



OPEN ACCESS

EDITED BY

Paul Horn,
Cincinnati Children's Hospital Medical
Center, United States

REVIEWED BY

Hyunwook Koh,
SUNY Korea, Republic of Korea
Kazuharu Harada,
Tokyo Medical University, Japan

*CORRESPONDENCE

Nikolay I. Nikolov
✉ n.nikolov@math.bas.bg

RECEIVED 17 November 2025

REVISED 28 January 2026

ACCEPTED 28 January 2026

PUBLISHED 23 February 2026

CITATION

Nikolov NI, Savov M and Palejev D (2026)

Multiple testing procedures under
positive dependency with block
structure.

Front. Appl. Math. Stat. 12:1748504.

doi: 10.3389/fams.2026.1748504

COPYRIGHT

© 2026 Nikolov, Savov and Palejev. This

is an open-access article distributed

under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).

The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Multiple testing procedures under positive dependency with block structure

Nikolay I. Nikolov^{1,2*}, Mladen Savov^{1,2} and Dean Palejev^{1,2}

¹Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria,

²Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

The classical Benjamini–Hochberg (B-H) method, widely used across various disciplines such as genetics, epidemiology, and social sciences, serves as an established procedure for controlling the false discovery rate (FDR) in multiple comparison scenarios. The B-H method assumes independence among tests, which often does not hold in large-scale dependent datasets. The Benjamini–Yekutieli (B-Y) adjustment controls the FDR under arbitrary dependence but is often very conservative and can lead to a reduction in statistical power. This paper investigates the performance of the B-H and B-Y procedures under specific positive block dependence structures. Two parametric forms of block dependence are considered to model the correlation among paired t -test statistics. Estimation algorithms induced by different matrix norms are developed for approximating the value of the unknown parameter. Modifications of existing multiple testing approaches are proposed by incorporating test dependence and enhancing their power through integration of Kolmogorov–Smirnov tests. Simulation studies are performed to demonstrate that the recommended methods preserve FDR control while improving power compared to traditional techniques.

KEYWORDS

Benjamini–Hochberg procedure, Benjamini–Yekutieli adjustment, false discovery rate, Kolmogorov–Smirnov test, multiple comparisons, p -values, statistical power

1 Introduction

One of the most common methods for controlling the false discovery rate (FDR) in multiple hypothesis testing is the Benjamini–Hochberg (B-H) algorithm introduced in [1], which is among the most cited scientific papers [2]. The B-H procedure finds application in various fields of natural and social sciences, including but not limited to neuroimaging [3–5], epidemiology [6, 7], gene ontology [8, 9], microbiology [10, 11], psychology [12, 13], geography [14, 15] and finance [16]. In genetics, it is the default FDR correction method for finding differentially expressed genes and integrated into the most widely used statistical software packages, such as limma [17], which is developed for both microarrays and RNA-seq data, as well as DESeq2 [18] and edgeR [19] for RNA-seq data only. However, gene studies and other areas with FDR applications typically involve large-scale dependent data that does not meet the assumptions for using the B-H method.

The effect of dependence on FDR analysis was studied by Benjamini and Yekutieli [20], who proposed an adjustment of the B-H method, denoted in this work by B-Y, that controls the FDR under arbitrary dependence among the p -values. However, the B-Y procedure is often too conservative in practice. Schwartzman and Lin [21] showed that strong correlations between the p -values result in high variability and low power of the B-H

and B-Y procedures. In particular, positive (negative) correlations can cause the empirical null distributions of z -values to become narrower (wider), which significantly affect the subsequent FDR adjustments. This shows the importance of creating new FDR methods designed to incorporate the structural dependence among tests. As highlighted by Chi et al. [22], one of the main challenges in designing new procedures under dependence is that there are many definitions of what it means to be dependent. One approach is to derive algorithms based on a particular parametrization of the dependence, e.g., equicorrelated test statistics proposed by Hartung [23]. Nevertheless, the sheer number of options for describing different types of dependences can be daunting, potentially accounting for the limited progress in this topic. As a more recent alternative, a generalization of the concept of p -values that is based on e -values, introduced by Vock and Wang [24], is suggested by Wang and Ramdas [25], as well as Zhao and Sun [26], for developing new FDR techniques. In this work we consider specific dependence structures with limited number of parameters as an initial investigation for the problem of controlling the FDR under dependent p -values.

In this paper we study the performance of the B-H and B-Y methods under positive block dependence and propose modifications of the FDR techniques. We consider two particular parametric forms of block dependence to model the relationship among paired t -test statistics. The first one is associated with a correlation matrix that depends exponentially on a single parameter, while the other is based on a linearly parameterized matrix that is positive definite for most parameter values, i.e., in most cases it is also a correlation matrix. To estimate the unknown parameter associated with the dependence structure, we develop algorithms based on different matrix norms. Additionally, we suggest to combine the presented procedures with Kolmogorov-Smirnov (K-S) tests, which aims to improve the power of the multiple testing algorithms.

The rest of the paper is organized as follows. In Section 2, we describe t -test procedures for paired samples under a general block dependence structure. We derive the asymptotic joint distribution of the t -test statistics in Section 2.2 and propose a transformation based on the associated dependence matrix. In Section 3, we consider multiple testing methods for controlling the FDR and propose modifications of the existing procedures by combining them with K-S tests. In Section 4, we make performance comparisons between the presented algorithms and show that our suggestions lead to an improvement in the average power and the FDR control. Finally, concluding remarks and a discussion for future work are provided in Section 5.

2 Dependencies between paired samples

We consider paired samples of n features over m individuals or pairs of individuals and denote by $(X_i^{(j)}, Y_i^{(j)})$ the paired observation of the j -th feature for the i -th individual or pair of individuals, where $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. For example, the vectors $\vec{X}_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(n)})$ and $\vec{Y}_i =$

$(Y_i^{(1)}, Y_i^{(2)}, \dots, Y_i^{(n)})$ can be obtained as measurements for the i -th individual before and after treatment, respectively, or as observations for the i -th individuals in a control and treatment group in a matched samples study. In the latter, $\mathbf{X} = \{X_i^{(j)}\}_{i=1, j=1}^{m, n}$ and $\mathbf{Y} = \{Y_i^{(j)}\}_{i=1, j=1}^{m, n}$ are the observation matrices in both groups, whereas $\mathbf{D} = \{X_i^{(j)} - Y_i^{(j)}\}_{i=1, j=1}^{m, n}$ is the matrix of differences. The problem of analyzing such samples arises, for example, in genetics, where each of n gene expression levels (each of n features) is tested for significance. In the next subsection we describe the sample distribution and dependence assumptions in this paper.

2.1 Dependence structure

Denote the paired differences of the sample vectors $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_m, \vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_m$, described above, by $\vec{D}_i = \vec{X}_i - \vec{Y}_i$, for $i = 1, 2, \dots, m$. Let us assume that the vectors $\vec{D}_1, \vec{D}_2, \dots, \vec{D}_m$ are independent and identically distributed (iid) with multivariate normal distribution, i.e., $\vec{D}_i \sim N(\vec{\mu}, \Sigma)$, for $i = 1, 2, \dots, m$, where $\vec{\mu}$ is the n -dimensional vector $(\mu_1, \mu_2, \dots, \mu_n)$ corresponding to the mean difference for each of the n features, whereas Σ is a n by n correlation matrix associated with the dependence between the n tested features.

Kim et al. [27] use block matrices to model the correlation matrix in its multivariate Gaussian model, which represents gene concentration levels. This structure assumes that groups of genes within the same regulatory pathway are correlated, while genes in different pathways are uncorrelated. Each block within the matrix corresponds to a distinct gene regulatory pathway, with specific variance and correlation coefficients defining the relationships among genes within that pathway. Motivated by the correlation form in [27], let us assume that Σ is a block matrix which can be expressed as

$$\Sigma = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{A}_k \end{pmatrix},$$

i.e., Σ has k blocks $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$. Note that in order to obtain such structure of the dependence matrix Σ , the tested features should be ordered (labeled) in a correct manner based on some prior information, e.g., in genetics one can apply knowledge from previous studies on the gene regulatory pathways. Dang et al. [28] use an interactive adjacency matrix to visualize binary relationships between proteins in biological pathways. To simplify complex networks, proteins with similar interaction patterns are grouped together, and these groups can then be “collapsed” into single nodes, effectively forming block matrices that represent relationships between groups rather than individuals. In case the number k and sizes of the blocks in Σ are unknown, Perreault et al. [29] propose an algorithm based on Kendall’s rank correlation for identifying different disjoint groups in Σ when the features in each block are equicorrelated. This block structure with exchangeable variables, i.e., with constant correlation for any two features within a block, is also considered by Hartung [23] in the context of combining dependent statistics. In our study, we propose the following two

generalizations of such dependence matrices, which allow the correlation coefficient in a block to decrease in an exponential or a linear manner.

