Michigan Technological University

Michigan Technological University
Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's Reports

2018

JOINT ANALYSIS OF MULTIPLE PHENOTYPES IN ASSOCIATION STUDIES

Xiaoyu Liang
Michigan Technological University, xiaoyuli@mtu.edu

Copyright 2018 Xiaoyu Liang

Recommended Citation

Liang, Xiaoyu, "JOINT ANALYSIS OF MULTIPLE PHENOTYPES IN ASSOCIATION STUDIES", Open Access Dissertation, Michigan Technological University, 2018.
https://doi.org/10.37099/mtu.dc.etdr/602

Follow this and additional works at: https://digitalcommons.mtu.edu/etdr

 Part of the Applied Statistics Commons, Biostatistics Commons, and the Statistical Methodology Commons

JOINT ANALYSIS OF MULTIPLE PHENOTYPES IN ASSOCIATION STUDIES

By

Xiaoyu Liang

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Mathematical Sciences

MICHIGAN TECHNOLOGICAL UNIVERSITY

2018

© 2018 Xiaoyu Liang

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:	*Shuanglin Zhang*
Dissertation Co-Advisor:	*Qiuying Sha*
Committee Member:	*Yeonwoo Rho*
Committee Member:	*Hairong Wei*
Department Chair:	*Mark S. Gockenbach*

Table of Contents

Preface

The first chapter “An Adaptive Fisher’s Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies” was published in Scientific Reports in October 2016. The second chapter “Joint Analysis of Multiple Phenotypes in Association Studies Using Allele-based Clustering Approach for Non-Normal Distributions” was invited for revision and the revision is under the review in Annals of Human Genetics. The third chapter “A Hierarchical Clustering Method for Dimension Reduction in Joint Analysis of Multiple Phenotypes” was accepted by Genetic Epidemiology in February 2018. For all three chapters, Dr. Shuanglin Zhang and Dr. Qiuying Sha designed researches, Dr. Shuanglin Zhang and Xiaoyu Liang performed statistical analyses, Dr. Shuanglin Zhang, Dr. Qiuying Sha, and Xiaoyu Liang wrote the manuscripts. In the first chapter, Zhenchuan Wang performed real data analysis. In the third chapter, Dr. Yeonwoo Rho was involved in designing the research.

Acknowledgements

I have gained a professional experience and unforgettable memories during my time at Michigan Technological University and the experience opened my eyes to other possibilities. I would like to express my deep appreciation to Michigan Technological University and the Department of Mathematical Sciences for giving me the opportunity to pursue an advanced degree in Statistics. It is my pleasure to acknowledge the roles of several individuals who were instrumental in the completion of my Ph.D. research.

First and foremost, I would like to express the deepest appreciation to my advisor Dr. Shuanglin Zhang, who continually and convincingly conveyed a spirit of adventure in regards to research and scholarship. He has been there providing his heartfelt support and guidance at all times and has given me invaluable guidance, inspiration, and suggestions in my quest for knowledge. He has given me all the freedom to pursue my research, while silently and non-obtrusively ensuring that I stay on course and do not deviate from the core of my research. I would like to express my deep gratitude to my co-advisor Dr. Qiuying Sha for her patient guidance, enthusiastic encouragement and persistent help of the research work. I have been extremely lucky to have supervisors who cared so much about my work, and who responded to my questions and queries so promptly. Without their assistance and dedicated involvement in every step throughout the progress, this dissertation would have never been accomplished. I would like to thank you very much for your support and understanding over these past four years.

I would also like to show gratitude to my committee member Dr. Hairong Wei. My grateful thanks are also extended to Dr. Yeonwoo Rho for her precious points in my third paper. It is my honor to have these distinguished professors on my committee.

Many influential colleagues and teachers deserve special thanks for their collaborations and instructions. I cannot name them all but must thank Mrs. Ann M. Humes and Mr. Jason Gregersen, Dr. Mark S. Gockenbach, and Dr. Kui Zhang. I would

like to thank Mrs. Ann M. Humes for her encouragement and support in my teaching experience. I would like to especially thank Mr. Jason Gregersen, who have been so helpful and cooperative in giving his support at all times to help me achieve my goal. I would like to thank our department chair, Dr. Mark S. Gockenbach, for his steadfast understanding and encouragement as I pursued this degree. I would also like to show gratitude to Dr. Kui Zhang, who was my fourth-year statistical consulting professor. His teaching style and enthusiasm for the topic made a strong impression on me.

Getting through my dissertation required more than academic support, and I have many people to thank for listening to and, at times, having to tolerate me over the past four years. I cannot begin to express my gratitude and appreciation for their friendship. Dr. Xuexia Wang, Li Guo, and Lilia Feng have been unwavering in their personal and professional support during the time I spent at the university. For many memorable evenings out and in, I must thank everyone above as well as Zhongyuan Hu, Yanfang Liu, Ruihao Huang, Joe Reath, and everyone in our research group. I would also like to thank Mrs. Abby Storm who opened both her home and heart to me when I first arrived in the city.

My acknowledgment would be incomplete without thanking the biggest source of my strength, my family. My grandmother, who offered her encouragement through phone calls every week. Special thanks are due to my parents Mrs. Yuping Fu and Mr. Gang Liang for their unwavering and unselfish love and support given to me at all times. Without their unconditional support and continuous expression of pride, I would never have made my achievements during the graduate study.

Abstract

Genome-wide association studies (GWAS) have become a very effective research tool to identify genetic variants of underlying various complex diseases. In spite of the success of GWAS in identifying thousands of reproducible associations between genetic variants and complex disease, in general, the association between genetic variants and a single phenotype is usually weak. It is increasingly recognized that joint analysis of multiple phenotypes can be potentially more powerful than the univariate analysis, and can shed new light on underlying biological mechanisms of complex diseases. Therefore, developing statistical methods to test for genetic association with multiple phenotypes has become increasingly important. This dissertation contains three chapters and the three chapters include three new methods we developed for jointly analyzing multiple phenotypes.

In the first chapter of this dissertation, we propose an Adaptive Fisher's Combination (AFC) method for joint analysis of multiple phenotypes in association studies. The AFC method combines p-values obtained in standard univariate GWAS by using the optimal number of p-values which is determined by the data. In the second chapter, we propose an Allele-Based Clustering (ABC) approach for the joint analysis of multiple non-normal phenotypes in association studies. In the ABC method, we consider the alleles at a SNP of interest as a dependent variable with two classes, and the correlated phenotypes as predictors to predict the alleles at the SNP of interest. In the third chapter, we develop a novel variable reduction method using hierarchical clustering method (HCM) for joint analysis of multiple phenotypes in association studies. HCM involves two steps. The first step applies a dimension reduction technique by using a representative phenotype for each cluster of phenotypes. Then, existing methods are used in the second step to test the association between genetic variants and the representative phenotypes rather than the individual phenotypes. We perform extensive simulations to evaluate performances of AFC, ABC, and HCM methods and compare the powers of our

methods with the powers of some existing methods. Our simulation studies show that the proposed methods have correct type I error rates and are either the most powerful test or comparable with the most powerful test. Finally, we illustrate our proposed methodologies AFC and HCM by analyzing whole-genome genotyping data from a lung function study. The results of real data analysis demonstrated that the proposed methods have great potential in GWAS on complex diseases with multiple phenotypes.

1 An Adaptive Fisher's Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies

Currently, the analyses of most genome-wide association studies (GWAS) have been performed on a single phenotype. There is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Therefore, using only one single phenotype may lose statistical power to identify the underlying genetic mechanism. There is an increasing need to develop and apply powerful statistical tests to detect association between multiple phenotypes and a genetic variant. In this study, we develop an Adaptive Fisher's Combination (AFC) method for joint analysis of multiple phenotypes in association studies. The AFC method combines p-values obtained in standard univariate GWAS by using the optimal number of p-values which is determined by the data. We perform extensive simulations to evaluate the performance of the AFC method and compare the power of our method with the powers of TATES, Tippett's method, Fisher's combination test, MANOVA, MultiPhen, and SUMSCORE. Our simulation studies show that the proposed method has correct type I error rates and is either the most powerful test or comparable with the most powerful test. Finally, we illustrate our proposed methodology by analyzing whole-genome genotyping data from a lung function study.

1.1 Background

To date, genome-wide association studies (GWAS) have become a tool of choice for the identification of genetic variants associated with complex human diseases. Currently, the analyses of most GWAS have been performed on a single phenotype. There is increasing evidence showing that pleiotropy, the effect of one variant on multiple phenotypes, is a widespread phenomenon in complex diseases (Sivakumaran et al., 2011; Wang et al., 2016b). Therefore, using only one single phenotype may lose statistical power to identify the underlying genetic mechanism. By taking into account the correlated structure of multiple phenotypes, we can not only discover genetic variants influencing multiple phenotypes that may lead to better understanding of etiology of complex human diseases

(He et al., 2013; Wang, 2014), but also can improve the statistical power by aggregating multiple weak effects and provide new biological insights by revealing pleiotropic variants (Amos & Laing, 1993; Jiang & Zeng, 1995; Schifano et al., 2013). Consequently, there is an increasing need to develop powerful statistical methods to detect association between multiple phenotypes and genetic variants.

Recently, several statistical methods for detecting association using multivariate phenotypes have been developed (Klei et al., 2008; Li et al., 2011; O'Reilly et al., 2012; Tang & Ferreira, 2012; van der Sluis et al., 2013; Yan et al., 2013). These methods can be divided into three groups: regression models, variable reduction method, and combining test statistics from univariate analysis (Yang & Wang, 2012). Regression models, such as linear mixed effects models, generalized mixed effects models, and generalized estimating equations, can be used to test the association between a genetic variant and multiple phenotypes. By using random effects to account for correlation among individuals, linear and generalized mixed effect models can model the covariance structure not only caused by correlated phenotypes, but also caused by population structure (Bates & DebRoy, 2004; Breslow & Clayton, 1993; Fitzmaurice & Laird, 1993; Laird & Ware, 1982; Yan et al., 2013). Generalized estimating equations collapse random effects and random residual errors in the model (Liang & Zeger, 1986). Existing variable reduction methods can be roughly divided into three categories, principal components analysis of phenotypes (PCP) (Aschard et al., 2014), canonical correlation analysis (CCA) (Tang & Ferreira 2012) and principal component of heritability (PCH) (Klei et al., 2008; Ott & Rabinowitz, 1999). The PCP approach applies a dimension reduction technique and tests for associations between genetic variants and the principle components of the phenotypes rather than the individual phenotypes. CCA provides a convenient statistical framework to simultaneously test the association between any number of quantitative phenotypes and any number of genetic variants genotyped across a gene or region of interest for unrelated individuals. For each genetic variant, the PCH approach reduces the phenotypes to a linear combination of phenotypes that has the highest heritability among all linear combinations of the phenotypes. Based on PCH,

several advanced methods have been proposed such as penalized PCH applicable to high-dimensional data (Wang et al., 2007a; Wang et al., 2007b) and principle components of heritability with coefficients maximizing the quantitative phenotype locus heritability (PCQH) (Ferreira & Purcell, 2009; Klei et al., 2008; Lange et al., 2002). The third group, combining test statistics from univariate tests, is to conduct univariate analysis on each phenotype, then combine the univariate test statistics (Yang et al., 2010). The Trait-based Association Test that uses Extended Simes procedure (TATES) (van der Sluis et al., 2013) belongs to this group. TATES combines p-values obtained in standard univariate GWAS while correcting for the correlation between p-values.

Motivated by TATES, in this article, we propose an Adaptive Fisher's Combination (AFC) method for joint analysis of multiple phenotypes in genetic association studies. We first test the association between each of the phenotypes and a genetic variant of interest using standard GWAS software. Then, AFC uses the optimal number of p-values which is determined by the data to test the association. Using extensive simulation studies, we evaluate the performance of the proposed method and compare the power of the proposed method with the powers of TATES, Tippett's method (Pesarin & Salmaso, 2010), Fisher's Combination test (FC) (Yang et al., 2016), Multivariate Analysis of Variance (MANOVA) (Cole et al., 1994), joint model of Multiple Phenotypes (MultiPhen) (O'Reilly et al., 2012), and Sum Score method (SUMSCORE) (van der Sluis et al., 2013). Our simulation studies show that the proposed method has correct type I error rates and is either the most powerful test or comparable with the most powerful tests. Finally, we illustrate our proposed methodology by analyzing whole-genome genotyping data from a lung function study.

1.2 Methods

1.2.1 Adaptive Fisher's Combination Method

Consider a sample of n unrelated individuals. Each individual has K phenotypes. Denote $Y_k = (y_{1k}, \dots, y_{nk})^T$ as the k^{th} phenotype of n individuals. Denote $X = (x_1, \dots, x_n)^T$ as the genotypic score of n individuals at a genetic variant of interest, where $x_i \in \{0, 1, 2\}$ is the number of minor alleles that the i^{th} individual carries at the genetic variant. We propose a new method to test the null hypothesis H_0: none of the K phenotypes are associated with the genetic variant.

We test the association between each phenotypic vector Y_k $(k = 1, 2, \dots, K)$ and the genotypic score X using a standard GWAS software (e.g. PLINK, Gen/ProbABEL, MaCH, SNPTEST, and FaST-LMM) (Aulchenko et al., 2007; Aulchenko et al., 2010; Li et al., 2009; Li et al., 2010; Lippert et al., 2011; Marchini et al., 2007; Purcell et al., 2007). Let $p_{1,}p_{2,\dots,}p_K$ denote the p-values obtained by the standard univariate GWAS. Based on these p-values, we propose an Adaptive Fisher's Combination (AFC) method to test the association between multiple phenotypes and the genetic variant. Let $p_{(k)}$ denote the k^{th} smallest p-value, $T_k = -\sum_{i=1}^{k}\left(\log p_{(i)}\right)$ $(k = 1, 2, \dots, K)$, and p_{T_k} denote the p-value of T_k. The statistic of AFC to test the association between the K phenotypes and the genetic variant is given by $T_{all} = \min_{1 \le k \le K} p_{T_k}$. We use the following permutation procedure to evaluate the p-values of T_k and T_{all}.

1. In each permutation, we randomly shuffle the genotypes and recalculate $p_{(1)}, \dots, p_{(K)}$ and $T_1, \dots, T_K$. Suppose that we perform B times of permutations. Let $T_k^{(b)}$ $(b = 0,1, \dots, B)$ denote the value of T_k based on the b^{th} permuted data, where $b = 0$ represents the original data.
2. We transfer $T_k^{(b)}$ to $p_{T_k}^{(b)}$ by

$$p_{T_k}^{(b)} = \frac{\#\{\mathrm{d}: T_k^{(d)} > T_k^{(b)} \; for \; d=0,1,\dots,B\}}{B} \qquad (1)$$

3. Let $T_{all}^{(b)} = \min_{1 \le k \le K} p_{T_k}^{(b)}$. Then, the p-value of T_{all} is given by

$$\frac{\#\{b: T_{all}^{(b)} < T_{all}^{(0)} \; for \; b=1,2,\ldots,B\}}{B} \tag{2}$$

As shown in Appendix, the null distributions of $p_1, p_2, \ldots, p_K$ and thus of T_{all} do not depend on the genetic variant being tested. Thus, the permutation procedure described above to generate an empirical null distribution of T_{all} needs to be done only once for a GWAS.

The R code of AFC is available at Shuanglin Zhang's homepage http://www.math.mtu.edu/~shuzhang/software.html.

1.2.2 Comparison of Methods

We compare the performance of our method with those of TATES (van der Sluis et al., 2013), Tippett's method (Pesarin & Salmaso, 2010), Fisher's Combination test (FC) (Yang et al., 2016), Multivariate Analysis of Variance (MANOVA) (Cole et al., 1994), joint model of Multiple Phenotypes (MultiPhen) (O'Reilly et al., 2012), and Sum Score method (SUMSCORE) (van der Sluis et al., 2013). Here we briefly introduce each of those methods using the notations in the method section.

TATES: Combine the K phenotype-specific p-values obtained in standard univariate GWAS to acquire one overall p-value, $\min_k \left(\frac{m_e p_{(k)}}{m_{e(k)}} \right)$, where m_e denotes the effective number of independent p-values of all K phenotypes, and $m_{e(k)}$ denotes the effective number of p-values among the top k p-values.

MANOVA: Consider a multivariate multiple linear regression model: $\boldsymbol{Y} = X\beta^T + \boldsymbol{\mathcal{E}}$, where $\boldsymbol{Y}$ denotes the $n \times K$ matrix of phenotypes, $\beta^T = (\beta_1, \ldots, \beta_K)$ is a vector of coefficients corresponding to the K phenotypes, and $\boldsymbol{\mathcal{E}}$ is the $n \times K$ matrix of random errors with each row of $\boldsymbol{\mathcal{E}}$ to be i.i.d. $MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{\mathcal{E}}$.

To test $H_0: \beta = 0$, the likelihood ratio test is equivalent to the Wilk's Lambda test statistic of MANOVA (Rencher, 2003), that is, $-2\log\Lambda = 2\left(l(\hat{\beta}, \hat{\boldsymbol{\Sigma}}) - l(0, \hat{\boldsymbol{\Sigma}}_0)\right) = n\log\frac{|\hat{\boldsymbol{\Sigma}}_0|}{|\hat{\boldsymbol{\Sigma}}|} = -n\log\left(\frac{|\boldsymbol{E}|}{|\boldsymbol{E}+\boldsymbol{H}|}\right)$. Here Λ denote the ratio of the likelihood function under H_0 to the likelihood function under H_1, $l(\beta, \boldsymbol{\Sigma})$ is the log-likelihood function, $\boldsymbol{H} = \hat{\beta}(X^TX)\hat{\beta}^T$ and $\boldsymbol{E} = \boldsymbol{Y}^T\boldsymbol{Y} - \hat{\beta}(X^TX)\hat{\beta}^T$, where $\hat{\beta} = \boldsymbol{Y}^TX(X^TX)^{-1}$ is the maximum likelihood estimator (MLE) of β, and $|\cdot|$ denotes the determinant of a matrix. Then the test statistic has an asymptotic χ^2_K distribution (Ray et al., 2016).

MultiPhen: By performing ordinal regression, MultiPhen develops a reversed analysis for joint analysis of multiple phenotypes by considering a genetic variant of interest $X = (x_1, \ldots, x_n)^T$ as a response variable, and the correlated phenotypes $Y_k = (y_{1k}, \ldots, y_{nk})^T$ as predictors.

SUMSCORE: Let T^k_{score} denote the score test statistic to test the association between the k^{th} phenotype and the genetic variant. The test statistic of SUMSCORE is given by $T_{SUMSCORE} = \sum_{k=1}^{K} T^k_{score}$. The p-value of $T_{SUMSCORE}$ is estimated using a permutation procedure.

Tippett: The test statistic of Tippett is given by $T_{Tippett} = \min_k p_k$. The p-value of $T_{Tippett}$ is estimated using a permutation procedure.

FC: The Fisher's combination test statistic is defined as $T_{FC} = \sum_{k=1}^{K} -2\log(p_k)$. Yang et al. (Yang et al., 2016) adopted three different approaches to calculate the p-value for correlated phenotypes. In this article, we calculate the p-value using a permutation procedure.

AFC, FC, and Tippett are closely related. Intuitively, when only one phenotype or very few phenotypes are associated with the variant, Tippett is more powerful than FC because in this case FC contains a lot of noises. When all phenotypes or a large proportion of the phenotypes are associated with the variant, FC is more powerful than

Tippett because in this case Tippett only uses the minimum p-value and loses information. AFC can be adaptive to the number of phenotypes associated with the variant.

1.3 Results

1.3.1 Simulation Studies

We generate genotype data at a genetic variant according to a minor allele frequency (MAF) under Hardy-Weinberg equilibrium. Phenotypes are generated similarly to that of van der Sluis et al. (van der Sluis et al., 2013). The phenotypic correlation structures mimic that of UK10K (UK10K Consortium, 2015), that is, the phenotypes are divided into several groups (factors) and the within-group correlation is larger than the between-group correlation. Denote $Y_k = (y_{1k}, \dots, y_{nk})^T$ as the k^{th} phenotype of n individuals and $X = (x_1, \dots, x_n)^T$ as the genotypic score of the n individuals at the genetic variant.

Scenario 1: considering one factor model with genetic variant effect on the factor. We first generate a common factor, $f = \beta X$, where f is the n by 1 common factor and β is the effect size. Then we simulate K phenotypes by

$$Y_k = af + \varepsilon_k \text{ for } k = 1,2, \dots, K, \tag{3}$$

where a is a factor loading, $\varepsilon_k = (\varepsilon_{1k}, \dots, \varepsilon_{nk})^{\boldsymbol{T}} \sim MVN(0, \boldsymbol{I_n})$, and $\boldsymbol{I_n}$ is the identity matrix.

