HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2018 June 05.

Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(521): 328–339. doi:10.1080/01621459.2016.1251930.

Multiple Testing of Submatrices of a Precision Matrix with Applications to Identification of Between Pathway Interactions

Yin Xia[1], **Tianxi Cai**[2], and **T. Tony Cai**[3]

[1]Department of Statistics, Fudan University and Department of Statistics & Operations Research, University of North Carolina at Chapel Hill

[2]Department of Biostatistics, Harvard School of Public Health, Harvard University

[3]Department of Statistics, The Wharton School, University of Pennsylvania

Abstract

Making accurate inference for gene regulatory networks, including inferring about pathway by pathway interactions, is an important and difficult task. Motivated by such genomic applications, we consider multiple testing for conditional dependence between subgroups of variables. Under a Gaussian graphical model framework, the problem is translated into simultaneous testing for a collection of submatrices of a high-dimensional precision matrix with each submatrix summarizing the dependence structure between two subgroups of variables.

A novel multiple testing procedure is proposed and both theoretical and numerical properties of the procedure are investigated. Asymptotic null distribution of the test statistic for an individual hypothesis is established and the proposed multiple testing procedure is shown to asymptotically control the false discovery rate (FDR) and false discovery proportion (FDP) at the pre-specified level under regularity conditions. Simulations show that the procedure works well in controlling the FDR and has good power in detecting the true interactions. The procedure is applied to a breast cancer gene expression study to identify between pathway interactions.

Keywords

Between pathway interactions; conditional dependence; covariance structure; false discovery proportion; false discovery rate; Gaussian graphical model; multiple testing; precision matrix; testing submatrices

1 Introduction

Simultaneous inference for the interactions among a large number of variables is an important problem in statistics with a wide range of applications. Many statistical methods have been proposed to infer about pairwise interactions (Ritchie et al., 2001; Chatterjee et al., 2006; Kooperberg and Ruczinski, 2005; Kooperberg and LeBlanc, 2008; Fan and Lv, 2008; Cai and Zhang, 2014; Cai and Liu, 2015, e.g). Most of the existing methods focus on marginal assessments of pairwise interactions without conditioning on the other variables. Such marginal methods may result in false identification of interactions due to the discrepancy between conditional and unconditional effects. When prior knowledge is available to group the variables of interest, it is often of interest to make simultaneous

inference for the interactions at the group level. For example, functionally related genes are often grouped into pathways and inferring about between pathway interactions is important as they represent a majority of the genetic interactions (Kelley and Ideker, 2005).

Motivated by applications in genomics, in this paper we propose methods to efficiently identify between group interactions while accounting for the joint effects from all other variables of interest. Under a Gaussian graphical model framework, we translate the problem of detecting between group interactions into the statistical problem of simultaneous testing of a collection of submatrices of a high-dimensional precision matrix. We first discuss the motivating problem of detecting between pathway interactions before presenting the framework for large-scale multiple testing of submatrices of a high-dimensional precision matrix.

1.1 Detection of Between Pathway Interactions

It is well known that genes interact functionally in networks to orchestrate cellular processes. Biological interactions of genes are often inferred based on co-expression networks since coexpressed genes tend to be functionally related or controlled by the same transcriptional regulatory elements (Weirauch, 2011). Throughout, we use the term gene-gene interaction to refer to *their biological interaction*, quantified by conditional co-expression (given all other genes), rather than statistical interaction unless specified otherwise. Accurately identifying important gene-gene interactions is a difficult task due to the high dimensionality of the feature space spanned by gene pairs. Particularly in genome-wide studies where the sample sizes are typically small compared to the number of interactions of interest, gene level analyses often produce results that are difficult to interpret or replicate.

One approach to improve the interpretability and reproducibility is to incorporate prior biological knowledge such as gene structure or protein-protein interaction network information to group functionally related genes into pathways and perform analysis at the pathway level. Throughout, we use the term *pathway* to refer generically a gene group under study, whether or not the group is indeed representing a metabolic or signaling pathway. A large number of knowledge bases have become available to assemble biologically meaningful gene groups (Xenarios et al., 2002; Rual et al., 2005; Matthews et al., 2009; Craven and Kumlien, 1999; Khatri et al., 2012). The knowledge bases provide prior information on biological processes, components, or structures in which individual genes and proteins are involved in. Analyzing high-throughput molecular measurements at the functional level is very appealing due to its potential in reducing the complexity of the problem and improving the power (Subramanian et al., 2005; Glazko and Emmert-Streib, 2009).

Detecting pathway level interactions is also biologically relevant because in order to produce appropriate physiological responses to both internal and external factors, pathways often need to function in a coordinated fashion due to the complex nature of biological systems. In addition, there is accumulating evidence that complex traits are often influenced by multiple groups of functional related genes through their dynamic interaction and coregulation (Jia et al., 2011). Therefore, the knowledge of pathway crosstalk network is helpful for inferring

the function of complex biological systems (Li et al., 2008). A wide range of between pathway crosstalk have been identified as critical for understanding many diseases including breast cancer, lung cancer, ovarian cancer, major depression disorder, and Alzheimer (Osborne et al., 2005; Shou et al., 2004; Jia et al., 2011; Liu et al., 2010; Pan, 2012; Puri et al., 2008).

In addition to applications to the identification of between genetic pathway interactions, the proposed procedures are also useful for other settings. Examples include interactions between biological markers when markers are measured at different time points with multiple measurements of each marker representing one group; and interactions between different brain regions when functional MRI measurements are taken over the entire brain with groups indexed by brain regions. We next describe our proposed framework for detecting between pathway interactions based on testing for submatrices of a high-dimensional precision matrix.

1.2 Multiple Testing of Submatrices of A Precision Matrix

Under a Gaussian graphical model framework, we formulate the problem of identifying between group interactions that account for joint effects from all genes of interest as the statistical problem of simultaneous testing of submatrices of a high-dimensional precision matrix. Let $\{\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n\}$ be a random sample consisting of n independent copies of a p dimensional Gaussian random vector $\boldsymbol{X} \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$. The precision matrix, which is the inverse of $\boldsymbol{\Sigma}$, is denoted by $\boldsymbol{\Omega} = (\omega_{i,j})$. It is well-known that the precision matrix is closely connected to the corresponding Gaussian graph $G = (V,E)$, which represents the conditional independence between components of $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$. Here V is the vertex set consisting of the p components $X_1, \ldots, X_p$ and E is the edge set consisting of ordered pairs (i, j), where $(i, j) \in E$ if there is an edge between X_i and X_j, indicating that X_i and X_j are conditionally dependent given $\{X_k, k \neq i, j\}$. It is a well-known fact that the conditional independence between X_i and X_j given all other variables is equivalent to $\omega_{i,j} = 0$. See, e.g., Lauritzen (1996).

Let $\mathcal{J}_1, \ldots, \mathcal{J}_M \subset \{1, \ldots, p\}$ be a collection of prespecified non-overlapping sets which index group memberships (e.g. pathway membership), we wish to test simultaneously the hypotheses of the conditional independence between any two gene groups given all remaining genes in the collection with proper control of the false discovery rate (FDR) and false discovery proportion (FDP) asymptotically. It follows from the above discussion that this multiple testing problem can be equivalently formulated as testing the hypotheses on the submatrices of the precision matrix $\boldsymbol{\Omega}$,

$$H_{0,m,h}: \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h} = 0 \text{ versus } H_{1,m,h}: \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h} \neq 0,\ 1 \leq m < h \leq M, \quad (1)$$

while controlling the FDR and FDP asymptotically. Hereafter, all results related to the FDR and FDP are studied in the asymptotic regime and we use FDR and FDP as simplifications for the expressions of asymptotic FDR and asymptotic FDP.

Simultaneous testing of between group interactions with FDR control is technically challenging, both in constructing a suitable test statistic and establishing its null distribution for testing the interactions between any two given groups and in developing a multiple testing procedure that accounts for the multiplicity and dependency with FDR control. To the best of our knowledge, there are no currently available methods with theoretical guarantees to infer about interactions between pre-specified gene groups that adjust for effects from a large number of other genes. Furthermore, no existing methods allow the testing for such group level interactions while properly controlling a desired FDR. Liu (2013) proposed a multiple testing procedure with the FDR control for the partial correlations under a Gaussian graphical model. Xia et al. (2015) considered the problem of identifying gene-by-gene interactions associated with a binary trait under a two-sample framework and proposed a procedure for testing the differential network by simultaneously testing entry-wise hypotheses with FDR control. These methods, which can identify the locations of individual gene-by-gene interactions, are however unable to detect the presence interactions between pairs of gene groups while controlling the FDR at the group level.

In this paper, we propose a novel multiple testing procedure for between group interactions that controls the FDR and FDP asymptotically at any pre-specified level $0 < \alpha < 1$. The simultaneous testing procedure is developed in two steps. In the first step, we construct a test statistic for testing the conditional independence of a given pair of variable groups $\mathcal{J}_m$ and $\mathcal{J}_h$, $H_{0,m,h}: \boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h} = 0$, with $m \neq h$. The test statistic is based on the Frobenius norm of a standardized submatrix estimate with unknown correlation structure. The estimation of this dependency structure is technically challenging, because correlations among the estimates of the entries of $\boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h}$ not only depend on the entries within the submatrix, but also largely depend on the entries outside of it. To incorporate this dependency structure, we estimate the eigenvalues of the correlation matrix of the entry estimates of a given submatrix $\boldsymbol{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h}$ through a Kronecker product by estimating the eigenvalues of two partial correlation submatrices $\boldsymbol{R}_{\mathcal{J}_m \times \mathcal{J}_m}$ and $\boldsymbol{R}_{\mathcal{J}_h \times \mathcal{J}_h}$ of $\boldsymbol{R} = \boldsymbol{D}^{-1/2}\boldsymbol{\Omega}\boldsymbol{D}^{-1/2}$, where $\boldsymbol{D}$ is the diagonal matrix of $\boldsymbol{\Omega}$. It is shown that the test statistic has asymptotically the same limiting null distribution as a mixture of χ^2_1 with the estimated correlation structure.

