Statistica Sinica **32** (2022), 89-108
doi:https://doi.org/10.5705/ss.202019.0425

A PERMUTATION TEST FOR TWO-SAMPLE MEANS AND SIGNAL IDENTIFICATION OF HIGH-DIMENSIONAL DATA

Efang Kong[1], Lengyang Wang[2], Yingcun Xia[2] and Jin Liu[2]

[1] *University of Electronic Science and Technology of China*
and [2] *National University of Singapore*

Abstract: Permutation tests are widely used in practice. However, these tests either need restrictive assumptions for validity, or are not applicable to high-dimensional data. This study considers permutation tests for high-dimensional mean comparisons. Here, in order to get around these restrictions, the test statistics are calculated based on pseudo samples generated using a "binning" procedure. The corresponding permutation tests are proved to be asymptotically consistent. We also consider a related problem for signal identification and establish the asymptotic properties of the tests. Simulation studies demonstrate the favorable performance of our methods compared with that of existing tests. Finally, the proposed method is applied to a genome-wide association study for seven complex human diseases to identify possible single nucleotide polymorphisms associated with the diseases.

Key words and phrases: Consistency of test, high-dimensional data, permutation tests, signal identification, test of mean-difference

1. Introduction

Testing the equality of the means of two random vectors based on random samples is a long-standing issue in multivariate analysis. The past two decades have witnessed increasing interest in this problem for high-dimensional settings. Existing methods are divided into two categories. The first group are based on the sum-of-squares of the sample mean differences; see, for example, Bai and Saranadasa (1996) and Chen and Qin (2010). These methods are generally more powerful against dense alternatives, in the sense that there is a large proportion of small to moderate component-wise differences. Those in the second group are based on the infinity norm of the mean differences; see, for example, Cai, Liu and Xia (2014), Xu et al. (2016), Chang et al. (2017), and Xue and Yao (2018). These methods are better suited to testing against sparse alternatives, that is, when there are only a few, but significant component-wise differences.

Corresponding author: Yingcun Xia, National University of Singapore, 119077 Singapore. E-mail: staxyc@nus.edu.sg.

This study focuses on permutation methods, which have served as a useful alternative to traditional methods for hypothesis testing; see Good (2005) and Ernest (2004) for a comprehensive review. The basic idea is to generate a reference distribution by recalculating a statistic for many permutations of the data. To illustrate, suppose the p-dimensional random vectors $X_1, \ldots, X_m$ are $\overset{i.i.d.}{\sim} P_1(\cdot)$ with mean μ^X and variance Σ^X, and $Y_1, \ldots, Y_n$ are $\overset{i.i.d.}{\sim} P_2(\cdot)$ with mean μ^Y and variance Σ^Y. Write $N = m + n$, and suppose that $m/N \to c$, for some constant $c \in (0, 1)$. Our interest is to test the null hypothesis

$$H_0 : \mu^X = \mu^Y.$$

Chung and Romano (2013) test H_0 using permutation methods for $p = 1$. The procedure is as follows. Write $Z^N = \{Z_1, \ldots, Z_N\}$, with $Z_i = X_i$, for $1 \le i \le m$, and $Z_{m+j} = Y_j$, for $1 \le j \le n$. Consider the standardized statistic

$$S_N(Z^N) = \frac{N^{1/2}(\bar{X}_m - \bar{Y}_n)}{\sqrt{(N/m)\hat{\sigma}_m^2(Z^N) + (N/n)\hat{s}_n^2(Z^N)}}, \tag{1.1}$$

where $\bar{X}_m$ and $\bar{Y}_n$ are the sample means of $\{Z_1, \ldots, Z_m\}$ and $\{Z_{m+1}, \ldots, Z_N\}$, respectively, and $\hat{\sigma}_m^2(Z^N)$ and $\hat{s}_n^2(Z^N)$ are the corresponding sample variances. Let G_N be the set of all permutations of $\{1, \ldots, N\}$. For any $\pi \in G_N$, let Z_π^N denote the rearranged Z^N through permutation π, and $Z_{\pi(i)}^N$, for $i = 1, \ldots, N$, be the ith entry of Z_π^N. Recompute $S_N(Z_\pi^N) \equiv S_N(Z_{\pi(1)}^N, \ldots, Z_{\pi(N)}^N)$, and let $\hat{R}_N^S(\cdot)$ denote the empirical distribution of $S_N(Z_\pi^N)$ evaluated at all $N!$ permutations of Z^N; that is,

$$\hat{R}_N^S(t) = \frac{1}{N!} \sum_{\pi \in G_N} I\{S_N(Z_\pi^N) \le t\}.$$

This empirical distribution $\hat{R}_N^S(\cdot)$, also referred to as the permutation distribution, is used as an approximation of the null (limiting) distribution of statistic (1.1), which in this case is given by $\Phi(\cdot)$, the distribution function of the standard normal $N(0, 1)$. We reject H_0 if $\hat{R}_N^S(S_N(Z^N)) \ge 1 - \alpha$. Chung and Romano (2013) proved that

$$\sup_{t \in R} |\hat{R}_N^S(t) - \Phi(t)| \to 0 \quad \text{in probability,}$$

and in this sense, the permutation procedure based on statistic (1.1) is considered to be consistent (valid). In general, however, the consistency of permutation tests should not be taken for granted. Indeed, Chung and Romano (2013) showed that the permutation test based on $S_N(Z^N) = \bar{X}_m - \bar{Y}_n$, that is (1.1) without

standardization, is inconsistent unless $c = 1/2$ or $\Sigma^X = \Sigma^Y$.

Clearly, in high-dimensional cases, in which the dimension p could far exceed the sample size, permutation tests based on the standardized statistic (1.1) are no longer applicable. In addition, the pre-pivoting method of Chung and Romano (2016), proposed for a multivariate setting, is not computationally feasible. This study intends to fill this gap by proposing a permutation procedure that is both asymptotically consistent and easy to implement, even for ultrahigh-dimensional data.

The rest of this paper is organized as follows. Section 2 begins with a basic formulation of the problem, and then presents results for the consistency of the permutation tests based on Hotelling's T^2-type statistics. These statistics require estimating the inverse of a covariance matrix, which renders their use impractical in a high-dimensional setting. Therefore, we describe alternative in Section 2.2, where we propose a "binning" procedure to produce pseudo samples, from which the test statistics are then derived. Section 3 applies the proposed tests to identify those variables that are the source of the difference between two high-dimensional means, which we refer to as signal identification. Some related theoretical results are also given. The numerical performance of the proposed methods and other existing methods are examined in Section 5 using simulation studies. Section 6 contains an empirical study of the genome-wide association for seven complex diseases using data from the Wellcome Trust Case Control Consortium (WTCCC). The assumptions needed for the asymptotic studies are given in the Appendix, and all technical proofs are relegated to the online Supplementary Material.