Scenario 2: considering 4 factor model with the genetic variant effect on the fourth factor, each factor has $\frac{K}{4}$ (K is a multiple of 4) phenotypes. We generate 4 correlated factors using $(f_1, f_2, f_3, f_4)^T \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \big(1 - \rho_{fa}\big)\boldsymbol{I} + \rho_{fa}\boldsymbol{A}$, $\boldsymbol{A}$ is a matrix with elements of 1, $\boldsymbol{I}$ is the identity matrix, and ρ_{fa} is the correlation between any two factors. Then, we transform the fourth factor f_4 to f_4^* by $f_4^* = f_4 + \beta X$ and simulate K phenotypes using

$$Y_k = \begin{cases} af_1 + \varepsilon_1 \; for \; k = 1, \dots, \frac{K}{4} \\ af_2 + \varepsilon_2 \; for \; k = \frac{K}{4} + 1, \dots, \frac{K}{2} \\ af_3 + \varepsilon_3 \; for \; k = \frac{K}{2} + 1, \dots, \frac{3K}{4} \\ af_4^* + \varepsilon_4 \; for \; k = \frac{3K}{4} + 1, \dots, K \end{cases} \quad (4)$$

where a is a factor loading, $\varepsilon_j = \left(\varepsilon_{1j}, \dots, \varepsilon_{nj}\right)^T \sim MVN(0, \boldsymbol{I_n})$ for $j = 1, \dots, 4$, and β is the effect size.

Scenario 3: considering two factor model with the genetic variant effect on the second factor, each factor has $\frac{K}{2}$ (K is a multiple of 2) phenotypes. We generate two correlated factors using $(f_1, f_2)^T \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \left(1 - \rho_{fa}\right)\boldsymbol{I} + \rho_{fa}\boldsymbol{A}$, $\boldsymbol{A}$ is a matrix with elements of 1, $\boldsymbol{I}$ is the identity matrix, and ρ_{fa} is the correlation between any two factors. Then, we transform the second factor f_2 to f_2^* by $f_2^* = f_2 + \beta X$ and simulate K phenotypes using

$$Y_k = \begin{cases} af_1 + \varepsilon_1 \; for \; k = 1, \dots, \frac{K}{2} \\ af_2^* + \varepsilon_2 \; for \; k = \frac{K}{2} + 1, \dots, K \end{cases} \quad (5)$$

where a is a factor loading, $\varepsilon_j = \left(\varepsilon_{1j}, \dots, \varepsilon_{nj}\right)^T \sim MVN(0, \boldsymbol{I_n})$ for $j = 1,2$, and β is the effect size.

Scenario 4: considering 4 factor model with genetic variant effect specific to the K^{th} phenotype, each factor has $\frac{K}{4}$ (K is a multiple of 4) phenotypes. By using the original factors f_1, f_2, f_3, f_4 in scenario 2, we simulate K phenotypes using

$$Y_k = \begin{cases} af_1 + \varepsilon_1 \; for \; k = 1, \dots, \frac{K}{4} \\ af_2 + \varepsilon_2 \; for \; k = \frac{K}{4} + 1, \dots, \frac{K}{2} \\ af_3 + \varepsilon_3 \; for \; k = \frac{K}{2} + 1, \dots, \frac{3K}{4} \\ af_4 + \varepsilon_4 \; for \; k = \frac{3K}{4} + 1, \dots, K-1 \\ a(f_4 + \beta X) + \varepsilon_4 \; for \; k = K \end{cases} \quad (6)$$

where a is a factor loading, $\varepsilon_j = \left(\varepsilon_{1j}, \dots, \varepsilon_{nj}\right)^T \sim MVN(0, \boldsymbol{I_n})$ for $j = 1, \dots, 4$, and β is the effect size.

Scenario 5: considering one factor model with the genetic variant effect specific to the K^{th} phenotype. We simulate K phenotypes by

$$Y_k = \begin{cases} af + \varepsilon_k \; for \; k = 1, \dots, K-1 \\ a(f + \beta X) + \varepsilon_k \; for \; k = K \end{cases} \quad (7)$$

where $f \sim MVN(0, \boldsymbol{I_n})$, a is a factor loading, $\varepsilon_k = (\varepsilon_{1k}, \dots, \varepsilon_{nk})^{\boldsymbol{T}} \sim MVN(0, \boldsymbol{I_n})$, and $\boldsymbol{I_n}$ is the identity matrix, and β is the effect size

Scenario 6: considering a network model, where the K phenotypes are correlated and the correlation structure is not due to one or multiple underlying common factors. We generate K phenotypes independent of genotypes for each individual by using $\left(\tilde{Y}_1, \dots, \tilde{Y}_K\right)^T \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \left(1 - \rho_{phe}\right)\boldsymbol{I} + \rho_{phe}\boldsymbol{A}$, $\boldsymbol{A}$ is a matrix with elements of 1, $\boldsymbol{I}$ is the identity matrix, and ρ_{phe} is the correlation between any two phenotypes. After generating $\tilde{Y}_1, \dots, \tilde{Y}_K$, let $Y_k = \beta X + \tilde{Y}_k + \varepsilon_k$, where $\varepsilon_k = (\varepsilon_{1k}, \dots, \varepsilon_{nk})^T \sim MVN(0, \boldsymbol{I_n})$.

In scenarios 2-5, the within-factor correlation is a^2 and between-factor correlaiton is $a^2\rho_{fa}$. To evaluate type I error of the proposed method, we generate phenotypic values independent of genotypes by assigning $\beta = 0$. To evaluate power, we generate phenotypic values according to the six scenarios described above.

1.3.2 Simulation Results

We use two sets of simulations to evaluate the type I error rates of the proposed method. The first set of simulations is normal simulation studies and includes 10,000 replicated samples for each sample size under each scenario. The p-values are estimated using 10,000 permutations. For 10,000 replicated samples, the 95% confidence intervals (CIs) for type I error rates at the nominal levels 0.01 and 0.001 are (0.008, 0.012) and (0.0004, 0.0016), respectively. The estimated type I error rates of the proposed test (AFC) are summarized in Table 1.1. From Table 1.1, we can see that most of the estimated type I error rates are within the 95% CIs and those type I error rates not within the 95% CIs are very close to the bound of the corresponding 95% CI, which indicates that the proposed method is valid.

The second set of simulations mimics GWAS. To be comparable with the real data analysis, we generate 6,000 unrelated individuals with 8 phenotypes at 10^6 variant sites under each scenario. The phenotypes are independent of genotypes. The MAF at each variant is a random number between 0.05 and 0.5. The null distributions of $T_1,\ldots,T_K$ and T_{all} are generated by 10^6 permutations using the genotypes at the first variant. We consider genotypes at 10^6 variants as 10^6 replicated samples. For 10^6 replicated samples, the 95% confidence intervals (CIs) for the type I error rates at the nominal levels 10^{-3}, 10^{-4}, and 10^{-5} are $(0.94 \times 10^{-3}, 1.06 \times 10^{-3})$, $(0.8 \times 10^{-4}, 1.20 \times 10^{-4})$, and $(0.38 \times 10^{-5}, 1.62 \times 10^{-5})$, respectively. The estimated type I error rates of the proposed test (AFC) are summarized in Table 1.2. From Table 1.2, we can see that all of the estimated type I error rates are within the 95% CIs, which indicates that the proposed method is valid.

For power comparisons, we consider (1) power as a function of the effect size under all six scenarios, and (2) power as a function of factorial correlation (ρ_{fa}) under scenarios 2-4 and power as a function of phenotypic correlation (ρ_{phe}) under scenario 6 because scenarios 1 and 5 are one factor model and thus have no ρ_{fa} and ρ_{phe} involved.

For Figures 1.1 and 1.2, the p-values of AFC, FC, Tippett, and SUMSCORE are estimated using 10,000 permutations and the p-values of TATES, MultiPhen, and MANOVA are estimated using asymptotic distributions. The powers of all tests are evaluated using 1,000 replicated samples at 0.1% significance level. For Figure 1.3, the p-values of AFC, FC, Tippett, and SUMSCORE are estimated using 10^7 permutations. The powers of all tests are evaluated using 1,000 replicated samples at 10^{-6} significance level.

Figure 1.1 gives the power comparisons of the 7 tests (AFC, TATES, Tippett, FC, MANOVA, MultiPhen, and SUMSCORE) for the power as a function of the effect size based on the six scenarios for 20 phenotypes. This figure shows that (1) AFC is either the most powerful test (genotypes directly impact on a portion of the phenotypes: scenarios 2-3) or comparable to the most powerful test (genotypes directly impact on all phenotypes or a single phenotype: scenarios 1, 4, 5, and 6); (2) TATES and Tippett have similar power and are much less powerful than other methods when genotypes directly impact on all phenotypes (scenarios 1 and 6); (3) MANOVA and MultiPhen have similar power and are much less powerful than other methods when genotypes directly impact on a portion of the phenotypes (scenarios 2-3); and (4) SUMSCORE and FC have similar power and are much less powerful than other methods when genotypes directly impact on a single phenotype (scenarios 4-5).

Power comparisons of the 7 tests for the power as a function of the factorial correlation (ρ_{fa}) under scenarios 2-4 and as a function of the phenotypic correlation (ρ_{phe}) under scenario 6 are given by Figure 1.2. This figure shows that under scenario 4, the powers of all tests do not change with the factorial correlation because only one phenotype is associated with the variant and thus the factorial correlation does not change the information of association between the variant and phenotypes. Under scenarios 2, 3 and 6, (1) the powers of SUMSCORE and FC decrease with the increasing of the factorial or phenotypic correlation because SUMSCORE and FC involve all phenotypes and thus information contained by all phenotypes will decrease with the increasing of the

factorial or phenotypic correlation; (2) the powers of TATES and Tippett do not change with the increasing of the factorial or phenotypic correlation because TATES and Tippett essentially only depend on the phenotype that has the strongest association with the variant; (3) under scenario 6, the power of AFC decreases with the increasing of the phenotypic correlation; under scenarios 2-3, the power of AFC does not change much with the factorial correlation; and (4) under scenario 6, the powers of MANOVA and MultiPhen decrease with the increasing of the phenotypic correlation; under scenarios 2-3, the powers of MANOVA and MultiPhen increase with the increasing of the factorial correlation, which is consistent with the results of Ray et al.[38]. We also give power comparisons of the 7 tests using a significance level of 10^{-6} with 10^7 permutations and 1,000 replicates for the power as a function of effect size (β) under scenario 2 (Figure 1.3). Figure 1.3 shows that the patterns of the power comparisons using significance level 10^{-6} are similar to that using a significance level of 0.1% in Figure 1.1 (scenario 2).

In summary, the proposed method has correct type I error rates and is either the most powerful test or comparable with the most powerful tests. No other methods have consistently good performance under the six simulation scenarios.

1.3.3 Real Data Analysis

Chronic obstructive pulmonary disease (COPD) is one of the most common lung diseases characterized by long term poor airflow and is a major public health problem (Nazir & Erbland, 2009). The COPDGene Study (Regan et al., 2011) (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000179.v1.p1) is a multi-center genetic and epidemiologic investigation to study COPD. This study is sufficiently large and appropriately designed for genome-wide association analysis of COPD. In this study, a total of more than 10,000 subjects have been recruited including non-Hispanic Whites (NHW) and African-Americans (AA). The participants in this study have completed a detailed protocol, including questionnaires, pre- and post-bronchodilator spirometry, high-resolution CT scanning of the chest, exercise capacity

(assessed by six-minute walk distance), and blood samples for genotyping. The participants were genotyped using the Illumina OmniExpress platform. The genotype data have gone through standard quality-control procedures for genome-wide association analysis detailed at http://www.copdgene.org/sites/default/files/GWAS_QC_Methodology_20121115.pdf. Variants with MAF<1% were excluded in the data set.

To evaluate the performance of our proposed method on a real data set, we applied all of the 7 methods to the COPDGene of NHW population to carry out GWAS of COPD-related phenotypes. Based on the literature studies of COPD (Chu et al, 2014; Han et al., 2011), we selected 7 key quantitative COPD-related phenotypes, including FEV1 (% predicted FEV1), Emphysema (Emph), Emphysema Distribution (EmphDist), Gas Trapping (GasTrap), Airway Wall Area (Pi10), Exacerbation frequency (ExacerFreq), Six-minute walk distance (6MWD), and 4 covariates, including BMI, Age, Pack-Years (PackYear) and Sex. EmphDist is the ratio of emphysema at -950 HU in the upper 1/3 of lung fields compared to the lower 1/3 of lung fields. Followed by Chu et al. (Chu et al, 2014), we did a log transformation on EmphDist in the following analysis. The correlation structure of the 7 COPD-related phenotypes is given in Figure 1.4. In the analysis, participants with missing data in any of the 11 variables were excluded. Therefore, a complete set of 5,430 individuals across 630,860 SNPs were used in the following analyses. In the analysis, we first adjusted each of the 7 phenotypes for the 4 covariates using linear models. Then, we performed the analysis based on the adjusted phenotypes.

To identify SNPs associated with the phenotypes, we adopted the commonly used genome-wide significance level 5×10^{-8}. The results were summarized in Table 1.3. There were total 14 SNPs in Table 1.3. All of the 14 SNPs had previously been reported to be in association with COPD by eligible studies (Brehm et al., 2011; Cho et al., 2010; Cho et al., 2014; Cui et al., 2014; Du et al., 2016; Hancock et al., 2010; Li et al., 2011b; Lutz, et al., 2015; Pillai, et al., 2009; Wilk et al., 2009; Wilk et al., 2012; Young et al., 2010; Zhang et al., 2011; Zhu et al., 2014). From Table 1.3, we can see that MultiPhen

identified 14 SNPs; MANOVA identified 13 SNPs; AFC identified 12 SNPs, FC and SUMSCORE identified 10 SNPs; and TATES and Tippett identified 9 SNPs. Among the five methods based on combining test statistics from univariate analysis (AFC, TATES, Tippett, FC, and SUMSCORE), AFC identified 2 or 3 more genome-wide significant SNPs than the other 4 methods.

1.4 Discussion

GWAS have identified many variants with each variant affecting 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. In this article, we developed a new method AFC to jointly analyze multivariate phenotypes in genetic association studies. We used extensive simulation studies as well as real data application to compare the performance of AFC with TATES, Tippett, FC, MANOVA, MultiPhen, and SUMSCORE. Our simulation results showed that AFC has correct type I error rates. With respect to power, AFC is either the most powerful test or has similar power with the most powerful test under a variety of simulation scenarios. Additionally, the real data analysis results demonstrated that the proposed method has great potential in GWAS on complex diseases with multiple phenotypes such as COPD.

AFC has several important advantages. First, it allows researchers to test genetic associations using standard GWAS software. Second, phenotypes of different types (e.g., dichotomous, ordinal, continuous) can be easily analyzed simultaneously. Third, since AFC is based on p-values obtained from standard univariate GWAS, it can not only test the association between multiple phenotypes and one genetic variant of interest, but also can test the association between multiple phenotypes and multiple genetic variants. For common variants, multiple-variant AFC can be applied based on p-values obtained in standard univariate GWAS for each variant and each phenotype. For rare variants, we can first combine genotypes of rare variants by giving different weights, hoping that we give

big weights to the variants having strong associations with the phenotypes. Then, we can apply AFC to test the association between the combined genotypes and multiple phenotypes. In conclusion, we showed that our proposed method provides a useful framework for joint analysis of multiple phenotypes in association studies.

It is well known that the effect sizes of identified variants are often small and that a large sample size is necessary to ensure sufficient power to detect such variants. A common strategy to increase sample size is to perform a meta-analysis by combining summary statistics from a series of studies. The proposed AFC method can be applied to meta-analysis. Suppose that there are L independent studies containing the variant of interest and each study has K phenotypes. Let $T_{1l}, \dots, T_{Kl}$ denote the summary statistics from the l^{th} study. Suppose that $T_l = (T_{1l}, \dots, T_{Kl})^T \sim N(0, \Sigma_l)$ under the null hypothesis, where Σ_l can be estimated from the values of T_l for all independent SNPs in the GWAS from the l^{th} study (Zhu et al., 2015b). Then, $T = (T_1^T, \dots, T_L^T)^T \sim N(0, \Sigma)$, where $\Sigma = diag(\Sigma_1, \dots, \Sigma_L)$. From T, we can calculate the corresponding p-values $P = (P_1^T, \dots, P_L^T)^T$, where $P_l = (p_{1l}, \dots, p_{Kl})^T$. The AFC test statistic is based on the p-values P. In the permutation procedure, we can generate T according to the distribution $N(0, \Sigma)$ and then we can calculate the p-values P in each permutation.

1.5 Tables and Figures

Table 1.1. The estimated type I error rates of the proposed method for MAF equals 0.3. α is the significance level. 10,000 replicates are used in the simulations.

α	Sample size	Scenario					
		1	2	3	4	5	6
0.01	1000	0.0088	0.0110	0.0105	0.0083	0.0083	0.0108
	2000	0.0095	0.0107	0.0094	0.0083	0.0098	0.0110
0.001	1000	0.0008	0.0015	0.0012	0.0008	0.0007	0.0012
	2000	0.0015	0.0014	0.0007	0.0009	0.0011	0.0014

Table 1.2. The estimated type I error rates of the proposed method that mimic GWAS. α is the significance level.

α	Scenario					
	1	**2**	**3**	**4**	**5**	**6**
1.00×10^{-3}	1.02×10^{-3}	1.06×10^{-3}	0.94×10^{-3}	1.03×10^{-3}	1.00×10^{-3}	1.05×10^{-3}
1.00×10^{-4}	1.03×10^{-4}	1.20×10^{-4}	0.80×10^{-4}	0.97×10^{-4}	1.20×10^{-4}	0.82×10^{-4}
1.00×10^{-5}	1.30×10^{-5}	1.10×10^{-5}	1.50×10^{-5}	1.40×10^{-5}	1.00×10^{-5}	0.50×10^{-5}

Table 1.3. Significant SNPs and the corresponding p-values in the analysis of COPDGene. The p-values of AFC, Tippett, FC, and SUMSCORE are evaluated using 109 permutations. The p-values of TATES, MANOVA, and MultiPhen are evaluated using asymptotic distributions. The graying out p-values indicate the p-values $> 5 \times 10^{-8}$.

Chr	Position	Variant identifier	AFC	TATES	Tippett	FC	MANOVA	MultiPhen	SUM-SCORE
4	145431497	rs1512282	1.10×10^{-8}	5.77×10^{-9}	8.00×10^{-9}	6.00×10^{-9}	1.69×10^{-9}	1.03×10^{-9}	2.00×10^{-8}
4	145434744	rs1032297	0	6.22×10^{-13}	0	0	6.52×10^{-14}	7.69×10^{-14}	0
4	145474473	rs1489759	0	2.52×10^{-16}	0	0	1.11×10^{-16}	1.22×10^{-16}	0
4	145485738	rs1980057	0	9.35×10^{-17}	0	0	6.68×10^{-17}	8.14×10^{-17}	0
4	145485915	rs7655625	0	1.64×10^{-16}	0	0	7.12×10^{-17}	9.13×10^{-17}	0
15	78882925	rs16969968	0	2.98×10^{-8}	4.90×10^{-8}	1.00×10^{-8}	1.32×10^{-11}	7.84×10^{-12}	3.33×10^{-8}
15	78894339	rs1051730	0	2.63×10^{-8}	4.20×10^{-8}	9.00×10^{-9}	1.41×10^{-11}	8.16×10^{-12}	1.00×10^{-8}
15	78898723	rs12914385	0	5.14×10^{-10}	0	0	1.76×10^{-12}	1.48×10^{-12}	0
15	78911181	rs8040868	0	2.40×10^{-9}	5.00×10^{-9}	0	2.74×10^{-12}	2.59×10^{-12}	0
15	78878541	rs951266	0	5.17×10^{-8}	8.10×10^{-8}	1.50×10^{-8}	1.77×10^{-11}	1.02×10^{-11}	2.15×10^{-8}
15	78806023	rs8034191	1.40×10^{-8}	1.02×10^{-7}	1.70×10^{-7}	9.50×10^{-8}	2.14×10^{-10}	7.74×10^{-11}	8.43×10^{-8}
15	78851615	rs2036527	2.90×10^{-8}	1.56×10^{-7}	2.41×10^{-7}	1.12×10^{-7}	3.99×10^{-10}	1.77×10^{-10}	2.01×10^{-7}
15	78826180	rs931794	6.30×10^{-8}	1.18×10^{-7}	1.94×10^{-7}	2.67×10^{-7}	2.35×10^{-10}	9.09×10^{-11}	3.32×10^{-7}
15	78740964	rs2568494	5.00×10^{-6}	2.88×10^{-5}	3.42×10^{-5}	1.34×10^{-5}	1.05×10^{-7}	4.23×10^{-8}	2.11×10^{-5}

Figure 1.1. Power comparisons of the 7 tests for power as a function of effect size (β) under the 6 scenarios. The total number of phenotypes is 20. The sample size is 1,000. MAF is 0.3. The factor loadings are 0.75. In scenarios 2, 3 and 4, the factorial correlation (ρ_{fa}) is 0.1. In scenario 6, the phenotypic correlation (ρ_{phe}) is 0.1. The powers are evaluated at 0.1% significance level.

