_{i,j} = (T_{i,j,1} - T_{i,j,2}) / (\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})^{1/2}$  as in (9). Let  $t$  be the threshold level such that  $H_{0,i,j}$  is rejected if  $|W_{i,j}| \geq t$ . Let  $\mathcal{H}_0 = \{(i, j) : \delta_{i,j} = 0, 1 \leq i < j \leq p\}$  be the set of true nulls. Denote by  $R_0(t) = \sum_{(i,j) \in \mathcal{H}_0} I(|W_{i,j}| \geq t)$  the total number of false positives, and by  $R(t) = \sum_{1 \leq i < j \leq p} I(|W_{i,j}| \geq t)$  the total number of rejections. The false discovery proportion and false discovery rate are defined as

$$\text{FDP}(t) = \frac{R_0(t)}{R(t) \vee 1}, \quad \text{FDR}(t) = E\{\text{FDP}(t)\}.$$

An ideal choice of  $t$  would reject as many true positives as possible while controlling the false discovery rate and false discovery proportion at the pre-specified level  $\alpha$ . That is, we select

$$t_0 = \inf \left\{ 0 \leq t \leq 2(\log p)^{1/2} : \text{FDP}(t) \leq \alpha \right\}.$$

Since  $\mathcal{H}_0$  is unknown, we can estimate  $\sum_{(i,j) \in \mathcal{H}_0} I\{|W_{i,j}| \geq t\}$  by  $2\{1 - \Phi(t)\}|\mathcal{H}_0|$  as in Liu (2013), where  $\Phi(t)$  is the standard normal cumulative distribution function. Note that  $|\mathcal{H}_0|$  can be estimated by  $(p^2 - p)/2$  due to the sparsity of  $\mathcal{H}_0$ . This leads to the following multiple testing procedure.

1. Calculate the test statistics  $W_{i,j}$ .
2. For given  $0 < \alpha < 1$ , calculate

$$\hat{t} = \inf \left\{ 0 \leq t \leq 2(\log p)^{1/2} : \frac{2\{1 - \Phi(t)\}(p^2 - p)/2}{R(t) \vee 1} \leq \alpha \right\}.$$

If  $\hat{t}$  does not exist, set  $\hat{t} = 2(\log p)^{1/2}$ .

3. For  $1 \leq i < j \leq p$ , reject  $H_{0,i,j}$  if and only if  $|W_{i,j}| \geq \hat{t}$ .

The following theorem shows that, under regularity conditions, the above procedure controls the false discovery proportion and false discovery rate at the pre-specified level  $\alpha$  asymptotically.

**Theorem 4**

Let

$$\mathcal{S}_\rho = \left\{ (i, j) : 1 \leq i < j \leq p, \frac{|\omega_{i,j,1} - \omega_{i,j,2}|}{(\theta_{i,j,1} + \theta_{i,j,2})^{1/2}} \geq (\log p)^{1/2 + \rho} \right\}.$$

Suppose for some  $\rho > 0$  and some  $\delta > 0$ ,  $|\mathcal{S}_\rho| \geq [1/\{(8\pi)^{1/2}\alpha\} + \delta](\log \log p)^{1/2}$ . Suppose that  $|A_\nu \cap \mathcal{H}_0| = o(p^\nu)$  for any  $\nu > 0$ , where  $A_\nu$  is given in Condition (C1). Assume that  $q_0 = |\mathcal{H}_0| \geq cp^2$  for some  $c > 0$ , and (4) and (5) hold. Let  $q = (p^2 - p)/2$ . Then under (C1) with  $p \rightarrow \infty$  for some  $c > 0$  and  $r > 0$ , we have

$$\lim_{(n,p) \rightarrow \infty} \frac{\text{FDR}(\hat{t})}{\alpha q_0/q} = 1, \quad \frac{\text{FDP}(\hat{t})}{\alpha q_0/q} \rightarrow 1$$

in probability, as  $(n, p) \rightarrow \infty$ .

The condition  $|\mathcal{L}_\rho| \geq [1/\{(8\pi)^{1/2}\alpha\} + \delta](\log \log p)^{1/2}$  in Theorem 4 is mild, since there are  $(p^2 - p)/2$  hypotheses in total and this condition only requires a few entries with the standardized difference having magnitude exceeding  $\{(\log p)^{1/2 + \rho/n}\}^{1/2}$  for some constant  $\rho > 0$ . The technical condition  $|A_{\mathcal{J}} \cap \mathcal{H}_0| = o(p^\nu)$  for any  $\nu > 0$  is to ensure that most of the regression residuals are not highly correlated with each other under the null hypotheses  $H_{0,i,j} : \delta_{i,j} = 0$ .

The basic idea for the proof of Theorem 4 is similar to that in Liu (2013). However, the setting here is more complicated as  $\omega_{i,j,1}$  and  $\omega_{i,j,2}$  are not necessarily zero under  $H_{0,i,j} : \delta_{i,j} = 0$ . So the coordinates of the regression residuals in (2) and (3) can be correlated with each other. Thus slightly stronger conditions are needed and the proof is more involved.

### 5. Simulation Study

The proposed testing procedures are easy to implement, and the Matlab code is available in the Supplementary Material. We carry out a simulation study to investigate the numerical performance, including the size and power, of the global test  $\Psi_\alpha$  and the false discovery rate controlled multiple testing procedure.

We first introduce the matrix models used in the simulations. Let  $D = (D_{i,j})$  be a diagonal matrix with  $D_{i,i} = \text{Unif}(0.5, 2.5)$  for  $i = 1, \dots, p$ . The following four models under the null,  $\Omega_1 = \Omega_2 = \Omega^{(m)} = (\omega_{i,j}^{(m)}) (m=1, \dots, 4)$ , are used to study the size of the tests.

**Model 1:**  $\Omega^{*(1)} = (\omega_{i,j}^{*(1)})$  where  $\omega_{i,i}^{*(1)} = 1, \omega_{i,i+1}^{*(1)} = \omega_{i+1,i}^{*(1)} = 0.6, \omega_{i,i+2}^{*(1)} = \omega_{i+2,i}^{*(1)} = 0.3$  and  $\omega_{i,j}^{*(1)} = 0$  otherwise.  $\Omega^{(1)} = D^{1/2} \Omega^{*(1)} D^{1/2}$ .

**Model 2:**  $\Omega^{*(2)} = (\omega_{i,j}^{*(2)})$  where  $\omega_{j,i}^{*(2)} = \omega_{i,j}^{*(2)} = 0.5$  for  $i = 10(k-1) + 1$  and  $10(k-1) + 2 \leq j \leq 10(k-1) + 10, 1 \leq k \leq p/10$ .  $\omega_{i,j}^{*(2)} = 0$  otherwise.  $\Omega^{(2)} = D^{1/2} (\Omega^{*(2)} + \delta I) / (1 + \delta) D^{1/2}$  with  $\delta = |\lambda_{\min}(\Omega^{*(2)})| + 0.05$ .

**Model 3:**  $\Omega^{*(3)} = (\omega_{i,j}^{*(3)})$  where  $\omega_{i,i}^{*(3)} = 1, \omega_{i,j}^{*(3)} = 0.8 \times \text{Bernoulli}(1, 0.05)$  for  $i < j$  and  $\omega_{j,i}^{*(3)} = \omega_{i,j}^{*(3)}$ .  $\Omega^{(3)} = D^{1/2} (\Omega^{*(3)} + \delta I) / (1 + \delta) D^{1/2}$  with  $\delta = |\lambda_{\min}(\Omega^{*(3)})| + 0.05$ .

**Model 4:**  $\Sigma^{*(4)} = (\sigma_{i,j}^{*(4)})$  where  $\sigma_{i,i}^{*(4)} = 1, \sigma_{i,j}^{*(4)} = 0.5$  for  $2(k-1) + 1 \leq i < j \leq 2k$ , where  $k = 1, \dots, [p/2]$  and  $\sigma_{i,j}^{*(4)} = 0$  otherwise.  $\Omega^{(4)} = D^{1/2} \{(\Sigma^{*(4)} + \delta I) / (1 + \delta)\}^{-1} D^{1/2}$  with  $\delta = |\lambda_{\min}(\Sigma^{*(4)})| + 0.05$ .

