

# Joint testing and false discovery rate control in high-dimensional multivariate regression

By YIN XIA

*Department of Statistics, School of Management, Fudan University, Shanghai 200433, China*  
xiayin@fudan.edu.cn

T. TONY CAI

*Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.*  
tcai@wharton.upenn.edu

AND HONGZHE LI

*Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.*  
hongzhe@upenn.edu

## SUMMARY

Multivariate regression with high-dimensional covariates has many applications in genomic and genetic research, in which some covariates are expected to be associated with multiple responses. This paper considers joint testing for regression coefficients over multiple responses and develops simultaneous testing methods with false discovery rate control. The test statistic is based on inverse regression and bias-corrected group lasso estimates of the regression coefficients and is shown to have an asymptotic chi-squared null distribution. A row-wise multiple testing procedure is developed to identify the covariates associated with the responses. The procedure is shown to control the false discovery proportion and false discovery rate at a prespecified level asymptotically. Simulations demonstrate the gain in power, relative to entrywise testing, in detecting the covariates associated with the responses. The test is applied to an ovarian cancer dataset to identify the microRNA regulators that regulate protein expression.

*Some key words:* Bias-corrected group lasso; Error rate control; Multiple phenotypes; Row-wise multiple testing.

## 1. INTRODUCTION

In genetic and genomic applications, multiple correlated phenotypes are often measured on the same individuals. Examples include genetic association studies, where many correlated phenotypic measurements are analysed jointly in the hope of increasing the power to detect causal genetic variants (Schifano et al., 2013; Zhou et al., 2015). In the study of gene expression, many genetic variants are associated with expression of multiple genes through trans-regulation (Grundberg et al., 2012). Identification of such trans-variants has been difficult due to limited sample sizes and multiple comparisons. In cancer genomic studies, microRNAs, or miRNAs, have been found to modulate many cellular processes, and protein-miRNA interactions are essential for post-transcriptional regulation in normal development and may be deregulated in cancer

(van Kouwenhove et al., 2011). The biological goal of these studies is to identify associations between miRNA and protein expression in order to understand the regulatory roles of miRNAs.

High-dimensional multivariate regression can be used in the applications mentioned above. One simple approach involves assessing the relationship between each response and each covariate individually and using a Bonferroni correction and false discovery rate control to account for the large number of tests conducted. This is often performed in genetic association analysis of gene expression data. Ruffieux et al. (2017) developed a Bayesian method for sparse multivariate regression models to select predictor-response pairs. Alternatively, one can apply a dimension reduction technique, such as principal component analysis of the responses, and test for association with the principal components rather than the individual responses (Suo et al., 2013). Classical Fisher combination tests have also been explored and applied (Yang et al., 2016).

By taking advantage of the similarity across multivariate responses, significant gains in statistical power can potentially be achieved in an association analysis. This motivates us to consider the following high-dimensional multivariate regression model, where  $D$  correlated responses are measured on  $n$  independent individuals, together with  $p$  covariates, where  $p$  can be much greater than  $n$ :

$$Y_{n \times D} = \mu_{n \times D} + X_{n \times p} B_{p \times D} + \Upsilon_{n \times D}, \quad (1)$$

where  $Y = (Y_{\cdot,1}, \dots, Y_{\cdot,D}) \in \mathbb{R}^{n \times D}$ , with  $Y_{\cdot,d} = (Y_{1,d}, \dots, Y_{n,d})^T$ , denotes the observed response matrix for  $n$  samples over  $D$  responses,  $\mu = (\mu_{\cdot,1}, \dots, \mu_{\cdot,D}) \in \mathbb{R}^{n \times D}$ , with  $\mu_{\cdot,d} = (\mu_{1,d}, \dots, \mu_{n,d})^T$ , is the mean response matrix, the rows of which are the same, and  $X = (X_{1,\cdot}^T, \dots, X_{n,\cdot}^T)^T \in \mathbb{R}^{n \times p}$  is the covariate matrix. In this model,  $B = (B_{\cdot,1}, \dots, B_{\cdot,D}) \in \mathbb{R}^{p \times D}$ , with  $B_{\cdot,d} = (B_{1,d}, \dots, B_{p,d})^T \in \mathbb{R}^p$ , represents the matrix of regression coefficients, where the  $i$ th row represents the regression coefficients of the  $i$ th covariate on the  $D$  responses, and  $D$  is fixed. Finally,  $\Upsilon = (\epsilon_{\cdot,1}, \dots, \epsilon_{\cdot,D}) \in \mathbb{R}^{n \times D}$  with  $\epsilon_{\cdot,d} = (\epsilon_{1,d}, \dots, \epsilon_{n,d})^T$ , where  $\{\epsilon_{k,d}\}$  are independent and identically distributed random variables with mean zero and variance  $\sigma_\epsilon^2$  which are independent of  $X$ .

To test whether the  $i$ th covariate is associated with any of the  $D$  responses, in this paper we develop an efficient procedure for simultaneously testing

$$H_{0,i} : B_{i,\cdot} = 0 \quad \text{versus} \quad H_{1,i} : B_{i,\cdot} \neq 0 \quad (i = 1, \dots, p) \quad (2)$$

while controlling the false discovery rate and false discovery proportion. Of particular interest is the scenario where the effects of the  $i$ th variable on each of the  $D$  responses share strong similarities; that is, if  $B_{i,d} \neq 0$ , then the rest of the entries in that row are more likely to be nonzero. Hence, a row-wise testing method using the group information should be more favourable than testing  $B$  column by column.

Currently there is significant interest in statistical inference for high-dimensional linear regression. In the ultra-sparse setting, Javanmard & Montanari (2013), van de Geer et al. (2014) and Zhang & Zhang (2014) considered constructing confidence intervals and testing for a given coordinate of a high-dimensional coefficient vector in a linear model, and proposed procedures based on the debiased lasso estimators. Cai & Guo (2017) studied adaptive confidence intervals for a general linear functional of the regression vector in the high-dimensional setting, and showed that the adaptivity depends on the sparsity of the regression and the loading vectors. Zhu & Bradic (2017) considered testing a single entry of the regression coefficient in the setting of nonsparse linear regression. These papers focused on inference for a given coordinate or a given linear functional, but simultaneous testing of all coordinates with false discovery rate control was

not considered. Furthermore, they only dealt with the regression setting with a single response, whereas here we consider joint testing of regression coefficients over multiple responses.

We develop a sum-of-squares-type test statistic for testing a given row of the regression coefficient matrix  $B$ , based on the covariance between the residuals from the fit of the regression model (1) and the corresponding  $pD$  inverse regression models introduced in § 2.2. To get an estimate of the error terms  $\Upsilon$  in (1), this model is reformulated and the group lasso algorithm applied to obtain a bias-corrected estimator of  $B$ , so as to make use of the group information. The test statistic is shown to have an asymptotic  $\chi^2$  distribution under the null hypothesis that the corresponding row of  $B$  is zero. A transformed statistic is introduced and its asymptotic power is studied. In addition, a new row-wise multiple testing procedure based on the normal quantile transformation is proposed and its theoretical error rate control investigated. It is shown numerically that for a range of settings, the row-wise testing procedure significantly outperforms the entrywise method. We apply our method to analyse an ovarian cancer genomic dataset. The results demonstrate that the proposed test provides a powerful tool for identifying the miRNA regulators that are associated with a set of important proteins related to cancer progression.

## 2. ESTIMATION AND ROW-WISE STATISTICAL TEST

### 2.1. Notation and definitions

We begin with some basic notation and definitions. For a vector  $a = (a_1, \dots, a_p)^T \in \mathbb{R}^p$ , define the  $l_q$ -norm by  $|a|_q = (\sum_{i=1}^p |a_i|^q)^{1/q}$  for  $1 \leq q \leq \infty$ . For subscripts, we use the convention that  $v_i$  stands for the  $i$ th entry of a vector  $v$  and  $M_{i,j}$  for the entry in the  $i$ th row and  $j$ th column of a matrix  $M$ . Throughout,  $n$  independent and identically distributed random samples  $\{Y_{k,d}, X_{k,\cdot} : k = 1, \dots, n; d = 1, \dots, D\}$  are observed, where  $X_{k,\cdot} = (X_{k,1}, \dots, X_{k,p})$  is a random vector with covariance matrix  $\Sigma$ . Define  $\Sigma^{-1} = \Omega = (\omega_{i,j})$ .

For any vector  $\mu_d \in \mathbb{R}^p$ , let  $\mu_{-i,d}$  denote the  $(p-1)$ -dimensional vector obtained by removing the  $i$ 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  $n \times p$  matrix  $A$ , let  $A_{i,-j}$  denote the  $i$ th row of  $A$  with its  $j$ th entry removed and let  $A_{-,j}$  denote the  $j$ th column of  $A$  with its  $i$ th entry removed. Let  $A_{-i,-j}$  denote the  $(n-1) \times (p-1)$  submatrix of  $A$  formed by removing its  $i$ th row and  $j$ th column, and let  $A_{-,j}$  denote the  $n \times (p-1)$  submatrix of  $A$  formed by removing the  $j$ th column. Let  $A_{i,\cdot}$  be the  $i$ th row of  $A$  and  $A_{\cdot,j}$  the  $j$ th column of  $A$ . Write  $\bar{A}_{\cdot,j} = n^{-1} \sum_{k=1}^n A_{k,j}$ ,  $\bar{A}_{-,j} = n^{-1} \sum_{k=1}^n A_{k,-j}$ ,  $\bar{A}_{\cdot,j} = (\bar{A}_{\cdot,j}, \dots, \bar{A}_{\cdot,j})_{n \times 1}^T$ , and  $\bar{A}_{(-,j)} = (\bar{A}_{-,j}^T, \dots, \bar{A}_{-,j}^T)_{n \times (p-1)}^T$ . Let  $\bar{A} = n^{-1} \sum_{k=1}^n A_{k,\cdot}$ .

For a matrix  $\Omega = (\omega_{i,j})_{p \times p}$ , the matrix elementwise infinity norm is defined by  $\|\Omega\|_\infty = \max_{1 \leq i,j \leq p} |\omega_{i,j}|$ . For a set  $\mathcal{H}$ ,  $|\mathcal{H}|$  denotes the cardinality. 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  $|a_n| \leq C|b_n|$  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  $\lim_{n \rightarrow \infty} a_n/b_n = 1$ .

### 2.2. Parameter estimation and construction of test statistics

In order to construct a row-wise testing procedure, an inverse regression approach is used to first obtain a nearly unbiased estimator of  $B$ . Based on model (1), for each  $i = 1, \dots, p$  and  $d = 1, \dots, D$ , we run linear regression by taking  $X_{k,i}$  as the response and  $(Y_{k,d}, X_{k,-i})$  as the covariates:

$$X_{k,i} = \alpha_{i,d} + (Y_{k,d}, X_{k,-i})\gamma_{i,d} + \eta_{k,i,d} \quad (k = 1, \dots, n) \quad (3)$$

where  $\eta_{k,i,d}$  has mean zero and variance  $\sigma_{i,d}^2$  and is uncorrelated with  $(Y_{k,d}, X_{k,-i})$ . The regression coefficients  $\gamma_{i,d} = (\gamma_{i,1,d}, \dots, \gamma_{i,p,d})^T$  satisfy

$$\gamma_{i,d} = -\sigma_{i,d}^2(-B_{i,d}/\sigma_\epsilon^2, B_{i,d}B_{-i,d}^\top/\sigma_\epsilon^2 + \Omega_{i,-i})^\top, \quad \sigma_{i,d}^2 = (B_{i,d}^2/\sigma_\epsilon^2 + \omega_{i,i})^{-1}.$$

Note that  $\text{cov}(\epsilon_{k,d}, \eta_{k,i,d}) = \text{cov}\{\epsilon_{k,d}, X_{k,i} - \alpha_{i,d} - (Y_{k,d}, X_{k,-i})\gamma_{i,d}\}$ . Because the  $\epsilon_{k,d}$  are independent of the random design matrix  $X$ ,  $r_{i,d} = \text{cov}(\epsilon_{k,d}, \eta_{k,i,d})$  can be expressed as  $-\gamma_{i,1,d} \text{cov}(\epsilon_{k,d}, Y_{k,d}) = -\gamma_{i,1,d} \sigma_\epsilon^2 = -\sigma_{i,d}^2 B_{i,d}$ . Thus the null hypotheses  $H_{0,i} : B_{i,\cdot} = 0$  are equivalent to

$$H_{0,i} : \sum_{d=1}^D (r_{i,d}/\sigma_{i,d}^2)^2 = 0 \quad (i = 1, \dots, p).$$

Test statistics can then be based on the estimates of  $\{r_{i,d}/\sigma_{i,d}^2 : i = 1, \dots, p; d = 1, \dots, D\}$ .

