

A COMPARISON OF PRINCIPAL COMPONENT METHODS BETWEEN MULTIPLE PHENOTYPE REGRESSION AND MULTIPLE SNP REGRESSION IN GENETIC ASSOCIATION STUDIES

BY ZHONGHUA LIU¹, IAN BARNETT² AND XIHONG LIN³

¹*Department of Statistics and Actuarial Science, The University of Hong Kong, zhliu@hku.hk*

²*Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, ibarnett@penncmedicine.upenn.edu*

³*Department of Biostatistics and Statistics, Harvard University, xlin@hsph.harvard.edu*

Principal component analysis (PCA) is a popular method for dimension reduction in unsupervised multivariate analysis. However, existing ad hoc uses of PCA in both multivariate regression (multiple outcomes) and multiple regression (multiple predictors) lack theoretical justification. The differences in the statistical properties of PCAs in these two regression settings are not well understood. In this paper we provide theoretical results on the power of PCA in genetic association testings in both multiple phenotype and SNP-set settings. The multiple phenotype setting refers to the case when one is interested in studying the association between a single SNP and multiple phenotypes as outcomes. The SNP-set setting refers to the case when one is interested in studying the association between multiple SNPs in a SNP set and a single phenotype as the outcome. We demonstrate analytically that the properties of the PC-based analysis in these two regression settings are substantially different. We show that the lower order PCs, that is, PCs with large eigenvalues, are generally preferred and lead to a higher power in the SNP-set setting, while the higher-order PCs, that is, PCs with small eigenvalues, are generally preferred in the multiple phenotype setting. We also investigate the power of three other popular statistical methods, the Wald test, the variance component test and the minimum p -value test, in both multiple phenotype and SNP-set settings. We use theoretical power, simulation studies, and two real data analyses to validate our findings.

1. Introduction. In view of the rapid progression of massive data in genetic and genomic studies, developing statistical machinery to effectively leverage the correlation present in the data to boost analysis power becomes increasingly important. In this paper we focus on jointly testing multiple correlated quantities in two important settings which frequently arise in genetic association studies: *the SNP-set setting*, which tests for the association between a single phenotype and a set of single nucleotide polymorphisms (SNPs) in a genetic construct, for example, gene or network (Li and Leal (2008), Wu et al. (2011)), and *the multiple phenotype setting*, which tests for the association between a single SNP and multiple phenotypes, that is, pleiotropy (Aschard et al. (2014), Solovieff et al. (2013)). Despite having a similar set of statistical methods developed separately in each case, we explore the subtle but important differences in the methodology across the two cases. Specifically, principal component analysis (PCA) is a popular unsupervised dimension reduction method frequently used in both settings. However, little is known about how the performance of PCA differs in multiple phenotype regression as compared to multiple SNP regression, and when a PCA should be used in practice in each case as opposed to other methods, as well as which PC should be used. We investigate this problem in this paper.

Received June 2018; revised November 2019.

Key words and phrases. Dimension reduction, principal component analysis, eigen-values, hypothesis testing, multiple phenotypes, minimum p -value test, SNP-set, variance-component test.

There is an increasing interest to analyze multiple phenotypes jointly in genetic association studies to detect pleiotropic effects, as the power gain of joint analysis of multiple phenotypes has been demonstrated by several previous studies (Liu and Lin (2018, 2019), Schifano et al. (2013), Solovieff et al. (2013), Stephens (2013), Zhou and Stephens (2014), Zhu et al. (2015)). PCA has been used as a popular tool for dimension reduction in multiple phenotype analysis (Aschard et al. (2014), Karasik et al. (2012), Suo et al. (2013), Zhang et al. (2012), Liu and Lin (2019)). Aschard et al. (2014) found that combining multiple related phenotypes through their principal components identified significant top SNPs that were missed by individual phenotype association tests. Liu and Lin (2019) further proposed a series of PC based association tests for multiple phenotype studies using summary statistics from genome-wide association studies (GWASs).

In the case of a univariate phenotype outcome, SNPs in a genetic construct, such as gene or pathway, can be tested jointly which boosts analysis power by combining effects across multiple SNPs while also reducing the multiple testing burden. While true for GWASs, SNP-set testing is more important in sequencing studies where rare variant associations would be near impossible to detect without grouping signals within a SNP-set (Lee et al. (2014), Li and Leal (2008)). Due to linkage disequilibrium (LD), it is frequently the case that adjacent SNPs are in LD and correlated, complicating their joint analysis. There have been numerous statistical techniques developed for the purpose of properly accounting for LD/correlation when testing multiple SNPs simultaneously (Conneely and Boehnke (2007), Han, Kang and Eskin (2009), Moskvina and Schmidt (2008), Wu et al. (2011)). Dimension reduction can be a useful tool in these settings particularly for larger SNP-sets, and principal component regression is a commonly used dimension reduction approach for removing correlation among SNPs in a SNP-set, for example, a gene (Wang and Abbott (2008), Karacaören et al. (2011)). For example, in Karacaören et al. (2011), principal component regression was used in a QTL analysis to account for the LD between significant markers in a set of SNPs that was regressed on gene expression.

Despite its frequent usage throughout genetic association studies, there is little theoretical analysis on when using PCs is beneficial and when it is not in regression analysis, and how its performance differs in the multiple phenotype setting as compared to the multiple SNP setting. This renders PCA at the risk of potential power loss if used inappropriately. In this paper we investigate the circumstances under which PCA is effective in both the SNP-set and multiple phenotype settings. In addition, we address two other popular approaches in genetic epidemiology, variance component tests and min p -value tests. Upon first glance, all of these statistical methods have strong similarities across the two settings, but we will illuminate some crucial statistical differences.

In particular, we find that in the SNP-set setting, principal components with large eigenvalues tend to have increased power, whereas the opposite holds true in the multiple phenotype setting. The power of these methods are compared both theoretically and through simulation. In addition, we apply the concepts developed in this paper to both a lipids GWAS, which studies how a SNP affects multiple lipids, as well as a Breast Cancer GWAS by performing a gene-based analysis to study how the SNPs in a gene affect breast cancer risk. Henceforth, we refer to the SNP-set setting as “1:K” with reference its one outcome and K predictors, and we similarly refer to the multiple phenotype setting as “K:1” for its K outcomes and one predictor. In the 1:K setting, K usually represents the number of SNPs in a gene, while in the K:1 setting, K represents the number of phenotypes and is usually chosen based on the scientific question under study.

Section 2 compares the theoretical power of principal component tests in both the K:1 and 1:K settings. Section 3 discusses the crucial role of principal components in variance component tests, while again highlighting the stark differences between the K:1 and 1:K

settings. Section 4 demonstrates these differences between the 1:K and K:1 settings through a series of power simulations. Finally, Section 5 shows these differences by comparing the analysis of two real data sets, one in the K:1 multiple phenotype setting (lipids GWAS) and one in the 1:K multiple SNP setting (breast cancer GWAS).

2. The power of PCA.

2.1. *The K:1 multiple phenotype setting.* In the K:1 multiple phenotype setting, we consider testing the association between a particular SNP and a group of correlated phenotypes, that is, study pleiotropy. Suppose that we have measured K continuous phenotypes on n genotyped individuals. We first consider a model without covariates and then discuss the extension of the results to the covariate case. For the i th individual, we can use the following model to link a particular genotype to the k th phenotype:

$$(2.1) \quad Y_{ik}^{(P)} = \beta_k^{(P)} G_i^{(P)} + \epsilon_{ik}^{(P)},$$

where $\epsilon_i^{(P)} = \{\epsilon_{i1}^{(P)}, \dots, \epsilon_{iK}^{(P)}\}^T \sim N(\mathbf{0}, \Sigma^{(P)})$, $Y_{ik}^{(P)}$ is the k th phenotype for the i th subject, $G_i^{(P)}$ is a scalar representing the genotype of a SNP of interest for the i th subject and $\beta_k^{(P)}$ is the regression coefficient for the genotype and the k th phenotype. We use the superscript (P) to indicate this setting represents multiple phenotype regression and the K:1 setting. For notation simplicity, we assume that the $Y_{ik}^{(P)}$ have been standardized and the $G_i^{(P)}$ have been centered and scaled such that $\sum_{i=1}^n \{G_i^{(P)}\}^2 = 1$. Denote $\beta^{(P)} = \{\beta_1^{(P)}, \dots, \beta_K^{(P)}\}^T$, $\mathbf{Y}_k^{(P)} = \{Y_{1k}^{(P)}, \dots, Y_{nk}^{(P)}\}^T$, and $\mathbf{G}^{(P)} = \{G_1^{(P)}, \dots, G_n^{(P)}\}^T$. To assess whether this particular SNP is associated with those K phenotypes, we can formulate this problem as the following hypothesis testing:

$$H_0 : \beta^{(P)} = 0 \quad \text{vs} \quad H_1 : \beta^{(P)} \neq 0.$$

Consider the following score statistics:

$$\mathbf{Z}^{(P)} = \{Z_1^{(P)}, Z_2^{(P)}, \dots, Z_K^{(P)}\}^T,$$

where $Z_k^{(P)} = \{\mathbf{G}^{(P)}\}^T \mathbf{Y}_k^{(P)}$ and its distribution is $Z_k^{(P)} \sim N\{\mu_k^{(P)}, 1\}$ where $\mu_k^{(P)} = \beta_k^{(P)}$. For each fixed genetic variant, we have a random vector $\mathbf{Z}^{(P)} = \{Z_1^{(P)}, \dots, Z_K^{(P)}\}^T \sim N(\beta^{(P)}, \Sigma^{(P)})$, where $\mu^{(P)} = \beta^{(P)}$ is defined as the signal vector and does not depend on the correlation matrix. In what follows, we assume that the correlation matrix $\Sigma^{(P)}$ of $\mathbf{Z}^{(P)}$ is known or can be consistently estimated. As sample sizes in GWAS are often large, the correlation matrix $\Sigma^{(P)}$ can be estimated very accurately using the GWAS phenotype-specific summary statistics and the bias and variance of the estimated $\Sigma^{(P)}$ have little effects on inference for $\beta^{(P)}$, and $\Sigma^{(P)}$ does not vary across SNPs (Liu and Lin (2018), Zhu et al. (2015)).