2. Permutation Tests for High-Dimensional Mean Comparison

We first introduce some notation. For any $v = (v_1, \ldots, v_p)^\top \in R^p$, let $|v|_\gamma = \{(|v_1|^\gamma + \cdots + |v_p|^\gamma)/p\}^{1/\gamma}$, for any $\gamma > 0$. In particular, $|v|_1 = (|v_1| + \cdots + |v_p|)/p$ stands for the L_1-norm, and $|v|_\infty = \max_{k=1,\ldots,p} |v_k|$ is the L_∞-norm. Write $\bar{X}_m = m^{-1}\sum_i X_i$, $\bar{Y}_n = n^{-1}\sum_j Y_j$, $\delta_N = (\delta_{N,1}, \ldots, \delta_{N,p})^\top = N^{1/2}(\bar{X}_m - \bar{Y}_n)$, and

$$\hat{\Sigma}_m^X = \frac{1}{m}\sum_i (X_i - \bar{X}_m)(X_i - \bar{X}_m)^\top, \quad \hat{\Sigma}_n^Y = \frac{1}{n}\sum_j (Y_j - \bar{Y}_n)(Y_j - \bar{Y}_n)^\top. \quad (2.1)$$

Denote by $\hat{\sigma}^2_{m,k}(X_1, \ldots, X_m)$ and $\hat{s}^2_{n,k}(Y_1, \ldots, Y_n)$, for $k = 1, \ldots, p$, the diagonal elements of $\hat{\Sigma}_m^X$ and $\hat{\Sigma}_n^Y$, respectively. Write $\Sigma(\bar{P}) = c\Sigma^X + (1-c)\Sigma^Y$.

2.1. Permutation tests based on Hotelling's T^2-type statistics

Write $\tilde{\Sigma} = c^{-1}\Sigma^X + (1-c)^{-1}\Sigma^Y$, the variance of δ_N, and suppose $\tilde{\Omega}_N = \tilde{\Omega}_N(Z^N)$ is an estimator of $\tilde{\Omega} = \tilde{\Sigma}^{-1}$. Then, in a manner similar to (1.1), define $e_N(Z^N) = \{\tilde{\Omega}_N\}^{1/2}\delta_N$ and

$$H^\gamma(Z^N) \equiv |e_N(Z^N)|_\gamma, \quad \gamma = 1 \text{ or } \infty. \tag{2.2}$$

Xu et al. (2016) considered using other values for γ, but in this study on permutation tests for high dimensions, we focus only on the cases where $\gamma = 1$ or ∞. In practice, these two choices should serve the purposes, because using $H^1(\cdot)$ is expected to be more powerful against dense alternatives, whereas $H^\infty(\cdot)$ works better against sparse alternatives. The latter also finds important applications in signal identification; see, for example, Benjamini and Hochberg (1995) and Jin and Cai (2007). For the permutation tests based on test statistics (2.2), we have the following result.

Theorem 1. *Suppose conditions* (C1)–(C5) *of Section* 4 *hold. Then,*

$$\text{under } H_0,\ \sup_{t\in R}\Big|\frac{1}{N!}\sum_{\pi\in G_N} I\{H^\infty(Z_\pi^N) < t\} - Pr\Big(H^\infty(Z^N) \le t\Big)\Big| \xrightarrow{p} 0, \tag{2.3}$$

where $\xrightarrow{p}$ *stands for convergence in probability. Parallel results hold for* $H^1(\cdot)$ *if conditions* (C1)–(C2), (C3′) *and* (C4)–(C5) *of Section* 4 *hold.*

In other words, the permutation tests based on (2.2) with $\gamma = 1$ and $\gamma = \infty$ are both consistent. However, in high-dimensional settings, these tests are difficult to implement, owing to challenges with estimating the high-dimensional precision matrix $\tilde{\Omega}$, if at all possible. A naive solution is to standardize (divide) the entries of δ_N by their marginal standard error. That is, with

$$v_{N,k}^2 = \frac{N}{m}\hat{\sigma}_{m,k}^2(X_1,\ldots,X_m) + \frac{N}{n}\hat{s}_{n,k}^2(Y_1,\ldots,Y_n), \tag{2.4}$$

consider the following test statistics:

$$S^1(Z^n) = p^{-1}\sum_{k=1}^p \Big|\frac{\delta_{N,k}}{v_{N,k}}\Big|, \quad S^\infty(Z^n) = \max_{1\le k\le p}\Big|\frac{\delta_{N,k}}{v_{N,k}}\Big|. \tag{2.5}$$

Theorem 2. *If conditions* (C1)–(C3) *and* (C6) *of Section* 4 *hold, then*

$$\sup_{t\in R}\Big|\frac{1}{N!}\sum_{\pi\in G_N} I\{S^\infty(Z_\pi^N) < t\} - Pr(|\Xi|_\infty < t)\Big| \xrightarrow{p} 0, \tag{2.6}$$

where Ξ *is a p-dimensional Gaussian, with covariance matrix given by* $[diag(\Sigma(\bar{P}))]^{-1/2}\Sigma(\bar{P})[diag(\Sigma(\bar{P}))]^{-1/2}$*, the correlation matrix associated with* $\Sigma(\bar{P})$*; on the other hand,*

$$\text{under } H_0, \quad \sup_{t\in R}\left|P\Big(S^{\infty}(Z^N)\le t\Big) - Pr(|\tilde{\Xi}|_{\infty} < t)\right| \to 0, \tag{2.7}$$

where $\tilde{\Xi}$ *is also a p-dimensional Gaussian, with covariance matrix given by* $[diag(\tilde{\Sigma})]^{-1/2}\tilde{\Sigma}[diag(\tilde{\Sigma})]^{-1/2}$*, the correlation matrix given by that of* $\tilde{\Sigma}$*. Parallel results hold for* $S^1(\cdot)$ *under conditions* (C1)–(C2)*,* (C3′)*, and* (C6)*.*

Because $\Sigma(\bar{P}) = c\Sigma^X + (1-c)\Sigma^Y$, permutation tests based on $S^{\gamma}(\cdot)$ are, in general, inconsistent, except when $\Sigma^X = \Sigma^Y$ or $c = 1/2$; this is also noted in Chung and Romano (2016) for the finite-dimension case. To correct the inconsistency associated with statistic (2.5), $S^{\infty}(\cdot)$, the permutation tests in Chung and Romano (2016) are coupled with a pre-pivoting procedure: for each permutation, bootstrapping is implemented to get an estimate of a "pre-pivoted" statistic. However, the significant computation required means this approach is not practically feasible. Moreover, their theoretical results were established only for the fixed-dimensional setting. Our solution is described in the next section.

2.2. A "binning" procedure and pseudo samples

The purpose of this procedure is to produce two pseudo samples of equal size. Without loss of generality, suppose $m > n$, such that $m = K \times n + k$, for some nonnegative integers K and k, with $0 \le k < n$. Thus, $K = [c/(1-c)]$, the integer part of $c/(1-c)$, and $k/n \to c/(1-c) - K$. Define

$$X_i' = X_i - \mu^X, \quad Y_j' = Y_j - \mu^X,\ i = 1,\dots,m,\ j = 1,\dots,n; \tag{2.8}$$

in practice, $\bar{X}_m$ can be used as a substitute for μ^X. The pseudo observations are then constructed as follows. If $k = 0$, define

$$X_i^* = \frac{n}{m}\sum_{j=(i-1)K+1}^{i\times K} X_j', \quad i = 1,\dots,n.$$

If $k > 0$, first randomly select k from $\{X_i^*,\, i = 1,\dots,n\}$, defined above, and assign each to one of the leftover $X'_{K\times n+i}$, for $i = 1,\dots,k$. Specifically, and without loss of generality, define

$$X_i^* := X_i^* + \frac{n}{m}X'_{K\times n+i}, \quad i = 1,\dots,k. \tag{2.9}$$

We call $\{X_1^*, \ldots, X_n^*\}$ and $\{Y_1', \ldots, Y_n'\}$ the pseudo samples. Note that although some of the pseudo observations X_i^* are derived from K original X_i, while others are derived from $K+1$ original X_i, these X_i^* are nevertheless identically distributed (see the proof of Theorem 3). More importantly, if the null hypothesis H_0 holds for the original observations X_i and Y_j, then it also holds for the pseudo samples, and vice versa. From now on, all steps involved in the permutation test are applied to these pseudo samples instead of the original X_i and Y_j.