In the second step, we construct a simultaneous testing procedure based on these test statistics. A major difficulty here is that the correlation structures of the entry estimates vary across different submatrices. Consequently the limiting null distributions of the test statistics for different submatrices are different. We introduce a normal quantile transformation for each test statistic, and the transformed test statistics are shown to have asymptotically the same distribution as the absolute value of a standard normal random variable under the null. Based on them, we develop a multiple testing procedure to account for the multiplicity in testing a large number of hypotheses so that the overall FDR and FDP are controlled.

Both the theoretical and numerical properties of the proposed procedure are investigated. The theoretical results show that, under regularity conditions, the proposed procedure asymptotically controls both the overall FDR and FDP at the pre-specified level. As a comparison, it is discussed in Section 4.3 that a direct application of the well-known B-H procedure (Benjamini and Hochberg, 1995) to the individual test statistics is not able to

control the FDP when the number of true alternatives is fixed. Simulation studies are carried out to examine the numerical performance of the multiple testing procedure in various settings. The results show that the procedure performs well numerically in terms of both the size and power of the test. We also consider a simulation setting that is similar to the breast cancer gene expression data analyzed in this paper by mimicking the true sizes of the gene groups in the breast cancer study. The result shows that the FDR is well controlled and this new group level based method significantly outperforms the alternative procedures.

Finally, we apply the proposed procedure to assess the between pathway interactions in a breast cancer gene expression study. Many of the identified interactions are consistent with those reported in the literature.

1.3 Structure of the Paper

The rest of the paper is organized as follows. We give a detailed construction of the statistic for testing a specific submatrix of a precision matrix in Section 2. The limiting null distribution of the test statistic and the theoretical properties of the testing procedure are obtained in Section 3. A multiple testing procedure for simultaneously assessing a collection of submatrices is proposed and its theoretical properties are established in Section 4. Simulation results demonstrating the performance of the proposed methods in finite sample are given in Section 5. In Section 6, we apply the new multiple testing procedure to a breast cancer gene expression study to identify between pathway interactions. A discussion on possible extensions is given in Section 7. All proofs are contained in the supplement Xia et al. (2016).

2 Testing A Given Submatrix

We consider in this section testing a given submatrix of the precision matrix $\mathbf{\Omega}$,

$$H_0: \mathbf{\Omega}_{\mathcal{I} \times \mathcal{J}} = 0 \quad \text{versus} \quad H_1: \mathbf{\Omega}_{\mathcal{I} \times \mathcal{J}} \neq 0, \quad (2)$$

under the framework of Section 1.2, where $\mathcal{I}$ and $\mathcal{J}$ index two non-overlapping gene groups. A rejection of H_0 means that at least one pair of variables from $\mathcal{I}$ and $\mathcal{J}$ are not conditionally independent from each other given all other variables. As the group information is considered as prior knowledge, performing analysis at the group level is more appealing than the entrywise procedure as discussed in Section 1. We shall construct a test statistic for H_0, corresponding to no interactions between gene groups $\mathcal{I}$ and $\mathcal{J}$ conditional on all other genes. Related works on testing for independence and conditional independence between random vectors can be found in, e.g., Gieser and Randles (1997); Um and Randles (2001); Beran et al. (2007); Su and White (2007, 2008); and Huang et al. (2010).

2.1 Notation and Definitions

Denote $\boldsymbol{A} \otimes \boldsymbol{B}$ the Kronecker product of matrix $\boldsymbol{A}$ and $\boldsymbol{B}$. For a vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}_p$, define the ℓ_q norm by $| \boldsymbol{\beta} |_q = (\sum_{i=1}^{p} | \beta_i |^q)^{1/q}$ for $1 \leq q \leq \infty$. For any vector $\boldsymbol{\mu}$ with

dimension $p \times 1$, let $\boldsymbol{\mu}_{-i}$ denote the $(p-1) \times 1$ vector by removing the i^{th} entry from $\boldsymbol{\mu}$. For a symmetric matrix $\boldsymbol{A}$, let $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the largest and smallest eigenvalues of $\boldsymbol{A}$. For any $p \times q$ matrix $\boldsymbol{A}$, $\boldsymbol{A}_{i,-j}$ denotes the i^{th} row of $\boldsymbol{A}$ with its j^{th} entry removed and $\boldsymbol{A}_{-i,j}$ denotes the j^{th} column of $\boldsymbol{A}$ with its i^{th} entry removed. $\boldsymbol{A}_{-i,-j}$ denotes the $(p-1)\times(q-1)$ submatrix of $\boldsymbol{A}$ with its i^{th} row and j^{th} column removed. $\boldsymbol{A}_{r\times c}$ denotes the submatrix of $\boldsymbol{A}$ corresponding to the row vector r and column vector c. For a $n\times p$ data matrix $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n)^\top$, denote an $n \times (p-1)$ matrix $\boldsymbol{U}_{\cdot,-i} = (\boldsymbol{U}_{1,-i}^\top, \ldots, \boldsymbol{U}_{n,-i}^\top)^\top$. Let $\bar{\boldsymbol{U}}_{\cdot,-i} = 1/n \sum_{k=1}^{n} \boldsymbol{U}_{k,-i}$ with dimension $1 \times (p-1)$, $\boldsymbol{U}_{(i)} = (U_{1,i}, \ldots, U_{n,i})^\top$ with dimension $n \times 1$, $\bar{\boldsymbol{U}}_{(i)} = (\bar{U}_i, \ldots, \bar{U}_i)^\top$ with dimension $n \times 1$, where $\bar{U}_i = 1/n \sum_{k=1}^{n} U_{k,i}$, and $\bar{\boldsymbol{U}}_{(\cdot,-i)} = (\bar{\boldsymbol{U}}_{\cdot,-i}^\top, \ldots, \bar{\boldsymbol{U}}_{\cdot,-i}^\top)^\top$ with dimension $n \times (p-1)$. For a matrix $\boldsymbol{\Omega} = (\omega_{i,j})_{p\times p}$, the matrix 1-norm is defined by $\|\boldsymbol{\Omega}\|_{L_1} = \max_{1 \le j \le p} \sum_{i=1}^{p} |\omega_{i,j}|$ and the matrix element wise infinity norm is defined to be $\|\boldsymbol{\Omega}\|_\infty = \max_{1\le i,j\le p} |\omega_{i,j}|$. For a set $\mathcal{H}$, denote $|\mathcal{H}|$ 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 $|a_n| \le C|b_n|$ holds for all n, write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$, and write $a_n \asymp b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

2.2 Testing Procedure

We shall first define a standardized estimate $W_{i,j}$ for each individual entry of the precision matrix, which is the one-sample version of the estimates proposed in Xia et al. (2015), then propose a novel test statistic $S_{\mathcal{G}\times\mathcal{J}}$ based on the sum of all possible $W_{i,j}^2$, for $(i,j) \in \mathcal{G} \times \mathcal{J}$.

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

$$X_{k,i} = \alpha_i + \boldsymbol{X}_{k,-i}\boldsymbol{\beta}_i + \varepsilon_{k,i}, \quad 1 \le k \le n, \quad (3)$$

where $\varepsilon_{k,i} \sim N(0, \sigma_{i,i} - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\Sigma}_{-i,i})$ is independent of $\boldsymbol{X}_{k,-i}$, and $\alpha_i = \mu_i - \boldsymbol{\Sigma}_{i,-i}\boldsymbol{\Sigma}_{-i,-i}^{-1}\boldsymbol{\mu}_{-i}$. The regression coefficient vector $\boldsymbol{\beta}_i$ and the error terms $\varepsilon_{k,i}$ satisfy

$$\boldsymbol{\beta}_i = -\omega_{i,i}^{-1}\boldsymbol{\Omega}_{-i,i} \quad \text{and} \quad r_{i,j} \equiv \mathrm{Cov}(\varepsilon_{k,i}, \varepsilon_{k,j}) = \omega_{i,j}/(\omega_{i,i}\omega_{j,j}).$$

As in Xia et al. (2015), we first develop an estimator of $\omega_{i,j}$ and then base the test on its bias corrected standardization. We begin by constructing estimators of $r_{i,j}$.

Let $\hat{\boldsymbol{\beta}}_i = (\hat{\beta}_{1,i}, \cdots, \hat{\beta}_{p-1,i})^\top$ be estimators of $\boldsymbol{\beta}_i$ satisfying max

$$\max_{1 \le i \le p} | \hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i |_1 = o_P\{(\log p)^{-1}\}, \quad (4)$$

$$\max_{1 \le i \le p} | \hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i |_2 = o_P\left\{(n \log p)^{-1/4}\right\}. \quad (5)$$

Such estimators can be obtained easily via the standard methods such as the Lasso and Dantzig Selector, see, e.g., Xia et al. (2015) Section 2.3. Specifically, if we use the Lasso estimator (see (18) in Section 5), then equations (4) and (5) can be satisfied under the condition (C1) in Section 3 and the sparsity condition $\max_{1 \le i \le p} |\boldsymbol{\beta}_i|_0 = o\{n^{1/2}/(\log p)^{3/2}\}$.

Define the fitted residuals by

$$\hat{\varepsilon}_{k,i} = X_{k,i} - \bar{X}_i - (\boldsymbol{X}_{k,-i} - \bar{\boldsymbol{X}}_{-i})\hat{\boldsymbol{\beta}}_i,$$

where $\bar{X}_i = \frac{1}{n}\sum_{k=1}^n X_{k,i}$, $\bar{\boldsymbol{X}}_{-i} = \frac{1}{n}\sum_{k=1}^n \boldsymbol{X}_{k,-i}$. A natural estimator of $r_{i,j}$ is the sample covariance between the residuals

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

However, when $i \neq j$, $\tilde{r}_{i,j}$ tends to be biased due to the correlation induced by the estimated parameters. Xia et al. (2015) proposed a bias corrected estimator of $r_{i,j}$ as

$$\hat{r}_{i,j} = -(\tilde{r}_{i,j} + \tilde{r}_{i,i}\hat{\beta}_{i,j} + \tilde{r}_{j,j}\hat{\beta}_{j-1,i}), \text{ for } 1 \le i < j \le p.$$

For $i = j$, we let $\tilde{r}_{i,i} = \tilde{r}_{i,i}$, which is a nearly unbiased estimator of $r_{i,i}$. For $1 \le i < j \le p$, a natural estimator of $\omega_{i,j}$ can then be defined by

$$T_{i,j} = \hat{r}_{i,j}/(\hat{r}_{i,i} \cdot \hat{r}_{j,j}).$$

Since $\{T_{i,j}, 1 \le i < j \le p\}$ are heteroscedastic and can possibly have a wide range of variability, we shall first standardize $T_{i,j}$. To estimate its variance, note that

$$\theta_{i,j} \equiv \mathrm{Var}(\varepsilon_{k,i}\varepsilon_{k,j}/(r_{i,i}r_{j,j}))/n = (1 + \rho_{i,j}^2)/(nr_{i,i}r_{j,j}),$$

where $\rho_{i,j}^2 = \beta_{i,j}^2 r_{i,i}/r_{j,j}$. Then $\theta_{i,j}$ can be estimated by $\hat{\theta}_{i,j} = (1 + \hat{\beta}_{i,j}^2 \hat{r}_{i,i}/\hat{r}_{j,j})/(n\hat{r}_{i,i}\hat{r}_{j,j})$.

