Taylor & Francis
Taylor & Francis Group

Check for updates

A Geometric Perspective on the Power of Principal Component Association Tests in Multiple Phenotype Studies

Zhonghua Liu[a] and Xihong Lin[b]

[a]Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong, China; [b]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

ABSTRACT

Joint analysis of multiple phenotypes can increase statistical power in genetic association studies. Principal component analysis, as a popular dimension reduction method, especially when the number of phenotypes is high dimensional, has been proposed to analyze multiple correlated phenotypes. It has been empirically observed that the first PC, which summarizes the largest amount of variance, can be less powerful than higher-order PCs and other commonly used methods in detecting genetic association signals. In this article, we investigate the properties of PCA-based multiple phenotype analysis from a geometric perspective by introducing a novel concept called principal angle. A particular PC is powerful if its principal angle is 0° and is powerless if its principal angle is 90°. Without prior knowledge about the true principal angle, each PC can be powerless. We propose linear, nonlinear, and data-adaptive omnibus tests by combining PCs. We demonstrate that the Wald test is a special quadratic PC-based test. We show that the omnibus PC test is robust and powerful in a wide range of scenarios. We study the properties of the proposed methods using power analysis and eigen-analysis. The subtle differences and close connections between these combined PC methods are illustrated graphically in terms of their rejection boundaries. Our proposed tests have convex acceptance regions and hence are admissible. The p-values for the proposed tests can be efficiently calculated analytically and the proposed tests have been implemented in a publicly available R package *MPAT*. We conduct simulation studies in both low- and high-dimensional settings with various signal vectors and correlation structures. We apply the proposed tests to the joint analysis of metabolic syndrome-related phenotypes with datasets collected from four international consortia to demonstrate the effectiveness of the proposed combined PC testing procedures. Supplementary materials for this article, including a standardized description of the materials available for reproducing the work, are available as an online supplement.

1. Introduction

In the past decade, Genome-Wide Association Studies (GWASs) have identified thousands of genetic variants associated with hundreds of human complex traits and diseases (Welter et al. 2014), as reported in the National Human Genome Research Institute and European Bioinformatics Institute's (NHGRI-EBI) catalog. By using the open-access NHGRI-EBI catalog, Sivakumaran et al. (2011) found abundant evidence of pleiotropy: 233 (16.9%) genes and 77 (4.6%) single nucleotide polymorphisms (SNPs) show pleiotropic effects, and the numbers are still growing over time. As detailed phenotype data from epidemiological studies, electronic health records (EHR), genome-wide omics profiling, and real-time mobile health devices are becoming rapidly available, there is an increasing interest in identifying cross-phenotype associations (Solovieff et al. 2013; Bush, Oetjens, and Crawford 2016), which hold great potentials for novel drug target discovery, drug repurposing, and informing precision medicine (Collins and Varmus 2015).

Our work is motivated by studying the genetic basis of metabolic syndrome (MetS) (Brown and Walker 2016). A set of clinical phenotypes are involved in the disease process of MetS.

Single-trait GWAS studies have been conducted to identify susceptible SNPs associated with each of those MetS-related phenotypes. The following four consortia studied the genetic architecture of the MetS traits. The International Consortium for Blood Pressure (ICBP) is an international effort to investigate blood-pressure genetics. It conducted a GWAS of systolic blood pressures (SBP) of 200,000 individuals of European descent (ICBP et al. 2011). The Global Lipids Genetics Consortium (GLGC) performed individual trait GWAS analysis of high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and triglycerides (TG; Teslovich et al. 2010). It examined the SNP–lipid associations in 188,578 European-ancestry individuals (Willer et al. 2013). The Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) represents a collaborative effort to combine data from multiple GWASs to identify genetic loci that impact glycemic and metabolic traits. The MAGIC study performed meta-analysis of 29 GWASs of Fasting Glucose (FG) from 58,074 nondiabetic participants, and 26 GWASs of Fasting Insulin (FI) from 51,750 nondiabetic participants (Manning et al. 2012), with both analyses adjusting for Body Mass Index (BMI). The Genetic

Investigation of ANthropometric Traits (GIANT) consortium investigates the genetic underpinning that modulates human body size and shape. It performed a GWAS analysis of Body Mass Index (BMI) using 339,224 individuals (Locke et al. 2015), and a BMI-adjusted GWAS analysis waist-hip-ratio (WHR) using 224,459 individuals of European ancestry (Shungin et al. 2015). Although those aforementioned studies identified the SNPs associated with each of the eight phenotypes, the single-trait analysis paradigm is likely to suffer from potential power loss for detecting the genetic variants associated with MetS by ignoring the fact that these clinical phenotypes of MetS are related and might share a common genetic basis.

It has been shown that joint analysis of multiple phenotypes together can increase statistical power to detect genetic variants. Numerous methods have been proposed for multiple phenotype analysis, see Solovieff et al. (2013) for a review. Examples include multivariate regression-based methods, which improve power under specific parametric assumptions, such as homogeneous effects across phenotypes, but are subject to power loss when these assumptions are violated (Schifano et al. 2013; Zhou and Stephens 2014); the p-value correction method TATES (van der Sluis et al. 2013), which accounts for between-phenotype correlation, has a good power in the presence of a very few association signals and can lose power otherwise. Furthermore, this method is subject to inflated Type I error rate by 12% (He, Avery, and Lin 2013). Zhu et al. (2015) proposed two tests, one for detecting homogeneous effects and another for detecting heterogenous effects based on a truncated test statistic. These tests were found to have good performance when their corresponding assumptions hold. In practice, researchers usually have little prior knowledge about which assumption holds, and hence it might be challenging to decide which test to use. Moreover, the p-value of the truncated test for detecting heterogenous effects could not be calculated analytically and requires Monte Carlo simulations, which are computationally expensive for genome-wide analysis of multiple phenotypes. Huang and Lin (2013) and Liu and Lin (2018) proposed variance component tests for multiple phenotypes. We will show that this variance component test is a special quadratic combination of PC test in this article.

Principal component analysis (PCA), as a popular dimension reduction technique, especially when the number of phenotypes is not small, is an appealing approach that transforms correlated phenotypes into orthogonal composite scores (Aschard et al. 2014). Although it has been empirically found that principal components (PCs) that explain a small amount of the total variance of the multiple phenotypes can be as powerful or even more powerful than the PCs that explain a large amount of the total variance of the multiple phenotypes (Aschard et al. 2014), however there is no theoretical explanation for this counter-intuitive phenomenon. It is also unclear which PCs should be used to achieve the best power for genetic association testing.

It is well known that the uniformly most powerful (UMP) test does not exist for composite hypothesis testing. The classical Wald test can lose substantial power when the first PC captures all the signals and also explains a large amount of the total variance. The canonical correlation analysis aims to find a best linear combination of the multiple phenotypes (Ferreira and Purcell 2009), and thus can perform poorly when the relationship between a genetic variant and multiple phenotypes is not linear. Therefore, there is a pressing need to develop effective powerful testing methods for multiple phenotype association studies.

Since multiple PCs are likely to contain association evidence, it could be advantageous to combine PCs together to achieve better power. There are several challenges on how to effectively combine association evidence across multiple PCs. First, the underlying genetic effects are unknown and can be heterogenous, that is, a genetic variant can have positive, negative, or null effects on different phenotypes. Second, the correlation structure among multiple phenotypes can be arbitrary, that is, phenotypes can be positively or negatively correlated with varying correlation strength. However, little is known in the literature about the effect of the correlation structure on the power of the PC-based tests. Furthermore, it is more challenging to effectively combine PCs in high-dimensional settings, such as in gene expression studies, because it is more complex to understand the interplay between the high-dimensional signal vectors and the between-phenotype correlation structures. Therefore, it is of significant interest to develop more powerful testing procedures by effectively combining PCs and taking into account the between-phenotype correlation structure, the effect size, and the direction of the genetic effects, in both low- and high-dimensional settings.

In this article, we aim to address these problems by developing robust and powerful PC-based methods for testing for genetic association with multiple phenotypes, as well as studying the effects of the between-phenotype correlation structures on the power of the proposed PC-based tests. This article makes the following contributions. First, we introduce a novel geometric concept called principal angle and show that a particular PC can be powerless if its principal angle is 90° and can be as powerful as the Oracle test if its principal angle is zero. In practical settings, any PC can be powerless if one has no prior knowledge about the true principal angles.

Second, we propose several data-driven methods to combine PCs to boost the power for testing for the association between a genetic variant and multiple phenotypes. We first propose the minimum PC p-value (PCMinP) and the Fisher's method by combining PC p-values (PCFisher) as testing statistics. We then propose linear and quadratic combinations of PCs weighted by the functions of eigenvalues. Specifically, we show that an inverse-eigenvalue weighted linear combination of PCs (PCLC) can be as powerful as the Oracle test when all the principal angles are equal to each other, but can lose power otherwise. Quadratic combinations of PCs are shown to be more robust than PCLC. We show that the classical Wald test and the recently proposed variance component score test (Huang and Lin 2013; Liu and Lin 2018) are special cases of the quadratic combinations of PCs. These two tests both favor the alternatives under which the last principal angle is zero. As we usually have no prior knowledge about the true signal direction in practice, we propose an omnibus test (PCO) which uses the data-driven method to best combine several linear and nonlinear PC tests together to achieve robust power performance under various alternatives.

Third, we perform eigen-analysis to investigate the effects of the between-phenotype correlation structure on the power performance of the PC-based tests. The subtle differences and

close connections between our proposed tests are compared graphically in terms of their rejection boundaries. Our proposed tests all have convex acceptance regions and hence are admissible (Birnbaum 1954, 1955). Fourth, the p-values of our proposed tests can be calculated analytically in a computationally efficient manner.

The Type I error rates of our proposed tests are shown to be well controlled by simulation studies. The powers of the proposed tests relative to several commonly used methods, such as the Wald test, the TATES method (van der Sluis et al. 2013), are compared using simulations in both low- and high-dimensional settings. The robust power performance of the proposed omnibus test PCO is demonstrated through simulations using a range of signal patterns and correlation structures. Finally, we applied our proposed tests to the aforementioned metabolic syndrome trait GWAS datasets and identified additional new genetic variants that were missed by the original univariate analyses. Those identified new SNPs might play important biological roles in the pathogenesis of MetS and can serve as potential candidates for future functional studies.

The remainder of this article is organized as follows. In Section 2, we describe our PC-based testing procedures and perform power analysis. In Section 3, the omnibus PC-based tests are proposed to improve robustness and power of the PC-based tests. In Section 4, we compare those tests in terms of their rejection boundaries and demonstrate their differences graphically. In Section 5, we perform eigen-analysis to investigate how the between-phenotype correlation structure affects the statistical powers of our proposed tests. In Section 6, we conduct simulation studies to evaluate the performance of our methods in both low- and high-dimensional settings. In Section 7, we apply our tests to the metabolic syndrome trait GWAS datasets. Finally, we conclude with discussions in Section 8.

2. The Principal Component Association Tests

Suppose that there are K correlated phenotypes denoted by $\mathbf{Y} = (Y_1, \ldots, Y_K)^T$. Traditional GWAS studies consist of hundreds of thousands of SNPs across the genome. One analyzes an SNP at a time for each phenotype separately. For a particular SNP, we have K correlated test statistics for testing for the presence of genetic effects, that is, Z-scores $\mathbf{Z} = (Z_1, \ldots, Z_K)^T$ that asymptotically follow a multivariate normal distribution with the covariance matrix $\mathbf{\Sigma}$, which is equal to the correlation matrix of $\mathbf{Y}$ conditional on other covariates included in the univariate analysis under the null (Liu and Lin 2018). In other words, the Z scores have already taken into account the effects of confounders, such as population stratification, and $\mathbf{\Sigma}$ is not the crude covariance of $\mathbf{Y}$ but the residual covariance after regressing $\mathbf{Y}$ on covariates. Although across the whole genome, different genetic variants could have different minor allele frequencies (MAF), however their association test statistics $\mathbf{Z}$ follow the same null distribution. This serves as the basis for consistently estimating $\mathbf{\Sigma}$ using the sample covariance matrix of the $\mathbf{Z}$-statistics across the genome under the null hypothesis (Zhu et al. 2015; Liu and Lin 2018).

For simplicity, we assume $\mathbf{\Sigma}$ is known for the ease of discussions hereafter. For a given dataset of sample size n, univariate analysis for each phenotype can be performed. For a particular genetic variant, we can obtain a K-dimensional vector of summary testing statistics $\mathbf{Z} \sim N(\boldsymbol{\beta}, \mathbf{\Sigma})$, where $\boldsymbol{\beta} \propto \sqrt{n}$ and n is the sample size for calculating $\mathbf{Z}$. We are interested in testing H_0: $\boldsymbol{\beta} = 0$ against H_a: $\boldsymbol{\beta} \neq 0$, where $\boldsymbol{\beta}$ is referred to as the signal vector. We would like to develop robust and powerful tests that are robust to the between-phenotype correlation structures and signal vector patterns, especially when the dimension of phenotypes is not small.

