2018

Statistical Methods for Analyzing Multivariate Phenotypes and Detecting Rare Variant Associations

Huanhuan Zhu
Michigan Technological University, hzhu1@mtu.edu

STATISTICAL METHODS FOR ANALYZING MULTIVARIATE PHENOTYPES AND DETECTING RARE VARIANT ASSOCIATIONS

By

Huanhuan Zhu

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Mathematical Sciences.

Department of Mathematical Sciences

Dissertation Co-Advisor:	*Qiuying Sha*
Dissertation Co-Advisor:	*Shuanglin Zhang*
Committee Member:	*Kui Zhang*
Committee Member:	*Laura E. Brown*
Department Chair:	*Mark S. Gockenbach*

Table of Contents

List of tables

List of figures

Preface

This dissertation is submitted for the degree of Doctor of Philosophy at Michigan Technological University. The research described herein was conducted under the supervision of Prof. Qiuying Sha and Prof. Shuanglin Zhang in the Department of Mathematical Sciences, Michigan Technological University, between September 2013 and March 2018.

This work is to the best of my knowledge original, except where references are made to previous work. Part of this work contains previously published material. The title of Chapter 1 is *Power comparisons of methods for joint association analysis of multiple phenotypes* and it was published in the Human Heredity (H. Zhu, S. Zhang, and Q. Sha, 2015, 80(3):144-152). The overall study was designed by Qiuying Sha. Huanhuan Zhu and Shuanglin Zhang conducted the statistical analyses. Huanhuan Zhu, Shuanglin Zhang, and Qiuying Sha drafted the manuscript. Chapter 2 entitled *A novel method to test associations between a weighted combination of phenotypes and genetic variants* was published in Plos One (H. Zhu, S. Zhang, and Q. Sha, 2018, 13(1): e0190788. https://doi.org/10.1371/journal.pone.0190788). Huanhuan Zhu, Shuanglin Zhang, and Qiuying Sha developed the methodology and wrote the original draft. Huanhuan Zhu and Shuanglin Zhang did formal analysis. Shuanglin Zhang and Qiuying Sha provided comments and review for editing. The title of Chapter 4 is *A novel statistical method for rare-variant association studies in general pedigrees* and it was published in BMC Proceedings (H. Zhu, Z. Wang, X. Wang, Q. Sha, 2016, 10(Suppl 7): 193–196). Qiuying Sha designed the overall study. Huanhuan Zhu and Zhenchuan Wang conducted statistical analyses, and Huanhuan Zhu, Xuexia Wang, and Qiuying Sha drafted the manuscript.

Chapter 3 is still in preparation for future publication and contains its own introduction, methods, simulation study, results and conclusion sections. Chapter 3 is a collaborative work of Huanhuan Zhu, Shuanglin Zhang, and Qiuying Sha.

Acknowledgements

At Michigan Technological University, I have obtained valuable experience in massive testing and high-dimensional data and analysis by effective collaborations with scientists in statistical genetics and biotechnology. I would like to express my deep appreciations to many professionals who helped me broaden my knowledge.

First of all, I express my sincere gratitude to the committee. I owe a great debt of gratitude to Dr. Qiuying Sha, who serves as my advisor and deserves my special appreciation for her dedicated advising and long-term support. Forever, I am motivated by her exponential insights and efforts in statistical genetics. I extend my appreciation to Dr. Shuanglin Zhang for his important advice and discussions on my research topics, to Dr. Kui Zhang and Dr. Laura Brown who carefully examined my dissertation and suggested substantial improvements. It is my honor to have these outstanding professors on my committee.

Many influential graduate students and professors deserve special thanks for their collaborations and instructions. I cannot name them all but must thank Dr. Xuexia Wang, Zhenchuan Wang, Dr. Xiao Zhang, Dr. Jianping Dong, and Dr. Mark S. Gockenbach.

Finally, I dedicate this dissertation to my family. Special thanks are due to my parents Mrs. Huaxiang Zhu and Mr. Huxiao Zhu, and my boyfriend Jinsheng Zhuang for their boundless love and dedication to me. Without their unconditional support and continuous expression of pride, I would never have made my achievements during the graduate study.

Abstract

This dissertation includes four papers with each distributed in one chapter.

In chapter 1, I compared the performance of eight multivariate phenotype association tests. The motivation to conduct this power comparison paper is as follows. For nearly 15 years, genome-wide association studies (GWAS) have been widely used to identify genetic variants associated with human diseases and traits. GWAS typically investigate genetic variants for a predefined phenotype, thus fail to identify weak but important effects. In recent years, many multivariate association tests have been developed. However, there is a lack of comprehensive summary of such kinds of approaches. To fill this important gap, I did this power comparison work. The results show that none of the methods is consistently more powerful than that of others. Relatively more powerful methods are still in large demanding.

In chapter 2, I proposed a Weighted Combination of multiple Phenotypes approach (WCmulP) for testing multiple correlated phenotypes and one genetic variant of interest. WCmulP linearly combines the multiple phenotypes with optimal weights such that the score test statistic is maximized. I compare WCmulP with other widely used tests and conduct extensive simulation studies as well as real data analysis to evaluate the performance of these methods. The results show that WCmulP outperforms the compared methods in most of the simulation scenarios and real data analysis.

As the availability of electronic health record (EHR), thousands of clinical phenotypes can be measured and collected systematically. As a result, the phenome-wide association studies (PheWAS) emerged to detect variants with a broad spectrum of phenotypes. However, the current PheWAS are intrinsically univariate test, which investigate the phenotype one at a time. Genuine PheWAS that simultaneously test the wide range of phenotypes need to be discovered. In chapter 3, I proposed a novel PheWAS approach, which referred to as PheCLC (PheWAS using clustering linear combination), to examine genetic variation associated with up to thousands of phenotypes. PheCLC jointly analyzes a wide spectrum of human phenotypes as well as classifies them into different categories based on the International Classification of Diseases (ICD) codes. The simulation results show that PheCLC certainly controls type I error rates and is much more powerful than the traditional multivariate approaches.

To date, GWAS have published thousands of common variants associated with human diseases. However, these common variants only contribute a small portion of the phenotypic variance. Many studies showed that rare variants could substantially explain missing heritability. In chapter 4, I derived a rare variant association study for family-based designs, where the rare variants can be enriched compared to population-based designs. I applied the proposed method as well as the other two family-based tests to the genetic analysis workshop 19 (GAW19) dataset and the results show that our method can identify more genes with power greater than 40% than the other two methods.

1 Chapter 1

Power Comparisons of Methods for Joint Association Analysis of Multiple Phenotypes

Background/Aims: Genome-wide association studies (GWAS) have identified many variants that each affects multiple phenotypes, which suggests that pleiotropic effects on human complex phenotypes may be widespread. Therefore, statistical methods that can jointly analyze multiple phenotypes in GWAS may have advantages over analyzing each phenotype individually. Several statistical methods have been developed to utilize such multivariate phenotypes in genetic association studies, however the performance of these methods under different scenarios is largely unknown. *Methods*: In this study, we evaluate the performance of some of the existing methods for association studies using multiple phenotypes, which include O'Brien's method, cross-validation method, optimal weight method, TATES, PCH, CCA, MANOVA and MultiPhen. We use simulation studies to compare the powers of these methods under a variety of scenarios, including different numbers of phenotypes, different values of between-phenotype correlation, different minor allele frequencies, and different mean and variance models. *Results*: Our simulation results show that there is no single method that has consistently good performance among all the scenarios. Each method has its own advantages and disadvantages. *Conclusion*: Our goal of this study is to provide researchers with useful guidelines on selecting statistical methods in the application of real data with multiple phenotypes.

1.1 Introduction

Currently, the analyses of most genome-wide association studies (GWAS) have been performed on a single phenotype. However, in the study of a complex disease, several correlated phenotypes may be measured for a disorder or its risk factors (Yang et al., 2010). For example, hypertension is measured by systolic blood pressures (SBP) and diastolic blood pressures (DBP), those two blood pressure indexes frequently display a linear relationship (Gavish et al., 2008); people's cognitive ability is usually measured by memory, intelligence, language, executive function, visual-spatial function, and motor, some of those six measurements might have relationship (Locke et al., 2006). The correlation among those multiple phenotypes can be leveraged to improve the power of genetic association tests to identify genetic markers associated with one or more phenotypes (Aschard et al., 2014).

It is not always clear how to best exploit the information to increase the power of detecting genetic markers that are associated with multiple phenotypes (Yang et al., 2010). One available method is the standard univariate association test, which performs one phenotype at a time (O'Reilly et al., 2012). However, analyzing each phenotype separately will suffer penalties from the multiple testing and result in a reduced power (Yang et al., 2010).

Recently, several methods are introduced to detect association using multivariate phenotypes (Yang et al., 2010; O'Reilly et al., 2012; O'Brien, 1984; van der Sluis et al., 2013; Klei et al., 2008; Ferreira and Purcell, 2009). These multivariate analyses of phenotypes offer several advantages over analyzing each phenotype separately (Aschard et al., 2014; Zhou and Stephens, 2014; Stephens, 2013; Yang and Wang, 2012; Solovieff et al., 2013; Galesloot et al., 2014). First, joint analyses of correlated phenotypes can exploit the correlation among phenotypes (Yang and Wang, 2012). Second, most multivariate procedures can perform a single test for association with a set of phenotypes, which reduces the number of performed tests and alleviates the multiple testing burden compared to analyzing all phenotypes separately (Klei et al., 2008; Zhu and Zhang, 2009). Finally, in case of presence of pleiotropy, where a single genetic marker is associated with multiple phenotypes, a multivariate analysis of phenotypes is more consistent with biology compared to cross-phenotype comparison of univariate analysis (Chavali et al., 2010). In summary, modeling multivariate phenotypes may increase the power over analyzing individual phenotype separately in genetic association studies (Yang et al., 2010; Klei et al., 2008; Lange et al., 2004).

Several methods to detect association using multivariate phenotypes have been introduced in recent years. These methods can be divided into three groups: combining test statistics from univariate analysis, variable reduction methods, and regression models (Yang and Wang, 2012). The first group includes O'Brien's method (OB) (O'Brien, 1984), cross-validation method (CV) (Yang et al., 2010), and Trait-based Association Test that uses Extended Simes procedure (TATES) (van der Sluis et al., 2013). Specifically, each method in this group is to perform univariate phenotype-genotype association test for each phenotype individually and then combine the test statistics from the univariate analysis (Yang et al., 2010; O'Brien, 1984; van der Sluis et al., 2013; Wei and Johnson, 1985). Variable reduction methods derive a single or a few new phenotypes that are linear combinations of the original phenotypes. The best-known method for variable reduction methods involves using one or more of the principal components of phenotypes (PCP) in place of the original phenotypes (Lan et al., 2003; Wang and Abbott, 2008). Building on the work of PCP, Klei et al. (2008) developed principal component of heritability (PCH) with coefficients maximizing the heritability of phenotypes. Another variable reduction method is canonical correlation analysis (CCA), which extracts the linear combination of phenotypes that explain the largest possible amount of the covariation between the genetic marker and all phenotypes (Ferreira and Purcell, 2009). Regression models, such as multivariate analysis of variance (MANOVA) and proportional odds logistic regression for joint model of multiple phenotypes (MultiPhen) (O'Reilly et al., 2012), can be used to analyze the association of a genetic marker with multivariate phenotypes.

Even though several methods have been developed to jointly analyze multiple phenotypes including the above mentioned methods, there is a lack of thorough comparison between those methods. Each of the methods attains its maximal power only in special circumstances (O'Reilly et al, 2012; van der Sluis et al., 2013). Therefore, a thorough

power comparison of these methods is a meaningful and worthy work. In this article, we compare the power performance of eight methods in a variety of models. These eight methods include OB (O'Brien, 1984), CV (Yang et al., 2010), optimal weight method (OW), TATES (van der Sluis et al., 2013), PCH (Klei et al., 2008), CCA (Ferreira and Purcell, 2009), MANOVA, and MultiPhen (O'Reilly et al., 2012). Our ultimate goal of this study is to provide researchers with useful guidelines on selecting statistical methods in the application of real data with multiple phenotypes.

1.2 Comparison of Methods

In this section, we briefly introduce each of those methods compared in this study. We use the following notations. Considering a sample of n unrelated individuals, each individual has been genotyped at a genetic marker. Assume that there are K possibly correlated phenotypes. Let $\boldsymbol{T} = (T_1, T_2, \dots, T_K)^T$ be a vector of K univariate test statistics, where T_k $(k = 1, 2, \dots, K)$ is a test statistic for testing the association between a genetic marker and the k^{th} phenotype. Assume that $\boldsymbol{T}$ asymptotically follows a multivariate normal distribution with mean $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_K)^T$ and (known or consistently estimated) covariance matrix $\boldsymbol{\Sigma}$. The null hypothesis is H_0: $\boldsymbol{\beta} = \boldsymbol{0}$, and the alternative hypothesis is H_1: at least one $\beta_k \neq 0$ for $k = 1, \dots, K$.

O'Brien's method (OB). *O'Brien* (1984) showed that if $\beta_1 = \beta_2 = \cdots = \beta_K$, the test statistic $\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}$ is the most powerful test among a class of test statistics that are linear combination of $T_1, T_2, \dots, T_K$, where $\boldsymbol{e} = (1,1 \dots,1)^T$ with length K (O'Brien, 1984; Wei and Johnson, 1985). Under the null hypothesis, the test statistic $\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}$ follows a normal distribution with mean 0 and variance $\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}$.

Cross-validation method (CV). CV (Yang et al., 2010) averages multiple sample splitting results. In detail, splitting the sample into two subsets, one is called training set for estimating weights and the other is called testing set for constructing final testing statistic. Let $\boldsymbol{T_W}$ and $\boldsymbol{T}$ denote the test statistic obtained from the training set and testing set, respectively. Then, the test statistic to test the association between a marker and multiple phenotypes is given by $S = \boldsymbol{T}_{\boldsymbol{W}}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}$. The splitting procedure is repeated R times (e.g., 100 times), the final test statistic is $\bar{S} = \frac{1}{R}\sum_{r=1}^{R} S_r$, where S_r is the test statistic S based on the r^{th} splitting procedure. The P-value of the final test statistic is calculated using a permutation procedure.

Optimal weight method (OW). The above two methods are seeking a weight vector $\boldsymbol{a}$ such that $\boldsymbol{a}^T\boldsymbol{T}$ can combine the univariate test statistics in a linear manner. $\boldsymbol{a} = \boldsymbol{\Sigma}^{-1}\boldsymbol{e}$ is proposed in the OB method while $\boldsymbol{a} = \boldsymbol{\Sigma}^{-1}\boldsymbol{T_W}$ is proposed in the CV method. For any vector $\boldsymbol{a}$ that may not depend on the data at hand, $W(\boldsymbol{a}) = \frac{(\boldsymbol{a}^T\boldsymbol{T})^2}{\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}}$ follows a chi-square distribution with degree of freedom 1 under the null hypothesis. If we consider to choose $\boldsymbol{a}$ that depends on the data at hand, we can find $\boldsymbol{a_0}$ such that $W(\boldsymbol{a_0}) = \max_{\boldsymbol{a}} W(\boldsymbol{a})$. We can use $W(\boldsymbol{a_0})$ as a test statistic, which we call it the optimal weight method (OW). If we denote $\boldsymbol{w} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{a}$, then the optimal weight statistic $W(\boldsymbol{a_0}) = \max_{\boldsymbol{a}} W(\boldsymbol{a}) = \max_{\boldsymbol{w}} \frac{\boldsymbol{w}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{w}} = \boldsymbol{T}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}$. Under the null hypothesis, $W(\boldsymbol{a_0}) = \boldsymbol{T}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}$ follows a chi-square distribution with degrees of freedom K.

Trait-based Association Test that uses Extended Simes procedure (TATES). TATES (van der Sluis et al., 2013) combines *P*-values obtained in standard univariate GWAS to acquire one phenotype-based *P*-value, while correcting for correlations between phenotypes. In detail, after obtaining the univariate *P*-values $p_1, \dots, p_K$ from the corresponding univariate association tests, those *P*-values are sorted in ascending order, $p_{(1)}, \dots, p_{(K)}$. Then, we can find the effective number of independent *P*-values of all K phenotypes, $m_{e,}$ and the effective number of *P*-values among the top k *P*-values, $m_{e(k)}$, where $k = 1, \dots, K$. Finally, the ultimate *P*-value is given by $Min\left(\frac{m_e p_{(k)}}{m_{e(k)}}\right)$.

Multivariate analysis of variance (MANOVA). Since we only consider one genetic marker in this study, one-way MANOVA is performed. Specifically, the Wilks' lambda test is used to conduct a hypothesis test. The Wilks' lambda test statistic is equivalent to the likelihood ratio test statistic, which is the ratio of the generalized variances $\frac{|\boldsymbol{E}|}{|\boldsymbol{H}+\boldsymbol{E}|}$, where $\boldsymbol{H}$ is the hypothesis sum of squares and cross product (SSCP) matrix, $\boldsymbol{E}$ is the error SSCP matrix, and $|\cdot|$ is the determinant of a matrix. The explicit forms of $\boldsymbol{H}$ and $\boldsymbol{E}$ are given by $\boldsymbol{H} = \widehat{\boldsymbol{\beta}}(\boldsymbol{X}^T\boldsymbol{X})\,\widehat{\boldsymbol{\beta}}^T, \boldsymbol{E} = \boldsymbol{Y}^T\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}(\boldsymbol{X}^T\boldsymbol{X})\,\widehat{\boldsymbol{\beta}}^T$, where $\widehat{\boldsymbol{\beta}} = \boldsymbol{Y}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}, \boldsymbol{X}$ is the $n \times 1$ vector of genotypes for all n individuals, and $\boldsymbol{Y}$ is the $n \times K$ matrix of phenotypes for all individuals. Under H_0, $-2\log\Lambda = -n\log\frac{|E|}{|H+E|}$ has an asymptotic χ^2_K distribution, where Λ denotes the ratio of the likelihood function under null hypothesis to the likelihood function under alternative hypothesis.

Principal components of heritability (PCH). Variable reduction approach derives a single or a few new phenotypes that are linear combinations of the original phenotypes. Existing methods include principal components analysis (PCA) where for the first component, the coefficients maximize the variance of the multivariate phenotypes (Yang and Wang, 2012), principal component of heritability (PCH) with coefficients maximizing the total heritability of the phenotypes (Ott and Rabinowitz, 1999). Recently, Klei et al. (2008) developed a PCH method, in which the sample is randomly split in a training set, which is used to construct the optimal linear combination of phenotypes from a heritability point of view, and a test set, which is used for association testing between genotype and the optimal linear combination of phenotypes. The test statistic is calculated repeatedly using random splits of the data. Ultimately, the statistic is derived from an integration of the individual test statistics.

Canonical correlation analysis (CCA). CCA is a multivariate generalization of the Pearson product-moment correlation (Hotelling, 1936). Ferreira and Purcell (Ferreira and Purcell, 2009) used CCA to measure the association between the genetic marker and phenotypes. CCA extracts the linear combination of phenotypes that explain the largest possible amount of the co-variation between the marker and all phenotypes. The test is based on Wilks' lambda and the corresponding *F*-approximation.

Joint model of multiple phenotypes (MultiPhen). MultiPhen (O'Reilly et al., 2012) inverts the general linear regression, in which the genotype is the response variable and all the phenotypes are independent variables. The genotype data is an allele count and is therefore modelled using ordinal regression. O'Reilly et al. (2012) used proportional odds logistic regression model which defines the class probabilities. A likelihood ratio test is performed to test the null hypothesis that none of the phenotypes have association with the genetic marker.

1.3 Simulation Study

To compare different methods, we investigate their type-I error rates and powers by simulation data sets with 1,000 unrelated individuals. To generate genotype data on a genetic marker, we assume that minor allele frequency (MAF) is 0.1 or 0.3 and assume Hardy-Weinberg Equilibrium. For each individual, we generate $K = 10$, 20 and 40 phenotypes. The K phenotypes of an individual are generated from the following model

$$\boldsymbol{y} = \boldsymbol{\mu}x + \boldsymbol{E} \tag{1.1}$$

where $\boldsymbol{y} = (y_1, \dots, y_K)^T$ are the phenotypic values of an individual; x is the genotypic score of the individual at the genetic marker; $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)^T$ are the genetic effects of the genetic marker on the phenotypes and their values depend on mean models (Table 1.1); $\boldsymbol{E} = \sqrt{\rho}\boldsymbol{B}\boldsymbol{u} + \sqrt{1-\rho}\boldsymbol{\varepsilon}$ is a vector with random effect $\boldsymbol{u}$ and random error $\boldsymbol{\varepsilon}$, where $\boldsymbol{B}$ is a $K \times n_u$ loading matrix, the values $\boldsymbol{B}$ and n_u depend on the variance models (Table 1.2); $\boldsymbol{u} = \left(u_1, \dots, u_{n_u}\right)^T \sim MVN(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_K)^T \sim MVN(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix. From the assumptions above, we can have $\boldsymbol{E} = (E_1, \dots, E_K)^T \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \rho\boldsymbol{B}\boldsymbol{B}^T + (1-\rho)\boldsymbol{I}$.

In our simulation studies, we consider six different mean models and two different variance models. Table 1.1 gives the values of the genetic effect $\boldsymbol{\mu}$ in different mean models. In mean model 1, the genetic marker has the same size and direction of effect on all phenotypes. In mean model 3, the genetic marker has effect on all the phenotypes, but has different directions of the effect. In mean models 2, 4, 5, and 6, the genetic marker has effect on part of the phenotypes, but has no effect on the rest.

The two different variance models are based on the correlation setting among all the phenotypes. Table 1.2 gives the values of n_u and $\boldsymbol{B}$ under the two different variance models. Under the first variance model, all of the phenotypes have correlation ρ with each other. Under the second variance model, the first half phenotypes have correlation ρ with each other, the second half phenotypes have correlation ρ, and there are no correlations between phenotypes in the first half and in the second half.

The heritability of genotypes to the k^{th} phenotype is given by

$$h^2(y_k) = \frac{\mathrm{var}(x)\mu_k^2}{\mathrm{var}(x)\mu_k^2 + 1} \approx \mathrm{var}(\boldsymbol{x})\mu_k^2 \tag{1.2}$$

The heritability of genotypes to the total K phenotypes is given by $h^2 = \sum_{k=1}^{K} h^2(y_k) \approx \mathrm{var}(\boldsymbol{x}) \sum_{k=1}^{K} \mu_k^2$. Then given the heritability h^2, we can calculate $\alpha_1, \alpha_2, \dots, \alpha_6$ for different mean models in Table 1.1.

For OB, CV, OW, and TATES, we use the score test statistic under the linear model as the univariate phenotype-genotype association test. For the type-I error rates, we assume $\boldsymbol{\mu} = \mathbf{0}$ such that the genetic marker is independent of all phenotypes. For power comparisons, we consider different values of heritability, different values of between-phenotype correlation, and different values of MAF.

1.4 Results

For type-I error rates evaluation, we consider different numbers of phenotypes, different MAFs, different variance models, and different values of significance levels. In the two variance models, we assume $\rho = 0.1$. In each simulation scenario, the sample size is 1,000, *P*-values of CV and PCH are estimated by 1,000 permutations, and *P*-values of other methods are calculated by asymptotic distributions. The type-I error rates are evaluated using 1,000 replicated samples. For 1,000 replicated samples, the 95% confidence intervals (CIs) for type-I error rates at nominal levels 0.05 and 0.01 are (0.0365, 0.0635) and (0.004, 0.016), respectively. Tables 1.3 and A.1.1-A.1.5 give the estimated type-I error rates of the eight methods for different numbers of phenotypes (10, 20, 40), different values of MAF (0.1, 0.3), different significance levels (0.01, 0.05), and two different variance models. From Table 1.3 and Tables A.1.1-A.1.5, we can see that except MultiPhen, almost all the estimated type-I error rates are within the 95% CIs, which indicates that the estimated type-I error rates are not significantly different from the nominal levels. Thus, the seven tests are all valid tests under our simulation study. From our results, we noticed that when the number of phenotypes is large (e.g., 40), the type-I error rates of MultiPhen are inflated. This phenomenon is also noticed by other studies. For example, Aschart et al. (2014) pointed out that Multiphen suffers from an inflated type-I error rates when the ratio of the number of phenotypes over the number of individuals is relatively large (>0.01). Because of this, we did not include MultiPhen in the power comparisons when the number of phenotypes is 40.

For power comparisons, we consider different values of MAF, different values of between-phenotype correlation ρ, different values of heritability, different numbers of phenotypes ($K = 10, 20, 40$), different mean models, and different variance models (see Figures 1.1-1.4 and Figures B.1.1-B.1.10). In each of the simulated scenarios, the sample size is 1,000, *P*-values of CV and PCH are estimated by 1,000 permutations, and *P*-values of other methods are calculated by asymptotic distributions. The power is evaluated using 1,000 replicated samples at a significance level of 0.05.

Our simulation results show the following patterns:

1. None of the considered methods are consistently most powerful under the simulation scenarios (Figures 1.1-1.4 and B.1.1-B.1.10).
2. OW, MANOVA, CCA, PCH, and MultiPhen have very similar power over all the simulation scenarios, and CV is consistently slightly less powerful than the five tests (Figures 1.1-1.4 and B.1.1-B.1.10). We call the five tests (OW, MANOVA, CCA, PCH, and MultiPhen) as group 1. The similar conclusion is also drawn in the following published works. van der Sluis et al. (2013) pointed out that under most circumstances, MultiPhen and MANOVA yield very similar results in terms of power. MANOVA is equivalent to CCA when CCA is applied to a single genetic marker at a time (Galesloot et al., 2014). The performance of Multiphen is similar to PCH and MANOVA when a small number of phenotypes are analyzed (Aschard

et al., 2014). According to O'Reilly et al. (2012), MultiPhen and CCA perform very similarly except in case of low MAF and non-normal phenotypes.