To obtain estimates of  $r_{i,d}$ , estimates of  $B$  and  $\gamma_{i,d}$  are first constructed. To utilize the row-wise similarity information, the estimator  $\hat{B} = (\hat{B}_{\cdot,1}, \dots, \hat{B}_{\cdot,D}) \in \mathbb{R}^{p \times D}$ , with  $\hat{B}_{\cdot,d} = (\hat{B}_{1,d}, \dots, \hat{B}_{p,d})^\top \in \mathbb{R}^p$ , is obtained by using the group lasso method proposed in [Yuan & Lin \(2006\)](#). Because the response variables and the covariates can be centred so that the observed mean is zero, without loss of generality we assume that  $\mu = 0$ . Specifically, write  $\text{vec}(Y) = (Y_{1,\cdot}, \dots, Y_{n,\cdot})^\top \in \mathbb{R}^{nD \times 1}$ ,  $\text{vec}(B) = (B_{1,\cdot}, \dots, B_{p,\cdot})^\top \in \mathbb{R}^{pD \times 1}$  and  $\text{vec}(\Upsilon) = (\epsilon_{1,\cdot}, \dots, \epsilon_{n,\cdot})^\top \in \mathbb{R}^{nD \times 1}$ . Model (1) can then be written as

$$\text{vec}(Y) = X \otimes I_{D \times D} \text{vec}(B) + \text{vec}(\Upsilon). \quad (4)$$

Let  $\text{vec}(\hat{B}) = (\hat{B}_{1,\cdot}, \dots, \hat{B}_{p,\cdot})^\top \in \mathbb{R}^{pD \times 1}$ . The group lasso estimator is defined as

$$\text{vec}(\hat{B}) = \arg \min \left\{ \frac{1}{2} \left\| \text{vec}(Y) - \sum_{j=1}^p (X \otimes I_{D \times D})_{\cdot, [(j-1)D+1]:jD} B_{j,\cdot} \right\|_2^2 + \lambda_n \sum_{j=1}^p \|B_{j,\cdot}\|_2 \right\}, \quad (5)$$

where  $\lambda_n$  is a tuning parameter. Selection of  $\lambda_n$  will be discussed in § 3.3. By Lemma A3 in the Appendix, if the row sparsity of  $B$  satisfies  $s(p) = o(n^{1/3}/\log p)$ , then

$$\max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_1 = O_p(a_{n1}), \quad \max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_2 = O_p(a_{n2}) \quad (6)$$

for some  $a_{n1}$  and  $a_{n2}$  such that

$$\max(a_{n1}a_{n2}, a_{n2}^2) = o\{(n \log p)^{-1/2}\}, \quad a_{n1} = o(1/\log p). \quad (7)$$

*Remark 1.* The estimator  $\hat{B}$  in (5) provides one way to estimate the regression coefficients  $B$ , and other methods can be applied to estimate it if the coefficients satisfy (6) and (7). On the other hand, row-wise estimation as in (5) is better than entrywise estimation because the entries in the same row of  $B$  have similar patterns.

Next, we construct estimates of  $\gamma_{i,d}$  in the inverse regression (3) that satisfy

$$\max_{1 \leq d \leq D} \max_{1 \leq i \leq p} |\hat{\gamma}_{i,d} - \gamma_{i,d}|_1 = O_p(a_{n1}), \quad \max_{1 \leq d \leq D} \max_{1 \leq i \leq p} |\hat{\gamma}_{i,d} - \gamma_{i,d}|_2 = O_p(a_{n2}), \quad (8)$$

which can be obtained via the lasso estimator ([Xia et al., 2015, 2018](#)); see § 3.3.

For  $k = 1, \dots, n$  and  $d = 1, \dots, D$ , define the residuals

$$\begin{aligned}\hat{\epsilon}_{k,\cdot} &= Y_{k,\cdot} - \bar{Y} - (X_{k,\cdot} - \bar{X})\hat{B}, \\ \hat{\eta}_{k,i,d} &= X_{k,i} - \bar{X}_i - \{Y_{k,d} - \bar{Y}_d, (X_{k,-i} - \bar{X}_{\cdot,-i})\}\hat{\gamma}_{i,d},\end{aligned}$$

where  $\hat{B}$  and  $\hat{\gamma}_{i,d}$  are estimators of  $B$  and  $\gamma_{i,d}$  that satisfy (7) and the following conditions by combining (6) and (8):

$$\begin{aligned}\max \left( \max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_1, \max_{1 \leq d \leq D} \max_{1 \leq i \leq p} |\hat{\gamma}_{i,d} - \gamma_{i,d}|_1 \right) &= O_p(a_{n1}), \\ \max \left( \max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_2, \max_{1 \leq d \leq D} \max_{1 \leq i \leq p} |\hat{\gamma}_{i,d} - \gamma_{i,d}|_2 \right) &= O_p(a_{n2}).\end{aligned}\tag{9}$$

We construct a nearly unbiased estimate of  $r_{i,d}$  as follows. Since  $r_{i,d} = \text{cov}(\epsilon_{k,d}, \eta_{k,i,d})$ , it can be estimated by the sample covariance between the error terms,  $R_{i,d} = n^{-1} \sum_{k=1}^n \epsilon_{k,d} \eta_{k,i,d}$ . By replacing the error terms with the residuals,  $R_{i,d}$  can be estimated by  $\tilde{r}_{i,d} = n^{-1} \sum_{k=1}^n \hat{\epsilon}_{k,d} \hat{\eta}_{k,i,d}$ . However, the bias induced by the estimated parameters exceeds the desired rate  $(n \log p)^{-1/2}$  and is not ignorable. By comparing  $\tilde{r}_{i,d}$  and  $R_{i,d}$ , Lemma A2 leads to

$$\tilde{r}_{i,d} = R_{i,d} + \tilde{\sigma}_\epsilon^2 (\gamma_{i,1,d} - \hat{\gamma}_{i,1,d}) + \tilde{\sigma}_{i,d}^2 (B_{i,d} - \hat{B}_{i,d}) + o_p\{(n \log p)^{-1/2}\}.$$

Define  $\hat{\sigma}_\epsilon^2 = (Dn)^{-1} \sum_{k=1}^n \sum_{d=1}^D \hat{\epsilon}_{k,d}^2$  and  $\hat{\sigma}_{i,d}^2 = n^{-1} \sum_{k=1}^n \hat{\eta}_{k,i,d}^2$  to be the empirical sample variances that satisfy

$$\max \left( |\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|, \max_{1 \leq i \leq p} |\hat{\sigma}_{i,d}^2 - \sigma_{i,d}^2| \right) = O_p\{(\log p/n)^{1/2}\}.$$

We have

$$\begin{aligned}\tilde{r}_{i,d} + \hat{\sigma}_\epsilon^2 \hat{\gamma}_{i,1,d} + \hat{\sigma}_{i,d}^2 \hat{B}_{i,d} &= n^{-1} \sum_{k=1}^n \{\epsilon_{k,d} \eta_{k,i,d} - E(\epsilon_{k,d} \eta_{k,i,d})\} \\ &\quad - \sigma_{i,d}^2 B_{i,d} (1 - \tilde{\sigma}_\epsilon^2 / \sigma_\epsilon^2 - \tilde{\sigma}_{i,d}^2 / \sigma_{i,d}^2) + o_p\{(n \log p)^{-1/2}\}.\end{aligned}$$

Thus, a debiased estimator for  $r_{i,d}$  can be defined by

$$\hat{r}_{i,d} = \tilde{r}_{i,d} + \hat{\sigma}_\epsilon^2 \hat{\gamma}_{i,1,d} + \hat{\sigma}_{i,d}^2 \hat{B}_{i,d}.$$

Under the null hypothesis  $H_{0,i}$ , the bias of  $\hat{r}_{i,d}$  is of order  $(n \log p)^{-1/2}$ , which is sufficiently small to construct the test statistics

$$T_{i,d} = \hat{r}_{i,d} / \hat{\sigma}_{i,d}^2 \quad (i = 1, \dots, p; d = 1, \dots, D).$$

Under  $H_{0,i}$ , we have  $|T_{i,d} - \tilde{U}_{i,d}| = o_p\{(n \log p)^{-1/2}\}$ , where

$$U_{i,d} = n^{-1} \sum_{k=1}^n \{\epsilon_{k,d} \eta_{k,i,d} - E(\epsilon_{k,d} \eta_{k,i,d})\}, \quad \tilde{U}_{i,d} = (B_{i,d} + U_{i,d}) / \sigma_{i,d}^2. \tag{10}$$

The variance of  $T_{i,d}$  can be approximated by

$$\theta_{i,d} \equiv \text{var}(\tilde{U}_{i,d}) = \text{var}(\epsilon_{k,d} \eta_{k,i,d} / \sigma_{i,d}^2) / n = (\sigma_\epsilon^2 / \sigma_{i,d}^2 + B_{i,d}^2) / n,$$

where  $\theta_{i,d}$  can be estimated by  $\hat{\theta}_{i,d} = (\hat{\sigma}_\epsilon^2 / \hat{\sigma}_{i,d}^2 + \hat{B}_{i,d}^2) / n$ . Define the standardized statistic

$$W_{i,d} = T_{i,d} / \hat{\theta}_{i,d}^{1/2} \quad (i = 1, \dots, p).$$

The final test statistic for (2) based on  $\{W_{i,d} : i = 1, \dots, p; d = 1, \dots, D\}$  is then defined as

$$S_i = \sum_{d=1}^D W_{i,d}^2 \quad (i = 1, \dots, p), \quad (11)$$

which will be studied in detail in § 2.3.

### 2.3. Asymptotic null distribution of $S_i$

To investigate the asymptotic null distribution of  $S_i$ , some regularity conditions are needed.

*Condition 1.* Assume that  $\log p = o(n^{1/5})$  and that for some constants  $C_0, C_1, C_2 > 0$ ,  $C_0^{-1} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C_0$ ,  $C_1^{-1} \leq \sigma_\epsilon^2 \leq C_1$ ,  $\|B\|_\infty \leq C_2$ , and  $\text{var}(Y_{k,d}) \leq C_2$ .

*Condition 2.* There exists some constant  $M > 0$  such that the quantities  $E\{\exp(M\epsilon_{k,d}^2)\}$  and  $\max_{\text{var}(a^\top X_{k,\cdot})=1} E[\exp\{M(a^\top X_{k,\cdot})^2\}]$  are finite.

*Condition 3.* Let  $\Lambda$  be the diagonal of  $\Omega$  and let  $(\xi_{i,j}) = R = \Lambda^{-1/2} \Omega \Lambda^{-1/2}$ . Assume that  $\max_{1 \leq i \leq j \leq p} |\xi_{i,j}| \leq \xi < 1$  for some constant  $0 < \xi < 1$ .

Condition 1, the eigenvalue condition, is common in high-dimensional settings (Cai et al., 2013; Liu, 2013; Xia et al., 2015). It implies that most of the variables are not highly correlated with each other and is sufficient to give Lemma A2 and to ensure error rate control in simultaneous inference, studied in § 3. The assumption  $\log p = o(n^{1/5})$  is sufficient for the simultaneous Gaussian approximations (A3) and (A4), as shown in the proof of Theorem 1. As we shall see from the numerical results in § 4,  $p$  can be much larger than  $n$  in practice. The same condition is commonly used (see, e.g., Cai et al., 2013; Xia et al., 2015), and stricter assumptions are also imposed, for example in Chang et al. (2017), where it was assumed that  $\log p = o(n^{1/7})$ . Condition 2 is a sub-Gaussian tail condition, and can be weakened to a polynomial tail condition if  $p < n^c$  for some constant  $c > 0$ . This condition is necessary for Gaussian approximation of the proposed test statistics. It is milder than the normal assumption used in the high-dimensional regression literature (Javanmard & Montanari, 2014; Zhang & Zhang, 2014). Condition 3 is also mild. For example, if  $\max_{1 \leq i \leq j \leq p} |\xi_{i,j}| = 1$ , then  $\Omega$  is singular.