For simplicity and clarity of the results in our simulation analysis, in Section 4 we assume that Σ has k blocks with the same size and the same structure. Thus, we consider the case $\mathbf{A}_1 = \mathbf{A}_2 = \dots = \mathbf{A}_k = \mathbf{A}$ with \mathbf{A} being an s by s correlation matrix, where $s = n/k$. We further assume that the common block matrix \mathbf{A} can be parameterized by a single parameter $\theta \in (0, 1)$. We suggest the following two forms for $\mathbf{A}(\theta)$:

$$\mathbf{A}^*(\theta) = \begin{pmatrix} 1 & \theta & \dots & \theta^{s-1} \\ \theta & 1 & \dots & \theta^{s-2} \\ \dots & \dots & \dots & \dots \\ \theta^{s-1} & \theta^{s-2} & \dots & 1 \end{pmatrix} = \left\{ \theta^{|i-j|} \right\}_{i,j=1}^s \quad \text{and}$$

$$\mathbf{A}'(\theta) = \begin{pmatrix} 1 & \theta & \frac{(q(s-2)-(q-1))\theta}{q(s-2)} & \dots & \frac{\theta}{q} \\ \theta & 1 & \theta & \dots & \frac{(s+q-3)\theta}{q(s-2)} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{\theta}{q} & \frac{(s+q-3)\theta}{q(s-2)} & \frac{(s+2q-4)\theta}{q(s-2)} & \dots & 1 \end{pmatrix}$$

$$= \left\{ 1 \{i \neq j\} \left[1 - \frac{(|i-j|-1)(q-1)}{q(s-2)} \right] \theta + 1 \{i = j\} \right\}_{i,j=1}^s, \quad (1)$$

where $1 \{ \cdot \}$ is the indicator function and q is a real constant $q \geq 1$. It is important to highlight that $\mathbf{A}^*(\theta)$ is positive definite for all values of $\theta \in (0, 1)$, whereas $\mathbf{A}'(\theta)$ may not be positive definite for some combinations of the values of $q \geq 1$, the parameter $\theta \in (0, 1)$ and the matrix size s . Hence, $\mathbf{A}^*(\theta)$, and consequently $\Sigma(\theta)$ based on $\mathbf{A}^*(\theta)$, is a correlation matrix for all $\theta \in (0, 1)$, while $\mathbf{A}'(\theta)$ may not be a correlation matrix for some $\theta \in (0, 1)$. However, $\mathbf{A}'(\theta)$ is the equicorrelated matrix in the limit case $q = 1$, i.e., $\mathbf{A}'(\theta)$ is positive definite for all $\theta \in (0, 1)$ when $q = 1$. For both structures $\mathbf{A}^*(\theta)$ and $\mathbf{A}'(\theta)$, the parameter θ is related to the dependence strength between the features in the block, whereas $\theta = 0$ means that the features are independent. The case $\theta = 1$ for $\mathbf{A}^*(\theta)$ corresponds to a deterministic linear relation between the features that belong to the same block.

The form $\mathbf{A}^*(\theta)$ is a Toeplitz type matrix, which is used often in stationary time series, [30, p. 311, Section 8.8.4], while $\mathbf{A}'(\theta)$ is a linear generalization of the equicorrelated matrix, considered in [23] and [29]. Under the dependence structure $\mathbf{A}^*(\theta)$, the correlation coefficients between the features decay exponentially, from θ for adjacent features to θ^{s-1} for the furthest features, while the correlations in a block modeled by $\mathbf{A}'(\theta)$ are decreasing in a linear manner from θ to θ/q . The equicorrelated dependence can be included in the form $\mathbf{A}'(\theta)$ as the limit case $q = 1$. For simplicity, in this paper we consider only the case $q = 2$, i.e., correlation changing from θ to $\theta/2$. It should be noted that when $q = 2$ the matrix $\mathbf{A}'(\theta)$ is not always positive definite for arbitrary $\theta \in (0, 1)$ and block size s , however in our simulation study the values of θ and s are chosen such that $\mathbf{A}'(\theta)$ is a correlation matrix, unless otherwise noted. In this work we further assume that all tests for the features within a given block are associated with either true null hypotheses or true alternatives. This corresponds to assuming that all genes in a pathway have similar behavior and consequently their expression levels are either significant or not. A general algorithm for estimating the value of θ by the observed data is presented in

Section 3, while simulation results under the two forms $\mathbf{A}^*(\theta)$ and $\mathbf{A}'(\theta)$ are given in Section 4. Next, we describe the t -test procedures for comparing the two samples \mathbf{X} and \mathbf{Y} .

2.2 Test statistics and asymptotic distribution

For the j -th feature, $j = 1, 2, \dots, n$, let us consider the Student's t -test for paired samples for testing $H_0 : \mu_j = 0$ against $H_1 : \mu_j \neq 0$. Let us denote the average of each column in the matrix of differences $\mathbf{D} = \{X_i^{(j)} - Y_i^{(j)}\}_{i=1, j=1}^{m, n}$ by

$$\bar{D}_j = \frac{1}{m} \sum_{i=1}^m (X_i^{(j)} - Y_i^{(j)}), \quad \text{for } j = 1, 2, \dots, n.$$

In the context of gene expression levels, this could be the differences between the logarithms of the expression levels, or equivalently, $\log(\text{fold change})$. Then, the Student's t -statistic for each feature is

$$T_m^{(j)} = \frac{\bar{D}_j \sqrt{m}}{\hat{\sigma}_m^{(j)}} \sim t(m-1), \quad \text{for } j = 1, 2, \dots, n,$$

where $\hat{\sigma}_m^{(j)} = \sqrt{\frac{\sum_{i=1}^m (D_i^{(j)} - \bar{D}_j)^2}{m-1}}$ is the sample estimation of the standard deviation of the j -th difference. The joint distribution of the test statistics $T_m^{(1)}, T_m^{(2)}, \dots, T_m^{(n)}$ has the following asymptotic property, given in the next proposition.

Proposition 1. Under H_0 and the assumptions in Subsection 2.1 with an arbitrary correlation (dependence) matrix Σ , the joint distribution of $T_m^{(1)}, T_m^{(2)}, \dots, T_m^{(n)}$ is asymptotically normal such that

$$\begin{pmatrix} T_m^{(1)} \\ T_m^{(2)} \\ \dots \\ T_m^{(n)} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} T^{(1)} \\ T^{(2)} \\ \dots \\ T^{(n)} \end{pmatrix} \sim N(\vec{0}, \Sigma), \quad \text{as } m \rightarrow \infty. \quad (2)$$

Proof. Let $\mathbf{B} = (b_{i,j})_{i,j=1}^m$ be a unitary matrix such that $b_{1,i} = m^{-1/2}$, for $i = 1, 2, \dots, m$. Then,

$$\mathbf{Z} = \mathbf{B}\mathbf{D} = \begin{pmatrix} \bar{D}_1 \sqrt{m} & \bar{D}_2 \sqrt{m} & \dots & \bar{D}_n \sqrt{m} \\ Z_2^{(1)} & Z_2^{(2)} & \dots & Z_2^{(n)} \\ \dots & \dots & \dots & \dots \\ Z_m^{(1)} & Z_m^{(2)} & \dots & Z_m^{(n)} \end{pmatrix},$$

such that

$$\begin{pmatrix} \bar{D}_1 \sqrt{m} \\ \bar{D}_2 \sqrt{m} \\ \dots \\ \bar{D}_n \sqrt{m} \end{pmatrix} \sim N(\vec{0}, \Sigma) \quad (3)$$

and

$$\begin{pmatrix} \bar{D}_j \sqrt{m} \\ Z_2^{(j)} \\ \dots \\ Z_m^{(j)} \end{pmatrix} \sim N(\vec{0}, \mathbf{I}_{m \times m}), \quad \text{for } j = 1, 2, \dots, n,$$

with $\mathbf{I}_{m \times m}$ being the identity matrix of size m .

Moreover, $\bar{D}_j \sqrt{m}$ is independent of $V_m^{(j)} = (m-1) \sum_{i=2}^m (Z_i^{(j)})^2$ with

$$\lim_{m \rightarrow \infty} \frac{V_m^{(j)}}{m-1} = 1, \quad \text{for } j = 1, 2, \dots, n.$$

Thus, the t -test statistics are

$$T_m^{(j)} = \frac{\bar{D}_j \sqrt{m}}{\sqrt{\frac{V_m^{(j)}}{m-1}}} \sim t(m-1), \quad \text{for } j = 1, 2, \dots, n.$$

Finally, by (3) and Slutsky's theorem it follows that

$$\lim_{m \rightarrow \infty} \begin{pmatrix} T_m^{(1)} \\ T_m^{(2)} \\ \dots \\ T_m^{(n)} \end{pmatrix} = \begin{pmatrix} T^{(1)} \\ T^{(2)} \\ \dots \\ T^{(n)} \end{pmatrix} \sim N(\vec{0}, \Sigma).$$

3 Multiple testing procedures

Let p_j be the p -value obtained from the t -test statistic $T_m^{(j)}$ for testing $H_0: \mu_j = 0$ against $H_1: \mu_j \neq 0$, where $j = 1, 2, \dots, n$. Denote by $p_{1:n} \leq p_{2:n} \leq \dots \leq p_{n:n}$ the associated ordered p -values. For a fixed FDR α , such that $0 < \alpha < 1$, the Benjamini–Hochberg (B-H) method rejects all hypotheses with p -values smaller or equal than $p_{R_n:n}$, where

$$R_n = \max \left\{ 1 \leq i \leq n : p_{i:n} \leq \frac{i}{n} \alpha \right\}$$

is the number of rejected hypotheses. The Benjamini–Yekutieli (B-Y) procedure rejects the null hypotheses for the tests corresponding to $p_{1:n}, p_{2:n}, \dots, p_{R_n^*:n}$, with

$$R_n^* = \max \left\{ 1 \leq i \leq n : p_{i:n} \leq \frac{i}{nc(n)} \alpha \right\},$$

and $c(n) = \sum_{i=1}^n i^{-1}$. For large values of n , the factor $c(n)$ can be approximated by $\log(n) + (2n)^{-1} + \gamma$, where γ is the Euler's constant ($\gamma \approx 0.5772$).