3. OB has the highest power among all methods when the genetic effects are homogeneous (mean model 1). However, this method reduces power significantly when genetic effects are heterogeneous, especially when opposite directions of the genetic effects exist or when genotypes impact a small portion of phenotypes (mean models 3-5). (Figures 1.1-1.4 and B.1.1-B.1.10)
4. Power comparisons of TATES with tests in group 1 and OB depend on the mean models, variance models, and the values of ρ. In general, TATES is the most powerful test when ρ is small and genotypes impact a very small portion of phenotypes. (Figures 1.1-1.4 and B.1.1-B.1.8)
5. The power of OB decreases with the increasing of ρ because OB involves all phenotypes and information contained by all phenotypes will be decreased with the increasing of ρ; the power of TATES is relatively robust to ρ because TATES essentially only depends on the phenotype that has the strongest association with the genotype; powers of tests in group 1 decrease or increase with the increasing of ρ depending on mean models and variance models. (Figures 1.3-1.4 and B.1.5-B.1.8)
6. Powers of all tests are robust to MAF. (Figures B.1.9-B.1.10)

The power is also evaluated at a significance level of 5×10^{-8} for six methods because the two methods (CV and PCH) that use permutations to calculate their p-values are computationally extensive. The patterns of power comparisons at significance level 5×10^{-8} (Figures B.1.11-B.1.14) are similar to those at significance level 0.05 (Figures 1.1-1.4). In summary, tests in group 1 have very similar power, and CV is slightly less powerful than tests in group 1. OB has the highest power among all methods when the genetic effects are homogeneous, but this method reduces power significantly when genetic effects are heterogeneous. In general, TATES is the most powerful test when between-phenotype correlation ρ is small and genotypes impact a very small portion of the phenotypes.

1.5 Discussion

In the study of a complex disease, several correlated phenotypes are often measured as risk factors for the disease (Aschard et al., 2014). Analyzing multiple disease-related phenotypes could potentially increase power to detect association of genetic markers with a disease. In recent years, several multivariate analyses of GWAS have been introduced. A thorough comparison between those methods is needed for researchers to choose the best and most appropriate method under a certain circumstance. In this study, we used simulated data to compare the performance of eight commonly used methods (OW, MANOVA, CCA, PCH, MultiPhen, OB, CV, and TATES) for testing association between multiple phenotypes and a genetic marker. Our simulation results showed that there is not a single method that performs best under all the simulated scenarios we considered, each method has its own pros and cons.

In our simulation studies, we did not study the effect of missing data. In the presence of missing data in the outcomes, dropping individuals with missing data in the analyses may result in power loss (Yang and Wang, 2012). Imputation can be used to impute missing genotype data or phenotype data. Missing phenotype data can be handled either by case-wise deletion (if data are missing above a pre-defined per-individual missingness threshold) or mean imputation (i.e. a missing phenotype is replaced by the corresponding sample mean) (Tang and Ferreira, 2012). Missing genotype data can be imputed using dedicated software and appropriate reference panels (e.g. HapMap).

Our study showed substantially different patterns of power comparisons among our simulated scenarios. Overall, OW, MANOVA, CCA, PCH, and MultiPhen have very similar power, and CV is slightly less powerful than the five tests. OB has the highest power among all methods when the genetic effects are homogeneous. TATES is the most powerful test when between-phenotype correlation ρ is small and genotypes impact a very small portion of the phenotypes. Because in practice, we do not know the number of phenotypes impacted by genotypes and we also do not know whether the genetic effects are homogeneous, we recommend that one can perform OB and one of the tests in group 1 when between-phenotype correlation ρ is large, and one can perform TATES, OB, and one of the five tests in group 1 when between-phenotype correlation ρ is small. We can also construct a robust test as follows. Let p_{TATES}, p_{OB}, and p_{OW} denote the *P*-values of TATES, OB, and OW, respectively. Then, we define the test statistic of the robust test as $T_{robust} = \min\{p_{TATES}, p_{OB}, p_{OW}\}$. However, the performance of the robust test needs further investigation.

1.6 Tables and Figures

Table 1.1. Six mean models used in the simulation studies

Mean models	Phenotype number	Effect size (direction of the effect)
1	1 $\vdots$ K	$\alpha_1\ (+)$ $\vdots$ $\alpha_1\ (+)$
2	1 $\vdots$ $\left[\frac{K}{2}\right]$ $\left[\frac{K}{2}\right]+1$ $\vdots$ K	0 $\vdots$ 0 $\alpha_2\ (+)$ $\vdots$ $\alpha_2\ (+)$
3	1 $\vdots$ $\left[\frac{K}{2}\right]$ $\left[\frac{K}{2}\right]+1$ $\vdots$ K	$\alpha_3\ (-)$ $\vdots$ $\alpha_3\ (-)$ $\alpha_3\ (+)$ $\vdots$ $\alpha_3\ (+)$
4	1 $\vdots$ $K-1$ K	0 $\vdots$ 0 $\alpha_4\ (+)$
5	1 $\vdots$ $\left[\frac{4K}{5}\right]$ $\left[\frac{4K}{5}\right]+1$ $\vdots$ K	0 $\vdots$ 0 $\left(\left[\frac{4K}{5}\right]+1\right)\alpha_5\ (+)$ $\vdots$ $K\alpha_5\ (+)$
6	1 $\vdots$ $\left[\frac{2K}{3}\right]$ $\left[\frac{2K}{3}\right]+1$ $\vdots$ K	0 $\vdots$ 0 $\left(\left[\frac{2K}{3}\right]+1\right)\alpha_6\ (+)$ $\vdots$ $K\alpha_6\ (+)$

Note: K is the total number of phenotypes; $[\cdot]$ is the floor function; $+\ (-)$ indicates that the direction of the genetic effect on the phenotype is positive (negative); given heritability, $\alpha_1, \alpha_2, \ldots, \alpha_6$ can be calculated based on formula (2.2) in Simulation section.

Table 1.2. The values of n_u and $\boldsymbol{B}$ in the two variance models.

Variance models	Values of $\boldsymbol{n_u}$ and $\boldsymbol{B}$
Model 1	$n_u = 1, \boldsymbol{B} = (1, 1, \dots, 1)^T$
Model 2	$n_u = 2, \boldsymbol{B} = \begin{pmatrix} \boldsymbol{e}_{\lceil \frac{K}{2} \rceil} & \boldsymbol{0}_{\lceil \frac{K}{2} \rceil} \\ \boldsymbol{0}_{K-\lceil \frac{K}{2} \rceil} & \boldsymbol{e}_{K-\lceil \frac{K}{2} \rceil} \end{pmatrix}$

Note: $\boldsymbol{e}_{\lceil \frac{K}{2} \rceil}$ is a column vector with all elements 1's, and length $\left\lceil \frac{K}{2} \right\rceil$; $\boldsymbol{0}_{\lceil \frac{K}{2} \rceil}$ is a column vector with all elements 0's, and length $\left\lceil \frac{K}{2} \right\rceil$; K is the total number of phenotypes considered in the simulation studies.

Table 1.3. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 20$, MAF is 0.3, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.053	0.011	0.048	0.011
CV	0.044	0.013	0.042	0.010
OW	0.049	0.013	0.044	0.013
TATES	0.050	0.011	0.052	0.011
MANOVA	0.049	0.013	0.047	0.013
CCA	0.049	0.012	0.049	0.013
PCH	0.051	0.010	0.044	0.014
MultiPhen	0.060	0.013	0.049	0.015

Figure 1.1. Power comparisons of the eight methods as a function of heritability for the six mean models under variance model 1. The total number of phenotypes is $K = 20$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

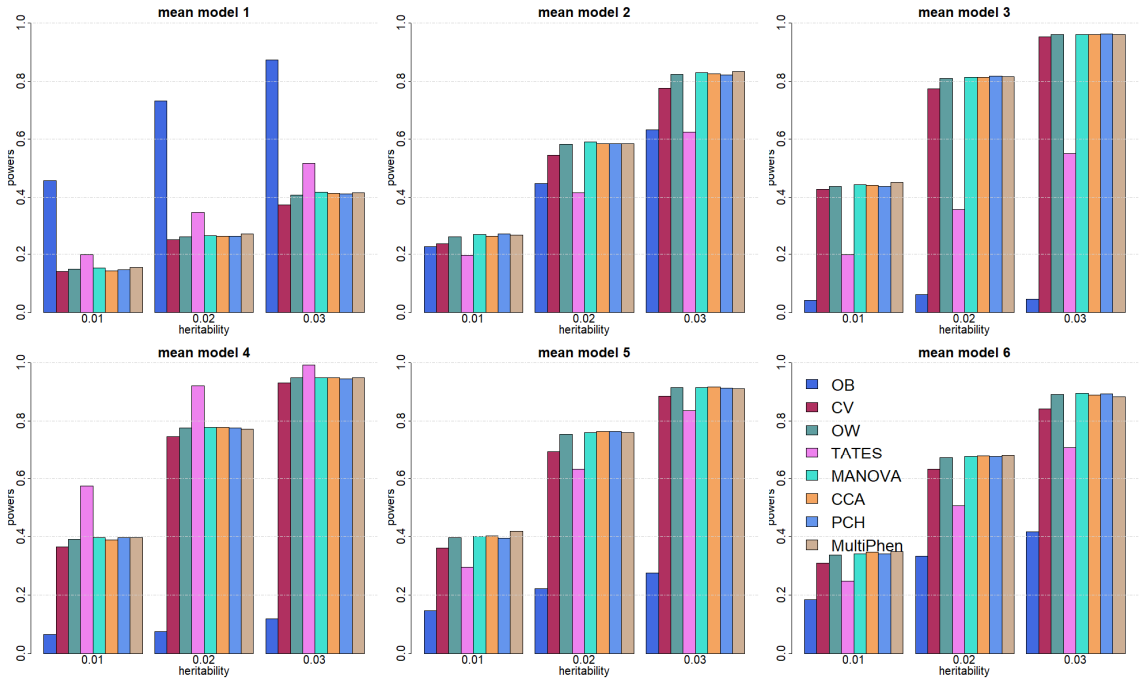


Figure 1.2. Power comparisons of the eight methods as a function of heritability for the six mean models under variance model 2. The total number of phenotypes is $K = 20$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

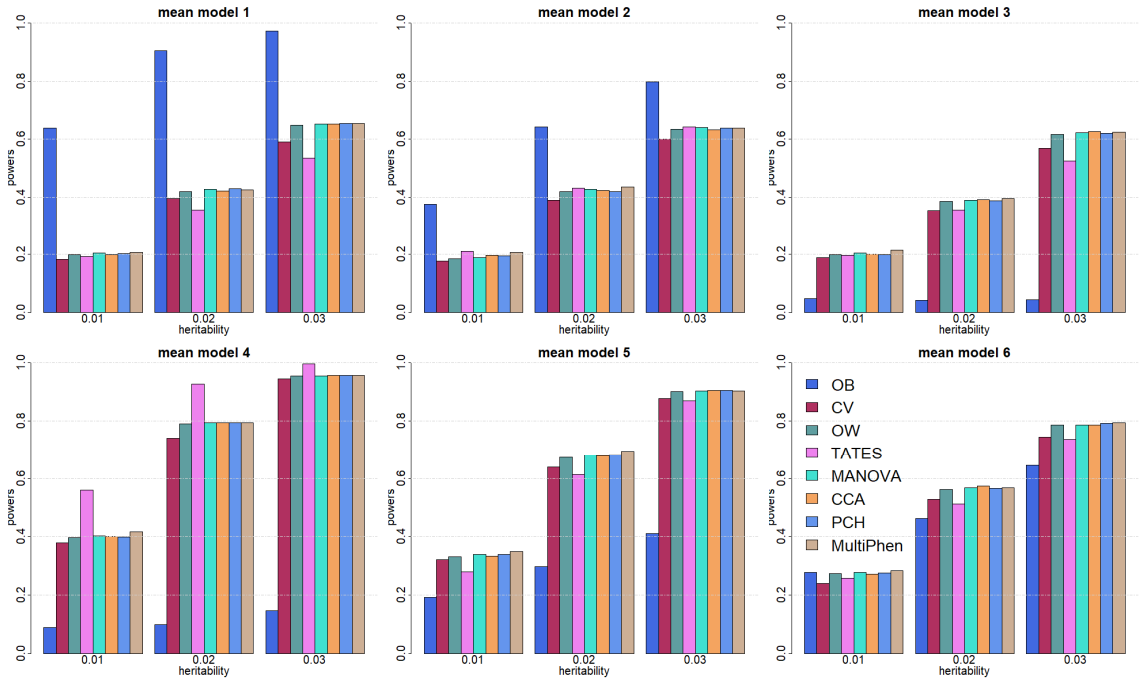


Figure 1.3. Power comparisons of the eight methods as a function of rho for the six mean models under variance model 1. The total number of phenotypes is $K = 20$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

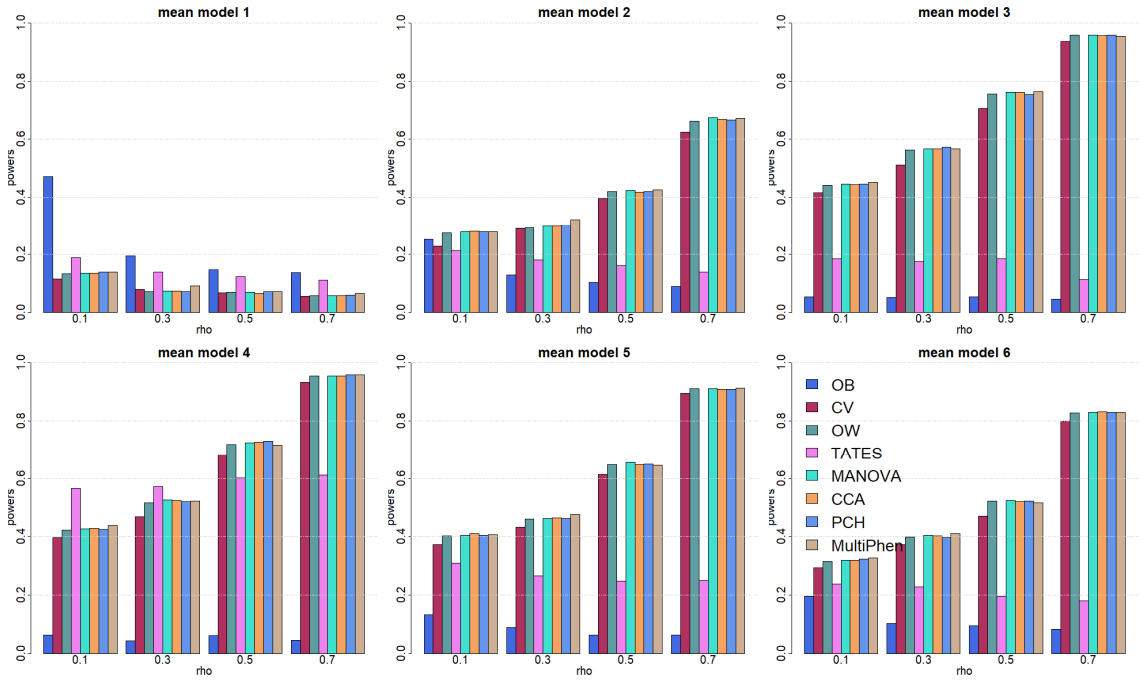
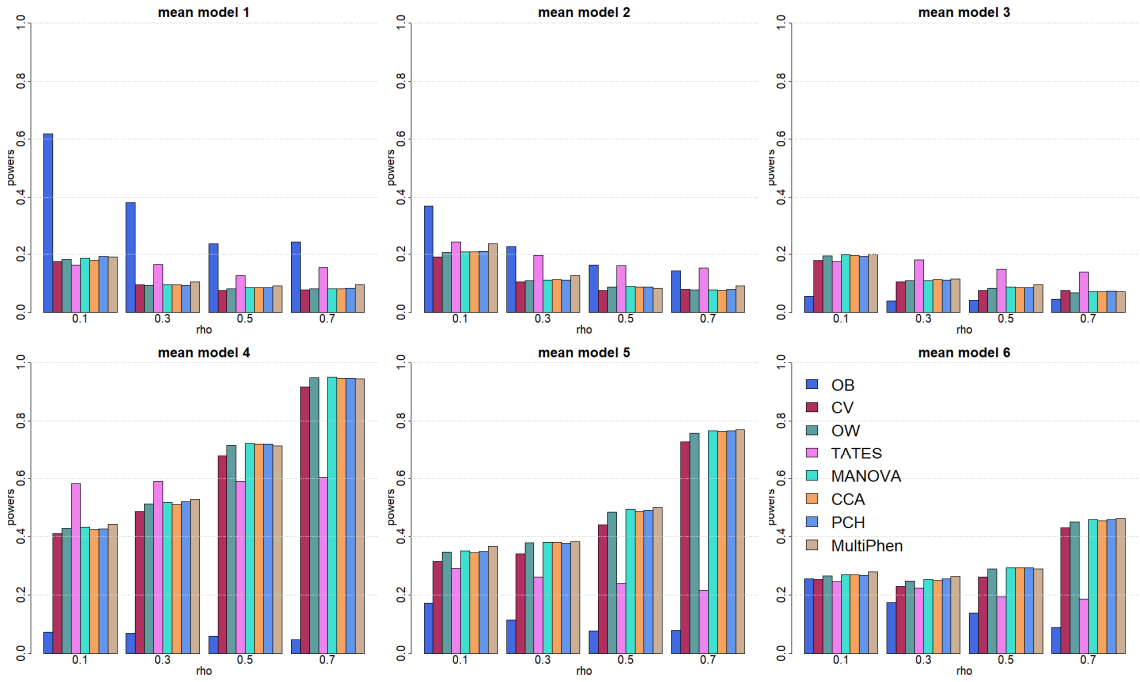


Figure 1.4. Power comparisons of the eight methods as a function of rho for the six mean models under variance model 2. The total number of phenotypes is $K = 20$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.



2 Chapter 2

A Novel Method to Test Associations between a Weighted Combination of Phenotypes and Genetic Variants

Many complex diseases like diabetes, hypertension, metabolic syndrome, et cetera, are measured by multiple correlated phenotypes. However, most genome-wide association studies (GWAS) focus on one phenotype of interest or study multiple phenotypes separately for identifying genetic variants associated with complex diseases. Analyzing one phenotype or the related phenotypes separately may lose power due to ignoring the information obtained by combining phenotypes, such as the correlation between phenotypes. In order to increase statistical power to detect genetic variants associated with complex diseases, we develop a novel method to test a weighted combination of multiple phenotypes (WCmulP). We perform extensive simulation studies as well as real data (COPDGene) analysis to evaluate the performance of the proposed method. Our simulation results show that WCmulP has correct type I error rates and is either the most powerful test or comparable to the most powerful test among the methods we compared. WCmulP also has an outstanding performance for identifying single-nucleotide polymorphisms (SNPs) associated with COPD-related phenotypes.

2.1 Introduction

Genome-wide association studies (GWAS) aim to discover genetic variants associated with complex diseases (O'Reilly et al., 2012; Yang and Wang, 2012). In GWAS, researchers often collect data on multiple correlated phenotypes to get a better understanding of the complex disease (Yang et al., 2010). Here are some examples of what diseases are measured by multiple phenotypes. In type 2 diabetes (T2D) studies data are usually collected on a number of risk factors and diabetes-related quantitative phenotypes. Hypertension is measured by systolic blood pressures (SBP) and diastolic blood pressures (DBP) (Yang and Wang, 2012), and the correlation coefficient between SBP and DBP was greater than 0.5 in 95% of patients (Gavish et al., 2008). The metabolic syndrome refers to the co-occurrence of insulin resistance, obesity, atherogenic dyslipidemia and hypertension, and these factors are associated and share underlying mediators, pathway and mechanisms (Huang, 2009). The correlations between multiple phenotypes can be leveraged to improve the power of genetic association tests to identify markers associated with one or more of the phenotypes (Aschard et al., 2014). The standard approach to analyze these multiple correlated phenotypes is to perform single-phenotype analyses separately and report the findings for each phenotype (O'Reilly et al., 2012). However, analyzing one phenotype at a time will suffer penalties from the multiple testing and result in a reduced power especially for GWAS (Yang et al., 2010). Recently, the joint analysis of multiple phenotypes has become popular because it can increase statistical power over

analyzing phenotypes separately in detecting genetic variants (Yang et al., 2010; Aschard et al., 2014).

There are three commonly used strategies to detect genetic associations between a genetic variant and multiple correlated phenotypes. The first one is combining test statistics (or p-values) from univariate analysis. This strategy first tests an association between each phenotype and a genetic variant individually and then combines the univariate analysis results, i.e. test statistics or p-values, by using different approaches. The O'Brien's method (O'Brien, 1984), sample splitting and cross-validation method (Yang et al., 2010), Trait-based Association Test that uses Extended Simes procedure (TATES) (van der Sluis et al., 2013), Unified Score-Based Association Test (USAT) (Ray et al., 2016), Fisher's Combination (Yang et al., 2016), and Adaptive Fisher's Combination (AFC) (Liang et al., 2016) belong to this strategy. The advantage of this strategy is its simplicity and is especially useful for analyzing different types of phenotypes such as continuous, dichotomous and survival (Yang and Wang, 2012). The second one is data reduction. This strategy derives a single or a few new phenotypes that are linear combinations of the original phenotypes. Existing methods include projection-based techniques and canonical correlation analysis (CCA). Projection-based approaches include principal components analysis (PCA) and principal component of heritability (PCH), where principal components (PCs) are built to maximize either the phenotypic variance or heritability (Yang and Wang, 2012; Aschard et al., 2014; Klei et al., 2008; Wang et al., 2016). Canonical correlation analysis (CCA) finds the linear combination of phenotypes that explain the largest possible amount of the correlation between the genetic variant and all multiple phenotypes (Ferreira and Purcell, 2009). Data reduction approaches are in general only applicable to multiple phenotypes consisting of all continuous phenotypes that are approximately normally distributed (Yang and Wang, 2012). The third strategy is regression models which include mixed effect models (Zhou and Stephens, 2014; Korte et al., 2012; Casale et al., 2015), the generalized estimating equation (GEE) (Zeger and Liang, 1986; Zhang et al., 2014), and reverse regression methods (O'Reilly et al., 2012; Yan et al., 2013; Wang et al., 2016). The linear mixed effects model (LME) and generalized linear mixed effects model (GLMM) are two commonly used mixed effects models, where the fixed effects are used for the genetic variant and random effects are used to account for phenotypic correlations. The GEE methods collapse the random effects and random residual errors in marginal regression models which are a class of models different from mixed effect models. The reverse regression methods take genotypes as the response variable and multiple phenotypes as predictors, such as the proportional odds logistic regression for joint model of multiple phenotypes (MultiPhen) (O'Reilly et al., 2012). Regression approaches are able to deal with a mixture of continuous, dichotomous, and survival phenotypes, but they are complicated and few available software were developed to implement these methods (Yang and Wang, 2012).

In this article, we developed a novel allele-based method for testing association between multiple phenotypes and a genetic variant. First, we take the allele at the genetic variant as

the response variable and the multiple phenotypes as predictors. Then, we present a new multivariate method that we refer to as WCmulP (Weighted Combination of multiple Phenotypes), inspired by TOW (Test for testing the effect of an Optimally Weighted combination of variants) procedure proposed by Sha et al. (2012) for rare variant association studies and allele-based aproach proposed by Majumdar et al. (2015). For each of the independent individuals, WCmulP linearly combines the multiple phenotypes to "one phenotype" by using the optimal weights proposed by Sha et al. (2012). Then we use the score test based on the logistic model to test the association between the genetic variant and the linear combination of phenotypes. Using extensive simulation studies, we compare the performance of WCmulP with some of the existing methods, MultiPhen (O'Reilly et al., 2012), O'Brien's method (O'Brien, 1984), TATES (van der Sluis et al., 2013), CCA (Ferreira and Purcell, 2009), and SHet (Zhu et al., 2015). Our results show that, in all of the simulation scenarios, WCmulP is either the most powerful test or comparable to the most powerful tests among the methods we compared. Finally, we evaluate the performance of our proposed method using a real data set, the COPDGene study from dbGaP.

2.2 Methods

We consider a sample of n unrelated individuals. Each individual has K possibly correlated phenotypes. Let $Y_{i,k}$ denote the k^{th} phenotype of the i^{th} individual. We propose to use an allele-based logistic regression model to test the association between a variant of interest and multiple phenotypes. For a genetic variant with two alleles, we use x_{2i-1} and x_{2i} to denote the coding of the two alleles of the i^{th} individual such that we use x_1 and x_2 to code the two alleles of the first individual, use x_3 and x_4 to code the two alleles of the second individual, and so on. For a variant with two alleles A and a, if the genotype of the i^{th} individual is AA, we define $x_{2i-1} = x_{2i} = 1$; if the genotype is aa, we define $x_{2i-1} = x_{2i} = 0$; and if the genotype is Aa, we define $x_{2i-1} = 1$ and $x_{2i} = 0$. We define the k^{th} phenotype corresponding to the two alleles x_{2i-1} and x_{2i} of the i^{th} individual as $y_{2i-1,k}$ and $y_{2i,k}$, where $y_{2i-1,k} = y_{2i,k} = Y_{i,k}$. Hence, the total number of observations in the allele-based data is $2n$. We model the relationship between alleles and multiple phenotypes using the inverse logistic regression model

$$\text{logit}(\pi_j) = \alpha + y_{j,1}\beta_1 + y_{j,2}\beta_2 + \cdots + y_{j,K}\beta_K, \quad j = 1, 2, \ldots, 2n, \tag{2.1}$$

where $\pi_j = \Pr\left(x_j = 1 \middle| \boldsymbol{Y}_j = \left(y_{j,1}, \ldots, y_{j,K}\right)^T\right)$, α is the intercept, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^T$ is a K-dimention vector of parameters. To test the association between multiple phenotypes and the variant is equivalent to test the null hypothesis $H_0: \boldsymbol{\beta} = \mathbf{0}$ under equation (2.1). We use the score test statistic given by Sha et al. (2011) to test $H_0: \boldsymbol{\beta} = \mathbf{0}$ under equation (2.1). The test statistic is

$$S = \boldsymbol{U}^T \boldsymbol{V}^{-1} \boldsymbol{U}, \tag{2.2}$$

where $\boldsymbol{U} = \sum_{j=1}^{2n}(x_j - \bar{x})\boldsymbol{Y}_j$, $\boldsymbol{V} = (1-\bar{x})\,\bar{x}\,\sum_{j=1}^{2n}(\boldsymbol{Y}_j - \overline{\boldsymbol{Y}})(\boldsymbol{Y}_j - \overline{\boldsymbol{Y}})^T$, $\bar{x} = \frac{1}{2n}\sum_{j=1}^{2n} x_j$, $\overline{\boldsymbol{Y}} = (\bar{y}_1, \ldots, \bar{y}_K)^T$ and $\bar{y}_k = \frac{1}{2n}\sum_{j=1}^{2n} y_{j,k}$ for $k = 1, \ldots, K$. The test statistic S asymptotically follows a chi-square distribution with K degrees of freedom.