Write $Z^n = \{Z_1, \ldots, Z_{2n}\}$, such that $Z_i = X_i^*$, $Z_{n+j} = Y_j'$, for $i, j = 1, \ldots, n$. Recall that $X_1^*, \ldots, X_n^*$ denote the first n elements of Z^n, and $Y_1', \ldots, Y_n'$ denote the remaining ones. Let $\bar{X}^* = n^{-1}\sum_{i=1}^n X_i^*$ and $\bar{Y}^* = n^{-1}\sum_{j=1}^n Y_j'$ be the two sample means. Write $\delta_n^* = (\delta_{n,1}^*, \ldots, \delta_{n,p}^*)^\top = n^{1/2}(\bar{X}^* - \bar{Y}^*)$, and consider the following simple test statistics:

$$S_0^1(Z^n) = |\delta_n^*|_1, \quad S_0^\infty(Z^n) = |\delta_n^*|_\infty. \tag{2.10}$$

Apparently, these statistics do not take into account the differences in the variations of the variables. Thus, an arguably improved alternative is such that

$$S_1^1(Z^n) = p^{-1}\sum_{k=1}^p \Big|\frac{\delta_{n,k}^*}{v_{n,k}^*}\Big|, \quad S_1^\infty(Z^n) = \max_{k=1,\ldots,p}\Big|\frac{\delta_{n,k}^*}{v_{n,k}^*}\Big|, \tag{2.11}$$

where $v_{n,k}^* = \{\hat{\sigma}_{n,k}^2(X_1^*, \ldots, X_n^*) + \hat{s}_{n,k}^2(Y_1', \ldots, Y_n')\}^{1/2}$ is the estimator of the variance of $\delta_{n,k}^*$. Denote by Z_π^n the rearranged Z^n through any given permutation $\pi \in G_{2n}$, and $S_1^\gamma(Z_\pi^n) \equiv S(Z_{\pi(1)}^n, \ldots, Z_{\pi(2n)}^N)$. The distribution of $S_1^\gamma(Z^n)$ is then given by the empirical distribution of $S_1^\gamma(Z_\pi^n)$, evaluated at all $(2n)!$ permutations of Z^n.

Theorem 3. *The permutation tests based on $S_0^\infty(\cdot)$ of* (2.10) *are consistent under conditions* (C1)–(C3) *of Section* 4. *Similarly, the permutation test based on $S_0^1(\cdot)$ is consistent under conditions* (C1)–(C2), *and* (C3′). *The same conclusions hold for the permutation tests based on $S_1^\infty(\cdot)$ or $S_1^1(\cdot)$ if condition* (C6) *of Section* 4 *also holds.*

Numerical evidence suggests that in terms of type-I error control, the tests based on $S_0^\gamma(\cdot)$ are more stable than those based on $S_1^\gamma(\cdot)$, especially when p is large. However, note that, in general, the latter possess better power, because they take into account the possibility of different marginal standard errors.

3. Signal Identification

Write $\delta_0 = (\delta_{01}, \ldots, \delta_{0p})^\top = \mu^X - \mu^Y$. Denote by $I_0 \subseteq \{1, \ldots, p\}$, such that

$$|\delta_{0k}| > 0, \ \forall k \in I_0; \quad |\delta_{0k}| = 0, \ \forall k \notin I_0.$$

This is referred to as the set of signals. The number of signals, that is, the cardinality of I_0, can increase with p.

Let $\tilde{t}_{n,p}(\cdot)$ stand for the permutation distribution function of $S_1^\infty = \max_{1\le k\le p} |\delta^*_{n,k}/v^*_{n,k}|$, and $\tilde{t}^{-1}_{n,p}(\cdot)$, its inverse. The significance level α_n is chosen such that $q_{\alpha_n}/(2\ln p)^{1/2} \to 1$ where $q_\alpha = -\ln(\pi) - 2\ln(-\ln(1-\alpha))$ is the $(1-\alpha)$ quantile of the type-I extreme value distribution $F(x) = \exp(-\exp\{-(\ln\pi + x)/2\})$. In other words, α_n is such that

$$\frac{\ln\{-\ln(1-\alpha_n)\}}{(\ln p)^{1/2}} \to \frac{-\sqrt{2}}{2}. \tag{3.1}$$

Consequently, the estimated set of signals is defined as

$$\hat{I}_n = \left\{ k : \left| \frac{\delta^*_{n,k}}{v^*_{n,k}} \right| > \tilde{t}^{-1}_{n,p}(1-\alpha_n),\ k = 1, \ldots, p \right\}.$$

Theorem 4. *Suppose conditions* (C1)–(C3) *and* (C6) *in Section* 4 *hold. If*

$$\liminf_{n,p\to\infty} \left(\frac{c}{s_1}\right)^{1/2} n^{1/2} (\ln p)^{-1/2} \min_{k\in I_0} |\delta_{0k}| \ge 2\sqrt{2}, \tag{3.2}$$

where s_1 *is as given in* (C2) *and* α_n *satisfies* (3.1)*, then as* $n, p \to \infty$,

$$Pr(\hat{I}_n = I_0) \to 1.$$

In other words, if the strength of the signals, measured using $\min_{k\in I_0} |\delta_{0k}|$, is sufficiently strong enough, the set of signals can be correctly identified in probability.

4. Notation and Assumptions

For any square matrix $M = [m_{ij}]$, $\|M\|_{(1,1)} = \max_j \sum_i |m_{ij}|$, where $\lambda_{max}(M)$ and $\lambda_{min}(M)$ denote the largest and smallest absolute eigenvalues, respectively of M. We assume the following conditions:

(C1) $\lim_{m\to\infty} m/N = c \in (0,1)$ and $c - m/N = O(N^{-1/2})$.

(C2) There exist constant $s_1 > s_0 > 0$, such that $s_0 \le \sigma^2_{kk}, s^2_{kk} \le s_1$.