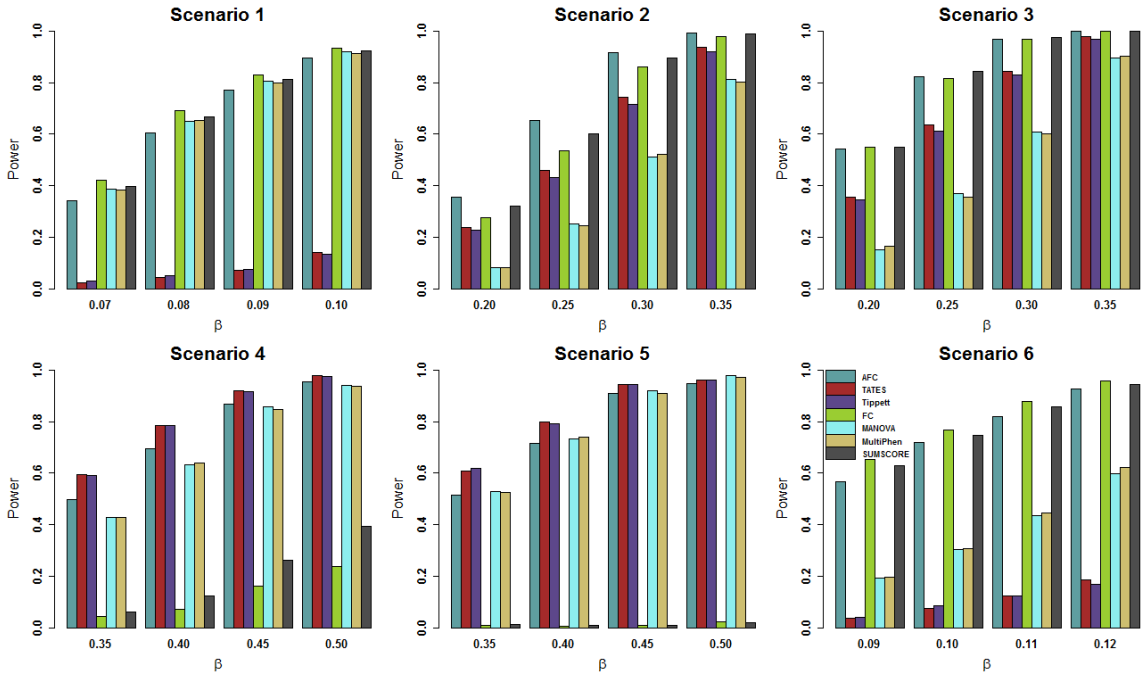


Figure 1.2. Power comparisons of the 7 tests for power as a function of factorial correlation (ρ_{fa}) under scenarios 2 to 4, and as a function of phenotypic correlation (ρ_{phe}) under scenario 6. The total number of phenotypes is 20. The sample size is 1,000. MAF is 0.3. The factor loadings are 0.75. In scenarios 2 and 3, the effect size (β) is 0.2. In scenario 4, the effect size (β) is 0.3. In scenario 6, the effect size (β) is 0.1. The powers are evaluated at 0.1% significance level.

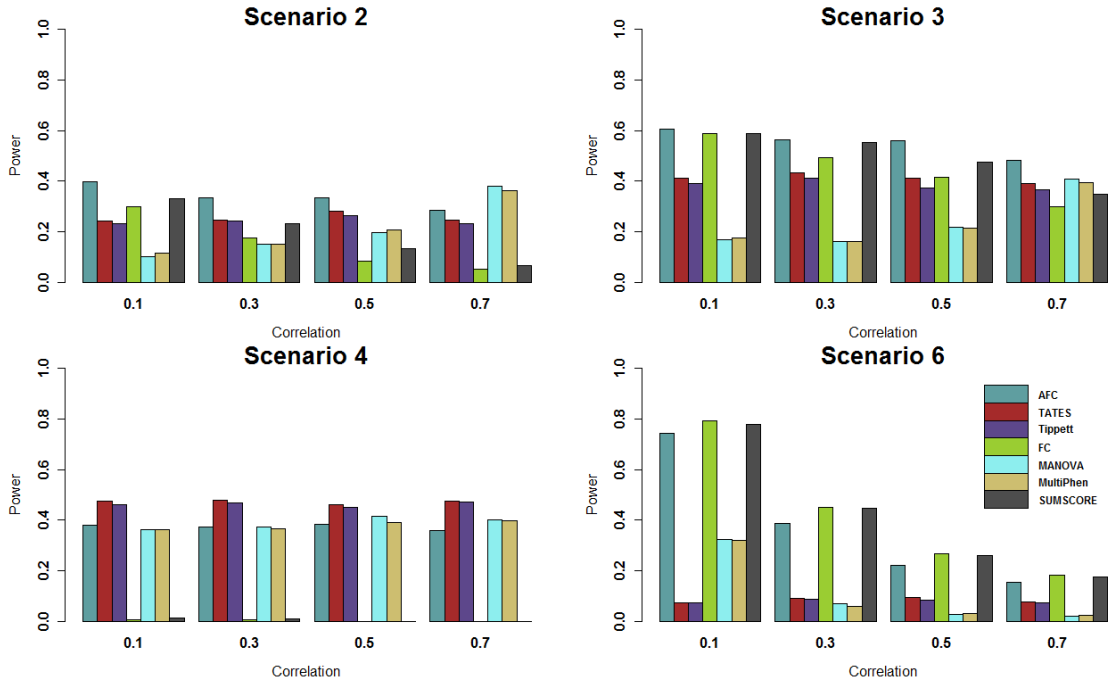


Figure 1.3. Power comparisons of the 7 tests for power as a function of effect size (β) under scenario 2. The total number of phenotypes is 20. The sample size is 1,000. MAF is 0.3. The factor loadings are 0.75. The factorial correlation (ρ_{fa}) is 0.1. The powers are evaluated at 10^{-6} significance level while p-values of AFC, FC, Tippet, and SUMSCORE are evaluated by 10^{7} permutations.

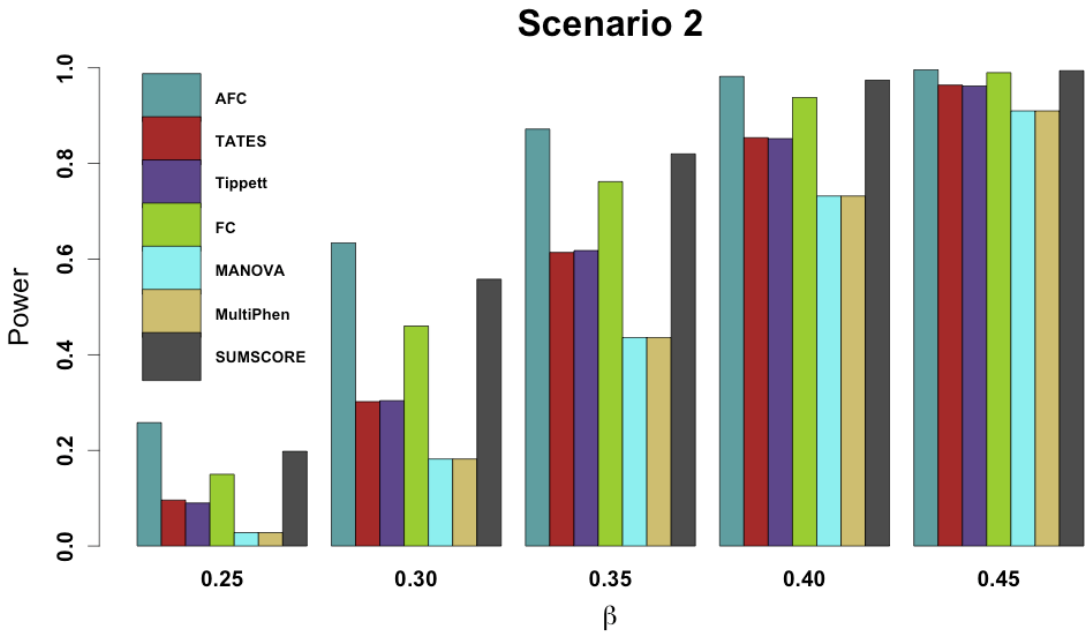
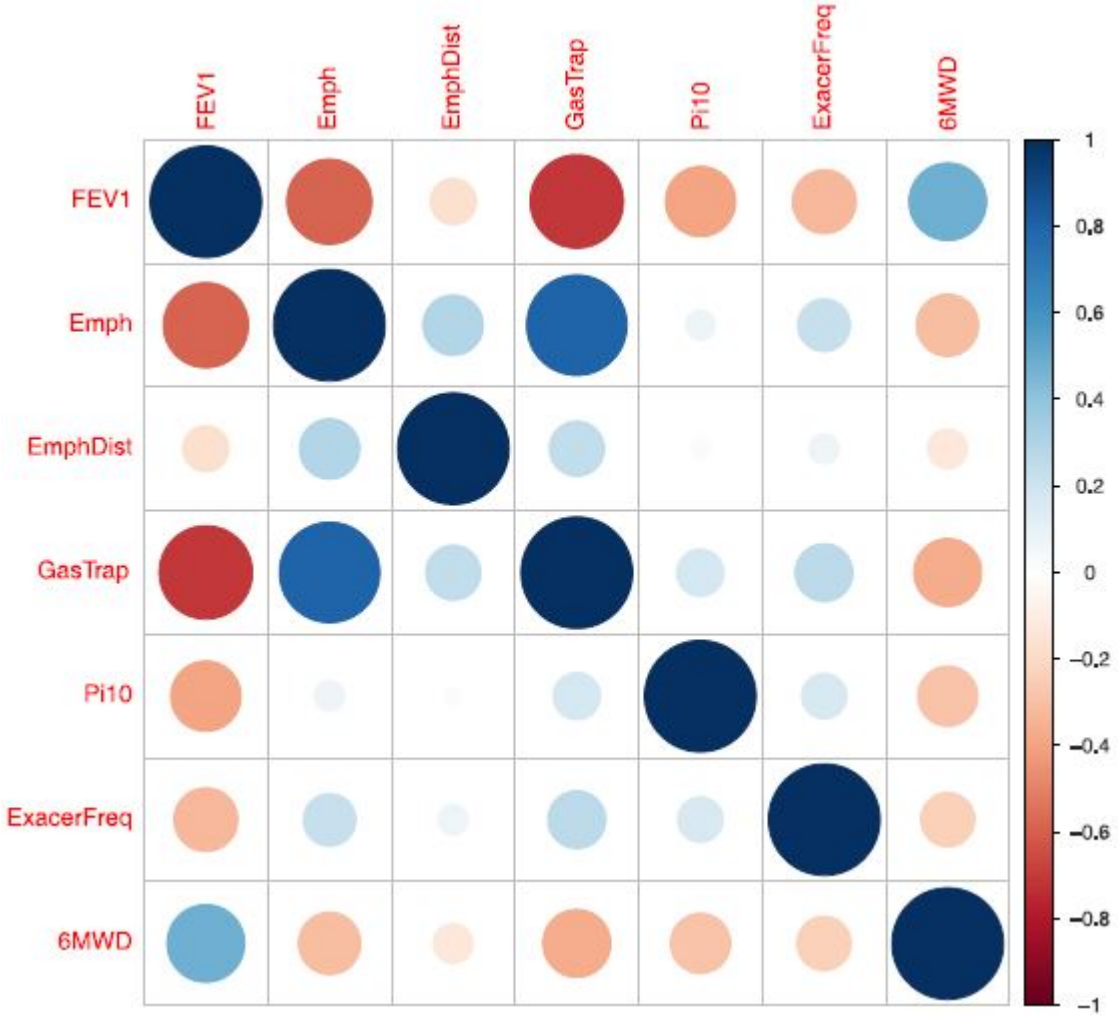


Figure 1.4. The correlation matrix plot of the 7 COPD-related phenotypes.



2 Joint Analysis of Multiple Phenotypes in Association Studies Using Allele-based Clustering Approach for Non-Normal Distributions

In the study of complex diseases, several correlated phenotypes are usually measured. There is also increasing evidence showing that testing the association between a single-nucleotide polymorphism (SNP) and multiple-dependent phenotypes jointly is often more powerful than analyzing only one phenotype at a time. Therefore, developing statistical methods to test for genetic association with multiple phenotypes has become increasingly important. In this study, we develop an Allele-Based Clustering (ABC) approach for the joint analysis of multiple non-normal phenotypes in association studies. In the ABC method, we consider the alleles at a SNP of interest as a dependent variable with two classes, and the correlated phenotypes as predictors to predict the alleles at the SNP of interest. We perform extensive simulation studies to evaluate the performance of the ABC method and compare the power of the ABC method with the powers of Adaptive Fisher's Combination test (AFC), Trait-based Association Test that uses Extended Simes procedure (TATES), Fisher's Combination test (FC), the standard MANOVA, and the joint model of Multiple Phenotypes (MultiPhen). Our simulation studies show that the proposed method has correct type I error rates and is much more powerful than other methods for some non-normal distributions.

2.1 Background

In the study of a complex disease, data on multiple phenotypes are often collected to have a better understanding of the disease. For example, cardiovascular disease (CVD) is characterized by high levels of both low-density serum lipoprotein levels (LDL) and systolic blood pressure (SBP) (Newman III et al., 1986; Majumdar et al., 2015); Chronic obstructive pulmonary disease (COPD) is one of the most common lung diseases characterized by reduced expiratory airflow, symptoms of cough, sputum production, and dyspnea (Casaburi et al., 2002); Thromboembolic disease is characterized by the

intermediate correlated phenotypes such as Factor VII, VIII, IX, XI, XII, and von Willebrand factor (Germain et al.,2011; Ray et al. 2016; Souto et al. 2000).

Statistical methods for detecting association using multivariate phenotypes have been developed. These methods can be roughly divided into two groups: univariate methods and multivariate methods. Univariate methods are based on univariate analyses and combine the results of univariate analyses. Multivariate methods are based on models that jointly model multiple phenotypes. Univariate methods are popular in practice because it is much easier to construct a univariate association test statistic than a multivariate association test statistic. Recently, several univariate methods have appeared to explore the genetic association with multiple phenotypes by considering the correlation structure among phenotypes (Kwak & Pan, 2016; Liang et al., 2016; van der Sluis et al., 2013; Yang et al., 2016). Kwak and Pan (2016) proposed an adaptive gene-based test and a pathway-based test for association analysis of multiple phenotypes with summary statistics from genome-wide association studies (GWAS). The Trait-based Association Test that uses Extended Simes procedure (TATES) combines p-values obtained in standard univariate GWAS while correcting for the correlation between p-values (van der Sluis et al., 2013). Adaptive Fisher's Combination test (AFC) combines p-values obtained in standard univariate GWAS by using the optimal number of p-values which is determined by the data (Liang et al., 2016).

Comparing univariate methods with multivariate methods, univariate methods may lose power since univariate methods ignore the extra information by analyzing multiple phenotypes in one unified analysis. Recently, several multivariate methods have been developed (Marchini et al., 2007; Galesloot et al., 2014; Zhou and Stephens, 2014; Korte et al., 2012; Casale et al., 2015; Zhang et al., 2014; Yan et al., 2013; O'Reilly et al., 2012; Tang and Ferreira, 2012; Aschard et al., 2014; Lange et al., 2004; Klei et al., 2008; Zhou et al., 2015). However, most of the multivariate methods are based on the assumption of normality for phenotypes. In this article, we develop a novel multivariate method, named an Allele-Based Clustering (ABC) approach, for the joint analysis of multiple non-normal phenotypes. The ABC approach is nonparametric in that it does not

assume any particular genetic models. Using extensive simulation studies, we evaluate the performance of the proposed method and compare the power of the proposed method with the powers of AFC (Liang et al., 2016), TATES (van der Sluis et al., 2013), Fisher's Combination test (FC) (Yang et al., 2016), MANOVA (Cole et al., 1994), and MultiPhen (O'Reilly et al., 2012). Our simulation studies show that the proposed method has correct type I error rates and is either the most powerful test or comparable with the most powerful tests for all simulation scenarios.

2.2 Methods

2.2.1 Allele-Based Clustering Method

Consider a sample of N unrelated individuals. Each individual has M phenotypes. We propose an allele-based method to test the null hypothesis H_0: none of the M phenotypes are associated with a SNP of interest. Denote $Y_{i,m}$ as the m^{th} phenotype of the i^{th} individual. For the SNP of interest, we use A (minor allele) and a to denote the two alleles of this SNP. Because each individual has two alleles at the SNP of interest, we use x_{2i-1} and x_{2i} to code the two alleles of the i^{th} individual. 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 that each allele of each individual has M phenotypes and define the m^{th} phenotype corresponding to the two alleles x_{2i-1} and x_{2i} of the i^{th} individual as $y_{2i-1,m}$ and $y_{2i,m}$, where $y_{2i-1,m} = y_{2i,m} = Y_{i,m}$. Hence, the total number of observations in the allele-based data is $2N$. We reindex the $2N$ allele-based data and use index j to denote the j^{th} allele-based data. Therefore, $y_{j,m}$ and x_j denote the j^{th} allele-based phenotypic and genotypic data.

We develop an Allele-Based Clustering (ABC) approach based on k-fold cross-validation for joint analysis of multiple phenotypes in association studies. In the ABC method, we consider the alleles at the SNP of interest as a dependent variable with two classes, and the correlated phenotypes as predictors to predict the alleles at the SNP of

interest. In the k-fold cross-validation (in this study, $k = 10$), the $2N$ allele-based phenotypic and genotypic data are divided into k mutually exclusive groups. We use each of the k groups as the testing set and the other $k - 1$ groups as the training set. For each pair of the testing and training sets, we use the training set to calculate the centers of the phenotypes corresponding to allele a and A, respectively, namely, $\bar{y}^{(0)} = \left(\bar{y}_1^{(0)}, \ldots, \bar{y}_M^{(0)}\right)^T$ and $\bar{y}^{(1)} = \left(\bar{y}_1^{(1)}, \ldots, \bar{y}_M^{(1)}\right)^T$ and use these centers to predict the alleles in the corresponding testing set, where $\bar{y}_m^{(0)}$ and $\bar{y}_m^{(1)}$ denote the mean values of the m^{th} phenotype for subsets $\{j: x_j = 0\}$ and $\{j: x_j = 1\}$ in the training set, $\{j: x_j = 0\}$ and $\{j: x_j = 1\}$ denote the subsets of index j in which $x_j = 0$ and $x_j = 1$, respectively. In the testing set, we calculate the Euclidean distance between y_j and $\bar{y}^{(0)}$, and y_j and $\bar{y}^{(1)}$, respectively. If y_j is closer to $\bar{y}^{(0)}$ than to $\bar{y}^{(1)}$, we predict the corresponding x_j as $\hat{x}_j = 0$; otherwise, we predict the corresponding x_j as $\hat{x}_j = 1$. In this way, each x_j has a predicted value $\hat{x}_j$ for $j = 1, \ldots, 2N$. For the SNP of interest, let q_0 and q_1 denote the allele frequency of a and A, respectively. For a given allele at the SNP of interest, the probability that we predict this allele by chance as allele a is q_0 and the probability that we predict this allele by chance as allele A is q_1. Therefore, we define the prediction accuracy as $T_{AC} = \frac{\#\{j: \hat{x}_j = x_j = 1\}}{q_1} + \frac{\#\{j: \hat{x}_j = x_j = 0\}}{q_0}$, where for a set D, $\#D$ denotes the number of elements in set D. We use T_{AC} as the test statistic to test the association between the M phenotypes and the SNP.

First, we delete those phenotypes that have weak associations with the SNP of interest based on the univariate analysis. For a given threshold t, $0 < t < 1$, we delete phenotypes with $p_m > t$ that means we select phenotypes with $p_m \leq t$, where p_m is the p-value to test the association between the m^{th} phenotype and the SNP. To test the association between the m^{th} phenotype and the SNP, we use the score test statistic under

the logistic model $\text{logit}(\pi_j) = \alpha_m + y_{j,m}\beta_m,\ j = 1, 2, \ldots, 2N$, where $\pi_j = \Pr(x_j = 1 | y_{j,m})$. The score test statistic to test the null hypothesis $H_0: \beta_m = 0$ is given by $S_m = U_m^2/V_m$, where $U_m = \sum_{j=1}^{2N} y_{j,m}(x_j - \bar{x})$, $V_m = \frac{1}{2N}\sum_{j=1}^{2N}(y_{j,m} - \bar{y}_m)^2 \sum_{j=1}^{2N}(x_j - \bar{x})^2$, $\bar{x} = \frac{1}{2N}\sum_{j=1}^{2N} x_j$, and $\bar{y}_m = \frac{1}{2N}\sum_{j=1}^{2N} y_{j,m}$. Under the null hypothesis, S_m follows a chi-square distribution with 1 degree of freedom. Let p_m denote the p-value of S_m to test the association between the m^{th} phenotype and the SNP of interest. For a given threshold t, $0 < t < 1$, we denote the prediction accuracy using the selected phenotypes as T_{AC_t}. Let P_t denote the p-value of the statistic T_{AC_t}. Our test statistic of ABC approach is given by

$$T_{ABC} = \min_t P_t.$$

T_{ABC} can be obtained by a simple grid search across a range of t. We choose grids $t_1, \ldots, t_S$ such that $0 \le t_1 < \quad < t_{S-1} < t_S \le 1$. In this study, we use $(t_1, \ldots, t_S) = (0.001, 0.01, 0.05, 1)$. By a grid search,

$$T_{ABC} = \min_t P_t = \min\{P_{t_1}, \ldots, P_{t_S}\}.$$

We use the following permutation procedure to evaluate the p-values of T_{ABC}.

