Test Gene-Environment Interactions for Multiple Traits in Sequencing Association Studies

Jianjun Zhang^a Qiuying Sha^b Han Hao^a Shuanglin Zhang^b
Xiaoyi Raymond Gao^{c-e} Xuexia Wang^a

^aDepartment of Mathematics, University of North Texas, Denton, TX, USA; ^bDepartment of Mathematical Sciences, Michigan Technological University, Houghton, MI, USA; ^cDepartment of Ophthalmology and Visual Science, The Ohio State University, Columbus, OH, USA; ^dDepartment of Biomedical Informatics, The Ohio State University, Columbus, OH, USA; ^eDivision of Human Genetics, The Ohio State University, Columbus, OH, USA

Keywords

Gene-environment interactions · Principal component analysis · Standardization analysis · Fisher's combination test

Abstract

Motivation: The risk of many complex diseases is determined by an interplay of genetic and environmental factors. The examination of gene-environment interactions (G×Es) for multiple traits can yield valuable insights about the etiology of the disease and increase power in detecting disease-associated genes. However, the methods for testing G×Es for multiple traits are very limited. **Method:** We developed novel approaches to test G×Es for multiple traits in sequencing association studies. We first perform a transformation of multiple traits by using either principal component analysis or standardization analysis. Then, we detect the effects of G×Es using novel proposed tests: testing the effect of an optimally weighted combination of G×Es (TOW-GE) and/or variable weight TOW-GE (VW-TOW-GE). Finally, we employ Fisher's combination test to combine the *p* values. **Results:** Extensive simulation studies show that the type I error rates

of the proposed methods are well controlled. Compared to the interaction sequence kernel association test (ISKAT), TOW-GE is more powerful when there are only rare risk and protective variants; VW-TOW-GE is more powerful when there are both rare and common variants. Both TOW-GE and VW-TOW-GE are robust to directions of effects of causal G×Es. Application to the COPD Gene Study demonstrates that our proposed methods are very effective. **Conclusions:** Our proposed methods are useful tools in the identification of G×Es for multiple traits. The proposed methods can be used not only to identify G×Es for common variants, but also for rare variants. Therefore, they can be employed in identifying G×Es in both genome-wide association studies and next-generation sequencing data analyses.

© 2020 S. Karger AG, Basel

Introduction

The risk of many complex diseases is determined by an interplay of genetic and environmental factors. For example, anthracyclines are one of the most effective classes of chemotherapeutic agents currently available for cancer

treatment. The therapeutic potential of anthracyclines, however, is limited because of their strong dose-dependent relation with progressive and irreversible cardiomyopathy leading to congestive heart failure. Two genes, HAS3 (hyaluronan synthase 3) and CELF4 (CUGBP Elav-like family member 4), exert modifying effects on anthracycline dose-dependent cardiomyopathy risk [1, 2].

Complex diseases are often characterized by many correlated traits that can better reflect their underlying mechanism. For example, hypertension can be characterized by systolic and diastolic blood pressure [3]; metabolic syndrome is evaluated by four component traits: high-density lipoprotein cholesterol, plasma glucose and type 2 diabetes, abdominal obesity, and diastolic blood pressure [4]; a person's cognitive ability is usually measured by tests in domains including memory, intelligence, language, executive function, and visual-spatial function [5]; and for chronic obstructive pulmonary disease (COPD), there are 7 key quantitative COPD-related phenotypes [6]. More and more large cohort studies have collected a broad array of correlated phenotypes to reveal the genetic components of many complex human diseases. Therefore, by jointly analyzing these correlated traits, we can not only gain more power by aggregating multiple weak effects, but also understand the genetic architecture of the disease of interest.

Evidence suggests that gene-environment interaction (G×E) effects may not only exist for a single trait but also for multiple traits of a disease. For example, a gene-by-stress genome-wide interaction analysis identifies EBF1 as a cardiovascular and metabolic risk gene. Gene EBF1 not only has a gene-by-stress interaction effect on hip circumference but also gene-by-stress interaction effects on waist circumference, BMI, fasting glucose, type 2 diabetes status, and common carotid intimal medial thickness [7].

Examination of G×E for multiple traits can yield valuable insights about the etiology of the disease and increase power in detecting disease-associated genes. Generally, a gene is the basic functional unit of inheritance. Thus, results from gene level association tests can be more readily integrated with downstream functional and pathogenic investigations than single marker tests. Next-generation sequencing technology provides potential opportunities to make directly testing both common and rare variants in a gene feasible [8]. Rare variants are usually defined as genetic variants with a minor allele frequency (MAF) <5%, which may play an important role in studying the etiology of complex human diseases. Numerous statistical methods have been developed for testing the main effect of rare variants, such as the sequence kernel association test (SKAT) [9], the combined multivariate and col-

lapsing (CMC) method [10], the weighted sum statistic (WSS) [11], and testing the effect of an optimally weighted combination of variants (TOW) [12].

The tests discussed above are designed to assess the association of the main effects of rare variants with a single trait. To our knowledge, limited methods have been developed for testing G×Es in sequencing association studies, especially for multiple traits. Existing methods for assessing common variants by environment interactions such as the gene-environment set association test (GESAT) [13] will be less powerful when it is naively applied to rare variants G×E analysis [14]. To test rare variants by environment interactions, Lin et al. [14] developed the interaction sequence kernel association test (ISKAT) to assess rare variants by environment interactions. When there is no prior information, ISKAT recommends Beta (MAF; 1, 25) as the weight which has the beta distribution density function with parameters 1 and 25 evaluated at the sample MAF. ISKAT may lose power when the MAFs of causal variants are not in the range (0.01, 0.035). Furthermore, existing methods for multiple-trait association tests have primarily focused on testing main effects of rare variants [15] or common variants [16, 17]. It is essential to develop novel methods for multiple traits to detect the G×Es for both rare and common variants.

In this article, we develop novel methods to test for rare and/or common variants by environment interactions on multiple traits in sequencing association studies. Our methods can be divided into three steps. We first use either principal component analysis or standardization analysis to release the correlation among the multiple traits. Then, we detect the effect of G×E for each transferred trait using novel proposed tests: testing the effect of an optimally weighted combination of G×E (TOW-GE) and/or variable weight TOW-GE (VW-TOW-GE). Both TOW-GE and VW-TOW-GE are robust to directions of effects of causal G×Es. Finally, we employ Fisher's combination test (FCT) to combine the *p* values of TOW-GE and/or VW-TOW-GE. Furthermore, we evaluate the performance of the proposed methods via simulation studies and real data analysis using the sequencing data from the COPDGene Study.