Under independence of the p -values, the FDR of the B-H procedure is $N\alpha/n$, where N is the number of true null hypotheses among all n tests, see [1]. For any arbitrary dependence, the FDR of the B-Y algorithm is less than α [20]. For the power of the respective method we have

$$\text{power}_n := \frac{R_n - V_n}{\max(1, n - N)} = \frac{R_n}{n} \left(\frac{1 - \alpha}{1 - \frac{N}{n}} + \alpha \right) + o_p(1),$$

where V_n is the number of false rejected nulls, cf. [31]. As the power and FDR of the B-H and B-Y procedures strongly depend on the underlying dependence, next we propose several adjustments and additions.

3.1 Transformation of the t -statistics

Let us consider the result of Proposition 1. In case the dependence correlation matrix Σ is known, then by (2) and the scaling property of the multivariate normal distribution:

$$(\Sigma)^{-1/2} \begin{pmatrix} T_m^{(1)} \\ T_m^{(2)} \\ \dots \\ T_m^{(n)} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Z^{(1)} \\ Z^{(2)} \\ \dots \\ Z^{(n)} \end{pmatrix} \sim N(\vec{0}, \mathbf{I}_{n \times n}), \quad \text{as } m \rightarrow \infty,$$

where $Z^{(1)}, Z^{(2)}, \dots, Z^{(n)}$ are iid standard normal variables. Thus, instead of applying the B-H method to the p -values corresponding to the t -test statistics $T_m^{(1)}, T_m^{(2)}, \dots, T_m^{(n)}$, we can apply the B-H procedure to the p -values obtained by the decorrelated statistics

$$\begin{pmatrix} S_m^{(1)} \\ S_m^{(2)} \\ \dots \\ S_m^{(n)} \end{pmatrix} = (\Sigma)^{-1/2} \begin{pmatrix} T_m^{(1)} \\ T_m^{(2)} \\ \dots \\ T_m^{(n)} \end{pmatrix}, \tag{4}$$

with $S_m^{(j)} \xrightarrow{d} N(0, 1)$ as $m \rightarrow \infty$, for $j = 1, 2, \dots, n$.

Since we assume that the correlation matrix Σ has a block structure, the statistic $S_m^{(j)}$ in (4) is a linear combination only of the statistics $T_m^{(i)}$ that belong to the same block as $S_m^{(j)}$, for $j = 1, 2, \dots, n$. Thus, the order of the blocks in transformation (4) is the same as in the raw t -test statistics. However, the statistic $S_m^{(j)}$ no longer corresponds to the j -th feature (gene) expression, but rather to a combination of features in the same block. Hence, if the B-H procedure detects $S_m^{(j)}$ as significant, this by itself does not necessarily mean that the j -th feature is significant. Because of our assumption that either the null hypothesis H_0 is true for all features in a specific block or H_1 holds for all features within a block, the FDR and power of a multiple testing procedure depend only on the number of rejected hypotheses within a given block instead of on the specific rejected tests. In Section 3.2, we propose a post-hoc procedure for identifying potentially missed significant features by the transformed B-H method.

In practice, the dependence matrix is unknown and needs to be estimated. Under our assumptions in Section 2.1, Σ can be parameterized by θ . Therefore, an estimation of Σ corresponds to an estimation $\hat{\theta}$ of the parameter θ . For finding $\hat{\theta}$, we propose a procedure based on minimizing a matrix norm of the difference between $\Sigma(\theta)$ and \hat{C} for all values $\theta \in (0, 1)$, where \hat{C} is the sample estimation of the correlation matrix Σ . For the matrix norm, based on our simulation results in Section 4.1, we suggest the Frobenius norm $\|\cdot\|_F$ or the Max norm $\|\cdot\|_M$, i.e., the value of θ to be estimated by

$$\hat{\theta} = \arg \min_{\theta \in (0,1)} \|\Sigma(\theta) - \hat{C}\|_F = \arg \min_{\theta \in (0,1)} \sqrt{\sum_{i=1}^n \sum_{j=1}^n [(\Sigma(\theta) - \hat{C})_{ij}]^2}, \tag{5}$$

or by

$$\hat{\theta} = \arg \min_{\theta \in (0,1)} \|\Sigma(\theta) - \hat{C}\|_M = \arg \min_{\theta \in (0,1)} \left\{ \max_{i,j \in \{1,2,\dots,n\}} |\Sigma(\theta) - \hat{C}|_{ij} \right\}.$$

Alternative matrix norms that are additionally considered in Section 4.1 are the column-sum norm $\|\cdot\|_1$ and the spectral norm $\|\cdot\|_2$, defined for $\Sigma(\theta) - \widehat{C}$ as

$$\begin{aligned} \|\Sigma(\theta) - \widehat{C}\|_1 &= \max_{j \in \{1, 2, \dots, n\}} \sum_{i=1}^n |\Sigma(\theta) - \widehat{C}|_{i,j} \quad \text{and} \\ \|\Sigma(\theta) - \widehat{C}\|_2 &= \max_{i \in \{1, 2, \dots, n\}} |s_i(\Sigma(\theta) - \widehat{C})|, \end{aligned}$$

where $s_i(\mathbf{M})$ is the i -th singular value of a matrix \mathbf{M} . Note that $\|\Sigma(\theta) - \widehat{C}\|_1$ is equivalent to the row-sum norm (infinity norm),

$$\|\Sigma(\theta) - \widehat{C}\|_\infty = \max_{i \in \{1, 2, \dots, n\}} \sum_{j=1}^n |\Sigma(\theta) - \widehat{C}|_{i,j},$$

since $\Sigma(\theta) - \widehat{C}$ is a symmetric matrix. Thus, in Section 4.1 we present the estimation results based only on $\|\cdot\|_1$. More details about the listed matrix norms can be found in [30, Section 3.9].

3.2 Combination with the Kolmogorov-Smirnov tests

Our simulation analysis in Section 4.2 shows that the FDR methods based on the p -values obtained from the transformed statistics $S_m^{(1)}, S_m^{(2)}, \dots, S_m^{(n)}$ in (4), with $\Sigma = \Sigma(\widehat{\theta})$, lead to a better control of the FDR in comparison to the FDR under the correlated t -statistics $T_m^{(1)}, T_m^{(2)}, \dots, T_m^{(n)}$. However, this performance improvement comes at the cost of reducing the power of the applied procedures. Therefore, next we propose an adjustment by combining those methods with the K-S tests for each block in Σ .

Let $p_1^{(l)}, p_2^{(l)}, \dots, p_s^{(l)}$ be the raw p -values corresponding to the test statistics of the features in the l -th block of Σ , for $l = 1, 2, \dots, k$. Since their distribution is standard uniform under H_0 and the independence assumption, to test their uniformity we consider the one-sided K-S distances

$$\begin{aligned} D_l^+ &= \sup_{x \in [0, 1]} (\widehat{F}_{p^{(l)}}(x) - x) = \max \left\{ \max_{i \in \{1, 2, \dots, s\}} \left(\frac{i}{s} - p_i^{(l)} \right), 0 \right\}, \\ l &= 1, 2, \dots, k, \end{aligned} \tag{6}$$

where $\widehat{F}_{p^{(l)}}(x)$ is the empirical distribution function of $p_1^{(l)}, p_2^{(l)}, \dots, p_s^{(l)}$, cf. [32, formula (4.3.3)]. Note that in (6) we consider only the one-sided differences $\widehat{F}_{p^{(l)}}(x) - x$, since the test statistics obtained under H_1 are associated with smaller p -values resulting in a peak of $\widehat{F}_{p^{(l)}}(x)$ around the origin $x = 0$, i.e., $\widehat{F}_{p^{(l)}}(x) > x$ for small values of x . Hence, in addition to the rejected hypotheses by the used FDR correction method, we propose to reject all hypotheses corresponding to the blocks with significant K-S statistics D_l^+ , $l = 1, 2, \dots, k$. In such a manner the K-S tests serve as a post-hoc method for identifying potentially missed significant features. However, this might change the FDR and does not guarantee a strict FDR control. Note that the K-S tests can also be used as an *ad-hoc* tool, i.e., applied independently of the multiple comparison procedure. For the simulations in Section 4.2, we use the K-S tests as a post-hoc exploratory tool rather than a procedure with guaranteed FDR control. Our simulation study indicates that it is

best to combine the K-S tests with the transformed B-H procedure based on the transformation (4) with $\Sigma(\widehat{\theta})$ and $\widehat{\theta}$ computed by (5). Thus, in Section 4.2 we present the K-S results only in combination with the transformed B-H algorithm.