(C3) $\ln(p) = O(n^{\alpha})$, $\alpha < 1/7$; there exist finite constants c_1, $c_2 > 0$, such that

$$E[|X_{i,k}|^{2+l})] \le c_1^l,\ E[|Y_{j,k}|^{2+l})] \le c_2^l,\ k = 1, \ldots, p,\ l = 1, 2;$$
$$E\left\{ \exp\left(\frac{X_{i,k}}{c_1} \right) \right\} \le 2,\quad E\left\{ \exp\left(\frac{Y_{j,k}}{c_2} \right) \right\} \le 2,\ k = 1, \ldots, p.$$

(C3′) $p = O(n^{\alpha})$, $\alpha < 1/7$; for $\nu = \{p^{-1/2}(v_1, v_2, \ldots, v_p)^{\top} : v_j = 1 \text{ or } -1\}$, $\tilde{X}_i = (v^{\top} X_i)_{v \in \nu}$, and $\tilde{Y}_j = (v^{\top} Y_j)_{v \in \nu}$, for $i = 1, \ldots, m$, and $j = 1, \ldots, n$, there exist finite constants $\tilde{c}_1 > 0$, $\tilde{c}_2 > 0$, such that

$$E[|\tilde{X}_{i,k}|^{2+l})] \le \tilde{c}_1^l,\ E[|\tilde{Y}_{j,k}|^{2+l})] \le \tilde{c}_2^l,\ k = 1, \ldots, 2^{p-1},\ l = 1, 2;$$
$$E\left\{ \exp\left(\frac{\tilde{X}_{i,k}}{\tilde{c}_1} \right) \right\} \le 2,\quad E\left\{ \exp\left(\frac{\tilde{Y}_{j,k}}{\tilde{c}_2} \right) \right\} \le 2,\ k = 1, \ldots, 2^{p-1}.$$

(C4) The eigenvalues of Σ^X and Σ^Y are bounded from both below and above by some constants $0 < c_3 < c_4$.

(C5) $\tilde{\Omega}_N$ is an estimate of $\tilde{\Omega} = \tilde{\Sigma}^{-1}$ that satisfies the following condition:

$$\|\{\tilde{\Omega}_N\}^{1/2} - \{\tilde{\Omega}\}^{1/2}\|_{(1,1)} = o_p(\{\ln p\}^{-1}); \tag{4.1}$$

similarly, for $\tilde{\Omega}_N = \tilde{\Omega}_N(Z_1, \ldots, Z_N)$, with $Z_1, \ldots, Z_N \overset{i.i.d.}{\sim} \bar{P} = cP_1(\cdot) + (1 - c)P_2(\cdot)$ (the mixture distribution), we have

$$\left\| \{\tilde{\Omega}_N\}^{1/2} - \left\{ \frac{\Sigma(\bar{P})}{c(1-c)} \right\}^{-1/2} \right\|_{(1,1)} = o_p(\{\ln p\}^{-1}). \tag{4.2}$$

(C6) $\hat{\sigma}_{m,k}^2$ and $\hat{s}_{n,k}^2$, for $k = 1, \ldots, p$, defined in (2.1), are consistent, and

$$\max_{1 \le k \le p} \left| \frac{\hat{\sigma}_{m,k}^2}{\sigma_{kk}^2} - 1 \right| = o_p\left(\frac{1}{\ln p} \right),\ \max_{1 \le k \le p} \left| \frac{\hat{s}_{n,k}^2}{s_{kk}^2} - 1 \right| = o_p\left(\frac{1}{\ln p} \right); \tag{4.3}$$

in a sense similar to (4.2), (4.3) also holds for the same statistic based on independent and identically distributed (i.i.d.) observations from the mixture distribution $\bar{P} = cP_1(\cdot) + (1 - c)P_2(\cdot)$.

Remarks. (C1) is taken from Chung and Romano (2013). (C2) and (C3) are found in Chernozhukov, Chetverikov and Kato (2017) to obtain a uniform bound over probabilities for hyperrectangles (see Proposition 2.1 therein); assumption (C3′) corresponds to those conditions in their Proposition 3.1, which concerns a uniform bound for probabilities over simple convex sets. Note that the latter

case requires a stricter rate on how large p can be relative to n. For simplicity, c_1 and c_2 are taken as finite here. It is possible to allow for infinite c_1 and c_2, but then a compromise must be made on how large $\ln p$ can be relative to n; refer to equation (9) of Chernozhukov, Chetverikov and Kato (2017) for an explicit expression that relates these two cases. (C4) is necessary for the anti-concentration inequality; see, for example, Proposition 4. Conditions (4.1) and (4.3) are adopted in Cai, Liu and Xia (2014) to derive the asymptotic power of the data-driven statistics, including $H^{\infty}(\cdot)$ of (2.2) and $S^{\infty}(\cdot)$ of (2.10) for two Gaussian populations. Kosorok and Ma (2007) give sufficient conditions for (4.3) to hold, one of which is that $\ln(p) = o(n^{\alpha})$, with $\alpha \in (0, 1/3]$.

5. Simulation Study

We choose to exclude those permutation tests based on Hotelling's T^2-type statistics of (2.2) from our numerical studies, owing to the heavy computational burden. The method of Chung and Romano (2016) is also excluded for the same reason. Instead, we focus on permutation tests based on statistics calculated for pseudo samples generated using the binning procedures $S_1^{\gamma}(\cdot)$ and $S_0^{\gamma}(\cdot)$ in (2.11) and (2.10), respectively. Other methods included in our comparison studies are those of Chen and Qin (2010) (CQ), Cai, Liu and Xia (2014)(CAI), Xu et al. (2016) (XLWP), and Xue and Yao (2018) (XY). The R package *"highmean"* is used for computations related to CQ, CAI, XLWP, and XY. Note that CQ uses only the L_2-norm, and CAI uses only the L_{∞}-norm. For signal identification, our method based on S_1^{∞} is also compared with that of Benjamini and Yekutieli (2001).

The sample sizes range from relatively small ($m = 75$, $n = 50$) to medium ($m = 300$, $n = 200$) to large ($m = 600$, $n = 400$); for the dimensionality, $p =$ 10, 100, or 1,000. Because it is computationally infeasible to evaluate all possible permutations, random permutations are usually used in practice, as first proposed by Dwass (1957). In our case, the permutation distribution is evaluated based on 2,500 (random) permutations. In addition, the empirical sizes of the tests are calculated based on 10,000 replications, and the empirical powers of each test is based on 2,000 replications.

The simulated data are generated according to the following model:

$$X_i = (x_{i,1}, \ldots, x_{ip})^{\top} + \mu^X, \text{ and } Y_j = (y_{i,1}, \ldots, y_{ip})^{\top} + \mu^Y, \tag{5.1}$$

where μ^X and μ^Y are two constant vectors and, for any given $i = 1, \ldots, m$, and $j = 1, \ldots, n$, $\{x_{i,k}, k = 1, 2, \ldots\}$ and $\{y_{j,k}, k = 1, 2, \ldots\}$ are stationary times series

such that

$$x_{i,k+1} = a_i x_{i,k} + \xi_k, \quad y_{j,k+1} = b_j y_{j,k} + \eta_k, \quad k = 1, 2, \ldots, \tag{5.2}$$

where ξ_k, η_k are independent random errors, and $\{a_i\}_{i=1}^m$, $\{b_j\}_{j=1}^n$ are hyper-parameters either fixed or random. This is implemented independently for all $i = 1, \ldots, m,$ and $j = 1, \ldots, n$. Using different specifications for a_i, b_j and ξ_k, η_k, we derive the following three models:

Model 1. a_i, b_k, for $i = 1, \ldots, m,$ and $k = 1, \ldots, n$ are i.i.d., following a uniform distribution on [0, 0.95]; $\xi_k \overset{i.i.d.}{\sim} N(0, 1)$ and $\eta_k \overset{i.i.d.}{\sim} N(0, 4)$. In this model, the X_i are distinctly distributed, as are the Y_i. However, the elements in both still have the same variance.