Materials and Methods

Methods

Consider *n* unrelated individuals sequenced in a region (a gene or a pathway) with *m* variants, where there are *p* rare variants and *K* measured continuous traits. For ease of presentation, we only consider a single environmental factor. We are interested in studying the *p* rare variants by environment interactions. The method

can be easily extended to the case where there is more than one environmental factor. For individual $i = 1, \dots, n$, let $\mathbf{Y}_i = (y_{i1}, \dots, y_{iK})^T$ denote the K continuous traits, $\mathbf{X}_i = (x_{i1}, \dots, x_{iq})^T$ denote the q covariates, $\mathbf{G}_i = (g_{i1}, \dots, g_{ip})^T$ denote genotypes for the p rare variants, and E_i denote the environmental factor. In our paper, we assume $p < n$. For simplicity, we assume a common set of covariates for all traits which can be easily extended to the case of differing covariates. Let $\mathbf{S}_i = (E_i g_{i1}, \dots, E_i g_{ip})^T$ be a vector of variants by environment interaction terms for the i -th individual.

The goal of our analysis is to study the association of these multiple phenotype traits with rare variants by environment interaction effects. Our method can be divided into the following three steps:

Step 1. Denote the correlation matrix of $\mathbf{Y}_i$ as $\Sigma = (\rho_{kl})$, where $\rho_{kl} = \text{Cor}(y_{ik}, y_{il})$ and define an $n \times K$ phenotype matrix $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_n)^T$. Two methods can be used to reduce the correlation among $y_{i1}, \dots, y_{iK}$:

- 1 We propose to use principal component (PC) analysis for multiple traits and obtain independent phenotype matrix $\mathbf{Y}^* = (\mathbf{Y}_1^*, \dots, \mathbf{Y}_K^*)^T$, where we denote $\mathbf{Y}_1^*$ as the first PC, $\mathbf{Y}_2^*$ as the second PC, and so on. According to the principal axis theorem in mechanics, we know that $\mathbf{Y}_1^*, \dots, \mathbf{Y}_K^*$ are independent.
- 2 If we assume the mean and variance of $\mathbf{Y}_i = (y_{i1}, \dots, y_{iK})^T$ are $\boldsymbol{\mu}$ and Σ , then we can also obtain the independent vector $\mathbf{Y}_1^*, \dots, \mathbf{Y}_K^*$ by standardization:

$$\mathbf{Y}^* = (\mathbf{Y} - \boldsymbol{\mu}) \Sigma^{-\frac{1}{2}}.$$

In practice, we use the unbiased estimators of $\boldsymbol{\mu}$ and Σ : $\hat{\boldsymbol{\mu}} = \bar{\mathbf{Y}}$ and $\hat{\Sigma} = \text{var}(\mathbf{Y})$, where $\bar{\mathbf{Y}}$ denotes the sample mean and $\text{var}(\mathbf{Y})$ denotes the sample covariance matrix.

In the Appendix, we prove that when multiple traits follow a multivariate normal distribution, if the multiple traits are uncorrelated, they are independent. For simplicity of symbols, we will use y_{ik} to denote y_{ik}^* , the elements of $\mathbf{Y}^*$, in the following paragraphs.

Step 2. We use the generalized linear model to model the relationship between the k -th trait values y_{ik} and covariates $\mathbf{X}_i$, genotypes $\mathbf{G}_i$, environmental factor E_i , and G×Es $\mathbf{S}_i$:

$$g(E(y_{ik} | \mathbf{X}_i, \mathbf{G}_i, E_i)) = \mathbf{X}_i^T \boldsymbol{\alpha}_1^k + E_i \boldsymbol{\alpha}_2^k + \mathbf{G}_i^T \boldsymbol{\alpha}_3^k + \mathbf{S}_i^T \boldsymbol{\beta}_k \quad (1)$$

$$= \tilde{\mathbf{X}}_i^T \boldsymbol{\alpha}_k + \mathbf{S}_i^T \boldsymbol{\beta}_k,$$

where $g(\cdot)$ be a canonical link function, $\boldsymbol{\alpha}_1^k, \boldsymbol{\alpha}_2^k, \boldsymbol{\alpha}_3^k$, and $\boldsymbol{\beta}_k$ are defined as $q \times 1$ coefficient vector of the covariate, the coefficient vector of the environmental factor, $p \times 1$ coefficient vector of the genotype, and $p \times 1$ coefficient vector of G×Es for the i -th individual and the k -th trait, respectively. Let $\tilde{\mathbf{X}}_i = (\mathbf{X}_i^T, E_i, \mathbf{G}_i^T)^T$ and $\boldsymbol{\alpha}_k = (\boldsymbol{\alpha}_1^k, \boldsymbol{\alpha}_2^k, \boldsymbol{\alpha}_3^k)^T$. For the k -th trait, we are interested in testing the null hypothesis $H_0: \boldsymbol{\beta}_k = 0$.

We develop a score test by treating $\boldsymbol{\alpha}_k$ as nuisance parameters and then adjust both the k -th trait value y_{ik} and $\mathbf{S}_i$ for the covariates $\mathbf{X}_i$, the genotypic score $\mathbf{G}_i$, and the environmental variable E_i by applying linear regression models. Specifically, $y_{ik} = \mathbf{X}_i^T \boldsymbol{\alpha}_{10}^k + E_i \boldsymbol{\alpha}_{20}^k + \mathbf{G}_i^T \boldsymbol{\alpha}_{30}^k + \tau_{ik}$ and $\mathbf{S}_i = \mathbf{X}_i^T \boldsymbol{\alpha}_{11} + E_i \boldsymbol{\alpha}_{21} + \mathbf{G}_i^T \boldsymbol{\alpha}_{31} + \nu_i$, where τ_{ik} and ν_i are normally distributed random error terms. Denote $\tilde{y}_{ik}$ as the residual of y_{ik} and $\tilde{\mathbf{S}}_i = (\tilde{s}_{i1}, \dots, \tilde{s}_{ip})$ as the residual of $\mathbf{S}_i$. Then, the relationship between $\tilde{y}_{ik}$ and $\tilde{\mathbf{S}}_i$ can be modeled by the generalized linear model:

$$g(E(\tilde{y}_{ik} | \tilde{\mathbf{S}}_i)) = \boldsymbol{\beta}_0^* + \tilde{\mathbf{S}}_i^T \boldsymbol{\beta}^*. \quad (2)$$

To test $H_0: \boldsymbol{\beta}_k = 0$ in equation 1 is equivalent to test $H_0: \boldsymbol{\beta}^* = 0$ in equation 2 [12]. To overcome the sparsity issue in G×E data and increase the power of a G×E association test, we extend the TOW method [12] and propose a novel test to test the effect of a weighted combination of G×Es,

$$\tilde{s}_i = \sum_{j=1}^p w_j \tilde{s}_{ij}.$$

Under the generalized linear model, the score test statistic given by Sha et al. [18] becomes:

$$S(w_1, \dots, w_p) = n \frac{\left(\sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k) (\tilde{s}_i - \bar{s}) \right)^2}{\sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k)^2 \sum_{i=1}^n (\tilde{s}_i - \bar{s})^2}$$

$$= n \frac{\left(\sum_{j=1}^p w_j \sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k) (\tilde{s}_{ij} - \bar{s}_j) \right)^2}{\sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k)^2 \sum_{i=1}^n (\tilde{s}_i - \bar{s})^2}.$$