This correlation matrix $\Sigma^{(P)}$ can be decomposed as

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

where $\mathbf{\Lambda}^{(P)}$ is a diagonal matrix $\text{diag}\{\lambda_1^{(P)}, \dots, \lambda_K^{(P)}\}$ whose elements represent the eigenvalues of $\Sigma^{(P)}$ and $\lambda_1^{(P)} \geq \lambda_2^{(P)} \geq \dots \geq \lambda_K^{(P)} > 0$, and $\mathbf{U}^{(P)}$ is the normalized orthogonal matrix whose k th column $\mathbf{u}_k^{(P)}$ represents the k th eigenvector associated with the k th largest eigenvalue. Then, the distribution of the portion of $\mathbf{Z}^{(P)}$ along the k th principal component is given by

$$V_k^{(P)} = \{\mathbf{u}_k^{(P)}\}^T \mathbf{Z}^{(P)} \sim N\{(\mathbf{u}_k^{(P)})^T \beta^{(P)}, \lambda_k^{(P)}\}, \quad 1 \leq k \leq K.$$

There is a relationship between the principal components of the score test statistics $\mathbf{Z}^{(P)}$ and the principal components of the K phenotypes: Let $\mathbf{Y}^{(P)} = (\mathbf{Y}_1^{(P)}, \dots, \mathbf{Y}_K^{(P)})$ be the $n \times K$ phenotype matrix and note that $V_k^{(P)} = (\mathbf{Y}^{(P)} \mathbf{u}_k^{(P)})^T \mathbf{G}^{(P)} = \{\text{PC}_k^{(P)}\}^T \mathbf{G}^{(P)}$, where $\text{PC}_k^{(P)} = \mathbf{Y}^{(P)} \mathbf{u}_k^{(P)}$ is the k th principal component of the K phenotypes. This suggests that if one performs PCA analysis of the K phenotypes and transforms the K phenotypes $(\mathbf{Y}_1^{(P)}, \dots, \mathbf{Y}_K^{(P)})$ to K independent phenotype PCs $\{\text{PC}_1^{(P)}, \dots, \text{PC}_K^{(P)}\}$, then $V_k^{(P)}$ is equal to the score statistic obtained by regressing the k th phenotype $\text{PC}_k^{(P)}$ on the genotype $\mathbf{G}^{(P)}$. This implies that whether a PC decomposition is done before or after testing, the resulting inference is the same.

Using the fact that $\|\mathbf{u}_k^{(P)}\|^2 = 1$, the noncentrality parameter (ncp) of the k th phenotype PC $V_k^{(P)}$ under the alternative hypothesis is

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

where $\theta_k^{(P)}$ is the k th principal angle defined as the angle between $\boldsymbol{\beta}^{(P)}$ and the eigenvector $\mathbf{u}_k^{(P)}$ first introduced by Liu and Lin (2019). Because the k th eigenvalue enters into the denominator of the k th ncp, excluding the effect of $\theta_k^{(P)}$, the first PC test statistic $V_1^{(P)}$ has the smallest ncp compared to the other PCs.

If we use $V_k^{(P)}$ as a testing statistic for the hypothesis $H_0 : \boldsymbol{\beta}^{(P)} = 0$ vs. $H_a : \boldsymbol{\beta}^{(P)} \neq 0$, then its theoretical power at significance level α is

$$\begin{aligned} \text{pr}(|V_k^{(P)}| > z_{1-\frac{\alpha}{2}} | H_a) &= \Phi\left(z_{\frac{\alpha}{2}} + \frac{\|\boldsymbol{\beta}^{(P)}\| \cdot |\cos(\theta_k^{(P)})|}{\sqrt{\lambda_k^{(P)}}}\right) \\ (2.2) \qquad \qquad \qquad &+ \Phi\left(z_{\frac{\alpha}{2}} - \frac{\|\boldsymbol{\beta}^{(P)}\| \cdot |\cos(\theta_k^{(P)})|}{\sqrt{\lambda_k^{(P)}}}\right), \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable, and z_α is its α percentile. The power of this test depends on both the k th principal angle and the k th eigenvalue. Specifically, the test statistic $V_k^{(P)}$ is powerful if $\theta_k^{(P)} = 0$ but is powerless if $\theta_k^{(P)} = 90^\circ$ (i.e., when $\mathbf{u}_k^{(P)}$ is orthogonal to $\boldsymbol{\beta}^{(P)}$), and the phenotype PCs with smaller eigenvalues, that is, $V_k^{(P)}$ with larger k , are more likely to be powerful. This geometric perspective offered by equation (2.2) demonstrates why the test based on the first PC, without taking the true direction of $\boldsymbol{\beta}$ into account, will have the least power of all the K PCs in the K:1 setting due to it having the largest eigenvalue.

2.2. The 1:K SNP-set setting. The difficulty of detecting single SNP effects in GWAS due to the tremendous multiple comparisons problem and weak SNP effects can be alleviated by changing the unit of analysis from a single SNP to meaningful groupings of SNPs. The most common choice of these groupings, or SNP-sets, is to combine the analysis of SNPs in the same gene. The number of SNPs in most GWAS is approximately 20 times larger than the number of genes, and therefore the multiple comparisons burden is greatly reduced by analyzing SNPs grouped by gene membership. Another benefit of analyzing SNP-set is the power gain by aggregating weak signal effects of multiple SNPs in a set, which is now becoming standard for rare variant association testings (Lee et al. (2014)).

A strong effort has been made to develop statistical methods for association testing in the SNP-set context. One primary difficulty is that often SNPs within the same gene are in LD.

One popular approach uses a principal component decomposition in order to orthogonalize the genotype matrix. We outline and analyze this approach more closely here.

Consider the 1:K linear model, where a scalar phenotype $Y_i^{(S)}$ is regressed on K SNPs in a SNP set $\mathbf{G}_i^{(S)} = \{G_{i1}^{(S)}, \dots, G_{iK}^{(S)}\}^T$ as

$$(2.3) \quad Y_i^{(S)} = \beta_1^{(S)} G_{i1}^{(S)} + \dots + \beta_K^{(S)} G_{iK}^{(S)} + \epsilon_i^{(S)},$$

where $\epsilon_i^{(S)} \sim N\{0, (\sigma^{(S)})^2\}$. We set $\mathbf{Y}^{(S)} = \{Y_1^{(S)}, \dots, Y_n^{(S)}\}^T$ to be the standardized phenotype vector and $\mathbf{G}^{(S)}$ to be the n by K matrix of genotypes with columns centered and scaled such that $\sum_{i=1}^n \{G_{ik}^{(S)}\}^2 = 1$ for each k . We here use the superscript (S) to indicate this setting represents 1:K SNP-set regression. Our goal is to test whether a group of SNPs $\mathbf{G}_i^{(S)}$ is associated with the scalar phenotype $Y_i^{(S)}$, that is, to test the null hypothesis $H_0 : \boldsymbol{\beta}^{(S)} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta}^{(S)} \neq \mathbf{0}$, where $\boldsymbol{\beta}^{(S)} = \{\beta_1^{(S)}, \dots, \beta_K^{(S)}\}^T$.

Based on model (2.3), the score test statistic for $\boldsymbol{\beta}^{(S)}$ is $\mathbf{Z}^{(S)} = \{\mathbf{G}^{(S)}\}^T \mathbf{Y}^{(S)} \sim N\{\boldsymbol{\mu}^{(S)}, \boldsymbol{\Sigma}^{(S)}\}$, where $\boldsymbol{\Sigma}^{(S)} = \{\mathbf{G}^{(S)}\}^T \mathbf{G}^{(S)}$ is a $K \times K$ correlation matrix of the K SNPs in a SNP set, and $\boldsymbol{\mu}^{(S)} = \boldsymbol{\Sigma}^{(S)} \boldsymbol{\beta}^{(S)}$. Note the covariance matrix $\boldsymbol{\Sigma}^{(S)}$ appears in both the mean and the covariance of $\mathbf{Z}^{(S)}$, while $\boldsymbol{\Sigma}^{(P)}$ appears only in the covariance of $\mathbf{Z}^{(P)}$.

Using spectral decomposition, we have

$$\boldsymbol{\Sigma}^{(S)} = \mathbf{U}^{(S)} \boldsymbol{\Lambda}^{(S)} \{\mathbf{U}^{(S)}\}^T,$$

where $\mathbf{U}^{(S)}$ is the eigenvector matrix whose k th column $\mathbf{u}_k^{(S)}$ represents the k th eigenvector of $\boldsymbol{\Sigma}^{(S)}$ and $\boldsymbol{\Lambda}^{(S)} = \text{diag}\{\lambda_1^{(S)}, \dots, \lambda_K^{(S)}\}$ is a diagonal matrix whose diagonal elements are the ordered eigenvalues of $\boldsymbol{\Sigma}^{(S)}$. Define $V_k^{(S)}$ as the portion of $\mathbf{Z}^{(S)}$ along the k th principal component,

$$(2.4) \quad V_k^{(S)} = \{\mathbf{u}_k^{(S)}\}^T \mathbf{Z}^{(S)} = \{\mathbf{G}^{(S)} \mathbf{u}_k^{(S)}\}^T \mathbf{Y}^{(S)} \sim N[\lambda_k^{(S)} \{\mathbf{u}_k^{(S)}\}^T \boldsymbol{\beta}^{(S)}, \lambda_k^{(S)}].$$

In fact, suppose we first perform principal component analysis on the genotype matrix $\mathbf{G}^{(S)}$ and obtain $\text{PC}_{ik}^{(S)} = \{\mathbf{G}_i^{(S)}\}^T \mathbf{u}_k^{(S)}$, which is the k th principal component of the SNP-set of the i th subject. Then, regress $\mathbf{Y}^{(S)}$ on these K orthogonal genotype principal components as

$$(2.5) \quad Y_i^{(S)} = \gamma_1^{(S)} \text{PC}_{i1}^{(S)} + \dots + \gamma_K^{(S)} \text{PC}_{iK}^{(S)} + \epsilon_i^*,$$

where $\epsilon_i^* \sim N(0, \sigma^{*2})$. Then testing whether the set of SNPs $\mathbf{G}_i^{(S)}$ is associated with the scalar phenotype $Y_i^{(S)}$, that is, testing the null hypothesis $H_0 : \boldsymbol{\beta}^{(S)} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta}^{(S)} \neq \mathbf{0}$ is equivalent to testing the null hypothesis $H_0 : \boldsymbol{\gamma}^{(S)} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\gamma}^{(S)} \neq \mathbf{0}$, where $\boldsymbol{\gamma}^{(S)} = \{\gamma_1^{(S)}, \dots, \gamma_K^{(S)}\}^T$. In other words, inference is the same for SNP-principal component regression as it is for the score tests on the original scale.