When K is large, the score test may lose power due to the large degrees of freedom. To overcome this problem, we combine the K phenotypes to one variable by using a linear combination of phenotypes, $y_j = \sum_{k=1}^{K} w_k y_{j,k}$, where $w_1, \ldots, w_K$ are the weights. With the linear combination of phenotypes $y_j = \sum_{k=1}^{K} w_k y_{j,k}$, the score test statistic in equation (2.2) becomes

$$S(w_1, \ldots, w_K) = 2n \frac{\left(\sum_{j=1}^{2n}(x_j - \bar{x})y_j\right)^2}{\sum_{j=1}^{2n}(x_j - \bar{x})^2 \sum_{j=1}^{2n}(y_j - \bar{y})^2}. \tag{2.3}$$

We propose to use the optimal weights proposed by Sha et al. (2012), that is, $w_k^o = \frac{\sum_{j=1}^{2n}(x_j-\bar{x})(y_{j,k}-\bar{y}_k)}{\sum_{j=1}^{2n}(y_{j,k}-\bar{y}_k)^2}$ for $k = 1,2,\dots,K$. Actually, the optimal weights $w_1^o,\dots,w_K^o$ maximize $S(w_1,\dots,w_K)$ in equation (2.3). With this optimally weighted combination of phenotypes $y_j^o = \sum_{k=1}^{K} w_k^o y_{j,k}$, the test statistic given in equation (1.3) becomes

$$S(w_1^o,\dots,w_K^o) = 2n \cdot \frac{\sum_{j=1}^{2n}(x_j-\bar{x})\left(y_j^o-\bar{y}^o\right)}{\sum_{j=1}^{2n}(x_j-\bar{x})^2}, \tag{2.4}$$

where $\bar{y}^o = \frac{1}{2n}\sum_{j=1}^{2n} y_j^o$. From equation (2.2) to equation (2.4), we reduced the dimension of the phenotypes from multivariate $(y_{j,k}, k = 1,\dots,K)$ to univariate (y_j^o) with optimal weights w_k^o such that equation (1.4) is the maximum of equation (2.3). Since $w_1^o,\dots,w_K^o$ are data-driven weights, $S(w_1^o,\dots,w_K^o)$ does not follow a chi-square distribution. We use a permutation procedure to evaluate the p-value of $S(w_1^o,\dots,w_K^o)$. In each permutation, we randomly shuffle the genotypes and keep the phenotypes unchanged. Since $\sum_{j=1}^{2n}\left(x_j-\bar{x}\right)^2$ does not change under each permutation, the test statistic $S(w_1^o,\dots,w_K^o)$ is equivalent to

$$T = \sum_{j=1}^{2n}\left(x_j-\bar{x}\right)\left(y_j^o-\bar{y}^o\right). \tag{2.5}$$

This test statistic T is our proposed test statistic to test the effect of the Weighted Combination of multiple Phenotypes (WCmulP).

The WCmulP method can also be extended to incorporate covariates. Suppose that there are p covariates. Let $Z_{i,l}$ denote the l^{th} covariate of the i^{th} individual. We define the l^{th} covariate corresponding to the two alleles x_{2i-1} and x_{2i} of the i^{th} individual as $z_{2i-1,l}$ and $z_{2i,l}$, where $z_{2i-1,l} = z_{2i,l} = Z_{i,l}$. We then adjust the phenotype value $y_{j,k}$ for the covariates by applying linear regressions. That is,

$$y_{j,k} = \alpha_{0,k} + \alpha_{1,k}z_{j,1} + \cdots + \alpha_{p,k}z_{j,p} + \tau_{j,k}.$$

Let $\tilde{y}_{j,k}$ denote the residuals of $y_{j,k}$ in the linear regression. We incorporate the covariate effects in WCmulP by replacing $y_{j,k}$ in equation (2.5) by $\tilde{y}_{j,k}$. With covariates, the statistic of WCmulP is defined as

$$T_{\text{WCmulP}} = T|_{y_{j,k}=\tilde{y}_{j,k}}.$$

2.3 Comparison of Methods

We compare the power of the proposed WCmulP with that of the following methods:

Score (Score test): the test statistic of Score is given by equation (2.2).

OB (O'Brien's method) (O'Brien, 1984): the test statistic of OB, $\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{T}_{\text{uni}}$, is a linear combination of univariate test statistics, and it is the most powerful test among a class of test statistics that are linear combination of $\boldsymbol{T}_{\text{uni}}$, where $\boldsymbol{T}_{\text{uni}}$ is the vector of the univariate test statistics, Σ is the covariance matrix of $\boldsymbol{T}_{\text{uni}}$, and $\boldsymbol{e} = (1,1 \dots,1)^T$ is a 1's vector with length K (the number of phenotypes).

MultiPhen (Joint model of Multiple Phenotypes) (O'Reilly et al., 2012): it uses the proportional odds logistic regression to model the genotype data as ordinal response and phenotypes as predictors. A likelihood ratio test is used to test the null hypothesis.

TATES (Trait-based Association Test that uses Extended Simes procedure) (van der Sluis et al., 2013): it combines univariate p-values to acquire one phenotype-based p-value, while correcting for correlations between phenotypes. The TATES p-value is given by $Min\left(\frac{m_e p_{(k)}}{m_{e(k)}}\right)$, where $p_{(k)}$ is the k^{th} $(k = 1, \dots, K)$ sorted p-value in ascending order, m_e and $m_{e(k)}$ are the effective numbers of independent p-values of all K phenotypes and k specified phenotypes, respectively. The effective numbers can be calculated from the correlation matrix of p-values.

CCA (Canonical Correlation Analysis) (Ferreira and Purcell, 2009): it extracts the linear combination of phenotypes that maximizes the correlations between linear combinations of phenotypes and genotypes at the variant of interest. The test is based on Wilks' lambda and the corresponding F-approximation.

SHet (Test for Heterogeneous genetic effects) (Zhu et al., 2015): The test statistic of SHet, S_{Het}, is based on S_{Hom}, which is the most powerful test statistic when the genetic effect is homogeneous. Both S_{Hom} and S_{Het} are quadratic combinations of the univariate test statistics. The test statistic of S_{Hom} is $S_{Hom} = \frac{e^T(RW)^{-1}T_{\text{uni}}\left(e^T(RW)^{-1}T_{\text{uni}}\right)^T}{e^T(WRW)^{-1}e}$, where R is the correlation matrix of T_{uni}, W is a diagonal matrix of weights for the univariate test statistics, and e is a 1's vector with length K (number of phenotypes). S_{Het} can be viewed as the maximum of S_{Hom}'s satisfying different thresholds. More specifically, given a threshold, only test statistics with absolute values that are greater than the threshold are used, R and W are therefore partially used corresponding to the selected test statistics. The p-values of S_{Het} can be evaluated by simulation.

2.4 Simulation Studies

Our simulations are similar to that of Wang et al. (2016). To evaluate the type I error rates and powers of our method, we simulate genotype-phenotype data sets for n unrelated individuals with total K phenotypes according to a variety of simulation scenarios. Specifically, genotype data at a genetic variant are simulated according to the minor allele frequency (MAF) under the assumption of Hardy-Weinberg equilibrium. We generate K phenotypes by the factor model

$$y = \lambda x + c\gamma f + \sqrt{1-c^2} \times \varepsilon, \tag{2.6}$$

where $y = (y_1, \ldots, y_K)^T$; x is the genotype score at the variant of interest; $\lambda = (\lambda_1, \ldots, \lambda_K)$ is the vector of effect sizes of the genetic variant on the K phenotypes; $f = (f_1, \ldots, f_R)^T \sim MVN(0, \Sigma)$, $\Sigma = (1-\rho)I + \rho A$, R is the number of factors, A is a matrix with elements of 1, I is the identity matrix, and ρ is the correlation between f_i and f_j for $i \neq j$; γ is a K by R matrix; c is a constant number; and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)^T$ is a vector of residuals, $\varepsilon_1, \ldots, \varepsilon_K$ are independent, and $\varepsilon_k \sim N(0,1)$ for $k = 1, \ldots, K$. Based on equation (2.6), we consider the following six models.

Model 1: There is only one factor and genotype has impacts on all traits with the same effect size. That is, $R = 1$, $\lambda = (\beta, \ldots, \beta)^T$, and $\gamma = (1, \ldots, 1)^T$.

Model 2: There are two factors and genotype has impacts on two factors with opposite effects. That is, $R = 2$, $\lambda = \left(\underbrace{-\beta, \ldots, -\beta}_{K/2}, \underbrace{\beta, \ldots, \beta}_{K/2} \right)^T$, and $\gamma = bdiag(D_1, D_2)$, where $D_i = \left(\underbrace{1, \ldots, 1}_{K/2} \right)^T$ for $i = 1, 2$, "$bdiag$" indicates the block diagonal matrix.

Model 3: There are two factors and genotype has impacts on one factor. That is, $R = 2$, $\lambda = \left(0, \ldots, 0, \underbrace{\beta, \ldots, \beta}_{K/2} \right)^T$, and $\gamma = bdiag(D_1, D_2)$, where $D_i = \left(\underbrace{1, \ldots, 1}_{K/2} \right)^T$ for $i = 1, 2$.

Model 4: There are four factors and genotype has impacts on one factor. That is, $R = 4$,

$\lambda = \left(0,\ldots,0,\underbrace{\beta,\ldots,\beta}_{K/4}\right)^T$, and $\gamma = bdiag(D_1, D_2, D_3, D_4)$, where $D_i = \left(\underbrace{1,\ldots,1}_{K/4}\right)^T$ for $i = 1,..,4$.

Model 5: There are four factors and genotype has impacts on two factors. That is, $R = 4$, $\lambda = \left(0,\ldots,0,\underbrace{-\beta,\ldots,-\beta}_{K/4},\underbrace{\beta,\ldots,\beta}_{K/4}\right)^T$, and $\gamma = bdiag(D_1, D_2, D_3, D_4)$, where $D_i = \left(\underbrace{1,\ldots,1}_{K/4}\right)^T$ for $i = 1,..,4$.

Model 6: There are four factors and genotype has impacts on three factors. That is, $R = 4$, $\lambda = \left(0,\ldots,0,\frac{2\beta}{K/4+1}\times 1,\frac{2\beta}{K/4+1}\times 2,\ldots,\frac{2\beta}{K/4+1}\times\frac{K}{4},\underbrace{-\beta,\ldots,-\beta}_{K/4},\underbrace{\beta,\ldots,\beta}_{K/4}\right)^T$, and $\gamma = bdiag(D_1, D_2, D_3, D_4)$, where $D_i = \left(\underbrace{1,\ldots,1}_{K/4}\right)^T$ for $i = 1,..,4$.

In the six models, the within-factor correlation is c^2 and the between-factor correlation is ρc^2. The structures of γ and $\mathrm{cov}(y|x)$ for different numbers of factors ($R = 1, 2, \text{and } 4$) when the number of phenotypes is 8 are given in Table A.2.2.

We also generate phenotypes with covariates effects. We refer to Sha et al. (2012) and Sun et al. (2016) by adding two covariates in equation (2.6) as $y = (0.5z_1 + 0.5z_2)e + \lambda x + c\gamma f + \sqrt{1-c^2}\times\varepsilon$, where z_1 is a continuous random variable generated from a standard normal distribution, z_2 is a binary random variable taking values of 0 and 1 with a probability of 0.5, and e is a K-dimensional vector with all elements being 1's. To evaluate type I error rates and powers, we consider $n = 1{,}000$ unrelated individuals, $MAF = 0.3$, and different numbers of phenotypes $K = 8,16$. To evaluate the type I error rates of all methods, we generate all phenotypes independent of genotypes by setting $\beta = 0$. We evaluate type I error rates at significance levels $\alpha = 0.001$ and 0.01 for all methods. To evaluate powers, we vary the values of β (within-factor correlation $c^2 = 0.5$ and between-factor correlation $\rho c^2 = 0.1$) and vary the values of within-factor correlation c^2 (0.3,0.5,...,0.9) (between-factor correlation $\rho c^2 = 0.1$ and $\beta = 0.1$,).

2.5 Simulation Results

To evaluate the type I error rates of WCmulP and other six methods, we consider different numbers of phenotypes, different significance levels, and different numbers of factors. In each simulation scenario, the p-values of WCmulP and SHet are estimated using 10,000 permutations, and the p-values of Score, MultiPhen, TATES, CCA and OB are estimated using their asymptotic distributions. The type I error rates of the seven methods are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals (CIs) for type I error rates of nominal levels 0.001 and 0.01 are $(0.00038, 0.00162)$ and $(0.008, 0.012)$, respectively. The estimated type I error rates of WCmulP and other six methods are summarized in Table 2.1 ($K = 8$) and Table 2.2 ($K = 16$). From these tables, we can see that all estimated type I error rates of WCmulP are within 95% CIs, which indicates that the proposed WCmulP is a valid test. The estimated type I error rates of SHet, Score, MultiPhen, TATES, CCA and OB are not significantly different from the nominal levels.

For power comparisons, we consider power as a function of genetic effect β (Figures 2.1-2.2) and power as a function of within-factor correlation c^2 (Figures 2.3-2.4). In each of the simulation scenario, the p-values of WCmulP and SHet are estimated using 1,000 permutations and the p-values of Score, MultiPhen, TATES, CCA and OB are estimated using their asymptotic distributions. The powers of the seven methods are evaluated using 1,000 replicated samples at a significance level of 0.01.

Our simulation results show that:

1. As expected, the powers of all methods increase as the genetic effect β increases in each model (Figures 2.1-2.2).
2. WCmulP is either the most powerful test or comparable to the most powerful tests in all six models (Figures 2.1-2.4).
3. As number of phenotypes increases from $K = 8$ to $K = 16$, WCmulP presents more obvious ascendancy than other methods.
4. SHet, Score, MultiPhen, and CCA have similar performance in all six models; we call these four tests as group 1.
5. OB is the most powerful test when the genetic effects are homogeneous (model 1). However, OB reduces power significantly when genetic effects are heterogeneous, especially when opposite directions of the genetic effects exist (models 2, 5-6) or when the genetic variant impacts only a small portion of phenotypes (model 4). This phenomenon was also observed by Zhu et al. (2015).
6. Power comparisons of TATES with tests in group 1 depend on the models. In general, TATES is more powerful than tests in group 1 when the genetic variant impacts on a portion of phenotypes (models 3 and 4).
7. In general, as the within-factor correlation c^2 increases, the powers of all methods decrease (Figures 2.3-2.4). TATES is relatively robust to c^2 because

it essentially only depends on the phenotype that has the strongest association with the genetic variant, as explained in Zhu et al. (2015).

We also considered using principal components (PCs) of the phenotypes instead of the original phenotypes to do power comparisons and the results are given in Figures B.2.1-B.2.4. We exclude PCs that explain less than 10^{-6} of the total variation. Using PCs of the phenotypes, we observe that: (1) WCmulP, Score, MultiPhen, and CCA have very similar powers in all six models (Figures B.2.1-B.2.4). We call these tests as group s1. The tests in group s1 are either the most powerful tests or comparable to the most powerful one; (2) SHet is less powerful than the tests in group s1; (3) OB is the least powerful method in all six models because PCs likely have effects with different directions; (4) TATES becomes the most powerful method when the genetic variant has effects on all phenotypes with the same absolute value of effect sizes (models 1 and 2) because in this case, one of the PCs may capture the most of association information.

We also compared the powers using a lower significance level 5×10^{-5} (Figure B.2.5). Figure B.2.5 shows that the pattern of the power comparisons by using significance level 5×10^{-5} is similar to that by using significance level 0.01 (Figure 2.1).

2.6 Real Data Analysis

Chronic obstructive pulmonary disease (COPD) refers to a group of diseases that cause airflow blockage and breathing-related problems. The Genetic Epidemiology of COPD Study (COPDGene) is a multicenter observational study designed to identify genetic factors associated with COPD, to define and characterize disease-related phenotypes, and to assess the association of disease-related phenotypes with the identified susceptibility genes (Regan et al., 2010). 10,192 participants (including 6,784 non-Hispanic Whites (NHW) and 3,408 African-Americans (AA)) are included in COPDGene. We selected 7 key quantitative COPD-related phenotypes and 4 covariates that are the same as those in Liang et al. (2016). The detailed description of these 7 phenotypes is in Table 2.3, and their correlation structure is given in Figure B.2.6. The four covariates include Body Mass Index, Age, Pack-Years (one pack-year is defined as smoking one pack per day for one year), and gender. A set of 5,430 NHW across 630,860 SNPs were used in the analysis after excluding subjects with missing data in any of the 11 variables.

We apply WCmulP and other six methods to both original 7 phenotypes (Table 2.4) and the principal components (PCs) of the phenotypes (Table A.2.1). PCs that explain less than 10^{-6} of the total variation are excluded. In this way, one PC is excluded and there are 6 PCs left. Using the first few PCs is also a dimension reduction method. Thus, using PCs of the phenotypes, WCmulP uses two dimension reduction methods: using the first few PCs and the weighted combination of those PCs. To identify SNPs significantly associated with the 7 COPD-related phenotypes and the top 6 PCs of the phenotypes, we use the genome-wide significance threshold of 5×10^{-8}. There are total 16 SNPs that are significant under at least one method (Table 2.4 and Table A.2.1). Those 16 SNPs have been reported being associated with the COPD-related phenotypes by previous studies (Pillai et al., 2009; Wilk et al., 2009; Wilk et al., 2012; Cho et al., 2010; Cho et al., 2010; 2012; 2014; Hancock et al., 2010; Young et al., 2010; Li et al., 2011; Zhang et al., 2011; Cui et al., 2014; Zhu et al., 2014; Lutz et al., 2015; Lee et al., 2015). From Table 2.4, we can see that MultiPhen identified the largest number of SNPs, 14 SNPs; WCmulP, SHet, Score, and CCA identified 13 SNPs; TATES identified 9 SNPs; and OB didn't identify any SNPs, that's likely because the true genetic effects of each SNP are heterogeneous for all phenotypes. From Table A.2.1, we can see that using PCs of the phenotypes, WCmulP identified all of the 16 SNPs; MultiPhen identified 15 SNPs; SHet, Score, and CCA identified 13 SNPs; TATES identified 4 SNPs; and OB identified 3 SNPs. In summary, the number of SNPs identified by WCmulP is comparable to the largest number of SNPs identified by other tests; and using PCs of phenotypes, WCmulP is the only method that identified all 16 SNPs. The results of the real data analysis are consistent with our simulation results.

2.7 Discussion

In this article, we developed WCmulP to perform multivariate analysis of multiple phenotypes in association studies based on the following reasons: (1) complex diseases are usually measured by multiple correlated phenotypes in genetic association studies; and (2) there is increasing evidence showing that studying multiple correlated phenotypes jointly may increase powers for detecting genetic variants that are associated with complex diseases. Our results show that WCmulP has correct type I error rates and is either the most powerful test or comparable to the most powerful tests among the seven tests we considered. None of the other methods showed consistent good performances under the simulation scenarios. OB is the most powerful test when the genetic effects are homogeneous, while it loses power dramatically when genetic effects are heterogeneous; especially when opposite directions of the genetic effects exist. SHet, Score, MultiPhen, and CCA have similar powers and they are less powerful than WCmulP in most scenarios. TATES is more powerful only when the genetic variant affects a portion of phenotypes. In addition, in the real data analysis, WCmulP identified 13 (out of 16) significant SNPs, 1 SNP less than the largest number of identified SNPs; using PCs of phenotypes, WCmulP is the only method that identified all 16 SNPs. The real data analysis results show that WCmulP has excellent performance in identifying SNPs associated with complex disease with multiple correlated phenotypes such as COPD.

In the context of association studies, it is important to correct for population stratification (PS). PS refers to allele frequency differences between populations unrelated to the outcome of interest, but due to systematic ancestry differences. PS can cause seriously confounded associations if not adjusted properly (Knowler et al., 1988; Lander and Schork, 1994). The principal component analysis (PCA) method (Chen et al., 2003; Zhang et al., 2003; Zhu et al., 2002; Price et al., 2006; Bauchet et al., 2007) and linear mixed model (LMM) approach (Kang et al., 2010; Zhang et al., 2010; Hoffman, 2013) have been used to adjust for population stratification. There are also other methods such as multidimensional scaling (MDS) (Li and Yu, 2008), the robust PCA based on resampling by half means (RPCA-RHM) (Liu et al., 2013), and the robust PCA based on the projection pursuit (RPCA-PP) (Liu et al., 2013), which are extension methods of the PCA approach. PCA identifies several top principal components of the genotype data matrix and uses them as covariates in the association analysis. We propose to use PCA to control for PS in our proposed method when samples from different populations are involved. However, the performance needs further investigations.

One disadvantage of WCmulP is that the test statistic does not have an asymptotic distribution and a permutation procedure is needed to calculate its p-value, which is time consuming compared to the methods whose test statistics have asymptotic distributions. The running time of WCmulP with 1,000 permutations on a data set with 5,000 individuals and 20 phenotypes on a laptop with 4 Intel(R) Cores(TM) i7-4790 CPU @ 3.6GHz and 4 GB memory is no more than 0.15s. To perform GWAS, we can first select genetic variants

that show evidence of association based on a small number of permutations (e.g. 1,000), and then a large number of permutations are used to test the selected significant genetic variants (Wang et al., 2016). Furthermore, WCmulP cannot be used for rare variant association studies, although recent studies have shown that complex diseases are caused by both common and rare variants (Kang et al., 2010; Bodmer and Bonilla, 2008; Pritchard and Cox, 2002; Teer and Mullikin, 2010; Walsh and King, 2007). How to extend WCmulP to rare variant association studies is our future work.

In our simulation studies, the numbers of phenotypes varied from 8 to 16 and the methods rely on all observations having fully observed phenotypes. However, in real data analysis, as the number of phenotypes increases the chance that missing at least one observation increases exponentially, especially in epidemiological and clinical research (Ali et al., 2011; Dahl et al., 2016). There are several approaches to handle missing phenotypes: deletion-based methods, simple replacement methods, and imputation methods (Ali et al., 2011). The most commonly used method for dealing with missing data is deletion-based method, in which observations with missing values are removed from the analysis (Ali et al., 2011). However, removal of observations with missing values will reduce sample size, thus resulting in power losses (Dahl et al., 2016). The simple replacement methods replace the missing values with plausible values for the variable with missing values, such as the sample mean (van der Sluis et al., 2013; Ali et al., 2011). It is a simple, unconditional method that does not depend on other variables. However, mean substitution approach may result in biased estimates where data are not missing completely at random (Ali et al., 2011). Imputation is a more sophisticated approach that fills in missing values with predicted values using model-based methods or conditional imputation, including multiple imputation (MI), multivariate normal imputation (MVNI), and fully conditional specification (FCS) (Ali et al., 2011; De Silva et al., 2017; Schafer, 1997; Carlin, 2015; Raghunathan et al., 2001; Van Buuren et al., 2006; Carpenter and Kenward, 2012). In MI, the incomplete dataset is generated multiple times and missing values are replaced by values drawn from a posterior distribution according to a suitable imputation model that utilizes the rest of the data (Ali et al., 2011; De Silva et al., 2017). MVNI fits a joint imputation model to all the variables containing missing values under the assumption that the variables follow a multivariate normal distribution (Schafer, 1997; Carlin, 2015). For each variable with missing values, FCS fits separate univariate regression models and iteratively cycles through the univariate regression models (Raghunathan et al., 2001; Van Buuren et al., 2006; Carpenter and Kenward, 2012). In our real data analysis, we removed 1354 observations with missing either phenotypes or covariates from 6784 samples. An alternative approach is to use mean substitution or imputation approaches to fill in the missing values.

2.8 Tables and Figures

Table 2.1. Estimated type I error rates for the seven methods under three simulation settings. The number of phenotypes is $K = 8$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulp and SHet are evaluated using 10,000 permutations. The type I error rate of all of the seven methods is evaluated using 10,000 replicated samples at a significance level of α.

	Type I error rates					
	$R = 1$		$R = 2$		$R = 4$	
Methods	$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.01$
WCmulP	0.0008	0.0097	0.0011	0.0091	0.0011	0.0104
SHet	0.0008	0.0106	0.0009	0.0093	0.0008	0.0104
Score	0.0006	0.0102	0.0008	0.0103	0.0004	0.0105
MultiPhen	0.0011	0.0106	0.0011	0.0105	0.0005	0.0107
TATES	0.0012	0.0094	0.0007	0.0121	0.0004	0.0106
CCA	0.0008	0.0107	0.0010	0.0099	0.0008	0.0107
OB	0.0007	0.0095	0.0016	0.0092	0.0013	0.0105

Note: R is the number of factors.

Table 2.2. Estimated type I error rates for the seven methods under three simulation settings. The number of phenotypes is $K = 16$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulp and SHet are evaluated using 10,000 permutations. The type I error rate of all of the seven methods is evaluated using 10,000 replicated samples at a significance level of α.