2.1. The Oracle Test for the Fixed Alternative Hypothesis

Under the fixed alternative hypothesis $\boldsymbol{\beta}$, the Uniformly Most Powerful (UMP) test is

$$\text{Oracle} = \boldsymbol{\beta}^T \mathbf{\Sigma}^{-1} \mathbf{Z}, \tag{1}$$

which directly follows from the Neyman–Pearson Lemma (Bittman et al. 2009). One can easily see that the Oracle test is a linear combination of $\mathbf{Z}$ with the coefficients depending on the true $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$. It is natural to view this hypothesis testing problem as a binary classification problem. We observe a vector $\mathbf{Z}$ and need to decide whether $\mathbf{Z}$ is from the null H_0 or the alternative H_a. This classification problem fits into the framework of linear discriminant analysis (LDA). In fact, this Oracle test can be viewed as the Fisher LDA (Fisher 1936), which is the Bayes optimal classifier (Bickel and Levina 2004) and provides the highest sensitivity uniformly at any given specificity (Su and Liu 1993). In practice, we do not know the true $\boldsymbol{\beta}$ and therefore we cannot perform this Oracle test. Nonetheless, we can use it as an *ideal* benchmark for power comparisons with those implementable tests.

Under the alternative hypothesis H_a: $\boldsymbol{\beta} \neq 0$, both $\boldsymbol{\beta}$ and correlation matrix $\mathbf{\Sigma}$ are unknown. Equation (1) implies that only a "smart" linear combination of individual Z-testing statistics that is as close as possible to the unknown true quantity $\mathbf{\Sigma}^{-1}\boldsymbol{\beta}$, can be as powerful as the Oracle test, but at the potential risk of being powerless if the linear combination is not "smart."

2.2. Single Principal Component Tests for the Composite Hypothesis

Consider the composite hypothesis H_0: $\boldsymbol{\beta} = 0$ versus H_a: $\boldsymbol{\beta} \neq 0$. Using spectral decomposition, we have

$$\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \sum_{k=1}^{K} \lambda_k \mathbf{u}_k \mathbf{u}_k^T,$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K > 0$ of $\mathbf{\Sigma}$, and $\mathbf{U}$ is the normalized orthogonal matrix whose kth column $\mathbf{u}_k$ is the kth eigenvector associated with the kth largest eigenvalue λ_k of $\mathbf{\Sigma}$. We also require that $\mathbf{U}$ is a proper rotation matrix, that is, $\det(\mathbf{U}) = 1$ (Pettofrezzo 1978). The K eigenvectors $\mathbf{u}_k$ ($k = 1, \ldots, K$) form an eigen-basis and hence constitute a new orthogonal coordinate system in which the kth coordinate direction corresponds to the kth principal component PC_k. It is straightforward to show that the distribution of PC_k is

$$\text{PC}_k = \mathbf{u}_k^T \mathbf{Z} \sim N(\mathbf{u}_k^T \boldsymbol{\beta}, \lambda_k), 1 \leq k \leq K.$$

As $||\mathbf{u}_k||^2 = 1$, the noncentrality parameter (*ncp*) of PC_k under the alternative is

$$ncp_k = \frac{(\mathbf{u}_k^T \boldsymbol{\beta})^2}{\lambda_k} = \frac{||\boldsymbol{\beta}||^2 \{\cos(\theta_k)\}^2}{\lambda_k},$$

where $\theta_k \in [0, 180°]$ is the angle between the signal vector $\boldsymbol{\beta}$ and the eigenvector $\mathbf{u}_k$ and is defined as the kth principal angle (PA), and $||\boldsymbol{\beta}|| = \sqrt{\sum_{k=1}^{K} \beta_k^2}$ which is defined as the overall signal magnitude. An underlying constraint for the principal angles is that $\sum_{k=1}^{K} \cos^2(\theta_k) = 1$, which will be used for power analysis later. If the kth principal component PC_k is used as a testing statistic for H_0: $\boldsymbol{\beta} = 0$ versus H_a: $\boldsymbol{\beta} \neq 0$, then its theoretical power at significance level α is

$$\text{Power} = \Phi(Z_{\frac{\alpha}{2}} + \sqrt{ncp_k}) + \Phi(Z_{\frac{\alpha}{2}} - \sqrt{ncp_k}),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function, and $Z_{\frac{\alpha}{2}}$ is its $\alpha/2$ percentile. If $\theta_k = 0$, then PC_k is as powerful as the Oracle test; however if $\theta_k = 90°$, then PC_k is powerless. This geometric perspective clearly explains why using any particular PC could be powerless to detect association signals in the situations where its principal angle is $90°$.

The principal angle of a PC measures the degree of similarity between the direction of the PC of multiple phenotypes and the direction of the genetic effect vector of an SNP on multiple phenotypes. When the principal angle of a PC is zero, it means that the direction of the PC completely aligns with the genetic effect direction, and is thus perfect for being used for summarizing multiple phenotypes into a scalar super-phenotype for detecting genotype-phenotype associations. If the principal angle of a PC is 90 degree, it means that the PC contains no information about the genetic effects and is thus not useful for detecting the genotype-multiple phenotype associations.

The power analysis for single PC test serves as the building blocks of the power analysis of combined PC-based tests. We observe that the power of single PC test depends not only on $\boldsymbol{\beta}$ but also $\boldsymbol{\Sigma}$ through its eigenvalues and eigenvectors, which will be investigated by eigen-analysis in Section 5. It should be noted that the PC directions of the Z-scores are often not the same as the PC directions of the original phenotypes $\mathbf{Y}$, as the Z-scores have taken the confounders into account.

2.3. The PCMinP Test

As the signal vector $\boldsymbol{\beta}$ is unknown in practice, one usually has no prior information about the true principal angles and thus cannot decide which PC to use for association testing. Hence, we propose to use the minimum principal component p-value as a testing statistic named PCMinP,

$$\text{PCMinP} = \min_{1 \leq k \leq K} p_k,$$

where p_k is the p-value based on PC_k. In fact, PCMinP is equivalent to using $\sup_{1 \leq k \leq K} |PC_k|/\sqrt{\lambda_k}$ as a test statistic, and hence can be viewed as a nonlinear combination of PCs. Because the K PCs are mutually independent, so the p-value of PCMinP can be easily computed as $p = 1 - (1 - \text{PCMinP})^K$.

Denote $\alpha^* = 1 - (1 - \alpha)^{1/K}$ where α is a prespecified significance level, then the power of PCMinP under the alternative is

$$\text{Power} = 1 - \prod_{k=1}^{K} \left[1 - \left\{ \Phi(Z_{\frac{\alpha^*}{2}} + \sqrt{ncp_k}) \right. \right.$$
$$\left. \left. + \Phi(Z_{\frac{\alpha^*}{2}} - \sqrt{ncp_k}) \right\} \right].$$

Suppose that $||\boldsymbol{\beta}||$ and λ_k are fixed, then the power of PCMinP is maximized when $\theta_K = 0$ and its maximal power is

$$\text{Power}_{\max} = 1 - (1 - \alpha)^{\frac{K-1}{K}} \left\{ 1 - \Phi\left(Z_{\frac{\alpha^*}{2}} + \frac{||\boldsymbol{\beta}||}{\sqrt{\lambda_K}}\right) \right.$$
$$\left. - \Phi\left(Z_{\frac{\alpha^*}{2}} - \frac{||\boldsymbol{\beta}||}{\sqrt{\lambda_K}}\right) \right\}.$$

This implies that PCMinP favors the alternatives under which the last PC captures all the signals. Furthermore, the power of PCMinP goes to 1 as $\lambda_K \to 0$. The power of PCMinP is minimized when $\cos^2(\theta_k) = \lambda_k/K$ ($k = 1, \ldots, K$), and the minimum power is

$$\text{Power}_{\min} = 1 - \left\{ 1 - \Phi\left(Z_{\frac{\alpha^*}{2}} + \frac{||\boldsymbol{\beta}||}{\sqrt{K}}\right) \right.$$
$$\left. - \Phi\left(Z_{\frac{\alpha^*}{2}} - \frac{||\boldsymbol{\beta}||}{\sqrt{K}}\right) \right\}^K.$$

The result follows directly from the inequality of arithmetic and geometric means. This implies that the worst situation for PCMinP is that all the PCs are equally powerful, for example, when multiple phenotypes are independent, for example, $\boldsymbol{\Sigma} = \mathbf{I}$.

2.4. The PCFisher Test

PCMinP aims to pick the most powerful PC direction and discards the other less powerful PCs. Hence, PCMinP does not fully use available information contained in all the PCs. We hereby propose to combine all the K independent principal component p-values using Fisher's method (Fisher 1932) with its null distribution given by

$$\text{PCFisher} = -2 \sum_{k=1}^{K} \log(p_k) \sim \chi_{2K}^2.$$

PCFisher can also be viewed as a nonlinear combination of the PCs,

$$\text{PCFisher} = -2 \sum_{k=1}^{K} \log\{1 - F_{\chi_1^2}(PC_k^2/\lambda_k)\}, \tag{2}$$

where $F_{\chi_1^2}(\cdot)$ represents the chi-squared cumulative distribution function with one degree of freedom. Equation (2) implies that PCFisher allocates larger weights to PCs with smaller eigenvalues. Therefore, PCFisher achieves its maximal power when $\theta_K = 0$ and achieves its minimal power when $\theta_1 = 0$ for fixed $||\boldsymbol{\beta}||$ and λ_k. The Fisher's p-value combination method is asymptotically Bahadur optimal (ABO), in the sense that the p-value of PCFisher converges to zero with the fastest rate under the alternative when the sample size goes to infinity (Bahadur 1967; Littell and Folks 1971, 1973).

2.5. The Test Based on a Linear Combination of PCs

Motivated by the inverse variance weighting method commonly used when combining independent tests (Mosteller, Bush, and

Green 1954; Liptak 1958), we obtain the following linear combination of PCs with each PC weighted by its inverse variance,

$$\text{PCLC} = \sum_{k=1}^{K} \frac{\text{PC}_k}{\lambda_k} \sim N(0, \sum_{k=1}^{K} \lambda_k^{-1}).$$

Under the alternative hypothesis, its noncentrality parameter is

$$ncp = \frac{\left(\sum_{k=1}^{K} \lambda_k^{-1} \mathbf{u_k^T} \boldsymbol{\beta}\right)^2}{\sum_{k=1}^{K} \lambda_k^{-1}} = \frac{||\boldsymbol{\beta}||^2 \left\{\sum_{k=1}^{K} \lambda_k^{-1} \cos(\theta_k)\right\}^2}{\sum_{k=1}^{K} \lambda_k^{-1}}.$$

We now study when the power of PCLC will be maximized and minimized with respect to θ_k, for any fixed $||\boldsymbol{\beta}||$ and λ_k. This can be formulated as the following constrained optimization problem:

$$\max_{\cos(\theta_k)} \quad \sum_{k=1}^{K} \lambda_k^{-1} \cos(\theta_k)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \cos^2(\theta_k) = 1.$$

Using the Lagrange multiplier method, we obtain that the power of PCLC is maximized when the principal angles satisfy the following conditions:

$$\cos^2(\theta_k) = \frac{\lambda_k^{-2}}{\sum_{k=1}^{K} \lambda_k^{-2}}, k = 1, \ldots, K. \tag{3}$$

In fact, we can rewrite PCLC as $\text{PCLC} = (\mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{J})^{\mathbf{T}}\mathbf{Z}$, where $\mathbf{J} = (1, \ldots, 1)^T$. Hence, PCLC will achieve its own maximal power when $\boldsymbol{\beta} \propto \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{J}$, which is equivalent to Equation (3). It can be easily seen that PCLC is powerless when $\boldsymbol{\beta} \perp \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{J}$ where positive and negative genetic effects are canceled out. The PCLC test can be as powerful as the Oracle test when $\boldsymbol{\beta} \propto \mathbf{U}\mathbf{J}$, or equivalently when all the principal angles are equal to each other, that is, $\cos^2(\theta_k) = 1/K (k = 1, \ldots, K)$. In other words, when all the K principal angles are the same, PCLC is more powerful than any other tests to detect such alternatives.

2.6. Quadratic Combination of PCs

PCLC is very sensitive to principal angles and can be powerless as shown theoretically in Section 2.5 and empirically in the simulation setting M5 in Table 3. To overcome this drawback, we propose to combine PCs using the following weighted quadratic function that weights the PCs by a function of the eigenvalues:

$$\text{PCQ}_\gamma = \sum_{k=1}^{K} \lambda_k^{1-\gamma} \left(\frac{\text{PC}_k}{\sqrt{\lambda_k}}\right)^2, 0 \le \gamma < +\infty,$$

where γ controls the relative importance of each PC in the quadratic combinations. For example, if PC_1 captures most of the signals, then we can choose smaller γ; while if PC_K captures most of the signals, then we can choose larger γ. Let $\mathbf{K}_\gamma = \mathbf{U}\boldsymbol{\Lambda}^{-\gamma}\mathbf{U}^T$ and denote the transformation as $\phi_\gamma(\mathbf{Z}) = \mathbf{U}^T\mathbf{Z}/\lambda^{\gamma/2}$, then PCQ_γ can be rewritten as $\text{PCQ}_\gamma = \mathbf{Z}^T\mathbf{K}_\gamma\mathbf{Z} = \langle\phi_\gamma(\mathbf{Z}), \phi_\gamma(\mathbf{Z})\rangle$, where $\langle\cdot, \cdot\rangle$ denotes the inner product in the transformed feature space. From this point of view, PCQ_γ is a

kernel-based testing statistic (Liu, Lin, and Ghosh 2007). The choice of γ is essentially a choice of kernel and reflects our prior belief in the true alternative. We show in this section that several commonly used tests with $\gamma = 0, 1, 2$ are special cases of PCQ_γ.