For global testing of  $H_0 : \delta = 0$ , the sample sizes are taken to be  $n_1 = n_2 = 100$ , while the dimension  $p$  varies over the values 50, 100, 200 and 400. For each model, data are generated from multivariate normal distributions with mean zero and covariance matrices  $\Sigma_1 = \Omega_1^{-1}$  and  $\Sigma_2 = \Omega_2^{-1}$ . The nominal significance level for all the tests is set at  $\alpha = 0.05$ .

To evaluate the power of the proposed tests, let  $U = (u_{i,j})$  be a matrix with eight random nonzero entries. The locations of four nonzero entries are selected randomly from the upper

triangle of  $U$ , each with a magnitude generated randomly and uniformly from the set  $[-2\omega(\log p/n)^{1/2}, -\omega(\log p/n)^{1/2}] \cup [\omega(\log p/n)^{1/2}, 2\omega(\log p/n)^{1/2}]$ , where  $\omega = \max_{1 \leq i \leq p} \omega_{i,i}^{(m)}$ . The other four nonzero entries in the lower triangle are determined by symmetry. We use the following four pairs of precision matrices  $(\Omega_1^{(m)}, \Omega_2^{(m)}) (m=1, \dots, 4)$ , to show the power of the test, where  $\Omega_1^{(m)} = \Omega^{(m)} + \delta I$  and  $\Omega_2^{(m)} = \Omega^{(m)} + U + \delta I$ , with  $\delta = |\min\{\lambda_{\min}(\Omega^{(m)} + U), \lambda_{\min}(\Omega^{(m)})\}| + 0.05$ . The actual sizes and powers in percentage for the four models, reported in Table 1, are estimated from 1000 replications.

Table 1 shows that the sizes of the global test  $\Psi_{\alpha_1}$  are close to the nominal level in all cases. This reflects the fact that the null distribution of the test statistic  $M_n$  is well approximated by its asymptotic distribution. The empirical sizes are slightly below the nominal level in some models, due to the correlation among the variables. Similar phenomena have also been observed in Cai et al. (2013) and are theoretically justified by their Proposition 1. Table 1 shows that the proposed test is powerful in all settings, although the two precision matrices differ only in eight entries with the magnitude of the difference of the order  $(\log p/n)^{1/2}$ .

In addition, we consider nearer alternatives by generating the nonzero entries randomly and uniformly from the set  $[-\omega(2 \log p/n)^{1/2}, \omega(2 \log p/n)^{1/2}]$ . The power results are summarized in Table 2. Under the nearer alternatives, the magnitude of the standardized difference of  $\Omega_1 - \Omega_2$  is smaller and as a result the power is lower.

More extensive simulation results are presented in the Supplementary Material. The proposed test significantly outperforms both that of Cai et al. (2013), which is powerful when  $\Sigma_1 - \Sigma_2$  is sparse under the alternative, and that of Li & Chen (2012), which is powerful when  $\Sigma_1 - \Sigma_2$  is dense under the alternative.

For simultaneous testing of the individual entries of the differential network with false discovery rate control, we select  $\lambda_{n_d, i, d}$  in (13) and (14) adaptively with the principle of making  $\sum_{(i,j) \in \mathcal{H}_0} I(|W_{i,j}| \geq t)$  and  $\{2 - 2\Phi(t)\} |\mathcal{H}_0|$  as close as possible. The algorithm is as follows.

1. For any given  $i \in \{1, \dots, p\}$ , let  $\lambda_{n_1, i, 1} = (s/20) (\sum_{i, i, 1} \log p/n_1)^{1/2}$  and  $\lambda_{n_2, i, 2} = (s/20) (\sum_{i, i, 2} \log p/n_2)^{1/2}$  for  $s = 1, \dots, 40$ . For each  $s$ , calculate  $\hat{\beta}_{i,d}^{(s)} (i=1, \dots, p)$  and  $d=1, 2$ . Based on the estimated regression coefficients, construct the corresponding standardized difference  $W_{i,j}^{(s)}$  for each  $s$ .

2. Choose

$$\hat{s} = \arg \min \sum_{l=1}^{10} \left( \frac{\sum_{1 \leq i \leq j \leq p} I\{|W_{i,j}^{(s)}| \geq \Phi^{-1}(1 - l[1 - \Phi\{(\log p)^{1/2}\}]/10)\}}{lp(p-1)[1 - \Phi\{(\log p)^{1/2}\}]/10} - 1 \right)^2.$$

The tuning parameters are chosen to be  $\lambda_{n_1,i,1} = \hat{s}/20(\sum_{i,i,1} \log p/n_1)^{1/2}$  and  $\lambda_{n_2,i,2} = \hat{s}/20(\sum_{i,i,2} \log p/n_2)^{1/2}$ .

Pairwise comparisons among these four models are considered. The sample sizes are  $n_1 = n_2 = 100$ , while the dimension  $p = 50, 100, \text{ and } 200$ . The false discovery rate level is set at  $\alpha_2 = 0.1$ , and the empirical false discovery rate and the power of false discovery rate control in percentage, summarized in Table 3, are estimated from 100 replications. We examine the power based on the average powers for 100 replications as follows

$$\frac{1}{100} \sum_{l=1}^{100} \frac{\sum_{(i,j) \in \mathcal{H}_1} I(|W_{i,j,l}| \geq \hat{t})}{|\mathcal{H}_1|},$$

where  $W_{i,j,l}$  denotes standardized difference for the  $l^{\text{th}}$  replication and  $\mathcal{H}_1$  denotes the nonzero locations. For all six cases, the false discovery rates are close to  $\alpha$  across all dimensions. For empirical power, the procedure is powerful when the dimension  $p$  is low, and retains high power for the comparisons between Model 1 and Models 2 and 4. However, for the comparison between Model 2 and Model 3, the power is low when dimension is high and this is because all of  $|\omega_{i,j,1} - \omega_{i,j,2}|/(\theta_{i,j,1}n_1 + \theta_{i,j,2}n_2)^{1/2}$  is smaller than 0.25 when  $p = 200$  and  $D = I$ . Similarly, most nonzero entries of the standardized difference for Model 2 and 4 are smaller than 0.24. Thus it is difficult to detect nonzero locations. Furthermore, under the same scenario,  $\omega_{i,j}/(\theta_{i,j}n)^{1/2}$  is always smaller than 0.16 for Model 3, and thus the detection becomes harder when we compare Model 3 with other models. Thus, the power results are not good when Model 3 is included in the comparison.

## 6. Real Data Analysis

The high throughput technology and massively parallel measurement of mRNA expression catalyzed a new area of genomic biomarkers. A number of prominent genomic markers have been identified to assist in predicting breast cancer patient survival in clinical practice, and increasingly, pharmacogenomic endpoints are being incorporated into the design of clinical trials (Olopade et al., 2008). Molecular pathways of pathogenesis for breast cancer have also been increasingly discovered and curated (Nathanson et al., 2001). However, the role of gene-by-gene interactions, within and across pathways, in breast cancer survival remain unclear. Here, we apply our procedures to identify gene-by-gene interactions important for breast cancer survival.