The following theorem gives the asymptotic null distribution of  $S_i$ , for any  $i = 1, \dots, p$ .

**THEOREM 1.** Suppose that Conditions 1 and 2 and the asymptotic conditions (9) and (7) hold. Then under  $H_{0,i} : B_{i,\cdot} = 0$ , for any  $t \in \mathbb{R}$ ,

$$\text{pr}(S_i \leq t) \rightarrow \text{pr}(\chi_D^2 \leq t)$$

as  $n, p \rightarrow \infty$ .

The normal quantile transformation of  $S_i$ , defined as

$$Q_i = \Phi^{-1}\{1 - \text{pr}(\chi_D^2 \geq S_i)/2\} \quad (12)$$

where  $\Phi$  is the standard normal cumulative distribution function, asymptotically has the same distribution as the absolute value of a standard normal random variable. A test  $\Phi_\alpha^i$  can be defined as

$$\Phi_\alpha^i = I\{Q_i \geq \Phi^{-1}(1 - \alpha/2)\}, \quad (13)$$

where the hypothesis  $H_{0,i} : B_{i,\cdot} = 0$  is rejected whenever  $\Phi_\alpha^i = 1$ .

#### 2.4. Asymptotic power

To analyse the asymptotic power of the test  $\Phi_\alpha^i$  given in (13), for a given row index  $i$ , define the class of regression coefficients

$$\mathcal{W}_i(\alpha, \nu) = \left\{ B : \sum_{d=1}^D \frac{B_{i,d}^2}{\theta_{i,d}} \geq (2 + \delta)(\Psi_{1-\alpha}^2 + \Psi_{1-\nu}^2) \right\} \quad (14)$$

for any constant  $\delta > 0$ , where  $\Psi_{1-\alpha}$  is the  $1 - \alpha$  quantile of  $\chi_D^2$ .

The next theorem shows that the test  $\Phi_\alpha^i$  is able to asymptotically distinguish the null parameter set in which  $B_{i,\cdot} = 0$  from  $\mathcal{W}_i(\alpha, \nu)$  for an arbitrarily small constant  $\delta > 0$ , with  $\nu \rightarrow 0$ .

**THEOREM 2.** *Suppose that Conditions 1 and 2 and the asymptotic conditions (9) and (7) hold. Then as  $n, p \rightarrow \infty$ , for any  $\delta > 0$ ,*

$$\inf_{B \in \mathcal{W}_i(\alpha, \nu)} \text{pr}(\Phi_\alpha^i = 1) \geq 1 - \nu.$$

Since  $\theta_{i,d}$  is of order  $1/n$ , Theorem 2 shows that the proposed test rejects the null hypothesis  $H_{0,i} : B_{i,\cdot} = 0$  with high probability for a large class of regression coefficients satisfying the condition that there exists one entry  $B_{i,d}$  having a magnitude larger than  $C/n^{1/2}$  for  $C = \{(2 + \delta)(C_0 C_1 + C_2^2)(\Psi_{1-\alpha}^2 + \Psi_{1-\nu}^2)\}^{1/2}$ , where  $C_0, C_1$  and  $C_2$  are given in Condition 1.

### 3. MULTIPLE TESTING WITH ERROR RATE CONTROL

#### 3.1. Multiple testing algorithm

In this section, a row-wise multiple testing procedure is introduced with error rate control for testing the  $p$  hypotheses

$$H_{0,i} : B_{i,\cdot} = 0 \quad \text{versus} \quad H_{1,i} : B_{i,\cdot} \neq 0 \quad (i = 1, \dots, p).$$

Let  $\mathcal{H} = \{1, \dots, p\}$ , let  $\mathcal{H}_0 = \{i : B_{i,\cdot} = 0, i \in \mathcal{H}\}$  be the set of true nulls, and let  $\mathcal{H}_1 = \mathcal{H} \setminus \mathcal{H}_0$  be the set of true alternatives. We are interested in cases where most of the rows of  $B$  consist of zeros, that is, where  $|\mathcal{H}_1|$  is small relative to  $|\mathcal{H}|$ .

Theorem 1 shows that  $S_i$  is asymptotically chi-squared distributed, and, as discussed in § 2.3, the normal quantile transformation of  $S_i$  defined as  $Q_i = \Phi^{-1}\{1 - \text{pr}(\chi_D^2 \geq S_i)/2\}$  has approximately the same distribution as the absolute value of a standard normal random variable under the null  $H_{0,i}$ . Let  $t$  be the threshold level such that  $H_{0,i}$  is rejected if  $Q_i \geq t$ . For any given  $t$ ,



denote the total number of false positives by  $R_0(t) = \sum_{i \in \mathcal{H}_0} I(Q_i \geq t)$  and the total number of rejections by  $R(t) = \sum_{i \in \mathcal{H}} I(Q_i \geq t)$ . Then 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  $t$  is

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

which would reject as many true positives as possible while controlling the false discovery proportion at the prespecified level  $\alpha$ . Here  $R_0(t)$  can be estimated by  $2\{1 - \Phi(t)\}|\mathcal{H}_0|$  and  $|\mathcal{H}_0|$  is bounded above by  $p$ . Thus we conservatively estimate  $|\mathcal{H}_0|$  by  $p$  because of the row sparsity of the regression coefficients. Therefore, we propose the following multiple testing algorithm.

*Step 1.* Construct  $S_i$  by (11); then calculate the row-wise test statistics  $Q_i$  by (12), for  $i \in \mathcal{H}$ .

*Step 2.* For a given  $0 \leq \alpha \leq 1$ , calculate

$$\hat{t} = \inf\left[0 \leq t \leq (2 \log p - 2 \log \log p)^{1/2} : \frac{2p\{1 - \Phi(t)\}}{R(t) \vee 1} \leq \alpha\right]. \quad (15)$$

If (15) does not exist, then set  $\hat{t} = (2 \log p)^{1/2}$ .

*Step 3.* For  $i \in \mathcal{H}$ , reject  $H_{0,i}$  if  $Q_i \geq \hat{t}$ .

### 3.2. Theoretical error rate control

We now turn our attention to the theoretical error rate control of the proposed multiple testing algorithm. For any  $i \in \mathcal{H}$ , define

$$\Gamma_i(\gamma) = \{j : j \in \mathcal{H}, |\xi_{i,j}| \geq (\log p)^{-2-\gamma}\},$$

where  $\xi_{i,j}$  is defined in Condition 3. The following theorem shows that the proposed multiple testing procedure controls the false discovery proportion and false discovery rate at the prespecified level  $\alpha$  asymptotically.

**THEOREM 3.** Assume  $p_0 = |\mathcal{H}_0| \asymp p$  and that (9) and (7) hold. Suppose there exists some  $\gamma > 0$  such that  $\max_{i \in \mathcal{H}_0} |\Gamma_i(\gamma)| = o(p^\tau)$  for any  $\tau > 0$ . Then under Conditions 1–3 with  $p \leq cn^r$  for some  $c > 0$  and  $r > 0$ , we have

$$\limsup_{(n,p) \rightarrow \infty} \text{FDR}(\hat{t}) \leq \alpha,$$

and for any  $\epsilon > 0$  we have

$$\lim_{(n,p) \rightarrow \infty} \text{pr}\{\text{FDP}(\hat{t}) \leq \alpha + \epsilon\} = 1.$$

**Remark 2.** The condition on  $|\Gamma_i(\gamma)|$  ensures that most of the row estimates of  $B$  are not highly correlated with each other for those indices belonging to the true nulls  $\mathcal{H}_0$ , so as to control the variance of  $R_0(t)$  in order to control the false discovery proportion.



When  $\hat{t}$  is not attained in the range  $[0, (2 \log p - 2 \log \log p)^{1/2}]$  as described in (15), it is thresholded at  $(2 \log p)^{1/2}$ . The following theorem states a weak condition to ensure the existence of  $\hat{t}$  in the range; as a result, the false discovery rate and false discovery proportion will converge to the prespecified level  $\alpha$ .

THEOREM 4. *Let*

$$\mathcal{S}_\rho = \{i \in \mathcal{H} : \text{there exists } d \text{ with } 1 \leq d \leq D \text{ such that } |B_{i,d}|/\theta_{i,d}^{1/2} \geq (\log p)^{1/2+\rho}\}.$$

*Suppose that for some  $\rho > 0$  and some  $\delta > 0$ ,  $|\mathcal{S}_\rho| \geq \{1/(\pi^{1/2}\alpha) + \delta\}(\log p)^{1/2}$ . Suppose there exists some  $\gamma > 0$  such that  $\max_{i \in \mathcal{H}_0} |\Gamma_i(\gamma)| = o(p^\tau)$  for any  $\tau > 0$ . Assume that  $p_0 = |\mathcal{H}_0| \geq cp$  for some  $c > 0$  and that (9) and (7) hold. Then under Conditions 1–3 with  $p \leq cn^r$  for some  $c > 0$  and  $r > 0$ ,*

$$\lim_{(n,p) \rightarrow \infty} \frac{\text{FDR}(\hat{t})}{\alpha p_0/p} = 1, \quad \frac{\text{FDP}(\hat{t})}{\alpha p_0/p} \rightarrow 1$$

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

*Remark 3.* The condition on  $|\mathcal{S}_\rho|$  requires that a few rows of  $B$  have one entry with magnitude exceeding  $(\log p)^{1/2+\rho}/n^{1/2}$  for some constant  $\rho > 0$ , among  $p$  hypotheses in total, and is thus a mild condition. It is critical to restrict  $t$  to the range  $[0, (2 \log p - 2 \log \log p)^{1/2}]$  and to threshold  $Q_i$  at  $(2 \log p)^{1/2}$  when  $\hat{t}$  is not in this range. First of all, when  $t \geq (2 \log p - 2 \log \log p)^{1/2}$ ,  $2p\{1 - \Phi(t)\} \rightarrow 0$  and is not even a consistent estimate of the false rejections  $R_0(t)$  because  $|R_0(t)/[2p\{1 - \Phi(t)\}] - 1| \not\rightarrow 0$  in probability as  $(n, p) \rightarrow \infty$ . Hence, if we use  $2p\{1 - \Phi(t)\}$  as an estimate of  $R_0(t)$  for all  $t \leq (2 \log p)^{1/2}$ , it may not be able to control  $\text{FDP}(\hat{t})$  with positive probability. Secondly, it is critical to threshold  $Q_i$  at  $(2 \log p)^{1/2}$  instead of  $(2 \log p - 2 \log \log p)^{1/2}$ . When  $t$  does not exist in the range, thresholding  $Q_i$  at  $(2 \log p - 2 \log \log p)^{1/2}$  will cause too many false rejections, and consequently  $\text{FDR}(\hat{t})$  cannot be controlled asymptotically at level  $\alpha$ .