When $\gamma = 0$, PCQ_γ has the following form:

$$\text{WI} = \text{PCQ}_0 = \sum_{k=1}^{K} \text{PC}_k^2 = \sum_{k=1}^{K} Z_k^2,$$

which follows from the fact that $\mathbf{U}$ is an isometric transformation and $\mathbf{U}\mathbf{U}^T = \mathbf{I}$. This choice of γ assumes a working independence (WI) relationship among the K Z-scores since $\mathbf{K}_\gamma$ reduces to an identity matrix. Under the null, WI follows a mixture of chi-squared distribution $\sum_j \lambda_j \chi_{1j}^2$, where λ_j are the eigenvalues of $\boldsymbol{\Sigma}$ and χ_{1j}^2 are independent χ_1^2 random variables. Hence, its p-value can be computed using the exact method (Davies 1980). At the significance level α, we reject the null hypothesis H_0: $\boldsymbol{\beta} = 0$ if $\sum_{k=1}^{K} \text{PC}_k^2 > C_\alpha$ where $P(\sum_{k=1}^{K} \text{PC}_k^2 > C_\alpha; H_0) = \alpha$. Thus, the acceptance region of WI is a K-dimensional ball with radius equal to $\sqrt{C_\alpha}$. Although the acceptance region of WI is spherically symmetric, however the probability distribution of $\mathbf{Z}$ is not spherically symmetric unless $\boldsymbol{\Sigma}$ is an identity matrix. Under the alternative, the power of WI favors the alternatives under which PC_1 captures all the signals. This is because PC_1 has the largest variance and hence signals from the PC_1 direction are more likely to fall outside of this ball-shape acceptance region (see Figure 1).

When $\gamma = 1$, PCQ_γ becomes the classical Wald test as

$$\text{Wald} = \text{PCQ}_1 = \sum_{k=1}^{K} \frac{\text{PC}_k^2}{\lambda_k} \sim \chi_K^2.$$

This can be easily shown using the fact that $\mathbf{Z}^T\boldsymbol{\Sigma}^{-1}\mathbf{Z} = (\mathbf{U}^T\mathbf{Z})^T\boldsymbol{\Lambda}^{-1}(\mathbf{U}^T\mathbf{Z}) = \sum_{k=1}^{K} \text{PC}_k^2/\lambda_k$. At the significance level α, its acceptance region is determined by $\sum_{k=1}^{K} \text{PC}_k^2/\lambda_k \le C_\alpha$, which is a K-dimensional ellipsoid and C_α is the $1-\alpha$ percentile of χ_K^2. Under the alternative, the distribution of the Wald test is a noncentral chi-squared distribution with noncentrality parameter $\text{ncp} = \sum_{k=1}^{K} ||\boldsymbol{\beta}||^2 \cos^2(\theta_k)/\lambda_k$. To know when the Wald test achieves its maximal power for any fixed $||\boldsymbol{\beta}||$ and λ_k, we can solve the following constrained optimization problem:

$$\max_{\cos^2(\theta_k)} \quad \sum_{k=1}^{K} \lambda_k^{-1} \cos^2(\theta_k)$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \cos^2(\theta_k) = 1.$$

Using standard linear programming technique, one can easily show that the power of the Wald test is maximized when $\theta_K = 0$, that is, when signals lie in the last PC direction, and minimized when $\theta_1 = 0$, that is, when signals lie in the first PC direction. Again, even though the Wald test achieves its maximal power when the last PC captures all the signals, this does not imply the Wald test is more powerful than its competitors under such alternatives.

When $\gamma = 2$, PCQ_γ is

$$\mathrm{VC} = \mathrm{PCQ}_2 = \sum_{k=1}^{K} \frac{\mathrm{PC}_k^2}{\lambda_k^2}.$$

In this case, PCQ_γ corresponds to the variance component (VC) score test $\mathrm{VC} = \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{Z}$ (Huang and Lin 2013; Liu and Lin 2018), which assumes that the β_k ($k = 1, \ldots, K$) follow a common distribution with mean 0 and variance τ and tests for $H_0 : \tau = 0$. The equivalence between VC and PCQ_2 can be seen by observing that $\mathrm{VC} = \mathbf{Z}^T \mathbf{U} \boldsymbol{\Lambda}^{-2} \mathbf{U}^T \mathbf{Z} = \sum_{k=1}^{K} \mathrm{PC}_k^2 / \lambda_k^2$. Compared with the Wald test, VC gives even more weight to the last PC and hence is more powerful than the Wald test when $\theta_K = 0$. VC follows a mixture of chi-squared distributions $\sum_{k=1}^{K} \lambda_k^{-1} \chi_{1k}^2$ under the null, where λ_k are the eigenvalues of $\boldsymbol{\Sigma}$ and χ_{1k}^2 are independent χ_1^2 random variables, so its p-value can be computed using the exact method (Davies 1980). The acceptance region of VC is also a K-dimensional ellipsoid but has a different shape from that of Wald as shown in Figure 1 in Section 4.

Here, we present a simple example to illustrate the power difference between the three quadratic tests: WI, Wald, and VC. Suppose we have a bivariate normal Z-scores with correlation $\rho = 0.8$. The first eigenvector is $\mathbf{u}_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and the second eigenvector is $\mathbf{u}_2 = (-1/\sqrt{2}, 1/\sqrt{2})$ by direct calculation. If $\boldsymbol{\beta} = (2.5, 2.5)^T$ which is in the direction of $\mathbf{u}_1$, then the powers of WI, Wald, and VC are 0.75, 0.65, 0.09, respectively; and if $\boldsymbol{\beta} = (-0.8, 0.8)^T$ which is in the direction of $\mathbf{u}_2$, then the powers of WI, Wald, and VC are 0.09, 0.61, 0.71, respectively. This shows that the Wald test is less powerful than WI when the first PC captures all the signals, and the Wald test is less powerful than VC when the last PC captures all the signals. More power comparisons among those three tests are provided in Table 3.

3. The Omnibus PC-Based Tests

3.1. Adaptive Quadratic Combination of PCs

The results in Section 2.6 show that a lack of prior knowledge about the true principal angle can lead to an unwise choice of γ, and the resulting test might have little power to detect the alternative. Therefore, we propose to choose γ in a data-adaptive fashion by choosing γ using the data that yields the smallest p-value, and then use this smallest p-value as a test statistic. In practice, it is computationally expensive to perform an exhaustive search for the optimal γ in the whole range. Instead, we restrict our search within $\gamma \in \{0, 1, 2\}$ and then pick the smallest p-value among WI, Wald, and VC as a testing statistic named PCAQ

$$\mathrm{PCAQ} = \min_\gamma p_\gamma,$$

where p_γ is the p-value of PCQ_γ for a given γ. Note that WI, Wald, and VC tests are correlated as they are calculated using the same data. Hence, their p-values p_γ are correlated. Calculations of the p-value of PCAQ need to take their correlations into account. Specifically, the p-value of PCAQ can be calculated as

$$p = 1 - P\{\min_\gamma X_\gamma > \Phi^{-1}(\mathrm{PCAQ})\}, \qquad (4)$$

where $X_\gamma = \Phi^{-1}(p_\gamma)$ and $\Phi^{-1}(\cdot)$ denotes the inverse standard normal cumulative distribution function.

Equation (4) can be efficiently computed using the following multivariate normal distribution function that has been implemented in the FORTRAN language (Genz 1992, 1993) and also wrapped in the R package *mvtnorm* (Genz et al. 2009). This computation requires an input of the correlation matrix $\mathbf{R_X}$ of the vector $(X_{\gamma=0}, X_{\gamma=1}, X_{\gamma=2})$ which only needs to be estimated once for the whole genome by the following algorithm:

1. Generate B random samples from $\mathbf{Z} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.
2. Compute the p-values of PCQ on the bth sample for $\gamma = 0, 1, 2$, $1 \le b \le B$.
3. Perform inverse-normal transformation $X_\gamma = \Phi^{-1}(p_\gamma)$ on the bth sample, where $\gamma = 0, 1, 2$.
4. Take the sample correlation matrix $\widehat{\mathbf{R}}_X$ across the B realizations of $\mathbf{X}_\gamma$.

In practice, one can take $B = 1000$ and this algorithm can provide a good estimate of $\mathbf{R}_X$ (in a few seconds) which can be used for computing the p-values for millions of SNPs in the whole genome.

3.2. The Omnibus PC-Based Test

The PCAQ test aims at constructing an optimal quadratic PC-based test. To construct an omnibus test across linear, quadratic, and other nonlinear tests, we can combine all the PC combination methods including PCMinP, PCFisher, PCLC, WI, Wald, and VC together by taking the minimum p-value of them as the omnibus test statistic named PCO. The p-values of those six tests are correlated as they are calculated using the same data. Similar to PCAQ, the p-value of PCO can also be computed by first performing an inverse-normal transformation of the p-value of the test statistic under consideration, then using a multivariate normal distribution function with the correlation matrix estimated using the same fast Monte Carlo simulation method described above. Compared to PCAQ, PCO combines three more nonquadratic tests and is expected to be more robust than PCAQ for various alternatives. However, a price PCO has to pay for combing more tests is that it might be slightly less powerful than PCAQ when quadratic combinations of PCs already have good power, for example, in the simulation setting M3 in Table 3. PCO is expected to be more powerful than PCAQ when any of PCLC, PCFisher, or PCMinP has better power than the quadratic combinations of PCs to detect the signals, as demonstrated in the simulation settings M4, M7, M12, M13, and M15 in Table 3.

4. Comparison of the Rejection Boundaries of the PC-Based Tests

In this section, we compare the proposed PC-based tests graphically in terms of their rejection boundaries. For the ease of illustration, we focus on the two-dimensional $(Z_1, Z_2)^T$ space as given in Figure 1. We also included the Oracle test for $\boldsymbol{\beta} = (1, 1)^T$ and $\boldsymbol{\beta} = (0, 1)^T$ assuming the true alternative is known. We set the correlation to be 0.6, so the two eigenvalues are $\lambda_1 = 1.6$ and $\lambda_2 = 0.4$, and the two corresponding eigenvectors are
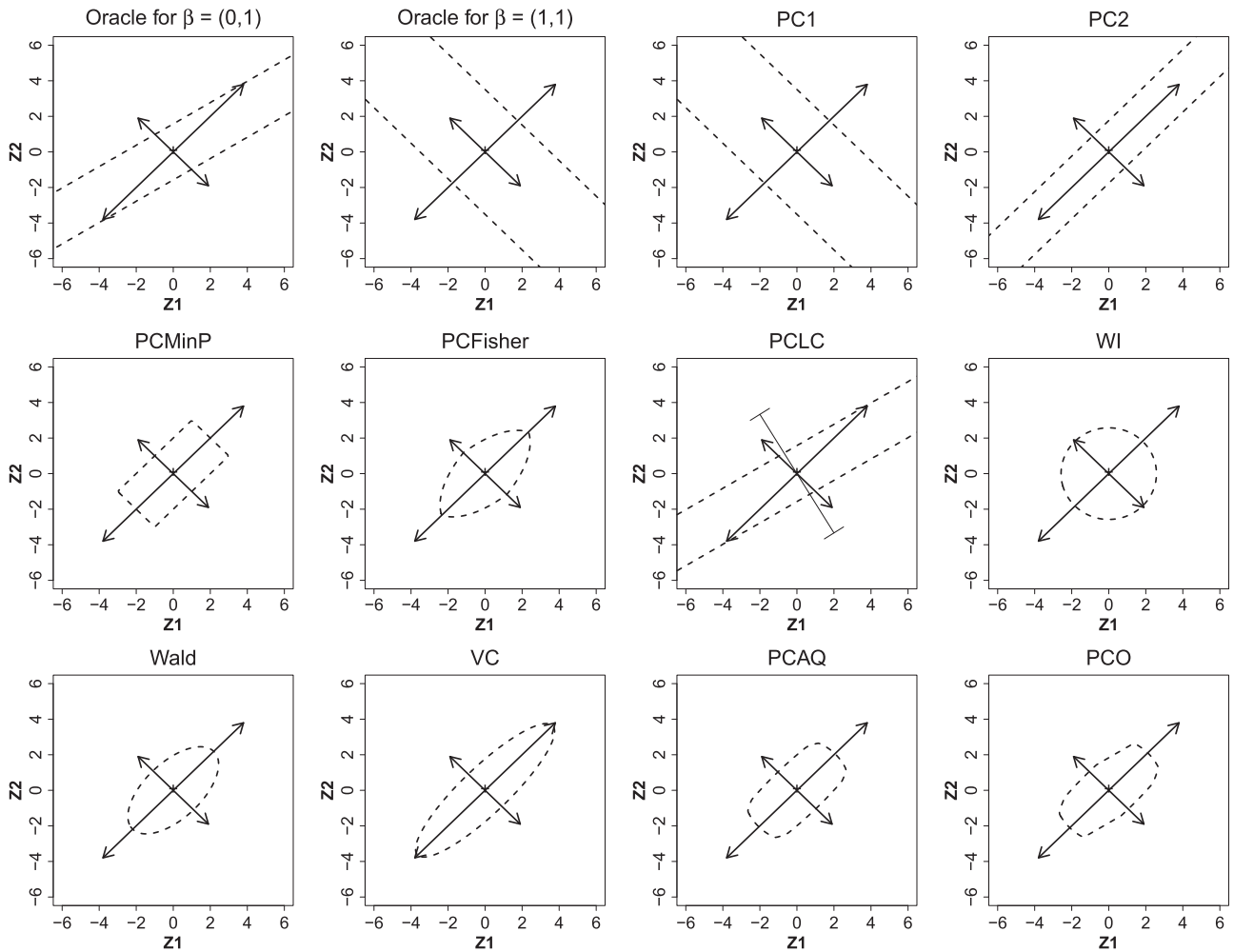
Figure 1. The rejection boundaries of the proposed PC-based tests for bivariate normal test statistics (Z_1, Z_2) with correlation equal to 0.6. The dashed lines or curves represent the boundaries that separate the acceptance and rejection regions at the significance level 0.05. The longer solid lines with arrows represent the PC_1 direction, the shorter solid lines with arrows represent the PC_2 direction, and the lengths of the longer and shorter solid lines with arrows are equal to $6\sqrt{\lambda_1}$ and $6\sqrt{\lambda_2}$, respectively. For PCLC, the added solid line with "T"-type arrows illustrates the direction for alternative β which is orthogonal to its rejection boundaries, where $\theta_1 = 76°$ and $\theta_2 = 14°$.