1. In each permutation, we randomly shuffle the genotypes and recalculate $T_{AC_{t_1}}, \ldots, T_{AC_{t_S}}$. Suppose that we perform B times of permutations. Let $T_{AC_{t_s}}^{(b)}$ ($b = 0, 1, \ldots, B$) denote the value of $T_{AC_{t_s}}$ based on the b^{th} permuted data, where $b = 0$ represents the original data.
2. We transfer $T_{AC_{t_s}}^{(b)}$ to $P_{t_s}^{(b)}$ by

$$P_{t_s}^{(b)} = \frac{\#\{d : T_{AC_{t_s}}^{(d)} > T_{AC_{t_s}}^{(b)} \text{ for } d = 0, 1, \ldots, B\}}{B} \quad (1)$$

3. Let $T_{ABC}^{(b)} = \min\left\{P_{t_1}^{(b)}, \ldots, P_{t_S}^{(b)}\right\}$. Then, the p-value of T_{ABC} is given by

$$\frac{\#\left\{b : T_{ABC}^{(b)} < T_{ABC}^{(0)} \text{ for } b = 1, 2, \ldots, B\right\}}{B} \tag{2}$$

2.2.2 Comparison of Methods

We compare the performance of our ABC method with those of AFC (Liang et al., 2016), TATES (van der Sluis et al., 2013), FC (Yang et al., 2016), MANOVA (Cole et al., 1994), and MultiPhen (O'Reilly et al., 2012).

2.3 Results

2.3.1 Simulation Studies

We generate genotype data at a SNP according to a minor allele frequency (MAF) (MAF=0.3 in all simulation studies) under Hardy-Weinberg equilibrium. Based on the genotype data, phenotypes are generated similar to that of Liang et al. (Liang et al., 2016) and van der Sluis et al. (van der Sluis et al., 2013). The phenotypic correlation structures (the phenotypes are divided into several groups (factors) and the within-group correlation is larger than the between-group correlation) are also similar to that of UK10K (UK10K Consortium, 2015).

In the following six scenarios, we use the following notations. Denote $Y_m = \left(Y_{1,m}, Y_{2,m}, \ldots, Y_{N,m}\right)^T$ as the m^{th} phenotype of N individuals and $G = (g_1, g_2, \ldots, g_N)^T$ as the genotypic score of N individuals at the SNP. Let R denote the number of factors and $(f_1, \ldots, f_R)^T \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a correlation matrix with all off-diagonal elements equal to ρ_{fa} and ρ_{fa} is the correlation between any two factors. β is the effect size; a is a factor loading; $\varepsilon_m = (\varepsilon_{1m}, \ldots, \varepsilon_{Nm})^{\boldsymbol{T}}$ and ε_{im} $(i = 1, \ldots, N; m = 1, \ldots, M)$ are random errors.

Scenario 1: we simulate M phenotypes by

$$Y_m = a\beta G + \varepsilon_m \text{ for } m = 1,2,\dots,M \tag{3}$$

Scenario 2: consider a 4-factor model ($R = 4$) with the SNP effect on the fourth factor. We simulate M phenotypes using

$$Y_m = \begin{cases} af_1 + \varepsilon_m & \text{for } m = 1,\dots,\frac{M}{4} \\ af_2 + \varepsilon_m & \text{for } m = \frac{M}{4}+1,\dots,\frac{M}{2} \\ af_3 + \varepsilon_m & \text{for } m = \frac{M}{2}+1,\dots,\frac{3M}{4} \\ a(f_4 + \beta G) + \varepsilon_m & \text{for } m = \frac{3M}{4}+1,\dots,M \end{cases} \tag{4}$$

Scenario 3: consider a 2-factor model ($R = 2$) with the SNP effect on the second factor. We simulate M phenotypes using

$$Y_m = \begin{cases} af_1 + \varepsilon_m & \text{for } m = 1,\dots,\frac{M}{2} \\ a(f_2 + \beta G) + \varepsilon_m & \text{for } m = \frac{M}{2}+1,\dots,M \end{cases} \tag{5}$$

Scenario 4: consider a 4-factor model ($R = 4$) with the SNP effect on the M^{th} phenotype. We simulate M phenotypes using

$$Y_m = \begin{cases} af_1 + \varepsilon_m & \text{for } m = 1,\dots,\frac{M}{4} \\ af_2 + \varepsilon_m & \text{for } m = \frac{M}{4}+1,\dots,\frac{M}{2} \\ af_3 + \varepsilon_m & \text{for } m = \frac{M}{2}+1,\dots,\frac{3M}{4} \\ af_4 + \varepsilon_m & \text{for } m = \frac{3M}{4}+1,\dots,M-1 \\ a(f_4 + \beta G) + \varepsilon_m & \text{for } m = M \end{cases} \tag{6}$$

Scenario 5: consider a 2-factor model ($R = 2$) with the SNP effect on the M^{th} phenotype. We simulate M phenotypes using

$$Y_m = \begin{cases} af_1 + \varepsilon_m & \text{for } m = 1, \dots, \frac{M}{2} \\ af_2 + \varepsilon_m & \text{for } m = \frac{M}{2} + 1, \dots, M-1 \\ a(f_2 + \beta G) + \varepsilon_m & \text{for } m = M \end{cases} \quad (7)$$

Scenario 6: consider an M-factor model ($R = M$) with the SNP effect on all phenotypes. We simulate M phenotypes using

$$Y_m = a(f_m + \beta G) + \varepsilon_m \text{ for } m = 1,2, \dots, M \quad (8)$$

In scenarios 2-6, the within-factor correlation is a^2 and between-factor correlation is $a^2\rho_{fa}$. In scenarios 1-6, we consider random errors having the following four non-normal distributions: two log-normal distributions ($\text{log-norm}(0,1)$ and $\sqrt{5}\text{log-norm}(0,1)$), Student's t-distribution ($t(2)$), and Inverse-gamma distribution (Inverse-gamma$(2,1)$). We also consider random errors following the standard normal distribution. To evaluate type I error rates, we generate phenotypic values independent of genotypes by assigning $\beta = 0$. To evaluate power, we generate phenotypic values according to the six scenarios described above. In type I error evaluations and power comparisons, MAF is 0.3, the total number of phenotypes (M) is 8, the sample size (N) is 1,000, the factor loading (a) is 0.75, and the factorial correlation (ρ_{fa}) conditional on genotypes is 0.1.

2.3.2 Simulation Results

In each simulation scenario, the p-values of the proposed test (ABC), AFC, and FC are estimated using 1,000 permutations and the p-values of TATES, MANOVA, and MultiPhen are estimated using asymptotic distributions.

For type I error evaluation, we consider different distributions of random errors, different correlation structures of phenotypes, and different significance levels. In each

simulation scenario, the type I error rates of all of the six tests are evaluated using 1,000 replicated samples. For 1,000 replicated samples, the 95% confidence intervals (CIs) for the type I error rates at the nominal levels 0.01 and 0.05 are (0.0038, 0.0162) and (0.0365, 0.0635), respectively. The estimated type I error rates for random errors following Student's $t(2)$ distribution, Inverse-gamma$(1,2)$ distribution, $\text{log-norm}(0,1)$ distribution, $\sqrt{5}\text{log-norm}(0,1)$ distribution, and the standard normal distribution are summarized in Tables 2.1 to 2.5, respectively. From these tables, we can see that most of the estimated type I error rates of ABC, AFC, FC, TATES, and MANOVA are within the 95% CIs and those type I error rates not within 95% CIs are very close to the bound of the corresponding 95% CI, which indicates that ABC, AFC, FC, TATES, and MANOVA are valid. MultiPhen for Inverse-gamma$(1,2)$ distribution has little inflated type I error rates.

The powers of all tests are evaluated using 1,000 replicated samples at 1% significance level. Figures 2.1 to 2.5 provide the power comparisons of the six tests (ABC, AFC, TATES, FC, MANOVA, and MultiPhen) for the power as a function of the effect size based on the six scenarios for 8 phenotypes. Figure 2.1 shows that when random errors follow a Student's t distribution, ABC is the most powerful test in all the scenarios and is much more powerful than other methods. Figure 2.2 shows that when random errors follow an Inverse-gamma distribution, ABC is the most powerful test in all the scenarios and is much more powerful than the other methods when genotypes directly impact on a portion of phenotypes (scenarios 2-5). Figure 2.3 shows that when random errors follow a Log-normal distribution, ABC is the most powerful test when genotypes directly impact on a portion of phenotypes (scenarios 2-5) and is comparable to the most powerful one when genotypes directly impact on all phenotypes (scenarios 1 and 6). To show the robustness of ABC to phenotype outliers, we generate random errors following a Log-normal distribution with a variance 5 times as that in Figure 2.3. As shown in Figure 2.4, increasing variance 5 times, the pattern of power comparisons is still the same as that in Figure 2.3. We also provide power comparisons when random errors follow a standard normal distribution (Figure 2.5). As expected, ABC loses some power

comparing with that when random errors follow non-normal distributions, but is still more powerful than TATES for scenarios 1 and 6, more powerful than FC for scenarios 4 and 5, and more powerful than MANOVA and MultiPhen for scenario 3.

2.4 Discussion

In GWAS for complex diseases, the association between a SNP and each phenotype is usually weak. Combining multiple related phenotypes can increase the power to identify causal SNPs, thus is a practically important area that requires methodology work. So far, existing methods for associations of multiple phenotypes primarily focus on the phenotypes with a normal distribution. However, when the normal assumption is violated, these methods may not control type I error rates or may lose power (Majumdar et al., 2015).

In this study, we introduced an Allele-Based Clustering (ABC) approach for joint analysis of multiple non-normal distributed phenotypes in association studies. The ABC approach is nonparametric in that it does not assume any particular genetic models and there are no parameters need to be estimated. We use a data-driven approach to select phenotypes that are associated with a SNP and use these phenotypes to predict the alleles at the SNP by using Euclidean distance. Besides Euclidean distance, we can also use other similarity or dissimilarity measures to predict the alleles at a SNP, such as Minkowski distance, Mahalanobis distance, correlation coefficient, and cosine measure.

We use extensive simulations to assess the performance of ABC. We are able to demonstrate that ABC has correct type I error rates and is much more powerful than other methods for some non-normal distributions. Therefore, our simulations show that ABC is especially useful for non-normal data and it truly adds to the geneticist's toolbox. Moreover, ABC does not estimate any parameters and does not assume any particular models. It provides another way for analyzing multiple phenotypes simultaneously. In the method section, we describe our methods without considering covariates. If there are

covariates need to be considered, we can use the following approach to incorporate covariates in the ABC approach. Suppose that we have p covariates. Let $(z_{i1}, \ldots, z_{ip})^T$ denote the covariates of the i^{th} individual. We can adjust the effects of covariates by applying the linear regression model $Y_{i,m} = a_0 + a_{1m}z_{i1} + \cdots + a_{pm}z_{ip} + \varepsilon_{im}$ and using the residual of $Y_{i,m}$ to replace $Y_{i,m}$ in the ABC approach. Although using the regular linear regression to regress effects of covariates out is more suited for normally distributed data, it can also be applied to non-normally distributed data. For examples, to correct for population stratification, Price et al. (2006) used the regular linear regression to regress effects of covariates out for a qualitative phenotype and genotypes (both a qualitative phenotype and genotypes are non-normally distributed data); to adjust effects of covariates in rare variant association studies, Sha et al. (2012) also used the regular linear regression to regress effects of covariates out for a qualitative phenotype and genotypes.

The computation time required for running ABC depends on the sample size, the number of phenotypes, the number of folds in cross-validation, and the number of permutations. The running time of ABC with 1,000 permutations on the data set with 10,000 individuals, 8 phenotypes, and 10-fold cross-validation on a laptop with 4 Intel Cores @ 2.00GHz and 4 GB memory is no more than 2.4s. To perform genome-wide studies, we can first select SNPs 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 variants.

2.5 Tables and Figures

Table 2.1. The estimated type I error rates of the 6 tests under Student's $t(2)$ distribution. MAF is 0.3. α is the significance level. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The number of replications is 1,000. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

α	Method	Scenario					
		1	**2**	**3**	**4**	**5**	**6**
0.01	**ABC**	0.009	0.010	0.015	0.012	0.008	0.011
	AFC	0.009	0.010	0.015	0.013	0.015	0.015
	TATES	0.007	0.007	0.011	0.007	0.007	0.007
	FC	0.009	0.011	0.011	0.008	0.011	**0.018**
	MANOVA	0.008	0.010	0.009	0.008	0.008	0.013
	MultiPhen	0.014	0.007	0.014	0.008	0.012	**0.019**
0.05	**ABC**	0.039	0.042	0.055	0.048	0.045	0.057
	AFC	0.050	0.058	0.051	0.050	0.049	0.049
	TATES	0.042	0.045	0.043	0.041	0.053	0.044
	FC	0.053	0.054	0.056	0.046	0.045	0.048
	MANOVA	0.050	0.051	0.046	0.04	0.044	0.046
	MultiPhen	0.062	0.052	0.063	0.042	0.058	**0.069**

Table 2.2. The estimated type I error rates of the 6 tests under Inverse-gamma(1,2) distribution. MAF is 0.3. α is the significance level. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The number of replications is 1,000. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

α	Method	Scenario					
		1	2	3	4	5	6
0.01	ABC	0.016	0.015	0.016	0.007	0.013	0.013
	AFC	0.012	**0.019**	0.009	0.010	0.014	0.007
	TATES	0.009	0.009	0.009	0.008	0.015	0.008
	FC	0.012	0.016	0.006	0.006	0.013	0.005
	MANOVA	0.009	0.009	0.008	0.009	0.014	0.004
	MultiPhen	0.010	**0.020**	0.013	0.013	**0.019**	**0.017**
0.05	ABC	0.063	0.056	0.059	0.041	0.054	0.050
	AFC	0.056	**0.065**	0.048	0.048	0.061	0.046
	TATES	0.040	0.050	**0.035**	**0.035**	0.051	0.042
	FC	0.055	0.061	0.046	0.052	0.062	0.046
	MANOVA	0.053	0.058	0.041	0.045	0.050	0.045
	MultiPhen	**0.068**	**0.084**	**0.072**	**0.063**	**0.074**	**0.069**

Table 2.3. The estimated type I error rates of the 6 tests under Log-norm$(0,1)$ distribution. MAF is 0.3. α is the significance level. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The number of replications is 1,000. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

α	Method	Scenario					
		1	2	3	4	5	6
0.01	ABC	0.014	0.012	0.01	0.007	0.009	0.006
	AFC	0.013	0.013	0.009	0.013	0.012	0.012
	TATES	0.011	0.013	0.006	0.006	0.008	0.008
	FC	0.008	0.010	0.009	0.013	0.008	0.010
	MANOVA	0.007	0.011	0.009	0.011	0.012	0.009
	MultiPhen	0.014	0.011	0.005	0.015	0.012	0.010
0.05	ABC	0.057	0.057	0.050	0.054	0.049	0.050
	AFC	0.040	0.06	0.038	0.056	0.040	0.046
	TATES	0.043	0.046	0.038	0.049	0.037	0.048
	FC	0.037	0.057	0.038	0.053	0.050	0.044
	MANOVA	0.039	0.055	0.042	0.048	0.041	0.041
	MultiPhen	0.056	0.055	0.056	0.057	0.045	0.060

Table 2.4. The estimated type I error rates of the 6 tests under $\sqrt{5}\log\text{-norm}(0,1)$ distribution. MAF is 0.3. α is the significance level. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The number of replications is 1,000. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

α	Method	Scenario					
		1	2	3	4	5	6
0.01	ABC	0.011	0.012	0.014	0.008	0.011	0.012
	AFC	0.015	0.013	0.010	0.014	0.008	0.014
	TATES	0.013	0.012	**0.017**	0.005	0.011	0.012
	FC	0.014	0.010	0.010	0.013	0.006	0.012
	MANOVA	0.014	0.009	0.008	0.013	0.006	0.011
	MultiPhen	**0.017**	0.014	0.014	**0.017**	0.011	0.012
0.05	ABC	0.052	0.058	0.054	0.045	0.049	0.049
	AFC	0.053	0.050	0.049	0.051	0.049	0.053
	TATES	0.043	0.056	0.053	0.039	0.048	0.052
	FC	**0.066**	0.055	0.044	0.050	0.045	0.049
	MANOVA	**0.064**	0.058	0.041	0.046	0.046	0.044
	MultiPhen	**0.067**	0.061	0.050	0.062	0.058	0.048

Table 2.5. The estimated type I error rates of the 6 tests under the standard normal distribution. MAF is 0.3. α is the significance level. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The number of replications is 1,000. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

α	Method	Scenario					
		1	2	3	4	5	6
0.01	ABC	0.007	0.012	0.011	0.009	0.013	0.008
	AFC	0.015	0.012	0.008	0.006	0.007	0.011
	TATES	0.012	0.015	0.009	0.012	0.005	0.009
	FC	0.013	0.014	0.009	0.007	0.007	0.007
	MANOVA	0.014	0.008	0.009	0.008	0.008	0.008
	MultiPhen	0.010	0.010	0.012	0.007	0.009	0.008
0.05	ABC	0.047	0.052	0.049	0.051	0.054	0.051
	AFC	0.062	0.048	0.049	0.042	0.054	0.045
	TATES	0.050	0.052	0.051	0.042	0.053	0.048
	FC	0.050	0.052	0.040	0.046	0.051	0.049
	MANOVA	0.052	0.050	0.055	0.051	0.057	0.044
	MultiPhen	0.055	0.055	0.060	0.048	0.056	0.044

Figure 2.1. Power comparisons of the 6 tests for power as a function of effect size (β) for Student's $t(2)$ distribution under the 6 scenarios. MAF is 0.3. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The powers are evaluated at 1% significance level.

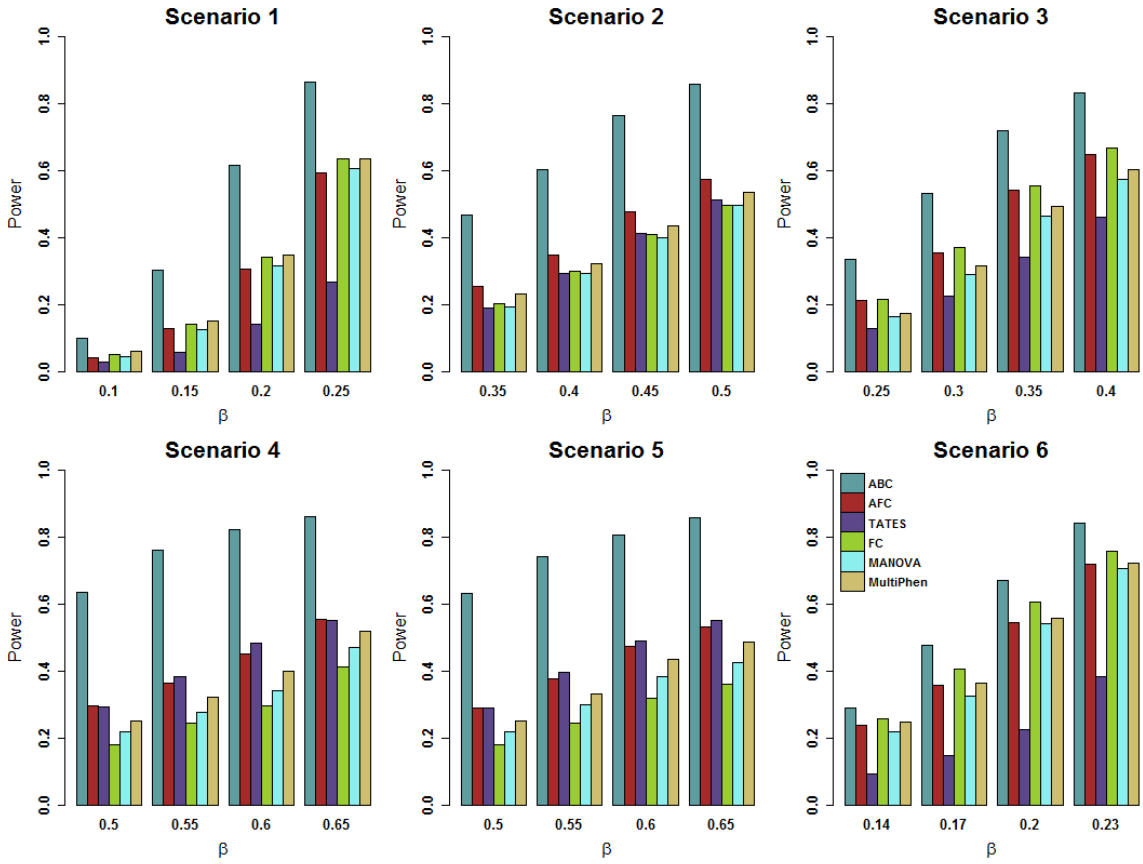


Figure 2.2. Power comparisons of the 6 tests for power as a function of effect size (β) for Inverse-gamma ($\alpha = 2, \beta = 1$) distribution under the 6 scenarios. MAF is 0.3. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The powers are evaluated at 1% significance level.