### 3.3. Algorithm details and tuning parameter selection

To obtain the row-wise test statistics  $Q_i$  in Step 1 of the above algorithm, we study the estimation of regression coefficients in models (1) and (3). As discussed in § 2.2, the regression coefficient matrix  $B$  can be estimated by first reformulating the model (1) as (4) and then applying the group lasso estimator of Yuan & Lin (2006) as in (5) with

$$\lambda_n = b\{8\hat{\sigma}_Y(1 + 5/2D^{-1} \log p)/(nD)\}^{1/2},$$

where  $\text{vec}(Y)$  and  $X$  are centred and  $\hat{\sigma}_Y$  is the sample variance of  $\text{vec}(Y)$ ; see, for example, Lounici et al. (2011). For the inverse regression models in (3),  $\gamma_{i,d}$  ( $d = 1, \dots, D$ ;  $i = 1, \dots, p$ ) is estimated by applying the lasso as follows:

$$\gamma_{i,d} = \Lambda_{i,d}^{-1/2} \arg \min_v \left( \frac{1}{2n} \left\| (Y_{\cdot,d}, X_{\cdot,-i}) - \{\bar{Y}_d, \bar{X}_{(-i)}\} \Lambda_{i,d}^{-1/2} v - (X_{\cdot,i} - \bar{X}_{\cdot,i}) \right\|_2^2 + \lambda_{i,n} |v|_1 \right), \quad (16)$$

where  $\lambda_{i,n} = b(\hat{\sigma}_{i,i} \log p/n)^{1/2}$  and  $\Lambda_{i,d} = \text{diag}(\hat{\sigma}_{Y_d}, \hat{\Sigma}_{-i,-i})$ , in which  $\hat{\sigma}_{Y_d}$  is the sample variance of  $Y_{k,d}$  and  $\hat{\Sigma} = (\hat{\sigma}_{i,j})$  is the sample covariance matrix of  $X_{k,\cdot}$ .

The tuning parameters  $\lambda_n$  and  $\lambda_{i,n}$  in (5) and (16) are selected adaptively using the data by matching the number of false rejections  $R_0(t)$  with its estimate  $2p\{1 - \Phi(t)\}$  as closely as possible. Since  $\mathcal{H}_0$  is unknown, the constant  $b$  in the tuning parameters is chosen to minimize

$$\int_c^1 \left[ \frac{\sum_{i \in \mathcal{H}} I\{Q_i^{(b)} \geq \Phi^{-1}(1 - \alpha/2)\}}{\alpha p} - 1 \right]^2 d\alpha,$$

where  $c > 0$  and  $Q_i^{(b)}$  is the statistic of the corresponding tuning parameter. A discretized criterion and the algorithm are summarized as follows.

*Step 1.* For  $b = 1, \dots, 40$  let  $\lambda_n = (b/20)[8\hat{\sigma}_Y\{1 + (5/2)\log p/D\}/(nD)]^{1/2}$  and  $\lambda_{i,n} = (b/20)(\hat{\sigma}_{i,i}\log p/n)^{1/2}$ . For each  $b$ , calculate  $\hat{B}^{(b)}$  and  $\hat{\gamma}_{i,d}^{(b)}$  ( $i = 1, \dots, p$ ;  $d = 1, \dots, D$ ). Based on the estimation of regression coefficients, construct the corresponding statistics  $S_i^{(b)}$  for each  $b$ .

*Step 2.* Choose  $\hat{b}$  to be the minimizer of

$$\hat{b} = \arg \min \sum_{s=1}^{10} \left[ \frac{\sum_{1 \leq i \leq p} I\{S_i^{(b)} \geq \Psi^{-1}(1 - s[1 - \Psi\{(2\log p)^{1/2}\}]/10)\}}{sp[1 - \Psi\{(2\log p)^{1/2}\}]/10} - 1 \right]^2.$$

Then the tuning parameters  $\lambda_n$  and  $\lambda_{i,n}$  are chosen by

$$\lambda_n = (\hat{b}/20)[8\hat{\sigma}_Y\{1 + (5/2)\log p/D\}/(nD)]^{1/2}, \quad \lambda_{i,n} = (\hat{b}/20)(\hat{\sigma}_{i,i}\log p/n)^{1/2}.$$

## 4. SIMULATION STUDIES

### 4.1. Data generation

The false discovery rate and power of the multiple testing procedure proposed in § 3.1 are evaluated by simulation. The data are generated by considering three matrix models, with covariates being a combination of continuous and discrete random variables. The proposed row-wise multiple testing procedure is compared with the entrywise testing method of Xia et al. (2018). The R code is available in the Supplementary Material.

The number of responses  $D$  is chosen to be 10. The design matrices  $X_{k\cdot}$  ( $k = 1, \dots, n$ ) are generated as in Xia et al. (2018), with some covariates continuous and others discrete. Let  $\Lambda = (\Lambda_{ij})$  be a diagonal matrix with  $\Lambda_{i,i} = \text{Un}(1, 3)$  for  $i = 1, \dots, p$ . The following three models are used to generate the design matrices.

Model 1 constructs  $\Omega^{*(1)} = (\omega_{ij}^{*(1)})$  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. Then the precision matrix is generated by  $\Omega^{(1)} = \Lambda^{1/2}\Omega^{*(1)}\Lambda^{1/2}$ .

Model 2 constructs  $\Omega^{*(2)} = (\omega_{ij}^{*(2)})$  where  $\omega_{i,j}^{*(2)} = \omega_{j,i}^{*(2)} = 0.5$  for  $i = 10(k-1) + 1$  and  $10(k-1) + 2 \leq j \leq 10(k-1) + 10$ , with  $1 \leq k \leq p/10$ , and  $\omega_{i,j}^{*(2)} = 0$  otherwise. Then the precision matrix is generated by  $\Omega^{(2)} = \Lambda^{1/2}(\Omega^{*(2)} + \delta I)/(1 + \delta)\Lambda^{1/2}$  with  $\delta = |\lambda_{\min}(\Omega^{*(2)})| + 0.05$ .

Model 3 constructs  $\Omega^{*(3)} = (\omega_{ij}^{*(3)})$  where  $\omega_{i,i}^{*(3)} = 1$ ,  $\omega_{i,j}^{*(3)} = 0.8 \times \text{Ber}(1, 0.05)$  for  $i < j$  and  $\omega_{j,i}^{*(3)} = \omega_{i,j}^{*(3)}$ . Then the precision matrix is generated by  $\Omega^{(3)} = \Lambda^{1/2}(\Omega^{*(3)} + \delta I)/(1 + \delta)\Lambda^{1/2}$  with  $\delta = |\lambda_{\min}(\Omega^{*(3)})| + 0.05$ .

For each of the three matrix models, independent and identically distributed samples  $X_{k,\cdot} \sim N(0, \Sigma^{(m)})$  ( $k = 1, \dots, n$ ) with  $m = 1, 2$  and  $3$  are obtained. Discrete covariates are simulated by  $l$  covariates of  $X_{k,\cdot,d}$  taking one of three discrete values,  $0, 1$  or  $2$ , with probability  $1/3$  each, where  $l$  is a random integer between  $\lfloor p/2 \rfloor$  and  $p$ . For the regression coefficient matrix  $B$ ,  $s$  nonzero rows are randomly selected, with  $s = 10, 20, 30$  and  $50$  for  $p = 50, 200, 500$  and  $1000$ , respectively. For the selected rows, we construct the regression coefficient matrix in the following two settings.

In setting 1, all entries in the nonzero rows are nonzero: the magnitudes of  $B_{i,d}$  are generated randomly from  $[-2(\log p/n)^{1/2}, -(\log p/n)^{1/2}] \cup [(\log p/n)^{1/2}, 2(\log p/n)^{1/2}]$  with  $n = 100$  and  $d = 1, \dots, D$ .

In setting 2, a small proportion of entries of the nonzero rows are nonzero: for each nonzero row, we randomly selected three nonzero locations  $\{l_1, l_2, l_3\}$  and generated the magnitudes of nonzero  $B_{i,d}$  randomly from  $[-4(\log p/n)^{1/2}, -2(\log p/n)^{1/2}] \cup [2(\log p/n)^{1/2}, 4(\log p/n)^{1/2}]$  with  $n = 100$  and  $d = l_1, l_2, l_3$ , and set  $B_{i,d} = 0$  otherwise.

The false discovery rate level is chosen as  $\alpha = 5\%$ . Based on 50 replications, the empirical false discovery rate and power are calculated by

$$\frac{1}{50} \sum_{l=1}^{50} \frac{\sum_{i \in \mathcal{H}_0} I(N_{i,l} \geq \hat{t})}{\sum_{i \in \mathcal{H}} I(N_{i,l} \geq \hat{t})}, \quad \frac{1}{50} \sum_{l=1}^{50} \frac{\sum_{i \in \mathcal{H}_1} I(N_{i,l} \geq \hat{t})}{|\mathcal{H}_1|},$$

where  $N_{i,l}$  denotes the normal quantile transformed statistics for the  $l$ th replication.

#### 4.2. Simulation results

Figure 1 shows that the new method controls the false discovery rate well with  $\alpha = 5\%$  for all three matrix models with randomly selected nonzero regression coefficients as described above, across the whole range of dimensions. In contrast, the entrywise method shows severe false discovery rate distortion when the same significance level  $\alpha$  is used. As an alternative, we can conservatively select the significance level for the entrywise method to be  $\alpha/D$ , for testing each of the  $D$  columns. The false discovery rate is well controlled by this corrected entrywise method for Models 1 and 2, but still suffers from distortion for Model 3 when the dimension is large.

Figure 1 shows that the proposed method has a clear power advantage over the two entrywise methods. Even for the uncorrected entrywise method, which has serious false discovery rate distortion, the empirical power is still much lower than that of our new procedure. When the dimension is large, i.e., when  $p = 1000$ , both entrywise methods suffer from trivial power, while the power of our method remains reasonably high.

In Fig. 2 one can observe similar behaviour to that in Fig. 1: the proposed method attains the desired error rate control in all models with various dimensions and sparsity levels. In setting 2, because only a small proportion of entries in the nonzero rows of the regression coefficient matrix  $B$  are nonzero, the entrywise method with the same significance level  $\alpha$  exhibits less severe false discovery rate distortion, though obvious deviation from the nominal level can still be seen, especially in Model 3. Furthermore, although the similarities between the columns of  $B$  are relatively weak in setting 2, the proposed method still shows significant power gain over the alternative procedures, especially as the dimension increases.

In summary, the numerical results suggest that when the regression coefficients share some row-wise similarities, groupwise testing is preferable to the entrywise method, in terms of both false discovery rate and power. The results of more simulations with weaker signal sizes are reported in the Supplementary Material.

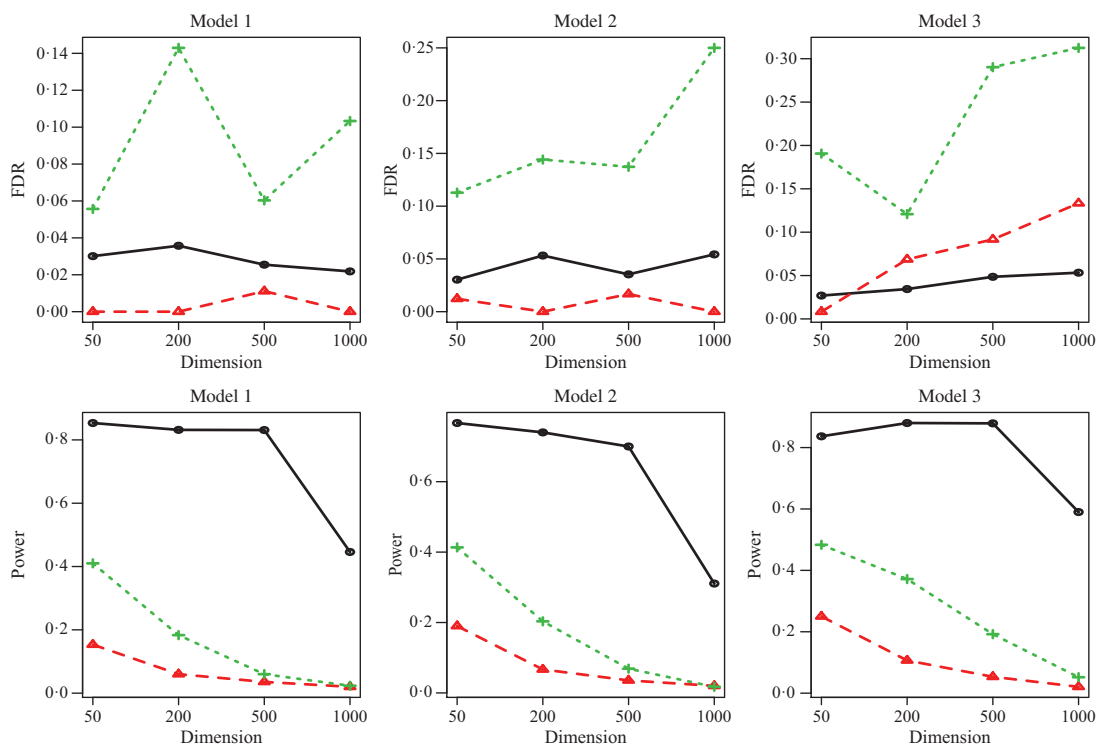


Fig. 1. Simulation setting 1: comparison of false discovery rate and power of the three methods, with  $\alpha = 5\%$ ; in each panel the solid line with circles represents the proposed row-wise method, the dotted line with crosses represents the entrywise method with significance level  $\alpha$ , and the dashed line with triangles represents the entrywise method with significance level  $\alpha/D$ .