The power of using $V_k^{(S)}$ as a testing statistic is

$$(2.6) \quad \begin{aligned} \text{pr}\{|V_k^{(S)}| > z_{1-\frac{\alpha}{2}}\} &= \Phi(z_{\frac{\alpha}{2}} + \sqrt{\lambda_k^{(S)}} \{\mathbf{u}_k^{(S)}\}^T \boldsymbol{\beta}^{(S)}) \\ &+ \Phi(z_{\frac{\alpha}{2}} - \sqrt{\lambda_k^{(S)}} \{\mathbf{u}_k^{(S)}\}^T \boldsymbol{\beta}^{(S)}). \end{aligned}$$

Analogous to the K:1 multiple phenotype regression setting, we have

$$\{\mathbf{u}_k^{(S)}\}^T \boldsymbol{\beta}^{(S)} = \|\boldsymbol{\beta}^{(S)}\| \cos(\theta_k^{(S)}),$$

where $\theta_k^{(S)}$ is the principal angle between the k th eigenvector of $\boldsymbol{\Sigma}^{(S)}$ and the effect vector $\boldsymbol{\beta}^{(S)}$. If $\theta_k^{(S)} = 90^\circ$, then $V_k^{(S)}$ is powerless, and if $\theta_k^{(S)} = 0$, then $V_k^{(S)}$ is powerful. This implies

that a larger eigenvalue $\lambda_k^{(S)}$ leads to larger power, and the power is highest when $\theta_1^{(S)} = 0$. This is the opposite of the conclusion found in the K:1 setting. This result shows that, for 1:K SNP-set regression, the PCs of SNPs with large eigenvalues have more power for detecting genetic association.

2.3. *Differences and connections between 1:K and K:1 problems.* In both 1:K multiple SNP regression and K:1 multiple phenotype settings, we aim to test for the overall association between one variable and K variables. In the 1:K SNP set setting, the PCA analysis is conducted on the genotype matrix to obtain orthogonal PC genotype scores, and then we can regress the scalar phenotype variable on those genotype PCs jointly. In the K:1 multiple phenotype setting, we first apply PCA on the multiple phenotypes and obtain orthogonal phenotype PCs and then regress each phenotype PC on the single SNP genotype variable.

There is an important difference when applying PCA in these two settings: correlation between explanatory variables has a different effect on inference than does correlation between outcome variables. In the K:1 setting, the K variables are multiple phenotypes which are put on the left-hand side of the regression model as outcomes, and the one explanatory variable is the SNP genotype which is put on the right-hand side of the regression model. In contrast, in the 1:K setting there are K explanatory variables, the multiple SNPs in the SNP-set, which are put on the right hand-side of the regression, while the one phenotype is still on the only outcome on the left-hand side of the regression model.

This difference leads to an asymmetry between the two settings. In the K:1 multiple phenotype regression setting, the correlation matrix among multiple phenotypes influences only the correlation matrix of the score testing statistics and has no influence on its mean vector, as the test statistic is $\mathbf{Z}^{(P)} \sim N\{\boldsymbol{\beta}^{(S)}, \boldsymbol{\Sigma}^{(P)}\}$. However, in the 1:K SNP-set regression setting the correlation matrix (LD) among the K SNPs influences both the mean vector and correlation matrix of the score testing statistics, as the test statistic is $\mathbf{Z}^{(S)} \sim N\{\boldsymbol{\Sigma}^{(S)}\boldsymbol{\beta}^{(S)}, \boldsymbol{\Sigma}^{(S)}\}$. This occurs because the Gaussian likelihood accounts for correlation between multivariate outcomes differently than it does for correlation between multiple explanatory variables, leading to the differences we see in the score equations and corresponding test statistics.

2.4. *Extension to allow for covariate adjustment.*

2.4.1. *The K:1 setting.* In many cases covariates are included in the model to adjust for potential confounding factors. Suppose we denote the q -dimensional centered covariate row vector for individual i as $\mathbf{X}_i^{(P)}$, then we can model the relationship between genotype and phenotype adjusted for covariates as

$$(2.7) \quad Y_{ik}^{(P)} = \mathbf{X}_i^{(P)} \boldsymbol{\alpha}_k^{(P)} + G_i^{(P)} \beta_k^{(P)} + \epsilon_{ik}^{(P)},$$

where $\boldsymbol{\alpha}_k^{(P)} = \{\alpha_{k1}, \dots, \alpha_{kq}\}^T$ are the covariate regression coefficients. Under $H_0 : \boldsymbol{\beta}^{(P)} = 0$, the score test statistics are

$$(2.8) \quad Z_k^{(P)} = (\mathbf{G}^{(P)})^T (\mathbf{Y}_k^{(P)} - \mathbf{X}^{(P)} \hat{\boldsymbol{\alpha}}_k^{(P)}),$$

where $\hat{\boldsymbol{\alpha}}_k^{(P)}$ is the maximum likelihood estimator under the null model H_0 . Using matrix notation, we have

$$(2.9) \quad Z_k^{(P)} = (\mathbf{G}^{(P)})^T [\mathbf{I} - \mathbf{X}^{(P)} ((\mathbf{X}^{(P)})^T \mathbf{X}^{(P)})^{-1} (\mathbf{X}^{(P)})^T] \mathbf{Y}_k^{(P)}.$$

Denote $\mathbf{a}^{(P)} = (\mathbf{G}^{(P)})^T [\mathbf{I} - \mathbf{X}^{(P)} ((\mathbf{X}^{(P)})^T \mathbf{X}^{(P)})^{-1} (\mathbf{X}^{(P)})^T]$ to be the row vector of length n , then we can rewrite equation (2.9) as $Z_k^{(P)} = \mathbf{a}^{(P)} \mathbf{Y}_k^{(P)}$, which reduces to the setting in Section 2.1.

2.4.2. *The 1:K setting.* To adjust for covariates in equation (2.3), we use $\boldsymbol{\alpha}^{(S)}$ as a column vector of q regression coefficients for the row vector $\mathbf{X}_i^{(S)}$ of q covariates

$$(2.10) \quad Y_i^{(S)} = \mathbf{X}_i^{(S)} \boldsymbol{\alpha}^{(S)} + \mathbf{G}_i^{(S)} \boldsymbol{\beta}^{(S)} + \epsilon_i^{(S)}.$$

The score test statistics are

$$(2.11) \quad \mathbf{Z}^{(S)} = (\mathbf{G}^{(S)})^T (\mathbf{Y}^{(S)} - \mathbf{X}^{(S)} \hat{\boldsymbol{\alpha}}^{(S)}),$$

where $\hat{\boldsymbol{\alpha}}^{(S)}$ is the maximum likelihood estimator under the null model $H_0 : \boldsymbol{\beta}^{(S)} = 0$. Similar to the K:1 setting, denote

$$\mathbf{a}^{(S)} = (\mathbf{G}^{(S)})^T [\mathbf{I} - \mathbf{X}^{(S)} ((\mathbf{X}^{(S)})^T \mathbf{X}^{(S)})^{-1} (\mathbf{X}^{(S)})^T]$$

to be the row vector of length n , then we can rewrite equation (2.9) as $\mathbf{Z}^{(S)} = \mathbf{a}^{(S)} \mathbf{Y}^{(S)}$, which can be treated in the same fashion of Section 2.2 and similar analysis can proceed.

3. The role of PCs in variance component tests. Principal components play a similar role in variance component score tests in linear mixed models. In the K:1 setting, if $\boldsymbol{\beta}^{(P)}$ from the multiple phenotype regression model (2.1) is such that each β_j is assumed to follow an arbitrary distribution with mean 0 and variance $\tau^{(P)}$, where $\tau^{(P)}$ is the variance component, then testing for $H_0 : \boldsymbol{\beta}^{(P)} = \mathbf{0}$ is equivalent testing for $H_0 : \tau^{(P)} = 0$ (Lin (1997)). Based on model (2.1), the variance component score test statistic for $H_0 : \tau^{(P)} = 0$ is (Huang and Lin (2013))

$$Q^{(P)} = \{\mathbf{Z}^{(P)}\}^T \{\boldsymbol{\Sigma}^{(P)}\}^{-1} \{\boldsymbol{\Sigma}^{(P)}\}^{-1} \mathbf{Z}^{(P)}.$$

This can also be rewritten as

$$(3.1) \quad \begin{aligned} Q^{(P)} &= \sum_{k=1}^K [\{\mathbf{Y}^{(P)} \mathbf{u}_k^{(P)}\}^T \mathbf{G}^{(P)} / \lambda_k^{(P)}]^2 \\ &= \sum_{k=1}^K [\mathbf{G}^T \{\mathbf{PC}_k^{(P)}\} / \lambda_k^{(P)}]^2 \\ &= \sum_{k=1}^K \{V_k^{(P)} / \lambda_k^{(P)}\}^2, \end{aligned}$$

where $\mathbf{PC}_k^{(P)}$, $V_k^{(P)}$ and $\lambda_k^{(P)}$ are all defined in Section 2.1. This suggests that the variance component multiple phenotype test statistic $Q^{(P)}$ can be viewed as an inverse squared eigenvalue weighted quadratic combination of the K phenotype PC score test statistics $V_k^{(P)}$.