For illustration, we consider 32 pathways from the molecular signature database that are related to breast cancer survival. Examples include the MAPK/ERK, WNT, TGF- $\beta$ , P13k-AKT-mTOR and ATRBRCA pathways. Existing literature has indicated that a defect in the MAPK pathway may lead to uncontrolled growth, which is a step necessary for the development of all cancers (Santen et al., 2002; Downward, 2003). Mutations or deregulated expression of genes in the Wnt pathway can induce cancer (Klaus & Birchmeier, 2008). The TGF- $\beta$  signaling pathway is critical to a plethora of cellular processes including cell proliferation, apoptosis and differentiation (Shi & Massagué, 2003). An increase in the TGF-

$\beta_2$  expression is associated with response to tamoxifen for breast cancer patients (Buck & Knabbe, 2006). The ATRBRCA pathway describes the role of BRCA1, BRCA2 and ATR in cancer susceptibility (Venkitaraman, 2002). BRCA1 and BRCA2 are the best-known genes linked to breast cancer risk. Hence, these pathways may play critical roles in breast cancer progression. To examine the interactions between genes in these pathways, we applied our procedure to a recent breast cancer gene expression study of 295 patients with primary breast carcinomas from the Netherlands Cancer Institute (van de Vijver et al., 2002). Out of the 32 pathways, there are a total of  $p = 754$  genes with available data in this study. The two populations we consider are the short term survivors, defined as those 78 patients who died within 5 years; and the long term survivors, defined as those 69 patients who survived more than 10 years. We are particularly interested in identifying gene pairs with interactive effects on the binary cancer survival trait using the proposed procedures. In this setting, the sparsity assumption about  $\beta_{i,k}$ 's is reasonable as it is generally believed that transcriptional regulation of a single gene is generally defined by a small set of regulatory elements (Segal et al., 2003; Dobra et al., 2004).

Based on our proposed procedures, we identified nine pairs of gene-by-gene interactions as significant at a false discovery rate level of 0.1. An interaction here does not simply indicate a co-expression between a pair of genes, but instead represents a difference between the co-expression patterns among the long terms survivors and among the short term survivors. As shown in Figure 1, the majority of the genes involved in these interactions belong to five major pathways, the MAPK, WNT, TGF- $\beta$ , Apoptosis, and ATRBRCA pathways, although many of these genes belong to multiple pathways. One pair of the identified interactions represent gene-by-gene interactions within pathways and the remaining eight pairs represent cross-talk between these pathways, some of which are previously documented. A total of five interactions are between the MAPK signaling pathway and the WNT and TGF- $\beta$ , Apoptosis, ATRBRCA and MTA3 pathways. These cross-talks are not surprising since MAPK modulates a wide range of processes including gene expression, mitosis, proliferation, metabolism and apoptosis (Wada & Penninger, 2004). Several recent studies suggest extensive crosstalk between WNT and MAPK signaling pathways in cancer. For example, hyper-activation of MAPK signaling results in down-regulation of the WNT signal transduction pathway in melanoma, suggesting a negative crosstalk between the two pathways; while in colorectal cancer, stimulating the WNT pathway leads to activation of the MAPK pathway through Ras stabilization, representing a positive crosstalk (Guardavaccaro & Clevers, 2012). The observed interactive effect between the WNT and MAPK pathways suggests that the cross-talk between these two pathways may play an important role in breast cancer survival. The interaction between the tumor suppressor gene BRCA2 and the MAPK pathway has been documented in experiments with prostate cancer cells with upregulation of BRCA2 linked to an increase in MAPK activity (Moro et al., 2007). In the WNT pathway, the WNT1 gene promotes cell survival in various cell types and it has been experimentally shown that blocking WNT1 signaling can induce apoptotic cell death (You et al., 2004). Thus the interaction between WNT1 gene and the PRKACB gene in the Apoptosis pathway may also be crucial for breast cancer.

## A. Appendix: Proofs

### A-1. Technical Lemmas

We prove the main results in this section. We begin by collecting technical lemmas proved in the supplementary material. The first lemma is the classical Bonferroni inequality.

#### Lemma A1 (Bonferroni inequality)

Let  $B = \cup_{t=1}^p B_t$ . For any  $k < [p/2]$ , we have

$$\sum_{t=1}^{2k} (-1)^{t-1} F_t \leq \text{pr}(B) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} F_t,$$

where  $F_t = \sum_{1 \leq i_1 < \dots < i_t \leq p} \text{pr}(B_{i_1} \cap \dots \cap B_{i_t})$ .

For  $d = 1, 2$ , let  $U_{i,j,d} = n_d^{-1} \sum_{k=1}^{n_d} (\varepsilon_{k,i,d} \varepsilon_{k,j,d} - E \varepsilon_{k,i,d} \varepsilon_{k,j,d})$ , and define

$\tilde{U}_{i,j,d} = (r_{i,j,d} - U_{i,j,d}) / (r_{i,i,d} r_{j,j,d})$  for  $1 \leq i < j \leq p$  and  $\tilde{U}_{i,i,d} = (r_{i,i,d} + U_{i,i,d}) / (r_{i,i,d} r_{i,i,d})$ .

#### Lemma A2

Suppose that Conditions (C1), (4) and (5) hold. Then

$$\max_{1 \leq i \leq p} |\tilde{r}_{i,i,d} - r_{i,i,d}| = O_p\{(\log p / n_d)^{1/2}\},$$

and

$$\tilde{r}_{i,j,d} = \tilde{R}_{i,j,d} - \tilde{r}_{i,i,d} (\hat{\beta}_{i,j,d} - \beta_{i,j,d}) - \tilde{r}_{j,j,d} (\hat{\beta}_{j-1,i,d} - \beta_{j-1,i,d}) + o_p\{(n_d \log p)^{-1/2}\},$$

for  $1 \leq i < j \leq p$ , where  $\tilde{R}_{i,j,d}$  is the empirical covariance between  $\{\varepsilon_{k,i,d} : k = 1, \dots, n_d\}$  and  $\{\varepsilon_{k,j,d} : k = 1, \dots, n_d\}$ . Consequently, uniformly in  $1 \leq i < j \leq p$ ,

$$\begin{aligned} \hat{r}_{i,j,d} - (\omega_{i,i,d} \hat{\sigma}_{i,i,d,\varepsilon} + \omega_{j,j,d} \hat{\sigma}_{j,j,d,\varepsilon} - 1) r_{i,j,d} &= -U_{i,j,d} + o_p\{(n_d \log p)^{-1/2}\}, \\ |\tilde{T}_{i,j,d} - \tilde{U}_{i,j,d}| &= O_p\{(\log p / n_d)^{1/2}\} r_{i,j,d} + o_p\{(n_d \log p)^{-1/2}\}, \end{aligned}$$

and uniformly in  $1 \leq i \leq p$ ,

$$|\tilde{T}_{i,i,d} - \tilde{U}_{i,i,d}| = o_p\{(n_d \log p)^{-1/2}\},$$

where  $\hat{r}_{i,j,d}$  is defined in (7),

$(\hat{\sigma}_{i,j,d,\varepsilon}) = (1/n_d) \sum_{k=1}^{n_d} (\varepsilon_{k,d} - \bar{\varepsilon}_d)(\varepsilon_{k,d} - \bar{\varepsilon}_d)^T$ ,  $\varepsilon_{k,d} = (\varepsilon_{k,1,d}, \dots, \varepsilon_{k,p,d})$  and

$\bar{\varepsilon}_d = n_d^{-1} \sum_{k=1}^{n_d} \varepsilon_{k,d}$ .



**Lemma A3**

Let  $X_k \sim \mathcal{N}(\mu_1, \Sigma_1)$  for  $k = 1, \dots, n_1$  and  $Y_k \sim \mathcal{N}(\mu_2, \Sigma_2)$  for  $k = 1, \dots, n_2$ . Define

$$\tilde{\Sigma}_1 = (\tilde{\sigma}_{i,j,1})_{p \times p} = \frac{1}{n_1} \sum_{k=1}^{n_1} (X_k - \mu_1)(X_k - \mu_1)^T, \quad \tilde{\Sigma}_2 = (\tilde{\sigma}_{i,j,2})_{p \times p} = \frac{1}{n_2} \sum_{k=1}^{n_2} (Y_k - \mu_2)(Y_k - \mu_2)^T.$$

Then, for some constant  $C > 0$ ,  $\tilde{\sigma}_{i,j,1} - \tilde{\sigma}_{i,j,2}$  satisfies the large deviation bound

$$\Pr \left[ \max_{(i,j) \in \mathcal{S}} \frac{(\tilde{\sigma}_{i,j,1} - \tilde{\sigma}_{i,j,2} - \sigma_{i,j,1} + \sigma_{i,j,2})^2}{\text{var}\{(X_{k,i} - \mu_{1,i})(X_{k,j} - \mu_{1,j})\}/n_1 + \text{var}\{(Y_{k,i} - \mu_{2,i})(Y_{k,j} - \mu_{2,j})\}/n_2} \geq x^2 \right] \leq C |\mathcal{S}| \{1 - \Phi(x)\} + O(p^{-1})$$

uniformly for  $0 < x < (8 \log p)^{1/2}$  and any subset  $\mathcal{S} \subseteq \{(i, j) : 1 \leq i \leq j \leq p\}$ .