## 5. REAL-DATA ANALYSIS

We apply the proposed testing procedures to an ovarian cancer dataset, with the goal of identifying the miRNA regulators that regulate the expression of ovarian cancer-related proteins. MicroRNAs are a family of small noncoding RNAs that regulate a wide array of biological processes, including carcinogenesis. In cancer cells, miRNAs have been found to be heavily dysregulated and affect the expression of genes and their protein products. To investigate the association between miRNA expression and protein expression in ovarian cancer, a total of 125 stage III and stage IV papillary serous primary ovarian cancer samples resected during debulking at the University of Turin were profiled for both miRNA expression and protein expression (Zsiros et al., 2015). A total of 3480 probes were profiled for miRNA expression and summarized as  $\log_2$  expression ratios. Among these, 3132 miRNAs were characterized and are used in our analysis. The protein expression data on these 125 samples were measured using reverse-phase protein arrays for a total of 195 proteins; reverse-phase protein array is a quantitative antibody-based technology that can be used to assess multiple protein markers in many samples in a cost-effective, sensitive and high-throughput manner (Li et al., 2013).

Since many of the proteins did not exhibit large variations in the ovarian cancer samples, they were first screened with the variance threshold set to 1, which resulted in the following 16 most variable proteins in our analysis: Annexin, Caveolin, Caveolin1, Claudin7, Cyclin.B1, E.cadherin, FAK.C, FAK.C1, HER2, HSP70, HSP70.1, IGFBP2, MAPK, p38, PARP, and PR.V. These proteins play functionally significant roles in ovarian cancer migration and invasion. For example, HER2 overexpression is a driving force in the carcinogenesis of several human cancers,

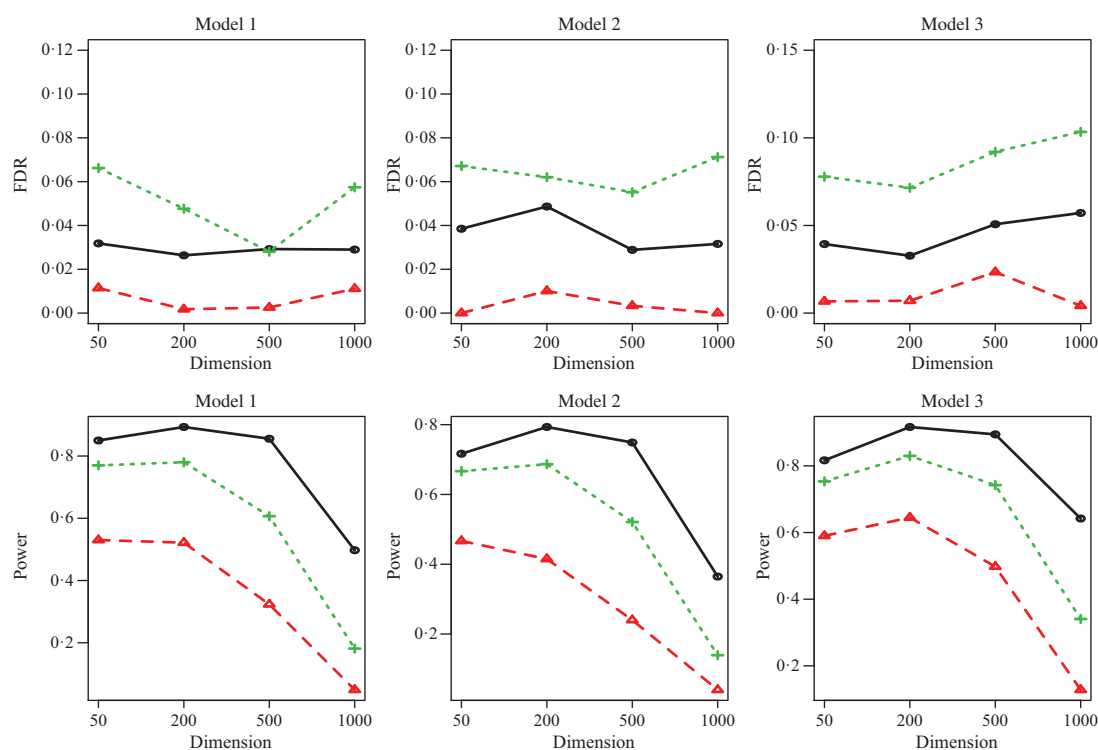


Fig. 2. Simulation setting 2: comparison of false discovery rate and power of the three methods, with  $\alpha = 5\%$ ; in each panel the solid line with circles represents the proposed row-wise method, the dotted line with crosses represents the entrywise method with significance level  $\alpha$ , and the dashed line with triangles represents the entrywise method with significance level  $\alpha/D$ .

including ovarian cancer (Pils et al., 2007). It has been reported that the overwhelming majority of human tumours overexpress members of the HSP70 family, and expression of these proteins is typically a marker of poor prognosis (Murphy, 2013). Insulin-like growth factor binding protein 2, IGFBP2, has been shown to enhance the invasion capacity of ovarian cancer cells (Lee et al., 2005). It is therefore important to identify the possible miRNA regulators that regulate the expression of such proteins in ovarian cancer.

Our goal is to identify the key protein regulators from the set of 3132 miRNAs using the data from these 125 ovarian cancer samples. Applying the proposed method, 25 miRNAs were identified as being associated with protein expression at a false discovery rate  $\alpha$ -level of 0.05; see Table 1. However, when the entrywise test was applied with an  $\alpha$ -level of 0.05/16, none of the miRNAs was selected. The identified miRNAs play various regulatory roles in cancer initiation and progression. Among them, miR-376a, miR-888, miR-187, miR-146a and miR-105 are associated with promotion of proliferation and metastasis, while miR-33b, miR-490-5p, miR-497, miR-596, miR-548-3p, miR-105, miR-185, miR-548i, miR-1247 and miR136 are associated with inhibition of cancer metastasis and progression. A few identified miRNAs, including miR-593, are also involved in protein kinase R, PKR, regulation and regulate cell proliferation. Table 1 lists supporting literature for the involvement of some of the identified miRNAs in ovarian cancer progression. These results show that miRNAs play important roles in regulating protein expression that are associated with ovarian cancer initiation and progression, and our proposed test provides a powerful tool for identifying such cancer-associated miRNAs.

For comparison, we also analysed this dataset using the method of Ruffieux et al. (2017), a Bayesian implementation of a sparse multivariate regression model that allows simultaneous

Table 1. *MicroRNAs that regulate protein expression in stage II to stage IV ovarian cancer, their biological functions, p-values from the proposed test, and relevant literature*

MicroRNA	Biological function	<i>p</i> -value	Reference
hsa-miR-376a	Promotes proliferation and metastasis	$3.65 \times 10^{-5}$	Yang et al. (2016)
hsa-miR-888	Cancer metastasis	$9.76 \times 10^{-6}$	Huang & Chen (2014)
hsa-miR-1288	Cancer location and pathological staging	$7.00 \times 10^{-6}$	Gopalan et al. (2014)
hsa-miR-33b	Inhibits cancer metastasis	$1.75 \times 10^{-4}$	Lin et al. (2015)
hsa-miR-490-5p	Tumour suppression	$6.86 \times 10^{-7}$	Lan et al. (2015)
hsa-miR-497	Inhibitory roles	$3.55 \times 10^{-4}$	Li et al. (2011)
mmu-miR-759	—	$3.22 \times 10^{-4}$	
has-RNU48	Tumour pathology and prognosis	$2.15 \times 10^{-4}$	Gee et al. (2011)
hsa-miR-596	Tumour suppression	$7.90 \times 10^{-5}$	Endo et al. (2013)
hsa-miR-187	Cancer progression	$5.49 \times 10^{-6}$	Chao et al. (2012)
has-RNU19	—	$1.68 \times 10^{-4}$	
hsa-miR-518f	—	$3.43 \times 10^{-5}$	
hsa-miR-548d-3p	Anti-oncogenic regulator	$9.73 \times 10^{-5}$	Shi et al. (2015)
hsa-miR-105	Inhibits tumour growth	$1.15 \times 10^{-4}$	Honeywell et al. (2013)
hsa-miR-185	Suppresses tumour growth	$2.92 \times 10^{-4}$	Imam et al. (2010)
hsa-miR-146a	Enhanced tumorigenic potential	$2.89 \times 10^{-4}$	Sandhu et al. (2014)
hsa-miR-548i	Anti-oncogenic regulator	$9.73 \times 10^{-5}$	Shi et al. (2015)
hsv1-miR-H6	—	$3.66 \times 10^{-4}$	
hsa-miR-129	Promotes apoptosis	$5.23 \times 10^{-7}$	Karaayvaz et al. (2013)
hsa-miR-1247	Inhibits cell proliferation	$2.13 \times 10^{-4}$	Shi et al. (2014)
hsa-miR-136	Inhibits cancer stem cell activity	$1.38 \times 10^{-4}$	Jeong et al. (2017)
hsa-miR-593	Regulates cell proliferation	$1.65 \times 10^{-6}$	Ito et al. (2011)
mmu-miR-105	Promotes metastasis	$1.15 \times 10^{-4}$	Zhou et al. (2014)
hsa-miR-886-5p	Associates with PKR and modulates its activity	$2.41 \times 10^{-4}$	Lee et al. (2011)
hsa-miR-553	—	$7.03 \times 10^{-5}$	

selection of predictors and associated responses. For each miRNA-protein pair, we obtained the marginal posterior inclusion probability. Among all the pairs, the maximum inclusion probability was 0.56, with only four pairs having inclusion probability over 0.25 and one pair having inclusion probability above 0.50. Using 400 permutations as suggested by Ruffieux et al. (2017), the procedure did not detect any miRNA-protein association at a false discovery rate level of 0.25. Since our test aims to test the association between a given miRNA and any of the proteins, for each miRNA we also took the maximum inclusion probability among all the proteins and calculated the false discovery rate using 400 permutations. Again, no association was identified for a false discovery rate level of 0.25. These results were similar to the entrywise test results.

#### ACKNOWLEDGEMENT

The research of Xia was supported in part by the National Natural Science Foundation of China, The Recruitment Program of Global Experts Youth Project from China, and Fudan University. Cai was supported in part by the U.S. National Science Foundation and National Institutes of Health. Li was partly supported by the National Cancer Institute. We thank Dr George Coukos for sharing the ovarian cancer datasets, and the associate editor and reviewer for very helpful comments.

#### SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes additional simulation results and the R code used in the simulations and for implementation of the proposed methods.

## APPENDIX

## Technical lemmas

In this section we prove the main theoretical results. We begin by stating lemmas that will be used in the proofs of the theorems. The following lemma was introduced in [Berman \(1962\)](#).

LEMMA A1. *If  $X$  and  $Y$  have a bivariate normal distribution with zero expectation, unit variance and correlation coefficient  $\rho$ , then*

$$\lim_{c \rightarrow \infty} \frac{\text{pr}(X > c, Y > c)}{\{2\pi(1 - \rho)^{1/2}c^2\}^{-1} \exp(-\frac{c^2}{1+\rho})(1 + \rho)^{1/2}} = 1$$

uniformly for all  $\rho$  such that  $|\rho| \leq \delta$ , for any  $\delta$  with  $0 < \delta < 1$ .