Because G×E for rare variants are essentially independent, we have:

$$\sum_{i=1}^n (\tilde{s}_i - \bar{s})^2 = \sum_{j=1}^p \sum_{l=1}^p w_j w_l \sum_{i=1}^n (\tilde{s}_{ij} - \bar{s}_j) (\tilde{s}_{il} - \bar{s}_l)$$

$$\approx \sum_{j=1}^p w_j^2 \sum_{i=1}^n (\tilde{s}_{ij} - \bar{s}_j)^2.$$

Thus, as a function of $(w_1, \dots, w_p)$, the score test statistic $S(w_1, \dots, w_p)$ reaches its maximum

$$S_0(w_1^0, \dots, w_p^0) = n \sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k) (\tilde{s}_i^0 - \bar{s}^0) / \sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k)^2,$$

when

$$w_j^0 = \sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k) (\tilde{s}_{ij} - \bar{s}_j) / \sum_{i=1}^n (\tilde{s}_{ij} - \bar{s}_j)^2$$

and

$$\tilde{s}_i^0 = \sum_{j=1}^p w_j^0 \tilde{s}_{ij}.$$

Similarly, we define the statistic to test the effect of the optimally weighted combination of G×Es (TOW-GE),

$$\sum_{j=1}^p w_j^0 \tilde{s}_{ij}$$

as:

$$T_{\text{TOW-GE}} = \sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k) (\tilde{s}_i^0 - \bar{s}^0), \quad (3)$$

which is equivalent to $S_0(w_1^0, \dots, w_p^0)$ when we use a permutation test to evaluate p values, where

$$\sum_{i=1}^n (\tilde{y}_{ik} - \bar{y}_k)^2$$

can be viewed as a constant.

The optimal weight w_j^0 is equivalent to

$$w_j^{0*} = \rho(\tilde{y}_k, \tilde{s}_j) / \sqrt{\sum_{i=1}^n (\tilde{s}_{ij} - \bar{\tilde{s}}_j)^2},$$

where $\rho(\tilde{y}_k, \tilde{s}_j)$ is the correlation coefficient between $\tilde{y}_k = (\tilde{y}_{1k}, \dots, \tilde{y}_{nk})$ and $\tilde{s}_j = (\tilde{s}_{1j}, \dots, \tilde{s}_{nj})$. From the expression of w_j^{0*} , we can see that it is proportional to $\rho(\tilde{y}_k, \tilde{s}_j)$, and thus, w_j^0 will put large weights to the G×Es that have strong associations with the trait and also adjust for the directions of the associations. Simultaneously, w_j^{0*} is proportional to

$$1 / \sqrt{\sum_{i=1}^n (\tilde{s}_{ij} - \bar{\tilde{s}}_j)^2},$$

w_j^0 will put large weights to G×Es with small variations which are common in G×Es for rare variants.

TOW-GE focuses mainly on rare variants by environment interactions, and it may lose power for both rare and common variants by environment interactions because it puts small weights on common variants by environment interactions. Thus, to test the G×E effects of both rare and common variants, we propose the following variable weight TOW-GE denoted as VW-TOW-GE. We first divide G×Es into two parts based on rare or common variants, and then, we apply TOW-GE to the two parts separately. Let

$$T_\lambda = \lambda \frac{T_r}{\sqrt{\text{var}(T_r)}} + (1 - \lambda) \frac{T_c}{\sqrt{\text{var}(T_c)}},$$

where T_r and T_c denote the test statistics of TOW-GE for G×E effects of rare and common variants, respectively. Denote p_λ as the p value of T_λ , and then, the test statistic of VW-TOW-GE is defined as $T_{VW-TOW-GE} = \min_{0 \leq \lambda \leq 1} p_\lambda$. Here, we use a finite grid search of λ to evaluate the statistics $T_{VW-TOW-GE}$. We divide the interval $[0, 1]$ into K subintervals of equal length ($\lambda_k = k/K$ for $k = 0, \dots, K$) and take the minimum p value of T_{λ_k} for a series of λ_k . The p value of $T_{VW-TOW-GE}$ can be evaluated by permutation tests following similar steps of permutation tests for VW-TOW as proposed by Sha et al. [12].

Step 3. For all K independent traits after being transformed, we obtain K marginal p values: $p_1, \dots, p_K$ from Step 2. Based on the study by Fisher [19], the FCT for K traits is given by:

$$T_{FCT} = -2 \sum_{i=1}^K \log p_i,$$

where p_i denotes the p value of the test statistic T_i for the i -th trait. Because these K traits are independent, the marginal p values are independent, and thus, T_{FCT} follows a χ^2 distribution with $2K$ degrees of freedom.

Comparison of Methods

We compared the performance of our proposed methods with ISKAT [14], the modified WSS for testing G×Es [11], and the modified CMC method for testing G×Es [10] in Step 2. In this study, the rank sum test used by WSS and the T^2 test used by CMC were replaced with the score test based on residuals $\tilde{y}_{ik}$ and $\tilde{s}_{ij}$.

Simulation Studies

The empirical Mini-Exome Genotype data provided by the GAW17 is used for simulation studies. The dataset contains genotypes of 697 unrelated individuals on 3,205 genes. Gene ELAVL4 in GAW17 was used to simulate the G×E effect on quantitative trait Q_1 which follows a normal distribution. Q_2 is a quantitative trait which has a weak correlation with Q_1 . We chose gene ELAVL4

in our simulation study. Gene ELAVL4 has 10 variants containing 8 rare variants and 2 common variants. Rare variants in the simulation are defined by a MAF < 0.05.

We consider four correlated traits ($K = 4$) with a compound-symmetry correlation matrix and consider two covariates: a standard normal covariate X_1 and a binary covariate X_2 with $P(X_2 = 1) = 0.5$. The environmental factor E is assumed to be continuous following standard normal distribution. We generate trait values by using the following four models:

$$\begin{aligned} Y_1 &= 0.5X_1 + 0.5X_2 + E\alpha_{11} + \mathbf{G}^T\alpha_{21} + \mathbf{S}^T\beta_1 + S_c\beta_1^c + \varepsilon_1 \\ Y_2 &= X_1 + X_2 + E\alpha_{12} + \mathbf{G}^T\alpha_{22} + \mathbf{S}^T\beta_2 + S_c\beta_2^c + \varepsilon_2 \\ Y_3 &= 0.5X_1 + 0.5X_2 + E\alpha_{13} + \mathbf{G}^T\alpha_{23} + \mathbf{S}^T\beta_3 + S_c\beta_3^c + \varepsilon_3 \\ Y_4 &= X_1 + X_2 + E\alpha_{14} + \mathbf{G}^T\alpha_{24} + \mathbf{S}^T\beta_4 + S_c\beta_4^c + \varepsilon_4, \end{aligned}$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ follow a normal distribution with mean of zero and variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$. We consider three scenarios for the correlation coefficient $\rho = 0.2, 0.5, 0.8$; $\alpha_1 = (0.015, 0.015, 0.03, -0.02)$; $\mathbf{S}$ is for the rare variants' G×Es, and S_c is one common variant's G×E. Specifically, in the case of rare variants' G×Es only, $\beta_i^c = 0$ for $i = 1, 2, 3, 4$.