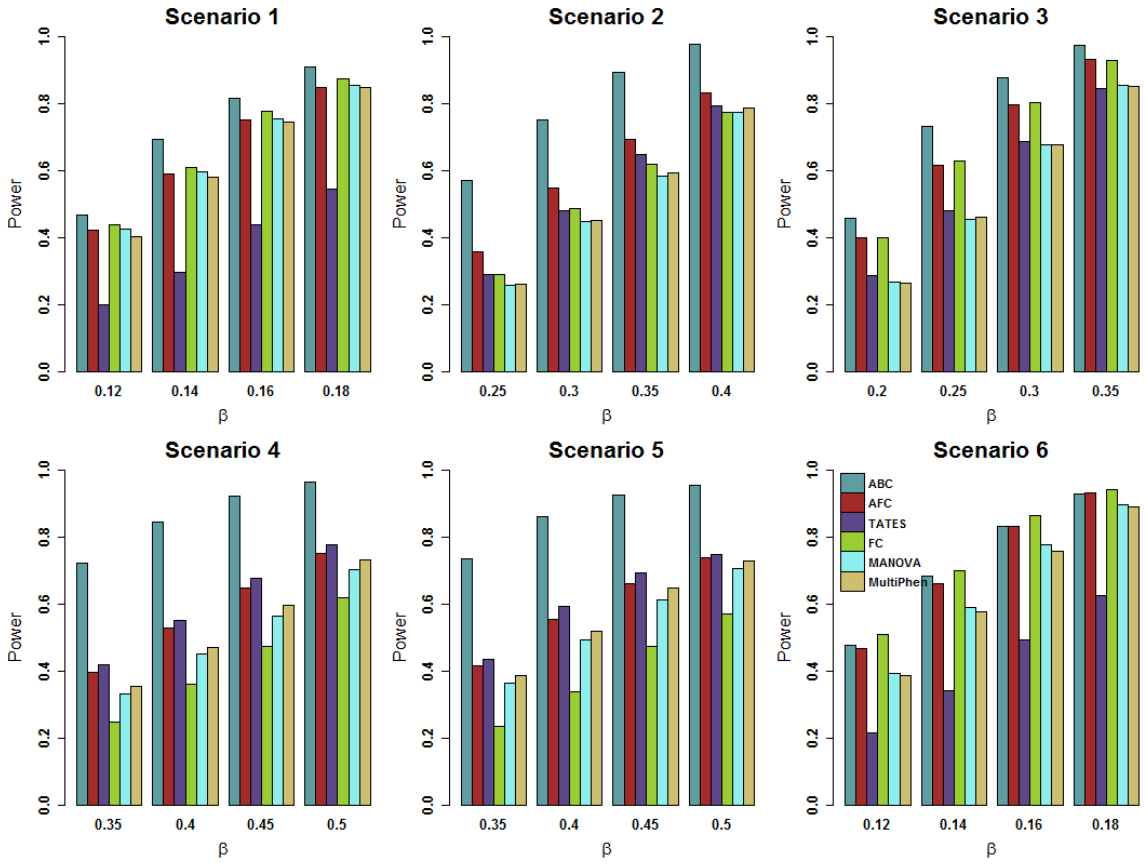


Figure 2.3. Power comparisons of the 6 tests for power as a function of effect size (β) for Log-norm $(0, 1)$ distribution under the 6 scenarios. MAF is 0.3. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The powers are evaluated at 1% significance level.

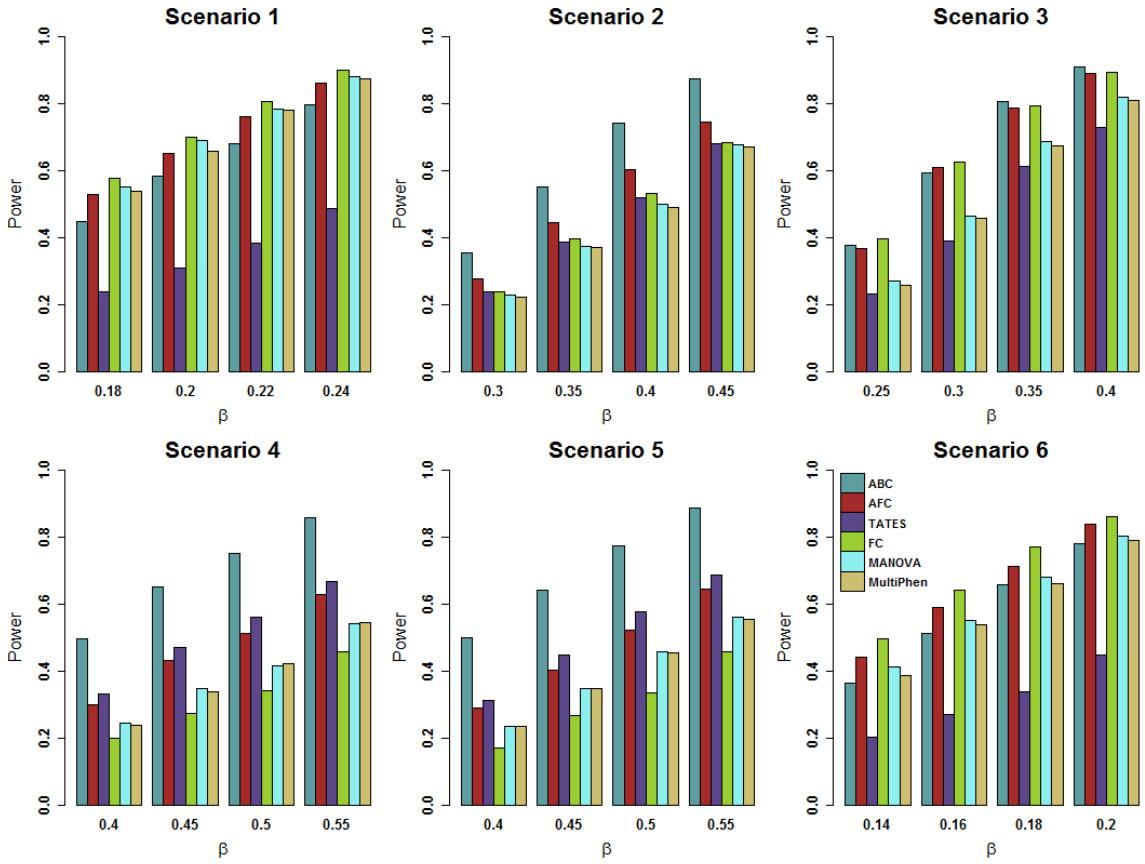


Figure 2.4. Power comparisons of the 6 tests for power as a function of effect size (β) for $\sqrt{5}\log\text{-norm}(0,1)$ distribution under the 6 scenarios. MAF is 0.3. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The powers are evaluated at 1% significance level.

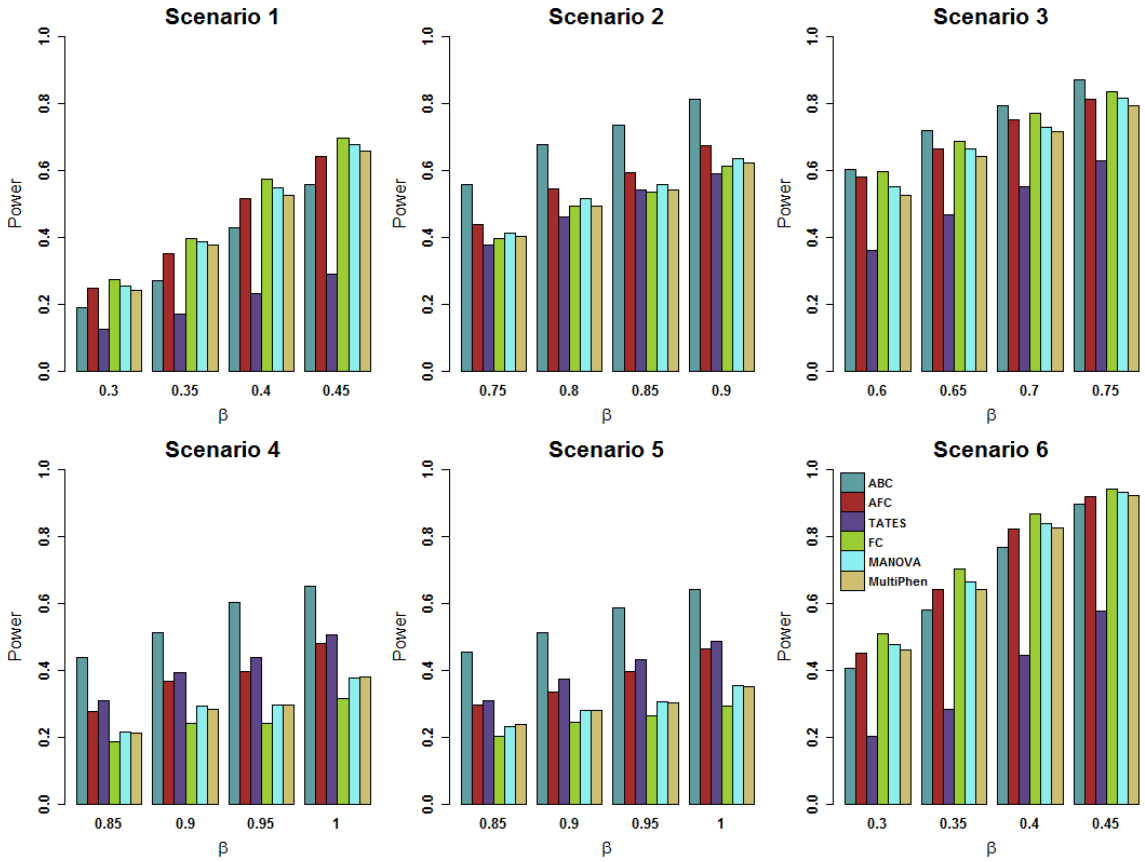
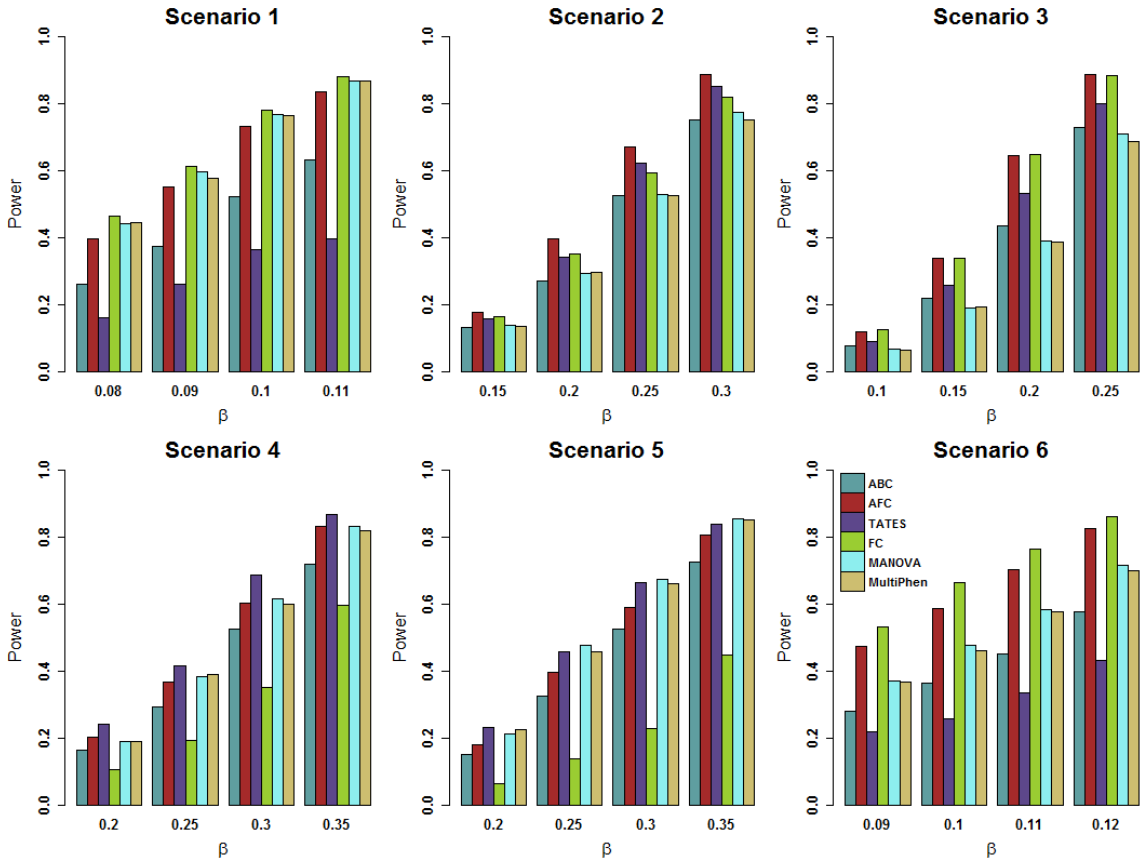


Figure 2.5. Power comparisons of the 6 tests for power as a function of effect size (β) for Normal $(0, 1)$ distribution under the 6 scenarios. MAF is 0.3. The total number of phenotypes (M) is 8. The sample size (N) is 1,000. The factor loading (a) is 0.75. The factorial correlation (ρ_{fa}) conditional on genotypes is 0.1. The powers are evaluated at 1% significance level.



3 A Hierarchical Clustering Method for Dimension Reduction in Joint Analysis of Multiple Phenotypes

Genome-wide association studies (GWAS) have become a very effective research tool to identify genetic variants of underlying various complex diseases. In spite of the success of GWAS in identifying thousands of reproducible associations between genetic variants and complex disease, in general, the association between genetic variants and a single phenotype is usually weak. It is increasingly recognized that joint analysis of multiple phenotypes can be potentially more powerful than the univariate analysis, and can shed new light on underlying biological mechanisms of complex diseases. In this study, we develop a novel variable reduction method using hierarchical clustering method (HCM) for joint analysis of multiple phenotypes in association studies. The proposed method involves two steps. The first step applies a dimension reduction technique by using a representative phenotype for each cluster of phenotypes. Then, existing methods are used in the second step to test the association between genetic variants and the representative phenotypes rather than the individual phenotypes. We perform extensive simulation studies to compare the powers of MANOVA, MultiPhen, and TATES using HCM with those of without using HCM. Our simulation studies show that using HCM is more powerful than without using HCM in most scenarios. We also illustrate the usefulness of using HCM by analyzing a whole-genome genotyping data from a lung function study.

3.1 Background

The successful applications of genome-wide association studies (GWAS) to numerous complex diseases established a large number of genetic associations (Lutz et al., 2017). Through GWAS, numerous genes have been shown to affect multiple phenotypes and yet the effect size on each phenotype is small for complex diseases (Yang et al., 2017). For example, multiple GWAS have found significant signals in the chromosome 15q25 region associated with lung cancer (Chen et al., 2015), chronic obstructive lung disease (COPD) (Cho et al., 2014), emphysema (Cho et al., 2015), and cigarette smoking

(Hancock et al., 2015).

Simultaneous testing of multiple phenotypes has been widely recognized as a valuable approach complementary to single phenotype tests. There are two main reasons: one is to increase statistical power, and the other is to shed light on underlying biology to possibly repurpose the use of existing drugs (Deng & Pan, 2017). Therefore, there is an increasing interest in joint analysis of multiple phenotypes with many new tests being recently proposed (Aschard et al., 2014; Casale et al., 2015; Cole et al., 1994; Galesloot et al., 2014; Klei et al., 2008; Korte et al., 2012; Lange et al., 2004; Liang et al., 2016; Marchini et al., 2007; O'Reilly et al., 2012; Tang & Ferreira, 2012; Wang et al., 2016; Yan et al., 2013; Zhang et al., 2014; Zhou & Stephens, 2014; Zhou et al., 2015; Zhu et al., 2015a).

Existing methods for joint analysis of multiple phenotypes roughly fall into three categories: regression methods, combining test statistics from univariate analyses, and variable reduction methods (Yang & Wang, 2012). In the first category, regression methods, there are three different approaches for analyzing the association of a genetic variant with multiple phenotypes: mixed effects models (Bates & DebRoy, 2004; Yan et al., 2013), generalized estimating equations (Liang & Zeger, 1986), and frailty models (Therneau et al., 2003). Tests that fall into the second category, combining test statistics from univariate analyses, conduct a univariate analysis first and then aggregate univariate test statistics. This approach is simple and feasible for meta-analyses (Schaid et al., 2016; Yang et al., 2016). Recently, many methods of combining test statistics from univariate analyses have been developed to explore the genetic association with multiple phenotypes by considering the correlation structure among phenotypes (Kwak & Pan, 2016; Liang et al., 2016; Van der Sluis et al., 2013; Yang et al., 2016). In the last category, tests based on variable reduction methods are roughly depending on three dimension reduction techniques. The first one is the principal component analysis of phenotypes (PCP) (Aschard et al., 2014). In PCP, the first few principal components (PCs) explaining most of the total phenotype variance are tested for association with a genetic variant, and the remaining components are not analyzed. However, Aschard et al.

(2014) showed that considering only the first few PCs often causes low power, whereas considering all PCs can improve the power. The second one is the canonical correlation analysis (CCA) (Tang & Ferreira, 2012). CCA searches for the linear combinations that maximize the correlation between two sets of multidimensional variables. It provides an efficient and powerful approach for both univariate and multivariate tests of association without the need for the permutation test. The last one is the principal component of heritability (PCH) (Klei et al., 2008; Ott & Rabinowitz, 1999; Wang et al., 2016a). PCH reduces multiple phenotypes to a linear combination of phenotypes that has the highest heritability among all linear combinations of phenotypes.

In this article, we develop a novel variable reduction method called hierarchical clustering method (HCM) for joint analysis of multiple phenotypes. HCM is a dimension reduction technique by using a representative phenotype for each cluster of phenotypes, then using existing methods for joint analysis of multiple phenotypes to test the association between a genetic variant of interest and the representative phenotypes rather than the individual phenotypes. One way to understand the dimension reduction technique of HCM is that when one cluster consists of highly positively correlated phenotypes, any linear combination of the phenotypes within this cluster can represent the cluster reasonably well (Bühlmann et al., 2013; Shah & Samworth, 2013). HCM does not require phenotypes themselves, it only requires a dissimilarity matrix of the phenotypes. This dissimilarity matrix can be estimated from the values of summary statistics using all independent single-nucleotide polymorphisms (SNPs) in a GWAS (Zhu et al., 2015b). We use extensive simulation studies to show the validity of the proposed two-step method and to investigate the power. In particular, the performance of three existing methods using HCM, multivariate analysis of variance (MANOVA) (Cole et al., 1994), joint model of multiple phenotypes (MultiPhen) (O'Reilly et al., 2012), and trait-based association test that uses extended simes procedure (TATES) (Van der Sluis et al., 2013), is compared with that of without using HCM. Our simulation studies show that MANOVA, MultiPhen, and TATES using HCM have correct type I error rates and are more powerful than MANOVA, MultiPhen, and TATES without using HCM in most

scenarios. We also apply MANOVA, MultiPhen, and TATES with and without using HCM to COPDGene data to further demonstrate the usefulness of HCM.

3.2 Methods

3.2.1 Hierarchical Clustering Method for Joint Analysis of Multiple Phenotypes

Consider a sample with n unrelated individuals. Each individual has K phenotypes. Denote $Y_k = (y_{1k}, \dots, y_{nk})^T$ as the k^{th} phenotype of n individuals and $\boldsymbol{Y} = (Y_1, \dots, Y_K)$ as the $n \times K$ phenotype matrix. Denote $X = (x_1, \dots, x_n)^T$ as the genotypic score of n individuals at a genetic variant of interest, where $x_i \in \{0, 1, 2\}$ is the number of minor alleles that the i^{th} individual carries at the genetic variant.

The proposed hierarchical clustering method (HCM) involves two steps. In the first step, we divide the K phenotypes into M clusters and use a representative phenotype for each of the M clusters. In the second step, we apply existing methods to the M representative phenotypes rather than directly to the individual phenotypes to test the association between phenotypes and the variant. In the first step, we need to find a partition $\mathcal{G}$ that partitions K phenotypes into M disjoint clusters $G_1, \dots, G_M$, where $\mathcal{G} = \{G_1, \dots, G_M\}$ with $\bigcup_{m=1}^{M} G_m = \{1, \dots, K\}$ and $G_m \cap G_\ell = \emptyset$ $(m \neq \ell)$. In this article, we use a hierarchical clustering strategy to cluster the phenotypes.

Strategies for hierarchical clustering generally fall into two types: agglomerative (bottom-up) and divisive (top-down). The agglomerative method starts with all phenotypes in their own cluster and merges the two clusters that have the smallest dissimilarity in each clustering iteration until there is only one single cluster left. The divisive method starts with all phenotypes in one cluster and splits the cluster into two that have the largest dissimilarity in each clustering iteration until all phenotypes are in their own cluster. Both methods can be described by a dendrogram which is frequently

used to illustrate the arrangement of the clusters produced by hierarchical clustering. We need a stopping criterion to cut the dendrogram into several clusters.