The following lemma is needed for false discovery rate control in Theorem 4.

**Lemma A4**

Let  $V_{i,j} = (U_{i,j,2} - U_{i,j,1}) \{\text{var}(\mathbf{e}_{k,i,1} \mathbf{e}_{k,j,1})/n_1 + \text{var}(\mathbf{e}_{k,i,2} \mathbf{e}_{k,j,2})/n_2\}^{-1/2}$ . Under the same conditions as in Theorem 4, we have for any  $\varepsilon > 0$  that,

$$\sum_{0 \leq t \leq t_p} \Pr \left[ \left| \frac{\sum_{(i,j) \in \mathcal{A}_\tau} \{I(|V_{i,j}| \geq t) - \Pr(|V_{i,j}| \geq t)\}}{2q_0 \{1 - \Phi(t)\}} \right| \geq \varepsilon \right] = o(1),$$

$$\int_0^{t_p} \Pr \left[ \left| \frac{\sum_{(i,j) \in \mathcal{A}_\tau} \{I(|V_{i,j}| \geq t) - \Pr(|V_{i,j}| \geq t)\}}{2q_0 \{1 - \Phi(t)\}} \right| \geq \varepsilon \right] dt = o(v_p),$$

where  $t_p = (4 \log p - \log_2 p - \log_3 p)^{1/2}$  and  $v_p = 1/\{\log p(\log_4 p)^2\}^{1/2}$ .

**A-2. Proof of Theorem 1**

Without loss of generality, throughout this section, we assume that  $\omega_{i,i,d} = 1$  for  $d = 1, 2$  and  $i = 1, \dots, p$ . Let  $A = \{(i, j) : 1 \leq i < j \leq p\}$ . (C1) implies  $|A_\tau| = o(p^{1/16})$ . To prove Theorem 1, we first show that the terms in  $A_\tau$  are negligible. Then we use Lemma 1, together with the Gaussian approximation technique, to show that

$$\Pr(\max_{(i,j) \in A \setminus A_\tau} W_{i,j}^2 - 4 \log p + \log \log p \leq t) \rightarrow \exp\{-(8\pi)^{-1/2} \exp(-t/2)\},$$

where  $W_{i,j}$  is defined in equation (9).

For  $d = 1, 2$ , let  $V_{i,j} = (U_{i,j,2} - U_{i,j,1}) \{\text{var}(\mathbf{e}_{k,i,1} \mathbf{e}_{k,j,1})/n_1 + \text{var}(\mathbf{e}_{k,i,d} \mathbf{e}_{k,j,d})/n_2\}^{1/2}$ , where  $\text{var}(\varepsilon_{k,i,d} \varepsilon_{k,j,d}) = r_{i,i,d} r_{j,j,d} (1 + \rho_{i,j,d}^2)$  with  $\rho_{i,j,d}^2 = \beta_{i,j,d}^2 = \beta_{i,j,d}^2 r_{i,i,d} / r_{j,j,d}$ . The proof of Lemma 2 yields

$$\max_{1 \leq i \leq p} |\hat{r}_{i,i,d} - r_{i,i,d}| = O_p\{(\log p/n)^{1/2}\}, \quad (\text{A1})$$

and  $\max_{1 \leq i \leq p} |\hat{r}_{i,i,d} - \tilde{R}_{i,i,d}| = o_p\{(n_d \log p)^{-1/2}\}$ , where  $n = \max\{n_1, n_2\}$ . Note that

$$\max_{1 \leq i \leq j \leq p} |\hat{\rho}_{i,j,d}^2 \hat{r}_{i,i,d} / \hat{r}_{j,j,d} - \rho_{i,j,d}^2| = o_p(1/\log p), \quad (A2)$$

and  $\max_{1 \leq i \leq j \leq p} |\omega_{i,i,d} \hat{\sigma}_{i,i,d,\varepsilon} + \omega_{j,j,d} \hat{\sigma}_{j,j,d,\varepsilon} - 2| = O_p\{(\log p/n)^{1/2}\}$ . Also note that for  $(i, j) \in A \setminus A_\tau$ , we have  $|\omega_{i,j,d}| = o\{(\log p)^{-1}\}$ . Then by Lemma 2, it is easy to see that, under conditions (C1), (4) and (5), we have, for  $(i, j) \in A \setminus A_\tau$ ,

$\max_{(i,j) \in A \setminus A_\tau} \|W_{i,j}\| - \|V_{i,j}\| = o_p\{(\log p)^{-1/2}\}$ . For  $(i,j) \in A_\tau$  as a result of Lemma 2, we have  $W_{i,j} = V_{i,j} + b_{i,j} + o_p(\log p^{-1/2})$ , where

$$b_{i,j} = 2\{\omega_{i,j}(\hat{\sigma}_{i,i,1,\varepsilon} - \hat{\sigma}_{i,i,2,\varepsilon}) + \omega_{i,j}(\hat{\sigma}_{j,j,1,\varepsilon} - \hat{\sigma}_{j,j,2,\varepsilon})\} / (\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})^{1/2},$$

$$(\hat{\sigma}_{i,j,d,\varepsilon}) = n_d^{-1} \sum_{k=1}^{n_d} (\varepsilon_{k,d} - \bar{\varepsilon}_d)(\varepsilon_{k,d} - \bar{\varepsilon}_d)^T, \varepsilon_{k,d} = (\varepsilon_{k,1,d}, \dots, \varepsilon_{k,p,d}) \text{ and } \bar{\varepsilon}_d = (1/n_d) \sum_{k=1}^{n_d} \varepsilon_{k,d}.$$

Note that

$$|b_{i,j}| \leq 2 \left( \frac{2\rho_{i,j}^2}{1+\rho_{i,j}^2} \right)^{\frac{1}{2}} \left[ \frac{|\tilde{\sigma}_{i,i,1,\varepsilon} - \tilde{\sigma}_{i,i,2,\varepsilon}|}{\{\text{var}(\varepsilon_{k,i,1}^2)/n_1 + \text{var}(\varepsilon_{k,i,2}^2)/n_2\}^{\frac{1}{2}}} + \frac{|\tilde{\sigma}_{j,j,1,\varepsilon} - \tilde{\sigma}_{j,j,2,\varepsilon}|}{\{\text{var}(\varepsilon_{k,j,1}^2)/n_1 + \text{var}(\varepsilon_{k,j,2}^2)/n_2\}^{\frac{1}{2}}} \right] + o\{(\log p)^{-1/2}\},$$

where  $\tilde{\sigma}_{i,i,d,\varepsilon} = n_d^{-1} \sum_{k=1}^{n_d} \varepsilon_{k,i,d}^2$ . Thus, we have

$$\text{pr}(\max_{(i,j) \in A_\tau} W_{i,j}^2 \geq 4\log p - \log \log p + t) \leq \text{Card}(A_\tau) \{\text{pr}(V_{i,j}^2 \geq \log p/8) + \text{pr}(b_{i,j}^2 \geq 2\log p)\} = o(1),$$

where the last equality is a direct result of Lemma 3. Thus it suffices to prove that

$$\text{pr}(\max_{(i,j) \in A \setminus A_\tau} V_{i,j}^2 - 4\log p + \log \log p \leq t) \rightarrow \exp\{-(8\pi)^{-1/2} \exp(-t/2)\}.$$