For type I error evaluation, we consider that there are no G×E effects in the aforementioned four models. Two scenarios are considered to evaluate the empirical type I error rates: (a) including main effects, we set the magnitudes of matrix $\alpha_2 = (\alpha_{21}^T, \alpha_{22}^T, \alpha_{23}^T, \alpha_{24}^T)^T$ as 0.3, and the sign for each α is randomly sampling from $(-1, 1)$; (b) no main effects ($\alpha_2 = 0$).

For power comparisons, we consider that there are G×E effects in the aforementioned four models in two different cases: (1) rare variants' G×Es only; (2) both rare and common variants' G×E effects. We consider two different scenarios for each case: (a) including main effects ($\alpha_2 \neq 0$); (b) no main effects ($\alpha_2 = 0$). We vary the number of non-zero elements in vector β_i , the proportion of non-zero elements in β_i that are positive, and the magnitudes of the non-zero β_{ij} . We set the magnitudes of the non-zero β_{ij} as $|\beta_{ij}| = c$ and increase c from 0.1 to 0.5.

Results

Simulation Results

The empirical type I error rates for both rare and common variants are shown in Table 1; and for rare variants only, they are shown in Appendix Table A1 (correlation coefficient $\rho = 0.5$). In each simulation scenario, the p values are estimated by 10,000 permutations, and the type I error rates are evaluated using 10,000 replicated samples. For 10,000 replicated samples, the 95% confidence intervals for type I error rates of nominal levels 0.05, 0.01, and 0.001 are (0.046, 0.054), (0.008, 0.012), and (0.0004, 0.0016), respectively. When there are (a) main effects ($\alpha_2 \neq 0$), TOW-GE, VW-TOW-GE, ISKAT, and WSS demonstrate well-controlled type I error rates, and the burden test CMC tends to have very conservative type I error rates (top panel of Table 1 and Appendix Table A1). When there are (b) no main effects ($\alpha_2 = 0$), all methods

Table 1. Type 1 error rates for both rare and common variants in the presence of main effects (top panel) and in the absence of main effects (bottom panel) for $n = 2,000$, where the type I error rates outside parentheses are based on PCA and the type I error rates in parentheses are based on STD analysis according to Step 1

α level	TOW-GE	ISKAT	WSS	CMC	VW-TOW-GE
<i>With main effect</i>					
0.05	0.054 (0.055)	0.053 (0.051)	0.054 (0.056)	0.035 (0.043)	0.051 (0.050)
0.01	0.012 (0.013)	0.011 (0.012)	0.011 (0.013)	0.007 (0.008)	0.011 (0.012)
0.001	0.0017 (0.0018)	0.0009 (0.0010)	0.0013 (0.0011)	0.0005 (0.0004)	0.0010 (0.0012)
<i>Without main effect</i>					
0.05	0.054 (0.055)	0.052 (0.054)	0.055 (0.056)	0.040 (0.037)	0.054 (0.060)
0.01	0.013 (0.011)	0.011 (0.010)	0.010 (0.011)	0.007 (0.006)	0.013 (0.014)
0.001	0.0012 (0.0018)	0.0019 (0.0017)	0.0018 (0.0019)	0.0018 (0.0017)	0.0009 (0.0015)

PCA, principal component analysis; STD, standardization; TOW-GE, testing the effect of an optimally weighted combination of G×E; ISKAT, interaction sequence kernel association test; WEE, weighted sum statistic; CMC, combined multivariate and collapsing method; VW-TOW-GE, variable weight TOW-GE.

control type I error rates very well (bottom panel of Table 1 and Appendix Table A1).

Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for testing rare variants' G×E effects on continuous traits with main effect are shown in Appendix Table A2, and without main effect, they are shown in Appendix Table A3 (correlation coefficient $\rho = 0.5$). From these two tables, we can see that the empirical power for all methods are increasing when sample size n increases. When sample size $n = 2,000$, we can easily see the difference regarding the power of the different methods.

The results based on PC analysis (PCA) for testing rare variants' G×E effects when including main effect are also given in Figure 1, and the results when not including main effect are given in Appendix Figure A1 (correlation coefficient $\rho = 0.5$). Figure 2 shows results when including main effect, and Appendix Figure A2 shows results when not including main effect for testing rare variants' G×E effects based on standardization analysis (correlation coefficient $\rho = 0.5$). The top, middle, and bottom panels of these figures correspond to the three scenarios in which there are 2, 6, and 8 non-zero β_{ijs} , respectively. The left and right panels of the figures correspond to the two cases in which 50% of the β_{ijs} are positive and 100% of the β_{ijs} are positive for the i -th trait, respectively. For each plot, we vary c , the magnitudes of the non-zero β_{ijs} . As shown in the four figures for the case where 50% of the β_{ijs} are positive, TOW-GE is more powerful than the other three tests. For the case where 100% of the β_{ijs} are positive, WSS is relatively more powerful than TOW-GE, and TOW-GE is more powerful than the other two tests.

The WSS is very sensitive to the directions of effects due to its simple aggregation of G×Es. Among the four tests (TOW-GE, ISKAT, WSS, and CMC) in the two different cases, CMC is the weakest test.

Power comparisons of the five tests (TOW-GE, VW-TOW-GE, ISKAT, WSS, and CMC) for both rare and common variants' G×E effects are given in Figure 3 (correlation coefficient $\rho = 0.05$). For each plot, we vary c from 0.02 to 0.1 and set 50% of the β_{ijs} as positive. Simultaneously, we set the coefficient of the common variant by environment interaction β_i^c as positive, and the magnitude of β_i^c as twice that of β_{ijs} . From Figure 3, we can see that the power of TOW-GE, VW-TOW-GE, and CMC outperform the other two tests, and WSS is the least powerful test. WSS loses power because it gives a common variant by environment interaction effect a very small weight. The method ISKAT, which is similar to SKAT, performs worse than our methods, and it will lose power when the MAFs of causal variants are not in the range (0.01, 0.035) [20].