Define the standardized statistics

$$W_{i,j} = T_{i,j}/(\hat{\theta}_{i,j})^{1/2}, \text{ for } 1 \le i < j \le p. \quad (7)$$

Finally, we propose the following test statistic for testing a given submatrix $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}}$,

$$S_{\mathcal{I}\times\mathcal{J}} = \sum_{(i,j)\in\mathcal{I}\times\mathcal{J}} W_{i,j}^2. \quad (8)$$

We detail in Section 3 statistical properties of the proposed test statistic.

3 Theories on Testing A Given Submatrix

In this section, we investigate the theoretical properties including the limiting null distribution and the asymptotic power. We first show that the null distribution of $S_{\mathcal{G}\times\mathcal{J}}$ converges to the distribution of a mixture of χ_1^2 variables as $(n, p) \to \infty$ and then demonstrate that the test based on $S_{\mathcal{G}\times\mathcal{J}}$ is powerful under a large collection of alternatives.

3.1 Asymptotic Null Distribution

Before studying the null distribution of $S_{\mathcal{G}\times\mathcal{J}}$, we first introduce the following condition on the eigenvalues of $\mathbf{\Omega}$, which is a common assumption in the high-dimensional setting (Cai et al., 2013; Xia et al., 2015; Liu, 2013).

(C1) Assume that $\log p = o(n^{1/5})$, and for some constant $C^0 > 0$, $C_0^{-1} \le \lambda_{\min}(\mathbf{\Omega}) \le \lambda_{\max}(\mathbf{\Omega}) \le C_0$. Suppose $|\mathcal{J}_m|$ does not depend on n and p for $1 \le m \le M$.

Let $\boldsymbol{D}$ be the diagonal of $\mathbf{\Omega}$ and let $(\eta_{i,j}) =: \boldsymbol{R} = \boldsymbol{D}^{-1/2}\mathbf{\Omega}\boldsymbol{D}^{-1/2}$. Under H_0, for (i_1, j_1), $(i_2, j_2) \in \mathcal{G} \times \mathcal{J}$, the covariance between the standardized statistics W_{i_1,j_1} and W_{i_2,j_2}, as defined in (7), is approximately equal to $\eta_{i_1,i_2}\eta_{j_1,j_2}$, and thus can be estimated by $\tilde{T}_{i_1,i_2}\tilde{T}_{j_1,j_2}$, where $\tilde{T} := (\tilde{T}_{i,j})_{p\times p}$ with $\tilde{T}_{i,j} = \hat{r}_{i,j}/\sqrt{\hat{r}_{i,i}\hat{r}_{j,j}}$. Thus, we shall estimate the covariance matrix of $\{W_{i,j}, (i,j) \in \mathcal{G}\times\mathcal{J}\}$ by the Kronecker product of $\tilde{T}_{\mathcal{G}\times\mathcal{G}}$ and $\tilde{T}_{\mathcal{J}\times\mathcal{J}}$. Let $\widehat{\mathbf{\Lambda}}_{\mathcal{I}} = (\hat{\lambda}_1^{\mathcal{I}}, \ldots, \hat{\lambda}_{|\mathcal{I}|}^{\mathcal{I}})^\top$ and $\widehat{\mathbf{\Lambda}}_{\mathcal{J}} = \left\{\hat{\lambda}_1^{\mathcal{J}}, \ldots, \hat{\lambda}_{|\mathcal{J}|}^{\mathcal{J}}\right\}^\top$ be the eigenvalues of $\tilde{T}_{\mathcal{G}\times\mathcal{G}}$ and $\tilde{T}_{\mathcal{J}\times\mathcal{J}}$ respectively. We then estimate the eigenvalues of the covariance matrix of $\{W_{i,j}, (i,j) \in \mathcal{G}\times\mathcal{J}\}$ by $\widehat{\mathbf{\Lambda}}^{\mathcal{I}\times\mathcal{J}} = (\hat{\lambda}_1^{\mathcal{I}\times\mathcal{J}}, \ldots, \hat{\lambda}_K^{\mathcal{I}\times\mathcal{J}})^\top$ which is the vectorized $\tilde{\Lambda}_{\mathcal{G}} \otimes \tilde{\Lambda}_{\mathcal{J}}$, where $K = |\mathcal{G}||\mathcal{J}|$. The following theorem states the asymptotic null distributions for $S_{\mathcal{G}\times\mathcal{J}}$.

Theorem 1—Suppose that (*C1*), (4) and (5) hold. Then under H_0: $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}} = 0$, for any given $t \in \mathbb{R}$, we have

$$\frac{P(S_{\mathcal{I}\times\mathcal{J}} \le t)}{P(\sum_{l=1}^{K} \hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}} Z_l^2 \le t)} \to 1, \quad (9)$$

as $(n, p) \to \infty$, where $(Z_1, \ldots, Z_K) \sim N(0, \mathbf{I}_{K\times K})$.

Remark 1—The difficulty of Theorem 1 comes from the fact that, though $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}} = 0$ under the null, the entries $\{e_{k,i}e_{k,j}, (i, j) \in \mathcal{G}\times\mathcal{J}\}$ can still be highly dependent with each other and their correlations depend on the entries outside of submatrix $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}}$. Thus, the distribution of $S_{\mathcal{G}\times\mathcal{J}}$ cannot be simply estimated by the chi-square distribution. Actually, if we use the chi-square approximation in the following FDR control procedure in Section 4, the choice of threshold level of each statistic will be too conservative and as the result the FDR cannot be controlled at the pre-specified level α, i.e., the FDR will be much larger than α.

It has been shown in the above theorem that $S_{\mathcal{G}\times\mathcal{J}}$ has different asymptotic distribution for different submatrix $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}}$. Thus, we introduce the normal quantile transformation of $S_{\mathcal{G}\times\mathcal{J}}$ as follows

$$N_{\mathcal{I}\times\mathcal{J}} = \Phi^{-1}\left\{1 - \mathrm{P}\left(\sum_{l=1}^{K} \hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}} Z_l^2 \ge S_{\mathcal{I}\times\mathcal{J}}\right)/2\right\},$$

where $\Phi(t) = \mathrm{P}(N(0, 1) \le t)$ is standard normal cumulative distribution function (cdf) and $S_{\mathcal{G}\times\mathcal{J}}$ is the observed value. Thus, we have $\mathrm{P}(|N(0, 1)| \ge N_{\mathcal{I}\times\mathcal{J}}) = \mathrm{P}(\sum_{l=1}^{K} \hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}} Z_l^2 \ge S_{\mathcal{I}\times\mathcal{J}})$. Since asymptotically $S_{\mathcal{G}\times\mathcal{J}}$ and $\sum_{l=1}^{K} \hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}} Z_l^2$ have the same distribution as studied in Theorem 1, thus $N_{\mathcal{G}\times\mathcal{J}}$ asymptotically has the same distribution as the absolute value of a standard normal random variable. We then define the test $\Phi_\alpha^{\mathcal{I}\times\mathcal{J}}$ by

$$\Phi_\alpha^{\mathcal{I}\times\mathcal{J}} = I\left\{N_{\mathcal{I}\times\mathcal{J}} \ge \Phi^{-1}(1-\alpha)\right\}. \quad (10)$$

The hypothesis H_0: $\mathbf{\Omega}_{\mathcal{G}\times\mathcal{J}} = 0$ is rejected whenever $\Phi_\alpha^{\mathcal{I}\times\mathcal{J}} = 1$.

Remark 2—The eigenvalues $\{\hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}}, l = 1, \ldots, K\}$ are calculated based on $\tilde{T}_{\mathcal{G}\times\mathcal{G}}$ and $\tilde{T}_{\mathcal{J}\times\mathcal{J}}$ as described earlier. Given the values of $\{\hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}}, l = 1, \ldots, K\}$, the distribution of the

mixture of χ_1^2 variables $\sum_{l=1}^{K} \hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}} Z_l^2$ can be approximated by a non-central chi-squared distribution with the parameters determined by the first four cumulants of the quadratic form, see, e.g., Liu et al. (2009). We will use this approximation in our numerical studies.

3.2 Asymptotic Power

We now turn to analyze the power of the test $\Phi_\alpha^{\mathcal{I}\times\mathcal{J}}$ given in (10). For a given pair of index sets $\mathcal{I}$ and $\mathcal{J}$, we shall first define the following class of precision matrices

$$\mathcal{W}_{\mathcal{I}\times\mathcal{J}}(\alpha,\beta)=\left\{\mathbf{\Omega}:\sum_{(i,j)\in\mathcal{I}\times\mathcal{J}}\frac{\omega_{i,j}^2}{\theta_{i,j}}\geq(2+\delta)(\Psi_{1-\alpha}^2+\Psi_{1-\beta}^2)\right\},\quad(11)$$

for any $\delta>0$, where $\Psi_{1-\alpha}$ is the $1-\alpha$ quantile of $\sum_{l=1}^{K}\hat{\lambda}_l^{\mathcal{I}\times\mathcal{J}}Z_l^2$ as defined in Theorem 1.

The next theorem shows that the test $\Phi_\alpha^{\mathcal{I}\times\mathcal{J}}$ is able to asymptotically distinguish the null parameter set in which $\mathbf{\Omega}_{\mathcal{I}\times\mathcal{J}}=0$ from $\mathcal{W}_{\mathcal{I}\times\mathcal{J}}(\alpha,\beta)$ for arbitrarily small constant $\delta>0$, with $\beta\to 0$.