We arrange the indices  $\{(i, j) : (i, j) \in A \setminus A_\tau\}$  in any ordering and set them as  $\{(i_m, j_m) : m = 1, \dots, q\}$  with  $q = \text{Card}(A \setminus A_\tau)$ . Let  $n_1/n_2 = K$  with  $K \geq 1$ ,  $\theta_{m,d} = \text{var}(\varepsilon_{i_m,d} \varepsilon_{j_m,d})$ , for  $d = 1, 2$  and define  $Z_{k,m} = (n_1/n_2) \{\varepsilon_{k,i_m,2} \varepsilon_{k,j_m,2} - E(\varepsilon_{k,i_m,2} \varepsilon_{k,j_m,2})\}$  for  $1 \leq k \leq n_2$ ,

$$Z_{k,m} = -\{\varepsilon_{k,i_m,1} \varepsilon_{k,j_m,1} - E(\varepsilon_{k,i_m,1} \varepsilon_{k,j_m,1})\} \text{ for } n_2 + 1 \leq k \leq n_1 + n_2,$$

$$V_m = (n_1^2 \theta_{m,2}/n_2 + n_1 \theta_{m,1})^{-1/2} \sum_{k=1}^{n_1+n_2} Z_{k,m} \text{ and}$$

$$\hat{V}_m = (n_1^2 \theta_{m,2}/n_2 + n_1 \theta_{m,1})^{-1/2} \sum_{k=1}^{n_1+n_2} \hat{Z}_{k,m}, \text{ where}$$

$$\hat{Z}_{k,m} = Z_{k,m} I(|Z_{k,m}| \leq \tau_n) - E\{Z_{k,m} I(|Z_{k,m}| \leq \tau_n)\}, \text{ and } \tau_n = 32K_1 \log(p+n). \text{ Note that}$$

$$\max_{(i,j) \in A \setminus A_\tau} V_{i,j}^2 = \max_{1 \leq m \leq q} V_m^2, \text{ and that}$$

$$\begin{aligned} & \max_{1 \leq m \leq q} n^{-1/2} \sum_{k=1}^{n_1+n_2} E[|Z_{k,m}| I\{|Z_{k,m}| \geq 32K_1 \log(p+n)\}] \\ & \leq Cn^{1/2} \max_{1 \leq k \leq n_1+n_2} \max_{1 \leq m \leq q} E[|Z_{k,m}| I\{|Z_{k,m}| \geq 32K_1 \log(p+n)\}] \\ & \leq Cn^{1/2}(p+n)^{-4} \max_{1 \leq k \leq n_1+n_2} \max_{1 \leq m \leq q} E[|Z_{k,m}| \exp\{|Z_{k,m}|/(8K_1)\}] \\ & \leq Cn^{1/2}(p+n)^{-4}. \end{aligned}$$

Hence,

$$\text{pr}\{\max_{1 \leq m \leq q} |V_m - \hat{V}_m| \geq (\log p)^{-1}\} \leq \text{pr}(\max_{1 \leq m \leq q} \max_{1 \leq k \leq n_1+n_2} |Z_{k,m}| \geq \tau_n) = O(p^{-1})$$

. By the fact that

$$|\max_{1 \leq m \leq q} V_m^2 - \max_{1 \leq m \leq q} \hat{V}_m^2| \leq 2 \max_{1 \leq m \leq q} |\hat{V}_m| \max_{1 \leq m \leq q} |V_m - \hat{V}_m| + \max_{1 \leq m \leq q} |V_m - \hat{V}_m|^2$$

, it suffices to prove that for any  $t \in \mathbb{R}$ , as  $n, p \rightarrow \infty$ ,

$$\text{pr}\left(\max_{1 \leq m \leq q} \hat{V}_m^2 - 4 \log p + \log \log p \leq t\right) \rightarrow \exp\left\{-(8\pi)^{1/2} \exp(-t/2)\right\}. \tag{A3}$$

By Lemma 1, for any integer  $l$  with  $0 < l < q/2$ ,

$$\begin{aligned} \sum_{d=1}^{2l} (-1)^{d-1} \sum_{1 \leq m_1 < \dots < m_d \leq q} \text{pr}\left(\bigcap_{j=1}^d F_{m_j}\right) & \leq \text{pr}\left(\max_{1 \leq m \leq q} \hat{V}_m^2 \geq y_p\right) \\ & \leq \sum_{d=1}^{2l-1} (-1)^{d-1} \sum_{1 \leq m_1 < \dots < m_d \leq q} \text{pr}\left(\bigcap_{j=1}^d F_{m_j}\right), \end{aligned} \tag{A4}$$

where  $y_p = 4 \log p - \log \log p + t$  and  $F_{m_j} = (\hat{V}_{m_j}^2 \geq y_p)$ . Let

$\tilde{Z}_{k,m} = \hat{Z}_{k,m} / (n_1 \theta_{m,2} / n_2 + \theta_{m,1})^{1/2}$  for  $m = 1, \dots, q$  and  $W_k = (\tilde{Z}_{k,m_1}, \dots, \tilde{Z}_{k,m_d})$  for  $1 \leq k \leq n_1 + n_2$ . Define  $|a|_{\min} = \min_{1 \leq i \leq d} |a_i|$  for any vector  $a \in \mathbb{R}^d$ . Then we have

$$\text{pr}\left(\bigcap_{j=1}^d F_{m_j}\right) = \text{pr}\left(\left|n_2^{-1/2} \sum_{k=1}^{n_1+n_2} W_k\right|_{\min} \geq y_p^{1/2}\right).$$

Then it follows from Theorem 1 in Zaitsev (1987) that

$$\begin{aligned} \text{pr}\left(\left|n_2^{-1/2} \sum_{k=1}^{n_1+n_2} W_k\right|_{\min} \geq y_p^{1/2}\right) & \leq \text{pr}\left\{|N_d|_{\min} \geq y_p^{1/2} - \varepsilon_n (\log p)^{-1/2}\right\} \\ & \quad + c_1 d^{5/2} \exp\left\{-\frac{n^{1/2} \varepsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}}\right\}, \end{aligned} \tag{A5}$$

where  $c_1 > 0$  and  $c_2 > 0$  are constants,  $\varepsilon_n \rightarrow 0$  which will be specified later and

$N_d = (N_{m_1}, \dots, N_{m_d})$  is a normal random vector with  $E(N_d) = 0$  and

$\text{cov}(N_d) = n_2/n_1 \text{cov}(W_1) + \text{cov}(W_{n_2+1})$ . Recall that  $d$  is a fixed integer which does not depend on  $n, p$ . Because  $\log p = o(n^{1/5})$ , we can let  $\varepsilon_n \rightarrow 0$  sufficiently slowly that, for any large  $M > 0$

$$c_1 d^{5/2} \exp \left\{ -\frac{n^{1/2} \varepsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}} \right\} = O(p^{-M}). \quad (\text{A6})$$

Combining (A4), (A5) and (A6) we have

$$\text{pr} \left( \max_{1 \leq m \leq q} \hat{V}_m^2 \geq y_p \right) \leq \sum_{d=1}^{2l-1} (-1)^{d-1} \sum_{1 \leq m_1 < \dots < m_d \leq q} \text{pr} \left\{ |N_d|_{\min} \geq y_p^{1/2} - \varepsilon_n (\log p)^{-1/2} \right\} + o(1). \quad (\text{A7})$$

Similarly, using Theorem 1 in Zaitsev (1987) again, we can get

$$\text{pr} \left( \max_{1 \leq m \leq q} \hat{V}_m^2 \geq y_p \right) \geq \sum_{d=1}^{2l} (-1)^{d-1} \sum_{1 \leq m_1 < \dots < m_d \leq q} \text{pr} \left\{ |N_d|_{\min} \geq y_p^{1/2} + \varepsilon_n (\log p)^{-1/2} \right\} - o(1). \quad (\text{A8})$$

We recall the following lemma, which is shown in the supplementary material of Cai et al. (2013).