Based on the definitions of  $U_{i,d}$  and  $\tilde{U}_{i,d}$  in (10), the following lemma is essentially proved in an unpublished 2014 paper by W. Liu and S. Luo, available upon request from the first author.

LEMMA A2. *Suppose that Conditions 1 and 2 and the asymptotic conditions (9) and (7) hold. Then*

$$\tilde{r}_{i,d} = n^{-1} \sum_{k=1}^n (\epsilon_{k,d} - \bar{\epsilon})(\eta_{k,i,d} - \bar{\eta}_{i,d}) + \tilde{\sigma}_\epsilon^2 (\gamma_{i,1,d} - \hat{\gamma}_{i,1,d}) + \tilde{\sigma}_{i,d}^2 (B_{i,d} - \hat{B}_{i,d}) + o_p\{(n \log p)^{-1/2}\}$$

and, consequently,

$$T_{i,d} = \tilde{U}_{i,d} + (\tilde{\sigma}_\epsilon^2 / \sigma_\epsilon^2 + \tilde{\sigma}_{i,d}^2 / \sigma_{i,d}^2 - 2) B_{i,d} + o_p\{(n \log p)^{-1/2}\},$$

where  $\tilde{\sigma}_\epsilon^2 = (Dn)^{-1} \sum_{k=1}^n \sum_{d=1}^D (\epsilon_{k,d} - \bar{\epsilon})^2$  and  $\tilde{\sigma}_{i,d}^2 = n^{-1} \sum_{k=1}^n (\eta_{k,i,d} - \bar{\eta}_{i,d})^2$  with  $\bar{\epsilon} = (Dn)^{-1} \sum_{k=1}^n \sum_{d=1}^D \epsilon_{k,d}$  and  $\bar{\eta}_{i,d} = n^{-1} \sum_{k=1}^n \eta_{k,i,d}$ . As a result, uniformly in  $i = 1, \dots, p$ ,

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

We apply the group lasso estimator (5) with  $\lambda = \{8\hat{\sigma}_Y(1 + A \log p/D)/(nD)\}^{1/2}$ , taking  $A > 5/2$ , where  $\text{vec}(Y)$  and  $X$  are centralized and  $\hat{\sigma}_Y$  is the sample variance of  $\text{vec}(Y)$ . Then, by Corollary 4.1 in [Lounici et al. \(2011\)](#), we have the following lemma.

LEMMA A3. *Consider the model (4). Under Conditions 1 and 2, if  $D = O(\log p)$ , then  $\hat{B}$  satisfies*

$$\max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_1 = O_p\{s(p)(\log p/n)^{1/2}\}, \quad \max_{1 \leq d \leq D} |\hat{B}_{\cdot,d} - B_{\cdot,d}|_2 = O_p[\{s(p) \log p/n\}^{1/2}],$$

where  $s(p) = \sum_{i=1}^p I(B_{i,\cdot} \neq 0)$  is the row sparsity of  $B$ .

## Proof of Theorem 1

Based on the definitions of  $U_{i,d}$  and  $\tilde{U}_{i,d}$  in (10), let  $V_{i,d} = U_{i,d}/(\sigma_{i,1}^2 \theta_{i,d}^{1/2})$ , where  $\theta_{i,d} = \text{var}(\tilde{U}_{i,d}) = \text{var}(\epsilon_{k,d} \eta_{k,i,d} / \sigma_{i,d}^2) / n = (\sigma_\epsilon^2 / \sigma_{i,d}^2 + B_{i,d}^2) / n$  ( $d = 1, \dots, D$ ). By Lemma 2 in [Xia et al. \(2015\)](#), under conditions (9) and (7) we have

$$|\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| = O_p\{(\log p/n)^{1/2}\}, \quad \max_{i,d} |\hat{\sigma}_{i,d}^2 - \sigma_{i,d}^2| = O_p\{(\log p/n)^{1/2}\}.$$

Therefore

$$\max_{i,d} |\hat{\theta}_{i,d} - \theta_{i,d}| = o_p\{1/(n \log p)\}. \quad (\text{A1})$$



Lemma A2, together with (A1), implies that uniformly in  $1 \leq i \leq p$ ,  $W_{i,d} = V_{i,d} + o_p\{(\log p)^{-1/2}\}$ ; thus we have that uniformly in  $1 \leq i \leq p$ ,  $S_i = \sum_{d=1}^D V_{i,d}^2 + o_p(1)$ . Hence, it suffices to show that  $\sum_{d=1}^D V_{i,d}^2$  converges to  $\chi_D^2$  in distribution. Define  $Z_{k,i,d} = \{\epsilon_{k,d}\eta_{k,i,d} - E(\epsilon_{k,d}\eta_{k,i,d})\}/\sigma_{i,d}^2$  for  $1 \leq k \leq n$ . Then

$$V_{i,d} = \sum_{k=1}^n Z_{k,i,d} / (n^2 \theta_{k,d})^{1/2}.$$

Without loss of generality, we assume  $\sigma_\epsilon^2 = \sigma_{i,d}^2 = 1$ . Define  $\hat{V}_{i,d} = \sum_{k=1}^n \hat{Z}_{k,i,d} / (n^2 \theta_{k,d})^{1/2}$ , where  $\hat{Z}_{k,i,d} = Z_{k,i,d} I(|Z_{k,i,d}| \leq \tau_n) - E\{Z_{k,i,d} I(|Z_{k,i,d}| \leq \tau_n)\}$  with  $\tau_n = (4/M) \log(p+n)$ . Note that

$$\begin{aligned} & \max_{1 \leq i \leq p} n^{-1/2} \sum_{k=1}^n E[|Z_{k,i,d}| I\{|Z_{k,i,d}| \geq (4/M) \log(p+n)\}] \\ & \leq Cn^{1/2} \max_{1 \leq k \leq n} \max_{1 \leq i \leq p} E[|Z_{k,i,d}| I\{|Z_{k,i,d}| \geq (4/M) \log(p+n)\}] \\ & \leq Cn^{1/2} (p+n)^{-2} \max_{1 \leq k \leq n} \max_{1 \leq i \leq p} E[|Z_{k,i,d}| \exp\{(M/2)|Z_{k,i,d}|\}] \\ & \leq Cn^{1/2} (p+n)^{-2}. \end{aligned}$$

Hence

$$\Pr\left\{\max_{1 \leq i \leq p} |V_{i,d} - \hat{V}_{i,d}| \geq (\log p)^{-1}\right\} \leq \Pr\left(\max_{1 \leq i \leq p} \max_{1 \leq k \leq n} |Z_{k,i,d}| \geq \tau_n\right) = O(p^{-1}).$$

Because of the fact that

$$\begin{aligned} \left| \max_{1 \leq i \leq p} \sum_{d=1}^D V_{i,d}^2 - \max_{1 \leq i \leq q} \sum_{d=1}^D \hat{V}_{i,d}^2 \right| & \leq 2D \max_{1 \leq i \leq p} \max_{1 \leq d \leq D} |\hat{V}_{i,d}| \max_{1 \leq i \leq q} \max_{1 \leq d \leq D} |V_{i,d} - \hat{V}_{i,d}| \\ & + D \max_{1 \leq i \leq q} \max_{1 \leq d \leq D} |V_{i,d} - \hat{V}_{i,d}|^2, \end{aligned}$$

it suffices to prove that for any  $t \in \mathbb{R}$ ,

$$\Pr\left(\sum_{d=1}^D \hat{V}_{i,d}^2 \leq t\right) \rightarrow \Pr(\chi_D^2 \leq t). \quad (\text{A2})$$

Let  $\tilde{Z}_{k,i,d} = \hat{Z}_{k,i,d} / \theta_{i,d}^{1/2}$  for  $i = 1, \dots, q$  and  $W_{k,i} = (\tilde{Z}_{k,i,1}, \dots, \tilde{Z}_{k,i,D})$  for  $1 \leq k \leq n$ . Then

$$\Pr\left(\sum_{d=1}^D \hat{V}_{i,d}^2 \geq t\right) = \Pr\left(\left|n^{-1/2} \sum_{k=1}^n W_{k,i}\right|_2^2 \geq t\right).$$

It follows from Theorem 1 in Zaitsev (1987) that

$$\Pr\left(\left|n^{-1/2} \sum_{k=1}^n W_{k,i}\right|_2^2 \geq t\right) \leq \Pr\{|N^{(D)}|_2^2 \geq t - \epsilon_n (\log p)^{-1/2}\} + c_1 d^{5/2} \exp\left\{-\frac{n^{1/2} \epsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}}\right\} \quad (\text{A3})$$

and that

$$\Pr\left(\left|n^{-1/2} \sum_{k=1}^n W_{k,i}\right|_2^2 \geq t\right) \geq \Pr\{|N^{(D)}|_2^2 \geq t + \epsilon_n (\log p)^{-1/2}\} - c_1 d^{5/2} \exp\left\{-\frac{n^{1/2} \epsilon_n}{c_2 d^3 \tau_n (\log p)^{1/2}}\right\}, \quad (\text{A4})$$

where  $c_1 > 0$  and  $c_2 > 0$  are constants,  $\epsilon_n \rightarrow 0$  will be specified later, and  $N^{(D)} = (N_1, \dots, N_D)$  is a normal random vector with  $E\{N^{(D)}\} = 0$  and  $\text{cov}\{N^{(D)}\} = \text{cov}(W_{k,i})$ . Because  $\Upsilon$  is independent of  $X$  and  $\eta_{k,i,d_2} = X_{k,i} - (Y_{k,d_2}, X_{k,-i})\gamma_{i,d_2}$  with  $Y_{k,d_2} = X_{k,B_{\cdot,d_2}} + \epsilon_{k,d_2}$ , we have that for  $d_1 \neq d_2$ ,  $\epsilon_{k,d_1}$  and  $\eta_{k,i,d_2}$  are independent of each other. On the other hand, by definition,  $\epsilon_{k,d}$  and  $\eta_{k,i,d} + \gamma_{i,1,d}\epsilon_{k,d}$  are independent of each other. Hence, under the null  $H_{0,i}$ , we have that the  $\epsilon_{k,d}$  are independent of  $\eta_{k,i,d}$ . Thus we have

$$\text{cov}(\epsilon_{k,d_1} \eta_{k,i,d_1}, \epsilon_{k,d_2} \eta_{k,i,d_2}) = E(\epsilon_{k,d_1} \eta_{k,i,d_1} \epsilon_{k,d_2} \eta_{k,i,d_2}) = 0.$$

So  $\text{cov}\{N^{(D)}\} = \text{cov}(W_{k,i}) = I_{D \times D}$ . This, together with (A3) and (A4), proves (A2), and Theorem 1 then follows.

### Proof of Theorem 2

By Lemma A2, we have

$$\max_{1 \leq d \leq D} \left| \frac{T_{i,d} - \{1 + o(1)\}ET_{i,d}}{\hat{\theta}_{i,d}^{1/2}} - U_{i,d} \right| = o_p\{(\log p)^{-1/2}\}.$$

Observe that

$$\sum_{1 \leq d \leq D} \frac{[1 + o(1)]ET_{i,d}^2}{\hat{\theta}_{i,d}} \leq 2 \left( S_i + \sum_{1 \leq d \leq D} \frac{[T_{i,d} - \{1 + o(1)\}ET_{i,d}]^2}{\hat{\theta}_{i,d}} \right).$$

Also note that  $E(T_{i,d}) = B_{i,d}\{1 + o(1)\} + o\{(n \log p)^{-1/2}\}$ . The result then follows from the definition (14) of the classes of precision matrices and Theorem 1.