In this study, we use the bottom-up hierarchical clustering method based on the dissimilarity matrix of the phenotypes. We define the dissimilarity matrix D with entries $d_{ij} = 1 - P_{ij}^{s}$, where P_{ij}^{s} is the ij^{th} entry of $P^{s}(\boldsymbol{Y})$ and $P^{s}(\boldsymbol{Y})$ is the sample correlation matrix of $\boldsymbol{Y} = (Y_1, \dots, Y_K)$. We choose the average linkage as the dissimilarity between two clusters. Hence, the dissimilarity between clusters G_m and G_ℓ is given by

$$h(G_m, G_\ell) = \frac{1}{|G_m| \cdot |G_\ell|} \sum_{i \in G_m, j \in G_\ell} d_{ij} \tag{1}$$

where $|G_m|$ denote the number of phenotypes in G_m. Using the bottom-up hierarchical clustering method, we start with each phenotype as a singleton cluster and then successively merge pairs of clusters with the smallest dissimilarity calculated by equation (1) until all clusters have been merged into a single cluster that contains all phenotypes. We refer the smallest dissimilarity in each iteration as the height of the merged cluster in the dendrogram. We determine the number of clusters in the HCM using a stopping criterion. The stopping criterion is similar to an established principle (Bühlmann et al., 2013). Let h_b denote the smallest dissimilarity between two clusters in iteration b ($b \geq 1$) or the height of iteration b. We define:

$$\hat{b} = \arg\max_{b \geq 1}(h_{b+1} - h_b) \tag{2}$$

Then, we choose the number of clusters identified at the iteration $\hat{b}$.

Before we define the representative phenotype for each cluster, we first scale each phenotype. We define the representative phenotype for the m^{th} cluster as the average phenotype values in the cluster, that is

$$\bar{Y}^{(m)} = \frac{1}{|G_m|} \sum_{k \in G_m} Y_k \,, m = 1, \dots, M \tag{3}$$

Let $\overline{\boldsymbol{Y}}$ denote the $n \times M$ design matrix whose m^{th} column is given by $\bar{Y}^{(m)}$. Then we apply existing methods to test the association between $\overline{\boldsymbol{Y}}$ and X.

3.2.2 Comparison of Methods

We compare the performance of MANOVA, MultiPhen, and TATES with using HCM with that of without using HCM. We refer the ones with using HCM as HCMANOVA, HCMultiPhen, and HCTATES, respectively. Since principal component analysis (PCA) is a popular dimension reduction method, we also compare the performance of MANOVA, MultiPhen, and TATES using HCM with that of using the first few PCs of the phenotypes. We choose the number of PCs that explain 95% of the total variance of the phenotypes. We refer MANOVA, MultiPhen, and TATES using PCs as PCMANOVA, PCMultiPhen, and PCTATES, respectively.

3.3 Results

3.3.1 Simulation Studies

To evaluate the type I error rates and powers of HCM, we generate genotypes at a genetic variant according to the minor allele frequency (MAF) under Hardy Weinberg equilibrium. Then, we generate K phenotypes by the factor model (Wang et al., 2016a)

$$y = \lambda x + c\gamma f + \sqrt{1 - c^2} \times \varepsilon \qquad (4)$$

where $y = (y_1, \dots, y_K)^T$; x is the genotype score at the variant; $\lambda = (\lambda_1, \dots, \lambda_K)^T$ is the vector of effect sizes of the genetic variant on the K phenotypes; f is a vector of factors, $f = (f_1, \dots, f_R)^T \sim MVN(0, \Sigma)$, $\Sigma = (1 - \rho)I + \rho A$, A is a matrix with elements of 1, I is the identity matrix, R is the number of factors, and ρ is the correlation between factors; γ is a K by R matrix; c is a constant number; and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)^T$ is a vector of residuals,

and $\varepsilon_1, \dots, \varepsilon_K$ are independent, and $\varepsilon_k \sim N(0,1)$ for $k = 1, \dots, K$. Based on equation (4), we consider the following four models with different number of factors and different number of factors affected by genotypes. In the four models, the within-factor correlation is c^2 and the between-factor correlation is ρc^2.

Model 1: There is only one factor and genotypes impact on all phenotypes. That is, $R = 1$, $\lambda = \beta(1,2, \dots, K)^T$, and $\gamma = (1, \dots, 1)^T$.

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

Model 3: There are five factors and genotypes impact on two factors. That is, $R = 5$, $\lambda = (\beta_{11}, \dots, \beta_{1k}, \beta_{21}, \dots, \beta_{2k}, \beta_{31}, \dots, \beta_{3k}, \beta_{41}, \dots, \beta_{4k}, \beta_{51}, \dots, \beta_{5k})^T$, and $\gamma = diag(D_1, D_2, D_3, D_4, D_5)$, where $D_i = \left(\underbrace{1, \dots, 1}_{K/5}\right)^T$ for $i = 1, \dots, 5$, $k = \frac{K}{5}$, $\beta_{11} = \cdots \beta_{1k} = \beta_{21} = \cdots = \beta_{2k} = \beta_{31} = \cdots = \beta_{3k} = 0$, $\beta_{41} = \cdots = \beta_{4k} = -\beta$, and $(\beta_{51}, \dots, \beta_{5k}) = \frac{2\beta}{k+1}(1, \dots, k)$.

Model 4: There are five factors and genotypes impact on four factors. That is, $R = 5$, $\lambda = (\beta_{11}, \dots, \beta_{1k}, \beta_{21}, \dots, \beta_{2k}, \beta_{31}, \dots, \beta_{3k}, \beta_{41}, \dots, \beta_{4k}, \beta_{51}, \dots, \beta_{5k})^T$, and $\gamma = diag(D_1, D_2, D_3, D_4, D_5)$, where $D_i = \left(\underbrace{1, \dots, 1}_{K/5}\right)^T$ for $i = 1, \dots, 5$, $k = \frac{K}{5}$, $\beta_{11} = \cdots \beta_{1k} = 0$, $\beta_{21} = \cdots = \beta_{2k} = \beta$, $\beta_{31} = \cdots = \beta_{3k} = -\beta$, $\beta_{41} = \cdots = \beta_{4k} = -\frac{2\beta}{k+1}(1, \dots, k)$, and $(\beta_{51}, \dots, \beta_{5k}) = \frac{2\beta}{k+1}(1, \dots, k)$.

To evaluate type I error rates, 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 MAF = 0.3, $c = \sqrt{0.5}$, and $\rho = 0.2$.

3.3.2 Simulation Results

For each model, we estimate the p-values of all test statistics using their asymptotic distributions.

For type I error evaluation, we consider different numbers of phenotypes, different significance levels, different sample sizes, and different models. For 10,000 replicated samples, the 95% confidence intervals (CIs) for type I error rates at the nominal levels 0.05, 0.01, and 0.001 are (0.0457, 0.0543), (0.008, 0.012), and (0.0004, 0.0016), respectively. The estimated type I error rates of HCMANOVA, HCMultiPhen, and HCTATES are summarized in Tables 3.1 to 3.3. The estimated type I error rates of PCMANOVA, PCMultiPhen, and PCTATES are summarized in Tables S1 to S3. From these tables, we can see that most of the estimated type I error rates are within the 95% CIs. In addition, the type I error rates outside of the 95% CIs are very close to the bounds of the corresponding 95% CI, which indicates that HCMANOVA, HCMultiPhen, HCTATES, PCMANOVA, PCMultiPhen, and PCTATES are valid tests.

For power comparisons, we consider different numbers of phenotypes and different models. The powers of all tests are evaluated based on 1000 replications and 5000 subjects at 5% significance level. Figure 3.1 and Figure 3.2 provide the power comparisons of the six tests (HCMANOVA, MANOVA, HCMultiPhen, MultiPhen, HCTATES, and TATES) for the power as a function of the effect size under the four models. We consider 20 phenotypes and 40 phenotypes in Figure 3.1 and Figure 3.2, respectively.

These two figures show that (1) when the effect sizes of the genetic variant on phenotypes show no groups (Model 1), HCMANOVA, HCMultiPhen, and HCTATES are slightly less powerful than MANOVA, MultiPhen, and TATES, respectively, because in most replications, HCM clusters each phenotype in a singleton cluster; (2) when the effect sizes show some groups and have the same direction (Model 2), HCMANOVA,

HCMultiPhen, and HCTATES are much more powerful than MANOVA, MultiPhen, and TATES, respectively; (3) when the effect sizes show some groups and have different directions (Models 3 and 4), HCMANOVA and HCMultiPhen are more powerful than MANOVA and MultiPhen, respectively, but HCTATES is less powerful than TATES; (4) HCMANOVA and HCMultiPhen have similar power; MANOVA and MultiPhen have similar power; (5) HCTATES and TATES are much less powerful than other methods when genotypes directly impact on all phenotypes (Model 1). Figures A.1 and A.2 provide the power comparisons of HCM with those of using PCs of phenotypes that explain 95% of the total variance. These figures show that using HCM as a dimension reduction method is more powerful than using PCs that explain 95% of the total variance. We also set up an additional simulation model (Model S1) to compare the powers of MANOVA, MultiPhen, and TATES with those of HCMANOVA, HCMultiPhen, and HCTATES (Figure A.3). Figure A.3 shows that under Model S1, HCMANOVA, HCMultiPhen, and HCTATES are more powerful than MANOVA, MultiPhen, and TATES, respectively.

In summary, the existing methods using HCM have correct type I error rates and are more powerful than or comparable with those without using HCM, and the existing methods using HCM are also more powerful than those using PCs of phenotypes as a dimension reduction method.

3.3.3 Real Data Analysis

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory disease including chronic bronchitis, emphysema, non-reversible asthma, and some forms of bronchiectasis. This disease is characterized by reduced maximum expiratory flow and slow forced emptying of the lungs (Siafakas et al., 1995). Despite being a treatable and preventable disease, it is still a major cause of morbidity and mortality. The prevalence continues to rise because of the worldwide epidemic of smoking (Nazir & Erbland,

2009). In this article, we demonstrated the application of the proposed method by conducting analysis on the data from the Genetic Epidemiology of COPD (COPDGene). The COPDGene Study was designed to investigate the underlying genetic factors of 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., 2011).

To evaluate the performance of our proposed method on a real data set, we applied the six methods to the COPDGene of non-Hispanic Whites population to carry out GWAS of COPD-related phenotypes. Similar to Liang et al., 2016, we selected 7 quantitative COPD-related phenotypes, including % predicted FEV1 (FEV1), Emphysema (Emph), Emphysema Distribution (EmphDist), Gas Trapping (GasTrap), Airway Wall Area (Pi10), Exacerbation Frequency (ExacerFreq), Six-minute Walk Distance (6MWD), and 4 covariates, including BMI, Age, Pack-Years (PackYear) and Sex. In our analysis, EmphDist is the ratio of emphysema at -950 HU in the upper 1/3 of lung fields compared to the lower 1/3 of lung fields. Followed by Chu et al. (Chu et al., 2014), we did a log transformation on EmphDist in the following analysis. We excluded participants with missing data in any of the 11 variables. There were total 5,430 individuals across 630,860 SNPs used in the analyses. We first adjusted each of the 7 phenotypes for the 4 covariates using linear models. Then, we performed the analysis based on the adjusted phenotypes. The detailed information can be found in Liang et al. (Liang et al., 2016).

Based on the correlation structure of the 7 COPD-related phenotypes given in Figure 4 in Liang et al. 2016, we changed the signs for phenotypes of FEV1 and 6MWD because the correlations of FEV1 and 6MWD with other 5 phenotypes are all negative. After changing the signs for the phenotypes of FEV1 and 6MWD, the pair-wise correlations among the 7 phenotypes are all positive.

To identify SNPs associated with the 7 COPD-related phenotypes, we adopted the commonly used genome-wide significance level 5×10^{-8} to account for multiple testing.

HCM divided the 7 phenotypes into 5 clusters (Figure 3.3). The first cluster contains three phenotypes including FEV1, Emph, and GasTrap. Each of the other four clusters contains only one phenotype. Table 3.4 summarized the significant SNPs identified by at least one method. There are 14 SNPs in Table 3.4. All of the 14 SNPs had previously been reported to be in association with COPD (Brehm et al., 2011; Cho et al., 2010 & 2014; Cui et al., 2014; Du et al., 2016; Hancock et al., 2010; Li et al., 2011; Lutz et al., 2015; Pillai et al., 2009; Wilk et al., 2009 & 2012; Young et al., 2010; Zhang et al., 2012; Zhu et al., 2014). From Table 3.4, we can see that HCMANOVA identifies 13 SNPs which are same as MANOVA; HCMultiPhen identifies 13 SNPs which are one less than MultiPhen; and HCTATES identifies 10 SNPs which are one more than TATES. The results of the real data analysis are consistent with our simulation results, that is, the existing methods using HCM are more powerful than or comparable with those without using HCM.

3.4 Discussion

In this study, we developed a HCM for joint analysis of multiple phenotypes in association studies. The proposed method is a dimension reduction technique by using a representative phenotype for each cluster of phenotypes. Applying HCM, we used existing methods to test the association between genetic variants and the representative phenotypes rather than the individual phenotypes.

HCM has several important advantages over other dimension reduction techniques. First, it can produce a dendrogram of the phenotypes, which may provide more information on the structure of phenotypes. Second, it is computationally fast and easy to implement. Third, it has the distinct advantage that any valid measure of distance can be used in the hierarchical clustering procedure. In fact, HCM does not require phenotypes themselves, it only requires a dissimilarity matrix of phenotypes. This dissimilarity matrix of phenotypes can be estimated from the values of summary statistics

using independent SNPs in a GWAS (Zhu et al., 2015b). Last, any linear combination of the phenotypes within each cluster can represent the cluster reasonably well when the cluster consists of highly positively correlated phenotypes (Bühlmann et al., 2013; Shah & Samworth, 2013).

We used extensive simulation studies as well as real data application to compare the performance of MANOVA, MultiPhen, and TATES with using HCM with that of without using HCM. Our simulation results showed that the three methods using HCM have correct type I error rates and are more powerful than or comparable with those without using HCM under a variety of simulation scenarios. Additionally, the real data analysis results demonstrated that HCM has great potential in GWAS with multiple phenotypes such as COPD. We also compared the proposed method with a popular dimension reduction method, PCA of phenotypes. Our simulation results showed that the three methods using HCM are more powerful than those using PCs of phenotypes.

In this study, we use the average phenotype in each cluster as the representative phenotype of the cluster. We can also use the first PC of the phenotypes in each cluster as a representative phenotype of the cluster. However, our simulation studies (Figure A.4 and Figure A.5) show that using the average as the representative has very similar performance as using the first PC as the representative. As we pointed out in the introduction section, any linear combination of the phenotypes within one cluster can represent the cluster reasonably well when the cluster consists of highly positively correlated phenotypes. The proposed method is more suitable for quantitative phenotypes. After scaling the phenotypes, the proposed method can be applied to binary or mixed traits. However, the performance of this approach for applying to binary or mixed traits needs further investigation.

3.5 Tables and Figures

Table 3.1. The estimated type I error rates of HCMANOVA. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.050	2000	0.0491	0.0479	0.0491	0.0512
		5000	0.0488	0.0507	0.0482	0.0491
	0.010	2000	0.0092	0.0084	0.0102	0.0108
		5000	0.0113	0.0107	**0.0077**	0.0089
	0.001	2000	0.0015	**0.0002**	0.0014	0.0009
		5000	0.0010	0.0013	0.0006	0.0009
40	0.050	2000	0.0482	0.0473	0.0495	0.0505
		5000	0.0509	0.0487	0.0500	0.0509
	0.010	2000	0.0091	0.0080	0.0098	0.0104
		5000	0.0120	0.0104	0.0100	0.0094
	0.001	2000	0.0013	0.0005	0.0009	0.0008
		5000	0.0012	0.0009	0.0005	0.0005

Table 3.2. The estimated type I error rates of HCMultiPhen. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.050	2000	0.0519	0.0474	0.0515	0.0501
		5000	0.0482	0.0479	0.0484	0.0513
	0.010	2000	0.0100	0.0082	0.0109	0.0104
		5000	0.0112	0.0111	0.0085	0.0100
	0.001	2000	**0.0018**	0.0006	0.0011	0.0008
		5000	0.0013	0.0013	0.0008	0.0006
40	0.050	2000	0.0513	0.0464	0.0512	0.0502
		5000	0.0539	0.0484	0.0496	0.0490
	0.010	2000	0.0112	**0.0078**	0.0104	0.0091
		5000	**0.0127**	0.0111	0.0096	0.0102
	0.001	2000	0.0013	0.0007	0.0009	0.0010
		5000	0.0013	0.0012	**0.0002**	0.0004

Table 3.3. The estimated type I error rates of HCTATES. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.050	2000	**0.0445**	0.0469	0.0509	0.0491
		5000	0.0452	0.0516	0.0498	0.0478
	0.010	2000	0.0105	0.0088	0.0119	0.0091
		5000	0.0096	0.0102	0.0087	0.0088
	0.001	2000	0.0016	0.0006	0.0013	0.001
		5000	0.0007	0.0012	0.0007	0.0013
40	0.050	2000	**0.0409**	0.0464	0.0519	0.0495
		5000	**0.0413**	0.0486	0.0488	0.0499
	0.010	2000	0.0088	0.0098	0.0095	0.0096
		5000	0.0097	0.011	0.0096	0.0094
	0.001	2000	0.0008	0.0005	0.001	0.0009
		5000	0.0012	0.0009	0.0013	0.0008

Table 3.4. Significant SNPs and the corresponding p-values in the analysis of COPDGene. The p-values of six tests are evaluated using asymptotic distributions. The bold p-values indicate the p-values $> 5 \times 10^{-8}$.

Chr	Position	Variant identifier	MANOVA	HCMANOVA	MultiPhen	HCMultiPhen	TATES	HCTATES
4	145431497	rs1512282	1.69×10^{-9}	1.61×10^{-10}	1.03×10^{-9}	9.68×10^{-11}	5.77×10^{-9}	4.76×10^{-9}
4	145434744	rs1032297	6.52×10^{-14}	5.16×10^{-15}	7.69×10^{-14}	6.67×10^{-15}	6.22×10^{-13}	5.52×10^{-13}
4	145474473	rs1489759	1.11×10^{-16}	7.46×10^{-18}	1.22×10^{-16}	9.32×10^{-18}	2.52×10^{-16}	0
4	145485738	rs1980057	6.68×10^{-17}	4.52×10^{-18}	8.14×10^{-17}	6.30×10^{-18}	9.35×10^{-17}	0
4	145485915	rs7655625	7.12×10^{-17}	4.98×10^{-18}	9.13×10^{-17}	7.33×10^{-18}	1.64×10^{-16}	0
15	78882925	rs16969968	1.32×10^{-11}	1.98×10^{-10}	7.84×10^{-12}	1.46×10^{-10}	2.98×10^{-8}	2.44×10^{-8}
15	78894339	rs1051730	1.41×10^{-11}	1.51×10^{-10}	8.16×10^{-12}	1.17×10^{-10}	2.63×10^{-8}	2.15×10^{-8}
15	78898723	rs12914385	1.76×10^{-12}	4.33×10^{-12}	1.48×10^{-12}	4.84×10^{-12}	5.14×10^{-10}	4.31×10^{-10}
15	78911181	rs8040868	2.74×10^{-12}	1.74×10^{-11}	2.59×10^{-12}	2.08×10^{-11}	2.40×10^{-9}	1.99×10^{-9}
15	78878541	rs951266	1.77×10^{-11}	3.04×10^{-10}	1.02×10^{-11}	2.24×10^{-10}	$\mathbf{5.17 \times 10^{-8}}$	4.21×10^{-8}
15	78806023	rs8034191	2.14×10^{-10}	2.95×10^{-9}	7.74×10^{-11}	1.70×10^{-9}	$\mathbf{1.02 \times 10^{-7}}$	$\mathbf{8.24 \times 10^{-8}}$
15	78851615	rs2036527	3.99×10^{-10}	6.62×10^{-9}	1.77×10^{-10}	4.05×10^{-9}	$\mathbf{1.56 \times 10^{-7}}$	$\mathbf{1.26 \times 10^{-7}}$
15	78826180	rs931794	2.35×10^{-10}	1.68×10^{-8}	9.09×10^{-11}	1.19×10^{-8}	$\mathbf{1.18 \times 10^{-7}}$	$\mathbf{9.58 \times 10^{-8}}$
15	78740964	rs2568494	$\mathbf{1.05 \times 10^{-7}}$	$\mathbf{6.35 \times 10^{-6}}$	4.23×10^{-8}	$\mathbf{5.54 \times 10^{-6}}$	$\mathbf{2.88 \times 10^{-5}}$	$\mathbf{7.44 \times 10^{-5}}$

Figure 3.1. Power comparisons of the six tests (HCMANOVA, MANOVA, HCMultiPhen, MultiPhen, HCTATES, and TATES) for the power as a function of effect size β for 20 quantitative phenotypes. MAF is 0.3. The sample size is 5000. The number of replication is 1000. The within-factor correlation is 0.5 ($c^2 = 0.5$) and the between-factor correlation is 0.1 ($\rho c^2 = 0.1$). The powers are evaluated at 5% significance level.