4 Simulation results

To compare the estimators $\widehat{\theta}$ of θ and the FDR methods given in Section 3, we perform a simulation study under various dependence configurations. The estimation algorithms and the FDR procedures were implemented using the R programming language [33], and the numerical simulations were performed on the Avitohol supercomputer at ICT-BAS, described in [34]. For each dependence structure and parameter setup, the results presented in this section are based on generating 1,020 samples. For the parameter settings, we combine different values of $n \in \{100, 500, 1000\}$, $m \in \{20, 50, 100, 150, 200\}$, $k \in \{5, 10, 20\}$, $\theta \in \{0.80, 0.90, 0.95\}$ and $\mu_j = \{0.5, 1, 2\}$, whereas for the correlation matrix Σ , we assume the structure of k blocks with equal size $s = n/k$ in the form $\mathbf{A}^*(\theta)$ or $\mathbf{A}'(\theta)$, for $q = 2$, defined in Section 2.1. It should be noted that in all settings in our simulations the matrix $\mathbf{A}'(\theta)$ is positive definite, i.e., it is a correlation matrix, except in the case when $n = 100, k = 20$ and $\theta = 0.95$. For the FDR procedures, we simulate exactly one block from the alternative hypothesis H_1 and generate the other $k - 1$ blocks in Σ under the null hypothesis $H_0: \mu_j = 0$. In all simulations, the significance level α is fixed to 0.05. To evaluate the performance of different estimators of θ , we use the average bias and the mean squared error (MSE) of the obtained estimations, while the FDR procedures are compared by computing the average FDR and power of all simulations in each setup.

4.1 Dependence structure estimation

Let us first analyze the estimators $\widehat{\theta}$ of the parameter θ based on different matrix norms, given in Section 3.1. To obtain various structures of the correlation matrix $\Sigma(\theta)$, we fix the sample size m and the number of features n to be 100, while varying the dependence strength $\theta \in \{0.80, 0.90, 0.95\}$ and the number of blocks $k \in \{5, 10, 20\}$. As a consequence, the size of each block $s = n/k$ also takes values in the set $\{5, 10, 20\}$. Since Σ does not depend on the vector of mean differences, we set μ_j to be 0, for $j = 1, 2, \dots, n$. For solving the optimization problem in (5) under different matrix norms we use the function `optimize()` of the built-in R *stats* package. The simulation results for the average bias (Avg. bias) and MSE of the estimators $\widehat{\theta}$ based on Frobenius $\|\cdot\|_F$, Max $\|\cdot\|_M$, Column-sum $\|\cdot\|_1$ and Spectral $\|\cdot\|_2$ norms are given in Table 1. The values obtained under the block form $\mathbf{A}^*(\theta)$ are presented on the left-hand side, whereas those under $\mathbf{A}'(\theta)$ on the right-hand side. The results under $\mathbf{A}'(\theta)$ for $n = 100, k = 20$ and $\theta = 0.95$ are missing, since in this case the $\mathbf{A}'(\theta)$ is not positive definite, i.e., it is not a correlation matrix.

The average bias results suggest that under the Frobenius norm the estimations of θ have uniformly the smallest absolute bias. The estimators $\widehat{\theta}$ based on $\|\cdot\|_1$ and $\|\cdot\|_2$ tend to overestimate the value

TABLE 1 Average simulation results for the performance of the estimation $\hat{\theta}$ with $n = 100, m = 100, k \in \{5, 10, 20\}$ and $\theta \in \{0.80, 0.90, 0.95\}$ under block dependence structure $A^*(\theta)$, left, and $A'(\theta)$, right.

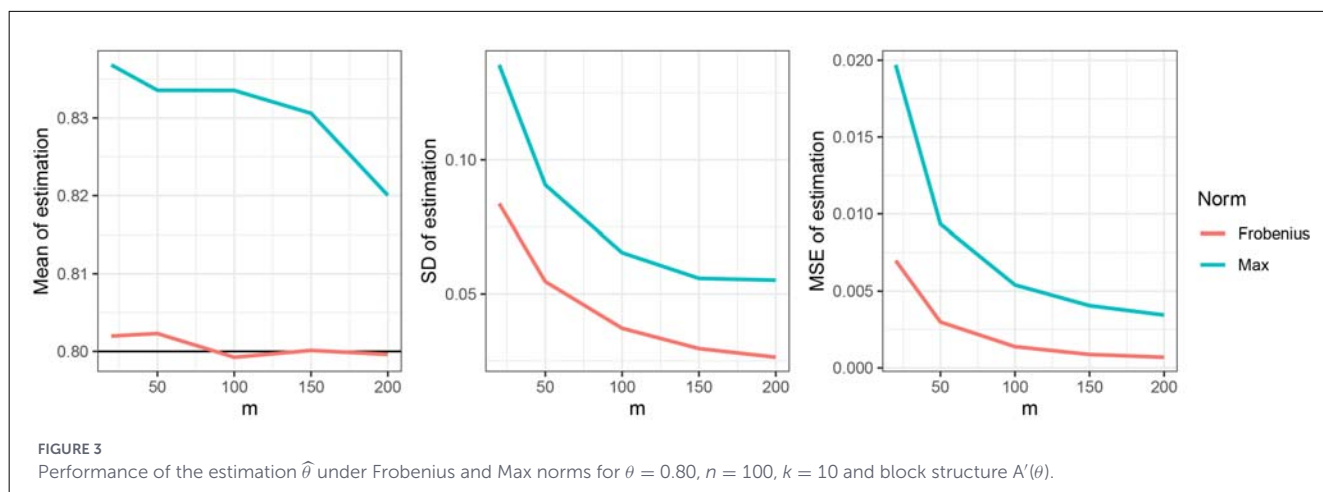
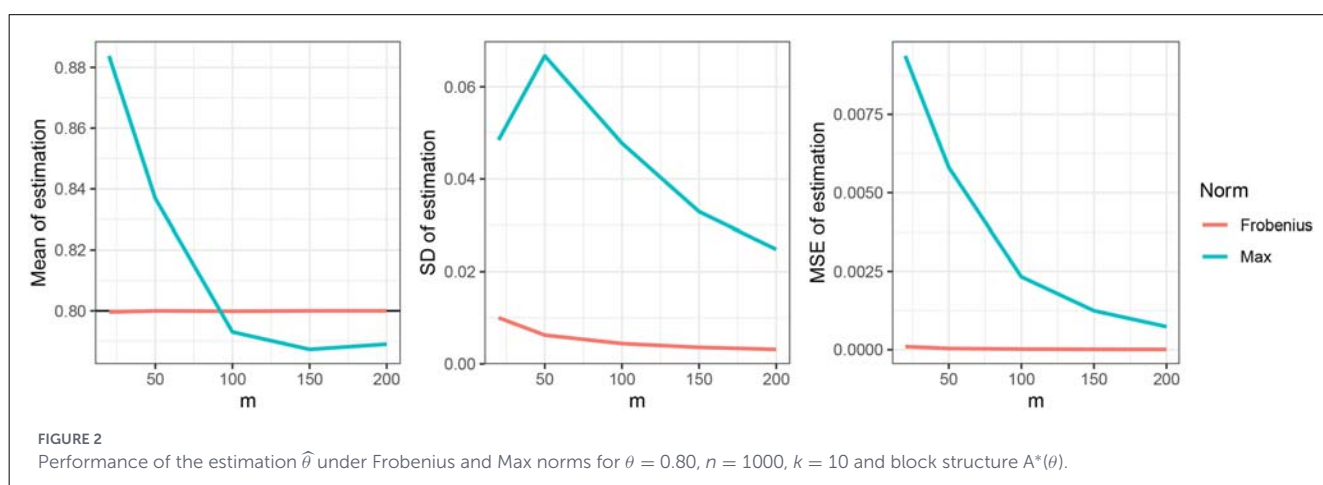
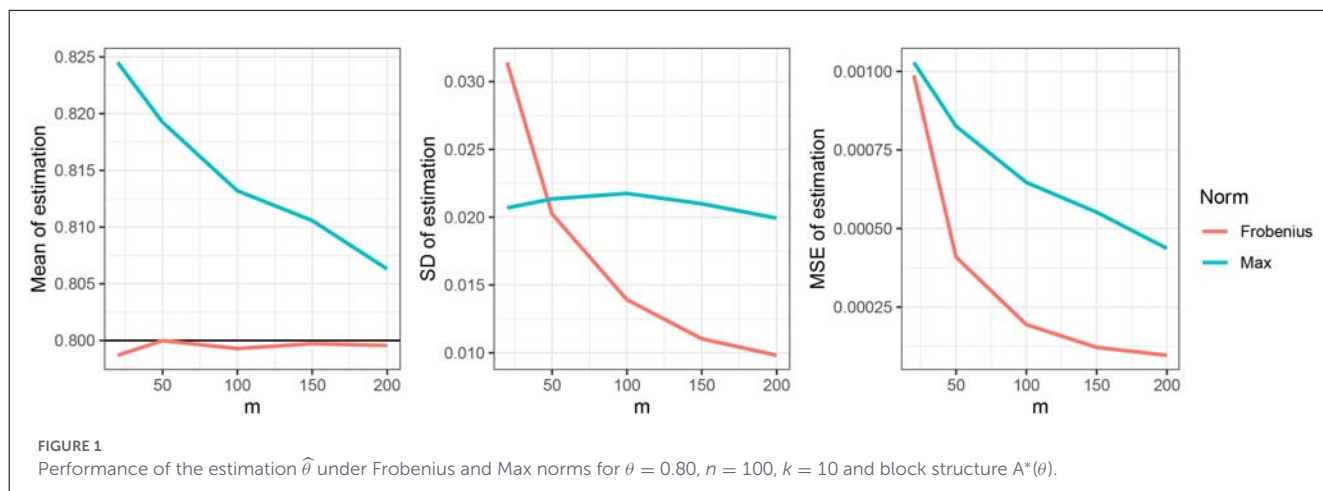
$k = 5$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.0004	0.0002	-0.0003	0.0001	-0.0003	0.0001	0.0010	0.0026	0.0011	0.0033	0.0012	0.0036
Max norm	-0.0001	0.0024	-0.0027	0.0001	-0.0104	0.0003	0.0293	0.0057	-0.0001	0.0016	-0.0018	0.0038
Column-sum norm	0.0196	0.0011	0.0088	0.0004	0.0052	0.0002	0.0434	0.0081	0.0407	0.0091	0.0385	0.0102
Spectral norm	0.0336	0.0015	0.0123	0.0004	0.0061	0.0002	0.0367	0.0056	0.0333	0.0063	0.0317	0.0068
$k = 10$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.0008	0.0002	-0.0006	0.0001	-0.0006	0.0001	-0.0008	0.0014	-0.0008	0.0017	-0.0009	0.0019
Max norm	0.0140	0.0006	-0.0187	0.0008	-0.0123	0.0005	0.0351	0.0054	-0.0008	0.0011	-0.0066	0.0030
Column-sum norm	0.0284	0.0020	0.0211	0.0011	0.0165	0.0008	0.0857	0.0157	0.0837	0.0168	0.0821	0.0175
Spectral norm	0.0559	0.0037	0.0319	0.0015	0.0232	0.0009	0.0979	0.0136	0.0925	0.0135	0.0900	0.0135
$k = 20$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.0002	0.0002	-0.0002	0.0002	-0.0002	0.0002	0.0001	0.0007	0.0001	0.0009	-	-
Max norm	-0.0316	0.0024	-0.0134	0.0005	-0.0044	0.0007	0.0428	0.0053	0.0016	0.0005	-	-
Column-sum norm	0.0619	0.0065	0.0588	0.0059	0.0542	0.0053	0.1441	0.0317	0.1448	0.0331	-	-
Spectral norm	0.1414	0.0212	0.1118	0.0137	0.0994	0.0112	0.2607	0.0733	0.2508	0.0691	-	-