The power of the multiple phenotype variance component test depends on both the K principal angles $\theta_k^{(P)}$ as well as the phenotype eigenvalues $\lambda_k^{(P)}$. Its power is more robust with respect to the direction of the true unknown $\boldsymbol{\beta}^{(P)}$ than the test based on a specific phenotype PC, in the sense that it can be powerful when the angle of the true $\boldsymbol{\beta}^{(P)}$ lies in between phenotype PCs. Since the last phenotype PC has the largest weight in $Q^{(P)}$, the phenotype variance component test in the K:1 setting implicitly assumes that the phenotype PCs with smaller eigenvalues capture more association signals and, hence, weight them more. In fact, it weights the last phenotype $\mathbf{PC}_K^{(P)}$ most. Similar assumptions are made by the multiple phenotype Wald test. If this underlying assumption is violated, like, for example, if the first PC actually has a small principal angle, then $Q^{(P)}$ will not be powerful.

The variance component test for a SNP set can be constructed similarly in the 1:K SNP set regression setting. Specifically, if $\boldsymbol{\beta}^{(S)}$ from model (2.3) is such that each $\beta_j^{(S)}$ is assumed

to follow an arbitrary distribution with mean 0 and variance $\tau^{(S)}$, where $\tau^{(S)}$ is the variance component, then testing for $H_0 : \boldsymbol{\beta}^{(S)} = \mathbf{0}$ is equivalent to testing for $H_0 : \tau^{(S)} = 0$. Based on model (2.3), this SNP-set variance component test takes the form of the sequence kernel association test statistic (SKAT) (Wu et al. (2011))

$$(3.2) \quad Q^{(S)} = \{\mathbf{Y}^{(S)}\}^T \mathbf{G}^{(S)} \{\mathbf{G}^{(S)}\}^T \mathbf{Y}^{(S)},$$

where a linear kernel is assumed and we have used the fact that $\mathbf{Y}^{(S)}$ is scaled to have a unit variance. This leads to the simplified form of

$$Q^{(S)} = \{\mathbf{Z}^{(S)}\}^T \mathbf{Z}^{(S)},$$

which can be rewritten as

$$(3.3) \quad \begin{aligned} Q^{(S)} &= \sum_{k=1}^K \lambda_k^{(S)} [\{\mathbf{G}^{(S)} \mathbf{u}_k^{(S)} / \sqrt{\lambda_k^{(S)}}\}^T \mathbf{Y}^{(S)}]^2 \\ &= \sum_{k=1}^K [\{\mathbf{P}\mathbf{C}_k^{(S)}\}^T \mathbf{Y}^{(S)}]^2 \\ &= \sum_{k=1}^K \{V_k^{(S)}\}^2. \end{aligned}$$

Going from equation (3.2) to equation (3.3) can be realized by noting that $\mathbf{G}^{(S)} (\mathbf{G}^{(S)})^T$ has the same eigenvalues as the component $(\mathbf{G}^{(S)})^T \mathbf{G}^{(S)}$, except that the corresponding normalized eigenvectors are $\mathbf{G}^{(S)} \mathbf{u}_k^{(S)} / \sqrt{\lambda_k^{(S)}}$ instead of $\mathbf{u}_k^{(S)}$. This can be seen by multiplying the left hand sides of the following equation by $\mathbf{G}^{(S)}$: $(\mathbf{G}^{(S)})^T \mathbf{G}^{(S)} \mathbf{u}_k^{(S)} = \lambda_k^{(S)} \mathbf{u}_k^{(S)}$. Hence, equation (3.3) suggests that $Q^{(S)}$ can be viewed as a quadratic combination of the K genotype PC score test statistics $V_k^{(S)}$ which have variance $\lambda_k^{(S)}$. This is in contrast to $Q^{(P)}$ of the K:1 multiple phenotype setting, where the inverse of the squared eigenvalues are used as weights in the sum.

To highlight the difference in the role of the eigenvalues in the two settings, it is interesting to note that, for multiple phenotypes, λ_k is squared in the denominator of the K:1 variance component test while it is only implicitly present through the variance of $V_k^{(S)}$ in the 1:K case, that is,

$$\begin{aligned} Q^{(P)} &= \sum_{k=1}^K (\lambda_k^{(P)})^{-2} \{V_k^{(P)}\}^2, \\ Q^{(S)} &= \sum_{k=1}^K \{V_k^{(S)}\}^2. \end{aligned}$$

The means are

$$\begin{aligned} E(Q^{(P)}) &= \sum_{k=1}^K (\lambda_k^{(P)})^{-2} [\lambda_k^{(P)} + \{(\mathbf{u}_k^{(P)})^T \boldsymbol{\beta}^{(P)}\}^2] \\ &= \sum_{k=1}^K (\lambda_k^{(P)})^{-1} + \sum_{k=1}^K (\lambda_k^{(P)})^{-2} \{(\mathbf{u}_k^{(P)})^T \boldsymbol{\beta}^{(P)}\}^2, \\ E(Q^{(S)}) &= \sum_{k=1}^K \lambda_k^{(S)} + \sum_{k=1}^K (\lambda_k^{(S)})^2 \{(\mathbf{u}_k^{(S)})^T \boldsymbol{\beta}^{(S)}\}^2. \end{aligned}$$

In contrast, for the Wald test in the K:1 multiple phenotype setting

$$\text{Wald}^{(P)} = \sum_{k=1}^K (\lambda_k^{(P)})^{-1} \{V_k^{(P)}\}^2,$$

and this is consistent with the Wald test in the 1:K SNP-set setting

$$\text{Wald}^{(S)} = \sum_{k=1}^K (\lambda_k^{(S)})^{-1} \{V_k^{(S)}\}^2.$$

So, the Wald tests put more weight on the last PC in both settings. However, differences arise when examining the ncp for each test. The ncp of the Wald test of multiple phenotypes is influenced more by smaller eigenvalues, while the ncp of the Wald test of a SNP set is influenced more by larger eigenvalues:

$$E(\text{Wald}^{(P)}) = K + \sum_{k=1}^K (\lambda_k^{(P)})^{-1} \{(\mathbf{u}_k^{(P)})^T \boldsymbol{\beta}^{(P)}\}^2,$$

$$E(\text{Wald}^{(S)}) = K + \sum_{k=1}^K \lambda_k^{(S)} \{(\mathbf{u}_k^{(S)})^T \boldsymbol{\beta}^{(S)}\}^2.$$

Comparing the means of the test statistics, for K:1 multiple phenotype regression the power of the variance component test is influenced more by the last PC than the Wald test. On the other hand, for 1:K SNP-set regression the power of the variance component test is influenced more by the first PC than the corresponding Wald test.

The fundamental reason for the difference between the variance component test statistics of the two settings is best understood by looking at the variance component (VC) score test statistics directly. For the 1:K setting, the variance component score test statistic in equation (3.2) is based on the loglikelihood of $\mathbf{Y}^{(S)}$, which is, apart from a constant

$$\ell(\boldsymbol{\beta}^{(S)} | \mathbf{Y}^{(S)}, \mathbf{G}^{(S)}) = -\frac{1}{2} \sum_{i=1}^n \{Y_i^{(S)} - G_i^{(S)} \boldsymbol{\beta}^{(S)}\}^2.$$

This likelihood does not contain $\boldsymbol{\Sigma}^{(S)}$, which appears only in the VC statistic (3.2) in its empirical form $(\mathbf{G}^{(S)})^T \mathbf{G}^{(S)}$. On the other hand, the likelihood of $\mathbf{Y}^{(P)}$ in the K:1 setting is, apart from a constant

$$\begin{aligned} & \ell(\boldsymbol{\beta}^{(P)} | \mathbf{Y}^{(P)}, \mathbf{G}^{(P)}) \\ &= -\frac{1}{2} \sum_{i=1}^n (\mathbf{Y}_i^{(P)} - G_i^{(P)} \boldsymbol{\beta}^{(P)})^T (\boldsymbol{\Sigma}^{(P)})^{-1} (\mathbf{Y}_i^{(P)} - G_i^{(P)} \boldsymbol{\beta}^{(P)}) - \frac{n}{2} \ln\{\boldsymbol{\Sigma}^{(P)}\}. \end{aligned}$$

Here, $\boldsymbol{\Sigma}^{(P)}$ enters directly into the likelihood itself. The stark difference in the two likelihood functions explains the differences in the variance component score tests as well as the PC and Wald tests.

4. Simulation studies. In this section we first conduct simulation studies to assess the estimation accuracy of the eigenvalues and eigenvectors of the correlation matrix $\boldsymbol{\Sigma}$ in the K:1 setting by using GWAS summary statistics. Second, we compare the finite sample performance of the PC-based tests in both the K:1 multiple phenotype setting and the 1:K multiple SNP setting. Specifically, we will show that the phenotype PCs with smaller eigenvalues are likely to play more important roles in the K:1 multiple phenotype association tests, while

the genotype PCs with larger eigenvalues are likely to play more important roles in the 1:K SNP-set association tests.

First, the correlation matrix $\Sigma^{(P)}$ in Section 5.1 is set to be a true correlation matrix, and its eigenvalues and eigenvectors can be calculated, respectively. Our goal is to estimate the eigenvalues and eigenvectors using GWAS summary Z-scores. One million three-dimensional multivariate normal Z-scores with mean zeros and correlation matrix equal to $\Sigma^{(P)}$ are generated mimicking GWAS Z-scores for one million SNPs. The correlation matrix can be estimated as $\widehat{\Sigma}^{(P)}$ using sample covariance matrix of those one million Z-score vectors (Liu and Lin (2018), Zhu et al. (2015)). The sample eigenvalues and eigenvectors can be computed as the eigenvalues and eigenvectors of the matrix $\widehat{\Sigma}^{(P)}$. To assess accuracy of estimation, we repeat this experiment for 1000 times. The bias of the three estimated eigenvalues is equal to the average of the difference between the true values and the estimated values over the 1000 experiments. The variability of the three estimated eigenvalues is calculated as the sample standard deviation of those estimated eigenvalues across the 1000 experiments.