$\mathbf{u}_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$ and $\mathbf{u}_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$, respectively. Suppose that the true alternative is $\boldsymbol{\beta} = (1, 1)^T$ which is in the same direction of $\mathbf{u}_1$, then PC_1 has the same rejection boundary as the Oracle test. Suppose that the true alternative is $\boldsymbol{\beta} = (0, 1)^T$ which is in the same direction of $\mathbf{UJ}$ where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2)$ is the eigenvector matrix and $\mathbf{J} = (1, 1)^T$, then PCLC has the same rejection boundary as the Oracle test. One can further deduce that PC_2 will have the same rejection boundary as the Oracle test if the true alternative is proportional to $\mathbf{u}_2$. We observe that all the proposed tests have convex acceptance regions. Hence, the proposed tests are all admissible (Birnbaum 1954, 1955). This implies that each test can be more powerful than its competitors for some alternatives but less powerful for others.

The rejection boundaries of PC_1 and PC_2 are all straight lines but are orthogonal to each other, indicating that these two PCs aim to detect orthogonal alternatives. PC_2 has a narrower gap between the two rejection boundary lines than that of PC_1, because PC_2 has a smaller variance (eigenvalue). The rejection boundaries of PCLC are also straight lines but are not orthogonal to either PC_1 or PC_2. The angle between the rejection boundary lines of PCLC and PC_1 is 14°. Hence, if the mean vector $\boldsymbol{\beta}$ also has angle 14° with PC_1 direction, then $\boldsymbol{\beta}$ is parallel

to the rejection boundaries of PCLC and will never be detected by PCLC. If $\boldsymbol{\beta}$ has angle 76° with PC_1 direction as shown by the solid line with "T"-type arrows, then $\boldsymbol{\beta}$ is orthogonal to the rejection boundaries of PCLC (shortest distance to the null) and will be detected by PCLC with its maximal power. However, this does not imply PCLC is more powerful than its competitors to detect the alternatives in the direction of the solid line with "T"-type arrows because PCLC is not as powerful as the Oracle test for such alternatives.

The rejection boundary of PCMinP is a tilted rectangle with the edge lengths proportional to $\sqrt{\lambda_k}, k = 1, 2$. PCMinP achieves its maximal power when $\boldsymbol{\beta}$ is in the PC_2 direction (shortest distance to the null), while achieves its minimal power when $\boldsymbol{\beta}$ points to the four corners, under which PC_1 and PC_2 have equal powers. The rejection boundaries of Wald and VC are both ellipses. However, the minor axis of VC is shorter than that of Wald, while the major axis of VC is longer than that of Wald. This implies that VC is more powerful than Wald when $\boldsymbol{\beta}$ is in the PC_2 direction. The rejection boundaries of PCFisher are similar to that of Wald, which well explains why they have similar powers as will be demonstrated in the simulation studies. The rejection boundary of WI is a circle and WI actually favors

alternatives in the PC_1 direction because it is more likely for the signals to fall outside of the rejection boundary of WI along the PC_1 direction compared to other PC directions.

We observe that the rejection boundaries of PCMinP, PCAQ, and PCO resemble each other because these three tests all use the minimum p-value as testing statistics across certain sets of tests and hence are data adaptive. The rejection boundary of PCAQ is smooth and does not have sharp angles like that of PCMinP. The rejection boundary of PCO is more bumpy than that of PCAQ since it combines linear and nonlinear tests. The rejection boundary comparisons well explain the differences and connections between the proposed PC-based tests, and illustrate that there is no UMP test for all the $\boldsymbol{\beta}$ directions.

5. Eigen-Analysis of Correlation Matrices and Their Effects on the PC-Based Tests

The results in Sections 2 and 3 show that the powers of the PC-based tests depend on the principal angles θ_k, the eigenvalues λ_k for a fixed norm of $\boldsymbol{\beta}$. To test for associations between an SNP and a set of multiple phenotypes, a question of practical interest is that how the PC-based tests perform in the presence of a mixture of signal and noise phenotypes, especially when signals are sparse. For example, in studying the effects of an SNP on a genetic pathway/network consisting of multiple gene expressions, it is common that an SNP affects some gene expressions but not others in the genomic pathway/network. In this section, we investigate how the correlation structure of $\boldsymbol{\Sigma}$ affects its eigenvalues and eigenvectors, and subsequently affects the powers of PC-based tests. Suppose that K_1 out of K phenotypes are associated with a genetic variant and $K_0 = K - K_1$ of them are not, that is, the signal vector $\boldsymbol{\beta}$ contains K_1 nonzeros (signals) and K_0 zeros (noises), denoted as $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_{K_1}, 0, \ldots, 0)$. For $\boldsymbol{\Sigma}$, we consider the following partitioned correlation matrix:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_2 \\ \boldsymbol{\Sigma}_2^T & \boldsymbol{\Sigma}_3 \end{pmatrix},$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3$ denote the correlation matrices among signal and noise phenotypes, respectively, and the (i, j)th element of $\boldsymbol{\Sigma}_2$ denotes the correlation between the ith signal phenotype and the jth noise phenotype, $1 \le i \le K_1, 1 \le j \le K_0$. We first obtain eigen-analysis results for special structured correlation matrices $\boldsymbol{\Sigma}$, and then consider more general situations.

5.1. Exchangeable Correlation Matrices

If $\boldsymbol{\Sigma}$ is exchangeable with correlation $\rho > 0$, then its eigenvalues are

$$\lambda_1 = (K - 1)\rho + 1; \lambda_k = 1 - \rho, k = 2, \ldots, K,$$

where the algebraic multiplicities of λ_1 and λ_k are one and $K - 1$, respectively. This implies that the eigenspace associated with λ_1 is of dimension one and can be spanned by eigenvector $\mathbf{u}_1^T = (\frac{1}{\sqrt{K}}, \frac{1}{\sqrt{K}}, \ldots, \frac{1}{\sqrt{K}})$, while the eigenspace associated with eigenvalue $\lambda_k, k = 2, \ldots, K$ is of dimension $K - 1$ and can be expressed as

$$E_{\lambda=1-\rho} = \{\mathbf{u} \in R^K : \sum_{k=1}^K u_k = 0\}.$$

Actually, there are infinitely many possible choices of the $K - 1$ eigenvectors in $E_{\lambda=1-\rho}$ when $K \ge 3$, and hence infinitely many possible choice of $PC_k, k = 2, \ldots, K$.

The first eigenvalue $(K - 1)\rho + 1$ is usually much larger than eigenvalue $1 - \rho$ for relatively large K. Such a correlation structure is related to the spiked population co-variance model (Johnstone 2001). The principal angle between $\boldsymbol{\beta}$ and $\mathbf{u}_1$ is 0 when $\boldsymbol{\beta} = (1, 1, \ldots, 1)^T c$ where c is a nonzero scalar, and is $90°$ when $\sum_{k=1}^K \beta_k = 0$. Therefore, PC_1 can best detect fully dense homogeneous signals, and its power decreases when the signals become sparser or in different directions. In addition, the power of PC_1 decreases when the correlation ρ increases. When signals are fully dense and homogeneous, the WI test will also have good power, but the Wald and VC tests might have low power. For example, in the simulation study, when $K = 40$ and $\rho = 0.2$, the WI test has power of 0.82 to detect fully dense and homogeneous signals $\boldsymbol{\beta} = (1.4, 1.4, \ldots, 1.4)^T$, but the Wald test has power of only 0.24 as shown in the setting M9 in Table 3. However, if the signals are heterogenous and $\sum_{k=1}^K \beta_k = 0$ with at least one β_k nonzero, $\boldsymbol{\beta}$ is in the eigen-space $E_{\lambda=1-\rho}$ and can be detected by Wald and VC with good power but not WI.

5.2. Block Diagonal Exchangeable Correlation Matrices

If $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3$ are exchangeable with correlations ρ_1, ρ_3, respectively, and the K_1 signal phenotypes are uncorrelated with the K_0 noise phenotypes, then the four unsorted eigenvalues of $\boldsymbol{\Sigma}$ and their algebraic multiplicities are

$$\lambda_1 = 1 + (K_1 - 1)\rho_1, \nu(\lambda_1) = 1;$$
$$\lambda_2 = 1 - \rho_1, \nu(\lambda_2) = K_1 - 1;$$
$$\lambda_3 = 1 + (K_0 - 1)\rho_3, \nu(\lambda_3) = 1;$$
$$\lambda_4 = 1 - \rho_3, \nu(\lambda_4) = K_0 - 1.$$

The signal phenotype eigen-spaces are

$$E_{\lambda_1} = \{\mathbf{u} \in R^K : \mathbf{u}^T = t(\underbrace{1/\sqrt{K_1}, \ldots, 1/\sqrt{K_1}}_{K_1}, \underbrace{0, \ldots, 0}_{K_0}), t \in R\},$$

$$E_{\lambda_2} = \{\mathbf{u} \in R^K : \sum_{k=1}^{K_1} u_k = 0, u_{K_1+1} = \cdots = u_K = 0\},$$

and the noise phenotype eigen-spaces are

$$E_{\lambda_3} = \{\mathbf{u} \in R^K : \mathbf{u}^T = t(\underbrace{0, \ldots, 0}_{K_1}, \underbrace{1/\sqrt{K_0}, \ldots, 1/\sqrt{K_0}}_{K_0}), t \in R\},$$

$$E_{\lambda_4} = \{\mathbf{u} \in R^K : \sum_{k=K_1+1}^{K} u_k = 0, u_1 = \cdots = u_{K_1} = 0\}.$$

Because the signal and noise phenotype eigen-spaces are orthogonal to each other, thus those K_0 PCs in the noise eigen-space are all powerless to detect any signals. Therefore, we only need to focus on discussing the powers of the K_1 PCs in the signal phenotype eigen-spaces. From Section 5.1, we know that the PC in E_{λ_1} can best detect homogeneous effects while the PCs in E_{λ_2} can best detect heterogenous effects. Actually, as long as some principal angle is zero, then that particular PC with zero principal angle in the signal eigen-space is as powerful as the Oracle test, regardless of signal sparsity.

5.3. Block Diagonal Correlation Matrices

We now consider more general situations where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_3$ are unstructured. By performing spectral decomposition on these two matrices, we have $\boldsymbol{\Sigma}_1 = \mathbf{U}_1 \boldsymbol{\Lambda}_1 \mathbf{U}_1^T$ and $\boldsymbol{\Sigma}_3 = \mathbf{U}_3 \boldsymbol{\Lambda}_3 \mathbf{U}_3^T$, where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_3$ are diagonal matrices with diagonal elements the eigenvalues, $\mathbf{U}_1$ and $\mathbf{U}_3$ are eigenvector matrices. If the signal and noise phenotypes are uncorrelated, then we have

$$\begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_3 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_3 \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_3 \end{pmatrix}^T.$$

Therefore, the PCs from the signal eigen-space allocate zero loadings for the noise phenotypes. In other words, if signal and noise phenotypes are uncorrelated, then the PCs from the signal eigen-space are not contaminated by any noise phenotypes and thus one particular PC from the signal eigen-space can be as powerful as the Oracle test if its principal angle is zero, regardless of signal sparsity.