In order to understand when and which approach should be utilized, power comparisons between PCA and Standardization (STD) analysis for TOW-GE are performed. The power of TOW-GE for testing rare variant by environment interaction effects on a continuous trait with main effect is shown in Figure 4, and the power of TOW-GE without main effect is shown in Appendix Figure A3 (correlation coefficient $\rho = 0.5$). As shown in these figures, TOW-GE using PCA is relatively more powerful than TOW-GE using STD.

We also assessed the performance of the proposed methods when the correlation coefficient among any pair of traits are $\rho = 0.2$ and 0.8, respectively. The results based on

Fig. 1. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome when there are main effects based on PCA. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

PCA for testing rare variants' G×E effects when including a main effect are given in Appendix Figure A4 for $\rho = 0.2$ and Appendix Figure A5 for $\rho = 0.8$, and the results when not including a main effect are given in Appendix Figure A6 for $\rho = 0.2$ and Appendix Figure A7 for $\rho = 0.8$. Appen-

dix Figure A8 for $\rho = 0.2$ and Appendix Figure A9 for $\rho = 0.8$ show the results when including a main effect, and Appendix Figure A10 for $\rho = 0.2$ and Appendix Figure A11 for $\rho = 0.8$ show the results when not including a main effect for testing rare variants' G×E effects based on STD analysis.

Fig. 2. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome when there are main effects based on STD analysis. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Power comparisons of the five tests (TOW-GE, VW-TOW-GE, ISKAT, WSS, and CMC) for both rare and common variants' G×E effects are given in Appendix Figure A12 for $\rho = 0.2$ and Appendix Figure A13 for $\rho = 0.8$. From the

aforementioned figures, we can see that all results for each scenario when $\rho = 0.2$ and $\rho = 0.8$ are consistent with the results for the corresponding scenario when $\rho = 0.5$.

Fig. 3. Power comparisons of five tests (TOW-GE, ISKAT, WSS, CMC, and VW-TOW-GE) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing both rare and common variant by environment interaction effects on a continuous outcome. **a, b** PCA. **c, d** STD analysis. **a, c** With main effect. **b, d** Without main effect.

In conclusion, TOW-GE is the most powerful test in the case of rare variants by environment interaction effects except for the case where 100% of the β_{ij} s are positive. VW-TOW-GE is the most powerful test or has similar power to the most powerful test in the case of both rare and common variants' G×E effects.

Real Data Analysis

COPD is one of the most common lung diseases mainly characterized by long-term poor airflow. COPD affects millions of Americans and is the third leading cause of disease-related death in the USA [21]. It is a complex disease influenced by genetic factors, environmental factors, and G×Es. It is well known that cigarette smoking is the main causal factor of COPD [22]. Several genes are suggested to play a role in the presence of a gene-by-smoking interaction. For example, Hersh et al. [23] reported that the SFTPB Thr131Ile polymorphism was associated with COPD, but only in the presence of a

gene with an environment interaction. SNP rs2292566 in gene EPHX1 was associated with COPD only in the presence of a gene-by-smoking (pack-years) interaction. Celedón et al. [24] showed that two SNPs in the promoter region of TGFB1 (rs2241712 and rs1800469) and one SNP in exon 1 of TGFB1 (rs1982073) were significantly associated with COPD among smokers in a COPD case-control study. Sandford and Silverman [22] reported that the GSTM1 gene was associated with severe chronic bronchitis in heavy smokers and an association of the TNF -308A allele with COPD was found in a Taiwanese population.

The COPDGene Study is a multi-center genetic and epidemiologic investigation to study COPD [25]. This study is sufficiently large and appropriately designed for a thorough analysis of COPD. In this study, we considered more than 5,000 non-Hispanic White (NHW) participants who had already completed a detailed protocol, including questionnaires, pre- and post-bronchodilator

Fig. 4. Power comparisons for PCA and STD analysis of the test TOW-GE for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome with main effect. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

spirometry, high-resolution CT scanning of the chest, exercise capacity (assessed by six-minute walk distance, 6MWD), and blood samples for genotyping. The participants were genotyped using the Illumina OmniExpress platform. The genotype data has 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. We performed genotype imputation using the 1000 Genomes Phase I v3 European reference panel.

Table 2. Summary of association analysis results of the COPD dataset

Gene name	TOW	TOW-GE	SKAT	WSS	CMC	VW-TOW
<i>Gene with main effects</i>						
SFTPb	0.0421 (0.0456)		0.3321 (0.3312)	0.2008 (0.3051)	0.7087 (0.8450)	0.0891 (0.2553)
EPHX1	0.5026 (0.5070)		0.3755 (0.3385)	0.9814 (0.9774)	0.8803 (0.9100)	0.5902 (0.3886)
GSTM1	0.0235 (0.0252)		0.3234 (0.3151)	0.6725 (0.7374)	0.3630 (0.3038)	0.0263 (0.0586)
HMOX1	0.1024 (0.0858)		0.0315 (0.0596)	0.0067 (0.0203)	0.9999 (0.9997)	0.3001 (0.1994)
TGFB1	0.0825 (0.0798)		0.1128 (0.1695)	0.4454 (0.4591)	0.2170 (0.3686)	0.4556 (0.1576)
TNF	0.2154 (0.2074)		0.2198 (0.2346)	0.0879 (0.2169)	0.7663 (0.7416)	0.3094 (0.2562)
<i>Gene-by-smoking interaction with main effect</i>						
SFTPb		0.3392 (0.2733)	0.6413 (0.6035)	0.6253 (0.5863)	0.8587 (0.8006)	0.4688 (0.6582)
EPHX1		0.5742 (0.5757)	0.5533 (0.5174)	0.3550 (0.3527)	0.9870 (0.9722)	0.8870 (0.9833)
GSTM1		0.5735 (0.5074)	0.2983 (0.2542)	0.9480 (0.9074)	0.3397 (0.3765)	0.5520 (0.5297)
HMOX1		0.4390 (0.4341)	0.2395 (0.2196)	0.2360 (0.1408)	0.9999 (0.9997)	0.5692 (0.4876)
TGFB1		0.4020 (0.4195)	0.6064 (0.5688)	0.4755 (0.5102)	0.9903 (0.9981)	0.3891 (0.4833)
TNF		0.8295 (0.8619)	0.6049 (0.6282)	0.1669 (0.2875)	0.0305 (0.0427)	0.8572 (0.8884)
<i>Gene-by-smoking interaction without main effect</i>						
SFTPb		0.0026 (0.0012)	0.1048 (0.0913)	0.1089 (0.0908)	0.3569 (0.4727)	0.0214 (0.0270)
EPHX1		0.2298 (0.2100)	0.2307 (0.2047)	0.9178 (0.9264)	0.4418 (0.4286)	0.3199 (0.3834)
GSTM1		0.0825 (0.0867)	0.1741 (0.1983)	0.6195 (0.7102)	0.3092 (0.3279)	0.0898 (0.1780)
HMOX1		0.0474 (0.0481)	0.0177 (0.0099)	0.0009 (0.0013)	0.9994 (0.9985)	0.0839 (0.0982)
TGFB1		0.0290 (0.0374)	0.0786 (0.0913)	0.4781 (0.5078)	0.1920 (0.2101)	0.1130 (0.0705)
TNF		0.2396 (0.2372)	0.2218 (0.2470)	0.0949 (0.1793)	0.2010 (0.2047)	0.4084 (0.3226)

The p value is calculated for gene with main effects (top panel), gene-by-smoking interaction with main effect (middle panel), gene-by-smoking interaction without main effect (bottom panel), where p values outside parentheses are based on PCA and p values in parentheses are based on STD analysis according to Step 1.