Theorem 2—Suppose that (*C1*), (4) and (5) hold. Then we have, for any constant $\delta>0$,

$$\inf_{\mathbf{\Omega}\in\mathcal{W}_{\mathcal{I}\times\mathcal{J}}(\alpha,\beta)}P\left(\Phi_\alpha^{\mathcal{I}\times\mathcal{J}}=1\right)\geq 1-\beta,\quad as\ n,p\to\infty\ .\quad(12)$$

Since $\theta_{i,j}$ is of order $1/n$, Theorem 2 shows that the proposed test rejects the null hypothesis H_0: $\mathbf{\Omega}_{\mathcal{I}\times\mathcal{J}}=0$ with high probability for a large class of precision matrices satisfying the condition that there exists one entry of the submatrix $\mathbf{\Omega}_{\mathcal{I}\times\mathcal{J}}$ having a magnitude larger than $C/n^{1/2}$ for $C=\{2(2+\delta)C_0^2(\Psi_{1-\alpha}^2+\Psi_{1-\beta}^2)\}^{1/2}$, where C_0 is given in Condition (C1).

4 Multiple Testing of Submatrices with FDR Control

In practice, there are typically many pathways under investigation and it is often of significant interest to identify which pairs of the pathways interact with each other. A natural approach to investigate interactions among the M pathways, indexed by $\{\mathcal{J}_m, m=1,\ldots,M\}$, is to carry out simultaneous testing of

$$H_{0,m,h}:\mathbf{\Omega}_{\mathcal{J}_m\times\mathcal{J}_h}=0\ \text{versus}\ H_{1,m,h}:\mathbf{\Omega}_{\mathcal{J}_m\times\mathcal{J}_h}\neq 0,\quad\text{for}\ 1\leq m<h\leq M,\quad(13)$$

where $\mathcal{J}_1, \ldots, \mathcal{J}_M \subset \{1, \ldots, p\}$ is a collection of pre-specified non-overlapping index sets. In this section, we introduce a multiple testing procedure with FDR and FDP control for testing a collection of $\mathcal{M} = M(M-1)/2$ hypotheses, and we shall assume that $\mathcal{M}$ is large. Let L_m denote the cardinality of $\mathcal{J}_m$ assumed to be independent of n or p for $1 \le m \le M$. Let $\mathcal{H} = \{(m, h): 1 \le m < h \le M\}$, $\mathcal{H}_0 = \{(m, h): \mathbf{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h} = 0, 1 \le m < h \le M\}$ be the set of true nulls and $\mathcal{H}_1 = \mathcal{H} \backslash \mathcal{H}_0$ be the set of true alternatives. We shall assume that $|\mathcal{H}_1|$ is relatively small compared to $|\mathcal{H}|$, and this assumption arises frequently in many contemporary applications.

4.1 Multiple Testing Procedure

Recall that the standardization of $T_{i,j}$ is defined by $W_{i,j} = T_{i,j}/(\hat{\theta}_{i,j})^{1/2}$ as in (7), and the test statistic $S_{\mathcal{J}_m \times \mathcal{J}_h}$ is defined based on $W_{i,j}$ as in (8). It has been shown in Theorem 1 that $S_{\mathcal{J}_m \times \mathcal{J}_h}$ has different asymptotic null distribution for different submatrix $\mathbf{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h}$. Thus, as discussed in Section 3.1, the normal quantile transformation of $S_{\mathcal{J}_m \times \mathcal{J}_h}$ is defined by

$$N_{\mathcal{J}_m \times \mathcal{J}_h} = \Phi^{-1}\left\{1 - \mathrm{P}\left(\sum_{l=1}^{L_m L_h} \tilde{\lambda}_l^{\mathcal{J}_m \times \mathcal{J}_h} Z_l^2 \ge S_{\mathcal{J}_m \times \mathcal{J}_h}\right)/2\right\},$$

and $N_{\mathcal{J}_m \times \mathcal{J}_h}$ approximately has the same distribution as the absolute value of a standard normal random variable under the null $H_{0,m,h}$. Let t be the threshold level such that $H_{0,m,h}$ is rejected if $N_{\mathcal{J}_m \times \mathcal{J}_h} \ge t$. For any given t, denote the total number of false positives by

$$R_0(t) = \sum_{(m,h) \in \mathcal{H}_0} I\{N_{\mathcal{J}_m \times \mathcal{J}_h} \ge t\}, \quad (14)$$

and the total number of rejections by

$$R(t) = \sum_{(m,h) \in \mathcal{H}} I\{N_{\mathcal{J}_m \times \mathcal{J}_h} \ge t\}. \quad (15)$$

The false discovery proportion (FDP) and false discovery rate (FDR) are defined as

$$\mathrm{FDP}(t) = \frac{R_0(t)}{R(t) \vee 1} \quad \text{and} \quad \mathrm{FDR}(t) = \mathrm{E}[\mathrm{FDP}(t)].$$

An ideal choice of t is

$$t_0 = \inf\left\{0 \le t \le \sqrt{2\log \mathcal{M}} : \frac{R_0(t)}{R(t) \vee 1} \le \alpha\right\},$$

which would reject as many true positives as possible while controlling the FDR at the prespecified level α. However, the total number of false positives, $R_0(t)$, is unknown as the set $\mathcal{H}_0$ is unknown. We propose to estimate $R_0(t)$ by $2(1 - \Phi(t))|\mathcal{H}_0|$ and simply estimate $|\mathcal{H}_0|$ by $\mathcal{M}$ because the number of true alternatives is relatively small. This leads to the following multiple testing procedure with FDR control.

1. Calculate test statistics $N_{\mathcal{J}_m \times \mathcal{J}_h}$.
2. For given $0 \le \alpha \le 1$, calculate

$$\hat{t} = \inf\left\{0 \le t \le \sqrt{2\log \mathcal{M} - 2\log\log \mathcal{M}} : \frac{2\mathcal{M}(1 - \Phi(t))}{R(t) \vee 1} \le \alpha\right\}. \quad (16)$$

   If (16) does not exist, then set $\hat{t} = \sqrt{2\log \mathcal{M}}$.
3. For $(m, h) \in \mathcal{H}$, reject $H_{0,m,h}$ if $N_{\mathcal{J}_m \times \mathcal{J}_h} \ge \hat{t}$.

4.2 Theoretical Properties

We now investigate the theoretical properties of the multiple testing procedure given above. For any $1 \le m \le M$, define

$$\Xi_m(\gamma) = \left\{h : 1 \le h \le M, h \ne m, \exists i \in \mathcal{J}_m, j \in \mathcal{J}_h \text{ s.t. } |\omega_{i,j}| \ge (\log \mathcal{M})^{-2-\gamma}\right\}.$$

The following theorem shows that, under regularity conditions, the above multiple testing procedure controls the FDR and FDP at the pre-specified level α asymptotically.

Theorem 3—Assume that $\mathcal{M}_0 =: |\mathcal{H}_0| \asymp \mathcal{M}$, and (4) and (5) hold. Suppose there exists some $\gamma > 0$ such that $\max_{1 \le m \le M} |\Xi_m(\gamma)| = o(M^\tau)$ for any $\tau > 0$. Then under *(C1) with* $p \le cn^r$ for some $c > 0$ and $r > 0$, we have

$$\overline{\lim}_{(n,\mathcal{M}) \to \infty} \mathrm{FDR}(\hat{t}) \le \alpha,$$

and for any $\varepsilon > 0$,

$$\lim_{(n,\mathcal{M}) \to \infty} P(\mathrm{FDP}(\hat{t}) \le \alpha + \varepsilon) = 1.$$

Remark 3—The technical condition on $|\Xi_m(\gamma)|$ is to ensure that most of the submatrices are not highly correlated with each other. In the special case when $\max_{1 \le m \le M} |\Xi_m(\gamma)| = 0$,

then all subgroups are weakly correlated with each other, i.e., $|\omega_{i,j}| \le (\log \mathcal{M})^{-2-\gamma}$ for all $i \in \mathcal{J}_m$, $j \in \mathcal{J}_h$ with $m \ne h$. Under this setting, it is shown in the supplement Xia et al. (2016) that the proposed multiple testing procedure performs asymptotically the same as the case when all submatrices are independent with each other. We do not need this strong condition, and the weaker condition $\max_{1\le m\le M} |\Xi_m(\gamma)| = o(M^\tau)$ for any $\tau > 0$ assumed in the theorem allows the number of highly correlated submatrices growing with M.

When $\hat{t}$ is not attained in the range $[0, \sqrt{2\log \mathcal{M} - 2\log\ \log \mathcal{M}}]$ as described in equation (16), we shall threshold it at $\sqrt{2\log \mathcal{M}}$. We state in the following corollary a condition to ensure the existence of $\hat{t}$ in the range, and as a result, the FDR and FDP will converge to the pre-specified level α.

Corollary 1—Let

$$\mathcal{S}_\rho = \left\{ (m-h) \in \mathcal{H} : \exists (i,j) \in \mathcal{J}_m \times \mathcal{J}_h \text{ such that } |\omega_{i,j}| / (\theta_{i,j})^{1/2} \ge (\log \mathcal{M})^{\frac{1}{2}+\rho} \right\}.$$

Suppose for some $\rho > 0$ and some $\delta > 0$, $|\mathcal{S}_\rho| \ge (\frac{1}{\sqrt{\pi\alpha}} + \delta)\sqrt{\log \mathcal{M}}$. Assume that $\mathcal{M}_0 =: |\mathcal{H}_0| \asymp \mathcal{M}$, and (4) and (5) hold. Suppose there exists some $\gamma > 0$ such that $\max_{1\le m\le M} |\Xi_m(\gamma)| = o(M^\tau)$ for any $\tau > 0$. Then, under (*C1*) with $p \le cn^r$ for some $c > 0$ and $r > 0$, we have

$$\lim_{(n,\mathcal{M})\to\infty} \mathrm{FDR}(\hat{t}) = \alpha, \quad \textit{and} \quad \mathrm{FDP}(\hat{t})/\alpha \to 1$$

in probability, as $(n,\mathcal{M}) \to \infty$.

Remark 4—The condition $|\mathcal{S}_\rho| \ge (\frac{1}{\sqrt{\pi\alpha}} + \delta)\sqrt{\log \mathcal{M}}$ in Corollary 1 is mild, since there are $\mathcal{M}$ hypotheses in total and this condition only requires a few submatrices having one entry with magnitude exceeding $(\log \mathcal{M})^{1/2+\rho}/n^{1/2}$ for some constant $\rho > 0$.