of θ , while those derived by $\|\cdot\|_F$ and $\|\cdot\|_M$ can have a positive or negative bias. However, in all cases the bias under $\|\cdot\|_F$ is less than 0.13% and under $\|\cdot\|_M$ is less than 5.35%, whereas for some simulation configurations the bias under $\|\cdot\|_1$ and $\|\cdot\|_2$ is up to 18.01% and 32.59%, respectively. In general, it can be concluded that the estimations derived by $\|\cdot\|_F$ and $\|\cdot\|_M$ have more stable bias performance compared to the $\|\cdot\|_1$ and $\|\cdot\|_2$ norms that have greater bias for larger number of blocks k . Furthermore, the $\|\cdot\|_1$ and $\|\cdot\|_2$ estimations are more sensitive to the dependence form, having smaller bias under $A^*(\theta)$ in comparison to $A'(\theta)$. Regarding the MSE, the estimations obtained by $\|\cdot\|_F$ and $\|\cdot\|_M$ clearly outperform those based on $\|\cdot\|_1$ and $\|\cdot\|_2$. Moreover, $\|\cdot\|_F$ estimations seem to be more stable in terms of the MSE, indicating robustness in both the bias and variance of the estimator. In our simulations the evaluation of $\|\cdot\|_F$ and $\|\cdot\|_M$ required less time and resources to the other matrix norms. Therefore, we suggest using the norms $\|\cdot\|_F$ and $\|\cdot\|_M$ rather than $\|\cdot\|_1$ and $\|\cdot\|_2$ for estimating the value of θ .

Let us now focus on the effect of the sample size m and the number of features n on the performance of $\hat{\theta}$ based on $\|\cdot\|_F$ and $\|\cdot\|_M$. The mean, standard deviation (SD) and MSE of estimations with $\|\cdot\|_F$ and $\|\cdot\|_M$ norms, for $\theta = 0.80, k = 10, m \in \{20, 50, 100, 150, 200\}$ and under block structure $A^*(\theta)$, are illustrated in Figure 1 ($n = 100$) and Figure 2 ($n = 1,000$). The corresponding results under the block matrix form $A'(\theta)$ are presented in Figure 3 ($n = 100$) and Figure 4 ($n = 1,000$). Based on our simulations, when the blocks of $\Sigma(\theta)$ follow the form $A'(\theta)$, it

seems that n has no effect on the estimation SD and MSE and there is a minor bias reduction of the $\|\cdot\|_M$ estimator for larger values of n . In contrast, under the block structure $A^*(\theta)$, increasing the size n leads to a notable deterioration of the performance of $\hat{\theta}$ derived by $\|\cdot\|_M$, while the mean bias, SD and MSE for the $\|\cdot\|_F$ estimator are reduced when n is large. As expected, in all simulation settings the estimation accuracy is improved by increasing the sample size m , since the correlation sample estimation \hat{C} is consistent and approximates $\Sigma(\theta)$ better for larger values of m . In general, the $\|\cdot\|_F$ estimator outperforms the one based on $\|\cdot\|_M$ in terms of the MSE measure, which summarizes both the estimation bias and variance. Thus, for transforming by (3) the test statistics obtained by the simulations in Section 4.2, we use the estimation $\hat{\theta}$ in (5), derived by the Frobenius norm.

To study the effect on the estimation accuracy of $\hat{\theta}$ when the block dependence is misspecified, we consider specific types of mismatches, namely estimations $\hat{\theta}$ of θ based on the block structure $A'(\theta)$, when the true simulated form is $A^*(\theta)$, and estimations $\hat{\theta}$ derived under $A^*(\theta)$, in the case of true underlying dependence simulated in the form $A'(\theta)$. The obtained average bias and MSE values under the same simulation setup as in Table 1 are given in Table 2. From the presented results we conclude that both the average bias and MSE of $\hat{\theta}$ are increased under misspecification of the dependence form, especially in the case when $\hat{\theta}$ is based on $A'(\theta)$, while the true simulated block structure is $A^*(\theta)$. The average bias of $\hat{\theta}$ derived by the column-sum norm $\|\cdot\|_1$ is up



to 51.51% (for $\theta = 0.80$ and $k = 5$). When θ is estimated under $A^*(\theta)$, whereas the true underlying dependence is $A'(\theta)$, the estimations $\hat{\theta}$ are closer to the true value of θ , with Frobenius and Max norms having average bias less than 6.58% when $\theta \geq 0.90$. Note that for all matrix norms and in both misspecification scenarios, the absolute average bias and MSE decrease for larger values of the correlation coefficient θ and the number of blocks k in $\Sigma(\theta)$. In addition, the estimator $\hat{\theta}$ based on the Frobenius

norm $\|\cdot\|_F$ seems more sensitive under misspecified dependence structure in comparison to the estimator $\hat{\theta}$ derived by the Max norm $\|\cdot\|_M$. Similar to the case of correctly specified form, the norms $\|\cdot\|_F$ and $\|\cdot\|_M$ lead to the smallest MSE of $\hat{\theta}$ in the most parameter settings. Nevertheless, the estimations based on $\|\cdot\|_F$ outperform the others for large values of θ and k . Therefore, we recommend Frobenius norm to be used for the estimation $\hat{\theta}$ of θ .