#### Lemma A5

For any fixed integer  $d \geq 1$  and real number  $t \in \mathbb{R}$ ,

$$\sum_{1 \leq m_1 < \dots < m_d \leq q} \text{pr} \left\{ |N_d|_{\min} \geq y_p^{1/2} \pm \varepsilon_n (\log p)^{-1/2} \right\} = \frac{1}{d!} \left\{ (8\pi)^{-1/2} \exp(-t/2) \right\}^d \{1 + o(1)\}. \quad (\text{A9})$$

It then follows from Lemma 5, (A7) and (A8) that

$$\begin{aligned} \limsup_{n,p \rightarrow \infty} \left( \max_{1 \leq m \leq q} \hat{V}_m^2 \geq y_p \right) &\leq \sum_{d=1}^{2l} (-1)^{d-1} \frac{1}{d!} \{(8\pi)^{-1/2} \exp(-t/2)\}^d \\ \liminf_{n,p \rightarrow \infty} \left( \max_{1 \leq m \leq q} \hat{V}_m^2 \geq y_p \right) &\geq \sum_{d=1}^{2l} (-1)^{d-1} \frac{1}{d!} \{(8\pi)^{-1/2} \exp(-t/2)\}^d \end{aligned}$$

for any positive integer  $l$ . By letting  $l \rightarrow \infty$ , we obtain (A3) and Theorem 1 is proved.

### A-3. Proof of Theorem 2

Let  $M_n^1 = \max_{1 \leq i \leq j \leq p} \{T_{i,j,1} - T_{i,j,2} - (\omega_{i,j,1} - \omega_{i,j,2})\}^2 / (\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})$ . It follows from the proof of Theorem 1 that  $\text{pr}(M_n^1 \leq 4 \log p - 2^{-1} \log \log p) \rightarrow 1$ , as  $n, p \rightarrow \infty$ . By (A1), (A2) and the inequalities  $\max_{1 \leq i \leq j \leq p} (\omega_{i,j,1} - \omega_{i,j,2})^2 / (\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2}) \leq 2M_n^1 + 2M_n$ , and  $\max_{1 \leq i \leq j \leq p} (\omega_{i,j,1} - \omega_{i,j,2}) / (\hat{\theta}_{i,j,1} + \hat{\theta}_{i,j,2})^{1/2} \geq 4(\log p)^{1/2}$ , we have  $\text{pr}(M_n \geq q_\alpha + 4 \log p - \log \log p) \rightarrow 1$  as  $n, p \rightarrow \infty$ .

### A-4. Proof of Theorem 3

To prove the lower bound result, we first construct the worst case scenario to test between  $\Omega_1$  and  $\Omega_2$ , and then apply the arguments as shown in Baraud (2002).

Let  $\mathcal{M}$  denote the set of all subsets of  $\{1, \dots, p\}$  with cardinality  $p^r$ , for  $r < 1/2$ . Let  $\hat{m}$  be a random subset of  $\{1, \dots, p\}$ , which is uniformly distributed on  $\mathcal{M}$ . We construct a class of  $\Omega_1$ ,  $\mathcal{N} = \{\Omega_{\hat{m}}, \hat{m} \in \mathcal{M}\}$ , such that  $\omega_{i,j} = 0$  for  $i \neq j$  and  $1/\omega_{i,i} - 1 = \rho 1_{i \in \hat{m}}$  for  $i, j = 1, \dots, p$  and  $\rho = c(\log p/n)^{1/2}$ , where  $c > 0$  will be specified later. Let  $\Omega_2 = I$  and  $\Omega_1$  be uniformly distributed on  $\mathcal{N}$ . Let  $\mu_\rho$  be the distribution of  $\Omega_1 - I$ . Note that  $\mu_\rho$  is a probability measure on  $\{\Delta \in \mathcal{S}(p^r) : \|\Delta\|_F^2 = p^r \rho^2\}$ , where  $\mathcal{S}(p^r)$  is the class of matrices with  $p^r$  nonzero entries. Let  $d\text{pr}_1(\{X_n, Y_n\})$  and  $d\text{pr}_2(\{X_n, Y_n\})$  be the functions with precision matrices  $\Omega_1$  and  $\Omega_2$  respectively, likelihood then we have

$$L_{\mu_\rho} = L_{\mu_\rho}(\{X_n, Y_n\}) = E_{\mu_\rho} \left\{ \frac{d\text{pr}_1(\{X_n, Y_n\})}{d\text{pr}_2(\{X_n, Y_n\})} \right\},$$

where  $E_{\mu_\rho}(\cdot)$  is the expectation on  $\Omega_1$ . By the arguments in Baraud (2002), it suffices to show that  $E(L_{\mu_\rho}^2) \leq 1 + o(1)$ . It is easy to check that

$$L_{\mu_\rho} = E_{\hat{m}} \left[ \prod_{i=1}^n \frac{1}{|\Sigma_{\hat{m}}|^{1/2}} \exp \left\{ -\frac{1}{2} Z_i^\top (\Omega_{\hat{m}} - I) Z_i \right\} \right],$$

where  $\Sigma_{\hat{m}} = \Omega_{\hat{m}}^{-1}$  and  $Z_1, \dots, Z_n \stackrel{i.i.d.}{\sim} N(0, I)$ . Thus, we have

$$E(L_{\mu\rho}^2) = E \left( \left( \binom{p}{k_p} \right)^{-1} \sum_{m \in \mathcal{M}} \left[ \prod_{i=1}^n \frac{1}{|\Sigma_m|^{1/2}} \exp\{-Z_i^T(\Omega_m - I)Z_i/2\} \right] \right)^2$$

$$= \left( \binom{p}{k_p} \right)^{-2} \sum_{m, m' \in \mathcal{M}} E \left[ \prod_{i=1}^n \frac{1}{|\Sigma_m|^{1/2}} \frac{1}{|\Sigma_{m'}|^{1/2}} \exp\{-Z_i^T(\Omega_m + \Omega_{m'} - 2I)Z_i/2\} \right]$$

Set  $\Omega_m + \Omega_{m'} - 2I = (a_{ij})$ . It is easy to show that  $a_{ij} = 0$  for  $i \neq j$ ,  $a_{jj} = 0$  if  $j \in (m \cup m')^c$ ,  $a_{jj} = 2/(1 + \rho) - 1$  if  $j \in m \cap m'$  and  $a_{jj} = 1/(1 + \rho) - 1$  if  $j \in m \setminus m'$  or  $m' \setminus m$ . Let  $t = |m \cap m'|$ . Then

$$E(L_{\mu\rho}^2) = \left( \binom{p}{k_p} \right)^{-1} \sum_{t=0}^{k_p} \binom{k_p}{t} \binom{p-k_p}{k_p-t} \frac{1}{(1+\rho)^{k_p n}} (1+\rho)^{(k_p-t)n} \left( \frac{1+\rho}{1-\rho} \right)^{tn/2}$$

$$\leq p^{k_p} (p-k_p)! / p! \sum_{t=0}^{k_p} \binom{k_p}{t} \left( \frac{k_p}{p} \right)^t \left( \frac{1}{1-\rho^2} \right)^{tn/2} = \left\{ 1 + \frac{k_p}{p(1-\rho)^{n/2}} \right\}^{k_p} (1+o(1)),$$

for  $r < 1/2$ . Thus, by letting  $c$  be sufficiently small, we have

$$E(L_{\mu\rho}^2) \leq \exp\{k_p \log(1 + k_p p^{c^2-1})\} (1+o(1)) \leq \exp(k_p^2 p^{c^2-1}) (1+o(1)) = 1+o(1).$$

### A-5. Proof of Theorem 4

We first show that  $\hat{t}$ , as defined in Section 4, is obtained in the range  $(0, 2(\log p)^{1/2})$ . Then we illustrate that  $R_0(t)$ , defined in Section 4, is close to  $2\{1 - \Phi(t)\}|\mathcal{H}_0$  by first showing the terms in  $A_\tau$  are negligible. We then focus on the set  $\mathcal{H}_0 | A_\tau$  and prove the result based on Lemma 4.