4.3 Differences with the B-H Procedure

In this section we first discuss the difference between our method and the Benjamini-Hochberg (B-H) procedure and then explain why in the multiple testing procedure it is critical to restrict t on the range $0 \le t \le \sqrt{2\log \mathcal{M} - 2\log\ \log \mathcal{M}}$ in equation (16) and to threshold $N_{\mathcal{J}_m \times \mathcal{J}_h}$ at $\sqrt{2\log \mathcal{M}}$ when $\hat{t}$ is not attained in the range.

Once the test statistic $N_{\mathcal{J}_m \times \mathcal{J}_h}$ for a given submatrix is developed, a natural approach to construct a procedure for simultaneously testing a collection of submatrices is to apply the well-known B-H procedure to the p-values $p_{m,h} = 2(1 - \Phi(N_{\mathcal{J}_m \times \mathcal{J}_h}))$, $1 \le m < h \le M$, computed from the transformed statistics $N_{\mathcal{J}_m \times \mathcal{J}_h}$. Applying the B-H procedure to these p values is equivalent to rejecting the null hypotheses $H_{0,m,h}$ whenever $N_{\mathcal{J}_m \times \mathcal{J}_h} \ge \hat{t}_{BH}$, where

$$\hat{t}_{BH} = \inf\left\{t \geq 0: \frac{2\mathscr{M}(1-\Phi(t))}{R(t)\vee 1} \leq \alpha\right\}. \quad (17)$$

Note that, the difference between our procedure and the B-H procedure is on the ranges of t in equations (16) and (17).

We first emphasize here that the restriction on the range $0 \leq t \leq \sqrt{2\log \mathscr{M} - 2\log\log \mathscr{M}}$ in our proposed procedure as defined in (16) is critical. When $t \geq \sqrt{2\log \mathscr{M} - \log\log \mathscr{M}}$, $2\mathscr{M}(1-\Phi(t)) \to 0$ is not even a consistent estimate of $R_0(t)$ because $|R_0(t)/\{2\mathscr{M}(1-\Phi(t))\} - 1| \not\to 0$ in probability as $(n, \mathscr{M}) \to \infty$. However, direct application of the B-H procedure to the p-values amounts to using $2\mathscr{M}(1-\Phi(t))$ as an estimate of $R_0(t)$ for all $t \geq 0$, and as a result it may not able to control the FDP with positive probability. For example, when the number of true alternatives is fixed, it is shown in Proposition 2.1 in Liu and Shao (2014) that the B-H procedure cannot control the FDP with positive probability. Thus, in order to control FDP, it is crucial to restrict t on the range $0 \leq t \leq \sqrt{2\log \mathscr{M} - 2\log\log \mathscr{M}}$.

When t is not attained in the range, it is also critical to threshold $N_{\mathscr{J}_m \times \mathscr{J}_h}$ at $\sqrt{2\log \mathscr{M}}$ instead of $\sqrt{2\log \mathscr{M} - 2\log\log \mathscr{M}}$. When t does not exist in the range, thresholding $N_{\mathscr{J}_m \times \mathscr{J}_h}$ at $\sqrt{2\log \mathscr{M} - 2\log\log \mathscr{M}}$ will cause too many false rejections, and consequently the FDR cannot be controlled asymptotically at level α. If the threshold level is increased to $\sqrt{2\log \mathscr{M}}$, the probability of false rejections can then be perfectly controlled asymptotically as shown in equation (3) of the supplement Xia et al. (2016).

To summarize, in order to control FDR and FDP, it is crucial to restrict t on the range $0 \leq t \leq \sqrt{2\log \mathscr{M} - 2\log\log \mathscr{M}}$ in equation (16), and when it is not attained in the range, to threshold $N_{\mathscr{J}_m \times \mathscr{J}_h}$ at $\sqrt{2\log \mathscr{M}}$.

5 Simulation Studies

We now turn to analyze the numerical performance of the proposed multiple testing procedure through simulation studies. We first investigate the size and power of the proposed method by considering three matrix models with a random selection of the size of submatrices. We then mimic the sizes of the pathways of the breast cancer dataset analyzed in Section 6 and study the numerical performance of the proposed multiple testing procedure in a setting that is similar to the real data application. Our method, which tests for the conditional dependence structure at a group level, is then compared with the entrywise testing method and the B-H procedure. We also compare the new method with the Bonferroni correction procedure and report the results in the supplement.

5.1 Simulation for Different Constructions of Submatrices

Our analysis is divided into two parts: the performance of the new test statistics for testing a given submatrix and the performance of the proposed multiple testing procedure. We first describe the construction of the submatrices. For a given precision matrix $\mathbf{\Omega}$, we randomly divide the upper triangular matrix of $\mathbf{\Omega}$ into $\mathscr{M}$ submatrices, where $\mathscr{M} = \lfloor p/s \rfloor (\lfloor p/s \rfloor - 1)/2$ and

$s = 2$ and 4. Thus the length of the index sets can range from 1 to $(p - \lfloor p/s \rfloor + 1)$. This is equivalent to grouping the genes into $\lfloor p/s \rfloor$ pathways and considering all possible conditional dependence structure between different pathways of different sizes.

The data $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ are generated from multivariate normal distribution with zero-mean and precision matrix $\boldsymbol{\Omega}$. Three choices of $\boldsymbol{\Omega}$ are considered:

- Model 1: $\boldsymbol{\Omega}^{*(1)} = (\omega_{i,j}^{*(1)})$ where $\omega_{i,i}^{*(1)} = 1, \omega_{i,i+1}^{*(1)} = \omega_{i+1,i}^{*(1)} = 0.5, \omega_{i,i+2}^{*(1)} = \omega_{i+2,i}^{*(1)} = 0.5$. For each of the submatrices as we constructed above, if it contains one of those entries, we make the first row of the submatrices equal to 0.5. Let $\omega_{i,j}^{*(1)} = 0$ otherwise. $\boldsymbol{\Omega}^{(1)} = \boldsymbol{D}^{1/2}(\boldsymbol{\Omega}^{*(1)} + \delta \boldsymbol{I})/(1 + \delta)\boldsymbol{D}^{1/2}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^{*(1)})| + 0.05$.
- Model 2: $\boldsymbol{\Omega}^{*(2)} = (\omega_{i,j}^{*(2)})$ where $\omega_{i,j}^{*(2)} = \omega_{j,i}^{*(2)} = 0.3$ for $i = 10(k-1) + 1$ and $10(k-1) + 2 \le j \le 10(k-1) + 10$, $1 \le k \le p/10$. $\omega_{i,j}^{*(2)} = 0$ otherwise. For each of the submatrices as we constructed above, if it contains less than three of those entries, we make the submatrices equal to 0. Let the first row of the submatrices which are closest to the diagonal equal to 0.3. $\boldsymbol{\Omega}^{(2)} = \boldsymbol{D}^{1/2}(\boldsymbol{\Omega}^{*(2)} + \delta \boldsymbol{I})/(1 + \delta)\boldsymbol{D}^{1/2}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^{*(2)})| + 0.05$.
- Model 3: $\boldsymbol{\Omega}^{*(3)} = (\omega_{i,j}^{(3)})$. For each of the two submatrices closest to the diagonal, as we constructed above, pick a random row and make the entries equal to 0.3. Let $\omega_{j,i}^{*(3)} = \omega_{i,j}^{*(3)}$. $\boldsymbol{\Omega}^{(3)} = \boldsymbol{D}^{1/2}(\boldsymbol{\Omega}^{*(3)} + \delta \boldsymbol{I})/(1 + \delta)\boldsymbol{D}^{1/2}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Omega}^{*(3)})| + 0.05$.

where $\boldsymbol{D} = (D_{i,j})$ is a diagonal matrix with $D_{i,i} = \text{Unif}(1, 3)$ for $i = 1, \ldots, p$.

For each generated dataset, we use the Lasso to estimate the regression coefficients $\boldsymbol{\beta}_i$:

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{D}_i^{-\frac{1}{2}} \arg\min_{\boldsymbol{u}} \left\{ \frac{1}{2n} \left| (\boldsymbol{X}_{-i} - \bar{\boldsymbol{X}}_{-i}) \boldsymbol{D}_i^{-1/2} \boldsymbol{u} - (\boldsymbol{X}_{(i)} - \bar{\boldsymbol{X}}_{(i)}) \right|_2^2 + \lambda_{n,i} |\boldsymbol{u}|_1 \right\}, \quad (18)$$

where $\boldsymbol{D}_i = \text{diag}(\hat{\boldsymbol{\Sigma}}_{-i,-i})$, and $\lambda_{n,i} = \kappa \sqrt{\hat{\sigma}_{i,i} \log p/n}$.

Performance for testing a given submatrix—We start by comparing our test based on the test statistic $S_{\mathcal{G} \times \mathcal{J}}$ with the entrywise testing of a given submatrix where the null hypothesis H_0: $\boldsymbol{\Omega}_{\mathcal{G} \times \mathcal{J}} = 0$ is rejected whenever $\max_{(i,j) \in \mathcal{G} \times \mathcal{J}} |W_{i,j}| \ge \Phi^{-1}(1 - \alpha/K)$. As our target is the FDR control of the multiple comparisons, we focus on the power comparisons of these two methods for a range of significance levels from 0 to $\alpha = 0.1/\mathcal{M}$. For illustration, we compare the performance of these two tests by testing against a randomly selected nonzero submatrix closest to the diagonal for Model 1 with $s = 4$. For each method, the sample size is taken to be $n = 200$, while the dimension p varies over the values 100, 200, 500 and 1000. For simplicity of the comparison, the tuning parameters $\lambda_{n,i}$ in (18) is

selected to be $\lambda_{n,i} = \sqrt{\hat{\sigma}_{i,i} \log p/n}$ for both methods. The power curves, illustrated in Figure 1, are estimated from 100 replications. We can see from the figure that the power of the new group method is much higher than the entrywise method, and the advantage becomes much clearer when the dimension of $\boldsymbol{\Omega}$ grows.

Comparison of the multiple testing procedures—We now compare the proposed group level FDR control procedure (Group) with three other methods: entrywise multiple testing method (Entrywise), B-H procedure (B-H) and Bonferroni correction procedure (Bonferroni).