For eigenvectors we use the normalized inner product (NIP) between two vectors to measure the accuracy of the estimation of the eigenvectors. For two vectors u, v , the normalized inner product is defined as $u \cdot v / \sqrt{u \cdot u} \sqrt{v \cdot v}$, where the \cdot notation denotes inner product operation between two vectors of the same dimension. If two vectors are identical, then the NIP is equal to 1. The biases of the estimation of the three eigenvalues are: 3.61×10^{-5} , -7.35×10^{-5} , -5.14×10^{-5} , and the standard errors are: 0.0021, 0.0016, 0.0005, respectively. The NIP for the three eigenvectors are: 0.9999910, 0.9999909, 0.9999995 and its standard errors are: 1.17×10^{-5} , 1.17×10^{-5} , 5.32×10^{-7} . Simulation results show that the eigenvalues and eigenvectors can be accurately estimated.

In the K:1 multiple phenotype regression setting, we show numerically that the phenotype variance component (VC) test $Q^{(P)}$, which is designed to test for the association of a SNP and multiple phenotypes without having a prior knowledge of the true direction of the effects $\beta^{(P)}$, gives more weights to the phenotype PCs with smaller eigenvalues. Hence, it favors the last phenotype PC-based test statistic $V_K^{(P)}$ more than the first phenotype PC-based test statistic $V_1^{(P)}$. Hence, the VC test in the K:1 setting will be most powerful if the last PC captures all the signals. Its performance is closer to the last phenotype PC-based test statistic $V_K^{(P)}$ than the first phenotype PC-based test statistic.

We considered an exchangeable correlation structure $\Sigma^{(S)} = (1 - \rho)\mathbf{I} + \rho\mathbf{11}^T$, where ρ specifies the strength of the correlation. To demonstrate the theoretical results, we consider two extreme cases of $\beta^{(P)}$: $\beta^{(P)} = 3 \cdot \mathbf{u}_1^{(P)}$ and $\beta^{(P)} = \mathbf{u}_K^{(P)}$. In both cases we vary ρ such that $\rho = 0.1, 0.2, \dots, 0.7, 0.8$. The simulations are based on 10^5 runs. We consider $K = 10$ phenotypes. The results are summarized in Figure 1.

Next, we consider the 1:K SNP-set regression setting using the same exchangeable correlation structure. To demonstrate the theoretical results, we consider two extreme cases of $\beta^{(S)}$: $\beta^{(S)} = \mathbf{u}_1^{(S)}$ and $\beta^{(S)} = 3 \cdot \mathbf{u}_K^{(S)}$. With $K = 10$ SNPs in a SNP set and simulations based on 10^5 runs, the results are summarized in Figure 1.

As shown in Figure 1, in the K:1 multiple phenotype setting the last phenotype PC test statistic $V_K^{(P)}$ has a behavior closer to the phenotype VC test than the first phenotype PC test statistic $V_1^{(P)}$. When the last principal angle $\theta_K^{(P)}$ is zero, $V_K^{(P)}$ and the VC test are powerful while $V_1^{(P)}$ has little power. When the first principal angle $\theta_1^{(P)}$ is zero, $V_1^{(P)}$ is powerful while $V_K^{(P)}$ has little power. Also, note that when $\beta^{(P)}$ is in the direction of the first PC, signal strength needs to be three times larger in order to reach comparable power levels to when $\beta^{(P)}$ is in the direction of the last PC. This supports our theoretical findings in Section 2. The performance of the K:1 VC test is closer to that of $V_K^{(P)}$ but has more power

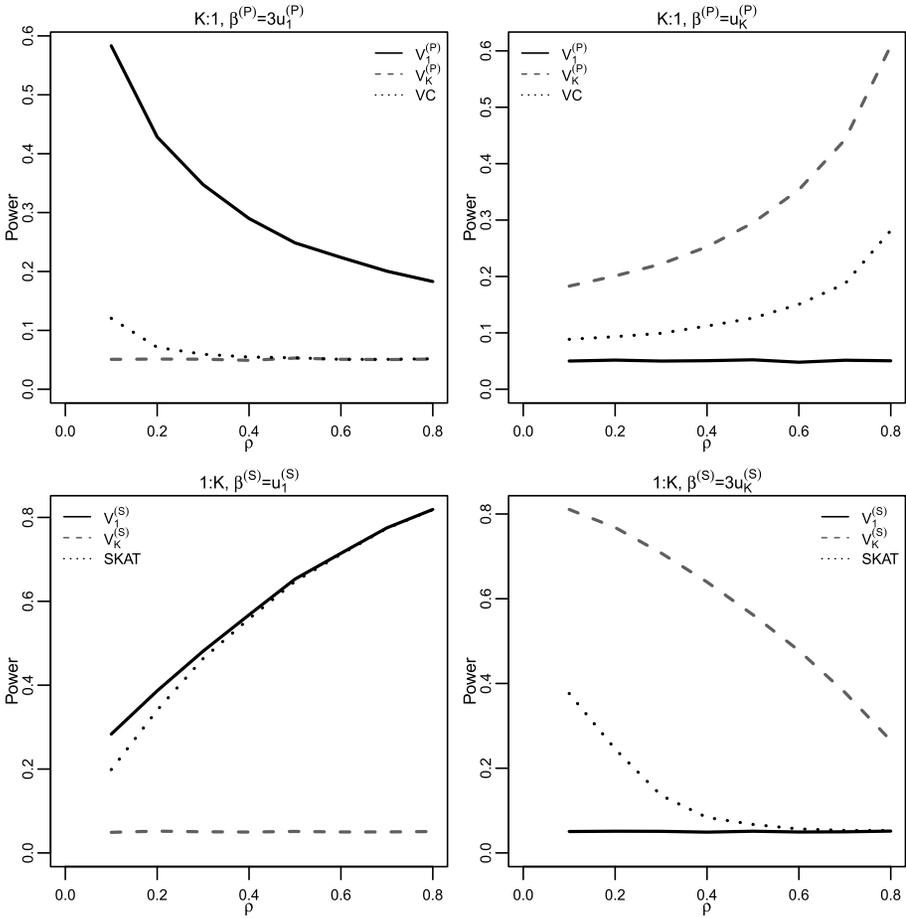


FIG. 1. Power simulations demonstrating the relationship between PC tests and variance component tests. Set $K = 10$, and consider the exchangeable correlation structure with a varying correlation coefficient ρ . For each $\beta^{(S)}$ (and $\beta^{(P)}$) and each ρ , 10^5 simulations are used to estimate the power of each method empirically.

when $\beta^{(P)} = 3u_1^{(P)}$ due to the (albeit small) contribution from the first PC. In this case, when the between-phenotype correlation ρ becomes larger, the phenotype VC test behaves more closely to $V_K^{(P)}$, reflecting the decrease in $\lambda_K^{(P)}$.

In the 1:K SNP-set regression setting, the first genotype PC-based test, $V_1^{(S)}$, has a behavior closer to the SNP-set variance component test, that is, the SKAT test, than does $V_K^{(S)}$. Specifically, when the first principal angle $\theta_1^{(S)} = 0$, both $V_1^{(S)}$ and SKAT are more powerful than $V_K^{(S)}$, although SKAT is slightly less powerful than $V_1^{(S)}$. When $\theta_K^{(S)} = 0$, $V_K^{(S)}$ is powerful while $V_1^{(S)}$ and SKAT have little power, although SKAT slightly outperforms $V_1^{(S)}$ here due to the small contribution from the last PC. Also, note that when $\beta^{(S)}$ is in the direction of the last PC, $\beta^{(S)}$ needs to be three times larger in order to reach comparable power levels to the case when $\beta^{(S)} = u_1^{(S)}$. As the between-SNP correlation (LD) ρ increases, the behavior of SKAT becomes more closer to $V_1^{(S)}$ due to the increase of $\lambda_1^{(S)}$ which gives the first PC a higher weight in the SKAT test statistic $Q^{(S)}$. In all, SKAT can be viewed as very similar to $V_1^{(S)}$ while being a slightly more robust when the signal is orthogonal to the first PC.

It is also interesting to compare the variance component tests and the principal component tests with the minimum p -value test (MinP) in the 1:K and K:1 settings. The MinP test statistic is defined as the minimum p -value of the K marginal test statistics $Z^{(S)}$ (or $Z^{(P)}$).

One uses the distribution of the MinP test statistics to calculate the p -value of this test. The MinP test is conceived to perform better when the signals are sparse (Conneely and Boehnke (2007)). However, we will show that the accuracy of this notion is highly dependent on the correlation structure. In fact, the MinP test might be less powerful than the VC tests and the PC tests in the presence of sparse signals due to correlation. In the 1:K SNP-set regression setting, an additional reason for this result is that even if $\beta^{(S)}$ is sparse, $\mu^{(S)} = \Sigma^{(S)}\beta^{(S)}$ may not be sparse. The opposite notion is also true: when $\beta^{(S)}$ (or $\beta^{(P)}$) is dense, it is not necessarily the case that VC and PC-based methods like $Q^{(S)}$ (or $Q^{(P)}$) outperform MinP.

We use the K:1 multiple phenotype regression setting to demonstrate these points for the dense case. Let $K = 10$ and $\Sigma^{(P)}$ be autocorrelated with $\rho = 0.5$. First, consider the case where $\beta^{(P)} = 1.9 \cdot [1, 1, 1, \dots, 1]^T$ such that $\theta_K^{(P)} = 90^\circ$. In this case MinP has 76.1% power, while the phenotype VC test $Q^{(P)}$ has 14.9% power, the Wald test has 74.6% power and $V_K^{(P)}$ has 5.0% power (Figure 2). In this dense signal case MinP far outperforms both $Q^{(P)}$ and $PC_K^{(P)}$. At first glance this result appears counter to common perceptions of the MinP test being ideal in sparse signal settings. Second, consider a different dense signal case, where $\beta^{(P)} \sim (1, -2, 3, -4, 5, -5, 4, -3, 2, -1)^T$ such that $\theta_K^{(P)} = 4.8^\circ$. In this case MinP has 10.8% power, while VC has 53.5% power, the Wald test has 36.6% power and $V_K^{(P)}$ has 82.2% power, which contradicts our first example and agrees with the prevailing intuition that MinP has lower power relative to VC and PC-based methods when the signal is dense. As this intuition originally emerged from uncorrelated setting, it cannot be directly applied to correlated data.