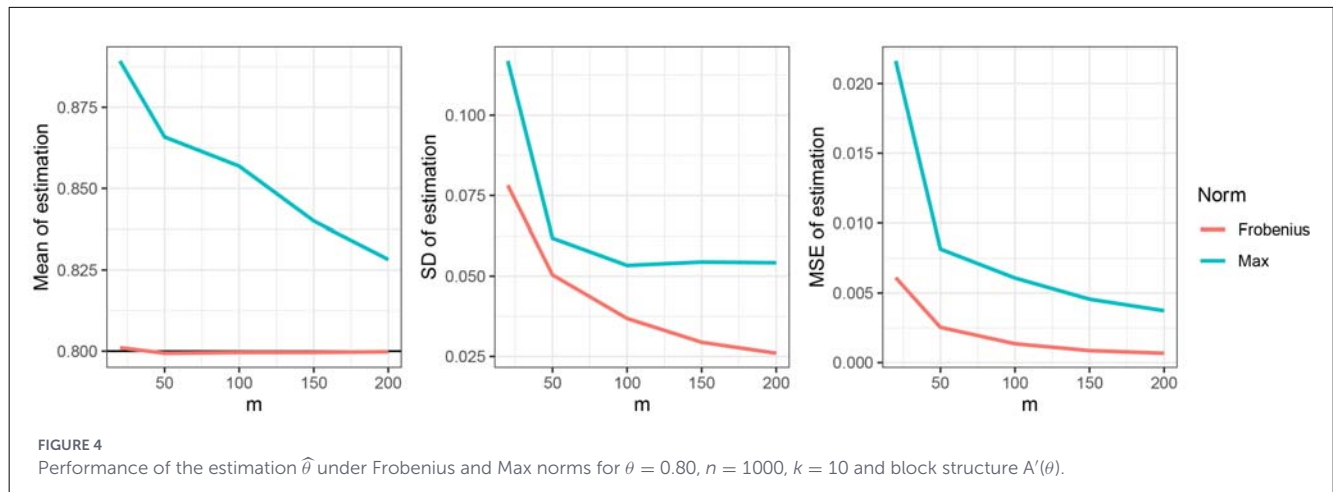


TABLE 2 Average simulation results for the performance of the estimation $\hat{\theta}$ with $n = 100, m = 100, k \in \{5, 10, 20\}$ and $\theta \in \{0.80, 0.90, 0.95\}$ under misspecified block dependence structure.

$k = 5$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.3879	0.1512	-0.2371	0.0580	-0.0801	0.0095	0.1450	0.0211	0.0593	0.0036	0.0154	0.0003
Max norm	-0.2455	0.0634	-0.1543	0.0281	-0.0298	0.0053	0.1311	0.0174	0.0496	0.0026	0.0072	0.0002
Column-sum norm	-0.4121	0.1792	-0.2251	0.0608	-0.0513	0.0103	0.1476	0.0220	0.0633	0.0042	0.0200	0.0006
Spectral norm	-0.2373	0.0581	-0.1685	0.0313	-0.0471	0.0070	0.1436	0.0208	0.0619	0.0040	0.0201	0.0006
$k = 10$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.1861	0.0355	-0.0571	0.0048	0.0405	0.0037	0.0944	0.0090	0.0209	0.0006	-0.0174	0.0004
Max norm	-0.1186	0.0179	-0.0165	0.0017	0.0226	0.0051	0.0787	0.0066	0.0036	0.0003	-0.0295	0.0014
Column-sum norm	-0.1280	0.0278	0.0230	0.0097	0.1299	0.0281	0.1148	0.0139	0.0405	0.0023	0.0009	0.0006
Spectral norm	-0.0545	0.0057	0.0381	0.0058	0.1343	0.0239	0.1189	0.0146	0.0484	0.0028	0.0121	0.0006
$k = 20$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE	Avg. bias	MSE
Frobenius norm	-0.0124	0.0009	0.0590	0.0045	0.1034	0.0118	0.0079	0.0003	-0.0467	0.0024	-	-
Max norm	0.0423	0.0051	0.0871	0.0127	0.2151	0.0506	-0.0204	0.0020	-0.0495	0.0041	-	-
Column-sum norm	0.1248	0.0272	0.2081	0.0574	0.2619	0.0875	0.0730	0.0081	0.0149	0.0030	-	-
Spectral norm	0.2494	0.0674	0.3162	0.1070	0.3668	0.1427	0.1467	0.0228	0.0894	0.0093	-	-

The values on the left are for $\hat{\theta}$ based on $A'(\theta)$ with true simulated dependence $A^*(\theta)$, while on the right are for $A^*(\theta)$ with true simulated structure $A'(\theta)$.

4.2 Performance of the multiple testing procedures

To compare the multiple testing procedures described in Section 3 under the block dependence structures $A^*(\theta)$ and $A'(\theta)$, we consider their mean FDR and power, obtained by averaging the associate results in each simulation. In Table 3 we present the average FDR and power of the B-H and B-Y procedures applied to the p -values corresponding to the raw and transformed t -statistics

in Sections 2.2 and 3.1, respectively, for fixed $\alpha = 0.05, n = 100, m = 100, k = 10$ and varying parameter values $\theta \in \{0.80, 0.90, 0.95\}$ and alternative differences $\mu_j \in \{1/2, 1, 2\}$. The transformed statistics $S_m^{(1)}, S_m^{(2)}, \dots, S_m^{(n)}$ in (4) are computed by estimating Σ with $\Sigma(\hat{\theta})$, where $\hat{\theta}$ is calculated by (5). Since the asymptotic normality of $S_m^{(j)}$ can be used only for large values of the sample size m , the distribution of $S_m^{(j)}$ under H_0 can deviate from $N(0, 1)$ when $m < 100$. Our experience shows that when $m = 20$ or $m = 50$ the distribution of $S_m^{(j)}$ is better approximated by $t(m-2)$

TABLE 3 Average simulation results for the performance of the B-H and B-Y procedures combined with the K-S tests and applied to the raw and transformed t -test statistics with $n = 100$, $m = 100$, $k = 10$, $\alpha = 0.05$, $\mu_j \in \{1/2, 1, 2\}$ and $\theta \in \{0.80, 0.90, 0.95\}$ under block dependence structure $A^*(\theta)$, left, and $A'(\theta)$, right.

$\mu_j = 1/2$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
B-H procedure	0.0400	0.6946	0.0346	0.6806	0.0293	0.6709	0.0369	0.6876	0.0361	0.6835	0.0322	0.6827
B-Y procedure	0.0080	0.4817	0.0059	0.4829	0.0042	0.4655	0.0085	0.4884	0.0073	0.4847	0.0041	0.4926
Transformed B-H	0.0513	0.0484	0.0617	0.0294	0.0685	0.0226	0.0674	0.0316	0.1676	0.0575	0.2392	0.0927
Transformed B-Y	0.0124	0.0186	0.0118	0.0108	0.0199	0.0079	0.0170	0.0106	0.1055	0.0294	0.1759	0.0493
Transformed B-H and K-S	0.0698	0.5618	0.0932	0.3470	0.1096	0.2500	0.1003	0.3864	0.1965	0.3928	0.3314	0.4950
$\mu_j = 1$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
B-H procedure	0.0397	1.0000	0.0376	1.0000	0.0328	1.0000	0.0405	0.9999	0.0397	1.0000	0.0353	1.0000
B-Y procedure	0.0079	0.9997	0.0061	0.9992	0.0044	0.9998	0.0062	0.9993	0.0065	1.0000	0.0058	0.9990
Transformed B-H	0.0494	0.4978	0.0505	0.3453	0.0628	0.2687	0.0566	0.3602	0.1402	0.3526	0.2013	0.3725
Transformed B-Y	0.0115	0.3207	0.0134	0.2046	0.0155	0.1483	0.0155	0.2061	0.0840	0.2117	0.1446	0.2297
Transformed B-H and K-S	0.0555	0.9993	0.0531	0.9943	0.0611	0.9794	0.0596	0.9975	0.1449	0.9964	0.2925	0.9875
$\mu_j = 2$	$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$		$\theta = 0.80$		$\theta = 0.90$		$\theta = 0.95$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
B-H procedure	0.0397	1.0000	0.0376	1.0000	0.0328	1.0000	0.0405	1.0000	0.0397	1.0000	0.0353	1.0000
B-Y procedure	0.0079	1.0000	0.0061	1.0000	0.0044	1.0000	0.0062	1.0000	0.0065	1.0000	0.0058	1.0000
Transformed B-H	0.0477	0.9626	0.0488	0.8947	0.0602	0.8475	0.0525	0.9175	0.1237	0.8704	0.1837	0.8728
Transformed B-Y	0.0099	0.9262	0.0107	0.8191	0.0136	0.7608	0.0135	0.8534	0.0685	0.7978	0.1224	0.7976
Transformed B-H and K-S	0.0731	1.0000	0.0738	1.0000	0.0840	1.0000	0.0805	1.0000	0.1624	1.0000	0.3003	1.0000

than by $N(0, 1)$ in terms of the critical values under H_0 , although $S_m^{(j)}$ is a standardized linear transformation of t -statistics with $m - 2$ degrees of freedom, which is not $t(m - 2)$ distributed. Since for $m \geq 100$ the Student's t distribution is practically identical to the standard normal distribution, in all simulations we have computed the p -values corresponding to $S_m^{(j)}$ by using the $t(m - 2)$ distribution. In addition, we present results for the combination of the B-H methods with the K-S tests, given in Section 3.2.