6. Simulation Studies

6.1. Type I Error Rates

Single PC tests and PCLC follow the standard normal distribution under the null so that their Type I error rates are always well controlled, hence we omit their Type I error results. Besides the Wald test, we evaluate the sizes of the proposed PC-based tests, including the p-value-based tests PCMinP and PCFisher, and quadratic tests WI and VC, and the omnibus tests PCAQ and PCO, at the nominal levels $\alpha = 0.05, 0.01, 0.001, 10^{-4}, 10^{-5}$, in view of the small significance levels that are of common interest in GWAS. For comparison purpose, we also included the p-value correction method TATES (van der Sluis et al. 2013) for comparison purpose which also only requires GWAS summary statistics. We first consider a low-dimensional unstructured covariance matrix $\boldsymbol{\Sigma}_{unK3}$ estimated from the global lipids data (Teslovich et al. 2010) for high-density lipoprotein cholesterol (HDL), total cholesterol (TC), and triglycerides (TG),

$$\boldsymbol{\Sigma}_{unK3} = \begin{bmatrix} 1.00 & 0.16 & -0.42 \\ 0.16 & 1.00 & 0.38 \\ -0.42 & 0.38 & 1.00 \end{bmatrix}. \quad (5)$$

We also consider a high-dimensional ($K = 100$) unstructured covariance matrix $\boldsymbol{\Sigma}_{unK100}$ generated using the algorithm described in Marsaglia and Olkin (1984) and the actual matrix is provided in the supplementary excel file. We generated 10 millions of multivariate normal samples of dimensions $K = 3$ and $K = 100$ with mean zeros and covariance matrices equal to $\boldsymbol{\Sigma}_{unK3}$ and $\boldsymbol{\Sigma}_{unK100}$, respectively. We found that the Type I error rates of the PC-based tests are well controlled at those nominal levels as summarized in Table 1. The Type I error rates of the p-value correction method TATES are slightly inflated, in line with previous findings by He, Avery, and Lin (2013).

6.2. Power Comparisons of the PC-Based Tests

As shown in Figure 1, different tests have different rejection boundaries and the power of each test depends on both the mean vector and the covariance matrix of $\mathbf{Z}$. We first provide

empirical evidence using bivariate phenotypes to show that the powers of the PC-based tests depend on the direction of the true $\boldsymbol{\beta}$ for a fixed between-phenotype correlation matrix, and no single test is most powerful for all directions of $\boldsymbol{\beta}$, while the omnibus tests are more robust.

Consider a bivariate standard normal $(Z_1, Z_2)^T$ with $\rho = 0.6$ and mean $\boldsymbol{\beta} = (\beta_1, \beta_2)^T \neq \mathbf{0}$ under the alternative. Using the polar coordinate system, we can rewrite $\boldsymbol{\beta} = r\{\cos(\phi), \sin(\phi)\}$, where $r \geq 0$ and $\phi \in [0, 360°]$. For illustrative purpose, we set $r = 2$. Then the power of each test is a function of ϕ only. We divide the interval $[0, 360°]$ equally into 72 subintervals specified by 73 grid points, $\phi_b = 0, 5°, \ldots, 360°, b = 1, 2, \ldots, 73$. For each ϕ_b, we generated one million standard bivariate normal samples with $\rho = 0.6$ and obtained one million p-values for each test. The power of each test for each ϕ_b is estimated by the proportion of p-values that are less than $\alpha = 0.05$. By connecting the 73 power points of each test, we obtain the power function curves in Figure 2.

We found a periodic pattern of the power curves with period equal to $180°$, and within each period, the power of each test is a function of ϕ, where ϕ specifies the direction of $\boldsymbol{\beta}$ in $\mathcal{R}^2$. PC_1 is as powerful as the Oracle test when $\phi = 45°, 225°$ or equivalently when $\boldsymbol{\beta} = \pm(\sqrt{2}, \sqrt{2})^T$, and consequently WI is almost as powerful as the Oracle test in such settings. Likewise, PC_2 is as powerful as the Oracle test when $\phi = 135°, 315°$ or equivalently when $\boldsymbol{\beta} = \pm(-\sqrt{2}, \sqrt{2})^T$, and VC is almost as powerful as the Oracle test (more powerful than Wald) in these settings. Wald and PCFisher have almost the same power curves. PCLC is as powerful as the Oracle test when $\phi = 90°, 270°$, or $\boldsymbol{\beta} = \pm(0, 2)^T$. The omnibus tests PCAQ and PCO are never as powerful as the ideal Oracle test, but are robust to the alternatives with little power loss compared to the Oracle test. When either the first or the last principal angle is zero, PCAQ and PCO can be more powerful than Wald.

In low-dimensional settings where $K = 3$ (M1–M5), we consider an unstructured correlation matrix $\boldsymbol{\Sigma}_{unK3}$ given in Equation (5) and the following five mean vectors: $\boldsymbol{\beta}_1 = (-1.94, 1.58, 2.87)$, $\boldsymbol{\beta}_2 = (2.31, 2.62, 0.1)$, $\boldsymbol{\beta}_3 = (0.99, -0.94, 1.18)$, $\boldsymbol{\beta}_4 = (0.94, 0.86, 1.92)$, $\boldsymbol{\beta}_5 = (0.79, 3.2, 0.16)$. We also consider $K = 8$ and an unstructured correlation matrix $\boldsymbol{\Sigma}_{unK8}$ given in Table 4 of Section 7, and three mean vectors: $\boldsymbol{\beta}_6 = 4.5\mathbf{u}_1 = (1.18, 0.69, 1.33, -2.39, 1.39, 2.66, 1.27, 0.52)$, $\boldsymbol{\beta}_7 = 3.5\mathbf{u}_4 = (-0.8, 1.04, 0.74, 0.39, 1.69, -0.01, -0.67, -2.55)$, and $\boldsymbol{\beta}_8 = 2.5\mathbf{u}_8 = (0.08, -0.02, -0.11, 1.54, -0.65, 1.83, -0.08, -0.24)$, where $\mathbf{u}_1, \mathbf{u}_4, \mathbf{u}_8$ denote the first, fourth, and eighth eigenvectors of $\boldsymbol{\Sigma}_{unK8}$, respectively.

In high-dimensional settings, we first consider $K = 40$ and an exchangeable correlation matrix $\boldsymbol{\Sigma}_{exK40}$ with off-diagonal correlation $\rho = 0.2$, and a fully dense and homogeneous signal vector $\boldsymbol{\beta}_9 = (1.4, 1.4, \ldots, 1.4)$ as in setting M9. In setting M10, the correlation matrix $\boldsymbol{\Sigma}_{bexK40}$ is block diagonal where the signal ($K_1 = 6$) and noise ($K_0 = 34$) blocks are exchangeable with correlations equal to 0.5 and 0.2, respectively, and the signal vector is sparse and $\boldsymbol{\beta}_{10} = (1.98, -1.51, -0.12, -0.12, -0.12, -0.12, 0, \ldots, 0)$ with 6 nonzero elements and 34 zero elements. We also consider $K = 100$ and an unstructured correlation matrix $\boldsymbol{\Sigma}_{unK100}$ (provided in supplementary excel file) generated using the Marsaglia and Olkin (1984) algorithm. In settings M11 and M12, consider

Table 1. Type I error rates estimated as the proportions of p-values less than significance level α in 10^7 simulation replications under the nulls in both low- and high-dimensional settings.

				Low-dimensional setting: $K = 3$, covariance matrix is Σ_{unK3}				
α	PCMinP	PCFisher	WI	Wald	VC	PCAQ	PCO	TATES
0.05	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.051
0.01	0.010	0.010	0.010	0.010	0.010	0.0099	0.0099	0.011
0.001	0.001	0.00099	0.001	0.001	0.001	0.001	0.00099	0.00106
10^{-4}	9.75×10^{-5}	1.01×10^{-4}	1.02×10^{-4}	1.01×10^{-4}	1.02×10^{-4}	9.81×10^{-5}	9.93×10^{-5}	1.08×10^{-4}
10^{-5}	1.03×10^{-5}	9.32×10^{-6}	8.33×10^{-6}	8.61×10^{-6}	8.52×10^{-6}	8.47×10^{-6}	9.71×10^{-6}	1.05×10^{-5}
				High-dimensional setting: $K = 100$, covariance matrix is Σ_{unK100}				
α	PCMinP	PCFisher	WI	Wald	VC	PCAQ	PCO	TATES
0.05	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.051
0.01	0.010	0.010	0.010	0.010	0.010	0.0099	0.0099	0.011
0.001	0.001	0.00099	0.001	0.001	0.001	0.001	0.00099	0.00106
10^{-4}	9.65×10^{-5}	1.01×10^{-4}	1.02×10^{-4}	1.01×10^{-4}	1.02×10^{-4}	9.86×10^{-5}	9.95×10^{-5}	1.06×10^{-4}
10^{-5}	1.02×10^{-5}	9.38×10^{-6}	8.46×10^{-6}	8.78×10^{-6}	8.99×10^{-6}	9.37×10^{-6}	9.78×10^{-6}	1.05×10^{-5}
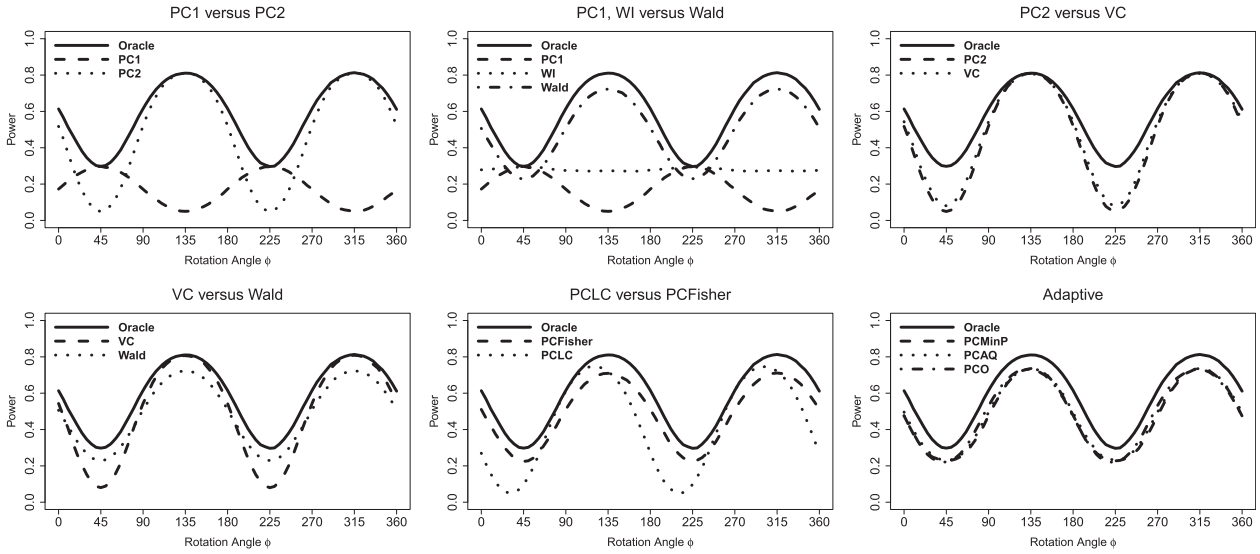
Figure 2. This figure shows the power curve of each PC test for alternatives $\beta = r(\cos(\phi), \sin(\phi))$ where $r = 2$ and $\phi \in [0, 360°]$ are the polar coordinates. The bivariate correlation is $\rho = 0.6$. This figure mirrors the rejection boundaries as shown in Figure 1.

two sparse signal vectors $\beta_{11} = (0.03, 0.04, -0.02, 0.05, -0.04, 0.01, 0.02, 0.09, -0.13, -0.02, 0, \ldots, 0)$ and $\beta_{12} = (-0.17, 0.4, -0.05, 0.19, -0.68, 0.21, 0.3, -0.28, 0.29, -0.11, 0, \ldots, 0)$ that both contain 10 nonzero elements and 90 zeros. In settings M13-M15, consider three dense signal vectors (provided in the supplementary excel file): β_{13} contains 90 signals and 10 noises, β_{14} and β_{15} are set to be proportional to the first and the eightieth eigenvectors of Σ_{unK100}, respectively.