	Type I error rates					
	$R = 1$		$R = 2$		$R = 4$	
Methods	$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.01$	$\alpha = 0.001$	$\alpha = 0.01$
WCmulP	0.0011	0.0089	0.0006	0.0094	0.0008	0.0098
SHet	0.0009	0.0098	0.0009	0.0126	0.0008	0.0088
Score	0.0010	0.0096	0.0011	0.0098	0.0010	0.0086
MultiPhen	0.0011	0.0096	0.0011	0.0121	0.0013	0.0103
TATES	0.0013	0.0110	0.0012	0.0102	0.0008	0.0104
CCA	0.0012	0.0097	0.0009	0.0111	0.0011	0.0089
OB	0.0011	0.0085	0.0006	0.0092	0.0007	0.0097

Table 2.3. Description of COPD-related phenotypes

Phenotypes	Descriptions
Gas Trapping (GasTrap)	Air trapping at -856 Hounsfield units (HU) on expiratory chest CT scan
Exacerbation Frequency (ExacerFreq)	Number of COPD exacerbations during the year before study enrollment
Emphysema (Emph)	% Emphysema at -950 HU
Airway Wall Area (Pi10)	Square root of the wall area of a hypothetical 10 mm internal perimeter airway
Emphysema Distribution (EmphDist)	Log ratio of emphysema at -950 HU in the upper 1/3 of lung fields compared to the lower 1/3 of lung fields
Six Minute Walk Distance (6MWD)	Measure of exercise capacity
FEV1	Observed FEV1 (liters)/predicted FEV1 (liters), with predicted values from Hankinson reference equations

Table 2.4. Significant SNPs and the corresponding p-values in the analysis of COPDGene. The p-values of WCmulP are evaluated using 10^9 permutations; the p-values of SHet are evaluated using 10^8 permutations. The p-values of Score, MultiPhen, CCA, TATES, and OB are evaluated using asymptotic distributions. The grayed-out p-values indicate the p-values $> 5 \times 10^{-8}$.

Chr	Variant identifier	WCmulP	SHet	Score	MultiPhen	CCA	TATES	OB
4	rs1512282	0	1.0E-08	1.90E-09	1.03E-09	1.69E-09	5.77E-09	0.339
4	rs1032297	0	0	5.55E-14	7.69E-14	6.52E-14	6.22E-13	0.452
4	rs1489759	0	0	1.11E-16	1.22E-16	1.11E-16	2.52E-16	0.483
4	rs1980057	0	0	1.11E-16	8.14E-17	0	9.35E-17	0.411
4	rs7655625	0	0	1.11E-16	9.13E-17	0	1.64E-16	0.478
15	rs16969968	0	0	1.91E-11	7.84E-12	1.32E-11	2.98E-08	0.986
15	rs1051730	1.00E-08	0	2.05E-11	8.16E-12	1.41E-11	2.63E-08	0.992
15	rs12914385	0	0	1.78E-12	1.48E-12	1.76E-12	5.14E-10	0.999
15	rs8040868	0	0	2.21E-12	2.59E-12	2.74E-12	2.40E-09	0.768
15	rs951266	2.00E-08	0	2.42E-11	1.02E-11	1.77E-11	5.17E-08	0.956
15	rs8034191	4.00E-08	1.0E-08	2.95E-10	7.74E-11	2.14E-10	1.02E-07	0.868
15	rs2036527	4.00E-08	1.0E-08	5.58E-10	1.77E-10	3.99E-10	1.56E-07	0.880
15	rs931794	4.80E-08	3.0E-08	3.13E-10	9.09E-11	2.35E-10	1.18E-07	0.913
15	rs2568494	7.18E-06	1.93E-06	1.22E-07	4.23E-08	1.05E-07	2.88E-05	0.269
15	rs17483721	8.12E-06	2.29E-06	2.26E-07	9.87E-08	2.11E-07	3.57E-05	0.308
15	rs17483929	8.15E-06	2.13E-06	1.65E-07	6.53E-08	1.50E-07	2.82E-05	0.347

Figure 2.1. Power comparisons of the seven methods as a function of β for the six models. The total number of phenotypes is $K = 8$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

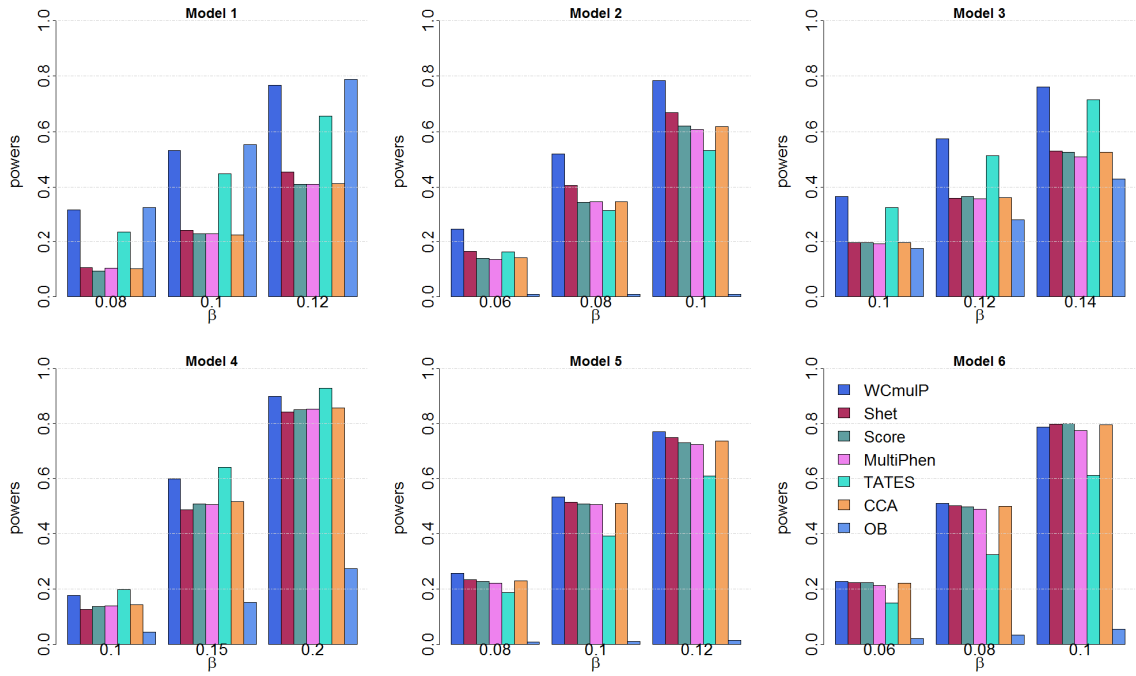


Figure 2.2. Power comparisons of the seven methods as a function of β for the six models. The total number of phenotypes is $K = 16$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

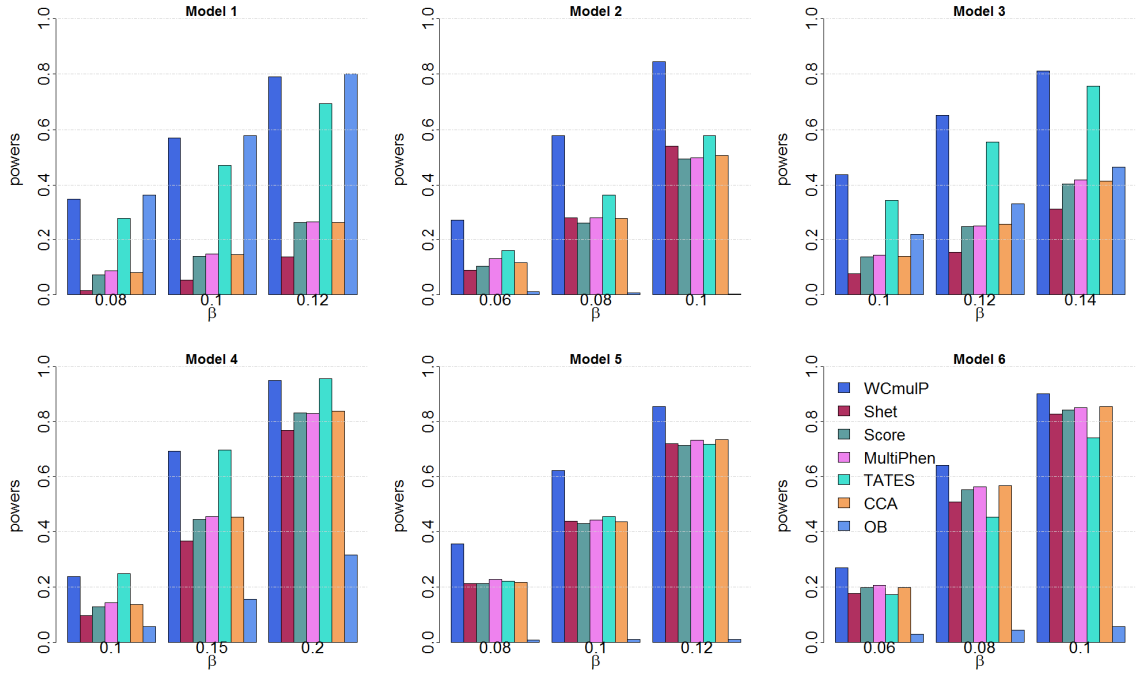


Figure 2.3. Power comparisons of the seven methods as a function of c^2 for the six models. The total number of phenotypes is $K = 8$, $\rho c^2 = 0.1$, $\beta = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations, the p-values of other methods are evaluated using asymptotic distribution. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

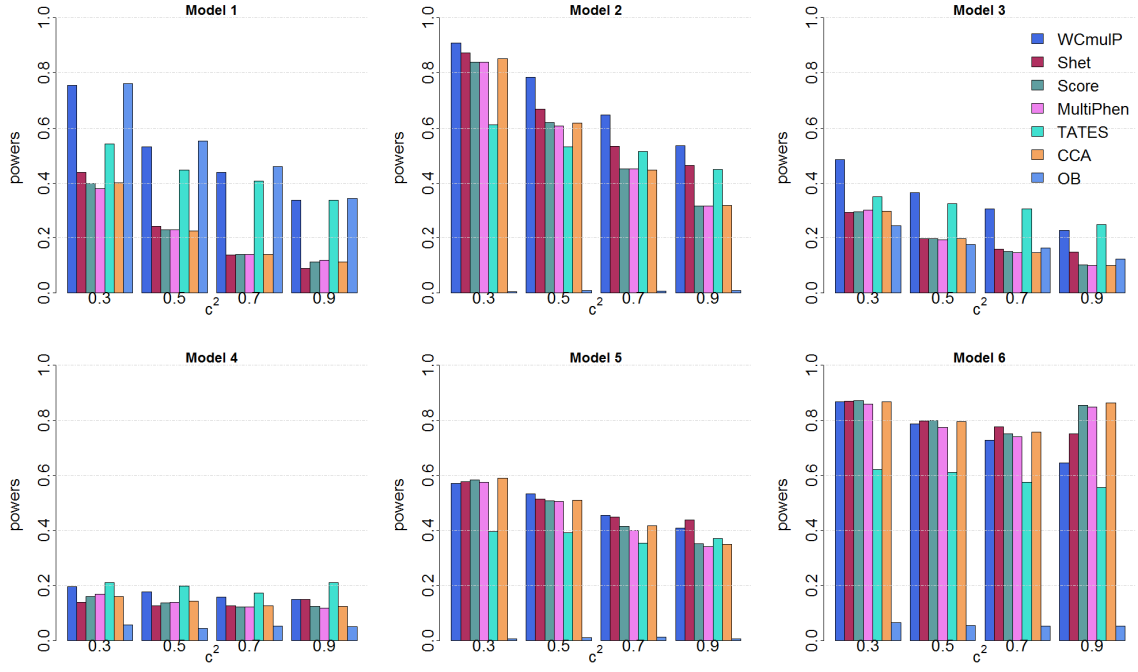
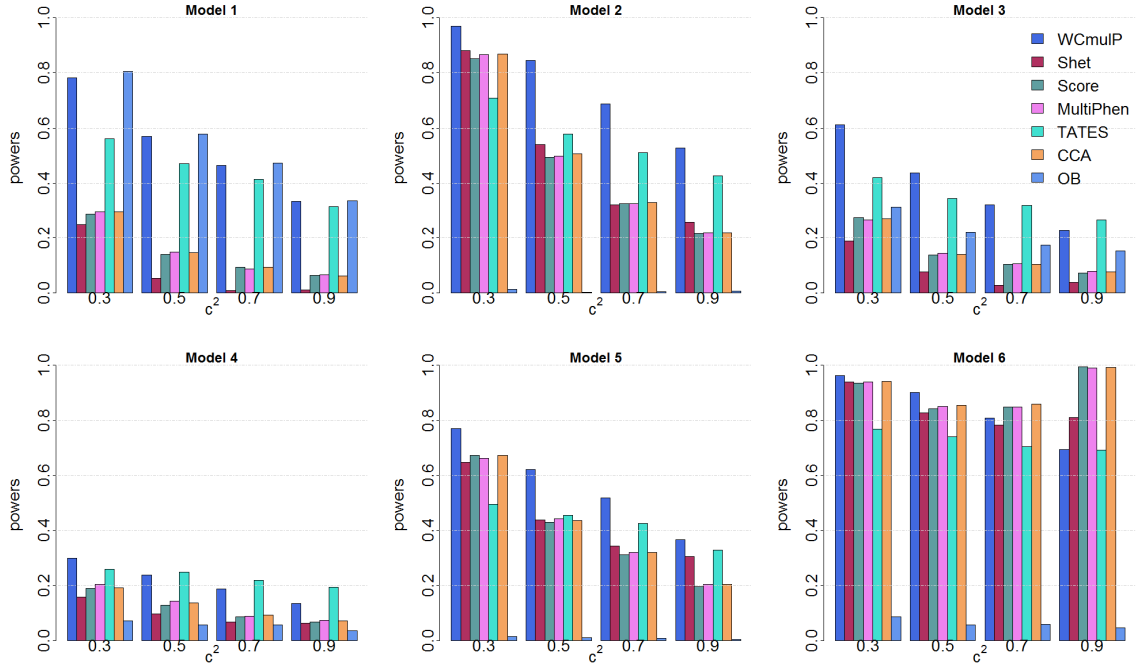


Figure 2.4. Power comparisons of the seven methods as a function of c^2 for the six models. The total number of phenotypes is $K = 16$, $\rho c^2 = 0.1$, $\beta = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations, the p-values of other methods are evaluated using asymptotic distribution. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.



3 Chapter 3

PheCLC: a Novel Statistical Method for Phenome-Wide Association Studies

Over the last decade, genome-wide association studies (GWAS) have been widely performed to identify genetic associations for many complex diseases. Typically, GWAS use a phenotype-to-genotype strategy, starting with a particular phenotype that is associated with genetic variants across the genome, and over 1,000 GWAS have been published linking thousands of statistically significant genetic variants to hundreds of human diseases and traits. A common limitation of GWAS is that they focus on only a single phenotype or a small set of phenotypes at a time. As a complement to GWAS, phenome-wide association studies (PheWAS) use a genotype-to-phenotype approach, beginning with a genotype to test for associations over a broad range of phenotypes. PheWAS were first demonstrated with electronic health record (EHR) data in 2010 and have already demonstrated their capacity to discover genetic association related to a wide range of diseases. In this article, we derived a novel and powerful multivariate method, which we referred as PheCLC, to test the association between a genetic variant with large numbers of phenotypes. Suppose that there is a certain number of phenotypic categories containing different phenotypes. PheCLC first calculates the p-values for testing the variant of interest and the phenotypes within each phenotypic category using a clustering linear combination method recently proposed by our group. Then, it combines the p-values obtained from the first step using the method similar to adaptive Fisher's combination method (Liang et al., 2016). We perform extensive simulation studies to compare the PheCLC method with other existing methods. The results show that our proposed PheCLC method controls the type I error rates very well and has outstanding performance over other methods.

3.1 Introduction

Over the last decade, as the completion of the Human Genome Project (Venter et al., 2001; Lander et al., 2001) and the HapMap Project (Frazer et al., 2007; International HapMap Consortium), our understanding of human genetic variation in the genome and its connection to human health were dramatically accelerated. As a result, genome-wide association studies (GWAS) emerged and have been widely used to identify genes or genetic variants that are associated with a single or a small number of human traits and diseases (Bush and Moore, 2012; Witte, 2010). By 2011, the US National Human Genome Research Institute (NHGRI) and the European Bioinformatics Institute (EMBL-EBI) published the GWAS Catalog (http://www.ebi.ac.uk/gwas/) highlighting a review of GWAS-identified variants, traits and studies (Hindorff et al., 2009). This review demonstrated that almost 5% of single nucleotide polymorphisms (SNPs) and almost 17% of genes are associated with more than one human traits (Sivakumaran et al., 2011). By 2013, over 1000 GWAS had been published linking up to 4000 statistically significant

genetic variants to over 500 human traits and diseases (Hebbring, 2014). In general, GWAS use a phenotype-to-genotype strategy, beginning with a single disease or a small number of diseases, to evaluate the associations between the pre-defined disease or diseases and hundreds of thousands, to over a million, genetic variants across the genome and to identify the significant genetic associations (Hebbring, 2014; Pendergrass et al., 2013; Denny et al., 2016). However, GWAS have several limitations. Frist, a general accepted significance threshold (5×10^{-8}) is commonly used in GWAS and the genetic variant with p-value which is less than this threshold is considered as significant (McCarthy et al., 2008; Risch and Merikangas, 1996). However, reaching this threshold by GWAS can be a challenge due to the burden of large-scale multiple testing. Second, the vast majority (more than 90% as reported by Paul et al., 2014) of significant SNPs are located in the intergenic regions (IGR) which are a subset of noncoding DNA. Identifying and interpreting their associations with human diseases is a major challenge (Hebbring, 2014). Third, GWAS SNPs are mainly tag SNPs (i.e. SNPs that have strong linkage disequilibrium (LD) with a causal variant) for common variants (Hindorff et al., 2009). Fourth, GWAS are examining the association with genetic variants with a limited number of traits and phenotypes (Pendergrass et al., 2012; Pendergrass et al., 2013; Bush et al., 2016; Denny et al., 2016). Therefore, most GWAS fail to identify clinically or biologically significant associations (Hebbring, 2014).

As an alternative approach to GWAS, phenome-wide association studies (PheWAS) is rapidly used to evaluate the impact of one or many genetic variants on a very broad range of phenotypes – the phenome (Pendergrass et al., 2012; Pendergrass et al., 2013; Hebbring, 2014). PheWAS came into view partly due to the availability of dense electronic health record (EHR) data, which is the most frequently used phenome. In 2010, PheWAS was first performed as a method with EHR data and published in Bioinformatics (Denney et al., 2010). Since then, more than half of PheWAS investigations have been demonstrated with EHR data (Denny et al., 2016). However, EHR-based phenotype data are generally collected for clinical use and may depict limited racial diversity (McCarty et al., 2011; Denny et al., 2010). Published PheWAS have been mainly implemented with the Electronic Medical Records and Genomics (Emerge) Network and Population Architecture using Genomics and Epidemiology (PAGE) I Network (Bush et al., 2016). Even though, PheWAS approach is still in an early stage, it has successfully displayed its capability of exploring the association between genetic variants and an extensive range of phenotypes. In the mechanics of conducting genetic association tests, current PheWAS are similar to the widely used GWAS, but from an inverse perspective. PheWAS use a genotype-to-phenotype strategy, starting with a genetic variant of interest to test for associations across the so-called phenome (Hebbring, 2014). Flipping the direction of inference in PheWAS, compared with GWAS, has several motivations. First, genetic variants can be systematically analyzed for their effects on clinical traits and diseases (Bush et al., 2016). Second, genetic variants have long been recognized as factors that influence human diseases and traits, and they may depend on environmental exposures and life stages (Ober and Vercelli, 2011), both of which can be involved in complete phenome. Third, many

conditions have known comorbidities and have multiple genetic factors that contribute to their etiology. Available methods for PheWAS include univariate association test, which conducts a single test of association between the genetic variant of interest and each phenotype at a time, and multivariate methods, which test the genetic variant and a large number of phenotypes jointly. Many GWAS reported that the multivariate tests are more powerful than the univariate tests. Here, we review the commonly adopted multivariate methods for PheWAS: proportional odds logistic regression for joint model of multiple phenotypes (MultiPhen) (O'Reilly et al., 2012), which regresses the genetic variant on multiple phenotypes, fits the proportional odds regression model and uses a likelihood ratio test to obtain the p-value; the Trait-based Association Test that uses Extended Simes procedure (TATES) (van der Sluis et al., 2013), in which univariate p-values are combined to acquire one phenotype-based p-value, while correcting for correlations between phenotypes; the Principal Component of Heritability association test (PCH) (Klei et al., 2008), which reduces the phenotypes to a single trait that has a higher heritability than any other linear combination of the phenotypes; canonical correlation analysis (CCA) (Ferreira and Purcell, 2009), which extracts the linear combination of phenotypes that explain the largest possible amount of the covariation between the genetic variant and all phenotypes; multivariate analysis of variance (MANOVA); and BIMBAM, which is a Bayesian model comparison and model averaging for multivariate regression (Stephens, 2013). Bush et al. (2016) compared the performance of those methods and reported that not a single approach performs best under all scenarios. Zhu et al. (2015) also compared most of the above methods (MultiPhen, TATES, PCH, CCA, and MANOVA) and pointed out that over all the simulation scenarios, MultiPhen, PCH, CCA, and MANOVA have very similar power when a small number of phenotypes are considered (≤ 20).

In this article, we propose a novel and powerful multivariate approach that we refer to as PheCLC (Phenome-wide association study that uses Clustering Linear Combination method). This method can deal with more than one thousand correlated or uncorrelated phenotypes with over two thousand individuals. We suppose that the whole phenome can be classified into numerous phenotypic categories and each category contains a certain number of phenotypes. PheCLC is a two-step approach. In the first step, we apply the clustering linear combination (CLC) method within each phenotypic category and derive a CLC p-value for testing the genetic variant of interest with all phenotypes in that category. The second step then combines all CLC p-values obtained from the first step by using Adaptive Fisher's Combination (AFC) method (Liang et al., 2016). In the simulation studies, we use the factor model (Wang et al., 2016) to generate thousands of phenotypes and evaluate the performance of our proposed PheCLC method, the results show that PheCLC has correct type I error rates and outperforms other methods that we compared with.

3.2 Methods

In our analyses, we consider a sample with n ($n \geq 2{,}000$) unrelated individuals, indexed by $i = 1,2,\ldots,n$. Each individual has K ($K \geq 1{,}000$) phenotypes in total and suppose these K phenotypes are from M phenotypic categories in which the effects of genetic variant are different. Suppose that there are K_m phenotypes in the m^{th} category, where $m = 1,2,\ldots,M$ and $K_1 + \cdots + K_M = K$. Let y_{imk} denote the k^{th} phenotype in the m^{th} category of the i^{th} individual and x_i denote the genotype at the variant of interest for the i^{th} individual. We can also incorporate covariates into the analyses. Suppose that there are p covariates, $z_{i1},\ldots,z_{ip}$, for the i^{th} individual, we adjust both the genotypes and phenotype values for the covariates using the method applied by Price et al. (2006) and Sha et al. (2012). That is, we regress both genotypes and phenotypes on the covariates through the following two linear models

$$y_{imk} = \alpha_{0mk} + \alpha_{1mk} z_{i1} + \ldots + \alpha_{pmk} z_{ip} + \varepsilon_{imk} \text{ and } x_i = \gamma_0 + \gamma_1 z_{i1} + \ldots + \gamma_p z_{ip} + \tau_i .$$

For simplification, in what follows, we assume that there are no covariates. We then use score statistics to test for association between the k^{th} phenotype in the m^{th} category and the variant of interest under the generalized linear model $g(E(y_{imk} \mid x_i)) = \beta_{mk}^0 + \beta_{mk}^1 x_i$. The score test statistic, T_{mk}, is given by

$$T_{mk} = U_{mk} \Big/ \sqrt{V_{mk}} ,$$

where $U_{mk} = \sum_{i=1}^{n} (y_{imk} - \bar{y}_{mk})(x_i - \bar{x})$ and $V_{mk} = \sum_{i=1}^{n} (y_{imk} - \bar{y}_{mk})^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ n$. Under the null hypothesis that there is no association between the genetic variant and the k^{th} phenotype in the m^{th} category (i.e. $\beta_{mk}^1 = 0$), T_{mk} asymptotically follows standard normal distribution. Following these univariate association tests, we obtained K_m such score test statistics in the m^{th} category. Next, we introduce the clustering linear combination (CLC) theorem and then apply CLC method to define an overall test statistic by combining the univariate test statistics in each category.

Theorem of Clustering Linear Combination: we assume that $T = (T_1, \ldots, T_k)^T \sim N(\beta, \Sigma)$, where $\beta = (\beta_1, \ldots, \beta_k)^T$. Suppose $\beta_1, \ldots, \beta_k$ can be divided into L clusters. That is, $\beta = \left(\theta_1 1_{k_1}^T, \ldots, \theta_L 1_{k_L}^T\right)^T$, $1_s = (1, \ldots, 1)^T$, and $k_1 + \cdots + k_L = k$. If the hierarchical clustering method can correctly cluster β, the most powerful test among all tests in the form $(CT)^T (C\Sigma C^T)^{-1} (CT)$ for an arbitrary $L \times k$ matrix C is given by Clustering Linear Combinations (CLC) test with statistic

$$CLC_L = (WT)^T (W\Sigma W^T)^{-1} (WT)$$

where $W = B^T\Sigma^{-1}$ and $B = diag(1_{k_1}, \dots, 1_{k_L})$.