### Proof of Theorem 3

We start by showing that  $\text{pr}[\sum_{i \in \mathcal{H}_0} I\{|Q_i| \geq (2 \log p)^{1/2}\} = 0] \rightarrow 1$  as  $(n, p) \rightarrow \infty$ , and then we focus on the event  $\{\hat{t} \text{ in (15) exists}\}$ . Next, we show the false discovery proportion result by dividing the null set into small subsets and controlling the variance of  $R_0(t)$  for each subset, and thus the false discovery rate result will also be proved. Note that

$$\text{pr} \left[ \sum_{i \in \mathcal{H}_0} I\{|Q_i| \geq (2 \log p)^{1/2}\} \geq 1 \right] \leq p_0 \max_{i \in \mathcal{H}_0} \text{pr}\{|Q_i| \geq (2 \log p)^{1/2}\}.$$

By Theorem 1,  $\text{pr}(\max_{i \in \mathcal{H}_0} \max_{1 \leq d \leq D} |W_{i,d} - \hat{V}_{i,d}| = o\{(\log p)^{-1/2}\}) = 1$ . Then, on the event  $\{\max_{i \in \mathcal{H}_0} \max_{1 \leq d \leq D} |W_{i,d} - \hat{V}_{i,d}| = o\{(\log p)^{-1/2}\}$ , by (A3), (A4) and the fact that  $G(t + o\{(\log p)^{-1/2}\})/G(t) = 1 + o(1)$  uniformly in  $0 \leq t \leq (2 \log p)^{1/2}$ , where  $G(t) = 2\{1 - \Phi(t)\}$ , we have

$$\text{pr} \left[ \sum_{i \in \mathcal{H}_0} I\{|Q_i| \geq (2 \log p)^{1/2}\} \geq 1 \right] \leq p_0 G\{(2 \log p)^{1/2}\} \{1 + o(1)\} = o(1).$$

Hence we shall focus on the event  $\{\hat{t} \text{ exists in the range } [0, (2 \log p - 2 \log \log p)^{1/2}]\}$ . By definition of  $\hat{t}$ , it is easy to show that

$$\frac{2\{1 - \Phi(\hat{t})\}p}{\max\{\sum_{i \in \mathcal{H}} I(|Q_i| \geq \hat{t}), 1\}} = \alpha.$$

Therefore it suffices to show that

$$\sup_{0 \leq t \leq (2 \log p - 2 \log \log p)^{1/2}} \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|Q_i| \geq t) - G(t)\}}{pG(t)} \right| \rightarrow 0$$

in probability. Let  $0 \leq t_0 < t_1 < \dots < t_b = t_p$  be such that  $t_i - t_{i-1} = v_p$  for  $1 \leq i \leq b-1$  and  $t_b - t_{b-1} \leq v_p$ , where  $v_p = \{\log p(\log_4 p)\}^{-1/2}$ . Thus we have  $b \sim t_p/v_p$ . For any  $t$  such that  $t_{i-1} \leq t \leq t_i$ ,

$$\frac{\sum_{i \in \mathcal{H}_0} I(|Q_i| \geq t_i)}{p_0 G(t_i)} \frac{G(t_i)}{G(t_{i-1})} \leq \frac{\sum_{i \in \mathcal{H}_0} I(|Q_i| \geq t)}{p_0 G(t)} \leq \frac{\sum_{i \in \mathcal{H}_0} I(|Q_i| \geq t_{i-1})}{p_0 G(t_{i-1})} \frac{G(t_{i-1})}{G(t)}.$$

So it suffices to prove that

$$\max_{0 \leq i \leq b} \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|Q_i| \geq t_i) - G(t_i)\}}{pG(t_i)} \right| \rightarrow 0$$

in probability. Define  $F_i = \sum_{1 \leq d \leq D} \hat{V}_{i,d}^2$  and  $M_i = \Phi^{-1}\{1 - \text{pr}(\chi_D^2 \geq F_i)/2\}$ . By the proof of Theorem 1, we have  $\max_{i \in \mathcal{H}_0} |S_i - F_i| = o_p(1)$ , because

$$\text{pr}\{\chi_D^2 \geq t + o(1)\}/\text{pr}\{\chi_D^2 \geq t\} = 1 + o(1)$$

for all  $0 \leq t \leq c \log p$  with any constant  $c > 0$ . Recall that  $Q_i = \Phi^{-1}\{1 - \text{pr}(\chi_D^2 \geq S_i)/2\}$ . Note that  $G[t + o\{(\log p)^{-1/2}\}]/G(t) = 1 + o(1)$  uniformly in  $0 \leq t \leq (2 \log p)^{1/2}$ . Based on the proof of Theorem 1, it suffices to show that

$$\max_{0 \leq i \leq b} \left| \frac{\sum_{i \in \mathcal{H}_0} [I(|M_i| \geq t_i) - G(t_i)]}{p_0 G(t_i)} \right| \rightarrow 0$$

in probability. Note that

$$\begin{aligned} \text{pr} \left[ \max_{0 \leq i \leq b} \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|M_i| \geq t_i) - G(t_i)\}}{p_0 G(t_i)} \right| \geq \epsilon \right] &\leq \sum_{i=1}^b \text{pr} \left[ \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|M_i| \geq t_i) - G(t_i)\}}{p_0 G(t_i)} \right| \geq \epsilon \right] \\ &\leq \frac{1}{v_p} \int_0^{t_p} \text{pr} \left\{ \left| \frac{\sum_{i \in \mathcal{H}_0} I(|M_i| \geq t)}{p_0 G(t)} - 1 \right| \geq \epsilon \right\} dt + \sum_{i=b-1}^b \text{pr} \left[ \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|M_i| \geq t_i) - G(t_i)\}}{p_0 G(t_i)} \right| \geq \epsilon \right]. \end{aligned}$$

Therefore, it suffices to show that for any  $\epsilon > 0$ ,

$$\int_0^{t_p} \text{pr} \left[ \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|M_i| \geq t) - \text{pr}(I(|M_i| \geq t))\}}{p_0 G(t)} \right| \geq \epsilon \right] dt = o(v_p). \quad (\text{A5})$$

Note that

$$\begin{aligned} E \left| \frac{\sum_{i \in \mathcal{H}_0} \{I(|M_i| \geq t) - \text{pr}(I(|M_i| \geq t))\}}{p_0 G(t)} \right|^2 \\ = \frac{\sum_{i,j \in \mathcal{H}_0} \{\text{pr}(|M_i| \geq t, |M_j| \geq t) - \text{pr}(|M_i| \geq t)\text{pr}(|M_j| \geq t)\}}{p_0^2 G^2(t)}. \end{aligned}$$

We divide the indices  $i, j \in \mathcal{H}_0$  into three subsets:  $\mathcal{H}_{01} = \{i, j \in \mathcal{H}_0 : i = j\}$ ,  $\mathcal{H}_{02} = \{i, j \in \mathcal{H}_0 : i \neq j, i \in \Gamma_j(\gamma) \text{ or } j \in \Gamma_i(\gamma)\}$ , which contains the highly correlated pairs, and  $\mathcal{H}_{03} = \mathcal{H}_0 \setminus (\mathcal{H}_{01} \cup \mathcal{H}_{02})$ . Then we have

$$\frac{\sum_{i,j \in \mathcal{H}_{01}} \{\text{pr}(|M_i| \geq t, |M_j| \geq t) - \text{pr}(|M_i| \geq t) \text{pr}(|M_j| \geq t)\}}{p_0^2 G^2(t)} \leq \frac{C}{p_0 G(t)}. \quad (\text{A6})$$

Note that  $\text{cov}(\epsilon_{k,d} \eta_{k,i,d}, \epsilon_{k,d} \eta_{k,j,d}) = E(\epsilon_{k,d}^2 \eta_{k,i,d} \eta_{k,j,d}) - E(\epsilon_{k,d} \eta_{k,i,d}) E(\epsilon_{k,d} \eta_{k,j,d})$ . Because  $\text{cov}(\epsilon_{k,d}, \eta_{k,i,d}) = -\sigma_{i,d}^2 B_{i,d}$ , we have  $E(\epsilon_{k,d} \eta_{k,i,d}) E(\epsilon_{k,d} \eta_{k,j,d}) = \sigma_{i,d}^2 \sigma_{j,d}^2 B_{i,d} B_{j,d}$ . Note that  $E(\epsilon_{k,d}^2 \eta_{k,i,d} \eta_{k,j,d})$  equals

$$\begin{aligned} & E\{\epsilon_{k,d}^2 (\eta_{k,i,d} + \epsilon_{k,d} \gamma_{i,1,d})(\eta_{k,j,d} + \epsilon_{k,d} \gamma_{j,1,d})\} \\ & - E\{\epsilon_{k,d}^2 (\eta_{k,i,d} + \epsilon_{k,d} \gamma_{i,1,d}) \epsilon_{k,d} \gamma_{j,1,d}\} - E\{\epsilon_{k,d}^3 \gamma_{i,1,d} \eta_{k,j,d}\}. \end{aligned}$$

Since  $\epsilon_{k,d}$  is uncorrelated with  $\eta_{k,i,d} + \epsilon_{k,d} \gamma_{i,1,d}$ , we have  $-\gamma_{i,1,d} \text{var}(\epsilon_{k,d}) = \text{cov}(\epsilon_{k,d} \eta_{k,i,d})$ . Thus,

$$E(\epsilon_{k,d}^2 \eta_{k,i,d} \eta_{k,j,d}) = \sigma_\epsilon^2 E\{(\eta_{k,i,d} + \epsilon_{k,d} \gamma_{i,1,d})(\eta_{k,j,d} + \epsilon_{k,d} \gamma_{j,1,d})\} - E(\epsilon_{k,d}^3 \gamma_{i,1,d} \eta_{k,j,d}).$$

Note that

$$E(\epsilon_{k,d}^3 \gamma_{i,1,d} \eta_{k,j,d}) = E\{\epsilon_{k,d}^3 \gamma_{i,1,d} (\eta_{k,j,d} + \epsilon_{k,d} \gamma_{j,1,d})\} - E(\epsilon_{k,d}^4 \gamma_{i,1,d} \gamma_{j,1,d}) = -3 \gamma_{i,1,d} \gamma_{j,1,d} \sigma_\epsilon^4$$

and that

$$\begin{aligned} & E\{(\eta_{k,i,d} + \epsilon_{k,d} \gamma_{i,1,d})(\eta_{k,j,d} + \epsilon_{k,d} \gamma_{j,1,d})\} \\ & = \text{cov}(\eta_{k,i,d}, \eta_{k,j,d}) + \gamma_{i,1,d} \text{cov}(\epsilon_{k,d}, \eta_{k,j,d}) + \gamma_{j,1,d} \text{cov}(\epsilon_{k,d}, \eta_{k,i,d}) + \gamma_{i,1,d} \gamma_{j,1,d} \sigma_\epsilon^2. \end{aligned}$$

We have  $\text{cov}(\epsilon_{k,d} \eta_{k,i,d}, \epsilon_{k,d} \eta_{k,j,d}) = (\omega_{i,j} \sigma_\epsilon^2 + 2B_{i,d} B_{j,d}) \sigma_{i,d}^2 \sigma_{j,d}^2$ .

Therefore, for  $i, j \in \mathcal{H}_0$ ,

$$\tilde{\xi}_{i,j,d} = \text{corr}(\epsilon_{k,d} \eta_{k,i,d}, \epsilon_{k,d} \eta_{k,j,d}) = \frac{(\omega_{i,j} \sigma_\epsilon^2 + 2B_{i,d} B_{j,d})}{\{(\omega_{i,i} \sigma_\epsilon^2 + 2B_{i,d}^2)(\omega_{j,j} \sigma_\epsilon^2 + 2B_{j,d}^2)\}^{1/2}} = \xi_{i,j}.$$

By the proof of Theorem 1 we also have, for  $d_1 \neq d_2$ ,  $\text{corr}(\epsilon_{k,d_1} \eta_{k,i,d_1}, \epsilon_{k,d_2} \eta_{k,j,d_2}) = 0$ . Thus we have  $|\text{corr}(M_i, M_j)| \leq \xi < 1$ , where  $\xi$  is defined in Condition 3. Hence, by Lemma A1 and Lemma 6.2 in Liu (2013),

$$\begin{aligned} & \frac{\sum_{i,j \in \mathcal{H}_{02}} \{\text{pr}(|M_i| \geq t, |M_j| \geq t) - \text{pr}(|M_i| \geq t) \text{pr}(|M_j| \geq t)\}}{p_0^2 G^2(t)} \\ & \leq C \frac{p^{1+\tau} t^{-2} \exp\{-t^2/(1+\xi)\}}{p^2 G(t)} \leq \frac{C}{p^{1-\tau} \{G(t)\}^{2\xi/(1+\xi)}}. \end{aligned} \quad (\text{A7})$$

It remains to consider the subset  $\mathcal{H}_{03}$ , in which  $M_i$  and  $M_j$  are weakly correlated with each other. It is easy to check that  $\max_{i,j \in \mathcal{H}_{03}} \text{pr}(|M_i| \geq t, |M_j| \geq t) = [1 + O\{(\log p)^{-1-\gamma}\}] G^2(t)$ .