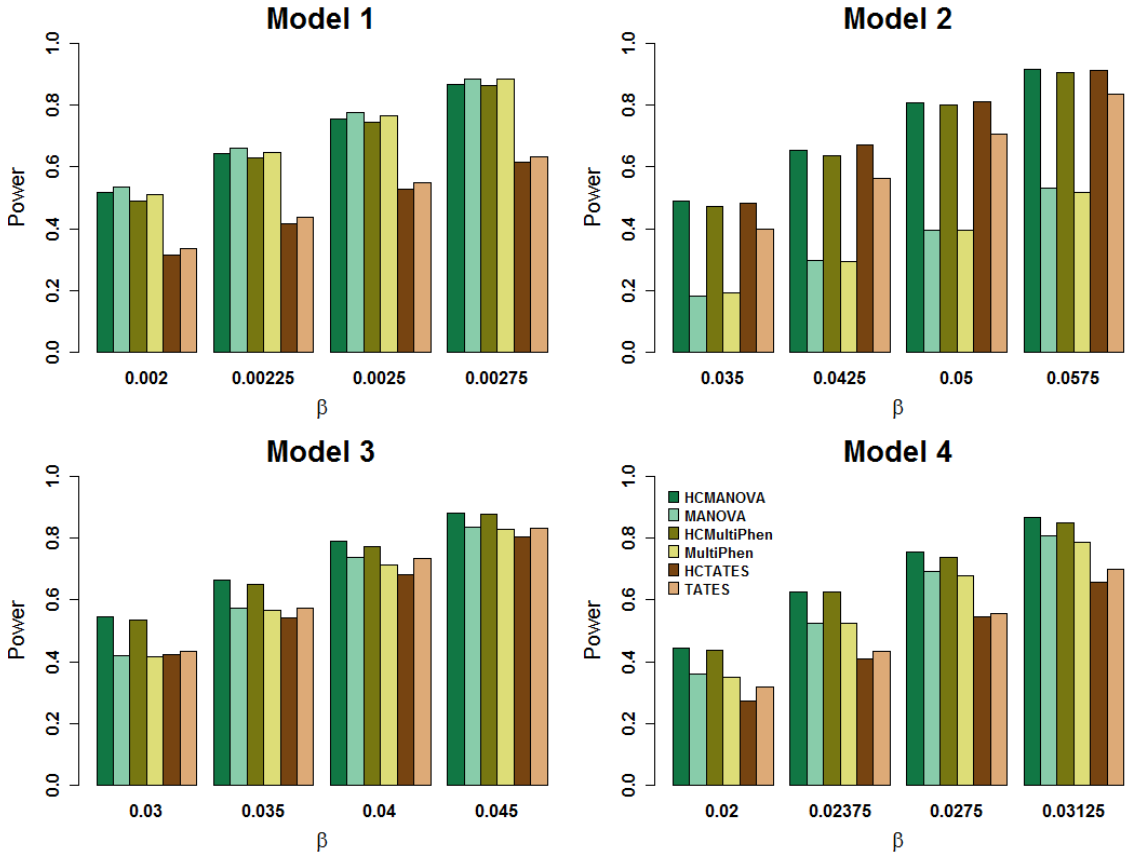


Figure 3.2. Power comparisons of the six tests (HCMANOVA, MANOVA, HCMultiPhen, MultiPhen, HCTATES, and TATES) for the power as a function of effect size β for 40 quantitative phenotypes. MAF is 0.3. The sample size is 5000. The number of replication is 1000. The within-factor correlation is 0.5 ($c^2 = 0.5$) and the between-factor correlation is 0.1 ($\rho c^2 = 0.1$). The powers are evaluated at 5% significance level.

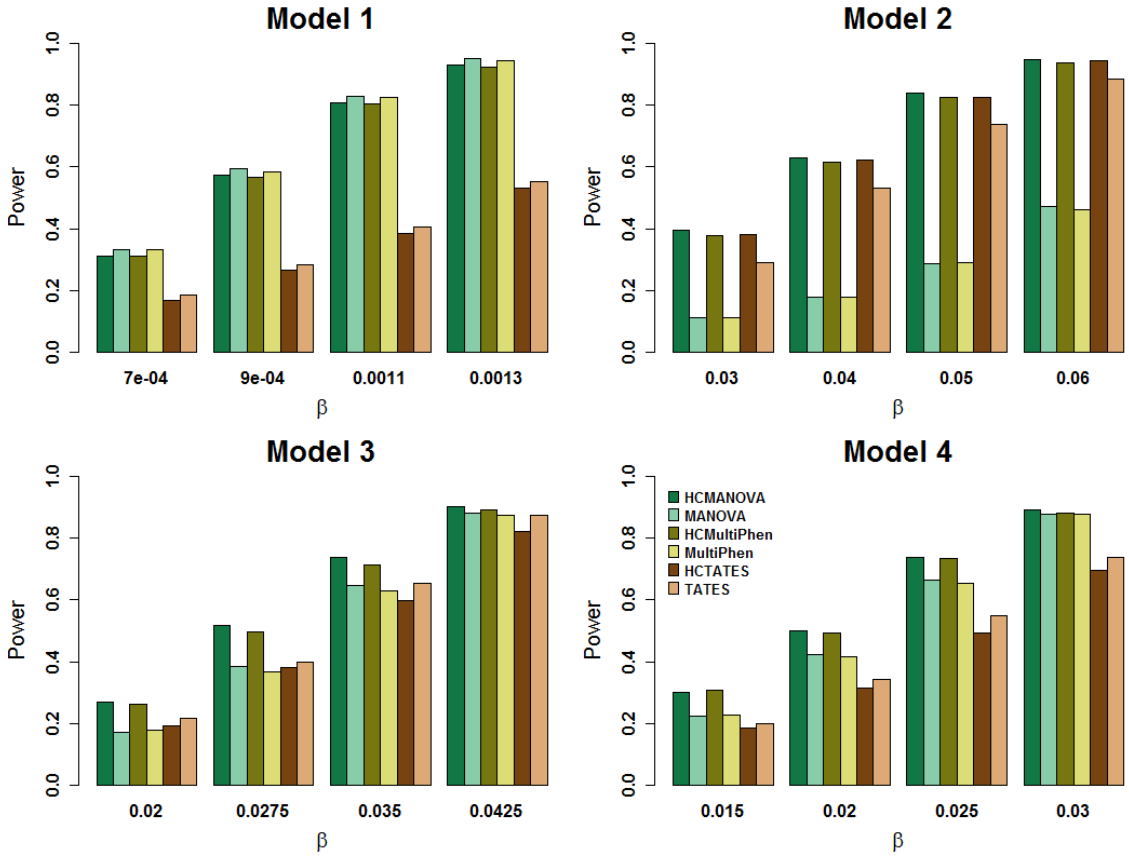
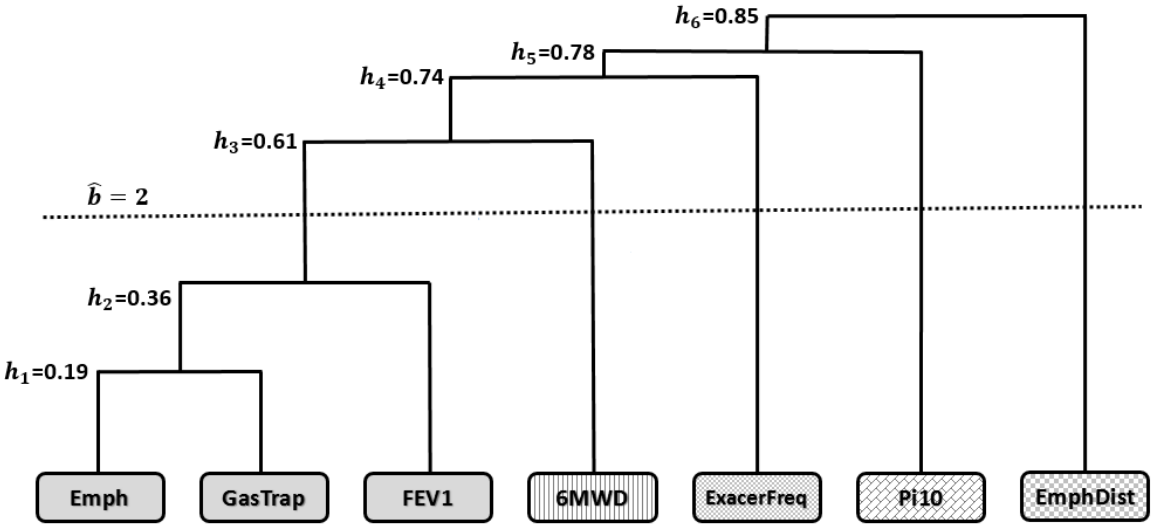


Figure 3.3. The dendrogram of the seven phenotypes in the COPDGene study.



4 Reference List

1. Amos, C. I., & Laing, A. E. (1993). A comparison of univariate and multivariate tests for genetic linkage. *Genetic Epidemiology*, 10(6), 671-676.
2. Aschard, H., Vilhjálmsson, B. J., Greliche, N., Morange, P. E., Trégouët, D. A., & Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *The American Journal of Human Genetics*, 94(5), 662-676.
3. Aulchenko, Y. S., Ripke, S., Isaacs, A., & Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23(10), 1294-1296.
4. Aulchenko, Y. S., Struchalin, M. V., & van Duijn, C. M. (2010). ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, 11(1), 134.
5. Bates, D. M., & DebRoy, S. (2004). Linear mixed models and penalized least squares. *Journal of Multivariate Analysis*, 91(1), 1-17.
6. Brehm, J. M., Hagiwara, K., Tesfaigzi, Y., Bruse, S., Mariani, T. J., Bhattacharya, S., ... & Cho, M. H. (2011). Identification of FGF7 as a novel susceptibility locus for chronic obstructive pulmonary disease. *Thorax*, 66(12), 1085-1090.
7. Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421), 9-25.
8. Bühlmann, P., Rütimann, P., van de Geer, S., & Zhang, C. H. (2013). Correlated variables in regression: clustering and sparse estimation. *Journal of Statistical Planning and Inference*, 143, 1835-1858.
9. Casaburi, R., Mahler, D. A., Jones, P. W., Wanner, A., San Pedro, G., ZuWallack, R. L., ... & Witek, T. (2002). A long-term evaluation of once-daily inhaled tiotropium in chronic obstructive pulmonary disease. *European Respiratory Journal*, 19(2), 217-224.
10. Casale, F. P., Rakitsch, B., Lippert, C., & Stegle, O. (2015). Efficient set tests for the genetic analysis of correlated traits. *Nature Methods*, 12(8), 755.

11. Chen, L. S., Hung, R. J., Baker, T., Horton, A., Culverhouse, R., Saccone, N., ... & Horsman, J. (2015). CHRNA5 risk variant predicts delayed smoking cessation and earlier lung cancer diagnosis—a meta-analysis. *Journal of the National Cancer Institute*, 107, djv100.

12. Cho, M. H., Boutaoui, N., Klanderman, B. J., Sylvia, J. S., Ziniti, J. P., Hersh, C. P., ... & Lange, C. (2010). Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature Genetics*, 42(3), 200.
13. Cho, M. H., McDonald, M. L. N., Zhou, X., Mattheisen, M., Castaldi, P. J., Hersh, C. P., ... & Lange, C. (2014). Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The Lancet Respiratory Medicine*, 2(3), 214-225.

14. Cho, M. H., Castaldi, P. J., Hersh, C. P., Hobbs, B. D., Barr, R. G., Tal-Singer, R., ... & Coxson, H. O. (2015). A genome-wide association study of emphysema and airway quantitative imaging phenotypes. *American Journal of Respiratory and Critical Care Medicine*, 192, 559-569.

15. Chu, J. H., Hersh, C. P., Castaldi, P. J., Cho, M. H., Raby, B. A., Laird, N., ... & Silverman, E. K. (2014). Analyzing networks of phenotypes in complex diseases: methodology and applications in COPD. *BMC Systems Biology*, 8(1), 78.
16. Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1994). How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin*, 115(3), 465.
17. Cui, K., Ge, X., & Ma, H. (2014). Four SNPs in the CHRNA3/5 alpha-neuronal nicotinic acetylcholine receptor subunit locus are associated with COPD risk based on meta-analyses. *PloS One*, 9(7), e102324.
18. Deng, Y., & Pan, W. (2017). Conditional analysis of multiple quantitative traits based on marginal GWAS summary statistics. *Genetic Epidemiology*, 41(5), 427-436.

19. Du, Y., Xue, Y., & Xiao, W. (2016). Association of IREB2 gene rs2568494 polymorphism with risk of chronic obstructive pulmonary disease: a meta-analysis. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 22, 177.
20. Ferreira, M. A., & Purcell, S. M. (2008). A multivariate test of association. *Bioinformatics*, 25(1), 132-133.
21. Fitzmaurice, G. M., & Laird, N. M. (1993). A likelihood-based method for analyzing longitudinal binary responses. *Biometrika*, 80(1), 141-151.
22. Galesloot, T. E., Van Steen, K., Kiemeney, L. A., Janss, L. L., & Vermeulen, S. H. (2014). A comparison of multivariate genome-wide association methods. *PloS One*, 9(4), e95923.
23. Germain, M., Saut, N., Greliche, N., Dina, C., Lambert, J. C., Perret, C., ... & Zelenika, D. (2011). Genetics of venous thrombosis: insights from a new genome wide association study. *PloS One*, 6(9), e25581.
24. Hancock, D. B., Eijgelsheim, M., Wilk, J. B., Gharib, S. A., Loehr, L. R., Marciante, K. D., ... & Schabath, M. B. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature Genetics*, 42(1), 45.
25. Hancock, D. B., Reginsson, G. W., Gaddis, N. C., Chen, X., Saccone, N. L., Lutz, S. M., ... & Stacey, S. N. (2015). Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Translational Psychiatry*, 5, e651.
26. Han, M. K., Kazerooni, E. A., Lynch, D. A., Liu, L. X., Murray, S., Curtis, J. L., ... & Anzueto, A. R. (2011). Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology*, 261(1), 274-282.
27. He, Q., Avery, C. L., & Lin, D. Y. (2013). A general framework for association tests with multivariate traits in large-scale genomics studies. *Genetic Epidemiology*, 37(8), 759-767.

28. Jiang, C., & Zeng, Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, 140(3), 1111-1127.
29. Klei, L., Luca, D., Devlin, B., & Roeder, K. (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic Epidemiology*, 32(1), 9-19.
30. Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., & Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, 44(9), 1066.
31. Kwak, I. Y., & Pan, W. (2016). Gene-and pathway-based association tests for multiple traits with GWAS summary statistics. *Bioinformatics*, 33(1), 64-71.
32. Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.
33. Lange, C., DeMeo, D. L., & Laird, N. M. (2002). Power and design considerations for a general class of family-based association tests: quantitative traits. *The American Journal of Human Genetics*, 71(6), 1330-1341.
34. Lange, C., Van Steen, K., Andrew, T., Lyon, H., DeMeo, D.L., Raby, B., Murphy, A., Silverman, E.K., MacGregor, A., Weiss, S.T. & Laird, N.M. (2004) 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*, 3, 1-27.
35. Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.
36. Liang, X., Wang, Z., Sha, Q., & Zhang, S. (2016). An Adaptive Fisher's Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies. *Scientific Reports*, 6, 34323.
37. Li, M. X., Gui, H. S., Kwan, J. S., & Sham, P. C. (2011a). GATES: a rapid and powerful gene-based association test using extended Simes procedure. *The American Journal of Human Genetics*, 88(3), 283-293.

38. Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods*, 8(10), 833.
39. Li, X., Howard, T. D., Moore, W. C., Ampleford, E. J., Li, H., Busse, W. W., ... & Fitzpatrick, A. M. (2011b). Importance of hedgehog interacting protein and other lung function genes in asthma. *Journal of Allergy and Clinical Immunology*, 127(6), 1457-1465.
40. Li, Y., Willer, C., Sanna, S., & Abecasis, G. (2009). Genotype imputation. *Annual Review of Genomics and Human Genetics*, 10, 387-406.
41. Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, 34(8), 816-834.
42. Lutz, S. M., Cho, M. H., Young, K., Hersh, C. P., Castaldi, P. J., McDonald, M. L., ... & Foreman, M. (2015). A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genetics*, 16(1), 138.
43. Lutz, S. M., Fingerlin, T. E., Hokanson, J. E., & Lange, C. (2017). A general approach to testing for pleiotropy with rare and common variants. *Genetic Epidemiology*, 41, 163-170.
44. Majumdar, A., Witte, J. S., & Ghosh, S. (2015). Semiparametric allelic tests for mapping multiple phenotypes: binomial regression and Mahalanobis distance. *Genetic Epidemiology*, 39(8), 635-650.
45. Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, 39(7), 906.
46. Nazir, S. A., & Erbland, M. L. (2009). Chronic obstructive pulmonary disease. *Drugs & Aging*, 26(10), 813-831.
47. Newman III, W. P., Freedman, D. S., Voors, A. W., Gard, P. D., Srinivasan, S. R., Cresanta, J. L., ... & Berenson, G. S. (1986). Relation of serum lipoprotein levels

and systolic blood pressure to early atherosclerosis. *New England Journal of Medicine*, 314(3), 138-144.

48. O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C., Elliott, P., Jarvelin, M. R., & Coin, L. J. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PloS One*, 7(5), e34861.
49. Ott, J., & Rabinowitz, D. (1999). A principal-components approach based on heritability for combining phenotype information. *Human Heredity*, *49*(2), 106-111.
50. Pesarin, F., & Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons.
51. Pillai, S. G., Ge, D., Zhu, G., Kong, X., Shianna, K. V., Need, A. C., ... & Ruppert, A. (2009). A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genetics*, 5(3), e1000421.
52. Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904.
53. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559-575.
54. Ray, D., Pankow, J. S., & Basu, S. (2016). USAT: A Unified Score-Based Association Test for Multiple Phenotype-Genotype Analysis. *Genetic Epidemiology*, 40(1), 20-34.
55. Regan, E. A., Hokanson, J. E., Murphy, J. R., Make, B., Lynch, D. A., Beaty, T. H., ... & Crapo, J. D. (2011). Genetic epidemiology of COPD (COPDGene) study design. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 7(1), 32-43.
56. Rencher, A. C. (2003). *Methods of multivariate analysis* (Vol. 492). John Wiley & Sons.

57. Schaid, D. J., Tong, X., Larrabee, B., Kennedy, R. B., Poland, G. A., & Sinnwell, J. P. (2016). Statistical methods for testing genetic pleiotropy. *Genetics*, 204, 483-497.

58. Schifano, E. D., Li, L., Christiani, D. C., & Lin, X. (2013). Genome-wide association analysis for multiple continuous secondary phenotypes. *The American Journal of Human Genetics*, 92(5), 744-759.

59. Shah, R. D., & Samworth, R. J. (2013). Discussion of 'Correlated variables in regression: clustering and sparse estimation'by Peter Bühlmann, Philipp Rütimann, Sara van de Geer and Cun-Hui Zhang. *Journal of Statistical Planning and Inference*, 143, 1866-1868.

60. Sha, Q., Wang, X., Wang, X., & Zhang, S. (2012). Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genetic Epidemiology*, 36(6), 561-571.

61. Siafakas, N. M., Vermeire, P., Pride, N. A., Paoletti, P., Gibson, J., Howard, P., ... & Postma, D. S. (1995). Optimal assessment and management of chronic obstructive pulmonary disease (COPD). The European Respiratory Society Task Force. *European Respiratory Journal*, 8, 1398-1420.

62. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., ... & Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. *The American Journal of Human Genetics*, 89(5), 607-618.

63. Souto, J. C., Almasy, L., Borrell, M., Blanco-Vaca, F., Mateo, J., Soria, J. M., ... & Blangero, J. (2000). Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. *The American Journal of Human Genetics*, 67(6), 1452-1459.

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

65. Therneau, T. M., Grambsch, P. M., & Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics*, 12, 156-175.

66. UK10K Consortium. (2015). The UK10K project identifies rare variants in health and disease. *Nature*, 526(7571), 82.

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

68. Wang, K. (2014). Testing genetic association by regressing genotype over multiple phenotypes. *PloS One*, 9(9), e106918.

69. Wang, Y., Fang, Y., & Jin, M. (2007a). A ridge penalized principal-components approach based on heritability for high-dimensional data. *Human Heredity*, 64(3), 182-191.

70. Wang, Y., Fang, Y., & Wang, S. (2007b). Clustering and principal-components approach based on heritability for mapping multiple gene expressions. In *BMC proceedings* (Vol. 1, No. 1, p. S121). BioMed Central.

71. Wang, Z., Sha, Q., & Zhang, S. (2016a). Joint analysis of multiple traits using" optimal" maximum heritability test. *PloS One*, 11, e0150975.

72. Wang, Z., Wang, X., Sha, Q., & Zhang, S. (2016b). Joint analysis of multiple traits in rare variant association studies. *Annals of Human Genetics*, 80(3), 162-171.

73. Wilk, J. B., Chen, T. H., Gottlieb, D. J., Walter, R. E., Nagle, M. W., Brandler, B. J., ... & O'Connor, G. T. (2009). A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genetics*, 5(3), e1000429.

74. Wilk, J. B., Shrine, N. R., Loehr, L. R., Zhao, J. H., Manichaikul, A., Lopez, L. M., ... & Loth, D. W. (2012). Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *American Journal of Respiratory and Critical Care Medicine*, 186(7), 622-632.

75. Yang, J. J., Li, J., Williams, L. K., & Buu, A. (2016). An efficient genome-wide association test for multivariate phenotypes based on the Fisher combination function. *BMC Bioinformatics*, 17(1), 19.

76. Yang, J. J., Williams, L. K., & Buu, A. (2017). Identifying Pleiotropic Genes in Genome-Wide Association Studies for Multivariate Phenotypes with Mixed Measurement Scales. *PloS One*, 12, e0169893.