For the m^{th} category, we apply CLC method to combine $T_{m1}, T_{m2}, \dots, T_{mK_m}$ and obtain one CLC test statistic. Let us say that we calculate L_0 (e.g., $L_0 = 10$) CLC test statistics, $CLC_1,\dots,CLC_{L0}$, in the m^{th} category. We then calculate the p-values of $CLC_1,\dots,CLC_{L0}$, $p_{m1},\dots,p_{mL0}$, and take the minimum value of these p-values, denoted $p_m = \min_{1\le l\le L0} p_{ml}$ ($m = 1, \dots, M$). Let $p_{(1)},\dots,p_{(M)}$ be order statistics of $p_1,\dots,p_M$ such that $p_{(1)} \le \dots \le p_{(M)}$. For any predefined integer L and cut points $1 \le c_1 \le \dots \le c_L \le M$ ($(c_1,\dots,c_L) = (1,2,\dots,10)$), we define the summation of negative $\log p_{(m)}$ at cut point c_l as

$$w_l = -\sum_{m=1}^{c_l} \log p_{(m)},\ l = 1,\dots,L.$$

Let P_l denote the p-value of w_l. Then, our proposed test statistic of PheCLC for testing the association between the genetic variant and all phenotypes is given by

$$T = \min_{1\le l\le L} P_l.$$

To calculate the p-value of T, we borrow the permutation procedure in Liang et al. (2016) and state this procedure as follows.

Step 1. In each permutation, we randomly shuffle the genotypes and recalculate $p_{(1)}, \dots, p_{(M)}$ and $w_1, \dots, w_L$. Suppose that we perform B times of permutations. Let $w_l^{(b)}$ ($b = 0,1, \dots, B$) denote the value of w_l based on the b^{th} permuted data, where $b = 0$ represents the original data.

Step 2. We transfer $w_l^{(b)}$ to $P_l^{(b)}$ by

$$P_l^{(b)} = \frac{\#\{d: w_l^{(d)} > w_l^{(b)} \ for\ d=0,1,\dots,B\}}{B}.$$

Step 3. Let $T^{(b)} = \min_{1\le l\le L} P_l^{(b)}$. Then, the p-value of T is given by

$$\frac{\#\{b: T^{(b)} < T^{(0)} \ for\ b=1,2,\dots,B\}}{B}.$$

As shown in Appendix of (Liang et al., 2016), the null distributions of $p_1, p_2, \dots, p_M$ and thus of T do not depend on the genetic variant being tested. Thus, the permutation procedure described above to generate an empirical null distribution of T needs to be done only once for many genetic variants.

3.3 Simulations

To evaluate the type I error rates and powers of our proposed method, we generate genotypes according to minor allele frequency (MAF) and assume Hardy Weinberg Equilibrium. Then, we generate K phenotypes by the factor model (Wang et al., 2016)

$$y = \lambda x + c\boldsymbol{\gamma} f + \sqrt{1-c^2} \times \varepsilon, \tag{3.1}$$

where $y = (y_1, \ldots, y_K)^T$; x is the genotype score at the variant of interest; $\lambda = (\lambda_1, \ldots, \lambda_K)$ is the vector of effect sizes of the genetic variant on the K traits; $f = (f_1, \ldots, f_M)^T \sim MVN(0, \Sigma)$, $\Sigma = (1-\rho) I + \rho A$, A is a matrix with elements of 1, I is the identity matrix, and ρ is the correlation between categories; $\boldsymbol{\gamma}$ is a K by M matrix; c is a constant number; and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)^T$ is a vector of residuals, and $\varepsilon_1, \ldots, \varepsilon_K$ are independent, and $\varepsilon_k \sim N(0,1)$ for $k = 1, \ldots, K$.

Based on equation (3.1), we consider six models:

Model 1: There are $M = 100$ categories and genotypes impact on one category. Let $k = \frac{K}{M}$. That is, $\lambda = \left(\lambda_1^T, \ldots, \lambda_M^T\right)^T$, and $\boldsymbol{\gamma} = diag(D_1, \ldots, D_M)$, where $D_i = \left(\underbrace{1, \ldots, 1}_{k}\right)^T$ and λ_i^T for $i = 1, \ldots, M$ are k dimensional vectors; $\lambda_1 = \cdots = \lambda_{M-1} = 0$ and $\lambda_M = \frac{2\beta}{k+1}(1, \ldots, k)$.

Model 2: There are $M = 100$ categories and genotypes impact on two categories. Let $k = \frac{K}{M}$. That is, $\lambda = \left(\lambda_1^T, \ldots, \lambda_M^T\right)^T$, and $\boldsymbol{\gamma} = diag(D_1, \ldots, D_M)$, where $D_i = \left(\underbrace{1, \ldots, 1}_{k}\right)^T$ and λ_i^T for $i = 1, \ldots, M$ are k dimensional vectors; $\lambda_1 = \cdots = \lambda_{M-2} = 0$, $\lambda_{M-1} = \frac{2\beta}{k+1}(1, \ldots, k)$, and $\lambda_M = \beta\left(\underbrace{1, \ldots, 1}_{k/2}, \underbrace{-1, \ldots, -1}_{k/2}\right)$.

Model 3: There are $M = 100$ categories and genotypes impact on five categories. Let $k = \frac{K}{M}$. That is, $\lambda = \left(\lambda_1^T, \ldots, \lambda_M^T\right)^T$, and $\boldsymbol{\gamma} = diag(D_1, \ldots, D_M)$, where $D_i = \left(\underbrace{1, \ldots, 1}_{k}\right)^T$ and

λ_i^T for $i=1,..,M$ are k dimensional vectors; $\lambda_1=\cdots=\lambda_{M-5}=0$, $\lambda_{M-4}=\frac{2\beta}{k+1}(1,\ldots,k)$, $\lambda_{M-3}=-\frac{2\beta}{k+1}(1,\ldots,k)$, $\lambda_{M-2}=\beta\left(\underbrace{1,\ldots,1}_{k/2},\underbrace{-1,\ldots,-1}_{k/2}\right)$, $\lambda_{M-1}=\beta(1,..,1)$, and $\lambda_M=-\beta(1,..,1)$.

Model 4: There are $M=50$ categories and genotypes impact on one category. Let $k=\frac{K}{M}$. That is, $\lambda=\left(\lambda_1^T,..,\lambda_M^T\right)^T$, and $\boldsymbol{\gamma}=diag(D_1,\ldots,D_M)$, where $D_i=\left(\underbrace{1,\ldots,1}_{k}\right)^T$ and λ_i^T for $i=1,..,M$ are k dimensional vectors; $\lambda_1=\cdots=\lambda_{M-1}=0$ and $\lambda_M=\frac{2\beta}{k+1}(1,\ldots,k)$.

Model 5: There are $M=50$ categories and genotypes impact on two categories. Let $k=\frac{K}{M}$. That is, $\lambda=\left(\lambda_1^T,..,\lambda_M^T\right)^T$, and $\boldsymbol{\gamma}=diag(D_1,\ldots,D_M)$, where $D_i=\left(\underbrace{1,\ldots,1}_{k}\right)^T$ and λ_i^T for $i=1,..,M$ are k dimensional vectors; $\lambda_1=\cdots=\lambda_{M-2}=0$, $\lambda_{M-1}=\frac{2\beta}{k+1}(1,\ldots,k)$, and $\lambda_M=\beta\left(\underbrace{1,\ldots,1}_{k/2},\underbrace{-1,\ldots,-1}_{k/2}\right)$.

Model 6: There are $M=50$ categories and genotypes impact on five categories. Let $k=\frac{K}{M}$. That is, $\lambda=\left(\lambda_1^T,..,\lambda_M^T\right)^T$, and $\boldsymbol{\gamma}=diag(D_1,\ldots,D_M)$, where $D_i=\left(\underbrace{1,\ldots,1}_{k}\right)^T$ and λ_i^T for $i=1,..,M$ are k dimensional vectors; $\lambda_1=\cdots=\lambda_{M-5}=0$, $\lambda_{M-4}=\frac{2\beta}{k+1}(1,\ldots,k)$, $\lambda_{M-3}=-\frac{2\beta}{k+1}(1,\ldots,k)$, $\lambda_{M-2}=\beta\left(\underbrace{1,\ldots,1}_{k/2},\underbrace{-1,\ldots,-1}_{k/2}\right)$, $\lambda_{M-1}=\beta(1,..,1)$, and $\lambda_M=-\beta(1,..,1)$.

To evaluate type I error rates of the proposed method, we let $\beta=0$. To evaluate powers, we let $\beta>0$. In the simulation studies for evaluation of type I error rates and powers, we set $K=1000$, $n=2000$, MAF = 0.3, $c^2=0.5$ and $\rho=0.2$.

3.4 Results

To evaluate the type I error rats of PheCLC and other three methods, we consider different numbers of categories. In each simulation scenario, the p-values of the proposed PheCLC are estimated using 200 permutations, and the p-values of TATES, MANOVA and OB are estimated using their asymptotic distributions. The type I error rates of all of the four methods are evaluated using 1,000 replicates. For 1,000 replicates, the 95% confidence intervals (CIs) for type I error rates of nominal levels 0.05 is (0.036, 0.063). The estimated type I error rates of PheCLC and other three methods are summarized in Table 3.1. Well-controlled type I error rates were observed for PheCLC, TATES, MANOVA and OB in all scenarios. Namely, the proposed PheCLC is proved to be a valid method.

To compare the powers of PheCLC with TATES, MANOVA and OB, we consider different sample sizes, different numbers of phenotypes for all six models. The power of the four methods at $\alpha = 0.05$ is compared in Figure 3.1. We summarize the power comparison results as follows.

1. PheCLC method outperformed TATAS, MANOVA and OB consistently in almost all scenarios.
2. OB is the most powerful test among a class of tests with linear combination of univariate test statistics being the overall test statistic when the genetic effects are equal to each other (Yang et al., 2010). While in all of the six scenarios we considered, both zero and nonzero values were assigned to the genetic effects, which explains why the powers of OB are very low.
3. When genotypes have effects on more than two categories, the powers of PheCLC are more than double that of TATES.

In most scenarios, MANOVA is more powerful than TATES, and its powers are at least 40% lower than that of PheCLC.

3.5 Discussion

Over the last decade, GWAS have identified thousands of genetic variants that are associated with human diseases and traits (Pendergrass et al., 2012; Pendergrass et al., 2013; Hebbring, 2014; Bush et al., 2016). Unfortunately, most GWAS have a limitation that they only focus on a pre-defined and a very limited number of phenotypes, as a result, they fail to identify clinically significant associations in which large numbers of phenotypes are involved (Pendergrass et al., 2012; Pendergrass et al., 2013; Denny et al., 2016; Bush et al., 2016). As a complementary approach to GWAS, PheWAS can examine the associations between genetic variants and a broad set of phenotypes and have demonstrated to be effective in identifying significant variants (Hebbring, 2014; Pendergrass and Ritchie, 2015; Pendergrass et al., 2015). In 2010, the first PheWAS was published in Bioinformatics and it was a proof-of-concept (PoC) study that was not designed to reveal new discoveries but to verify known GWAS-significant associations and to validate the approach (Denny et al., 2010; Denny et al., 2011; Hebbring et al., 2013; Liao et al., 2013; Pendergrass et al., 2013; Ritchie et al., 2013; Shameer et al., 2014; Bush et al., 2016). Unlike GWAS that have been widely used for nearly a decade, PheWAS approach is still in its infancy. Even so, they have already demonstrated their potentiality to identify important genetic associations. PheWAS can be used to explore pleiotropy of genetic variants in a broad spectrum of phenotypes, discover interrelationships between phenotypes, and characterize the genetic architecture of many complex traits (Pendergrass et al., 2012).

However, the PheWAS approaches have presented several challenges (Pendergrass et al., 2012; Pendergrass et al., 2013; Hebbring 2014; Denny et al., 2016; Bush et al., 2016). The biggest challenge of PheWAS is interpretations. For example, when there exists a significant association between a genetic variant and multiple phenotypes, multiple possible interpretations must be taken into consideration, such as true pleiotropy, false phenotype distinction because of mis-characterization of the true phenotype, and confounded phenotype relationships (Bush et al., 2016). Furthermore, the phenome is not universally defined and PheWAS have been limited by how well the human phenome can be defined (Hebbring, 2014). The International Classification of Disease (ICD) codes are an internationally recognized, standardized coding system used to define disease status and the two widely used versions are the ninth and tenth (ICD9/10) (Denny et al., 2010; Hebbring, 2014). Unfortunately, not every ICD code is equal (McCarty et al., 2007), and it's infeasible to manually assess the validity of the whole phenome for all patients. Second, like GWAS, PheWAS are hypothesis-generating approaches that are challenged by multiple hypotheses (Hebbring, 2014). However, unlike GWAS in which a significance threshold 5.0×10^{-8} is commonly used, a generally accepted threshold has not emerged due to the variations in the number of phenotypes and genotypic variants in PheWAS (Bush et al., 2016). One way to correct for multiple testing is Bonferroni correction; however it may not be an appropriate comparison adjustment due to the non-independence of the phenotypes as well as the variants (Hebbring, 2014; Verma et al., 2016). In addition,

differences across populations may affect the ability to validate findings (Hebbring, 2014). For example, it's likely to see different results from a population with European ancestry compared with a population with African ancestry due to significant differences in the linkage disequilibrium structure and allele frequencies between the two populations (Hebbring, 2014).

The PheWAS concept is not new, but it was not developed as a method until very recently due to the limited availability of cohorts with multiple phenotypes (Bush et al., 2016). Examples of available tests for PheWAS include MultiPhen (O'Reilly et al., 2012), PCH (Klei et al., 2008), TATES (van der Sluis et al., 2013), CCA (Ferreira and Purcell, 2009), MANOVA, and BIMBAM (Stephens, 2013). In this article, we proposed PheCLC to test the association between a genetic variant and a wide spectrum of phenotypes. PheCLC takes into account the possibility that phenotypes are from different phenotypic categories, which is common in PheWAS due to the very large number of phenotypes, and conducts genetic association tests in each category. Then, the p-values obtained from all categories are combined using AFC method (Liang et al., 2016) in order to derive the PheCLC p-value. Even though PheCLC has already demonstrated its validity and capacity in the simulation studies, there are still some improvements in this method. Bush et al. (2016) pointed out the necessity of using meta-analysis for PheWAS, and how to apply PheCLC to meta-analysis becomes one of our future projects. In addition, we will extend PheCLC approach to perform pathway-based analysis. As Hebbring (2014) reported that, the pathways may play an essential role in many disease aetiologies.

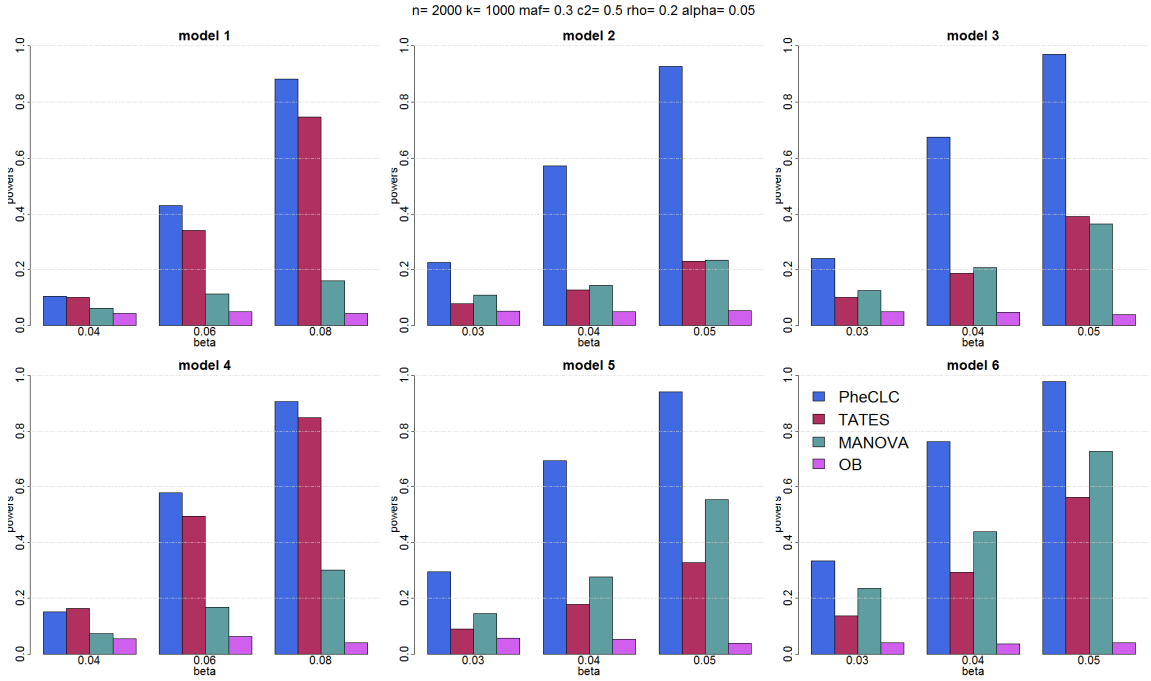
To summarize, PheCLC has well-controlled type I error rates and is more powerful than the currently available methods for PheWAS in most scenarios that we studied. PheCLC allows researchers to test genetic associations for a large number of phenotypes that are from different phenotypic categories.

3.6 Tables and Figures

Table 3.1. Estimated type I error rates for the four methods with different number of categories. The sample size is 2000, the total number of phenotypes is K=1000, rho is 0.2, c^2=0.5 and MAF is 0.3. The type I error rate is evaluated using 1,000 replicated samples at a significance level of 0.05.

Methods	Type I error rates	
	$M = 100$	$M = 50$
PheCLC	0.053	0.058
TATES	0.058	0.038
MANOVA	0.040	0.058
OB	0.038	0.040

Figure 3.1. Power comparisons of the four methods as a function of beta for the six models. The sample size is 2000, total number of phenotypes is $K = 1000$, MAF is 0.3, c2 is 0.5, and rho is 0.2. The power of all of the four methods is evaluated using 1,000 replicated samples at a significance level of 0.05.



4 Chapter 4

A Novel Statistical Method for Rare-variant Association Studies in General Pedigrees

Both population-based and family-based designs are commonly used in genetic association studies to identify rare variants that underlie complex diseases. For any type of study design, the statistical power will be improved if rare variants can be enriched in the samples. Family-based designs, with ascertainment based on phenotype, may enrich the sample for causal rare variants and thus can be more powerful than population-based designs. Therefore, it is important to develop family-based statistical methods that can account for ascertainment. In this paper, we develop a novel statistical method for rare-variant association studies in general pedigrees for quantitative traits. This method uses a retrospective view that treats the treat as fixed and the genotype as random, which allows us to account for complex and undefined ascertainment of families. We then apply the newly developed method to the Genetic Analysis Workshop 19 data set and compare the power of the new method with two other methods for general pedigrees. The results show that the newly proposed method increases power in most of the cases we consider, more than the other methods.

4.1 Background

There is increasing interest in detecting associations between rare variants and complex traits. Although statistical methods to detect common variant associations are well developed, these variant-by-variant methods may not be optimal for detecting associations with rare variants as a result of allelic heterogeneity as well as the extreme rarity of individual variants (Li and Leal, 2008). Recently, several statistical method for detecting associations of rare variants were developed for population-based designs, including the cohort allelic sums test (Morgenthaler and Thilly, 2007), the combined multivariate and collapsing method (Li and Leal, 2008), the weighted sum statistic (Madsen and Browning, 2009), the variable minor allele frequency threshold method (Price et al., 2010), the adaptive sum test (Han and Pan, 2010), the set-up method (Hoffmann et al., 2010), the sequence kernel association test (Wu et al., 2011), and the test for optimally weighted combination of variants (Sha et al., 2012).

Meanwhile, quite a few statistical methods for rare variant association studies have been developed for family-based designs. For any type of study design, the statistical power will be improved if rare variants can be enriched in the samples. If one parent has a copy of a rare allele, half of the offspring are expected to carry it, and, hence, variants that are rare in the general population could be very common in certain families (Shi and Rao, 2011). Therefore, family-based designs may plan an important role in rare-variant association studies. Because of the importance of family-based designs in rare-variant association

studies, several family-based rare-variant association methods for quantitative traits (Liu and Leal, 2012; Chen et al., 2013; Svishcheva et al., 2014) and for qualitative traits (Zhu et al., 2010; Feng et al., 2011; Zhu and Xiong, 2012) have been developed. However, most of these methods were developed under the assumption of random ascertainment and family-based designs with random ascertainment may not yield enrichment of rare variants. To analyzing the sequencing data in general pedigrees provided by Genetic Analysis Workshop 19 (GAW19), we proposed a novel method to the GAW19 data set, we compared the power of the proposed method with that of two popular methods for family-based designs.

4.2 Methods

Consider a sample of n pedigrees with n_i members in the i^{th} pedigree and a genomic region with M variants. Let y_{ij} and $g_{ij} = \left(g_{ij1}, \dots, g_{ijM}\right)^T$ denote the trait value and genotypes of the M variants in the genomic region for the j^{th} individual in the i^{th} pedigree. Let $x_{ij} = \sum_{m=1}^{M} w_m g_{ijm}$ denote the weighted combination of genotypes at the M variants, where $w = (w_1, \dots, w_M)^T$ is a weight function.

For given genotypes, we assume that $y_{ij} \sim N\left(\alpha + x_{ij}\beta, \sigma^2\right)$. Using the notation $g_i = \left(g_{i1}, \dots, g_{in_i}\right)^T$, the retrospective likelihood is given by

$$RL = \prod_{i=1}^{n} \Pr\left(g_i \middle| y_{i1}, \dots, y_{in_i}\right)$$

$$= \prod_{i=1}^{n} \frac{\Pr\left(y_{i1}, \dots, y_{in_i} | g_i\right) \Pr(g_i)}{\sum_{g_i^*} \Pr\left(y_{i1}, \dots, y_{in_i} \middle| g_i^*\right) \Pr(g_i^*)}$$

$$= \prod_{i=1}^{n} \frac{\exp\left(-\sum_{j=1}^{n_i} \left(y_{ij} - \alpha - x_{ij}\beta\right)^2 / 2\sigma^2\right) \Pr(g_i)}{\sum_{g_i^*} \exp\left(-\sum_{j=1}^{n_i} \left(y_{ij} - \alpha - x_{ij}^*\beta\right)^2 / 2\sigma^2\right) \Pr(g_i^*)}$$

where $\sum_{g_i^*}$ represents the summation of all possible genotypes. Based on the RL, the score test statistic for testing the null hypothesis $H_0: \beta = 0$ is given by

$$T_{score} = U^2 / V \tag{4.1}$$

where $U = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}\right)\left(y_{ij} - \bar{y}\right), V = w^t \Sigma w \sum_{i=1}^{n} y_i^T \Phi_i y_i, y_i = \left(y_{i1}, \dots, y_{in_i}\right)^T, \bar{y} = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} y_{ij}, \Phi_i$ is twice the kinship coefficient of the i^{th} pedigree, and $\Sigma = cov(g_{11}, g_{11})$ is the covariance matrix of the multiple variant genotype of one individual. Σ can be estimated by $\hat{\Sigma} = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \sum_{j=1}^{n_i} g_{ij}$. It is worth pointing out that T_{score} is equivalent to the quantitative version of the retrospective likelihood score statistic proposed by Schaid et al (2013).

Because rare variants are essentially independent, following Pan (2000) and Sha et al (2012), we replace $\hat{\Sigma}$ by $\hat{\Sigma}_0 = diag\left(\hat{\Sigma}\right)$. Then, the score test statistic T_{score} becomes

$$T_0(w) = w^T u u^T / w^T \hat{\Sigma}_0 w \sum_{i=1}^{n} y_i^T \Phi y_i$$

where $u = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (g_{ij} - \bar{g})(y_{ij} - \bar{y})$. As a function of w, $T_0(w)$ reaches its maximum when $w = \hat{\Sigma}_0^{-1} u$ and the maximum value of $T_0(w)$ is $u^T \hat{\Sigma}_0^{-1} u / \sum_{i=1}^{n} y_i^T \Phi y_i$. We define the statistic of optimally weighted score test (OW-score) as

$$T_{OW-score} = u^T \hat{\Sigma}_0^{-1} u / \sum_{i=1}^{n} y_i^T \Phi y_i = \sum_{m=1}^{M} (u_m^2 / \sigma_{mm}) / \sum_{i=1}^{n} y_i^T \Phi y_i$$

where σ_{mm} is the $(m,m)^{th}$ element of $\hat{\Sigma}_0$ and u_m is the m^{th} element of u. Under the null hypothesis, $T_{OW-score}$ is asymptotically distributed as a mixture of independent χ^2 statistics (Liu et al., 2007; Liu et al., 2009). Alternatively, the distribution of $T_{OW-score}$ can be approximated by a Satterwaite approximation for the distribution of quadratic forms (Wu et al., 2011; Kwee et al., 2008; Liu et al., 2008) or a scaled χ^2 distribution (Schaid et al., 2013). We propose to approximate the distribution of $T_{OW-score}$ by a scaled χ^2 distribution with the scale δ and degree of freedom d estimated by the expectation and variance of $T_{OW-score}$. Note that $u \sim N(0, \Sigma \sum_{i=1}^{n} y_i^T \Phi y_i)$. We have $\hat{\mu}_T = \hat{E}(T_{OW-score}) = trace(\hat{\Sigma} \hat{\Sigma}_0^{-1})$ and $\hat{\sigma}_T^2 = \widehat{var}(T_{OW-score}) = 2trace(\hat{\Sigma} \hat{\Sigma}_0^{-1} \hat{\Sigma} \hat{\Sigma}_0^{-1})$. Then, the scale δ is estimated as $\hat{\delta} = \hat{\sigma}_T^2 / (2 \hat{\mu}_T)$ and the degree of freedom d is estimated as $\hat{d} = 2 \hat{\mu}_T^2 / \hat{\sigma}_T^2$.

We compare the performance of our OW-score with (a) WS-score, the score test given by equation (4.1) with weight given by Madsen and Browning (2009) and (b) famSKAT, family-based sequence kernel association test given by Chen et al (2013).