Based on our simulations, the B-Y algorithm is very conservative with average FDR much smaller than the significance level $\alpha = 0.05$ and the FDR results of the other multiple testing procedures. This is in agreement with the way it is constructed, see [20]. Moreover, in all simulation setups the B-H method has the largest average power compared to the other multiple testing approaches. Thus, under our dependence assumptions, the B-Y is outperformed by the B-H algorithm. However, the average FDR of the B-H procedure is less than $\alpha = 0.05$ as well. Since the dependence structure of Σ is part of the normalization (4), the transformed B-H and B-Y techniques lead to an increase of the FDR, which is also shown by the results in Table 3. Similar to the transformations applied to the raw p -values, the transformed B-H controls the FDR in a better manner compared to the transformed B-Y. Nevertheless, this improvement in the FDR comes at the expense of average power. More specifically, both

methods (transformed B-H and B-Y) suffer from low power in case $\mu_j \in \{1/2, 1\}$, i.e., when the alternative H_1 is closer to the null hypothesis H_0 . In order to fix this disadvantage, we suggest to combine the transformed B-H procedure with the K-S tests, which allows a balance to be achieved between controlling the FDR and preserving power. Further note that the raw B-H and B-Y procedures have almost identical FDR and power results for $\mu_j = 1$ and $\mu_j = 2$, indicating that their performance is stable above a certain threshold. Regarding the influence of the dependence strength, associated with θ , from our simulations it can be concluded that increasing the value of θ , i.e., increasing the correlation between the features in a block, has a negative effect on the FDR control and reduces the average power for all multiple testing methods.

Let us concentrate now on the performance of the multiple testing procedures under various values of the sample size m . Our simulation results for $n = 100$, $k = 10$, $\theta = 0.80$, $\mu_j = 2$ and $m \in \{20, 50, 100, 150, 200\}$ are illustrated in Figures 5, 6 under block structures $A^*(\theta)$ and $A'(\theta)$, respectively. Along with the average FDR and power, we also provide their standard deviation (SD), estimated by the simulation values. It seems that for all testing procedures the FDR stabilizes when $m \geq 100$. The transformed B-H technique shows the best control of the FDR for $m \geq 100$, with good approximation of the significance level $\alpha = 0.05$ and smaller

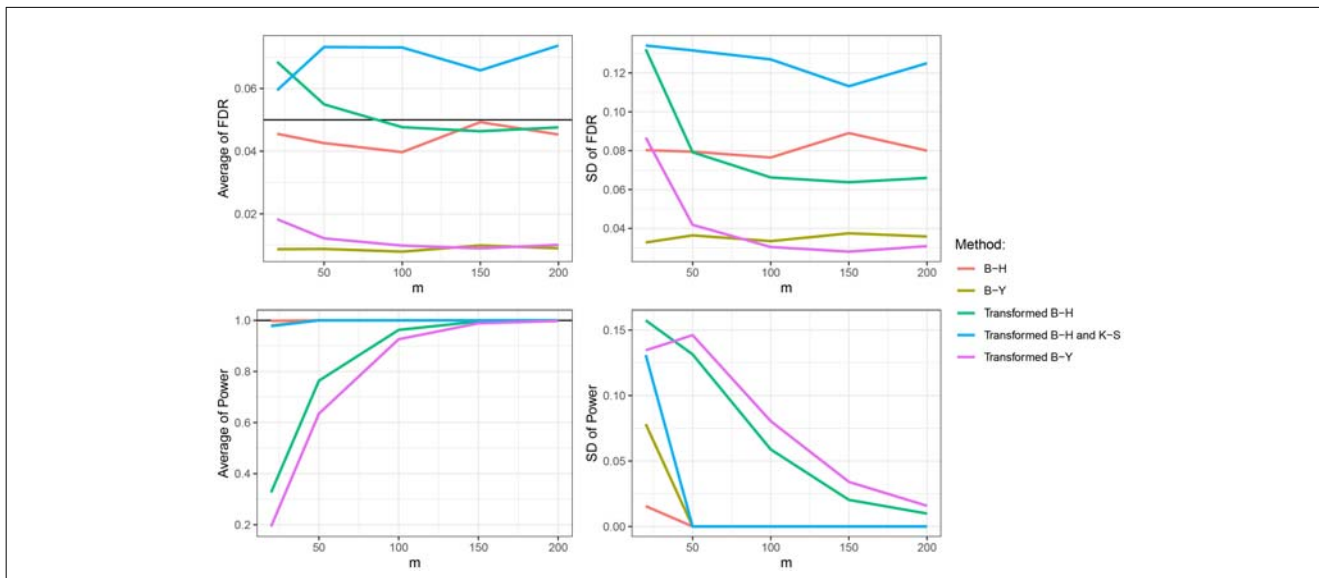


FIGURE 5 Performance of the B-H and B-Y procedures applied to the raw and transformed t -test statistics for $\theta = 0.80$, $n = 100$, $k = 10$, $\alpha = 0.05$, $\mu_j = 2$ and block structure $A^*(\theta)$.

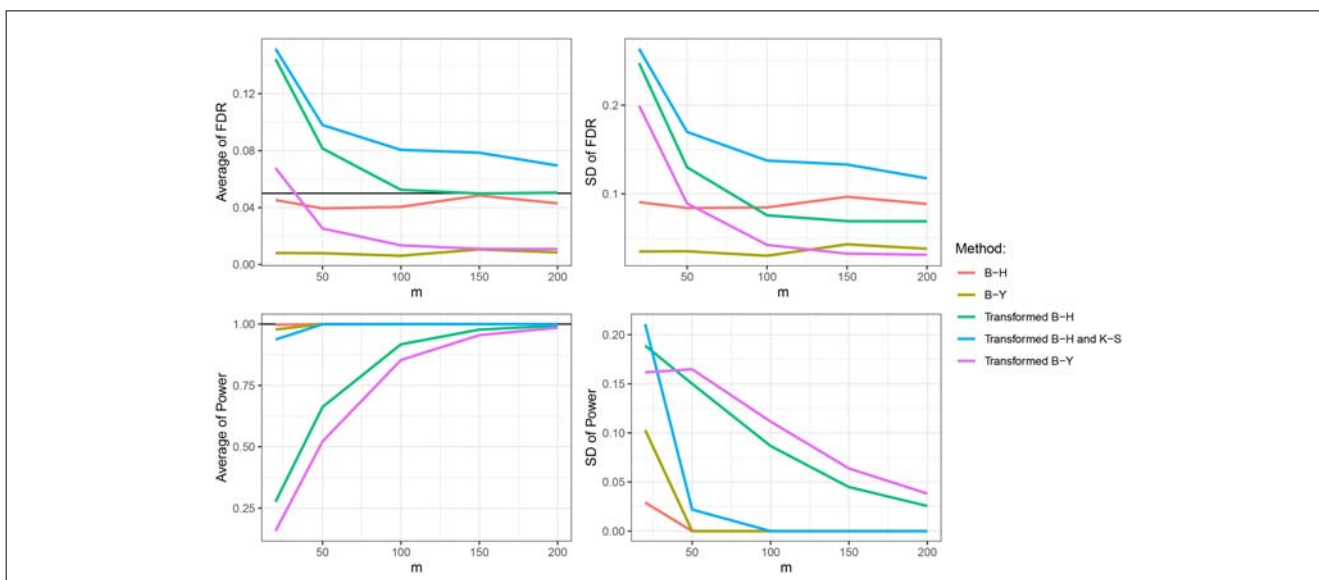


FIGURE 6 Performance of the B-H and B-Y procedures applied to the raw and transformed t -test statistics for $\theta = 0.80$, $n = 100$, $k = 10$, $\alpha = 0.05$, $\mu_j = 2$ and block structure $A'(\theta)$.

SD than the one based on the raw B-H algorithm. One can observe that the SD of FDR for the B-Y and transformed B-Y methods are very small since the corresponding FDR values are close to 0 and thus have a small room for variation. Since the transformed B-H and B-Y have an unsatisfactory average power when the small sample size m is small, it appears that for $m < 100$ the transformed schemes should be combined with the K-S tests. However, the raw B-H procedure outperforms the others in terms of average power when $m < 100$, whereas the transformed B-H method has the best balance of FDR control and power for larger samples ($m \geq 100$). Hence, in case the alternative hypothesis H_1 is closer to the null H_0 , i.e., μ_j is closer to 0, we recommend to combine the transformed approaches with the K-S tests when the sample size m is not large

enough. In addition, we should mention that simulation results, obtained for the number of features $n \in \{500, 1000\}$ and which are not included in this paper, suggest similar conclusions as those under $n = 100$.