Model 2. The same as Model 1, but the even-indexed elements of X_i and Y_i are multiplied by two. Thus, elements in X_i and Y_i have different variances.

Model 3. $a_i \equiv -0.2, b_j \equiv 0.7$, and $\xi_k \sim t(3)$, $\eta_k \sim 2t(3)$, where $t(3)$ is the t-distribution. Thus, the generated data are heavy-tailed.

In the study on empirical sizes, $\mu^X = \mu^Y = 0$; when comparing the empirical power of various tests, we keep $\mu^Y = 0$, and consider two designs for $\mu^X = (\mu_1^X, \ldots, \mu_p^X)^\top$.

(i) Dense alternatives: $\mu_1^X, \ldots, \mu_p^X \overset{i.i.d.}{\sim}$ uniform $[0, c_{n,p}]$, with $c_{n,p} = s/(p^{0.25} \times \min(m, n)^{0.5})$, and $s = 6, 9, 11, 14$, which specifies the overall signal-to-noise ratio.

(ii) Sparse alternatives: with $s = 7, 8, 9, 10$, randomly select $0.2 \times p^{0.5}$ elements from $\{\mu_1^X, \ldots, \mu_p^X\}$, and assign to them the value $c_{n,p} = s/\min(m, n)^{0.5}$; the unselected entries remain zero.

Note that the strength of the signals varies with the sample sizes and the dimension; we adopt such a design in order to evaluate how the empirical power of the various tests is affected by different sample sizes and dimensions.

The empirical sizes of the tests are summarized in Table 1 (significance level 1%). For the two columns under the label $S_0^\gamma(\cdot)$, L_1 corresponds to $\gamma = 1$, and L_∞ corresponds to $\gamma = \infty$. The same format applies to the columns under XLWP and $S_1^\gamma(\cdot)$. In both tables, the numbers in small bold font show empirical sizes that deviate from the nominal level by more than 20%. First, the permutation tests based on $S_1^\gamma(\cdot)$ and $S_0^\gamma(\cdot)$ control the type-I error better than all other methods for nearly all models, and especially so when the sample size is small ($n = 50$

Table 1. Empirical sizes (%) of different methods (nominal size = 1%)

			CQ	XLWP		XY	S_0^γ		S_1^γ	
model	n	p	L_2	L_2	L_∞	L_∞	L_1	L_∞	L_1	L_∞
1	50	10	**2.09**	**3.10**	**0.60**	0.91	0.94	0.95	0.97	0.98
		100	1.20	**1.64**	**0.67**	**0.63**	1.01	0.82	1.01	0.98
		1,000	1.04	**1.68**	0.81	**0.52**	**1.22**	0.96	**1.21**	0.95
	200	10	**2.02**	**2.91**	**0.60**	1.07	1.05	0.99	1.07	0.99
		100	**1.23**	**1.55**	**0.60**	0.86	0.94	0.96	0.98	0.84
		1,000	0.98	1.12	**0.52**	0.82	1.11	1.04	1.08	0.97
	400	10	**1.79**	**3.59**	**0.55**	**0.72**	0.95	**0.76**	0.92	**0.76**
		100	1.19	**2.19**	**0.62**	0.97	0.94	0.96	0.91	1.03
		1,000	1.10	**1.28**	0.84	0.91	1.11	0.92	1.08	0.98
2	50	10	**2.42**	**4.63**	0.80	0.90	1.03	0.86	1.04	0.93
		100	**1.47**	**3.30**	0.98	0.83	1.13	0.99	**1.32**	**1.30**
		1,000	1.17	**5.34**	**1.43**	**0.62**	**1.27**	0.84	**1.68**	**1.39**
	200	10	**2.46**	**3.05**	**0.54**	1.06	0.87	0.95	1.06	1.08
		100	**1.46**	**2.31**	**0.64**	0.95	1.04	0.96	0.99	1.02
		1,000	1.10	**1.21**	1.09	1.03	1.09	1.17	1.05	1.00
	400	10	**2.10**	**3.63**	**0.55**	**0.73**	0.86	**0.72**	0.87	**0.71**
		100	**1.43**	**2.22**	**0.62**	0.88	0.98	0.86	1.01	0.96
		1,000	1.12	**1.26**	0.84	1.05	1.18	1.01	1.12	1.07
3	50	10	**2.55**	**3.18**	**0.65**	**0.76**	1.08	1.02	1.18	1.18
		100	**1.54**	**1.84**	1.12	**0.19**	1.20	1.08	**1.51**	**1.35**
		1,000	0.97	**3.56**	**2.14**	**0.01**	1.02	1.03	**1.55**	**1.67**
	200	10	**2.22**	**3.26**	**0.64**	0.80	0.95	1.07	0.88	0.94
		100	**1.46**	**2.20**	**0.74**	**0.31**	1.04	0.98	1.07	1.01
		1,000	1.04	1.16	1.09	**0.01**	1.00	0.99	**1.25**	**1.23**
	400	10	**2.38**	**3.12**	**0.49**	0.94	1.01	1.12	1.01	1.06
		100	**1.42**	**2.36**	**0.52**	**0.56**	1.04	1.12	0.95	1.02
		1,000	0.87	1.03	**0.63**	**0.00**	0.91	0.87	1.01	1.18

* CAI and L_∞ of XLWP are almost identical, and thus are not reported. Values that deviate more than 20% from the nominal level are highlighted in small bold font.

or $n = 75$). Furthermore, we observe that the performance of $S_1^\gamma(\cdot)$ is slightly hampered by the low efficiency in the variance estimation when p is large and n is small; this is consistent with the remarks after Theorem 3.

On the other hand, although CQ controls the type-I error quite well at the 5% significance level (not reported here), it does so less well when the nominal level is at 1%, unless the sample size is sufficiently large; see Table 1. As p increases to 1,000, the performance of CQ improves, which is consistent with the fact that its asymptotic (null) distribution is derived when $p \to \infty$. The performance of