For each setup, we generated 10^5 multivariate normal samples with mean equal to β and correlation matrix equal to Σ and obtained 10^5 p-values for each test. The empirical power was calculated as the proportion of p-values less than $\alpha = 0.05$. We summarize the power results in Table 3. The results show that whenever $\theta_k = 0$, then PC_k is as powerful as the Oracle test as shown in the low-dimensional settings from M1 to M3. PC_1 requires a larger overall signal magnitude $||\beta||$ to have comparable power as that of PC_3, simply because PC_1 has a larger variance. As expected, WI is more powerful than Wald and VC whenever PC_1 is the Oracle test, while VC is more powerful than WI and Wald whenever the last PC is the Oracle test. PCLC is as powerful as the Oracle test as shown in the setting M4 where

the three principal angles are equal to each other, and PCLC is powerless in the setting M5 where the signal vector β is parallel to the rejection boundary of PCLC. The TATES method can have comparable power to PCO in settings M1, M2, and M5, but it can perform poorly in settings M3 and M4. By contrast, the PCO always has good power in all those five settings M1–M5. In M6, the PC_1 test attains the Oracle power, and hence WI and PCMinP are both more powerful than the Wald test, the adaptive omnibus tests PCAQ and PCO also outperform Wald. In setting M7 where we set the fourth principal angle to zero, and thus the Wald test outperform WI and VC in this setting, but is still less powerful than PCMinP. PCO outperforms Wald because PCMinP is one of its combining component. In M8 where the last principal angle is set to be zero, the last PC is the Oracle test and VC is more powerful than Wald. PCAQ and PCO also outperform Wald. The TATES method has comparable power to PCO in M6 but performs poorly than PCO in M7–M8.

In high-dimensional setting M9, PC_1 has the Oracle power to detect fully dense homogeneous signals simply because the first principal angle is zero, while both PCFisher and Wald have very low power in this setting. The TATES method is

Table 2. Parameter configurations for power comparison in simulation studies. The numbers in the three columns correspond to principal angles $\theta_1, \theta_2, \theta_K$ are in the unit of degree. The last column represents the power of the Oracle test.

Setup	K	β	Σ	$\|\beta\|$	θ_1	θ_2	θ_K	Oracle
M1	3	β_1	Σ_{unK3}	3.8	0	90	90	0.85
M2	3	β_2	Σ_{unK3}	3.5	90	0	90	0.89
M3	3	β_3	Σ_{unK3}	1.8	90	90	0	0.85
M4	3	β_4	Σ_{unK3}	2.3	54.7	54.7	54.7	0.79
M5	3	β_5	Σ_{unK3}	3.3	71.2	27.5	109.7	0.91
M6	8	β_6	Σ_{unK8}	4.5	0	90	90	0.93
M7	8	β_7	Σ_{unK8}	3.5	90	90	90	0.94
M8	8	β_8	Σ_{unK8}	2.5	90	90	0	0.92
M9	40	β_9	Σ_{exK40}	8.85	0	90	90	0.83
M10	40	β_{10}	Σ_{bexK40}	2.5	90	90	0	0.94
M11	100	β_{11}	Σ_{unK100}	0.18	90	90.3	3.6	0.92
M12	100	β_{12}	Σ_{unK100}	1.0	90	90	90	0.97
M13	100	β_{13}	Σ_{unK100}	3.38	90	90	90	0.97
M14	100	β_{14}	Σ_{unK100}	15.0	0	90	90	0.88
M15	100	β_{15}	Σ_{unK100}	2.0	90	90	90	0.97

Table 3. Powers estimated as the proportions of p-values less than significance level $\alpha = 0.05$ in 10^5 replications under various alternatives. The setups are described in Table 2.

					Low-dimensional setting: $K = 3$ and $K = 10$							
Setup	PC_1	PC_2	PC_K	PCMinP	PCFisher	PCLC	WI	Wald	VC	PCAQ	PCO	TATES
M1	0.85	0.05	0.05	0.76	0.72	0.23	0.81	0.74	0.31	0.76	0.75	0.77
M2	0.05	0.89	0.05	0.80	0.75	0.30	0.79	0.77	0.44	0.73	0.77	0.77
M3	0.05	0.05	0.85	0.74	0.70	0.67	0.23	0.72	0.84	0.81	0.77	0.28
M4	0.17	0.23	0.59	0.53	0.63	0.79	0.42	0.65	0.68	0.65	0.72	0.43
M5	0.14	0.75	0.45	0.75	0.82	0.05	0.73	0.82	0.71	0.79	0.78	0.78
M6	0.93	0.05	0.05	0.76	0.61	0.14	0.84	0.67	0.32	0.80	0.75	0.71
M7	0.05	0.05	0.05	0.81	0.66	0.23	0.67	0.72	0.68	0.66	0.75	0.59
M8	0.05	0.05	0.92	0.76	0.61	0.34	0.33	0.67	0.83	0.78	0.74	0.34

					High-dimensional setting: $K = 40$ and $K = 100$							
Setup	PC_1	PC_2	PC_K	PCMinP	PCFisher	PCLC	WI	Wald	VC	PCAQ	PCO	TATES
M9	0.83	0.05	0.05	0.42	0.21	0.06	0.82	0.24	0.06	0.75	0.68	0.59
M10	0.05	0.05	0.94	0.64	0.30	0.12	0.10	0.36	0.55	0.45	0.56	0.20
M11	0.05	0.05	0.92	0.53	0.18	0.66	0.05	0.22	0.92	0.85	0.83	0.05
M12	0.05	0.05	0.05	0.75	0.24	0.09	0.05	0.31	0.07	0.19	0.63	0.06
M13	0.06	0.05	0.05	0.73	0.23	0.06	0.07	0.28	0.05	0.19	0.61	0.09
M14	0.88	0.05	0.05	0.42	0.15	0.04	0.86	0.18	0.05	0.77	0.71	0.70
M15	0.05	0.05	0.07	0.72	0.24	0.06	0.06	0.30	0.07	0.20	0.58	0.06

less powerful than the PCO in M9. In M10 where the signals are sparse, we additionally considered the MinP test defined as the minimum p-values across all the K original Z-testing statistics as in Conneely and Boehnke (2007), which is designed to detect sparse signals. The power of MinP (not reported in Table 3) is 0.20, smaller than the powers of PCFisher, Wald, VC, PCAQ, and PCO. This surprising result demonstrates that PC-based tests can outperform MinP for detecting sparse signals by leveraging on the between-phenotype correlation structures. The TATES method also has very low power in M10. In high-dimensional settings M11 and M12 where the signals are sparse, the MinP and the TATES methods are almost powerless while PCO has very good power to detect these two sparse signals. The Wald test also performs poorly in these two settings. For dense signals in settings M13 and M15, the TATES method is almost powerless and the Wald test also has very low power, while the PCO test still has good power. In setting M14 where we set the first principal angle to be zero, PC_1 is the Oracle test and the WI test is thus very powerful, while the Wald test has very low power. The TATES method has comparable power to the PCO test in M14. We found that the TATES method has similar performance to the WI test, which can be explained by the similarities between the rejection boundaries of these two

tests. The rejection boundary of the TATES method is provided in Figure S1 in Section S1 of the supplementary materials.

7. Joint Analysis of Multiple Metabolic Syndrome-Related Phenotypes

We are interested in detecting the genetic associations between individual SNPs and multiple phenotypes of metabolic syndrome using the GWAS summary statistics of the MetS-related phenotypes from the four international consortia described in the Introduction section. The GWAS summary statistics from these four consortia are publicly available. The website links for those datasets are provided in Section S2 in the supplementary materials. However, the individual level phenotype and genotype data are not directly accessible. Hence, any multiple phenotype analysis method that requires individual level data cannot be applied. The single-trait GWAS analysis performed by the four international consortia might miss susceptible SNPs that are associated with MetS, even with very large sample sizes, because the genetic effects of common variants are usually small.

To increase analysis power for identifying additional SNPs associated with MetS, we applied the proposed PC-based testing

Table 4. The correlation matrix Σ of single-trait GWAS Z-scores estimated using the MetS GWAS summary statistics datasets. The eigen-values of Σ are: 1.75, 1.26, 0.99, 0.95, 0.94, 0.82, 0.75, 0.54.

Σ	BMI	FG	FI	HDL	LDL	TG	WHR	SBP
BMI	1	−0.02	−0.04	−0.2	0.05	0.16	−0.01	−0.03
FG	−0.02	1	0.2	−0.02	0.01	0.05	0.03	0.08
FI	−0.04	0.2	1	−0.11	0.03	0.15	0.12	0.08
HDL	−0.2	−0.02	−0.11	1	−0.09	−0.42	−0.11	0
LDL	0.05	0.01	0.03	−0.09	1	0.24	0.06	0
TG	0.16	0.05	0.15	−0.42	0.24	1	0.15	0.07
WHR	−0.01	0.03	0.12	−0.11	0.06	0.15	1	0.06
SBP	−0.03	0.08	0.08	0	0	0.07	0.06	1

Table 5. The numbers of newly detected SNPs (not reported by the original GWAS studies) that reached the genome-wide significance at $\alpha = 5 \times 10^{-8}$ by joint analysis of the eight MetS-related traits (BMI, FG, FI, HDL, LDL, TG, WHR, SBP) using the proposed PC-based tests and the TATES method before and after LD pruning.

LD Pruning	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	
Before	23	7	16	6	29	32	19	108	
After	7	2	3	1	5	4	1	26	
LD Pruning	PCMinP	PCFisher	PCLC	WI	Wald	VC	PCAQ	PCO	TATES
Before	123	404	7	210	458	476	682	581	3
After	25	60	3	42	65	69	103	98	1

procedures and the TATES method (van der Sluis et al. 2013) to the GWAS summary statistics data by jointly analyzing the eight MetS-related traits described in the Introduction. They include Body Mass Index (BMI), Fasting Glucose (FG), Fasting Insulin (FI), High-Density Lipoprotein cholesterol (HDL), Low-Density Lipoprotein cholesterol (LDL), triglycerides (TG), Waist-hip-ratio (WHR), and Systolic Blood Pressure (SBP). We first merged these four GWAS summary statistics datasets using the common 1,999,568 SNPs shared by the four datasets. We then performed our proposed PC-based tests using these univariate Z-scores, and also applied the TATES method on the univariate p-values. The correlation matrix Σ among these MetS-related traits was estimated by the sample correlation matrix across approximately independent SNPs after LD pruning (Zhu et al. 2015; Liu and Lin 2018), and is provided in Table 4.

The QQ plots for the GWAS analysis using the PCA-based methods and the TATES method are provided in Section S2 of the supplementary materials and show the proposed methods had good genomic control. If an SNP was associated with any of the eight MetS-related phenotypes, then it should have been reported previously in the single trait analysis published in the literature. Therefore, the newly detected SNPs are those SNPs that were detected by multiple phenotype analysis but were missed by previous single-trait analysis performed by the four international consortia. In other words, the p-values of those newly detected SNPs are not genome-wide significant in any of the eight single-trait GWAS studies. Since the identified SNPs might be in linkage disequilibrium (LD) with each other, we performed LD pruning to obtain almost independent SNPs using the LD threshold $r^2 < 0.01$ within each 500kb region by PLINK (Purcell et al. 2007). After LD pruning, we greatly reduced the numbers of newly detected significant SNPs, indicating that many newly detected SNPs are in LD with each other. The numbers of new SNPs detected by each test are summarized in Table 5.

In what follows, we only report and discuss the identified SNPs after LD pruning. The last PC detected 26 SNPs after LD pruning, more than the other seven PCs. As expected, VC

detected more SNPs than Wald and WI. PCAQ, which combines WI, Wald, and VC, detected 103 SNPs, while PCO detected 98 SNPs. PCO detected slightly fewer SNPs than PCAQ because PCLC, PCFisher, and PCMinP did not contribute more new SNPs in addition to the WI, Wald, and VC tests. Note by combining more tests, PCO pays a higher price than PCAQ in the p-value adjustment. This is because PCO takes the smallest p-value of the six tests as the test statistic, while PCAQ only uses the smallest p-value of the three tests (WI, Wald, and VC) as the test statistic. Many SNPs can be detected by more than one test. For example, 95 SNPs can be detected by both PCAQ and PCO as shown in Figure 3. It can also be seen from Figure 3 that VC and Wald detected 40 SNPs in common, while VC detected 29 SNPs that Wald failed to detect, and Wald detected 25 SNPs that VC failed to detect. This illustrates that each test can be more powerful than the others in some scenarios because of its admissibility property. It should be noted that our proposed adaptive omnibus tests PCAQ and PCO detected more SNPs than nonadaptive tests, demonstrating their robust performance.

The TATES method only detected three new SNPs missed by the original single-trait GWAS studies. They are rs9600212 ($p = 1.08 \times 10^{-13}$), rs9592962 ($p = 1.45 \times 10^{-12}$), rs9592961 ($p = 7.12 \times 10^{-12}$), within a 2kb intronic region of gene KLF12 on chromosome 13 before LD pruning. The most significant SNP rs9600212 was retained after LD pruning. These three SNPs were also detected by PCFisher, WI, Wald, VC, PCAQ, and PCO. Note that KLF12 was found to be associated with the duration of the Q, R, and S waves (QRS duration), which measures the duration of ventricular muscle depolarization seen on a typical electrocardiogram and hence might play a role in affecting heart functions (Sotoodehnia et al. 2010).