We use the 1:K SNP-set setting to demonstrate the sparse case. We here set all SNPs to have pairwise correlation $\rho = 0.8$ with one another, except for the first SNP which is independent of all others. We simulate power at the $\alpha = 0.05$ level according to model (2.3) with $\sigma^2 = 8$, $K = 10$ and $n = 1000$. First, consider the ideal case for MinP where $\beta^{(S)} = [1, 0, 0, \dots, 0]^T$ and, therefore, $\mu^{(S)} = [1, 0, 0, \dots, 0]^T$. In this case MinP has 46.7% power, while the SKAT $Q^{(S)}$ has 8.2% power, the Wald test has 7.2% power and $V_1^{(S)}$ has 5.0% power. The poor performance of $Q^{(S)}$ and $V_1^{(S)}$ is due to the angle between the first eigenvector of $\Sigma^{(S)}$ and $\beta^{(S)}$ being orthogonal ($\theta_1^{(S)} = 90^\circ$) (Figure 3). If $\beta^{(S)} \sim [0, 31, 32, \dots, 39]^T$, then MinP has 59.2% power, while $Q^{(S)}$ has 61.2% power and $V_1^{(S)}$ has 61.8% power. The improvement of $Q^{(S)}$ and $V_1^{(S)}$ is due to a smaller angle between $\beta^{(S)}$ and the first eigenvector of $\Sigma^{(S)}$

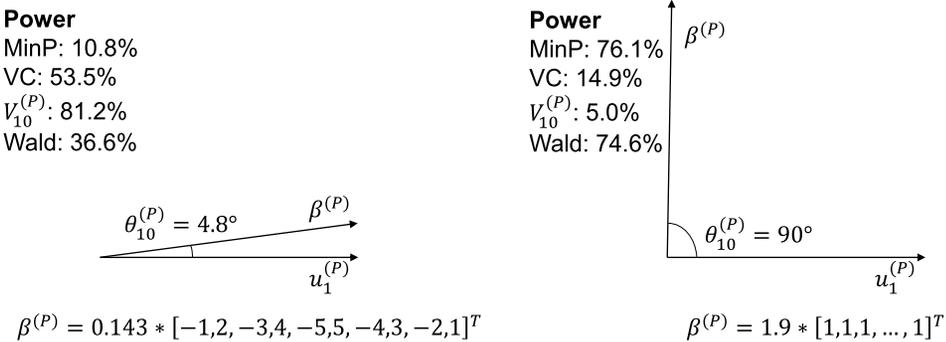


FIG. 2. The effect of $\theta_1^{(K)}$, the angle between $\beta^{(P)}$ and the last eigenvector $u_K^{(P)}$, on the power of association tests in the K:1 multiple phenotype regression setting. Both $\beta^{(P)}$ and $u_K^{(P)}$ are vectors in R^{10} , where there are $K = 10$ phenotypes being tested. The correlation structure $\Sigma^{(P)}$ is autocorrelated with $\rho = 0.5$. Power is calculated using 10^5 simulations under the null distribution for each test at the significance level 0.05.

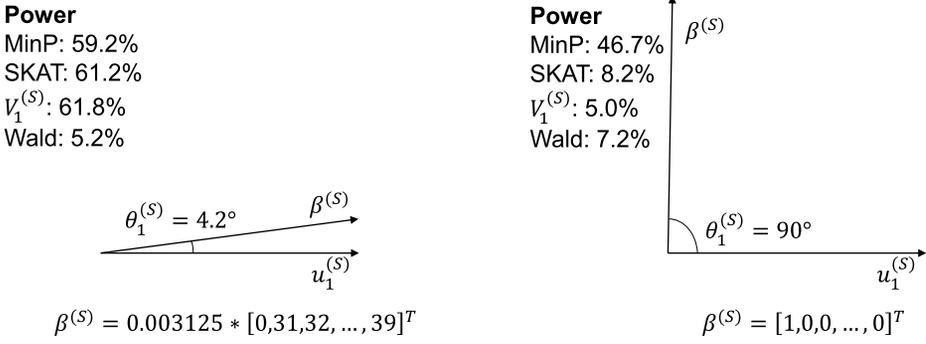


FIG. 3. The effect of $\theta_1^{(S)}$, the angle between $\beta^{(S)}$ and the first eigenvector $u_1^{(S)}$, on the power of association tests in the 1:K SNP-set regression setting. Both $\beta^{(S)}$ and $u_1^{(S)}$ are vectors in R^{10} , where there are $K = 10$ SNPs in a SNP-set. The correlation structure $\Sigma^{(S)}$ is exchangeable among nine SNPs with $\rho = 0.8$, and they are independent of one SNP. Power is calculated using 10^5 simulations under the null distribution for each test at the significance level 0.05.

($\theta_1^{(S)} = 4.2^\circ$). Based on these results it is clear that notions of signal sparsity/density and its effect on statistical power must always be conditional on the correlation and signal structure.

Our simulation study results highlight the importance of using principal angles and eigenvalues to understand the power of PC-based tests in the K:1 and 1:K settings. We have demonstrated that the tests based on the first PC are more powerful in general in the 1:K setting while tests based on the the last PC are more powerful in general in the K:1 setting. We have also illustrated that although signal sparsity plays a role in determining the power of a test, using only signal sparsity to categorize the performance of PC-based tests and the MinP test without also considering the correlation structure is not appropriate.

5. Real data analysis. In this section we compare the PC-based tests in K:1 and 1:K settings by analyzing two real data sets. The first data set is in the K:1 setting and contains GWAS summary statistics of three lipids traits where we are interested in studying individual SNP effects on multiple lipid phenotypes. The second data set is in the 1:K setting and is a breast cancer GWAS data set where we are interested in performing gene-level analysis of breast cancer risk across the genome.

5.1. *The K:1 multiple phenotype setting.* Coronary artery disease is a leading cause of death in the U.S. and worldwide. Serum concentrations of high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL), total cholesterol (TC) and triglycerides (TG) are important risk factors for coronary artery disease and are therapeutic targets for drug development. Genetic analysis of those lipids levels can help identify genetic determinants of abnormal lipids levels. The publicly available global lipids GWAS data set contains the summary statistics (Z-scores) for these lipids traits calculated from more than 100,000 individuals of European ancestry (Teslovich et al. (2010)). Since LDL and TC are highly correlated with a correlation 0.88, we restrict the analysis to three phenotypes HDL, LDL and TG since LDL is of more clinical interest than TC.

The total number of genotyped or imputed SNPs is 269,1421, and for each SNP there are three phenotype-specific Z-scores corresponding to HDL, LDL and TG. The estimated correlation matrix between the three Z-scores using the sample correlation matrix over all the independent SNPs after LD pruning is as follows (Zhu et al. (2015)):

$$\Sigma^{(P)} = \begin{pmatrix} 1.00 & -0.08 & -0.42 \\ -0.08 & 1.00 & 0.27 \\ -0.42 & 0.27 & 1.00 \end{pmatrix}.$$

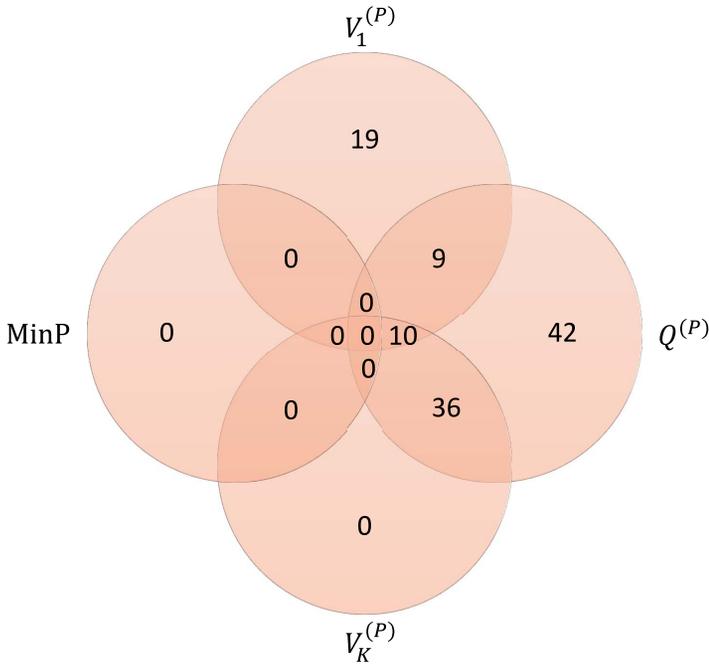


FIG. 4. The number of novel independent SNPs detected by PC-based tests for the global lipids GWAS data in the K:1 setting. With $K = 3$ different lipids, HDL, LDL and TG, multiple phenotype tests based on SNP-level association tests were performed: first and last PCs, VC and MinP.

The three eigenvalues of $\Sigma^{(P)}$ are $\lambda^{(P)} = (1.54, 0.93, 0.53)$, and the three corresponding eigenvectors are $\mathbf{u}_1^{(P)} = (0.59, -0.43, -0.68)$, $\mathbf{u}_2^{(P)} = (-0.53, -0.84, 0.07)$, $\mathbf{u}_3^{(P)} = (0.60, -0.32, 0.73)$. We apply several methods to analyze the data including the phenotype PC tests, the phenotype VC test $Q^{(P)}$ test and the MinP test to jointly analyze the three correlated Z-scores and to obtain corresponding overall p -values for testing the association between each SNP and the three lipids. The p -value of MinP cannot be more significant than the smallest individual p -value corresponding to each lipid which is one drawback of MinP in the sense that MinP cannot detect any additional SNPs. Due to linkage disequilibrium (LD), the SNPs identified might be in LD with each other. To obtain independent signals, we performed LD pruning using the LD threshold $r^2 < 0.01$ within 500kb region (Purcell et al. (2007)). Figure 4 presents the numbers of independent SNPs identified by each test. The first PC identified 38 SNPs, while 19 of them were also identified by the VC test; The third PC identified 46 SNPs, and all of them were also identified by the VC test.