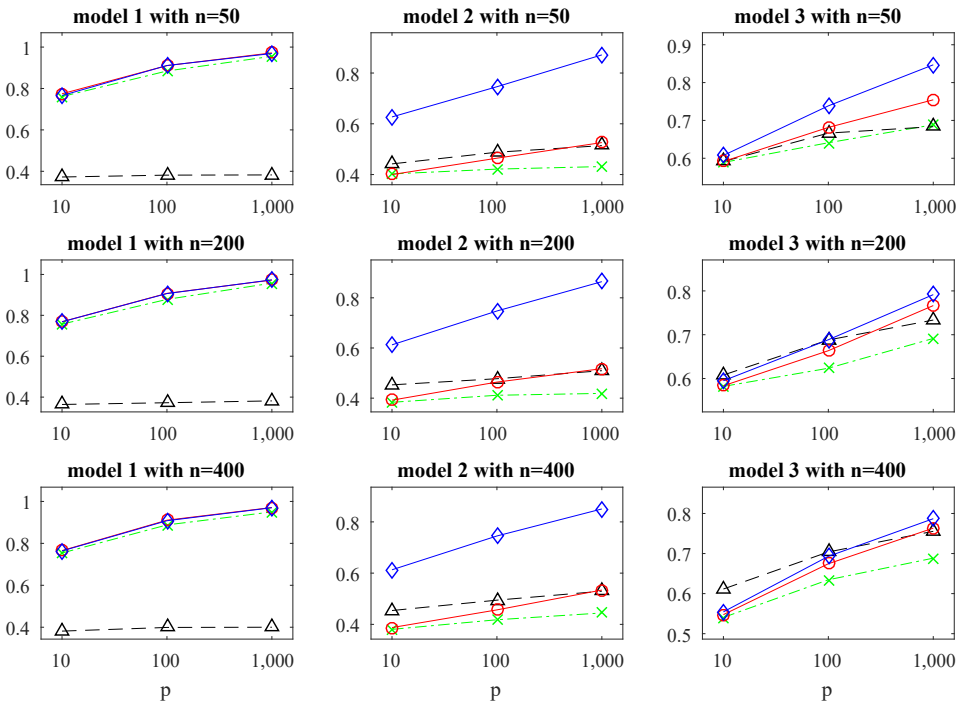

Figure 1. Simulation results with dense signals. In each panel, the dash-dot line with a cross represents CQ, the dashed line with a triangle represents XLWP with the L_2-norm, the solid line with a circle represents S_0^1, and the solid line with a diamond represents S_1^1.

XLWP with the L_2-norm is similar to that of CQ, while the performance of XLWP with the L_∞-norm is mostly too conservative. Even though XY can produce a reasonable type-I error, it does not fare well with heavy-tailed distributions (Model 3) and moderate dimensions. In addition, one can observe from Table 1 that the type-I error of XLWP with the L_2-norm tends to be inflated when p is small or moderate, for example, 10 or 100. Thus, it is not surprising that XLWP possesses a higher empirical power than other methods do in these settings, as seen in Figures 1 and 2 for Model 2.

Figure 1 shows the statistical power of the various tests against dense alternatives, with p varying from 10 to 1,000 at a significance level of 5%. We draw the following conclusions. The performance of XLWP differs dramatically across the models. Specifically, it has decent power with Models 2 and 3, but has very low power for Model 1. Recall that XLWP incurs excessive type-I errors when the dimension p is small (10 or 100). Therefore, we should be cautious with the high power of XLWP with Models 2 and 3, because this very likely comes at the

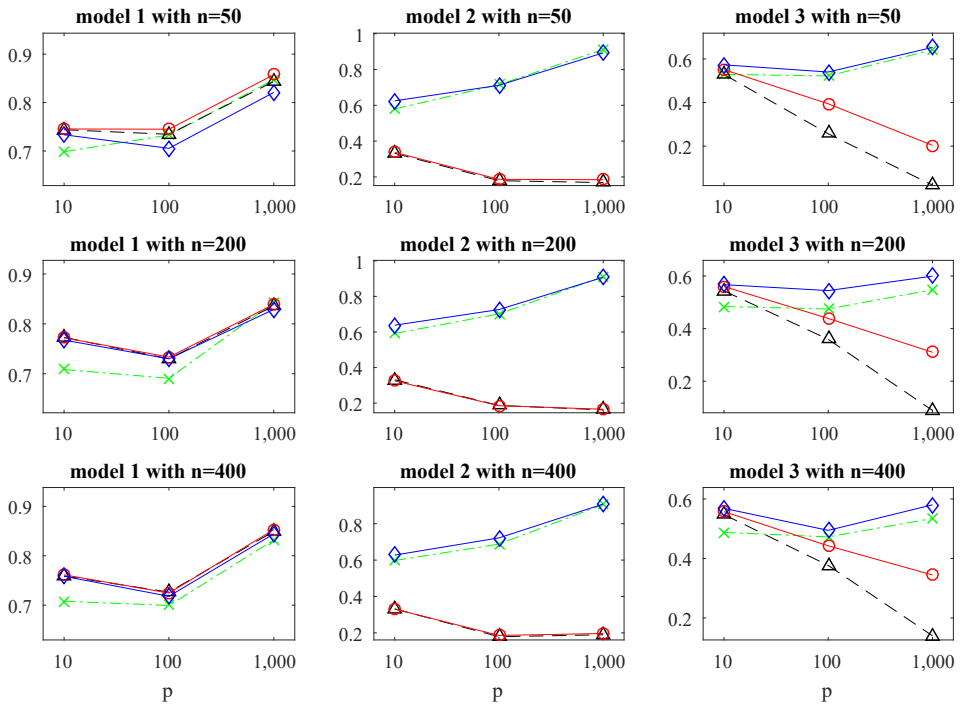

Figure 2. Simulation results with sparse signals. In each panel, the dash-dot line with a cross represents XLWP with L_∞-norm (or CLX), the dashed line with a triangle represents XY, the solid line with a circle represents S_0^∞, and the solid line with a diamond represents S_1^∞.

price of an inflated type-I error. In contrast, the permutation test based on $S_1^1(\cdot)$ is always among the best performers across all of the models.

Figure 2 depicts the changes in power for all four tests against sparse alternatives, for which the L_∞-norm is expected to fare better. Because the power of CAI and XLWP with the L_∞-norm are very similar in all settings considered, we only report those for XLWP. For Models 1 and 2, the methods perform similarly, except for XY, which is significantly worse than the other three when p is less than 100. Note that the high statistical power of XLWP is the consequence of the aforementioned unduly high type-I error. Similarly to $S_1^\gamma(\cdot)$, CAI and XLWP take into account the possible differences across the variances. However, surprisingly, their performance for Model 3 seems to contradict the conclusions drawn about their theoretical properties, especially when compared with $S_1^\gamma(\cdot)$. CAI and XLWP also suffer from low power for Model 3, possibly owing to the difficulty of estimating the covariance matrices for heavy-tailed data. For Model 3, XY and S_1^∞ substantially outperform XLWP and S_0^∞ in terms of power. However, XY

Table 2. The average of true discoveries (FDR in parentheses) based on 10,000 replications at a significance level of 5%.

			$\#I_0 = 0$		$\#I_0 = 8$	
model	p	n	BY	S_1^∞	BY	S_1^∞
1 ($\delta = 5$)	100	200	–(0.0097)	–(0.0502)	0.8760(0.0098)	1.2690(0.0322)
		500	–(0.0105)	–(0.0486)	0.8710(0.0074)	1.2610(0.0330)
		1,000	–(0.0104)	–(0.0486)	0.8700(0.0086)	1.2220(0.0232)
	10,000	200	–(0.0056)	–(0.0544)	0.8070(0.0082)	1.2900(0.0320)
		500	–(0.0054)	–(0.0530)	0.8020(0.0080)	1.3170(0.0252)
		1,000	–(0.0050)	–(0.0494)	0.7810(0.0024)	1.2570(0.0239)
3 ($\delta = 10$)	100	200	–(0.0079)	–(0.0495)	1.6180(0.0081)	2.0770(0.0323)
		500	–(0.0076)	–(0.0503)	1.3880(0.0097)	1.8670(0.0311)
		10,000	–(0.0079)	–(0.0511)	1.3880(0.0088)	1.8540(0.0228)
	10,000	200	–(0.0039)	–(0.0596)	1.7910(0.0049)	2.5190(0.0205)
		500	–(0.0044)	–(0.0520)	1.5750(0.0028)	2.2860(0.0214)
		1,000	–(0.0042)	–(0.0505)	1.5380(0.0033)	2.2620(0.0233)

– no true signals

achieves the same statistical power as that of S_1^∞, at the expense of an inflated type-I error.