Note that it is possible that an SNP can be detected by single-trait analysis but might not be detected by multiple trait analysis using PC-based tests or the TATES method. For example, SNP rs6129779 on chromosome 20 was found to be associated with LDL ($p = 4.04 \times 10^{-9}$) and has been reported by the GLGC (Willer et al. 2013), but it was not detected using the proposed

(a)

(b)

(c)

(d)
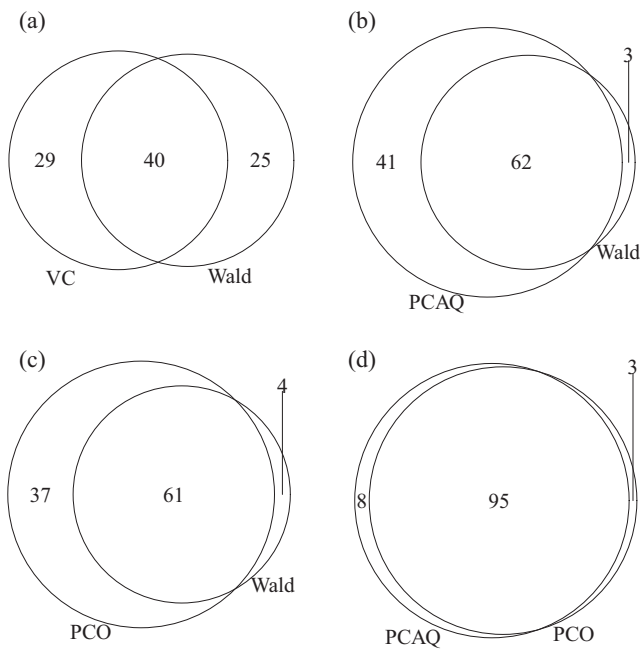
Figure 3. Venn diagrams for the overlapping SNPs detected by Wald, VC, PCAQ, and PCO tests after LD pruning.

PC-based tests or the TATES method. This is because there is only one weak signal mixed with seven noises, and thus joint analysis using either PC-based tests or the TATES method might not be able to detect this rare and weak signal with sufficient power.

We now take a subset of the newly detected SNPs presented in Table 6 to illustrate the differences and connections of our proposed PC-based tests. For the 10 SNPs in Table 6, none of their phenotype-specific p-values reached the genome-wide significance threshold, so those 10 SNPs were not identified by the original single-trait analysis performed by the four international consortia. We also estimated the eight empirical principal angles

(provided in the supplementary excel file) for each of the 10 SNPs by calculating the angles between the Z-score vector and the eigenvectors of the correlation matrix in Table 4 to help illustrate the concept of principal angle in this real data example. For SNP rs355838, its first principal angle was estimated as $41.8°$ and the p-value of PC1 was 1.53×10^{-9}, and all the other seven principal angles for this SNP were more closer to $90°$ and the p-values of all the other seven PCs were not genome-wide significant. Intuitively, this means that the genetic effect vector of SNP rs355838 is more similar to the first PC direction and less similar to the other PC directions. Using the first PC will more likely detect this association signal. As a result, the p-value of WI, which has a similar performance to PC1, for detecting SNP rs355838 is more significant than those of PCFisher, Wald, and VC. Biologically, SNP rs355838 is located in an intronic region of gene COBLL1, which was reported as a pleiotropic gene that was associated with metabolic syndrome and inflammation by Kraja et al. (2014). Specifically, this SNP was found to be associated with at least one metabolic trait and one inflammatory marker.

The last principal angle of SNP rs8321 was estimated to be $38°$, and all the other seven principal angles for this SNP were more closer to $90°$. The p-value of the last PC was 1.71×10^{-11}, and hence VC was more significant than Wald, WI, PCFisher, PCLC, and PCMinP. Another example is SNP rs308971 whose first principal angle was estimated to be $29.26°$ and the p-value of the first PC was 4.39×10^{-8}, while the p-values of all the other seven PCs were not genome-wide significant, because their principal angles are closer to $90°$. For this SNP rs308971, the p-value of WI test was 2.48×10^{-10} while the Wald and VC tests were not even genome-wide significant. This gene SYN2 was found to be related to type 2 diabetes (Zeggini et al. 2008).

As for SNP rs9394279, the p-value of PCLC was 1.4×10^{-8} while the p-values of WI, Wald, VC, and PCAQ were not genome-wide significant. The PCO test which contains

Table 6. p-Values of a selected subset of new SNPs detected by PC-based tests. The p-values of PCAQ and PCO for the first SNP rs355838 are reported as $< 10^{-15}$ due to the numerical precision limits of the R package *mvtnorm*. CHR represents chromosome number.

SNP	CHR	Gene	FG	FI	HDL	LDL	TG	WHR	BMI	SBP
rs355838	2	COBLL1	1.78E-01	1.56E-07	4.10E-07	3.05E-05	1.21E-04	1.70E-07	7.80E-06	1.13E-01
rs8321	6	ZNRD1	1.12E-01	2.96E-01	1.69E-05	1.91E-01	1.05E-06	1.10E-01	3.87E-01	3.19E-01
rs5754352	22	UBE2L3	5.21E-02	9.09E-01	6.57E-08	3.33E-02	9.11E-03	5.70E-01	5.93E-04	2.88E-01
rs308971	3	SYN2	5.25E-03	2.78E-06	3.27E-03	2.36E-02	3.51E-05	1.50E-04	5.54E-01	1.29E-02
rs2269928	11	C11orf9	4.64E-01	4.76E-01	7.63E-02	9.59E-08	1.32E-06	3.20E-01	7.87E-01	2.50E-01
rs10744777	12	ALDH2	9.99E-01	8.40E-01	1.46E-02	1.56E-07	2.18E-02	8.90E-01	1.18E-01	6.24E-06
rs11717195	3	ADCY5	2.68E-07	1.20E-01	4.77E-04	3.77E-02	4.52E-01	2.60E-01	1.34E-04	5.05E-01
rs6485702	11	LRP4	4.60E-02	8.54E-02	6.88E-08	4.68E-04	2.40E-07	2.10E-01	4.72E-01	2.99E-01
rs6810027	3	NISCH	4.11E-01	1.06E-02	1.07E-06	1.67E-02	3.09E-02	1.20E-07	8.68E-02	2.37E-01
rs9394279	6	intergenic	5.02E-01	2.62E-03	2.92E-05	2.96E-02	6.46E-01	1.60E-01	1.56E-01	2.27E-01

SNP	CHR	Gene	PCLC	PCFisher	PCMinP	WI	Wald	VC	TATES	PCAQ	PCO
rs355838	2	COBLL1	1.42E-03	1.10E-16	1.23E-08	9.62E-18	1.28E-16	5.14E-11	6.15E-06	<E-15	<E-15
rs8321	6	ZNRD1	3.05E-06	1.25E-09	1.37E-10	1.28E-05	1.50E-10	9.30E-14	2.78E-04	9.44E-14	1.82E-13
rs5754352	22	UBE2L3	3.69E-05	2.04E-09	8.74E-07	4.28E-06	1.54E-09	7.13E-11	2.69E-05	7.13E-11	1.40E-10
rs308971	3	SYN2	1.73E-02	2.32E-06	3.51E-07	2.48E-10	2.32E-07	1.32E-03	7.29E-05	2.48E-10	4.92E-10
rs2269928	11	C11orf9	9.57E-02	5.71E-09	3.71E-05	7.67E-06	4.61E-09	5.07E-10	6.98E-06	5.38E-10	1.03E-09
rs10744777	12	ALDH2	1.11E-02	6.14E-10	1.43E-06	1.90E-07	6.63E-10	7.54E-09	2.24E-05	6.73E-10	1.22E-09
rs11717195	3	ADCY5	1.78E-06	1.18E-09	8.52E-04	1.69E-07	2.18E-09	4.71E-08	8.51E-06	2.18E-09	2.33E-09
rs6485702	11	LRP4	2.15E-01	7.79E-07	8.47E-07	2.07E-09	3.65E-07	5.21E-04	5.01E-06	2.16E-09	5.35E-09
rs6810027	3	NISCH	2.56E-03	1.69E-08	2.34E-05	4.74E-09	1.72E-08	1.70E-06	4.21E-06	4.94E-09	9.58E-09
rs9394279	6	intergenic	1.40E-08	2.36E-06	4.59E-03	1.34E-04	3.18E-06	1.71E-06	4.23E-04	2.48E-06	2.98E-08

PCLC as one combining component has p-value 2.98×10^{-8}. This demonstrates that PCO which combines PCLC, PCMinP, PCFisher, WI, Wald, and VC all together is more robust than any of the individual components, and is also more robust than PCAQ which only combines three tests: WI, Wald, and VC. We can also see from Table 6 that whenever any of WI, Wald, and VC is significant, then the p-value of PCAQ is slightly more significant than PCO. This is because PCLC, PCMinP, and PCFisher contribute little or none information in addition to WI, Wald, and VC when the latter three tests can already capture the signal, and in this case PCAQ will perform slightly better than PCO. However, as in case of SNP rs9394279, WI, Wald, VC and hence PCAQ failed to detect this signal, but PCO was able to detect it. This is because PCLC can capture this signal. Those identified new SNPs provide potential candidates for future functional studies to better understand their biological roles in the etiology of metabolic syndrome.

8. Discussion

In this article, we proposed a series of principal component-based testing procedures to detect genetic associations between an SNP and multiple phenotypes in GWAS studies. These methods are implemented in our software package MPAT (multiple phenotype association tests). Contrary to the common notion and practice of PCA analysis which usually retains the top few PCs that explain most of the variability in the data for dimension reduction to be used for testing for genetic effects with multiple phenotypes, we found that the higher-order PCs can be more powerful than the top PCs for association analysis. This counter-intuitive phenomenon can be well explained by the novel geometric concept of principal angle first introduced in this article. Theoretically, a particular PC is powerful if its principal angle is zero and powerless if its principal angle is $90°$. Prior to the introduction of the novel concept principal angle, the power of PC-based tests for the multivariate normal means depends on the mean vector (K parameters) and the correlation matrix ($K(K - 1)/2$ parameters). With the help of principal angle, the power of PC-based tests only depends on the K principal angles, K eigenvalues, and the overall signal strength. Hence, the complexity of power analysis for PC-based tests reduces from quadratic to linear order in the number of phenotypes. However, the principal angles are generally unknown in practical settings. One cannot choose a particular PC based on estimated principal angles and then use that cherry-picked PC for inference, because this approach will incur data snooping bias and the Type I error rate will be inflated. Actually, the proposed PCMinP test correctly adjusts for this cherry-picking process and provides a valid inference.

Effective combination of PCs for multiple phenotype genetic association testing depends on the K eigenvalues and the K unknown principal angles. We proposed linear, nonlinear, and adaptive omnibus combinations of PCs to achieve robust power. PCLC is an inverse-eigenvalue weighted linear combination of PCs and can be as powerful as the Oracle test when all the principal angles are equal to each other, but can lose power otherwise. In the worst case, PCLC is powerless when the signal vector is parallel to its rejection boundaries. PCMinP is expected

to perform well when there exists one principal angle equal to zero, but can lose power when the signal vector lies in the middle of all the PC directions. The PCFisher test combines all the mutually independent principal component p-values using Fisher's method, which can be more powerful than PCMinP when the signal vector lies in the middle of all the PCs but can be less powerful than PCMinP when some principal angle is equal to zero. We further proposed three quadratic combinations of PCs: WI, Wald, and VC. Surprisingly, the classical Wald test and the variance component score test using the linear mixed model framework (Huang and Lin 2013; Liu and Lin 2018) are two special cases of weighted quadratic combinations of PCs. Using convex optimization, we found that the Wald test achieves its maximal power when the last principal angle is zero and minimal power when the first principal angle is zero. The VC test is more powerful than the Wald test when the last principal angle is zero and can be less powerful otherwise. The WI test is more powerful than both the Wald and VC tests when the first principal angle is zero but can be less powerful otherwise. None of them is robust to the unknown principal angles.

The adaptive quadratic test PCAQ is more robust than the WI, Wald, and VC tests. As demonstrated by the simulation studies, PCMinP and PCLC can be more powerful than PCAQ in some situations. This suggests that an omnibus test that combines all these six tests together would be even more robust than PCAQ. The p-values of PCAQ and PCO can both be calculated analytically by numerical integration. This is advantageous when analyzing a large number of phenotypes with millions of SNPs across the whole genome, as the principle angles are likely to change from one SNP to another and a powerful test for one SNP might not be powerful for another SNP. All the proposed testing procedures have been implemented in a publicly available R package *MPAT*. The connections and subtle differences between those PC-based tests were illustrated graphically in terms of their rejection boundaries. The theoretical conditions under which each PC-based test can be more powerful than the traditional Wald test are as follows: the principal angles $\theta_k = 0$ for PC_k, PCMinP and PCO; $\cos^2(\theta_k) = 1/K$ for PCLC; $\theta_1 = 0$ for WI; $\theta_K = 0$ for VC; $\theta_1 = 0$; or $\theta_K = 0$ for PCAQ.