Under the condition of Theorem 4, we have  $\sum_{1 \leq i < j \leq p} I\{|W_{i,j}| \geq 2(\log p)^{1/2}\} = [1/\{(8\pi)^{1/2} \alpha\} + \delta](\log_2 p)^{1/2}$ , with probability going to one. Hence we have with probability going to one,

$$\frac{(p^2 - p)/2}{\max\{\sum_{1 \leq i < j \leq p} I\{|W_{i,j}| \geq 2(\log p)^{1/2}\}, 1\}} \leq \frac{p^2 - p}{2} \left\{ \frac{1}{(8\pi)^{1/2} \alpha} + \delta \right\}^{-1} (\log_2 p)^{-1/2}.$$

Let  $t_p = (4 \log p - \log_2 p - \log_3 p)^{1/2}$ . Because  $1 - \Phi(t_p) \sim 1/\{(2\pi)^{1/2} t_p\} \exp(-t_p^2/2)$ , we have  $\text{pr}(1 \leq \hat{t} \leq t_p) \rightarrow 1$  according to the definition of  $\hat{t}$  in the false discovery rate control algorithm in Section 4. Note that, for  $0 \leq \hat{t} \leq t_p$ , we have

$$\frac{2\{1 - \Phi(\hat{t})\}(p^2 - p)/2}{\max\{\sum_{1 \leq i < j \leq p} I\{|W_{i,j}| \geq 2(\log p)^{1/2}\}, 1\}} = \alpha.$$

Thus to prove Theorem 4, it suffices to prove that

$\left| \sum_{(i,j) \in \mathcal{H}_0} \{I(|W_{i,j}| \geq t) - G(t)\} / \{q_0 G(t)\} \rightarrow 0\right.$  in probability, for  $0 < t \leq \{4 \log p + \alpha(\log p)\}^{1/2}$ , where  $G(t) = 2\{1 - \Phi(t)\}$ . Now we consider two cases.

1. If  $t = \{4 \log p + \alpha(\log p)\}^{1/2}$ , the proof of Theorem 1 yields that

$\text{pr}(\max_{(i,j) \in A_\tau} W_{i,j}^2 \geq t^2) = o(1)$ . Thus, it suffices to prove that

$\left| \sum_{(i,j) \in \mathcal{H}_0} \{I(|W_{i,j}| \geq t) - G(t)\} / \{q_0 G(t)\} \rightarrow 0\right.$  in probability. For  $(i,j) \in \mathcal{H}_0 \setminus A_\tau$ , we have from the proof of Theorem 1 that  $\max_{1 \leq i < j \leq p} |W_{i,j} - V_{i,j}| = o_p\{(\log p)^{-1/2}\}$ . Thus, it suffices to show that

$$\left| \frac{\sum_{(i,j) \in \mathcal{H}_0 \setminus A_\tau} \varepsilon_{i,j}(t)}{q_0 G(t)} \right| \rightarrow 0 \tag{A10}$$

in probability, where  $\varepsilon_{i,j}(t) = I(|V_{i,j}| \geq t) - G(t)$ .

2. If  $t = (C \log p)^{1/2}$  with  $C < 4$ , we have

$$\left| \frac{\sum_{(i,j) \in A_\tau \cap \mathcal{H}_0} \{I(|W_{i,j}| \geq t) - I(|V_{i,j}| \geq t)\}}{q_0 G(t)} \right| \leq \frac{2|A_\tau \cap \mathcal{H}_0|}{O(p^{2-C/2})} \rightarrow 0$$

in probability. Thus, it is again enough to show that

$$\left| \frac{\sum_{(i,j) \in \mathcal{H}_0 \setminus A_\tau} \varepsilon_{i,j}(t)}{q_0 G(t)} \right| \rightarrow 0 \tag{A11}$$

in probability. Define  $\tilde{\mathcal{H}}_0 = \mathcal{H}_0 \setminus A_\tau$ . Let  $0 = t_0 < \dots < t_m = t_p$  such that  $t_l - t_{l-1} = v_p$  for  $l = 1, \dots, m-1$  and  $t_m - t_{m-1} = v_p$ . Thus we have  $m \sim t_p/v_p$ . For any  $t$  such that  $t_{l-1} \leq t \leq t_l$ , we have

$$\frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} I(|V_{i,j}| \geq t_l)}{q_0 G(t_l)} \frac{G(t_l)}{G(t_{l-1})} \leq \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} I(|V_{i,j}| \geq t_l)}{q_0 G(t)} \leq \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} I(|V_{i,j}| \geq t_{l-1})}{q_0 G(t_{l-1})} \frac{G(t_{l-1})}{G(t)}$$

Thus it suffices to prove  $\max_{0 \leq l \leq m} \left| \sum_{(i,j) \in \tilde{\mathcal{H}}_0} \varepsilon_{i,j}(t_l) \right| / \{q_0 G(t_l)\} \rightarrow 0$  in probability. Note that

$$\begin{aligned} \text{pr} \left\{ \max_{0 \leq l \leq m} \left| \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} \varepsilon_{i,j}(t_l)}{q_0 G(t_l)} \right| \geq \varepsilon \right\} &\leq \sum_{l=1}^m \text{pr} \left\{ \left| \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} \varepsilon_{i,j}(t_l)}{q_0 G(t_l)} \right| \geq \varepsilon \right\} \\ &\leq \frac{1}{v_p} \int_0^{t_p} \text{pr} \left\{ \left| \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} \varepsilon_{i,j}(t)}{q_0 G(t)} \right| \geq \varepsilon \right\} dt + \sum_{l=m-1}^m \text{pr} \left\{ \left| \frac{\sum_{(i,j) \in \tilde{\mathcal{H}}_0} \varepsilon_{i,j}(t_l)}{q_0 G(t_l)} \right| \geq \varepsilon \right\}. \end{aligned}$$

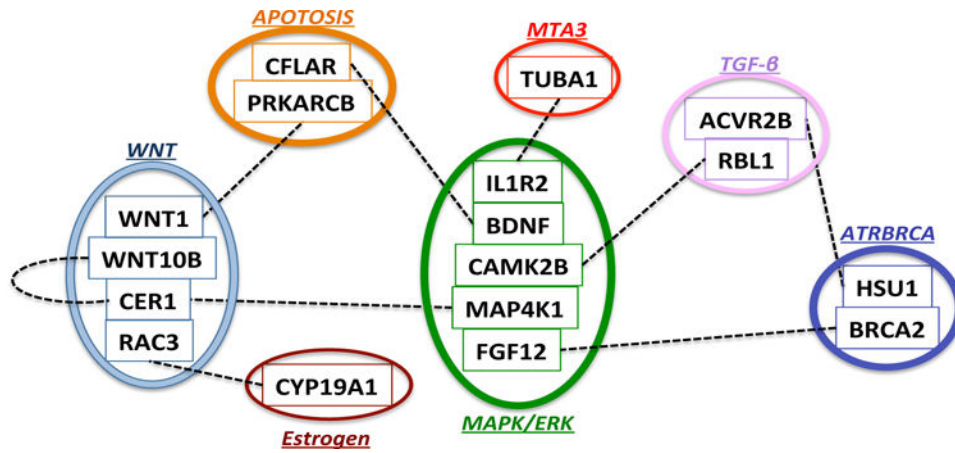
Thus by (A5) with  $d = 1$  and Lemma 4, Theorem 4 is proved.