Thus we have

$$\frac{\sum_{i,j \in \mathcal{H}_{03}} \{\text{pr}(|M_i| \geq t, |M_j| \geq t) - \text{pr}(|M_i| \geq t) \text{pr}(|M_j| \geq t)\}}{p_0^2 G^2(t)} = O\{(\log p)^{-1-\gamma}\}. \quad (\text{A8})$$

Equation (A5) follows by combining (A6), (A7) and (A8).

## Proof of Theorem 4

Under the conditions of Theorem 4, we have

$$\sum_{i \in \mathcal{H}} I\{|Q_i| \geq (2 \log p)^{1/2}\} \geq \{1/(\pi^{1/2}\alpha) + \delta\}(\log p)^{1/2}$$

with probability tending to 1. Hence, with probability approaching 1, we have

$$\frac{p}{\sum_{i \in \mathcal{H}} I\{|Q_i| \geq (2 \log p)^{1/2}\}} \leq p\{1/(\pi^{1/2}\alpha) + \delta\}^{-1}(\log p)^{-1/2}.$$

Let  $t_p = (2 \log p - 2 \log \log p)^{1/2}$ . 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 proposed multiple testing algorithm in § 3.1; that is,  $\text{pr}(\hat{t} \text{ exists in } [0, t_p]) \rightarrow 1$ . Theorem 4 then follows from the proof of Theorem 3.

## REFERENCES

- BERMAN, S. M. (1962). A law of large numbers for the maximum in a stationary Gaussian sequence. *Ann. Math. Statist.* **33**, 93–7.
- CAI, T., LIU, W. & XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Am. Statist. Assoc.* **108**, 265–77.
- CAI, T. T. & GUO, Z. (2017). Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *Ann. Statist.* **45**, 615–46.
- CHANG, J., ZHOU, W., ZHOU, W.-X. & WANG, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73**, 31–41.
- CHAO, A., LIN, C. Y., LEE, Y. S., TSAI, C. L., WEI, P. C., HSUEH, S., WU, T. I., TSAI, C. N., WANG, C. J., CHAO, A. S. et al. (2012). Regulation of ovarian cancer progression by microRNA-187 through targeting Disabled homolog-2. *Oncogene* **31**, 764–75.
- ENDO, H., MURAMATSU, T., FURUTA, M., UZAWA, N., PIMKHAOKHAM, A., AMAGASA, T., INAZAWA, J. & KOZAKI, K. (2013). Potential of tumor-suppressive miR-596 targeting LGALS3BP as a therapeutic agent in oral cancer. *Carcinogenesis* **34**, 560–9.
- GEE, H. E., BUFFA, F. M., CAMPS, C., RAMACHANDRAN, A., LEEK, R., TAYLOR, M., PATIL, M., SHELDON, H., BETTS, G., HOMER, J. et al. (2011). The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br. J. Cancer* **104**, 1168–77.
- GOPALAN, V., PILLAI, S., EBRAHIMI, F., SALAJEGHEH, A., LAM, T. C., LE, T. K., LANGSFORD, N., HO, Y. H., SMITH, R. A. & LAM, A. K. (2014). Regulation of microRNA-1288 in colorectal cancer: Altered expression and its clinicopathological significance. *Molec. Carcinogen.* **53**, Suppl. 1 E36–44.
- GRUNDBERG, E., SMALL, K. S., HEDMAN, A. K., NICA, A. C., BUIL, A., KEILDSON, S., BELL, J. T., YANG, T.-P., MEDURI, E., BARRETT, A. et al. (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genet.* **44**, 1084–9.
- HONEYWELL, D., CABRITA, M., ZHAO, H., DIMITROULAKOS, J. & ADDISON, C. (2013). miR-105 inhibits prostate tumour growth by suppressing CDK6 levels. *PLoS ONE* **8**, e70515.
- HUANG, S. & CHEN, L. (2014). miR-888 regulates side population properties and cancer metastasis in breast cancer cells. *Biochem. Biophys. Res. Commun.* **450**, 1234–40.
- IMAM, J., BUDDAVARAPU, K., LEE-CHANG, J., GANAPATHY, S., CAMOSY, C., CHEN, Y. & RAO, M. (2010). MicroRNA-185 suppresses tumor growth and progression by targeting the Six1 oncogene in human cancers. *Oncogene* **29**, 4971–9.
- ITO, T., SATO, F., KAN, T., CHENG, Y., DAVID, S., AGARWAL, R., PAUN, B. C., JIN, Z., OLARU, A. V., HAMILTON, J. P. et al. (2011). Polo-like kinase 1 regulates cell proliferation and is targeted by miR-593\* in esophageal cancer. *Int. J. Cancer* **129**, 2134–46.
- JAVANMARD, A. & MONTANARI, A. (2013). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *IEEE Trans. Info. Theory* **60**, 6522–54.
- JAVANMARD, A. & MONTANARI, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Mach. Learn. Res.* **15**, 2869–909.
- JEONG, J., KANG, H., KIM, T., KIM, G., HEO, J., KWON, A., KIM, S., JUNG, S. & AN, H. (2017). MicroRNA-136 inhibits cancer stem cell activity and enhances the anti-tumor effect of paclitaxel against chemoresistant ovarian cancer cells by targeting Notch3. *Cancer Lett.* **386**, 168–78.
- KARAAYVAZ, M., ZHAI, H. & JU, J. (2013). miR-129 promotes apoptosis and enhances chemosensitivity to 5-fluorouracil in colorectal cancer. *Cell Death Dis.* **4**, e659.

- LAN, G., L., Y., XIE, X., PENG, L. & WANG, Y. (2015). MicroRNA-490-5p is a novel tumor suppressor targeting c-FOS in human bladder cancer. *Arch. Med. Sci.* **11**, 561–9.
- LEE, E. J., MIRCEAN, C., SHMULEVICH, I., WANG, H., LIU, J., NIEMISTÖ, A., KAVANAGH, J. J., LEE, J. H. & ZHANG, W. (2005). Insulin-like growth factor binding protein 2 promotes ovarian cancer cell invasion. *Molec. Cancer* **4**, 7.
- LEE, K., KUNKEAW, N., JEON, S. H., LEE, I., JOHNSON, B. H., KANG, G. Y., BANG, J. Y., PARK, H. S., LEELAYUWAT, C. & LEE, Y. S. (2011). Precursor miR-886, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity. *RNA* **17**, 1076–89.
- LI, D., ZHAO, Y., LIU, C., CHEN, X., QI, Y., JIANG, Y., ZOU, C., ZHANG, X., LIU, S., WANG, X. et al. (2011). Analysis of MiR-195 and MiR-497 expression, regulation and role in breast cancer. *Clin. Cancer Res.* **17**, 1722–30.
- LI, J., LU, Y., AKBANI, R., JU, Z., ROEBUCK, P. L., LIU, W., YANG, J.-Y., BROOM, B. M., VERHAAK, R. G. W., KANE, D. W. et al. (2013). TPCA: A resource for cancer functional proteomics data. *Nature Meth.* **10**, 1046–7.
- LIN, Y., LIU, A. Y., FAN, C., ZHENG, H., LI, Y., ZHANG, C., WU, S., YU, D., HUANG, Z., LIU, F. et al. (2015). MicroRNA-33b inhibits breast cancer metastasis by targeting HMGA2, SALL4 and Twist1. *Sci. Rep.* **5**, 9995.
- LIU, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.* **41**, 2948–78.
- LOUNICI, K., PONTIL, M., VAN DE GEER, S. & TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39**, 2164–204.
- MURPHY, M. E. (2013). The HSP70 family and cancer. *Carcinogenesis* **34**, 1181–8.
- PILS, D., PINTER, A., REIBENWEIN, J., ALFANZ, A., HORAK, P., SCHMID, B. C., HEFLER, L., HORVAT, R., REINTHALLER, A., ZEILLINGER, R. et al. (2007). In ovarian cancer the prognostic influence of HER2/neu is not dependent on the CXCR4/SDF-1 signalling pathway. *Br. J. Cancer* **96**, 485–91.
- RUFFIEUX, H., DAVIDSON, A. C., HAGER, J. & IRINCHEEVA, I. (2017). Efficient inference for genetic association studies with multiple outcomes. *Biostatistics* **18**, 618–36.
- SANDHU, R., REIN, J., D'ARCY, M., HERSCHKOWITZ, J. I., HOADLEY, K. A. & TROESTER, M. A. (2014). Overexpression of miR-146a in basal-like breast cancer cells confers enhanced tumorigenic potential in association with altered p53 status. *Carcinogenesis* **35**, 2567–75.
- SCHIFANO, E. D., LI, L., CHRISTIANI, D. C. & LIN, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *Am. J. Hum. Genet.* **92**, 744–59.
- SHI, S., LU, Y., QIN, Y., LI, W., CHENG, H., XU, Y., XU, J., LONG, J., LIU, L., LIU, C. et al. (2014). miR-1247 is correlated with prognosis of pancreatic cancer and inhibits cell proliferation by targeting neuropilins. *Curr. Molec. Med.* **14**, 316–27.
- SHI, Y., QIU, M., WU, Y. & HAI, L. (2015). miR-548-3p functions as an anti-oncogenic regulator in breast cancer. *Biomed. Pharmacother.* **75**, 111–6.
- SUO, C., TOULOPOULOU, T., BRAMON, E., WALSHE, M., PICCHIONI, M., MURRAY, R. & OTT, J. (2013). Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC Bioinformatics* **14**, 151.
- VAN DE GEER, S., BÜHLMANN, P., RITOV, Y. & DEZEURE, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.* **42**, 1166–202.
- VAN KOUWENHOVE, M., KEDDE, M. & AGAMI, R. (2011). MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nature Rev. Cancer* **11**, 644–56.
- XIA, Y., CAI, T. & CAI, T. T. (2015). Testing differential networks with applications to the detection of gene-gene interactions. *Biometrika* **102**, 247–66.
- XIA, Y., CAI, T. & CAI, T. T. (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Statist. Sinica* **28**, 63–92.
- YANG, J. J., LI, J., WILLIAMS, L. K. & BUU, A. (2016). An efficient genome-wide association test for multivariate phenotypes based on the Fisher combination function. *BMC Bioinformatics* **17**, 19.
- YUAN, M. & LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Statist. Soc. B* **68**, 49–67.
- ZAITSSEV, A. Y. (1987). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Prob. Theory Rel. Fields* **74**, 535–66.
- ZHANG, C.-H. & ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Statist. Soc. B* **76**, 217–42.
- ZHOU, J. J., CHO, M. H., LANGE, C., LUTZ, S., SILVERMAN, E. K. & LAIRD, N. M. (2015). Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability. *Hum. Hered.* **79**, 93–104.
- ZHOU, W., FONG, M. Y., MIN, Y., SOMLO, G., LIU, L., PALOMARES, M. R., YU, Y., CHOW, A., O'CONNOR, S. T. F., CHIN, A. R. et al. (2014). Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell* **25**, 501–15.
- ZHU, Y. & BRADIC, J. (2017). Significance testing in non-sparse high-dimensional linear models. *arXiv:1610.02122v3*.
- ZSIROS, E., DUTTAGUPTA, P., DANGAJ, D., LI, H., FRANK, R., GARRABRANT, T., HAGEMANN, I. S., LEVINE, B. L., JUNE, C. H., ZHANG, L. et al. (2015). The ovarian cancer chemokine landscape is conducive to homing of vaccine-primed and CD3/CD28–costimulated T cells prepared for adoptive therapy. *Clin. Cancer Res.* **21**, 2840–50.

[Received on 2 May 2017. Editorial decision on 14 November 2017]