Overall, the permutation tests based on $S_0^\gamma(\cdot)$ and $S_1^\gamma(\cdot)$, with $\gamma = 1$ against dense alternatives and $\gamma = \infty$ against sparse alternatives, deliver better results than the other methods in terms of both empirical size and power. Between $S_0^\gamma(\cdot)$ and $S_1^\gamma(\cdot)$, the former has better control over the type-I error, especially when the sample size is small, but the latter usually achieves higher power.

Next, we evaluate the performance of the permutation tests based on S_1^∞ in terms of signal identification, as described in Section 3. If the set of signals is empty, the empirical false discovery rate (FDR) should be no more than a pre-set value, α (set to 5%, in this case). Otherwise, fix $\#I_0 = 8$ with the exact locations of the eight signals distributed randomly among $\{1, \ldots, p\}$, with the strength of the signals given by

$$\frac{\delta \times \ln(\log(p))}{\sqrt{n}},$$

which changes with n and p. We compare our method S_1^∞ with that of Benjamini and Yekutieli (2001), denoted as BY. When $\#I_0 = 0$ (the first two columns of Table 2), it is obvious that our method S_1^∞ has very good control over the FDR, whereas BY tends to be overly conservative. On the other hand, when $\#I_0 = 8$ (the last two columns of Table 2), our method S_1^∞ is able to identify more true signals than BY does.

6. Analysis of WTCCC Data Set

Genome-wide association studies (GWAS) are widely used to identify risk genetic variants by genotyping millions of single nucleotide polymorphisms (SNPs) in large cohorts. A traditional GWAS analysis uses a single-variant analysis that does take into account the LD structure among the SNPs, but suffers from a heavy burden of multiple testing. Thus, the results from such analyses are usually conservative. Because the proposed method is based on S_1^∞, it can explicitly control the FDR. We applied S_1^∞ to seven traits from the WTCCC, including bipolar disorder (BPD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type-1 diabetes (T1D), and type-2 diabetes (T2D) (Burton et al. (2007)). We performed strict quality control on the samples from WTCCC using PLINK (Purcell et al. (2007)) and GCTA (Yang et al. (2011)). First, we removed individuals with missing genotypes higher than 0.02. For each trait case and two shared control data sets, we removed SNPs with minor allele frequencies less than 0.05 and SNPs with a missing rate larger than 0.01. Then, we combined cases with controls for each trait, and removed SNPs with p-values less than 0.001 for the Hardy–Weinberg equilibrium test. Pairs of subjects with an estimated relatedness greater than 0.025 were identified, and one subject from these pairs was removed. After the quality control, we have 1,959 cases and 2,992 controls over 308,093 SNPs for CAD, 1,970 cases and 2,992 controls over 307,741 SNPs for CD, 1,994 cases and 2,992 controls over 307,357 SNPs for T1D, and 1,969 cases and 2,992 controls over 305,394 SNPs for T2D. We applied the permutation test with S_1^∞ to the data; the resulting Manhattan plots are shown in Figure 3. The analysis for each disease takes around 16 minutes on a Windows console with a 2.30 GHz Intel Xeron CPU E5-2697.

For the significance level 1%, we summarize our findings as follows. For CAD, S_1^∞ identified 15 SNPs, all from genes *AL359922.1* and *CDKN2B-AS1* within the band 9p21.3. These two genes have previously been reported to be associated with CAD (van der Harst and Verweij (2018); Lee et al. (2013)). For CD, S_1^∞ identified 39 SNPs, of which 21 SNPs are within six gene regions, where all six genes have been reported to be associated with CD in previous studies (Julià et al. (2013); de Lange et al. (2017); Liu et al. (2015)). For T1D, S_1^∞ identified 369 SNPs, and 173 SNPs were within 83 genes, among which 23 genes have been previously reported to be associated with T1D, including *ERBB3*, *CLEC16A*, and *DDR1* (Plagnol et al. (2011); Hakonarson et al. (2007); Tomer et al. (2015)). For T2D, S_1^∞ identified 13 SNPs within two gene regions, where both genes have previously been reported, that is, *TCF7L2* and *FTO* (Hackinger et al. (2018);

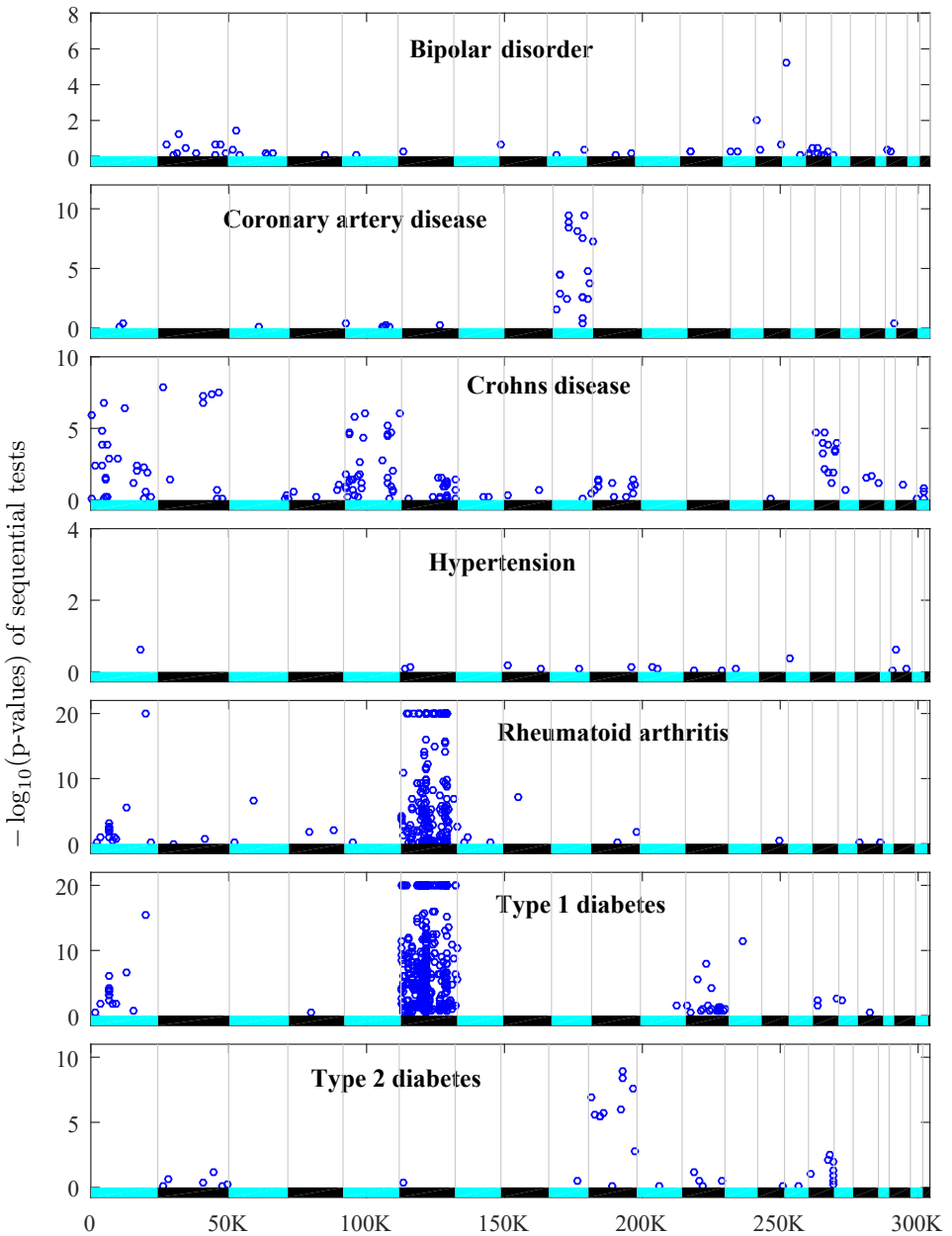

Figure 3. For each of seven diseases, -log10 of the test p-value for quality-control-positive SNPs (values bigger than 20 are censored at 20) are plotted against the position of SNPs arranged in according to the chromosomes in black and gray.