## References

- Anderson, TW. An Introduction To Multivariate Statistical Analysis. 3rd. New York: Wiley-Interscience; 2003.
- Baraud Y. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*. 2002; 8:577–606.
- Breiman L. Random forests. *Mach Learn*. 2001; 45:5–32.
- Buck MB, Knabbe C. TGF-Beta signaling in breast cancer. *Ann N Y Acad Sci*. 2006; 1089:119–126. [PubMed: 17261761]
- Cai T, Liu W, Xia Y. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J Am Statist Assoc*. 2013; 108:265–277.
- Chapman J, Clayton D. Detecting association using epistatic information. *Genet Epidemiol*. 2007; 31:894–909. [PubMed: 17654599]
- Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet*. 2006; 79:1002–1016. [PubMed: 17186459]
- Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. *J R Statist Soc B*. 2014; 76:373–397.
- Dobra A, Hans C, Jones B, Nevins JR, Yao G, West M. Sparse graphical models for exploring gene expression data. *J Multivariate Anal*. 2004; 90:196–212.
- Downward J. Targeting RAS signalling pathways in cancer therapy. *Nat Rev Cancer*. 2003; 3:11–22. [PubMed: 12509763]
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010; 11:446–450. [PubMed: 20479774]
- Fan J, Lv J. Sure independence screening for ultra-high dimensional feature space (with discussion). *J R Statist Soc B*. 2008; 70:849–911.
- Gregg JP, Lit L, Baron CA, Hertz-Picciotto I, Walker W, Davis RA, Croen LA, Ozonoff S, Hansen R, Pessah IN, et al. Gene expression changes in children with autism. *Genomics*. 2008; 91:22–29. [PubMed: 18006270]
- Guardavaccaro D, Clevers H. Wnt/ $\beta$ -Catenin and MAPK Signaling: Allies and enemies in different battlefields. *Sci Signal*. 2012; 5 pe15.
- Hu VW, Sarachana T, Kim KS, Nguyen A, Kulkarni S, Steinberg ME, Luu T, Lai Y, Lee NH. Gene expression profiling differentiates autism case–controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Res*. 2009; 2:78–97. [PubMed: 19418574]
- Klaus A, Birchmeier W. Wnt signalling and its impact on development and cancer. *Nat Rev Cancer*. 2008; 8:387–398. [PubMed: 18432252]
- Kooperberg C, Leblanc M. Increasing the power of identifying gene  $\times$  gene interactions in genomewide association studies. *Genet Epidemiol*. 2008; 32:255–263. [PubMed: 18200600]
- Kooperberg C, Ruczinski I. Identifying interacting SNPs using Monte Carlo logic regression. *Genet Epidemiol*. 2005; 28:157–170. [PubMed: 15532037]
- Li J, Chen SX. Two sample tests for high-dimensional covariance matrices. *Ann Statist*. 2012; 40:908–940.
- Li KC, Palotie A, Yuan S, Bronnikov D, Chen D, Wei X, Choi OW, Saarela J, Peltonen L. Finding disease candidate genes by liquid association. *Genome Biol*. 2007; 8:R205. [PubMed: 17915034]
- Liu W. Gaussian graphical model estimation with false discovery rate control. *Ann Statist*. 2013; 41:2948–2978.
- Marchini J, Donnelly P, Cardon L. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet*. 2005; 37:413–417. [PubMed: 15793588]
- Mechanic L, Luke B, Goodman J, Chanock S, Harris C. Polymorphism Interaction Analysis (PIA): a method for investigating complex gene-gene interactions. *BMC Bioinformatics*. 2008; 9:146. [PubMed: 18325117]



- Moore J. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn.* 2004; 4:795–803. [PubMed: 15525222]
- Moro L, Arbin AA, Marra E, Greco M. Constitutive activation of MAPK/ERK inhibits prostate cancer cell proliferation through upregulation of BRCA2. *Int J Oncol.* 2007; 30:217–224. [PubMed: 17143532]
- Nathanson K, Wooster R, Weber B. Breast cancer genetics: what we know and what we need. *Nat Med.* 2001; 7:552–556. [PubMed: 11329055]
- Olopade O, Grushko T, Nanda R, Huo D. Advances in Breast Cancer: Pathways to Personalized Medicine. *Clin Cancer Res.* 2008; 14:7988. [PubMed: 19088015]
- Phillips PC. Epistasis: the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet.* 2008; 9:855–867. [PubMed: 18852697]
- Ritchie M, Hahn L, Roodi N, Bailey L, Dupont W, Parl F, Moore J. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet.* 2001; 69:138–147. [PubMed: 11404819]
- Santen RJ, Song RX, Mcpherson R, Kumar R, Adam L, Jeng MH, Yue W. The role of mitogen-activated protein (MAP) kinase in breast cancer. *J Steroid Biochem Mol Biol.* 2002; 80:239–256. [PubMed: 11897507]
- Schott JR. A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput Stat Data An.* 2007; 51:6535–6542.
- Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 2003; 34:166–176. [PubMed: 12740579]
- Shi Y, Massagué J. Mechanisms of TGF- $\beta$  signaling from cell membrane to the nucleus. *Cell.* 2003; 113:685–700. [PubMed: 12809600]
- Srivastava MS, Yanagihara H. Testing the equality of several covariance matrices with fewer observations than the dimension. *J Multivariate Anal.* 2010; 101:1319–1329.
- van de Vijver M, He Y, Van't Veer L, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med.* 2002; 347:1999–2009. [PubMed: 12490681]
- van't Veer LJ, Dai H, Van De Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, Van Der Kooy K, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002; 415:530–536. [PubMed: 11823860]
- Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell.* 2002; 108:171–182. [PubMed: 11832208]
- Wada T, Penninger JM. Mitogen-activated protein kinases in apoptosis regulation. *Oncogene.* 2004; 23:2838–2849. [PubMed: 15077147]
- You L, He B, Uematsu K, Xu Z, Mazieres J, Lee A, McCormick F, Jablons DM. Inhibition of wnt-1 signaling induces apoptosis in  $\beta$ -catenin-deficient mesothelioma cells. *Cancer Res.* 2004; 64:3474–3478. [PubMed: 15150100]
- Zaitsev AY. On the gaussian approximation of convolutions under multidimensional analogues of Bernstein's inequality conditions. *Probab Theory Rel.* 1987; 74:535–566.
- Zerba K, Ferrell R, Sing C. Complex adaptive systems and human health: the influence of common genotypes of the apolipoprotein E (ApoE) gene polymorphism and age on the relational order within a field of lipid metabolism traits. *Hum Genet.* 2000; 107:466–475. [PubMed: 11140944]



**Fig. 1.** Identified gene-by-gene interactions for the breast cancer example. The dashed lines between gene-pairs represent detected interactions. Genes inside each circle belong to the same pathway whose name is also shown.

**Table 1**

Empirical sizes and powers (%) for global testing with  $\alpha_1 = 0.05$ ,  $n_1 = n_2 = 100$ , and 1000 replications.

$p$	Model 1	Model 2	Model 3	Model 4
Size				
50	3.8	3.9	5.4	4.4
100	3.6	4.4	4.1	3.8
200	3.4	3.6	3.7	3.9
400	3.5	3.7	3.6	3.5
Power				
50	100	98.7	95.6	81.6
100	99.7	96.6	95.1	77.8
200	93.1	88.2	93.6	72.1
400	86.3	73.1	77.7	70.7

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Empirical power (%) for global testing under nearer alternatives.

$p$	Model 1	Model 2	Model 3	Model 4
Power under nearer alternative				
50	90.3	71.6	58.9	20.6
100	89.4	70.3	60.8	22.8
200	81.9	55.2	54.2	21.7
400	73.5	54.7	57.7	17.5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3** Empirical false discovery rate and power (%) with  $\alpha_2 = 0.1$ ,  $n_1 = n_2 = 100$ , and 100 replications.

$p$	Models 1, 2	Models 1, 3	Models 2, 3	Models 1, 4	Models 2, 4	Models 3, 4
Empirical False Discovery Rate						
50	10.5	11.0	12.6	12.2	11.5	10.2
100	9.5	10.0	12.1	11.8	11.4	9.5
200	9.7	10.4	11.2	11.7	11.6	10.3
Power						
50	67.9	65.6	35.7	55.0	30.2	26.1
100	64.2	38.3	19.3	51.4	25.1	18.2
200	61.1	20.6	17.1	46.1	21.7	11.3