To demonstrate the roles of principal angles on the power of the phenotype PC tests, we also compute the third empirical principal angle $\hat{\theta}_3^{(P)}$ which is estimated as the angle between $\mathbf{Z}^{(P)}$ and the third eigenvector of the correlation matrix of $\mathbf{Z}^{(P)}$. Table 1 summarizes the SNPs with the smallest principal angles $\hat{\theta}_3^{(P)}$ as well as the SNPs with top five largest principal angles $\hat{\theta}_3^{(P)}$. The first five SNPs, which have the smallest principal angles $\hat{\theta}_3^{(P)}$, can be detected by $V_3^{(P)}$ and $Q^{(P)}$ but not by MinP. They all have the third principal angles less than 29° . The VC test $Q^{(P)}$ gives more weight to $V_3^{(P)}$ and also has very significant p -values. However, for the last five SNPs, which have the largest principal angles $\hat{\theta}_3^{(P)}$, their effects cannot be detected by $V_3^{(P)}$ or $Q^{(P)}$.

This real data example well illustrates our theoretical finding that the variance component test favors the last PC in the K:1 setting and that the tests that favor the last PC generally

TABLE 1

Joint Analysis of the GWAS summary statistics of three lipids levels in the K:1 multiple phenotype setting. The three lipids levels include HDL, LDL and TG. We use $\mathbf{Z}^{(P)}$ as an estimate of the direction of β . The angle between $\mathbf{Z}^{(P)}$ and $\mathbf{u}_k^{(P)}$ is estimated as $\hat{\theta}_k^{(P)}$. SNPs with the smallest five principal angles $\hat{\theta}_3^{(P)}$ as well as the SNPs with the five largest principal angles $\hat{\theta}_3^{(P)}$ are displayed

SNP	CHR	MinP pval	$V_1^{(P)}$ pval	$V_3^{(P)}$ pval	$Q^{(P)}$ pval	$\hat{\theta}_1^{(P)}$	$\hat{\theta}_3^{(P)}$
rs7257916	19	2.19E-05	8.24E-01	2.90E-18	1.38E-19	87.52	6.21
rs13210634	6	5.58E-04	3.68E-01	9.61E-09	3.89E-09	75.51	20.56
rs804267	8	6.49E-05	1.72E-01	6.08E-09	2.90E-09	68.23	21.79
rs531117	11	1.51E-04	3.77E-01	1.25E-08	2.95E-09	76.25	25.92
rs1233489	6	7.83E-04	3.44E-01	2.07E-08	3.69E-09	75.44	28.98
rs6124245	20	1.42E-06	3.45E-03	9.04E-01	4.89E-04	48.84	89.10
rs2168711	18	1.14E-07	1.30E-06	8.28E-01	1.79E-04	29.86	88.69
rs12673863	7	1.11E-07	3.61E-03	8.28E-01	1.31E-04	51.95	88.45
rs7826687	8	8.08E-07	5.56E-09	7.32E-01	1.11E-03	4.35	88.03
rs17625826	2	1.22E-04	9.58E-08	5.49E-01	1.70E-03	11.64	86.29

are more powerful. It is clear that the principal angles provide a geometric perspective on the power of PC-based tests.

5.2. *The 1:K SNP-set setting.* To illustrate the performance of genotype PC-based tests relative to principal angles for SNP-set analysis of a phenotype in the 1:K setting, we analyzed the 1145 postmenopausal women of European ancestry with breast cancer and 1142 controls from the Cancer Genetic Markers of Susceptibility (CGEM) GWAS (Hunter et al. (2007)). These women were genotyped at 528,173 loci using an Illumina HumanHap500 array. A logistic regression model is used for performing gene-level analysis by controlling for the covariates—age, postmenopausal hormone usage and the top three principal components to correct for population stratification (Price et al. (2006)). Let \mathbf{X} be the matrix of these covariates (with an additional column of 1s) and $\mathbf{W} = \text{diag}\{\hat{\pi}_1(1 - \hat{\pi}_1), \dots, \hat{\pi}_n(1 - \hat{\pi}_n)\}$ where $\hat{\pi}_i = E(Y_i^{(S)} | H_0, \mathbf{X}^{(S)})$ so that $\mathbf{P} = \mathbf{W} - \mathbf{W}\mathbf{X}^{(S)}((\mathbf{X}^{(S)})^T \mathbf{W}\mathbf{X}^{(S)})^{-1}(\mathbf{X}^{(S)})^T \mathbf{W}$. The marginal test statistics are

$$(5.1) \quad Z_j^{(S)} = \frac{(\mathbf{G}_j^{(S)})^T (\mathbf{Y}^{(S)} - \hat{\boldsymbol{\pi}})}{\sqrt{(\mathbf{G}_j^{(S)})^T \mathbf{P} \mathbf{G}_j^{(S)}}},$$

where we estimate $\text{cov}(Z_j^{(S)}, Z_k^{(S)}) = \sigma_{jk}^{(S)}$, the (j, k) th component of $\boldsymbol{\Sigma}^{(S)}$ by

$$\hat{\sigma}_{jk}^{(S)} = \frac{(\mathbf{G}_j^{(S)})^T \mathbf{P} \mathbf{G}_k^{(S)}}{\sqrt{(\mathbf{G}_j^{(S)})^T \mathbf{P} \mathbf{G}_j^{(S)}} \sqrt{(\mathbf{G}_k^{(S)})^T \mathbf{P} \mathbf{G}_k^{(S)}}},$$

where $\mathbf{G}_j^{(S)}$ and $\mathbf{G}_k^{(S)}$ are the j th and k th column vectors of \mathbf{G} , respectively.

We perform gene-level analysis across the genome, testing 15,890 genes with an average of 18 SNPs per gene, including the 20kb regions flanking each gene. We perform gene-level SNP set analysis using genotype PC tests $V_k^{(S)}$, SKAT ($Q^{(S)}$) and MinP, and demonstrate the effect of the first principal angle $\theta_1^{(S)}$ on the performance of these tests.

We estimate $\beta^{(S)}$ by its MLE and estimate the first principal angle $\theta_1^{(S)}$ using the angle between $\hat{\beta}^{(S)}$ and $\mathbf{u}_1^{(S)}$. We compare in Table 2 the performance of $V_1^{(S)}$, $V_K^{(S)}$, $Q^{(S)}$ and MinP among the top gene hits. Despite lacking genome-wide significance, Figure 5 shows

TABLE 2

Top breast cancer GWAS gene hits for competing methods compared to principal angles in the 1:K setting. Though $\beta^{(S)}$ is unknown, we use $\hat{\beta}^{(S)}$ as an estimate of $\beta^{(S)}$. The angle between $\hat{\beta}^{(S)}$ and $\mathbf{u}_k^{(S)}$ is $\hat{\theta}_k^{(S)}$. The top 15 most significant genes, as determined by the smallest of the p -values of the four methods (first and last PCs, VC and MinP), are listed and sorted in increasing order by their $\theta_1^{(S)}$

Gene	K	MinP pval	$V_1^{(S)}$ pval	$V_K^{(S)}$ pval	$Q^{(S)}$ pval	$\hat{\theta}_1^{(S)}$	$\hat{\theta}_K^{(S)}$
STXBP1	9	4.00E-04	1.52E-04	9.25E-02	1.52E-03	6.3	88.9
PTCD3	12	2.66E-04	4.36E-05	8.45E-01	1.13E-04	7.7	89.9
POLR1A	16	4.00E-04	6.37E-05	1.54E-01	3.43E-04	9.0	90.0
TBK1	11	1.28E-03	3.59E-05	2.24E-01	5.44E-05	10.8	89.1
C11orf49	24	2.57E-03	3.11E-04	8.50E-01	3.84E-04	11.5	90.0
VAPB	25	7.53E-03	2.50E-04	1.34E-02	4.93E-04	15.7	89.8
XPOT	9	1.07E-03	2.04E-04	9.45E-02	1.59E-04	17.1	89.3
C1orf71	25	1.65E-02	1.19E-04	8.37E-01	6.40E-04	18.4	89.9
CNGA3	26	1.19E-03	1.33E-04	5.22E-01	1.36E-04	28.4	89.9
FGFR2	35	1.00E-04	8.15E-05	1.32E-01	3.88E-05	28.9	89.3
TMEM175	10	5.00E-04	5.05E-03	2.87E-04	2.96E-03	32.9	88.4
ZNF19	6	3.71E-02	4.55E-02	2.12E-04	3.54E-04	36.5	86.1
VWA3B	51	1.39E-03	1.23E-03	2.82E-01	2.10E-04	39.3	90.0
DACT1	9	1.26E-01	1.59E-01	1.55E-04	1.32E-01	44.9	75.6
DGKQ	9	3.31E-04	5.01E-01	1.22E-05	7.33E-03	80.1	87.0

the similarity of performance between $V_1^{(S)}$ and SKAT while showing the differences when compared to both MinP and $V_K^{(S)}$. The performance of $V_K^{(S)}$ was strongest when $\hat{\theta}_K^{(S)}$ was further from 90° . When $\hat{\theta}_1^{(S)}$ is small ($< 20^\circ$), then $\beta^{(S)}$ is likely to share much of its direction

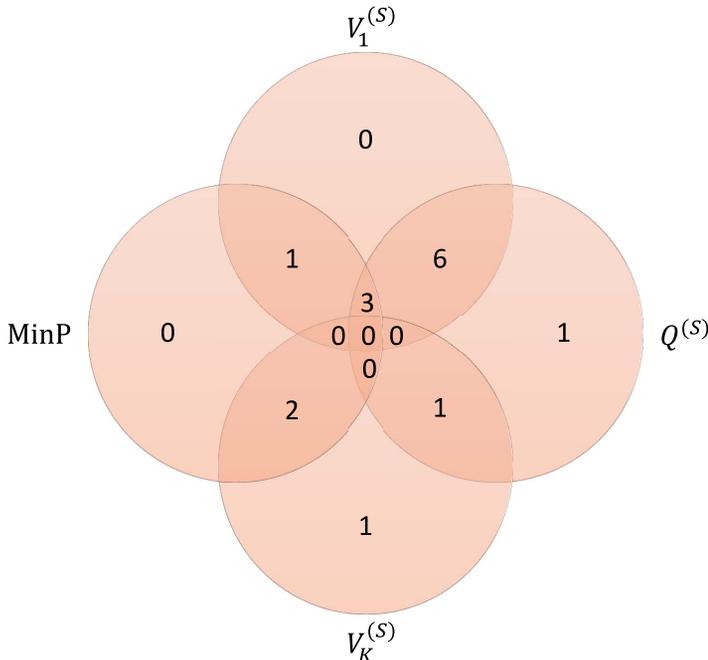


FIG. 5. Venn Diagram for 1:K method similarity in the detection of the top genes in the breast cancer GWAS. The top 15 genes were determined by sorting the smallest p -values of all four methods: first and last PCs, SKAT and MinP. Among these genes, p -values less than 10^{-3} were counted for each method to produce the diagram.