For the new method, as described in Section 4, we select the tuning parameters $\lambda_{n,i}$ in (18) adaptively by the data with the principle of making $\Sigma_{(m,h)\in\mathcal{H}_0} /(N_{\mathcal{J}_m \times \mathcal{J}_h} \geq t)$ and $(2 - 2\Phi(t))|\mathcal{H}_0|$ as close as possible. The algorithm is similar as Xia et al. (2015) and is summarized as follows.

1. Let $\lambda_{n,i} = b/20\sqrt{\hat{\boldsymbol{\Sigma}}_{i,i} \log p/n}$ for $b = 1, \cdots, 40$. For each b, calculate $\hat{\boldsymbol{\beta}}_i^{(b)}$, $i = 1, \cdots, p$. Based on the estimation of regression coefficients, construct the corresponding standardized transformed statistics $N_{\mathcal{J}_m \times \mathcal{J}_h}^{(b)}$ for each b.

2. Choose $\hat{b}$ as the minimizer of

$$\sum_{d=1}^{10} \left(\frac{\sum_{(m,h) \in \mathcal{H}} I(N_{\mathcal{J}_m \times \mathcal{J}_h}^{(b)} \geq \Phi^{-1}(1 - d(1 - \Phi(\sqrt{\log \mathcal{M}}))/10))}{d(1 - \Phi(\sqrt{\log \mathcal{M}}))/10 \cdot 2\mathcal{M}} - 1 \right)^2.$$

The tuning parameters $\lambda_{n,i}$ are then chosen to be

$$\lambda_{n,i} = \hat{b}/20\sqrt{\hat{\boldsymbol{\Sigma}}_{i,i} \log p/n}.$$

We examine the power of the new method based on the average powers for 100 replications,

$$\frac{1}{100} \sum_{r=1}^{100} \frac{\sum_{(m,h) \in \mathcal{H}_1} I\{N_{\mathcal{J}_m \times \mathcal{J}_{h,r}} \geq \hat{t}\}}{|\mathcal{H}_1|}, \quad (19)$$

where r denotes the r-th replication.

For the entrywise multiple testing method, we select the tuning parameters $\lambda_{n,i}$ adaptively using the principle as described in Section 5 in Xia et al. (2015). We applied the multiple testing procedure as developed in Section 4 of Xia et al. (2015) by restricting t on the range $[0, \sqrt{4 \log p - 2 \log \log p}]$ and threshold $|W_{i,j}|$ at $\sqrt{4 \log p}$ if $\hat{t}$ is not attained in the range. We then examine the empirical FDR by

$$\frac{1}{100}\sum_{r=1}^{100}\frac{\sum_{(m,h)\in\mathcal{H}_0} I\{\max_{(i,j)\in\mathcal{J}_m\times\mathcal{J}_h}|W_{i,j}|\geq\hat{t}\}}{\sum_{(m,h)} I\{\max_{(i,j)\in\mathcal{J}_m\times\mathcal{J}_h}|W_{i,j}|\geq\hat{t}\}},$$

and the empirical power by

$$\frac{1}{100}\sum_{r=1}^{100}\frac{\sum_{(m,h)\in\mathcal{H}_1} I\{\max_{(i,j)\in\mathcal{J}_m\times\mathcal{J}_h}|W_{i,j}|\geq\hat{t}\}}{|\mathcal{H}_1|}.$$

We apply the Bonferroni correction procedure to the new test statistics and calculate its power based on (19), with $\hat{t}$ obtained by setting $\alpha_B = \alpha/\mathcal{M}$. The power of the B-H procedure applied to $\max_{(i,j)\in\mathcal{Q}\times\mathcal{J}}|W_{i,j}|$ are calculated by (19) with no restriction on the range of $\hat{t}$.

We apply all procedures to these three models with $s = 2$ and 4. For each method, the sample size is taken to be $n = 200$, while the dimension p varies over the values 100, 200, 500 and 1000. The FDR level is set at $\alpha = 0.1$ and $\alpha = 0.01$ respectively, and the empirical FDRs and powers, summarized in Tables 1 and 2, are estimated from 100 replications. The standard errors of the estimated powers are much smaller than the powers themselves and are thus not reported.

The average numbers of conditionally dependent ("true interaction") and conditionally independent ("no interaction") pairs of subgroups with 100 replications are summarized in Table 3. It can be seen that the number of "true interactions" is relatively small compared to the total number of pairs of subgroups in all cases, as we assumed in Section 4.1.

The results in Table 1 show that the empirical FDRs of the new group level method are well maintained under the target FDR level and are reasonably close to α for almost all settings. The standard errors of the FDP are small in most cases, especially when the dimension grows. They are slightly larger in the cases when $\alpha = 0.01$, mainly due to the fact that the estimation error of the standard deviation of FDP is of the order $1/I^{1/2}$ with $I = 100$. As a comparison, the empirical FDRs of the entrywise method have serious distortion in most of the scenarios, especially when $s = 4$, in which case the empirical FDRs can be even larger than 4α. The empirical FDRs of the B-H procedure are well under control in most cases. However, its standard errors are much larger than the standard errors of the proposed method in many cases, which coincides with the discussion in Section 4.3. The numerical results also show that the Bonferroni correction procedure is much more conservative than the other two methods, and the detailed analysis is summarized in the supplement Xia et al. (2016).

Table 2 shows that the empirical powers of our proposed method for all these models are very high under various constructions of submatrices. In particular, it outperforms the entrywise testing method and the B-H procedure. Especially when the dimension is high, the powers of the new method are much higher than the other methods under all scenarios. Furthermore, the power gain of the new group level testing procedure over the entrywise testing method is significant when the dimension is high. Especially for model 3 when $s = 4$,

the empirical powers of the new procedure are more than twice the entrywise testing method. This is because the advantage of the group level testing becomes more significant when the signals spread across various submatrices as in Model 3. We can see from the table that the empirical power of the new method gets smaller when the dimension p grows. This is because of the fact that we keep the magnitude of $\omega_{i,j}$ invariant for various range of dimensions.

5.2 Simulation by Mimicking the Sizes of Gene Groups

We now consider a simulation setting that is similar to the breast cancer data application given in Section 6. The submatrices of the precision matrix $\mathbf{\Omega}$ is constructed by mimicking the sizes of the 249 gene groups used in the breast cancer application, with parameter values $p = 1624$, $n = 295$ and $\mathcal{M} = 30876$. The sizes of the gene groups range from 1 to 110, and the corresponding sizes of the off-diagonal submatrices range from 1×1 to 97×110. For the diagonal submatrices $\mathbf{\Omega}^*_{\mathcal{J}_m \times \mathcal{J}_m} := (\omega^*_{m,i,j})$ with sizes $L_m \times L_m$, $m = 1, \ldots, 249$, which describe the conditional dependency within the pathways, we let $\omega^*_{m,i,i} = 1, \omega^*_{m,i,i+1} = \omega^*_{m,i+1,i} = 0.8$ if $L_m \geq 2$, $\omega^*_{m,i,i+2} = \omega^*_{m,i+2,i} = 0.6$ if $L_m \geq 3$, and $\omega^*_{m,j,i} = \omega^*_{m,i,j}$. For each of the non-diagonal submatrices $\mathbf{\Omega}^*_{\mathcal{J}_M \times \mathcal{J}_{m+1}}$ and $\mathbf{\Omega}^*_{\mathcal{J}_m \times \mathcal{J}_{m+2}}$, we randomly pick one row and let $\min\{10, |\mathcal{J}_{m+1}|\}$ and $\min\{10, |\mathcal{J}_{m+2}|\}$ random entries of $\omega^*_{i,j}$ in the rows equal to 0.5 respectively. We then construct the precision matrix as $\mathbf{\Omega} = \boldsymbol{D}^{1/2}(\mathbf{\Omega}^* + \delta \boldsymbol{I})/(1+\delta)\boldsymbol{D}^{1/2}$, with $\delta = \lambda_{\min}(\mathbf{\Omega}^*) + 0.05$. The FDR level is set at $\alpha = 0.1$ and $\alpha = 0.01$ respectively.

By mimicking the gene group sizes, we apply the proposed method in Section 4.1, the entrywise testing procedure, the B-H procedure and the Bonferroni correction procedure as described in Section 5.1. The empirical FDR and power results are summarized in Table 4, and the performance of the Bonferroni method is reported in the supplement. The empirical FDR of the new method is equal to 0.062 when $\alpha = 0.1$ and is equal to 0.006 when $\alpha = 0.01$, and thus both are close to the corresponding pre-specified level. Similarly as in Section 5.1, the B-H procedure has larger standard errors than the new procedure, while the entrywise multiple testing procedure has serious FDR distortion. For the empirical powers, it is shown in Table 4 that, the new testing procedure is more powerful than all the other methods.

6 Analysis of Breast Cancer Gene Expression Data

In this section, we apply the multiple testing procedure developed in Section 4 to identify between pathway interactions based on a breast cancer gene expression study as described in van't Veer et al. (2002), to further illustrate the merit of the procedure.

This study consists of 295 subjects with primary breast carcinomas whose gene-expression levels (in log scale) are measured at cancer diagnosis. For illustration, we consider $M = 70$ breast cancer related pathways, including several major signaling pathways, assembled based on existing literature (Osborne et al., 2005; Pan, 2012, e.g.). These pathways consist of $p = 1624$ unique genes, from the molecular signature database. Examples include the

MAPK signaling, WNT signaling, TGF-β signaling, calcium signaling, cell communication, p53 signaling and breast cancer estrogen signaling pathways. Note that many of the pathways have overlapping genes while our method requires group indices to be non-overlapping since two groups with shared genes are obviously dependent of each other. To remove the influence of such trivial dependence, we further partitioned the 70 pathways into 249 non-overlapping gene subgroups whose sizes range from 1 to 110 with an average of 6.5. The algorithm used for such partitioning aims to identify the smallest number of non-overlapping subgroups that can cover all the genes under consideration. The partitioning algorithm begins with creating an $M \times p$ index matrix, $\mathbb{I} = [\mathbf{I}_1, \ldots, \mathbf{I}_p]$. For $m = 1, \ldots, M$ and $q = 1, \ldots, p$, the (m, q)th element of $\mathbb{I}$ is set to 1 if the qth gene belongs to the mth pathway, and 0 otherwise. Then the subgroups are indexed by the unique values of $\{\mathbf{I}_1, \ldots, \mathbf{I}_p\}$.