Tabassum et al. (2013)).

Interestingly, in our analysis of the seven diseases using the WTCCC data, we identified many "new" SNPs not reported in the original study of the Burton et al. (2007), but that were detected in later studies. These SNPs and their corresponding studies are listed in Table S1 in the Supplementary Material. Statistically, it is more interesting to note that these other are based on either much larger cohorts or other populations. This indicates clearly the efficiency of our method in identifying weak signals (SNPs) associated with a disease.

Supplementary Material

The online Supplementary Material includes a brief introduction to the high-dimensional central limit theorem and some propositions used in the proofs. It also contains technical proofs for Theorems 1–4 as well as additional results for the analysis of the WTCCC data set.

Acknowledgments

We are grateful to the assistant editor and three referees for their meticulous review, valuable comments, and constructive suggestions. EK was supported by a grant from the National Natural Science Foundation of China (Nos. 11771066 and 11931014). YX was partially supported by the National Natural Science Foundation of China (Nos. 72033002 and 11931014), and Fundamental Research Funds for the Central Universities of China (ZYGXJ097), and AcRF grant R-155-000-220-114 of the National University of Singapore. JL was partially supported by AcRF Tier 2 [MOE2016-T2-2-029, MOE2018-T2-1-046, MOE2018-T2-2-006] from the Ministry of Education, Singapore, and block fund R-913-200-098-263 from the Duke-NUS Medical School.

References

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica* **6**, 311–329.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* **57**, 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165–1188.

Bickel, P. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

Burton, P. R., Calyton, D. G., Cardon, L. and Craddock, N. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**,

661–678.

Cai, T. and Liu, W. D. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106**, 672–684.

Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **76**, 349–372.

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics* **73**, 1300–1310.

Chen, S. and Qin, Y. (2010). A two sample test for high dimensional data with applications to gene-set testing. *The Annals of Statistics* **38**, 808–835.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* **162**, 47–70.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability* **45**, 2309–2352.

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics* **41**, 484–507.

Chung, E. and Romano, J. P. (2016). Multivariate and multiple permutation tests. *Journal of Econometrics* **193**, 76–91.

de Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A. et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics* **49**, 256.

Dizier, M. H., Demenais, F. and Mathieu, F. (2017). Gain of power of the general regression model compared to Cochran-Armitage Trend tests: Simulation study and application to bipolar disorder. *BMC Genetics* **18**, 24.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *The Annals of Mathematical Statistics* **28**, 181–187.

Ernest, N. (2004). Permutation methods: A basis for exact inference. *Statistical Science* **19**, 676–685.

Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer-Verlag, New York.

Hackinger, S., Prins, B., Mamakou, V., Zengini, E., Brčić, L., Serafetinidis, I. et al. (2018). Evidence for genetic contribution to the increased risk of type 2 diabetes in schizophrenia. *Translational Psychiatry* **8**, 252.

Hakonarson, H., Grant, S. F., Bradfield, J. P., Marchand, L., Kim, C. E., Glessner, J. T. et al. (2007). A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448**, 591–594.

Hoeffding, W. (1952). The large-sample power of tests based on permutations of observations. *The Annals of Mathematical Statistics* **23**, 169–192.

Jin, J. and Cai, T. (2007). Estimating the null and the proportion of non-null effects in large-scale multiple comparisons. *Journal of the American Statistical Association* **102**, 495-506.

Julià, A., Domènech, E., Ricart, E., Tortosa, R., García-Sánchez, V., Gisbert, J. P. et al. (2013). A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. *Gut* **62**, 1440–1445.

Kosorok, M. and Ma, S. (2007). Marginal asymptotics for the "large p, small n" paradigm: With applications to microarray data. *The Annals of Statistics* **35**, 1456–1486.

Lee, J. Y., Lee, B. S., Shin, D. J., Park, K. W., Shin, Y. A., Kim, K. J. et al. (2013). A genome-wide association study of a coronary artery disease risk variant. *Journal of Human Genetics* **58**, 120–126.

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A. et al. (2018). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47**, 979.

Plagnol, V., Howson, J. M., Smyth, D. J., Walker, N., Hafler, J. P., Wallace, C. et al. (2011). Genome-wide association analysis of autoantibody positivity in type 1 diabetes cases. *PLoS Genetics* **7**, e1002216.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575.

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association* **85**, 686–692.

Tabassum, R., Chauhan, G., Dwivedi, O. P., Mahajan, A., Jaiswal, A., Kaur, I. et al. (2013). Genome-wide association study for type 2 diabetes in Indians identifies a new susceptibility locus at 2q21. *Diabetes* **62**, 977–986.

Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians.* Springer-Verlag, Berlin Heidelberg.

Tomer, Y., Dolan, L. M., Kahaly, G., Divers, J., D'Agostino Jr, R. B., Imperatore, G. et al. (2015). Genome wide identification of new genes and pathways in patients with both autoimmune thyroiditis and type 1 diabetes. *Journal of Autoimmunity* **60**, 32–39.

van der Harst, P. and Verweij, N. (2018). Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circulation Research* **122**, 433–443.

van Hulzen, K. J., Scholz, C. J., Franke, B., Ripke, S., Klein, M., McQuillin, A. et al. (2017). Genetic overlap between attention-deficit/hyperactivity disorder and bipolar disorder: Evidence from genome-wide association study meta-analysis. *Biological Psychiatry* **82**, 634–641.

Xu, G., Lin, L., Wei, P. and Pan, W. (2016). An adaptive two-sample test for high dimensional means. *Biometrika* **103**, 609–624.

Xue, K. and Yao, F. (2018). Distribution and correlation free two-sample test high dimensional means. *The Annals of Statistics* **48**, 1304–1328.

Yang, J., Lee, S. H., Goddard, M. E. and Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76–82.

Efang Kong

University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China.

E-mail: kongefang@uestc.edu.cn

Lengyang Wang

National University of Singapore, 119077 Singapore.

E-mail: e0321268@u.nus.edu

Yingcun Xia

National University of Singapore, 119077 Singapore.

E-mail: staxyc@nus.edu.sg

Jin Liu
National University of Singapore, 119077 Singapore.
E-mail: jin.liu@duke-nus.edu.sg

(Received November 2019; accepted May 2020)