with $\mathbf{u}_1^{(S)}$. The p -values for both $Q^{(S)}$ and $V_1^{(S)}$ tend to be more significant than the p -value using MinP, although MinP still performs adequately. In contrast, when $\hat{\theta}_1^{(S)}$ is large ($> 70^\circ$), then $\beta^{(S)}$ is likely to be nearly orthogonal to $\mathbf{u}_1^{(S)}$. This indicates that, even when the signal is strong, $V_1^{(S)}$ and $Q^{(S)}$ generally have almost no power while MinP has relatively small p -values. Overall, MinP appears to be robust to the value of $\hat{\theta}_1^{(S)}$, whereas $Q^{(S)}$ and $V_1^{(S)}$ are very sensitive and require the principal angle $\theta_1^{(S)}$ to be small. This sensitivity to $\theta_1^{(S)}$ reflects the important role of the principal angle in the theoretical power of equation (2.6). In addition, the smallest overall p -values occur for genes with small $\theta_1^{(S)}$ which supports the conclusion that tests based on the first PC or giving more weights to the first PC are more powerful in the 1:K setting.

6. Discussions. In this paper we investigate the connections and differences between the association tests used in the analysis of multiple phenotypes in the K:1 setting and SNP-sets in the 1:K setting. In the multiple phenotype setting, we aim to assess the associations between a genetic variant and a group of correlated phenotypes; while in the SNP-set setting, we aim to assess the associations between a phenotype and a group of SNPs. At the first glance, these two problems seem to be the same statistically since they both aim at detecting an association between one variable and a group of variables. However, these two problems are very different in terms of the performances of various tests. Principal components based testing procedures have distinctive performance in the two settings. The higher order PC with small eigenvalues are generally preferred in multiple phenotype association studies, while the lower order PC with large eigenvalues are generally preferred in the SNP-set setting. A similar result holds for variance component tests in both settings. The variance component test $Q^{(P)}$ for multiple phenotypes gives the highest weight to the last PC and, therefore, implicitly assumes that the last PC is most informative, while $Q^{(S)}$ for SNP-set testing favors the first PC. The VC tests are more robust than single PC based tests, as they combine evidence across all the PCs and allow the unknown principal angles to be in between PCs. In practical data analysis settings, the relatively more robust variance component tests are recommended for use, such as the SKAT type test for SNP-set level association studies.

In both settings the correlation structures either among a group of SNPs or among multiple phenotypes play an important role in the performance of each test. Although the investigation of the statistical powers of a test in high dimensional settings is commonly categorized into dense and sparse regimes, this simple categorization can be misleading for correlated tests. A permutation of the signal vector, which would not alter sparsity, can drastically influence the power of many methods in the presence of correlation, such as for MinP and variance component tests. Indeed, correlation structure, signal strength and signal sparsity are all important factors in high-dimensional testing. Further research is required to investigate how they interplay in their effects on statistical powers of PC-based tests and other tests.

Jolliffe (1982) demonstrated using data examples that principal components with small variances can be as important as the ones with large variances in principal component regression; however, he did not provide any theoretical explanation for the observed empirical phenomenon. We have established the importance of principal angles, the angle between the principal components and the regression coefficient vector, in determining the power of many popular association tests.

It is worth noting that for the multiple phenotype problem ($K : 1$ setting), we regress multiple phenotypes on individual SNPs, following the traditional GWAS strategy. Our goal in this paper is not to identify which phenotypes that drive associations but to study whether and how to combine multiple phenotypes can boost power for detecting genetic association with a SNP. By modeling marginal SNP associations with multiple phenotypes, it is possible

that the detected associations could be due to the correlation/linkage disequilibrium (LD) of a detected SNP with a different causal SNP nearby. This is an issue present with the entire traditional GWAS paradigm of analyzing one SNP at a time. Indeed, association does not necessarily mean causation. Fine mapping is an active area of research with the goal of identifying causal variants of genetic associations. The fine mapping effort becomes more feasible as the field moves into whole genome sequencing studies, such as the Genome Sequencing Program (GSP) of the National Human Genome Research Institute and the Trans-Omics Precision Medicine Program (TOPMed) of the National Heart, Lung and Blood Institute.

Though in this paper we considered only the 1:K and K:1 settings, of increasing interest is the K:J setting, where the association between K outcomes and J predictors is sought. This leads to a scenario where \mathbf{Z} is a K by J matrix as opposed to the K-dimensional vector like it is in the K:1 and 1:K settings. The models considered in this paper cannot be adequately adapted to this K:J setting, so future research is needed to allow for the analysis of both multiple phenotypes and SNP-sets simultaneously.

Acknowledgments. This research is supported by NIH Grant Grants R35-CA197449, P01-CA134294, U01-HG009088, U19-CA203654 and R01-HL113338. We thank the reviewers for helpful and constructive suggestions.

The first and the second author contribute equally.

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. *Am. J. Hum. Genet.* **94** 662–676.
- CONNELLY, K. N. and BOEHNKE, M. (2007). So many correlated tests, so little time! Rapid adjustment of P-values for multiple correlated tests. *Am. J. Hum. Genet.* **81** 1158–1168.
- HAN, B., KANG, H. M. and ESKIN, E. (2009). Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genet.* **5** e1000456. <https://doi.org/10.1371/journal.pgen.1000456>
- HUANG, Y.-T. and LIN, X. (2013). Gene set analysis using variance component tests. *BMC Bioinform.* **14** 210.
- HUNTER, D. J., KRAFT, P., JACOBS, K. B., COX, D. G., YEAGER, M., HANKINSON, S. E., WACHOLDER, S., WANG, Z., WELCH, R. et al. (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* **39** 870–874.
- JOLLIFFE, I. T. (1982). A note on the use of principal components in regression. *Appl. Stat.* 300–303.
- KARACAÖREN, B., SILANDER, T., ÁLVAREZ-CASTRO, J. M., HALEY, C. S. and DE KONING, D. J. (2011). Association analyses of the MAS-QTL data set using grammar, principal components and Bayesian network methodologies. In *BMC Proceedings* **5** S8. BioMed Central Ltd.
- KARASIK, D., CHEUNG, C. L., ZHOU, Y., CUPPLES, L. A., KIEL, D. P. and DEMISSIE, S. (2012). Genome-wide association of an integrated osteoporosis-related phenotype: Is there evidence for pleiotropic genes? *J. Bone Miner. Res.* **27** 319–330.
- LEE, S., ABECASIS, G. R., BOEHNKE, M. and LIN, X. (2014). Rare-variant association analysis: Study designs and statistical tests. *Am. J. Hum. Genet.* **95** 5–23.
- LI, B. and LEAL, S. M. (2008). Methods for detecting associations with rare variants for common diseases: Application to analysis of sequence data. *Am. J. Hum. Genet.* **83** 311–321.
- LIN, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84** 309–326. MR1467049 <https://doi.org/10.1093/biomet/84.2.309>
- LIU, Z. and LIN, X. (2018). Multiple phenotype association tests using summary statistics in genome-wide association studies. *Biometrics* **74** 165–175. MR3777937 <https://doi.org/10.1111/biom.12735>
- LIU, Z. and LIN, X. (2019). A geometric perspective on the power of principal component association tests in multiple phenotype studies. *J. Amer. Statist. Assoc.* **114** 975–990. MR4011752 <https://doi.org/10.1080/01621459.2018.1513363>
- MOSKIVINA, V. and SCHMIDT, K. M. (2008). On multiple-testing correction in genome-wide association studies. *Genet. Epidemiol.* **32** 567–573.
- PRICE, A. L., PATTERSON, N. J., PLENGE, R. M., WEINBLATT, M. E., SHADICK, N. A. and REICH, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38** 904–909.

- PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81** 559–575.
- SCHIFANO, E. D., LI, L., CHRISTIANI, D. C. and LIN, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *Am. J. Hum. Genet.* **92** 744–759.
- SOLOVIEFF, N., COTSAPAS, C., LEE, P. H., PURCELL, S. M. and SMOLLER, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nat. Rev. Genet.* **14** 483–495.
- STEPHENS, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* **8** e65245. <https://doi.org/10.1371/journal.pone.0065245>
- SUO, C., TOULOPOULOU, T., BRAMON, E., WALSH, M., PICCHIONI, M., MURRAY, R. and OTT, J. (2013). Analysis of multiple phenotypes in genome-wide genetic mapping studies. *BMC Bioinform.* **14** 151. <https://doi.org/10.1186/1471-2105-14-151>
- TESLOVICH, T. M., MUSUNURU, K., SMITH, A. V., EDMONDSON, A. C., STYLIANOU, I. M., KOSEKI, M., PIRRUCCELLO, J. P., RIPATTI, S., CHASMAN, D. I. et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466** 707–713.
- WANG, K. and ABBOTT, D. (2008). A principal components regression approach to multilocus genetic association studies. *Genet. Epidemiol.* **32** 108–118.
- WU, M. C., LEE, S., CAI, T., LI, Y., BOEHNKE, M. and LIN, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89** 82–93.
- ZHANG, F., GUO, X., WU, S., HAN, J., LIU, Y., SHEN, H. and DENG, H. W. (2012). Genome-wide pathway association studies of multiple correlated quantitative phenotypes using principle component analyses. *PLoS ONE* **7** e53320.
- ZHOU, X. and STEPHENS, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies *Nat. Methods* **11** 407–409.
- ZHU, X., FENG, T., TAYO, B. O., LIANG, J., YOUNG, J. H., FRANCESCHINI, N., SMITH, J. A., YANEK, L. R., SUN, Y. V. et al. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96** 21–36.