Applying our proposed methods with target false discovery rate of 0.01, we identified 494 between subgroup interactions out of the 30876 possible subgroup pairs. These between subgroup interactions can be mapped to 311 unique between pathway interactions and 18 within pathway interactions. The top pathways with highest numbers of interactions with other pathways include MAPK signaling, calcium signaling, gycan structures biosynthesis, WNT signaling, cell communication, TGF-β signaling and breast cancer estrogen pathways. The MAPK signaling pathway has interactions with 92 gene subgroups which corresponds to 31 pathways including TGF-β, MTOR, P53, WNT, and ERBB signaling pathways. The WNT signaling pathway interacts with 25 other pathways including TGF-β, MTOR, MPAK and breast cancer estrogen signaling. The TGF-β signaling pathway interacts with 21 other pathways including MAPK, p53, WNT and calcium signaling.

Many of these interactions have been previously documented. For example, experimental data suggest that inhibition of mTORC1 leads to MAPK pathway activation (Carracedo et al., 2008). The interaction between TGF-β and WNT pathways has been known for a long time and is probably the most extensively studied. At the organism level, TGF-β interacts with many other pathways at every stage of life from birth to death. During embryonic development, the complex but delicate interactions between the TGF-β, WNT, MAPK, and other pathways are important for a range of processes including body patterning, stem cell maintenance and cell fate determination (Guo and Wang, 2008). Kouzmenko et al. (2004) showed the first direct evidence of interaction between WNT and estrogen signaling pathways via functional interaction between β-catenin and ERα.

To examine whether these 70 breast cancer pathways are enriched with interactions, we randomly selected 50 sets of 70 pathways of similar sizes as the breast cancer pathways from the C2 pathway gene sets curated from various online databases (available from the Broad Institute). For each of the 70 randomly selected pathways, we performed the same analysis as the breast cancer pathways by first partitioning them into non-overlapping subgroups and then applied our method to identify significant between subgroup interactions. To determine whether the 70 breast cancer pathways are enriched with between subgroup interactions relative to these randomly selected pathways, we calculate the proportion of between subgroup interactions deemed as significant at the FDR level of 0.01. Across the 50 randomly selected pathways, the average proportion of significant pairs was 0.011 with standard deviation 0.002. The proportion of significant pairs we identified in the

breast cancer data is 0.016, which is 2.5 standard deviations higher than the mean of proportions from those 50 random sets. The results suggest that the selected 70 pathways are indeed enriched with “interaction” pairs.

7 Discussions

We proposed in this paper a multiple testing procedure under the Gaussian graphical models for detecting between group interactions. The proposed method can potentially be extended in several directions. We discuss in this section two of these possible extensions.

7.1 Extension to Gaussian Copula Graphical Models

In the present paper, the problem of identifying the conditional between group interactions is translated to the problem of multiple testing of submatrices of a high-dimensional precision matrix $\mathbf{\Omega}$ under the Gaussian graphical model framework. The main reason for the success of this approach is that the conditional independence between two non-overlapping groups of variables is equivalent to the corresponding submatrix of $\mathbf{\Omega}$ being 0. This approach can be extended to more general settings of the semiparametric Gaussian copula graphical models where the population distribution is non-Gaussian, see Liu et al. (2012) and Xue and Zou (2012). The semiparametric Gaussian copula model assumes that the variables follow a joint normal distribution after a set of unknown marginal monotonic transformations. It would be interesting to develop a multiple testing procedure and investigate its properties under the semiparametric Gaussian copula graphical models. Detailed analysis is involved and is an interesting topic for future research.

7.2 The Two-Sample Case

We have focused on the one-sample case in this paper. It is also of interest to study the two-sample case where the goal is to discover the changes in the conditional dependence between pathway interactions under two different disease settings. In the one-sample case studied in this paper, $\mathbf{\Omega}_{\mathcal{J}_m \times \mathcal{J}_h} = 0$ under $H_{0,m,h}$. Thus the null is simple but the technical details of deriving the limiting distribution of a given submatrix is still very involved because the correlation structure of $\{W_{i,j}, (i,j) \in \mathcal{J}_m \times \mathcal{J}_h\}$ largely depends on the entries outside of the submatrix of interest. In the two-sample case, we wish to test the hypotheses $H_{0,m,h}: \mathbf{\Omega}^{(1)}_{\mathcal{J}_m \times \mathcal{J}_h} = \mathbf{\Omega}^{(2)}_{\mathcal{J}_m \times \mathcal{J}_h}$. Under the null hypothesis $H_{0,m,h}$, each submatrix is not necessary a zero matrix. So the null is composite, consequently the dependence structures of the suitable test statistics depend on the entries both inside and outside of the submatrices of direct interest. The two-sample case is technically even more challenging and we leave it as future work.

Table 5: Empirical FDRs (standard errors) (%) with $n = 200$, $\alpha = 0.1$ and 0.01 respectively, 100 replications.

Table 6: Empirical powers (%) with $n = 200$, $\alpha = 0.1$ and 0.01 respectively, 100 replications.

Table 7: Empirical FDRs (SEs) and powers (%) by mimicking the real data with $\alpha = 0.1$ and $\alpha = 0.01$ respectively, based on 100 replications.

Acknowledgments

The research of Tony Cai was supported in part by NSF Grants DMS-1208982 and DMS-1403708, and NIH Grant R01 CA127334.

The research of Yin Xia was supported in part by "The Recruitment Program of Global Experts" Youth Project from China, the startup fund from Fudan University and NSF Grant DMS-1612906.

The research of Tianxi Cai was supported in part by NIH Grants R01 GM079330, P50 MH106933, and U54 HG007963.

References

Anderson, TW. An Introduction To Multivariate Statistical Analysis. 3. Wiley-Intersceince; New York: 2003.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B. 1995; 57:289–300.

Beran R, Bilodeau M, de Micheaux PL. Nonparametric tests of independence between random vectors. Journal of Multivariate Analysis. 2007; 98(9):1805–1824.

Cai TT, Liu W. Large-scale multiple testing of correlations. Journal of the American Statistical Association. 2015:110. (to appear).

Cai TT, Liu W, Xia Y. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. J Amer Statist Assoc. 2013; 108(501):265–277.

Cai TT, Zhang A. Inference on high-dimensional differential correlation matrix. 2014 arXiv preprint arXiv:1408.5907.

Carracedo A, Ma L, Teruya-Feldstein J, Rojo F, Salmena L, Alimonti A, Egia A, Sasaki AT, Thomas G, Kozma SC, et al. Inhibition of mtorc1 leads to mapk pathway activation through a pi3k-dependent feedback loop in human cancer. The Journal of clinical investigation. 2008; 118(9):3065. [PubMed: 18725988]

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. The American Journal of Human Genetics. 2006; 79(6):1002–1016. [PubMed: 17186459]

Craven M, Kumlien J. Constructing biological knowledge bases by extracting information from text sources. ISMB. 1999; 1999:77–86.

Fan J, Lv J. Sure independence screening for ultra-high dimensional feature space (with discussions). Journal of the Royal Statistical Society, Series B. 2008; 70:849–911.

Gieser PW, Randles RH. A nonparametric test of independence between two vectors. Journal of the American Statistical Association. 1997; 92(438):561–567.

Glazko GV, Emmert-Streib F. Unite and conquer: univariate and multivariate approaches for finding differentially expressed gene sets. Bioinformatics. 2009; 25(18):2348–2354. [PubMed: 19574285]

Guo X, Wang XF. Signaling cross-talk between tgf-β/bmp and other pathways. Cell research. 2008; 19(1):71–88.

Huang TM, et al. Testing conditional independence using maximal nonlinear conditional correlation. The Annals of Statistics. 2010; 38(4):2047–2091.

Jia P, Kao CF, Kuo PH, Zhao Z. A comprehensive network and pathway analysis of candidate genes in major depressive disorder. BMC systems biology. 2011; 5(Suppl 3):S12.

Kelley R, Ideker T. Systematic interpretation of genetic interactions using protein networks. Nature biotechnology. 2005; 23(5):561–566.

Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS computational biology. 2012; 8(2):e1002375. [PubMed: 22383865]

Kooperberg C, LeBlanc M. Increasing the power of identifying gene× gene interactions in genome-wide association studies. Genetic epidemiology. 2008; 32(3):255–263. [PubMed: 18200600]

Kooperberg C, Ruczinski I. Identifying interacting snps using monte carlo logic regression. Genetic epidemiology. 2005; 28(2):157–170. [PubMed: 15532037]

Kouzmenko AP, Takeyama K-i, Ito S, Furutani T, Sawatsubashi S, Maki A, Suzuki E, Kawasaki Y, Akiyama T, Tabata T, et al. Wnt/β-catenin and estrogen signaling converge in vivo. Journal of Biological Chemistry. 2004; 279(39):40255–40258. [PubMed: 15304487]

Lauritzen, SL. Graphical models. Oxford University Press; 1996.

Li Y, Agarwal P, Rajagopalan D. A global pathway crosstalk network. Bioinformatics. 2008; 24(12): 1442–1447. [PubMed: 18434343]

Liu H, Han F, Yuan M, Lafferty J, Wasserman L, et al. High-dimensional semiparametric gaussian copula graphical models. Ann Statist. 2012; 40(4):2293–2326.

Liu H, Tang Y, Zhang HH. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Comput Stat Data Anal. 2009; 53(4):853–856.

Liu W. Gaussian graphical model estimation with false discovery rate control. Ann Statist. 2013; 41(6):2948–2978.

Liu W, Shao QM. Phase transition and regularized bootstrap in large scale t-tests with false discovery rate control. Ann Statist. 2014; 42(5):2003–2025.