4.3 Results

We applied our proposed method as well as the WS-score test and famSKAT to the simulated data from GAW19. All tests were conducted on 849 individuals, from 20 pedigrees, that had no missing genotypes or phenotypes. Sex, age, blood pressure medication status, and smoking status were considered as covariates in this study. We were aware of the underlying simulation model.

There are two related phenotypes, systolic blood pressure (SBP) and diastolic blood pressure (DBP), at three time points. We considered the average of DBP at three time points as the phenotype of interest in our analysis. We compared the power of the three tests (OW-score, WS-score, and famSKAT) to detect association between each of the top 14 genes that influence the phenotype of interest. We used the variants between the first functional single nucleotide polymorphism (SNP) and the last functional SNP in each gene in our analysis. We did not consider CABP2 because the power of the three tests are essentially the same due to the only one variant in this gene. To adjust the effects of the covariates on the phenotype of interest, we first applied a linear model by regressing the phenotype of interest on the covariates: sex, the average of age, the average of blood pressure medication status, and the average of smoking status. The power comparisons based on the 200 replicated data sets are given in Table 4.1. Significance level is assessed at 5%. This table shows that the OW-score test identified three genes with power greater than 40%, famSKAT identified 1 gene with power greater than 40%, and the WS-score test could not identify any genes with power greater than 40%. OW-score and famSKAT have different power mainly because they use different weights. Let w_m and W_m denote the weights, rescaled to the interval (0, 1), of the OW-score test and famSKAT for the m^{th} variant. Then, $w_m > W_m$ when minor allele frequency (MAF) is less than 0.01; $w_m \leq W_m$ when MAF is in the interval (0.01, 0.05); $w_m > W_m$ when MAF is greater than 0.05. The OW-score test has much higher power than famSKAT for RAI1 and REPIN1 because none of the MAFs of the causal variants in RAI1 and REPIN1 are in the interval (0.01, 0.05).

We also evaluated the type I error rate of the proposed OW-score test. To evaluate the type I error, we used 1000 blocks (100 variants in each block) from chromosome 5 that are far from causal variants. In each block, we applied the OW-score test to each of the 200 replicates to test association between genotypes and the phenotype of interest. We obtained one p-value for each replicate and each block. The type I errors of the proposed test were 0.04887, 0.00921, and 0.00131 at significance levels of 0.05, 0.01, and 0.001, respectively. We also considered the average of SBP at three time points as the phenotype of interest, which yielded similar results.

4.4 Discussion

Next-generation sequencing technologies make directly testing rare variant association possible. However, the development of powerful statistical methods for rare variant associations studies is still underway. In this article, we proposed a novel statistical method for rare-variant association studies based on general pedigrees for quantitative traits. The application to the GAW19 data set showed that the proposed method has correct type I error rate and is more powerful than the other two methods against which our method was compared.

We described our method for quantitative traits. For qualitative traits, we can derive a score test similar to that given by equation (4.1). However, the performance of the proposed method for qualitative traits requires further investigation. Like many statistical methods for rare variant associations studies, the proposed method can consider phenotype measurement at only one time point. Statistical methods based on sequence data have been developed for unrelated individuals that have phenotype measurements at multiple time points (Wang et al., 2014). From a statistical standpoint, modeling using longitudinal phenotypes is more informative than using phenotypes at a single time point and thus can increase the power of an association test (Wang et al., 2014; Furlotte et al., 2012). Our future work includes extension of the proposed method to longitudinal phenotypes.

4.5 Conclusions

In this article, we developed a novel statistical method for rare variant association studies in general pedigrees (randomly ascertainment pedigrees or ascertained pedigrees). Application to GAW19 data set showed that the newly proposed method is more powerful than the other two methods in most of the cases. Our new method uses a retrospective view, which allows us to account for complex and undefined ascertainment of families. The GAW19 data is based on randomly ascertained pedigrees. Results of applying our method to GAW19 data showed that the proposed method has correct type I error based on random ascertainment. When random ascertainment is violated and ascertainment is based on trait values, the proposed method is expected to have correct type I error. If pedigrees are ascertained because of extreme trait values, the proposed method is expected to have higher power than methods based on randomly ascertained pedigrees.

4.6 Table

Table 4.1. Power comparisons of the three tests using the average of DBP at three time points as phenotypes. Significance level is assessed at 5%.

Genes	$T_{OW-score}$	$T_{WS-score}$	FamSKAT
CGN	0.135	0	0.035
FLT3	0.005	0	0.08
LEPR	0.05	0.015	0.065
MAP4	0.175	0.185	**0.425**
MTRR	**0.465**	0.005	0.06
NRF1	0	0.005	0.035
PTTG1IP	0.02	0.145	0.06
RAI1	**0.845**	0.005	0.155
REPIN1	**0.915**	0.05	0.085
SLC35E2	0.005	0	0.05
TNN	0	0	0.035
ZFP37	0	0.005	0.005
ZNF443	0.01	0.015	0.195
ZNF544	0.005	0.015	0.06

5 Reference List

Ali, A., et al. Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. British journal of cancer 2011;104(4):693.

Aschard, H., et al. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. The American Journal of Human Genetics 2014;94(5):662-676.

Bauchet, M., et al. Measuring European population stratification with microarray genotype data. The American Journal of Human Genetics 2007;80(5):948-956.

Bodmer, W. and Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. Nature genetics 2008;40(6):695.

Bush, W.S. and Moore, J.H. Genome-wide association studies. PLoS computational biology 2012;8(12):e1002822.

Bush, W.S., Oetjens, M.T. and Crawford, D.C. Unravelling the human genome–phenome relationship using phenome-wide association studies. Nature Reviews Genetics 2016;17(3):129.

Carlin, J. Multiple imputation: a perspective and historical overview. Handbook of Missing Data 2015.

Carpenter, J. and Kenward, M. Multiple imputation and its application. John Wiley & Sons; 2012.

Casale, F.P., et al. Efficient set tests for the genetic analysis of correlated traits. Nature methods 2015;12(8):755.

Chavali, S., et al. Network properties of human disease genes with pleiotropic effects. BMC systems biology 2010;4(1):78.

Chen, H., Meigs, J.B. and Dupuis, J. Sequence kernel association test for quantitative traits in family samples. Genetic epidemiology 2013;37(2):196-204.

Chen, H.S., et al. Qualitative Semi-Parametric Test for Genetic Associations in Case-Control Designs Under Structured Populations. Annals of human genetics 2003;67(3):250-264.

Cho, M.H., et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nature genetics 2010;42(3):200.

Cho, M.H., et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Human molecular genetics 2011;21(4):947-957.

Cho, M.H., et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. The lancet Respiratory medicine 2014;2(3):214-225.

Consortium, I.H. A haplotype map of the human genome. Nature 2005;437(7063):1299.

Consortium, I.H. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851.

Cronin, R.M., et al. Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Frontiers in genetics 2014;5:250.

Cui, K., Ge, X. and Ma, H. Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. PloS one 2014;9(7):e102324.

Dahl, A., et al. A multiple-phenotype imputation method for genetic studies. Nature genetics 2015;47(3):466.

De Silva, A.P., et al. A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. BMC medical research methodology 2017;17(1):114.

Denny, J.C., Bastarache, L. and Roden, D.M. Phenome-wide association studies as a tool to advance precision medicine. Annual review of genomics and human genetics 2016;17:353-373.

Denny, J.C., et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome-and phenome-wide studies. The American Journal of Human Genetics 2011;89(4):529-542.

Denny, J.C., et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. Bioinformatics 2010;26(9):1205-1210.

Feng, T., Elston, R.C. and Zhu, X. Detecting rare and common variants for complex traits: sibpair and odds ratio weighted sum statistics (SPWSS, ORWSS). Genetic epidemiology 2011;35(5):398-409.

Ferreira, M.A. and Purcell, S.M. A multivariate test of association. Bioinformatics 2008;25(1):132-133.

Furlotte, N.A., Eskin, E. and Eyheramendy, S. Genome-Wide Association Mapping With Longitudinal Data. Genetic epidemiology 2012;36(5):463-471.

Galesloot, T.E., et al. A comparison of multivariate genome-wide association methods. PloS one 2014;9(4):e95923.

Gavish, B., Ben-Dov, I.Z. and Bursztyn, M. Linear relationship between systolic and diastolic blood pressure monitored over 24 h: assessment and correlates. Journal of hypertension 2008;26(2):199-209.

Han, F. and Pan, W. A data-adaptive sum test for disease association with multiple common or rare variants. Human heredity 2010;70(1):42-54.

Hancock, D.B., et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nature genetics 2010;42(1):45.

Hebbring, S.J. The challenges, advantages and future of phenome-wide association studies. Immunology 2014;141(2):157-165.

Hebbring, S.J., et al. A PheWAS approach in studying HLA-DRB1* 1501. Genes and immunity 2013;14(3):187.

Hindorff, L.A., et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences 2009;106(23):9362-9367.

Hoffman, G.E. Correcting for population structure and kinship using the linear mixed model: theory and extensions. PloS one 2013;8(10):e75707.

Hoffmann, T.J., Marini, N.J. and Witte, J.S. Comprehensive approach to analyzing rare genetic variants. PloS one 2010;5(11):e13584.

Hotelling, H. Relations between two sets of variates. Biometrika 1936;28(3/4):321-377.

Huang, P.L. A comprehensive definition for metabolic syndrome. Disease models & mechanisms 2009;2(5-6):231-237.

Kang, H.M., et al. Variance component model to account for sample structure in genome-wide association studies. Nature genetics 2010;42(4):348.

Klei, L., et al. Pleiotropy and principal components of heritability combine to increase power for association analysis. Genetic epidemiology 2008;32(1):9-19.

Knowler, W.C., et al. Gm3; 5, 13, 14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. American journal of human genetics 1988;43(4):520.

Korte, A., et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature genetics 2012;44(9):1066.

Kwee, L.C., et al. A powerful and flexible multilocus association test for quantitative traits. The American Journal of Human Genetics 2008;82(2):386-397.

Lan, H., et al. Dimension reduction for mapping mRNA abundance as quantitative traits. Genetics 2003;164(4):1607-1614.

Lander, E.S., et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

Lander, E.S. and Schork, N.J. Genetic dissection of complex traits. Science 1994;265(5181):2037-2048.

Lange, C., et al. A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. Statistical applications in genetics and molecular biology 2004;3:Article17.

Lee, J.H., et al. IREB2 and GALC are associated with pulmonary artery enlargement in chronic obstructive pulmonary disease. American journal of respiratory cell and molecular biology 2015;52(3):365-376.

Li, B. and Leal, S.M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. The American Journal of Human Genetics 2008;83(3):311-321.

Li, Q. and Yu, K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genetic epidemiology 2008;32(3):215-226.

Li, X., et al. Importance of hedgehog interacting protein and other lung function genes in asthma. Journal of Allergy and Clinical Immunology 2011;127(6):1457-1465.

Liang, X., et al. An Adaptive Fisher's Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies. Scientific reports 2016;6:34323.

Liao, K.P., et al. Associations of autoantibodies, autoimmune risk alleles, and clinical diagnoses from the electronic medical records in rheumatoid arthritis cases and non–rheumatoid arthritis controls. Arthritis & Rheumatology 2013;65(3):571-581.

Liu, D., Ghosh, D. and Lin, X. Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. BMC bioinformatics 2008;9(1):292.

Liu, D., Lin, X. and Ghosh, D. Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models. Biometrics 2007;63(4):1079-1088.

Liu, D.J. and Leal, S.M. A unified framework for detecting rare variant quantitative trait associations in pedigree and unrelated individuals via sequence data. Human heredity 2012;73(2):105-122.

Liu, H., Tang, Y. and Zhang, H.H. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. Computational Statistics & Data Analysis 2009;53(4):853-856.

Liu, L., et al. Robust methods for population stratification in genome wide association studies. BMC bioinformatics 2013;14(1):132.

Locke, D.E., et al. Relationship of indicators of neuropathology, psychopathology, and effort to neuropsychological results in patients with epilepsy or psychogenic non-epileptic seizures. Journal of clinical and experimental neuropsychology 2006;28(3):325-340.

Lutz, S.M., et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. BMC genetics 2015;16(1):138.

Madsen, B.E. and Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. PLoS genetics 2009;5(2):e1000384.

Majumdar, A., Witte, J.S. and Ghosh, S. Semiparametric allelic tests for mapping multiple phenotypes: binomial regression and Mahalanobis distance. Genetic epidemiology 2015;39(8):635-650.

McCarthy, M.I., et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature reviews genetics 2008;9(5):356.

McCarty, C.A., et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. BMC medical genomics 2011;4(1):13.

McCarty, C.A., et al. Healthy People 2010 disease prevalence in the Marshfield Clinic Personalized Medicine Research Project cohort: opportunities for public health genomic research. 2007.

Morgenthaler, S. and Thilly, W.G. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 2007;615(1):28-56.

O'Reilly, P.F., et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PloS one 2012;7(5):e34861.

Ober, C. and Vercelli, D. Gene–environment interactions in human disease: nuisance or opportunity? Trends in genetics 2011;27(3):107-115.

O'Brien, P.C. Procedures for comparing samples with multiple endpoints. Biometrics 1984:1079-1087.

Ott, J. and Rabinowitz, D. A principal-components approach based on heritability for combining phenotype information. Human heredity 1999;49(2):106-111.

Pan, W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genetic epidemiology 2009;33(6):497-507.

Paul, D.S., Soranzo, N. and Beck, S. Functional interpretation of non-coding sequence variation: concepts and challenges. Bioessays 2014;36(2):191-199.

Pendergrass, S. and Ritchie, M.D. Phenome-wide association studies: leveraging comprehensive phenotypic and genotypic data for discovery. Current genetic medicine reports 2015;3(2):92-100.

Pendergrass, S.A., et al. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS genetics 2013;9(1):e1003087.

Pendergrass, S.A., et al. Visually integrating and exploring high throughput phenome-wide association study (PheWAS) results using PheWAS-View. BioData mining 2012;5(1):5.

Pendergrass, S.A., et al. Phenome-wide association studies: embracing complexity for discovery. Human heredity 2015;79(3-4):111-123.

Pillai, S.G., et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS genetics 2009;5(3):e1000421.

Price, A.L., et al. Pooled association tests for rare variants in exon-resequencing studies. The American Journal of Human Genetics 2010;86(6):832-838.

Price, A.L., et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 2006;38(8):904.

Pritchard, J.K. and Cox, N.J. The allelic architecture of human disease genes: common disease–common variant… or not? Human molecular genetics 2002;11(20):2417-2423.

Raghunathan, T.E., et al. A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey methodology 2001;27(1):85-96.

Ray, D., Pankow, J.S. and Basu, S. USAT: A Unified Score-Based Association Test for Multiple Phenotype-Genotype Analysis. Genetic epidemiology 2016;40(1):20-34.

Regan, E.A., et al. Genetic epidemiology of COPD (COPDGene) study design. COPD: Journal of Chronic Obstructive Pulmonary Disease 2011;7(1):32-43.

Risch, N. and Merikangas, K. The future of genetic studies of complex human diseases. Science 1996;273(5281):1516-1517.

Ritchie, M.D., et al. Genome-and Phenome-Wide Analyses of Cardiac Conduction Identifies Markers of Arrhythmia RiskClinical Perspective. Circulation 2013;127(13):1377-1385.

Schafer, J.L. Analysis of incomplete multivariate data. CRC press; 1997.

Schaid, D.J., et al. Multiple genetic variant association testing by collapsing and kernel methods with pedigree or population structured data. Genetic epidemiology 2013;37(5):409-418.

Sha, Q., et al. Detecting association of rare and common variants by testing an optimally weighted combination of variants. Genetic epidemiology 2012;36(6):561-571.

Sha, Q., Zhang, Z. and Zhang, S. An improved score test for genetic association studies. Genetic epidemiology 2011;35(5):350-359.

Shameer, K., et al. A genome-and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Human genetics 2014;133(1):95-109.

Shi, G. and Rao, D. Optimum designs for next-generation sequencing to discover rare variants for common complex disease. Genetic epidemiology 2011;35(6):572-579.

Sivakumaran, S., et al. Abundant pleiotropy in human complex diseases and traits. The American Journal of Human Genetics 2011;89(5):607-618.

Solovieff, N., et al. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics 2013;14(7):nrg3461.

Stephens, M. A unified framework for association analysis with multiple related phenotypes. PloS one 2013;8(7):e65245.

Sun, J., et al. A method for analyzing multiple continuous phenotypes in rare variant association studies allowing for flexible correlations in variant effects. European journal of human genetics 2016;24(9):1344.

Svishcheva, G.R., Belonogova, N.M. and Axenovich, T.I. FFBSKAT: fast family-based sequence kernel association test. PloS one 2014;9(6):e99407.

Tang, C.S. and Ferreira, M.A. A gene-based test of association using canonical correlation analysis. Bioinformatics 2012;28(6):845-850.

Teer, J.K. and Mullikin, J.C. Exome sequencing: the sweet spot before whole genomes. Human molecular genetics 2010;19(R2):R145-R151.

Van Buuren, S., et al. Fully conditional specification in multivariate imputation. Journal of statistical computation and simulation 2006;76(12):1049-1064.

Van der Sluis, S., Posthuma, D. and Dolan, C.V. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS genetics 2013;9(1):e1003235.

Venter, J.C., et al. The sequence of the human genome. science 2001;291(5507):1304-1351.

Verma, A., et al. eMERGE Phenome-Wide Association Study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. BMC medical genomics 2016;9(1):32.

Walsh, T. and King, M.-C. Ten genes for inherited breast cancer. Cancer cell 2007;11(2):103-105.

Wang, K. and Abbott, D. A principal components regression approach to multilocus genetic association studies. Genetic epidemiology 2008;32(2):108-118.

Wang, S., et al. Detecting association of rare and common variants by testing an optimally weighted combination of variants with longitudinal data. In, BMC proceedings. BioMed Central; 2014. p. S91.

Wang, Z., Sha, Q. and Zhang, S. Joint analysis of multiple traits using" optimal" maximum heritability test. PloS one 2016;11(3):e0150975.

Wang, Z., et al. Joint analysis of multiple traits in rare variant association studies. Annals of human genetics 2016;80(3):162-171.

Wei, L. and Johnson, W.E. Combining dependent tests with incomplete repeated measurements. Biometrika 1985;72(2):359-364.

Wilk, J.B., et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS genetics 2009;5(3):e1000429.

Wilk, J.B., et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. American journal of respiratory and critical care medicine 2012;186(7):622-632.

Witte, J.S. Genome-wide association studies and beyond. Annual review of public health 2010;31:9-20.

Wu, M.C., et al. Rare-variant association testing for sequencing data with the sequence kernel association test. The American Journal of Human Genetics 2011;89(1):82-93.

Yan, T., et al. Genetic association with multiple traits in the presence of population stratification. Genetic epidemiology 2013;37(6):571-580.

Yang, J.J., et al. An efficient genome-wide association test for multivariate phenotypes based on the Fisher combination function. BMC bioinformatics 2016;17(1):19.

Yang, Q. and Wang, Y. Methods for analyzing multivariate phenotypes in genetic association studies. Journal of probability and statistics 2012;2012.

Yang, Q., et al. Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. Genetic epidemiology 2010;34(5):444-454.

Young, R., et al. Chromosome 4q31 locus in COPD is also associated with lung cancer. European Respiratory Journal 2010;36(6):1375-1382.

Zeger, S.L. and Liang, K.-Y. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986:121-130.

Zhang, J., et al. Nicotinic acetylcholine receptor variants associated with susceptibility to chronic obstructive pulmonary disease: a meta-analysis. Respiratory research 2011;12(1):158.

Zhang, S., Zhu, X. and Zhao, H. On a semiparametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. Genetic epidemiology 2003;24(1):44-56.

Zhang, Y., et al. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. NeuroImage 2014;96:309-325.

Zhang, Z., et al. Mixed linear model approach adapted for genome-wide association studies. Nature genetics 2010;42(4):355.

Zhou, X. and Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature methods 2014;11(4):407.

Zhu, A.Z., et al. Association of CHRNA5-A3-B4 SNP rs2036527 With Smoking Cessation Therapy Response in African-American Smokers. Clinical Pharmacology & Therapeutics 2014;96(2):256-265.

Zhu, H., Zhang, S. and Sha, Q. Power comparisons of methods for joint association analysis of multiple phenotypes. Human heredity 2015;80(3):144-152.

Zhu, W. and Zhang, H. Why do we test multiple traits in genetic association studies? Journal of the Korean Statistical Society 2009;38(1):1-10.

Zhu, X., et al. Detecting rare variants for complex traits using family and unrelated data. Genetic epidemiology 2010;34(2):171-187.

Zhu, X., et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. The American Journal of Human Genetics 2015;96(1):21-36.

Zhu, X., et al. Association mapping, using a mixture model for complex traits. Genetic epidemiology 2002;23(2):181-196.

Zhu, Y. and Xiong, M. Family-based association studies for next-generation sequencing. The American Journal of Human Genetics 2012;90(6):1028-1045.

Appendix A: Supplementary Tables

Table A.1.1. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 10$, MAF is 0.3, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.052	0.013	0.046	0.011
CV	0.056	0.006	0.042	0.010
OW	0.047	0.006	0.046	0.009
TATES	0.060	0.008	0.040	0.007
MANOVA	0.047	0.006	0.046	0.009
CCA	0.045	0.009	0.046	0.011
PCH	0.048	0.006	0.048	0.010
MultiPhen	0.049	0.007	0.049	0.012

Table A.1.2. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 40$, MAF is 0.3, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.038	0.007	0.060	0.011
CV	0.051	0.010	0.058	0.014
OW	0.045	0.011	0.056	0.015
TATES	0.045	0.012	0.051	0.011
MANOVA	0.046	0.011	0.063	0.016
CCA	0.046	0.012	0.063	0.016
PCH	0.045	0.011	0.057	0.017
MultiPhen	0.055	0.015	0.076	0.015

Table A.1.3. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 10$, MAF is 0.1, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.051	0.005	0.056	0.014
CV	0.043	0.009	0.040	0.009
OW	0.044	0.006	0.041	0.009
TATES	0.048	0.010	0.047	0.007
MANOVA	0.045	0.007	0.041	0.010
CCA	0.042	0.008	0.039	0.010
PCH	0.045	0.006	0.039	0.011
MultiPhen	0.046	0.008	0.044	0.009

Table A.1.4. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 20$, MAF is 0.1, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.043	0.010	0.046	0.008
CV	0.058	0.012	0.050	0.010
OW	0.059	0.013	0.049	0.008
TATES	0.047	0.008	0.052	0.018
MANOVA	0.059	0.014	0.051	0.008
CCA	0.059	0.014	0.052	0.009
PCH	0.059	0.014	0.048	0.007
MultiPhen	0.065	0.013	0.054	0.011

Table A.1.5. Estimated type-I error rates for the eight methods under two variance models. The total number of phenotypes is $K = 40$, MAF is 0.1, and the sample size is 1,000. α is the significance level.

	Type-I error rates			
	Variance model 1		Variance model 2	
Methods	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
OB	0.048	0.009	0.050	0.009
CV	0.044	0.007	0.053	0.013
OW	0.037	0.004	0.055	0.012
TATES	0.042	0.010	0.058	0.014
MANOVA	0.042	0.004	0.060	0.017
CCA	0.046	0.005	0.064	0.015
PCH	0.039	0.006	0.059	0.017
MultiPhen	0.053	0.008	0.072	0.016

Table A.2.1. Significant SNPs and the corresponding p-values in the analysis of COPDGene using the principal components (PCs) of phenotypes. The p-values of WCmulP are evaluated using 10^9 permutations, the p-values of SHet are evaluated using 10^8 permutations. The grayed-out p-values indicate the p-values $> 5 \times 10^{-8}$.

Chr	Variant identifier	WCmulP	SHet	Score	MultiPhen	CCA	TATES	OB
4	rs1512282	0	0	1.33E-09	9.28E-10	1.19E-09	2.85E-06	1.01E-04
4	rs1032297	0	0	1.18E-13	2.19E-13	1.40E-13	2.58E-09	6.91E-07
4	rs1489759	0	0	2.22E-16	3.14E-16	2.22E-16	3.19E-12	2.32E-08
4	rs1980057	0	0	1.11E-16	2.16E-16	1.11E-16	9.21E-13	1.72E-08
4	rs7655625	0	0	2.22E-16	2.68E-16	1.11E-16	1.80E-12	1.63E-08
15	rs16969968	0	0	7.84E-12	3.93E-12	5.42E-12	6.56E-07	1.99E-03
15	rs1051730	0	0	8.23E-12	4.02E-12	5.63E-12	5.43E-07	1.35E-03
15	rs12914385	0	0	5.60E-13	5.10E-13	5.53E-13	4.94E-07	4.64E-05
15	rs8040868	0	0	8.47E-13	1.10E-12	1.05E-12	3.42E-07	2.58E-04
15	rs951266	0	0	9.82E-12	4.85E-12	7.16E-12	6.65E-07	2.96E-03
15	rs8034191	0	0	1.04E-10	3.19E-11	7.56E-11	4.19E-06	6.25E-03
15	rs2036527	0	0	2.13E-10	7.79E-11	1.52E-10	9.02E-06	5.36E-03
15	rs931794	0	0	1.04E-10	3.29E-11	7.71E-11	2.47E-05	1.66E-02
15	rs2568494	4.10E-08	6.0E-08	6.97E-08	2.58E-08	6.01E-08	3.01E-04	3.69E-02
15	rs17483721	5.00E-08	2.2E-07	1.25E-07	6.10E-08	1.17E-07	2.86E-04	2.64E-02
15	rs17483929	4.00E-08	9.0E-08	8.89E-08	3.89E-08	8.08E-08	2.94E-04	3.28E-02

Table A.2.2. The structures of γ and $\mathrm{cov}(y)$ for different numbers of factors when number of phenotypes is 8.