COPD, chronic obstructive pulmonary disease. For all other abbreviations, see footnote of Table 1.

Bold values represent significant p values for significance level 0.05.

Based on the literature studies of COPD [26, 27], 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), and 6MWD plus 3 covariates, including BMI, age, and sex, and 1 environmental factor (Pack-Years). EmphDist is the ratio of emphysema at -950 HU in the upper 1/3 of the lung fields compared to the lower 1/3 of the lung fields where we did a log transformation on EmphDist in the following analysis, referred also to Chu et al. [26]. In the analysis, participants with missing data in any of these phenotypes were excluded.

To evaluate the performance of our proposed method on a real dataset, we applied all 5 methods (TOW-GE, ISKAT, WSS, CMC, and VW-TOW-GE) to 6 COPD-associated genes (SFTPb, EPHX1, GSTM1, HMOX1, TGFB1, and TNF) through an interaction with cigarette smoking. In the analysis, we excluded the extreme low frequency SNPs ($MAF < 0.001$) and individuals with

missing values in any of the 7 phenotypes and 3 covariates. We considered three different scenarios: (1) gene main effects, (2) gene-by-smoking interaction with main effect, and (3) gene-by-smoking interaction without main effect. When we considered only main effects of genes, we used 5 popular genetic association testing methods (TOW, SKAT, WSS, CMC, and VW-TOW) in Step 2. We adopted 10,000 permutations for our methods and used 0.05 as the significance level. The results are summarized in Table 2. From Table 2, we can see that two genes (SFTPb and GSTM1) are identified only by our methods, and gene HMOX1 is identified by the SKAT and WSS methods for scenario (1) gene main effects; gene TNF is only identified by the CMC method for scenario (2) gene-by-smoking interaction with main effect; three genes (SFTPb, HMOX1, and TGFB1) are identified by our methods and the HMOX1 gene is also identified by the ISKAT and WSS methods for scenario (3) gene-by-smoking interaction without main effect. Our method identified the most significant genes, especially in scenario (3).

Discussion

Complex diseases are often characterized by correlated traits. For example, hypertension is defined by two levels according to the American College of Cardiology/American Heart Association (ACC/AHA) guidelines: (1) elevated blood pressure, with a systolic blood pressure between 120 and 129 mm Hg and a diastolic blood pressure <80 mm Hg, and (2) stage 1 hypertension, with a systolic blood pressure of 130–139 mm Hg or a diastolic blood pressure of 80–89 mm Hg [28]. Combining multiple correlated traits in one genetic association test can not only enhance our understanding of the etiology of the disease but also increase power in testing G×E effects. In this paper, we propose a new framework to test the association between G×Es and multiple traits. The proposed method can be divided into three steps: (1) reduce correlation among traits; (2) test the G×E effect for each transferred trait; and (3) combine *p* values of all transferred traits by FCT. PCA and STD analysis are applied to reduce the correlation among the traits in Step 1. Then, we apply the extended TOW method to test rare variants by environment interaction (TOW-GE) and the VW-TOW method to test both rare and common variants by environment interaction (VW-TOW-GE) in Step 2. We use the simulation studies to demonstrate that the proposed methods outperform comparable methods in most of the scenarios with or without main effect. In addition, we test G×E effects for 7 related phenotypes of COPD. Our proposed methods verified the largest number of genes especially for gene-by-smoking interactions without main effect.

We noticed that CMC loses power when the regions tested are large because of the increased degrees of freedom for Hotelling's T^2 , such as gene HMOX1 in the COPD data (13,854 bases). The CMC method is powerful when the directions of G×E effects of all causal variants are the same, and the gene tested is short, such as gene TNF (2,770 bases).

In this paper, we focus on the joint analysis of multiple continuous traits with G×E effects. Our results show that the proposed methods, TOW-GE or VW-TOW-GE, are the most powerful tests or have similar power to the most powerful test when compared with competing methods. It will be interesting to generalize the methods to study the joint test of mixed type of phenotypes, which is analytically and computationally more challenging. TOW-GE is similar to TOW, which is derived for independent variants. Since common variants

within a gene are usually correlated, it will be of future research interest to assess the performance of TOW-GE for common variant's G×E effects.

Disclosure Statement

The authors declare no conflicts of interest.

Funding Sources

Q. Sha was supported by the National Human Genome Research Institute of the National Institutes of Health (NIH) under award number R15HG008209. X. Raymond Gao was supported by NIH (Bethesda, MD, USA) grants R01EY027315 and RF1AG060472. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. X. Wang was supported by the University of North Texas Foundation which was contributed by Dr. Linda Truitt Creagh. The content of this paper is solely the responsibility of the authors and does not necessarily represent the views of the University of North Texas Foundation and Dr. Linda Truitt Creagh. The Genetic Analysis workshops are supported by NIH grant R01 GM031575 from the National Institute of General Medical Sciences. Preparation of the Genetic Analysis Workshop 17 Simulated Exome Data Set was supported in part by NIH R01 MH059490 and used sequencing data from the 1000 Genomes Project (www.1000genomes.org). This research used data generated by the COPDGene study (phs000179/HMB and phs000179/DS-CS-RD), which was supported by NIH grants U01HL089856 and U01HL089897. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the NIH. The COPDGene project is also supported by the COPD Foundation through contributions made by an Industry Advisory Board comprised of Pfizer, AstraZeneca, Boehringer Ingelheim, Novartis, and Sunovion. A superior high-performance computing infrastructure at University of North Texas was used in obtaining the results presented in this publication.

Author Contributions

Study design and formal analysis: J.Z. and X.W.; Methodology: J.Z. and X.W.; Visualization: Q.S.; Writing – original draft: J.Z. and X.W.; Writing – review and editing: all authors.

Appendix

When multiple traits follow a multivariate normal distribution, if the multiple traits are uncorrelated, they are independent.