Liu ZP, Wang Y, Zhang XS, Chen L. Identifying dysfunctional crosstalk of pathways in various regions of alzheimer's disease brains. BMC systems biology. 2010; 4(Suppl 2):S11. [PubMed: 20840725]

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. Reactome knowledgebase of human biological pathways and processes. Nucleic acids research. 2009; 37(suppl 1):D619–D622. [PubMed: 18981052]

Osborne CK, Shou J, Massarweh S, Schiff R. Crosstalk between estrogen receptor and growth factor receptor pathways as a cause for endocrine therapy resistance in breast cancer. Clinical cancer research. 2005; 11(2):865s–870s. [PubMed: 15701879]

Pan XH. Pathway crosstalk analysis based on protein-protein network analysis in ovarian cancer. Asian Pacific Journal of Cancer Prevention. 2012; 13(8):3905–3909. [PubMed: 23098491]

Puri N, Salgia R, et al. Synergism of egfr and c-met pathways, cross-talk and inhibition, in non-small cell lung cancer. Journal of carcinogenesis. 2008; 7(1):9. [PubMed: 19240370]

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(1):138–147. [PubMed: 11404819]

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein–protein interaction network. Nature. 2005; 437(7062):1173–1178. [PubMed: 16189514]

Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H, Schiff R. Mechanisms of tamoxifen resistance: increased estrogen receptor-her2/neu cross-talk in er/her2–positive breast cancer. Journal of the National Cancer Institute. 2004; 96(12):926–935. [PubMed: 15199112]

Su L, White H. A consistent characteristic function-based test for conditional independence. Journal of Econometrics. 2007; 141(2):807–834.

Su L, White H. A nonparametric hellinger metric test for conditional independence. Econometric Theory. 2008; 24(04):829–864.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(43):15545–15550. [PubMed: 16199517]

Um Y, Randles RH. A multivariate nonparametric test of independence among many vectors. Journal of Nonparametric Statistics. 2001; 13(5):699–708.

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(6871):530–536. [PubMed: 11823860]

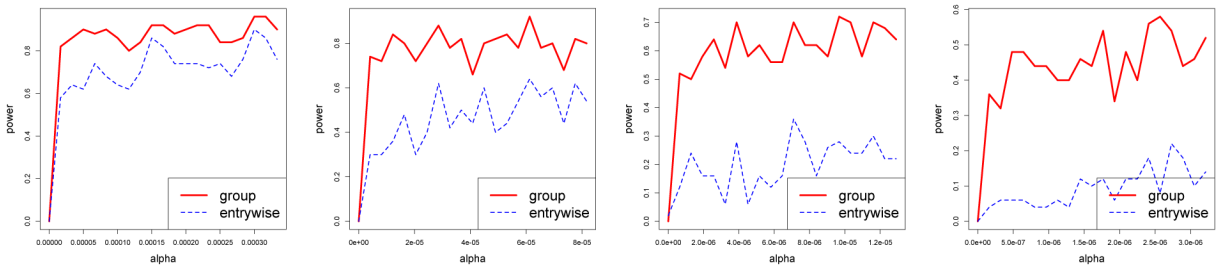
Weirauch MT. Gene coexpression networks for the analysis of dna microarray data. Applied statistics for network biology: methods in systems biology. 2011:215–250.

Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research. 2002; 30(1):303–305. [PubMed: 11752321]

Xia Y, Cai T, Cai TT. Testing differential networks with applications to the detection of gene-gene interactions. Biometrika. 2015; 102:247–266. [PubMed: 28502988]

Xia Y, Cai T, Cai TT. Supplement to “Multiple Testing of Submatrices of a Precision Matrix with Applications to Identification of Between Pathway Interactions”. Technical report. 2016

Xue L, Zou H. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. Ann Statist. 2012; 40(5):2541–2571.



(a) p=100 (b) p=200 (c) p=500 (d) p=1000

Figure 1.

Power comparisons of the group method (red, solid) and entrywise method (blue, dash) for testing a given nonzero submatrix. 100 replications.

Table 1

Empirical FDRs (standard errors) (%) with $n = 200$, $\alpha = 0.1$ and 0.01 respectively, 100 replications.

p	method	$\alpha = 10\%$						$\alpha = 1\%$					
	s	2			4			2			4		
	Models	1	2	3	1	2	3	1	2	3	1	2	3
		Empirical FDR (SE) (in %)											
	Group	8.9 (3.4)	9.5 (4.3)	9.0 (2.9)	4.6 (3.3)	8.6 (6.0)	9.7 (5.6)	0.8 (1.2)	1.0 (1.0)	0.6 (0.8)	0.2 (0.5)	0.8 (1.0)	0.4 (1.1)
100	Entrywise	24.5 (4.6)	15.7 (5.0)	12.7 (3.7)	36.2 (7.6)	24.1 (8.3)	16.3 (6.1)	2.7 (2.3)	1.3 (1.8)	1.0 (1.2)	4.0 (4.1)	1.8 (3.3)	1.3 (5.5)
	B-H	10.0 (3.6)	12.7 (5.5)	9.8 (4.2)	8.3 (5.0)	12.4 (7.4)	11.7 (6.8)	0.8 (1.3)	1.2 (1.9)	0.8 (1.1)	0.4 (1.5)	1.5 (5.6)	1.1 (3.5)
	Group	8.8 (2.5)	9.4 (3.5)	8.7 (2.5)	5.8 (3.3)	8.5 (4.5)	8.7 (3.9)	0.6 (0.5)	0.9 (0.5)	0.8 (0.8)	0.3 (0.4)	0.8 (0.2)	0.8 (0.7)
200	Entrywise	23.5 (3.5)	14.9 (3.7)	13.6 (3.0)	33.1 (7.0)	23.0 (5.5)	15.1 (5.8)	2.1 (1.4)	1.0 (1.2)	0.9 (1.0)	3.1 (2.8)	1.3 (2.3)	0.7 (3.1)
	B-H	9.8 (2.9)	11.7 (3.8)	9.2 (3.1)	8.8 (4.2)	11.9 (4.6)	12.1 (5.5)	0.8 (0.9)	0.9 (1.9)	0.8 (1.1)	0.8 (1.5)	1.4 (2.5)	1.0 (4.2)
	Group	7.9 (1.5)	9.9 (2.3)	8.1 (1.5)	5.4 (1.8)	8.7 (2.9)	9.3 (2.7)	0.8 (0.4)	0.9 (0.7)	0.7 (0.5)	0.5 (0.6)	0.7 (0.8)	0.9 (0.8)
500	Entrywise	19.4 (2.2)	15.0 (2.3)	12.5 (2.1)	24.6 (4.9)	20.6 (4.1)	16.2 (5.3)	1.6 (1.0)	0.9 (0.9)	0.8 (0.8)	1.6 (1.4)	1.1 (1.6)	1.1 (2.1)
	B-H	8.3 (2.0)	11.3 (2.6)	9.2 (2.1)	8.7 (3.4)	11.9 (4.2)	11.6 (4.9)	0.8 (0.6)	1.1 (1.1)	0.9 (1.1)	0.7 (1.0)	1.2 (2.3)	1.4 (3.5)
	Group	7.9 (1.2)	9.8 (1.8)	8.7 (1.4)	6.0 (1.7)	9.0 (2.0)	10.0 (2.1)	0.7 (0.3)	0.9 (0.4)	0.7 (0.4)	0.4 (0.4)	0.8 (0.5)	1.0 (0.7)
1000	Entrywise	16.7 (2.0)	14.5 (1.8)	13.1 (2.1)	20.9 (3.5)	20.0 (2.9)	16.5 (5.4)	1.3 (0.7)	1.3 (0.9)	1.1 (1.0)	1.7 (1.7)	1.2 (1.6)	0.8 (2.0)
	B-H	7.8 (1.3)	11.8 (2.3)	10.1 (1.9)	9.5 (2.7)	12.0 (2.8)	12.7 (6.2)	0.6 (0.5)	1.1 (1.0)	0.8 (1.1)	0.9 (1.2)	1.3 (2.3)	0.8 (2.1)

Table 2

Empirical powers (%) with $n = 200$, $\alpha = 0.1$ and 0.01 respectively, 100 replications.

p	method	$\alpha = 10\%$						$\alpha = 1\%$					
	s	2			4			2			4		
	Models	1	2	3	1	2	3	1	2	3	1	2	3
100	Group	93.8	88.5	84.6	95.2	84.4	73.5	87.8	74.2	69.4	92.3	68.7	54.4
	Entrywise	92.8	90.3	85.2	92.8	85.1	67.3	83.4	71.9	66.8	85.9	56.3	28.5
	B-H	92.8	88.3	84.6	92.9	81.1	67.2	82.3	66.3	64.5	84.0	50.9	31.2
200	Group	90.4	83.3	72.5	92.6	77.9	58.8	82.6	66.8	56.1	87.9	62.7	42.6
	Entrywise	87.6	84.8	71.7	87.3	76.0	45.5	93.6	60.8	47.6	73.6	40.8	15.0
	B-H	87.6	81.4	70.9	86.9	71.3	43.3	72.2	54.6	46.7	71.1	33.8	14.1
500	Group	84.3	72.3	56.6	85.9	67.4	41.4	75.1	55.4	40.7	76.0	53.4	27.2
	Entrywise	78.7	70.8	50.8	72.5	60.0	24.2	60.0	43.3	25.6	48.3	25.9	5.0
	B-H	77.8	66.0	50.4	70.0	52.5	20.3	57.5	36.3	23.8	43.3	19.3	3.0
1000	Group	80.1	63.3	46.4	80.7	59.9	31.9	69.9	57.2	31.8	69.2	46.8	21.1
	Entrywise	71.2	59.2	37.1	60.7	48.5	14.5	49.3	31.5	15.5	33.5	18.5	3.1
	B-H	69.8	53.1	35.4	56.6	39.9	9.6	45.8	23.9	12.7	27.3	10.6	1.2

Table 3

Average numbers of true interactions and no interactions based on 100 replications.

s	2			4			2			4		
Models	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
p	**True interactions**						**No interactions**					
100	72	54	96	29	28	46	1153	1171	1129	271	272	254
200	146	110	196	60	58	96	4803	4839	4754	1165	1167	1129
500	373	279	496	154	147	246	30752	30846	30629	7596	7603	7504
1000	748	560	996	311	297	496	124002	124190	123754	30814	30828	30629

Table 4

Empirical FDRs (SEs) and powers (%) by mimicking the real data with α = 0.1 and α = 0.01 respectively, based on 100 replications.

p	method	α = 10%		α = 1%	
		FDR (SE) (in %)	Power (in %)	FDR (SE) (in %)	Power (in %)
1624	Group	6.2 (1.4)	47.2	0.5 (0.3)	33.6
	Entrywise	26.0 (2.7)	39.9	2.5 (1.5)	26.0
	B-H	11.1 (2.0)	44.9	1.1 (1.0)	19.8