A comparison between the results shown in Figures 5, 6 suggests that the presented methods have larger average power and FDR that is closer to the desired level α under dependence $A^*(\theta)$ than under the block form $A'(\theta)$. This is also supported by the results for $\theta = 0.80$ in Table 3 and is due the fact that when $\theta = 0.80$ the matrix $A^*(\theta)$ is closer to the identity matrix, which corresponds to independence between all features. However, for larger values of θ the structure $A^*(\theta)$ presents stronger correlation than the form $A'(\theta)$, leading to a better FDR control and larger

power of the multiple testing procedures under $A'(\theta)$ compared to those under $A^*(\theta)$, as shown by the results for $\theta \in \{0.90, 0.95\}$ in Table 3.

5 Concluding discussion

In this paper we considered the B-H and B-Y multiple testing procedures for controlling the FDR under positive block dependence. We introduced two parametric block dependence structures to capture correlations between paired t -test statistics. For estimating the value of the unknown parameter, we proposed algorithms based on different matrix norms. Our simulation results showed that estimations based on the Frobenius norm outperform those derived under Max, column-sum and spectral norms in terms of MSE and SD, with an average bias less than 0.13% in most of the parameter settings.

Our simulation study on the performance of the classical B-H and B-Y methods under dependence indicated that these multiple testing procedures have inadequate FDR control. Moreover, their ability to maintain the intended FDR decreases as the dependence strength grows. To address this limitation, we proposed a transformation of the t -test statistics by incorporating the estimated dependence structure. From our simulations we concluded that the B-H and B-Y algorithms applied to the transformed statistics have better control of the FDR in comparison to the traditional approaches. However, this improvement in the FDR comes at the cost of reducing the average power. Therefore, we recommend to combine those procedures with K-S tests, especially when they have small statistical power. Although there is no guarantee for a strict FDR control, in our simulation setup we achieved a balance between FDR control and test power by combining both methods.

The number of scenarios for describing different types of dependence and significance structures is extremely high, probably resulting in having a small number of papers about FDR procedures under dependence [22]. In this study we describe an initial approach to modeling such structures and we focus only on basic dependence structures with limited number of parameters and fixed blocks of significant variables. We assume that the correlation matrix between the tested statistics have a block structure that can be exponentially or linearly parameterized by a single parameter. We further assume that all variables within a given block are associated with either true null hypotheses or true alternatives, which corresponds to the notion of the genes in a particular genetic pathway being highly correlated. These simplifications are necessary, given that this is an initial approach for the problem, but somehow limit the practical applications of the suggested models and methods. If the dependence structure assumptions are not valid, the transformed statistics may produce incorrect p -values due to change of their distribution. This can lead to false rejections and to the failure of a multiple testing procedure to control the FDR. Thus, more complex scenarios that include more general cases or assumptions are a natural extension of this work.

As a future work it would be worth extending the presented techniques in this paper to broader dependence structures and real-world applications. It would be interesting to develop an algorithm for identifying different disjoint groups in a correlation matrix, as a generalization of the method proposed by Perreault et al. [29]. In

addition to the K-S tests, combining multiple comparison methods with other nonparametric tests can be explored.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

NIN: Formal analysis, Methodology, Visualization, Conceptualization, Writing – original draft, Software, Investigation. MS: Formal analysis, Funding acquisition, Supervision, Writing – review & editing. DP: Software, Investigation, Writing – review & editing, Methodology, Supervision, Conceptualization.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was financed by the European Union - NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project No. BG-RRP-2.004-0008.

Acknowledgments

Some computations in this work were performed on the supercomputer Avitohol described in Atanassov et al. [34]. The authors acknowledge the provided access to the infrastructure purchased under the National Roadmap for RI, financially coordinated by the MES of the Republic of Bulgaria, grant No. D01-98/26.06.2025.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Statist Soc: Series B.* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Van Noorden R, Maher B, Nuzzo R. The top 100 papers. *Nature.* (2014) 514:550–3. doi: 10.1038/514550a
- Bennett CM, Wolford GL, Miller MB. The principled control of false positives in neuroimaging. *Soc Cogn Affect Neurosci.* (2009) 4:417–22. doi: 10.1093/scan/nsp053
- Lindquist MA. The statistical analysis of fMRI data. *Statist Sci.* (2008) 23:439–64. doi: 10.1214/09-STS282
- Schwartzman A, Dougherty RF, Taylor JE. False discovery rate analysis of brain diffusion direction maps. *Ann Appl Stat.* (2008) 2:153. doi: 10.1214/07-AOAS133
- Boffetta P, McLaughlin JK, La Vecchia C, Tarone RE, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst.* (2008) 100:988–95. doi: 10.1093/jnci/djn191
- Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol.* (2014) 67:850–7. doi: 10.1016/j.jclinepi.2014.03.012
- Ge W, Fazal Z, Jakobsson E. Using optimal F-measure and random resampling in gene ontology enrichment calculations. *Front Appl Mathem Statist.* (2019) 5:20. doi: 10.3389/fams.2019.00020
- Klopfenstein DV, Zhang L, Pedersen BS, Ram F, Warwick Vesztracy A, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. *Scient Reports.* (2018) 8:10872. doi: 10.1038/s41598-018-28948-z
- Lutz KC, Jiang S, Neugent ML, De Nisco NJ, Zhan X, Li Q, et al. Survey of statistical methods for microbiome data analysis. *Front Appl Mathem Statist.* (2022) 8:884810. doi: 10.3389/fams.2022.884810
- Xiao J, Cao H, Chen J. False discovery rate control incorporating phylogenetic tree increases detection power in microbiome-wide multiple testing. *Bioinformatics.* (2017) 33:2873–81. doi: 10.1093/bioinformatics/btx311
- Mayo DG. Significance tests: vitiated or vindicated by the replication crisis in psychology? *Rev Philos Psychol.* (2021) 12:101–20. doi: 10.1007/s13164-020-00501-w
- Shrout PE, Rodgers JL. Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annu Rev Psychol.* (2018) 69:487–510. doi: 10.1146/annurev-psych-122216-011845
- Benjamini Y, Heller R. False discovery rates for spatial signals. *J Am Stat Assoc.* (2007) 102:1272–81. doi: 10.1198/016214507000000941
- Caldas de Castro M, Singer BH. Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association. *Geograp Anal.* (2006) 38:180–208. doi: 10.1111/j.0016-7363.2006.00682.x
- Barras L, Scaillet O, Wermers R. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *J Finance.* (2010) 65:179–216. doi: 10.1111/j.1540-6261.2009.01527.x
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* (2015) 43:e47–e47. doi: 10.1093/nar/gkv007
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* (2014) 15:1–21. doi: 10.1186/s13059-014-0550-8
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616
- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Statist.* (2001) 29:1165–88. doi: 10.1214/aos/1013699998
- Schwartzman A, Lin X. The effect of correlation in false discovery rate estimation. *Biometrika.* (2011) 98:199–214. doi: 10.1093/biomet/asq075
- Chi Z, Ramdas A, Wang R. Multiple testing under negative dependence. *Bernoulli.* (2025) 31:1230–55. doi: 10.3150/24-BEJ1768
- Hartung J. A note on combining dependent tests of significance. *Biometrical J: J Mathem Methods Biosci.* (1999) 41:849–55.
- Vovk V, Wang R. E-values: Calibration, combination and applications. *Annals Statist.* (2021) 49:1736–54. doi: 10.1214/20-AOS2020
- Wang R, Ramdas A. False Discovery Rate Control with E-values. *J Royal Statist Soc Series B: Statist Methodol.* (2022) 84:822–52. doi: 10.1111/rssb.12489
- Zhao Z, Sun W. False discovery rate control for structured multiple testing: asymmetric rules and conformal Q-values. *J Am Stat Assoc.* (2025) 120:805–17. doi: 10.1080/01621459.2024.2359739
- Kim E, Ivanov I, Hua J, Lampe J, Hullar M, Chapkin R, et al. The model-based study of the effectiveness of reporting lists of small feature sets using RNA-Seq data. *Cancer Inform.* (2017) 16:1–15. doi: 10.1177/1176935117710530
- Dang TN, Murray P, Forbes AG. PathwayMatrix: visualizing binary relationships between proteins in biological pathways. In: *BMC Proceedings*, vol. 9. Cham: Springer (2015). p. S3.
- Perreault S, Duchesne T, Nešlehová JG. Detection of block-exchangeable structure in large-scale correlation matrices. *J Multivariate Analy.* (2019) 169:400–22. doi: 10.1016/j.jmva.2018.10.009
- Gentle JE. *Matrix Algebra: Theory, Computations, and Applications in Statistics*. New York: Springer. (2007).
- Chi Z. On the performance of FDR control: constraints and a partial solution. *Ann Stat.* (2007) 35:1409–31. doi: 10.1214/009053607000000037
- Gibbons JD, Chakraborti S. *Nonparametric Statistical Inference*, 5th ed. Chapman and Hall/CRC. (2010).
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team. (2025). Available online at: <https://www.R-project.org/> (Accessed January 28, 2026).
- Atanassov E, Gurov T, Ivanovska S, Karaivanova A. Parallel Monte Carlo on Intel MIC architecture. *Procedia Comput Sci.* (2017) 108:1803–10. doi: 10.1016/j.procs.2017.05.149