	$R = 1$	$R = 2$	$R = 4$
γ	$\begin{pmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{pmatrix}$	$\begin{pmatrix}1&0\\1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\end{pmatrix}$
$\mathrm{cov}(y)$	$\begin{pmatrix}1&c^2&\cdots&c^2\\c^2&1&\cdots&c^2\\\vdots&\vdots&\ddots&\vdots\\c^2&c^2&\cdots&1\end{pmatrix}_{8\times8}$	$\begin{pmatrix}1&c^2&c^2&c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\c^2&1&c^2&c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\c^2&c^2&1&c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\c^2&c^2&c^2&1&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&1&c^2&c^2&c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&c^2&1&c^2&c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&c^2&c^2&1&c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&c^2&c^2&c^2&1\end{pmatrix}$	$\begin{pmatrix}1&c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\c^2&1&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&1&c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&c^2&1&\rho c^2&\rho c^2&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&1&c^2&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&c^2&1&\rho c^2&\rho c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&1&c^2\\\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&\rho c^2&c^2&1\end{pmatrix}$

Appendix B: Supplementary Figures

Figure B.1.1. Power comparisons of the eight methods as a function of heritability for the six mean models under variance model 1. The total number of phenotypes is $K = 10$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

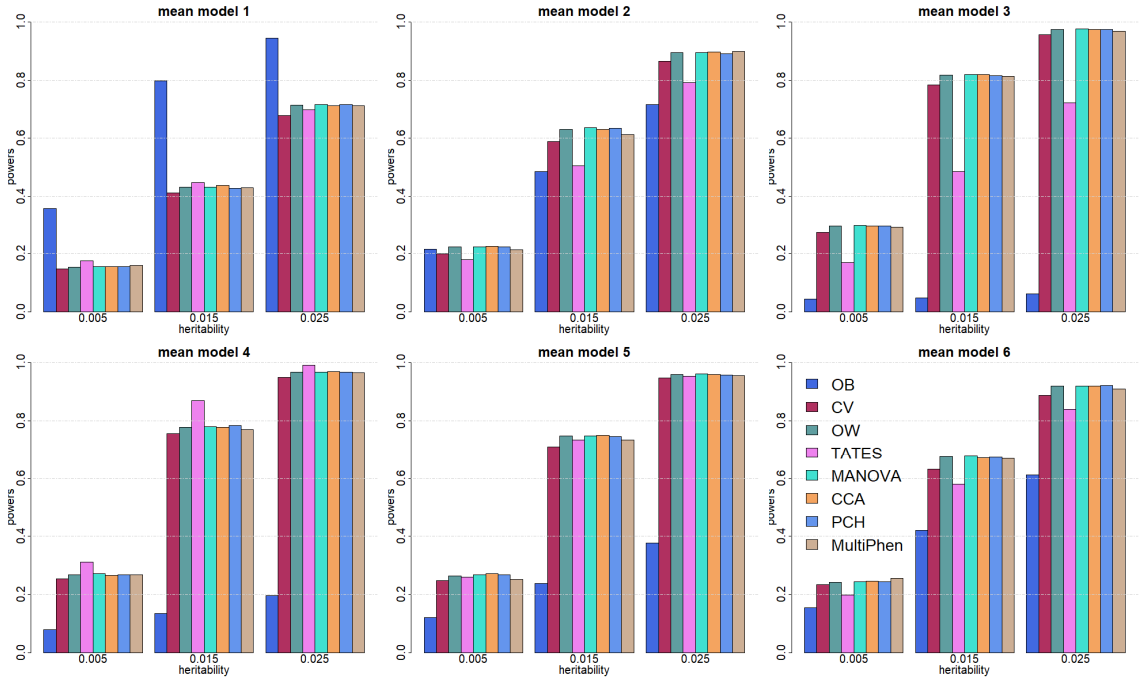


Figure B.1.2. Power comparisons of the eight methods as a function of heritability for the six mean models under variance model 2. The total number of phenotypes is $K = 10$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

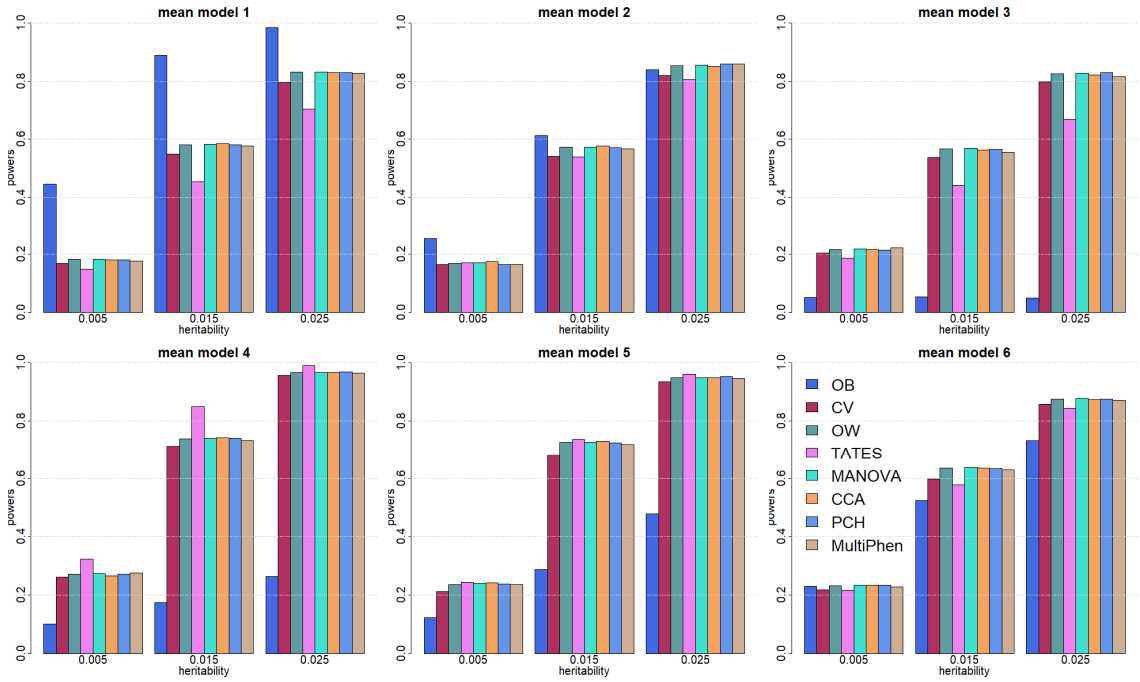


Figure B.1.3. Power comparisons of the seven methods as a function of heritability for the six mean models under variance model 1. The total number of phenotypes is $K = 40$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

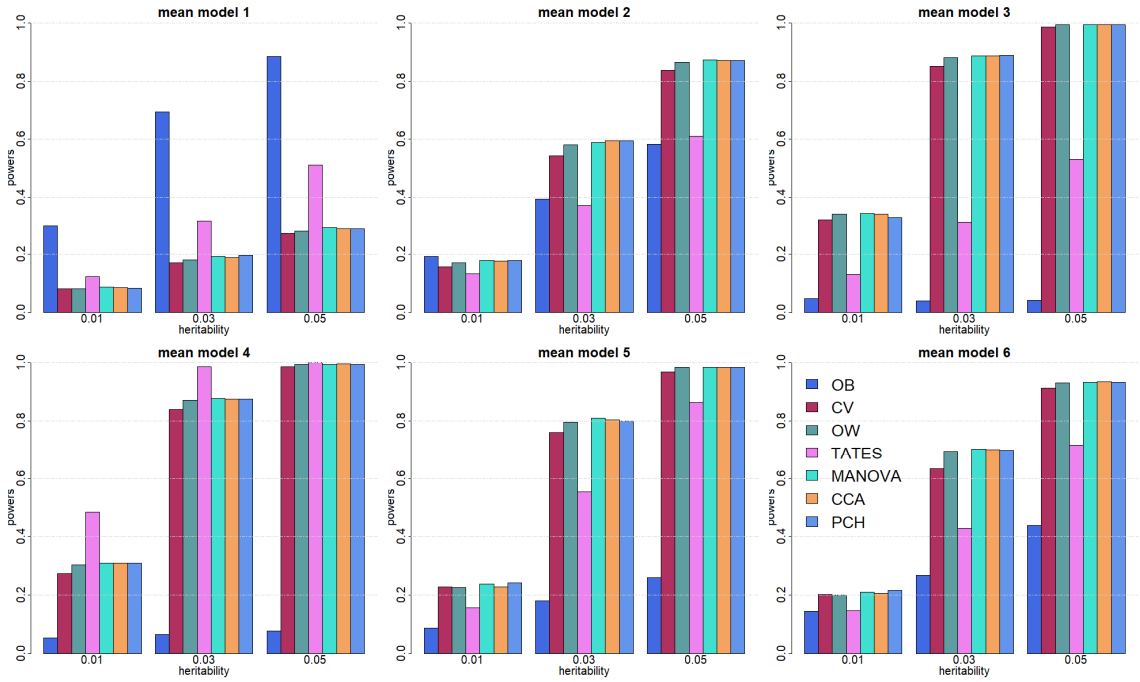


Figure B.1.4. Power comparisons of the seven methods as a function of heritability for the six mean models under variance model 2. The total number of phenotypes is $K = 40$, rho is 0.1, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

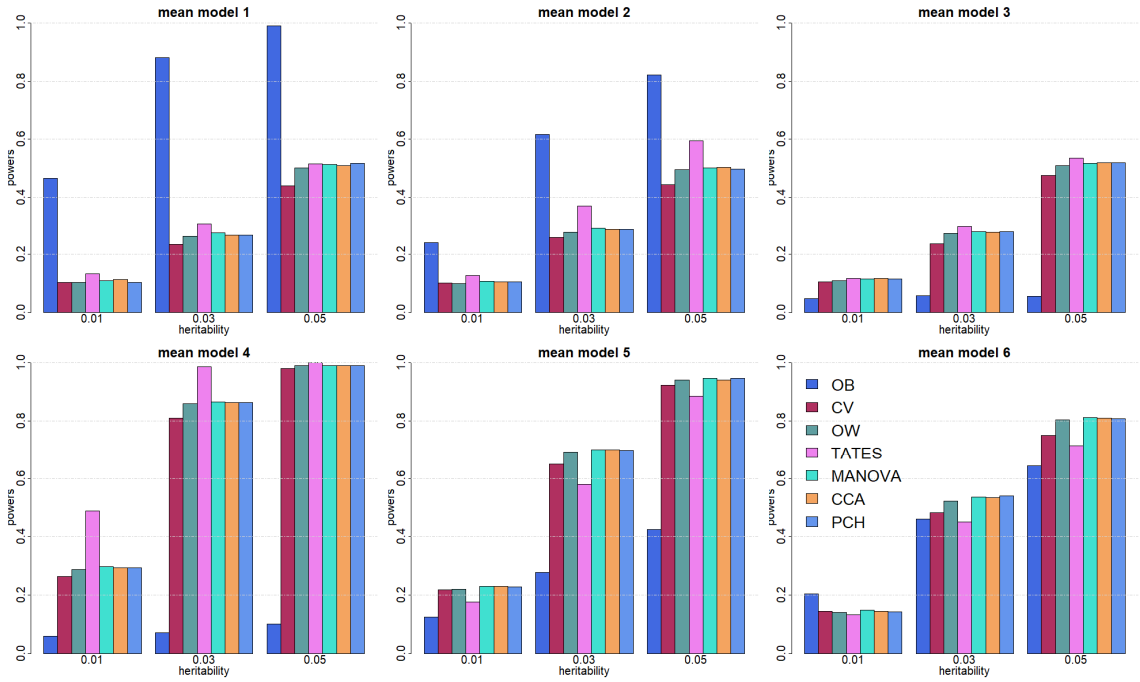


Figure B.1.5. Power comparisons of the eight methods as a function of rho for the six mean models under variance model 1. The total number of phenotypes is $K = 10$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

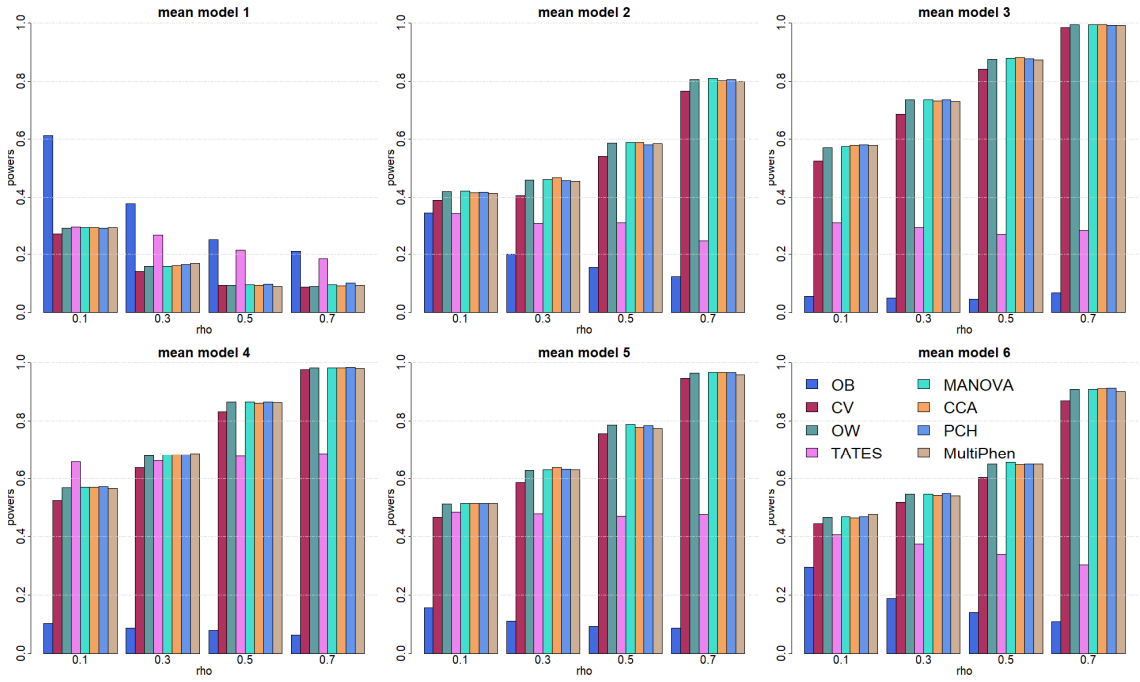


Figure B.1.6. Power comparisons of the eight methods as a function of rho for the six mean models under variance model 2. The total number of phenotypes is $K = 10$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

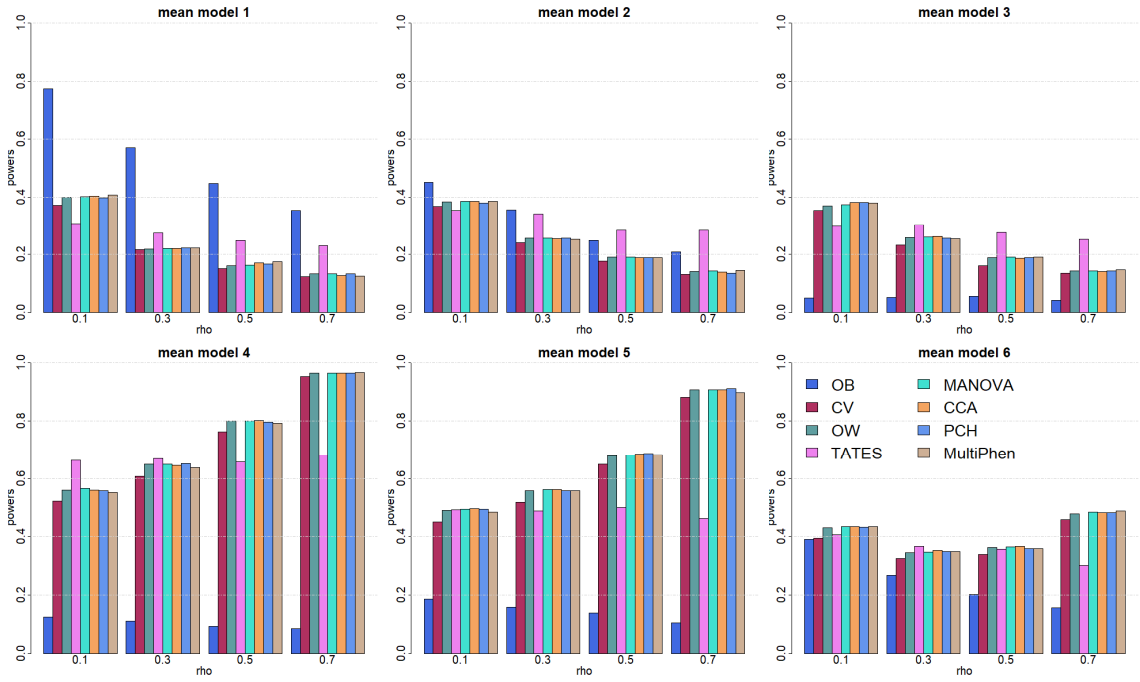


Figure B.1.7. Power comparisons of the seven methods as a function of rho for the six mean models under variance model 1. The total number of phenotypes is $K = 40$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

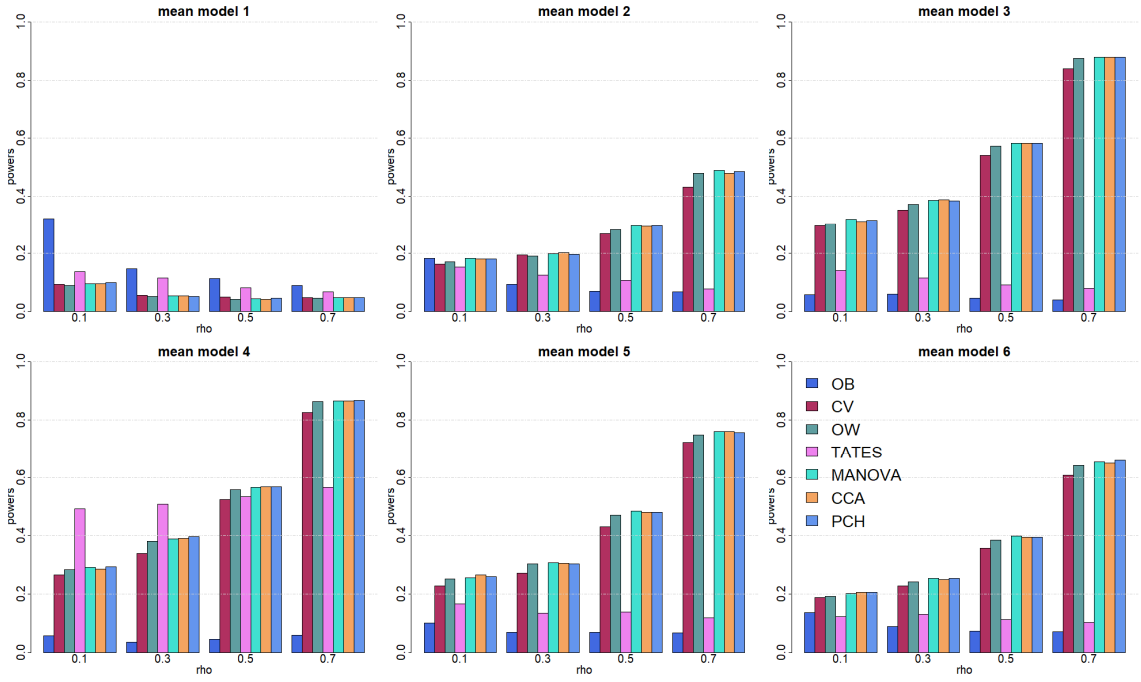


Figure B.1.8. Power comparisons of the seven methods as a function of rho for the six mean models under variance model 2. The total number of phenotypes is $K = 40$, heritability is 0.01, MAF is 0.3, and the sample size is 1,000. Significance is assessed at the 5% level.

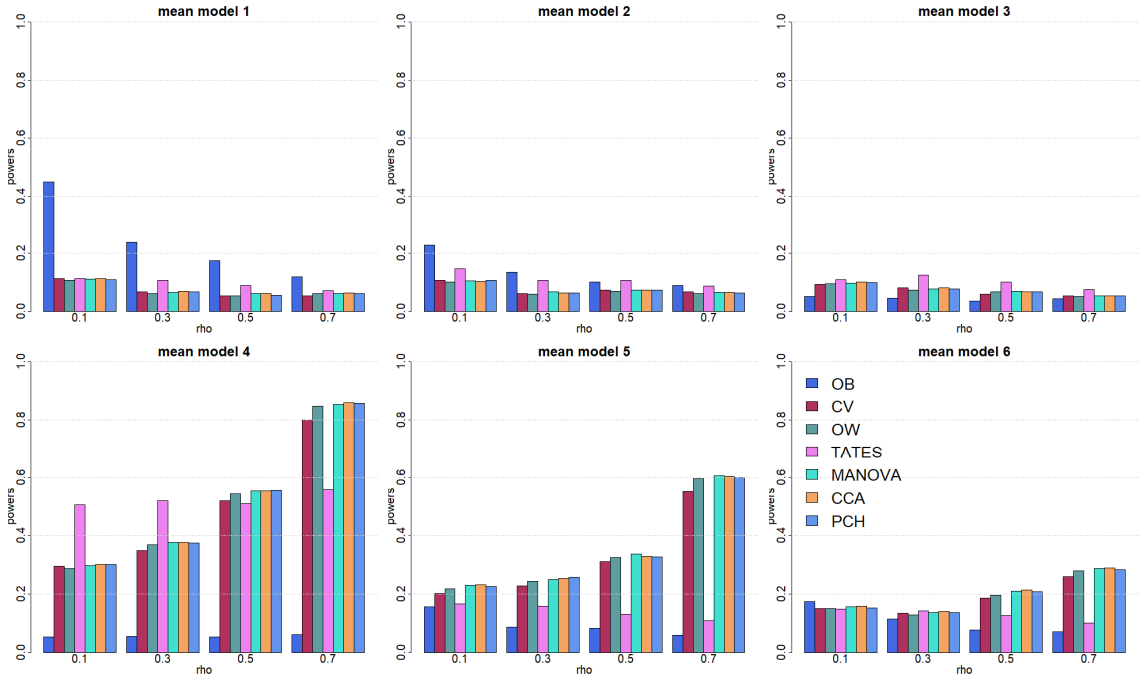


Figure B.1.9. Power comparisons of the eight methods as a function of MAF for the six mean models under variance model 1. The total number of phenotypes is $K = 20$, heritability is 0.02, rho is 0.1, and the sample size is 1,000. Significance is assessed at the 5% level.

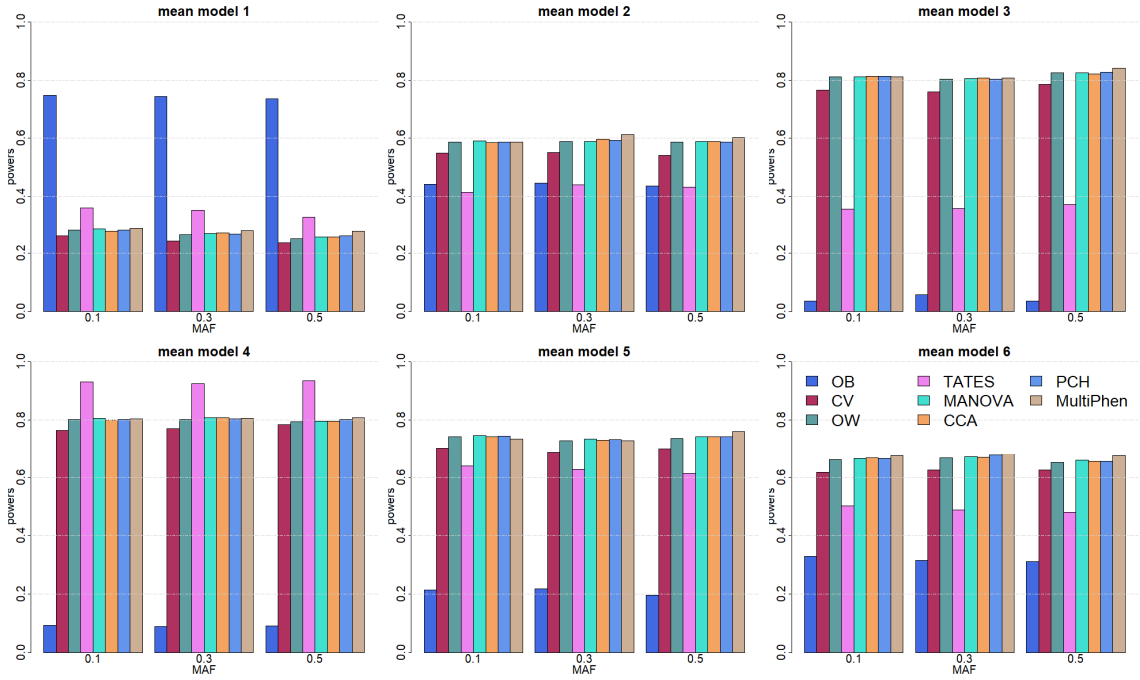


Figure B.1.10. Power comparisons of the eight methods as a function of MAF for the six mean models under variance model 2. The total number of phenotypes is $K = 20$, heritability is 0.02, rho is 0.1, and the sample size is 1,000. Significance is assessed at the 5% level.