77. Yang, Q., & Wang, Y. (2012). Methods for analyzing multivariate phenotypes in genetic association studies. *Journal of Probability and Statistics*, 2012.

78. Yang, Q., Wu, H., Guo, C. Y., & Fox, C. S. (2010). Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic Epidemiology*, 34(5), 444-454.

79. Yan, T., Li, Q., Li, Y., Li, Z., & Zheng, G. (2013). Genetic association with multiple traits in the presence of population stratification. *Genetic Epidemiology*, 37(6), 571-580.

80. Young, R. P., Whittington, C. F., Hopkins, R. J., Hay, B. A., Epton, M. J., Black, P. N., & Gamble, G. D. (2010). Chromosome 4q31 locus in COPD is also associated with lung cancer. *European Respiratory Journal*, 36(6), 1375-1382.

81. Zhang, J., Summah, H., Zhu, Y. G., & Qu, J. M. (2011). Nicotinic acetylcholine receptor variants associated with susceptibility to chronic obstructive pulmonary disease: a meta-analysis. *Respiratory Research*, 12(1), 158.

82. Zhang, Y., Xu, Z., Shen, X., Pan, W., & Alzheimer's Disease Neuroimaging Initiative. (2014). Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, 96, 309-325.

83. Zhou, J. J., Cho, M. H., Lange, C., Lutz, S., Silverman, E. K., & Laird, N. M. (2015). Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability. *Human Heredity*, 79(2), 93-104.

84. Zhou, X., & Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature Methods*, 11(4), 407.

85. Zhu, A. Z., Zhou, Q., Cox, L. S., David, S. P., Ahluwalia, J. S., Benowitz, N. L., & Tyndale, R. F. (2014). Association of CHRNA5-A3-B4 SNP rs2036527 With Smoking Cessation Therapy Response in African-American Smokers. *Clinical Pharmacology & Therapeutics*, 96(2), 256-265.

86. Zhu, H., Zhang, S., & Sha, Q. (2015a). Power comparisons of methods for joint association analysis of multiple phenotypes. *Human Heredity*, *80*, 144-152.

87. Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., ... & Chen, W. (2015b). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *The American Journal of Human Genetics*, 96(1), 21-36.

Appendix

A.1 An Adaptive Fisher's Combination Method for Joint Analysis of Multiple Phenotypes in Association Studies

Without loss of generality, we assume that all phenotypes are quantitative. We use the linear model $y_{ij} = \beta_{0j} + \beta_{1j} x_i + \varepsilon_i$ to relate the j^{th} phenotype and the genotype. Let T_j denote the score test statistic to test the null hypothesis $H_0 : \beta_{1j} = 0$. Then, T_j is given by

$$T_j = U_j \Big/ \sqrt{V_j},$$

where $U_j = \sum_{i=1}^{n} y_{ij}(x_i - \bar{x})$ and $V_j = \sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ n$. Under the null hypothesis, the statistic T_j asymptotically follows a standard normal distribution. It is reasonable to assume that $T = \left(T_1, \ldots, T_K\right)^T$ follows a multivariate normal distribution with mean 0 and covariance matrix Σ under the null hypothesis[1]. Note that $\frac{1}{n}\sum_{i=1}^{n} (y_{ij} - \bar{y}_j)^2 \to \mathrm{var}\left(y_{1j}\right) = \sigma_j^2$, where σ_j^2 is the variance of the j^{th} phenotype. We have

$$\mathrm{cov}\left(T_l, T_s\right) \to \frac{E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \left(y_{il} - E\left(y_{1l}\right)\right)\left(y_{js} - E\left(y_{1s}\right)\right)\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right) \right)}{\sigma_l \sigma_s \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

$$= \frac{\mathrm{cov}\left(y_{1l}, y_{1s}\right) \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{\sigma_l \sigma_s \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} = \frac{\mathrm{cov}\left(y_{1l}, y_{1s}\right)}{\sigma_l \sigma_s} = \rho_{ls}$$

as $n \to \infty$, where ρ_{ls} denotes the correlation coefficient between the l^{th} phenotype and the s^{th} phenotype. We can see that under null hypothesis, the distribution of

$T = (T_1, \ldots, T_K)^T$, and thus distributions of $p_1, \ldots, p_K$ and T_{all}, are independent of genotypes.

Using the same arguments as above, we can show that if $p_1, \ldots, p_K$ come from z score statistics or Wald statistics, distributions of $p_1, \ldots, p_K$ and T_{all} are also independent of the genotype (Zhu et al., 2015b).

A.2 A Hierarchical Clustering Method for Dimension Reduction in Joint Analysis of Multiple Phenotypes

We use PCMANOVA, PCMultiPhen, and PCTATES to represent MANOVA, MultiPhen, and TATES applying to the first few principal components (PCs) of phenotypes that explain 95% of the total variance, respectively; we use HCMANOVA-PC1, HCMultiPhen-PC1, and HCTATES-PC1 to represent HCMANOVA, HCMultiPhen, and HCTATES using the first PC of the phenotypes in each cluster as a representative of the cluster, respectively.

In power comparisons (Figures A.1 – A.5), we use the following set up: MAF is 0.3; the sample size is 5000; the number of replication is 1000; and the significance level is 5%. In Figures A.1, A.2, A.4, and A.5, the within-factor correlation is 0.5 ($c^2 = 0.5$) and the between-factor correlation is 0.1 ($\rho c^2 = 0.1$).

Simulation Model S1: phenotypes $y = (y_1, \ldots, y_K)^T$ are generated according to multivariate normal distribution $MVN_K(\mu, \Sigma)$ with $\mu = x\beta(1, \ldots, K)^T$ and $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = \rho^{|i-j|}$ and x is the genotype score at the variant.

Table A.1. The estimated type I error rates of PCMANOVA. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.05	2000	0.0499	0.0529	0.0494	0.0477
		5000	0.0493	0.0462	0.0495	0.0495
	0.01	2000	0.0105	0.011	0.0108	0.0092
		5000	0.0098	0.0104	0.0101	0.0101
	0.001	2000	0.0012	0.001	0.0008	0.0009
		5000	0.0008	0.001	0.0012	0.0014
40	0.05	2000	0.0467	0.0521	0.0491	0.0476
		5000	0.05	0.0498	0.0492	0.0529
	0.01	2000	0.0099	0.0112	0.0088	0.0099
		5000	0.01	0.0111	0.0097	0.011
	0.001	2000	0.0013	0.0008	0.0011	0.0009
		5000	0.0013	0.001	0.0011	0.0015

Table A.2. The estimated type I error rates of PCMultiPhen. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.05	2000	0.0522	0.0523	0.0507	0.0488
		5000	0.0509	0.0492	0.0507	0.0477
	0.01	2000	0.0111	0.0114	0.0107	0.0106
		5000	0.0103	0.0104	0.0113	0.0104
	0.001	2000	0.0013	0.0008	0.0007	0.001
		5000	0.0007	0.0009	0.0009	0.0014
40	0.05	2000	0.0541	**0.0566**	**0.0561**	0.0528
		5000	0.051	0.0537	0.0504	0.0539
	0.01	2000	0.0118	0.0112	0.0095	0.0106
		5000	0.0103	0.0115	0.0096	0.011
	0.001	2000	**0.0017**	0.0009	0.0014	0.0009
		5000	0.0012	0.0012	0.0006	0.0012

Table A.3. The estimated type I error rates of PCTATES. MAF is 0.3. K is the number of phenotypes. α is the significance level. 10,000 replicates are used in the simulations. The type I error rates in bold indicate the values out of the bounds of the 95% CIs.

K	α	Sample Size	Model			
			1	2	3	4
20	0.05	2000	0.0476	0.0487	0.0474	0.0505
		5000	0.0472	0.053	0.047	0.0485
	0.01	2000	0.0085	0.0103	0.0105	0.0098
		5000	0.0105	0.0094	0.0084	0.0099
	0.001	2000	0.0005	0.0013	0.0015	0.0008
		5000	0.0011	0.0012	0.0008	0.0009
40	0.05	2000	0.0524	0.0484	0.0491	0.0541
		5000	0.0503	0.0501	0.051	0.0492
	0.01	2000	0.0088	0.0102	0.0088	0.0108
		5000	**0.0128**	0.0096	0.0104	0.0099
	0.001	2000	0.0013	0.0012	0.0014	0.0016
		5000	0.0007	0.0012	0.0007	0.0012

Figure A.1. Power comparisons of the six tests (HCMANOVA, PCMANOVA, HCMultiPhen, PCMultiPhen, HCTATES, and PCTATES) for the power as a function of effect size β for 20 quantitative phenotypes.

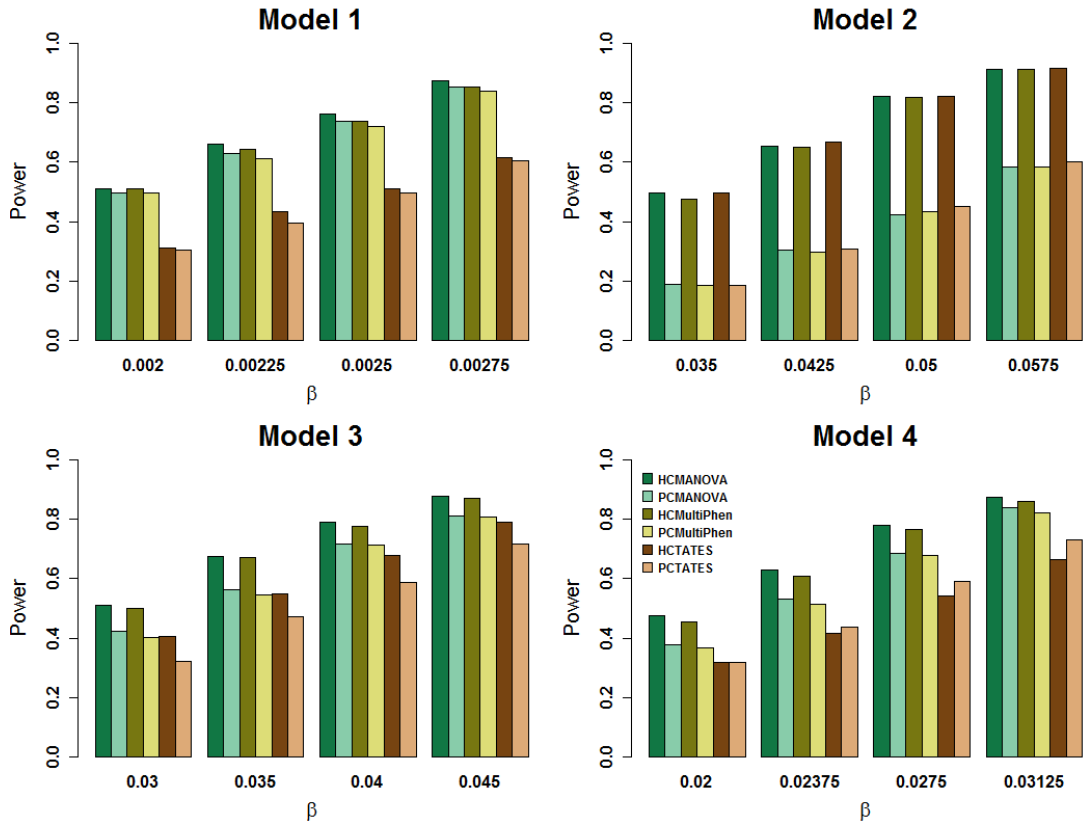


Figure A.2. Power comparisons of the six tests (HCMANOVA, PCMANOVA, HCMultiPhen, PCMultiPhen, HCTATES, and PCTATES) for the power as a function of effect size β for 40 quantitative phenotypes.

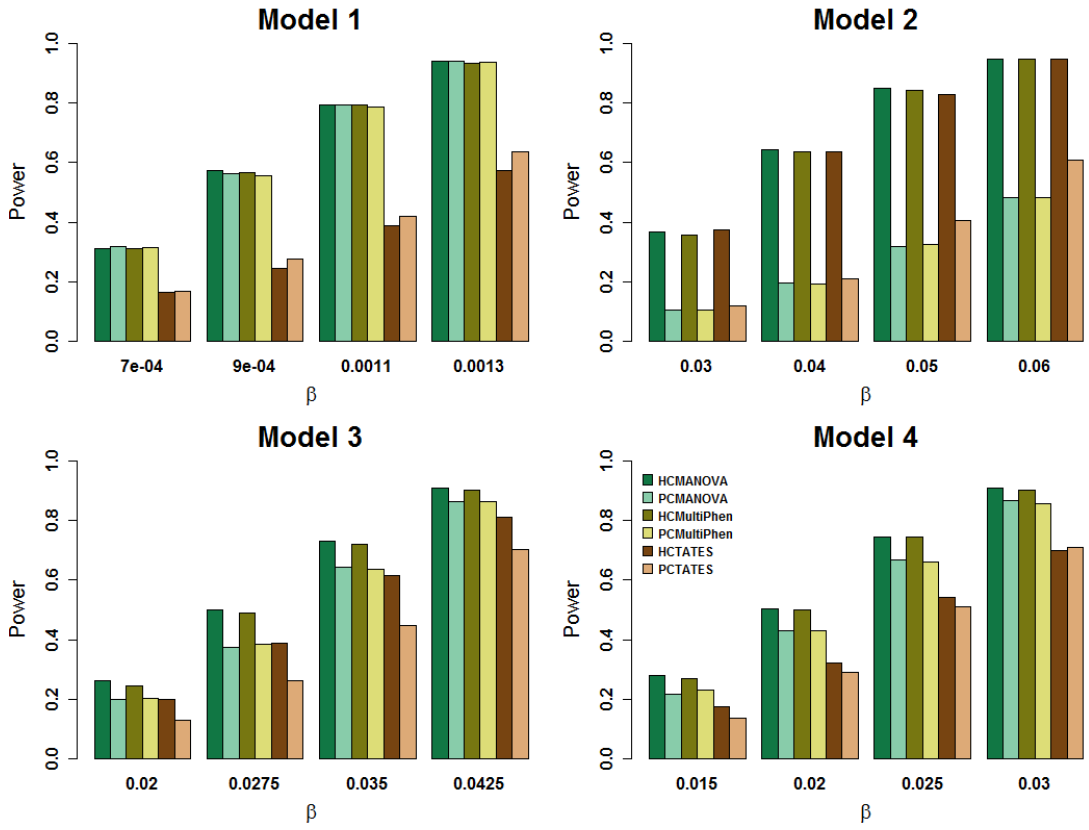


Figure A.3. Power comparisons of the six tests (HCMANOVA, MANOVA, HCMultiPhen, MultiPhen, HCTATES, and TATES) for the power as a function of effect size β for 20 quantitative phenotypes under Model S1.

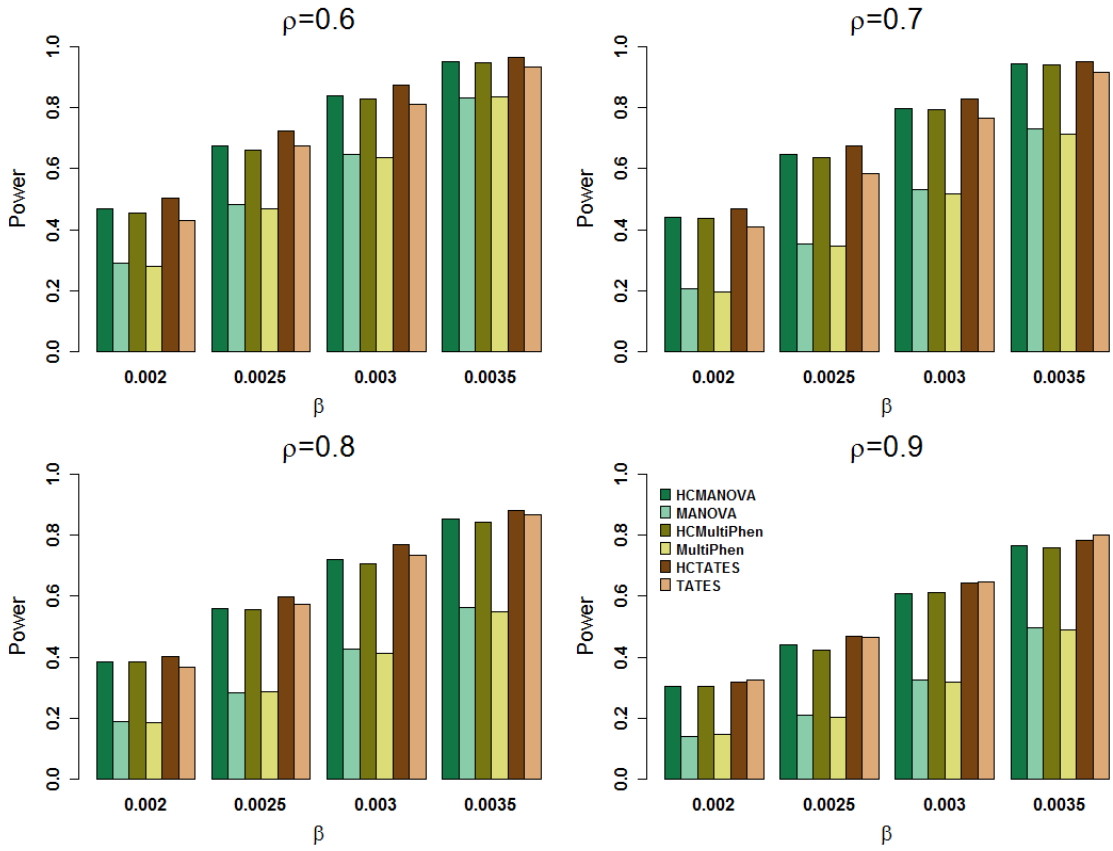


Figure A.4. Power comparisons of the six tests (HCMANOVA, HCMANOVA-PC1, HCMultiPhen, HCMultiPhen-PC1, HCTATES, and HCTATES-PC1) for the power as a function of effect size β for 20 quantitative phenotypes

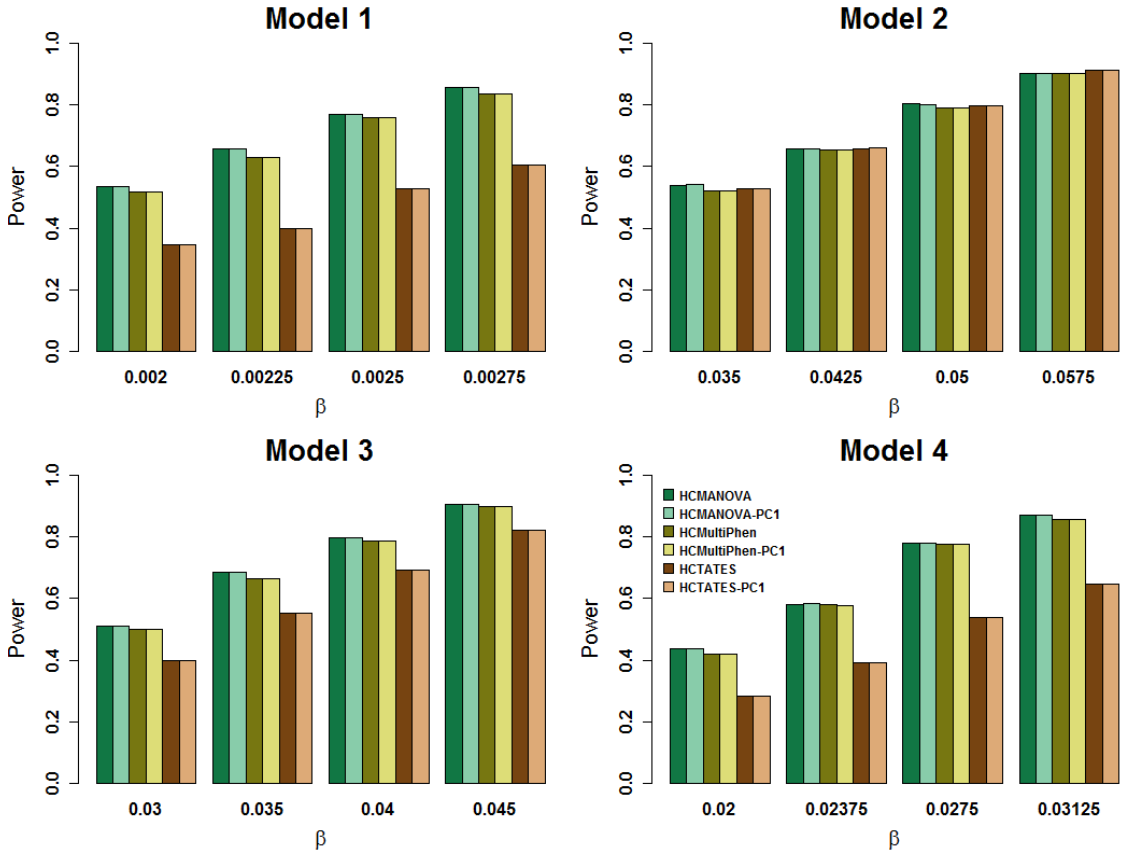


Figure A.5. Power comparisons of the six tests (HCMANOVA, HCMANOVA-PC1, HCMultiPhen, HCMultiPhen-PC1, HCTATES, and HCTATES-PC1) for the power as a function of effect size β for 40 quantitative phenotypes.