Proof

We assume that a multiple traits vector $\mathbf{X} = (X_1, \dots, X_n)$ follows multiple normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ and variance $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$, the joint probability density function of $\mathbf{X}$ is:

$$\begin{aligned}
f_X(\mathbf{x}) &= (2\pi)^{-\frac{n}{2}} \left| \Sigma \right|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma (\mathbf{x} - \boldsymbol{\mu}) \right] \\
&= (2\pi)^{-\frac{n}{2}} (\sigma_1^2, \dots, \sigma_n^2)^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2} \right] \\
&= \prod_{i=1}^n \left[\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) \right] \\
&= \prod_{i=1}^n f_{x_i}(x_i).
\end{aligned}$$

The joint probability density function of $\mathbf{X}$ equals the multiplication of the density of each trait. Therefore, when $\mathbf{X} = (X_1, \dots, X_n)$ are jointly normal, if multiple traits are uncorrelated, they are independent.

Table A1. Type I error rates for rare variants in the presence of main effects (top panel) and in the absence of main effects (bottom panel) for $n = 2,000$, where the type I error rates outside parentheses are based on PCA and the type I error rates in parentheses are based on STD analysis according to Step 1

α level	TOW-GE	ISKAT	WSS	CMC
<i>With main effect</i>				
0.05	0.055 (0.055)	0.060 (0.058)	0.057 (0.054)	0.024 (0.028)
0.01	0.012 (0.013)	0.011 (0.010)	0.012 (0.014)	0.004 (0.005)
0.001	0.0017 (0.0016)	0.0015 (0.0017)	0.0015 (0.0013)	0.0005 (0.0004)
<i>Without main effect</i>				
0.05	0.054 (0.055)	0.060 (0.062)	0.053 (0.059)	0.052 (0.053)
0.01	0.011 (0.013)	0.016 (0.014)	0.009 (0.013)	0.010 (0.013)
0.001	0.0012 (0.0018)	0.0016 (0.0016)	0.0014 (0.0018)	0.0017 (0.0013)

For abbreviations, see footnote of Table 1.

Table A2. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for different sample sizes (n) at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome with main effect, where we set 50% of the β_{ij} s are positive, where the power outside parentheses is based on PCA and the power in parentheses is based on STD analysis

	$ \beta $	TOW-GE	ISKAT	WSS	CMC
<i>With main effect</i>					
$n = 1,000$	0.1	0.091 (0.086)	0.085 (0.073)	0.087 (0.080)	0.024 (0.026)
	0.2	0.153 (0.117)	0.148 (0.122)	0.148 (0.107)	0.066 (0.057)
	0.3	0.348 (0.235)	0.278 (0.221)	0.314 (0.233)	0.135 (0.079)
	0.4	0.574 (0.399)	0.444 (0.348)	0.464 (0.346)	0.253 (0.159)
	0.5	0.802 (0.611)	0.543 (0.470)	0.623 (0.490)	0.357 (0.235)
$n = 2,000$	0.1	0.091 (0.080)	0.109 (0.087)	0.093 (0.079)	0.050 (0.041)
	0.2	0.292 (0.203)	0.243 (0.179)	0.181 (0.131)	0.113 (0.069)
	0.3	0.589 (0.384)	0.471 (0.378)	0.355 (0.238)	0.261 (0.171)
	0.4	0.903 (0.729)	0.555 (0.493)	0.538 (0.420)	0.459 (0.298)
	0.5	0.986 (0.916)	0.627 (0.587)	0.673 (0.561)	0.646 (0.478)
$n = 3,000$	0.1	0.123 (0.097)	0.129 (0.107)	0.092 (0.082)	0.062 (0.049)
	0.2	0.385 (0.240)	0.357 (0.267)	0.179 (0.131)	0.145 (0.088)
	0.3	0.808 (0.601)	0.542 (0.474)	0.372 (0.274)	0.377 (0.232)
	0.4	0.982 (0.908)	0.640 (0.597)	0.582 (0.452)	0.648 (0.485)
	0.5	0.999 (0.990)	0.695 (0.644)	0.717 (0.604)	0.841 (0.701)
$n = 4,000$	0.1	0.141 (0.108)	0.154 (0.117)	0.093 (0.085)	0.062 (0.043)
	0.2	0.536 (0.339)	0.448 (0.354)	0.204 (0.148)	0.203 (0.137)
	0.3	0.936 (0.786)	0.630 (0.580)	0.406 (0.286)	0.468 (0.310)
	0.4	0.999 (0.981)	0.639 (0.588)	0.591 (0.457)	0.790 (0.634)
	0.5	1.000 (0.999)	0.750 (0.692)	0.767 (0.659)	0.916 (0.815)

For abbreviations, see footnote of Table 1.

Table A3. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for different sample sizes (n) at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome without main effect, where we set 50% of the β_{ij} s are positive, where the power outside parentheses is based on PCA and the power in parentheses is based on STD analysis

	$ \beta $	TOW-GE	ISKAT	WSS	CMC
<i>Without main effect</i>					
$n = 1,000$	0.1	0.076 (0.083)	0.099 (0.092)	0.063 (0.058)	0.049 (0.043)
	0.2	0.127 (0.099)	0.162 (0.117)	0.102 (0.092)	0.081 (0.076)
	0.3	0.323 (0.209)	0.307 (0.231)	0.203 (0.134)	0.174 (0.114)
	0.4	0.573 (0.394)	0.466 (0.366)	0.309 (0.227)	0.292 (0.210)
	0.5	0.810 (0.617)	0.528 (0.460)	0.429 (0.333)	0.383 (0.272)
$n = 2,000$	0.1	0.084 (0.073)	0.095 (0.079)	0.075 (0.071)	0.074 (0.069)
	0.2	0.283 (0.191)	0.254 (0.191)	0.150 (0.111)	0.164 (0.104)
	0.3	0.588 (0.389)	0.453 (0.346)	0.239 (0.166)	0.290 (0.214)
	0.4	0.912 (0.730)	0.581 (0.519)	0.394 (0.295)	0.461 (0.344)
	0.5	0.991 (0.921)	0.659 (0.612)	0.532 (0.405)	0.606 (0.458)
$n = 3,000$	0.1	0.117 (0.092)	0.127 (0.098)	0.081 (0.083)	0.092 (0.070)
	0.2	0.391 (0.236)	0.375 (0.277)	0.149 (0.111)	0.198 (0.136)
	0.3	0.819 (0.605)	0.546 (0.467)	0.296 (0.211)	0.403 (0.283)
	0.4	0.986 (0.915)	0.646 (0.613)	0.479 (0.354)	0.629 (0.489)
	0.5	1.000 (0.996)	0.720 (0.657)	0.612 (0.497)	0.801 (0.693)
$n = 4,000$	0.1	0.142 (0.113)	0.153 (0.110)	0.087 (0.074)	0.092 (0.079)
	0.2	0.546 (0.344)	0.425 (0.327)	0.167 (0.129)	0.245 (0.180)
	0.3	0.952 (0.803)	0.632 (0.571)	0.344 (0.237)	0.502 (0.369)
	0.4	0.999 (0.980)	0.689 (0.622)	0.501 (0.387)	0.765 (0.632)
	0.5	1.000 (1.000)	0.792 (0.718)	0.668 (0.541)	0.878 (0.784)

For abbreviations, see footnote of Table 1.