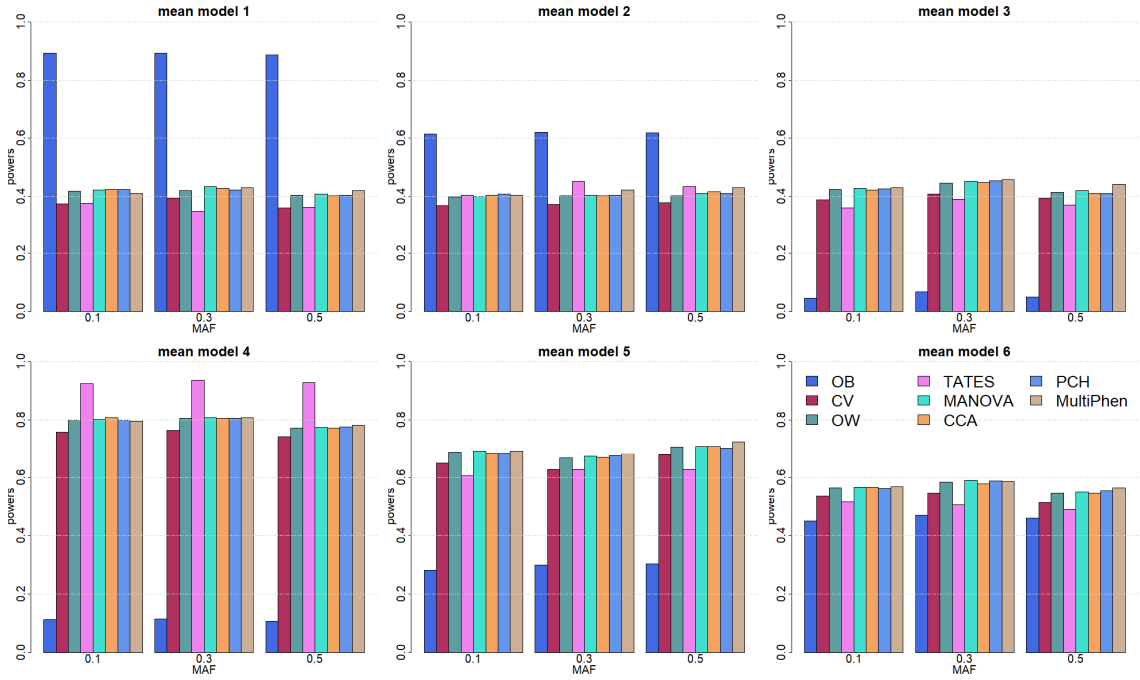


Figure B.1.11. Power comparisons of the six methods as a function of heritability for the six mean models under variance model 1. The total number of phenotypes is $K = 20$, rho is 0.1, MAF is 0.3, and sample size is 1000. Significance level is 5×10^{-8}.

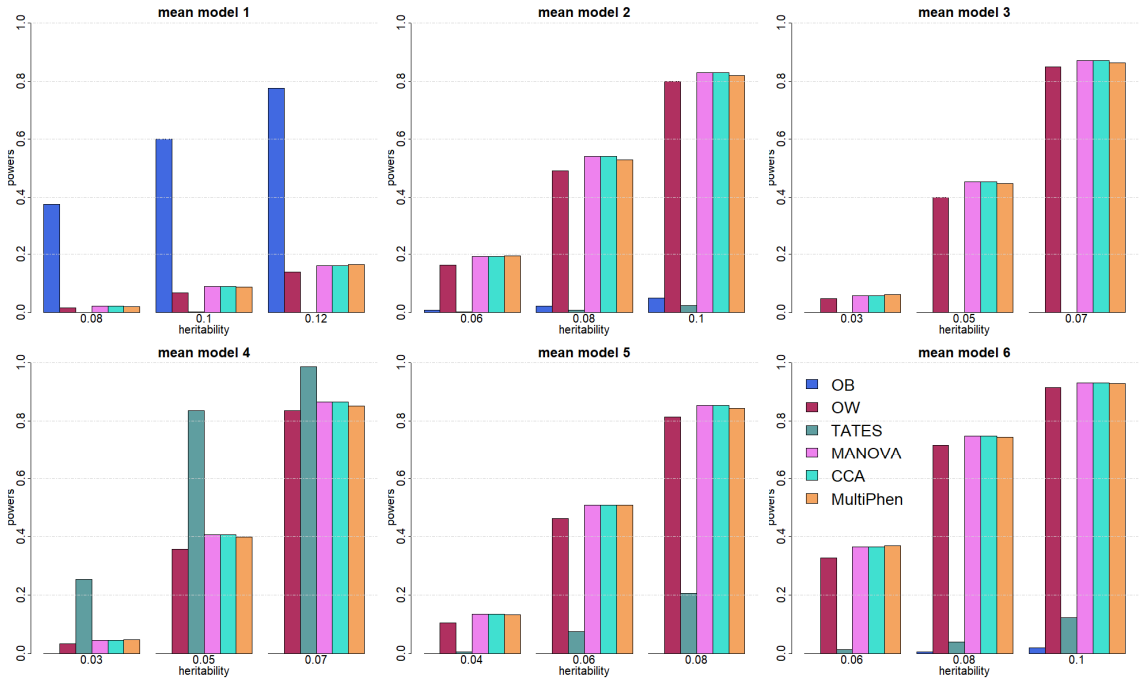


Figure B.1.12. Power comparisons of the six methods as a function of heritability for the six mean models under variance model 2. The total number of phenotypes is $K = 20$, rho is 0.1, MAF is 0.3, and sample size is 1000. Significance level is 5×10^{-8}.

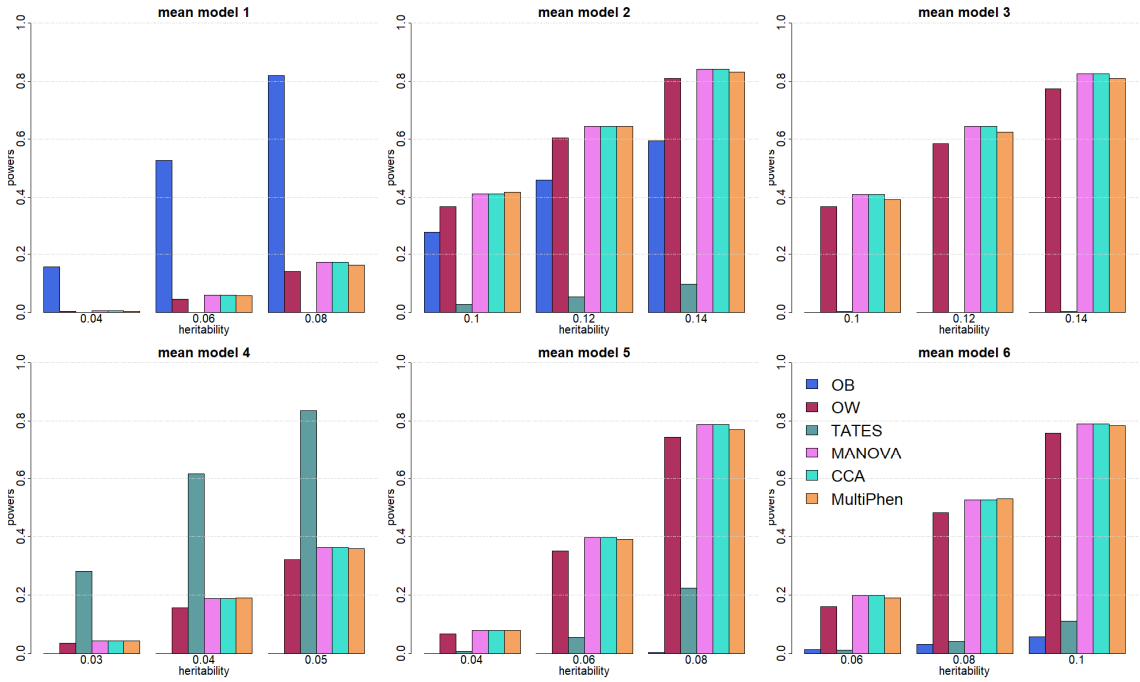


Figure B.1.13. Power comparisons of the six methods as a function of rho for the six mean models under variance model 1. The total number of phenotypes is $K = 20$, MAF is 0.3, and sample size is 1000. Significance level is 5×10^{-8}.

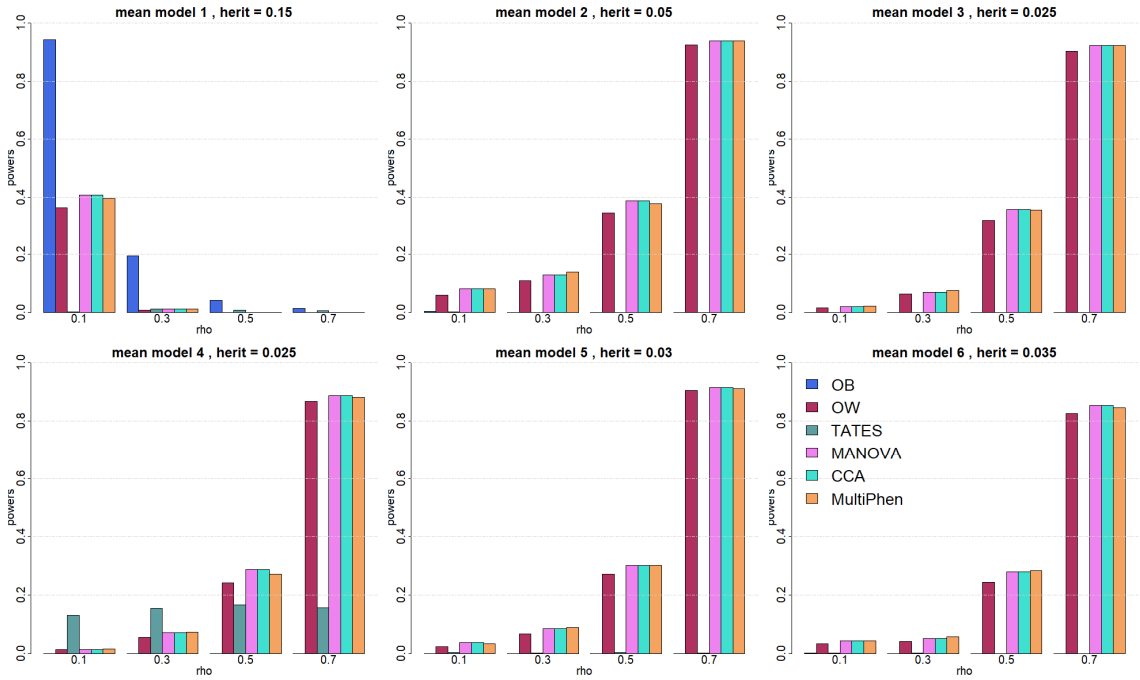


Figure B.1.14. Power comparisons of the six methods as a function of rho for the six mean models under variance model 2. The total number of phenotypes is $K = 20$, MAF is 0.3, and sample size is 1000. Significance level is 5×10^{-8}.

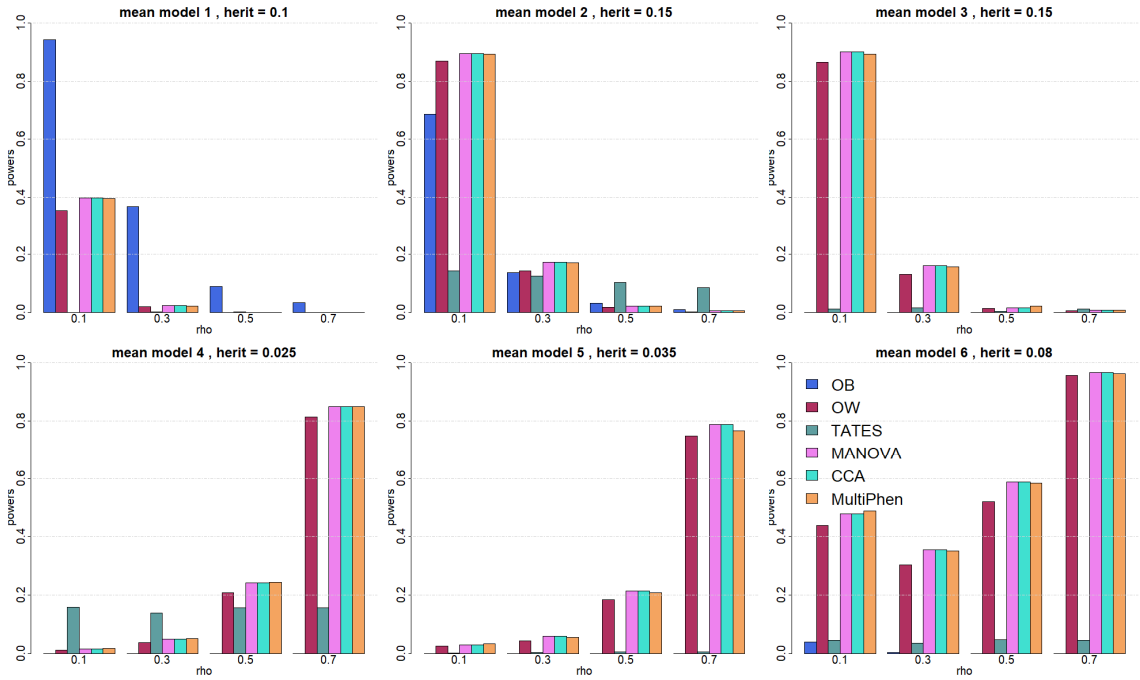


Figure B.2.1. Power comparisons of the seven methods as a function of β for the six models using the principal components (PCs) of phenotypes. The total number of phenotypes is $K = 8$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

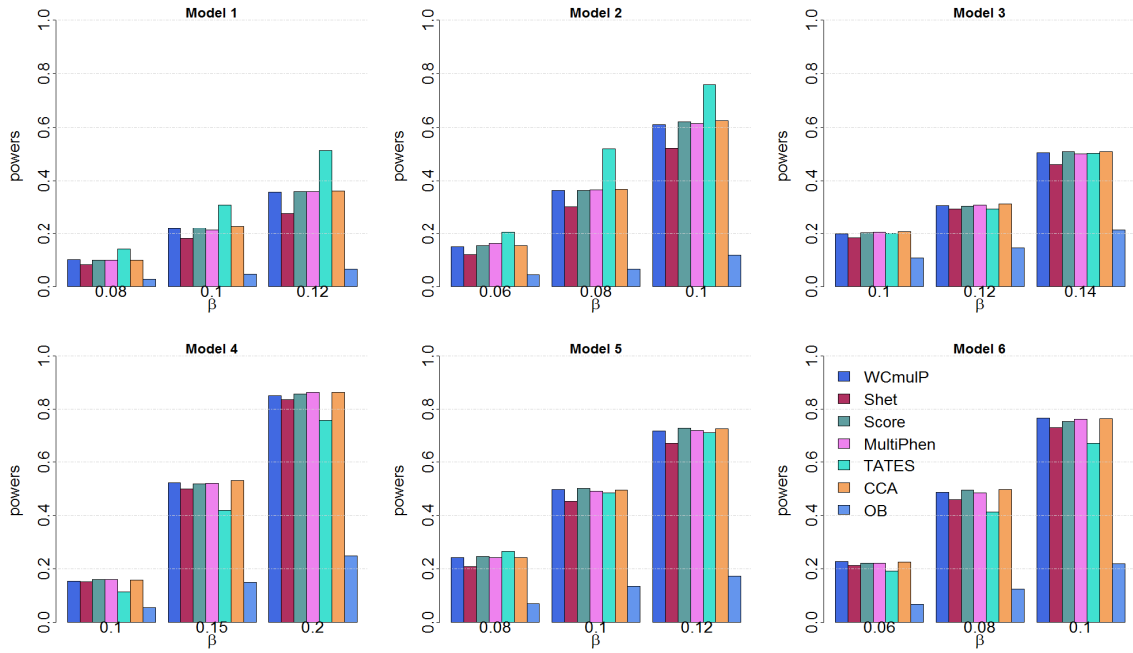


Figure B.2.2. Power comparisons of the seven methods as a function of β for the six models using the principal components (PCs) of phenotypes. The total number of phenotypes is $K = 16$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations. The power of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

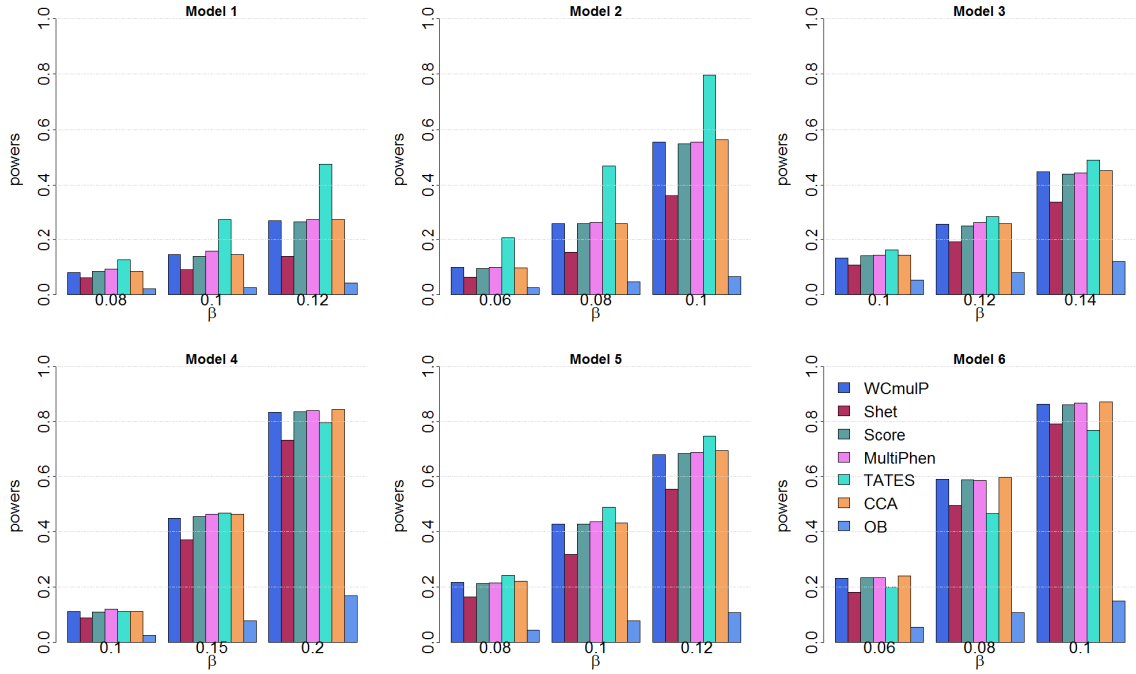


Figure B.2.3. Power comparisons of the seven methods as a function of c^2 for the six models using the principal components (PCs) of the phenotypes. The total number of phenotypes is $K = 8$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations, the p-values of other methods are evaluated using asymptotic distribution. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

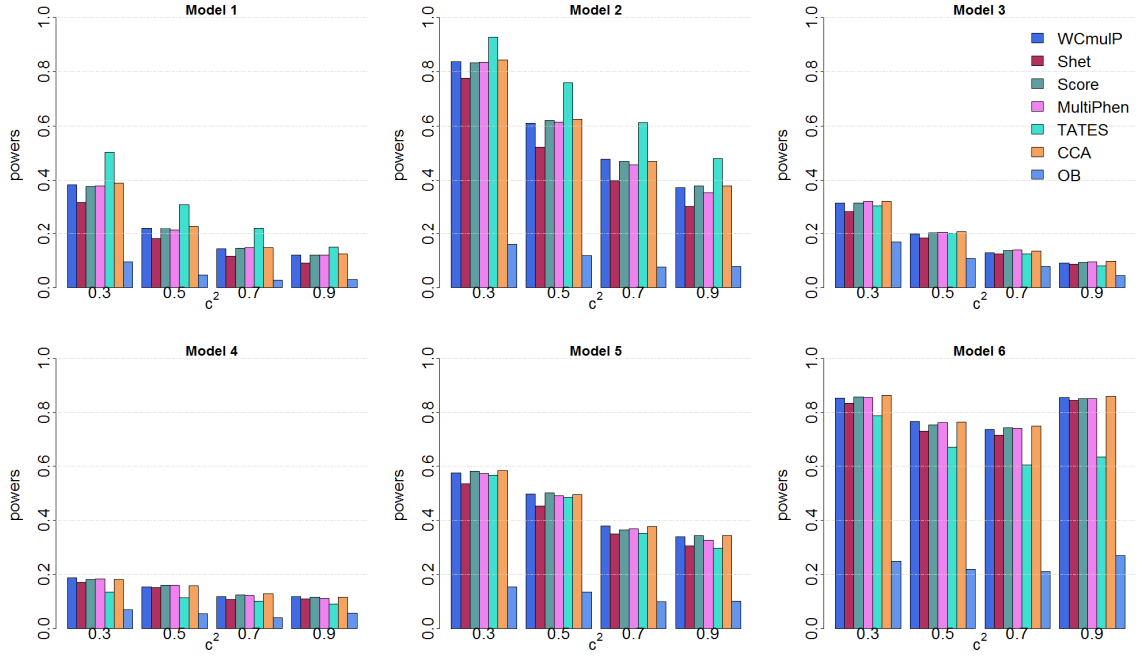


Figure B.2.4. Power comparisons of the seven methods as a function of c^2 for the six models using the principal components (PCs) of the phenotypes. The total number of phenotypes is $K = 16$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 1,000 permutations, the p-values of other methods are evaluated using asymptotic distribution. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 0.01.

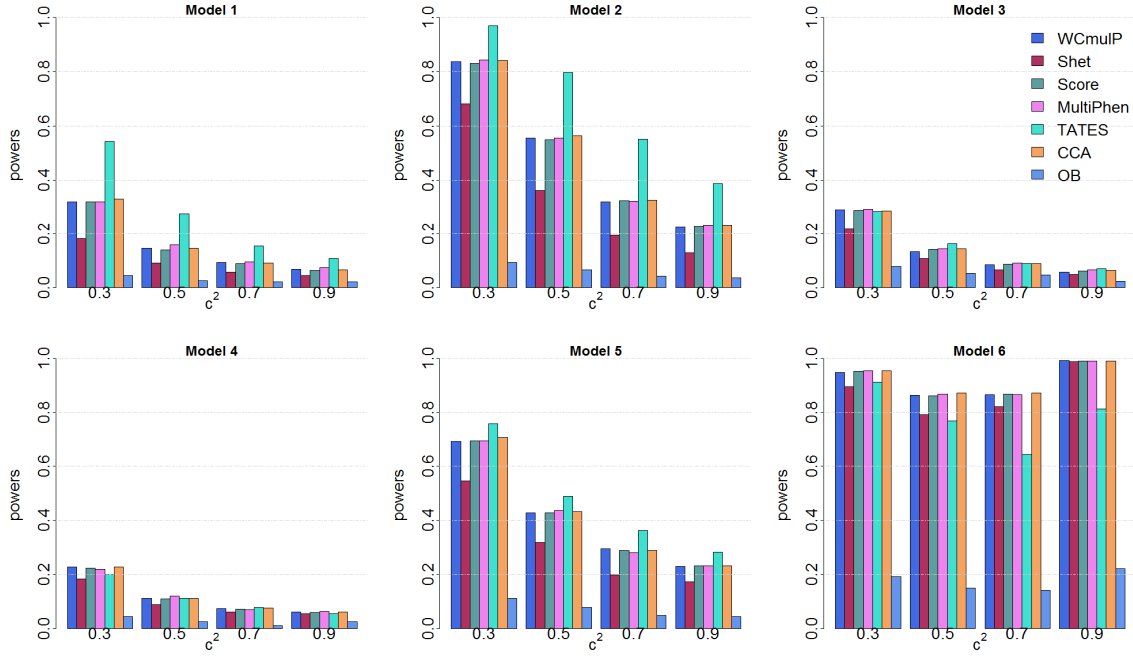


Figure B.2.5. Power comparisons of the seven methods as a function of β for the six models. The total number of phenotypes is $K = 8$, $c^2 = 0.5$, $\rho c^2 = 0.1$, and $MAF = 0.3$. The p-values of WCmulP and SHet are evaluated using 100,000 permutations. The power of all of the seven methods is evaluated using 1,000 replicated samples at a significance level of 5×10^{-5}.

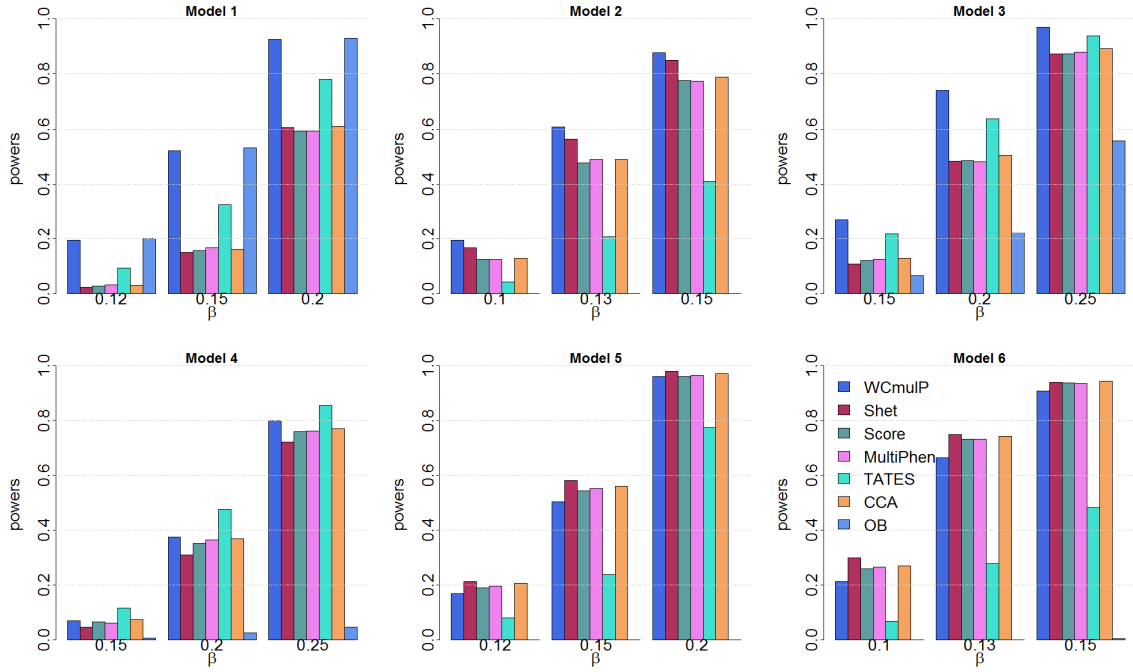


Figure B.2.6. The correlation matrix plot of the 7 COPD-related phenotypes.