The eigen-analysis section investigates how the correlation structures among multiple phenotypes can influence the eigenvalues and eigenvectors of the correlation matrix, and subsequently affect the powers of PC-based tests. From eigen-analysis and simulation studies, we found that the PCO test can outperform the MinP and the TATES tests for the detection of sparse signals, especially when the dimension is high. The classical Wald test can perform poorly in high-dimension settings as discovered by our eigen-analysis and demonstrated empirically by simulation studies, whereas the omnibus test PCO can still have good power in those settings. The eigen-analysis highlights the importance of the correlation structures in affecting the powers of PC-based tests for detecting both sparse and dense signals. The eigen-analysis also shows that caution is needed for PC-based multiple phenotype analysis in the presence of highly correlated phenotypes. In such cases, the covariance matrix of multiple phenotypes is close to be singular, and the eigenvalues of the last few PCs are likely to be very small, making some tests that combine PCs, such as the Wald test, unstable. One can either remove some of highly correlated phenotypes before

performing multiple phenotype tests, or remove the last few PCs with very small eigenvalues before combining PCs to construct tests. For the former, one can select biologically meaningful phenotypes for a joint analysis in collaboration with domain scientists. At the same time, statistical consideration of power and numerical stability also should be taken into account when selecting phenotypes into analysis. Further research is needed on how to truncate PCs using selective inference theory (Choi, Taylor, and Tibshirani 2014) and then use those truncated PCs to construct valid and powerful tests by balancing the power and the numerical stability of the tests.

In this post-GWAS era, there are increasing amounts of GWAS summary statistics for multiple phenotypes publicly available on dbGAP (*http://www.ncbi.nlm.nih.gov/gap*) and other places. Our methods and software provide a cost-effective way to analyze such datasets to discover novel biology by borrowing information across multiple phenotypes. We demonstrated the usefulness of our methods by analyzing multiple metabolic syndrome-related clinical phenotypes with datasets collected from four international consortia. This real data example illustrates that the PCO test has robust power to detect additional novel loci underlying metabolic syndrome, outperforming the existing TATES method. It is of future research interest to apply our tests to higher-dimensional practical settings, for instance, in the studies of the genetic basis of gene expression levels or DNA methylation levels in a biological pathway/network when such datasets are available. When individual level data are available for both phenotypes and genotypes in the future, it would be practically interesting to compare the performances of our PC-based tests with other multiple phenotypes methods as discussed in Galesloot et al. (2014).

PCA is just one-dimension reduction method for transforming the correlated Z-statistics into uncorrelated ones using spectral-decomposition of the correlation matrix. There exist other methods for de-correlating correlated Z-statistics, such as the Cholesky decomposition. It would be interesting to explore the differences and connections between the testing statistics obtained from eigen-decomposition and Cholesky decomposition. With the increased availability of phenome-wide data, there will be a greater demand for analyzing multiple phenotypes in sequencing studies especially using electronic medical record data and molecular phenotype data. Since popular region-based association testing statistics for rare variants are not normally distributed, for example, the SKAT test statistic follows a mixture of chi-squared distributions (Lee, Wu, and Lin 2012), our current PC-based tests are not directly applicable for the detection of associations between rare variants and multiple phenotypes. More future work is needed to extend the current PCA framework for joint analysis of multiple phenotypes in GWAS to multiple phenotype analysis in sequencing association studies to detect rare variant effects.

Supplementary Materials

The supplementary pdf file contains simulation results and additional real data analysis results. The supplementary excel file contains the unstructured correlation matrix of dimension 100×100, and the $\boldsymbol{\beta}$ vectors of length 100 in simulation settings M11–M15, and the estimated principal angles for the 10 SNPs in Table 6.

References

Aschard, H., Vilhjálmsson, B. J., Greliche, N., Morange, P.-E., Trégouët, D.-A., and Kraft, P. (2014), "Maximizing the Power of Principal-Component Analysis of Correlated Phenotypes in Genome-wide Association Studies," *The American Journal of Human Genetics*, 94, 662–676. [2]

Bahadur, R. R. (1967), "Rates of Convergence of Estimates and Test Statistics," *The Annals of Mathematical Statistics*, 38, 303–324. [4]

Bickel, P. J., and Levina, E. (2004), "Some Theory for Fisher's Linear Discriminant Function, 'Naive Bayes', and Some Alternatives When There are Many More Variables than Observations," *Bernoulli*, 10, 989–1010. [3]

Birnbaum, A. (1954), "Combining Independent Tests of Significance," *Journal of the American Statistical Association*, 49, 559–574. [3,7]

—— (1955), "Characterizations of Complete Classes of Tests of Some Multiparametric Hypotheses, with Applications to Likelihood Ratio Tests," *The Annals of Mathematical Statistics*, 26, 21–36. [3,7]

Bittman, R. M., Romano, J. P., Vallarino, C., and Wolf, M. (2009), "Optimal Testing of Multiple Hypotheses with Common Effect Direction," *Biometrika*, 96, 399–410. [3]

Brown, A. E., and Walker, M. (2016), "Genetics of Insulin Resistance and the Metabolic Syndrome," *Current Cardiology Reports*, 18, 75. [1]

Bush, W. S., Oetjens, M. T., and Crawford, D. C. (2016), "Unravelling the Human Genome-Phenome Relationship using Phenome-wide Association Studies," *Nature Reviews Genetics*, 17, 129–145. [1]

Choi, Y., Taylor, J., and Tibshirani, R. (2014), "Selecting the Number of Principal Components: Estimation of the True Rank of a Noisy Matrix," *The Annals of Statistics*, 45, 2590–2617. arXiv:1410.8260. [15]

Collins, F. S., and Varmus, H. (2015), "A New Initiative on Precision Medicine," *New England Journal of Medicine*, 372, 793–795. [1]

Conneely, K. N., and Boehnke, M. (2007), "So many Correlated Tests, so Little Time! Rapid Adjustment of P values for Multiple Correlated Tests," *The American Journal of Human Genetics*, 81, 1158–1168. [11]

Davies, R. B. (1980), "Algorithm AS 155: The Distribution of a Linear Combination of Chi-Squared Random Variables," *Journal of the Royal Statistical Society*, Series C, 29, 323–333. [5,6]

Ferreira, M. A. R., and Purcell, S. M. (2009), "A Multivariate Test of Association," *Bioinformatics*, 25, 132–133. [2]

Fisher, R. A. S. (1932), *Statistical Methods for Research Workers* (Vol. 5, 7th ed., *Biological Monographs and Manuals*), Edinburgh, London: Oliver and Boyd. [4]

—— (1936), "The use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, 7, 179–188. [3]

Galesloot, T. E., van Steen, K., Kiemeney, L. A., Janss, L. L., and Vermeulen, S. H. (2014), "A Comparison of Multivariate Genome-Wide Association Methods," *PloS one*, 9, e95923. [15]

Genz, A. (1992), "Numerical Computation of Multivariate Normal Probabilities," *Journal of Computational and Graphical Statistics*, 1, 141–149. [6]

—— (1993), "Comparison of Methods for the Computation of Multivariate Normal Probabilities," *Computing Science and Statistics*, 25, 400–400. [6]

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2009), "mvtnorm: Multivariate Normal and t Distributions," *R Package Version 0.9-8*, available at *http://CRAN. R-project. org/package= mvtnorm* [6]

He, Q., Avery, C. L., and Lin, D.-Y. (2013), "A General Framework for Association Tests with Multivariate Traits in Large-Scale Genomics Studies," *Genetic Epidemiology*, 37, 759–767. [2,9]

Huang, Y.-T., and Lin, X. (2013), "Gene Set Analysis using Variance Component Tests," *BMC Bioinformatics*, 14, 210. [2,6,14]

ICBP et al. (2011), "Genetic Variants in Novel Pathways Influence Blood Pressure and Cardiovascular Disease Risk," *Nature*, 478, 103–109. [1]

Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Components Analysis," *Annals of statistics*, 29, 295–327. [8]

Kraja, A. T., Chasman, D. I., North, K. E., Reiner, A. P., Yanek, L. R., Kilpeläinen, T. O., Smith, J. A., Dehghan, A., Dupuis, J., Johnson, A. D., et al. (2014), "Pleiotropic Genes for Metabolic Syndrome and Inflammation," *Molecular Genetics and Metabolism*, 112, 317–338. [13]

Lee, S., Wu, M. C., and Lin, X. (2012), "Optimal Tests for Rare Variant Effects in Sequencing Association Studies," *Biostatistics*, 13, 762–775. [15]

Liptak, T. (1958), "On the Combination of Independent Tests," *Magyar Tud Akad Mat Kutato Int Kozl*, 3, 171–197. [5]

Littell, R. C., and Folks, J. L. (1971), "Asymptotic Optimality of Fisher's Method of Combining Independent Tests," *Journal of the American Statistical Association*, 66, 802–806. [4]

——— (1973), "Asymptotic Optimality of Fisher's Method of Combining Independent Tests II," *Journal of the American Statistical Association*, 68, 193–194. [4]

Liu, D., Lin, X., and Ghosh, D. (2007), "Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models," *Biometrics*, 63, 1079–1088. [5]

Liu, Z., and Lin, X. (2018), "Multiple Phenotype Association Tests using Summary Statistics in Genome-wide Association Studies," *Biometrics*, 74, 165–175. [2,3,6,12,14]

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., et al. (2015), "Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology," *Nature*, 518, 197–206. [2]

Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I., et al. (2012), "A Genome-Wide Approach Accounting for Body Mass Index Identifies Genetic Variants Influencing Fasting Glycemic Traits and Insulin Resistance," *Nature Genetics*, 44, 659–669. [1]

Marsaglia, G., and Olkin, I. (1984), "Generating Correlation Matrices," *SIAM Journal of Statistical Computation*, 5, 470–475. [9]

Mosteller, F., Bush, R. R., and Green, B. F. (1954), *Selected Quantitative Techniques*, Cambridge, MA: Addison-Wesley. [5]

Pettofrezzo, A. J. (1978), *Matrices and Transformations*, New York: Courier Corporation. [3]

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C. (2007), "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses," *The American Journal of Human Genetics*, 81, 559–575. [12]

Schifano, E. D., Li, L., Christiani, D. C., and Lin, X. (2013), "Genome-Wide Association Analysis for Multiple Continuous Secondary Phenotypes," *The American Journal of Human Genetics*, 92, 744–759. [2]

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., et al. (2015), "New Genetic loci Link Adipose and Insulin Biology to Body Fat Distribution," *Nature*, 518, 187–196. [2]

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J. F., and Campbell, H. (2011), "Abundant Pleiotropy in Human Complex Diseases and Traits," *The American Journal of Human Genetics*, 89, 607–618. [1]

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., and Smoller, J. W. (2013), "Pleiotropy in Complex Traits: Challenges and Strategies," *Nature Reviews Genetics*, 14, 483–495. [1,2]

Sotoodehnia, N., Isaacs, A., De Bakker, P. I., Dörr, M., Newton-Cheh, C., Nolte, I. M., Van Der Harst, P., Müller, M., Eijgelsheim, M., Alonso, A., et al. (2010), "Common Variants in 22 loci are Associated with QRS Duration and Cardiac Ventricular Conduction," *Nature Genetics*, 42, 1068–1076. [12]

Su, J. Q., and Liu, J. S. (1993), "Linear Combinations of Multiple Diagnostic Markers," *Journal of the American Statistical Association*, 88, 1350–1355. [3]

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010), "Biological, Clinical and Population Relevance of 95 Loci for Blood Lipids," *Nature*, 466, 707–713. [1,9]

van der Sluis, S., Posthuma, D., and Dolan, C. V. (2013), "TATES: Efficient Multivariate Genotype-Phenotype Analysis for Genome-Wide Association Studies," *PLoS Genetics*, 9, e1003235. [2,3,9,12]

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014), "The NHGRI GWAS Catalog, a Curated Resource of SNP-trait Associations," *Nucleic Acids Research*, 42, D1001–D1006. [1]

Willer, C. J., Consortium, G. L. G., et al. (2013), "Discovery and Refinement of loci Associated with Lipid Levels," *Nature Genetics*, 45, 1274–1283. [1,12]

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., de Bakker, P. I., Abecasis, G. R., Almgren, P., Andersen, G., et al. (2008), "Meta-Analysis of Genome-Wide Association Data and Large-Scale Replication Identifies Additional Susceptibility loci for Type 2 Diabetes," *Nature Genetics*, 40, 638–645. [13]

Zhou, X., and Stephens, M. (2014), "Efficient Multivariate Linear Mixed Model Algorithms for Genome-Wide Association Studies," *Nature Methods*, 11, 407–409. [2]

Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., Smith, J. A., Yanek, L. R., Sun, Y. V., Edwards, T. L., et al. (2015), "Meta-analysis of Correlated Traits via Summary Statistics from GWASs with an Application in Hypertension," *The American Journal of Human Genetics*, 96, 21–36. [2,3,12]