Fig. A1. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome when there are no main effects based on PCA. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A2. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome when there are no main effects based on STD analysis. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A3. Power comparisons for PCA and STD analysis of the test TOW-GE for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by environment interaction effects on a continuous outcome with no main effect. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A4. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are main effects based on PCA for $p = 0.2$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A5. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are main effects based on PCA for $p = 0.8$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A6. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are no main effects based on PCA for $p = 0.2$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A7. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are no main effects based on PCA for $p = 0.8$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A8. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are main effects based on STD analysis for $p = 0.2$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A9. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are main effects based on STD analysis for $p = 0.8$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A10. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are no main effects based on STD analysis for $p = 0.2$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A11. Power comparisons of the four tests (TOW-GE, ISKAT, WSS, and CMC) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing rare variant by gene environment interaction effects on a continuous outcome when there are no main effects based on STD analysis for $p = 0.8$. **a, b** 2 non-zero β s. **c, d** 6 non-zero β s. **e, f** 8 non-zero β s. **a, c, e** 50% of the β s are positive; 50% of the β s are negative. **b, d, f** 100% of the β s are positive; 0% of the β s are negative.

Fig. A12. Power comparisons of five tests (TOW-GE, ISKAT, WSS, CMC, and VW-TOW-GE) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing both rare and common variant by environment interaction effects on a continuous outcome for $p = 0.2$. **a, b** PCA. **c, d** STD analysis. **a, c** With main effect. **b, d** Without main effect.

Fig. A13. Power comparisons of five tests (TOW-GE, ISKAT, WSS, CMC, and VW-TOW-GE) for $n = 2,000$ at the $\alpha = 0.05$ level of significance for testing both rare and common variant by environment interaction effects on a continuous outcome for $p = 0.8$. **a, b** PCA. **c, d** STD analysis **a, c** With main effect. **b, d** Without main effect.

References

- Wang X, Liu W, Sun CL, Armenian SH, Hakonarson H, Hageman L, et al. Hyaluronan synthase 3 variant and anthracycline-related cardiomyopathy: a report from the children's oncology group. *J Clin Oncol*. 2014 Mar; 32(7):647–53.
- Wang X, Sun CL, Quiñones-Lombraña A, Singh P, Landier W, Hageman L, et al. CELF4 variant and anthracycline-related cardiomyopathy: a children's oncology group genome-wide association study. *J Clin Oncol*. 2016 Mar;34(8):863–70.
- Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al.; Wellcome Trust Case Control Consortium. Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet*. 2009 Jun;41(6):666–76.
- Zabaneh D, Balding DJ. A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS One*. 2010 Aug; 5(8):e11961.
- Yang Q, Wang Y. Methods for analyzing multivariate phenotypes in genetic association studies. *J Probab Stat*. 2012; DOI: 10.1155/2012/652569.
- Zhang JJ, Zhao ZH, Guo X, Guo B, Wu BL. Powerful statistical method to detect disease-associated genes using publicly available genome-wide association studies summary data. *Genet Epidemiol*. 2019 Dec;43(8):941–51.
- Singh A, Babyak MA, Nolan DK, Brummett BH, Jiang R, Siegler IC, et al. Gene by stress genome-wide interaction analysis and path analysis identify EBF1 as a cardiovascular and metabolic risk gene. *Eur J Hum Genet*. 2015 Jun;23(6):854–62.
- Andrés AM, Clark AG, Shimmin L, Boerwinkel E, Sing CF, Hixson JE. Understanding the accuracy of statistical haplotype inference with sequence data of known phase. *Genet Epidemiol*. 2007 Nov;31(7):659–71.
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011 Jul;89(1): 82–93.
- Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 2008 Sep;83(3):311–21.
- Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*. 2009 Feb; 5(2):e1000384.
- Sha Q, Wang X, Wang X, Zhang S. Detecting association of rare and common variants by testing an optimally weighted combination of variants. *Genet Epidemiol*. 2012 Sep;36(6): 561–71.
- Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics*. 2013 Sep;14(4):667–81.
- Lin X, Lee S, Wu MC, Wang C, Chen H, Li Z, et al. Test for rare variants by environment interactions in sequencing association studies. *Biometrics*. 2016 Mar;72(1):156–64.
- Wu B, Pankow JS. Sequence kernel association test of multiple continuous phenotypes. *Genet Epidemiol*. 2016 Feb;40(2):91–100.
- Ferreira MA, Purcell SM. A multivariate test of association. *Bioinformatics*. 2009 Jan; 25(1):132–3.
- Wu B, Pankow JS. Statistical methods for association tests of multiple continuous traits in genome-wide association studies. *Ann Hum Genet*. 2015 Jul;79(4):282–93.
- Sha Q, Zhang Z, Zhang S. An improved score test for genetic association studies. *Genet Epidemiol*. 2011 Jul;35(5):350–9.
- Fisher RA. Statistical methods for research workers. In: Kotz S, Johnson NL, editors. Breakthroughs in statistics. Springer series in statistics (Perspectives in statistics). New York: Springer; 1992. p. 66–70.
- Yang X, Wang S, Zhang S, Sha Q. Detecting association of rare and common variants based on cross-validation prediction error. *Genet Epidemiol*. 2017 Apr;41(3):233–43.
- Murphy TF, Sethi S. Chronic obstructive pulmonary disease: role of bacteria and guide to antibacterial selection in the older patient. *Drugs Aging*. 2002;19(10):761–75.
- Sandford AJ, Silverman EK. Chronic obstructive pulmonary disease. 1: susceptibility factors for COPD the genotype-environment interaction. *Thorax*. 2002 Aug;57(8):736–41.
- Hersh CP, Demeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol*. 2005 Jul;33(1):71–8.
- Celedón JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, et al. The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Hum Mol Genet*. 2004 Aug; 13(15):1649–56.
- Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic epidemiology of COPD (COPDGene) study design. *COPD*. 2010 Feb;7(1):32–43.
- Chu JH, Hersh CP, Castaldi PJ, Cho MH, Raby BA, Laird N, et al. Analyzing networks of phenotypes in complex diseases: methodology and applications in COPD. *BMC Syst Biol*. 2014 Jun;8(1):78.
- Han MK, Kazerooni EA, Lynch DA, Liu LX, Murray S, Curtis JL, et al.; COPDGene Investigators. Chronic obstructive pulmonary disease exacerbations in the COPDGene study: associated radiologic phenotypes. *Radiology*. 2011 Oct;261(1):274–82.
- Whelton PK, Carey RM, Aronow WS, Casey DE Jr, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2018 May;71(19):e127–248.