Published in final edited form as:

Stat Med. 2019 March 30; 38(7): 1230–1244. doi:10.1002/sim.8037.

A Powerful and Data-adaptive Test for Rare-variant-based Gene-environment Interaction Analysis

Tianzhong Yang^{1,2}, Han Chen^{3,4}, Hongwei Tang⁵, Donghui Li⁵, and Peng Wei^{1,*}

¹Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

²Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, TX 77030, USA

³Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, TX 77030, USA

⁴Center for Precision Health, School of Public Health and School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

⁵Departments of Gastrointestinal Medical Oncology and Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Abstract

As whole-exome/genome sequencing data become increasingly available in genetic epidemiology research consortia, there is emerging interest in testing the interactions between rare genetic variants and environmental exposures that modify the risk of complex diseases. However, testing rare-variant-based gene-by-environment interactions (GxE) is more challenging than testing the genetic main effects due to the difficulty in correctly estimating the latter under the null hypothesis of no GxE effects and the presence of neutral variants. In response, we have developed a family of powerful and data-adaptive GxE tests, called “aGE” tests, in the framework of the adaptive powered score test, originally proposed for testing the genetic main effects. Using extensive simulations, we show that aGE tests can control the type I error rate in the presence of a large number of neutral variants or a nonlinear environmental main effect, and the power is more resilient to the inclusion of neutral variants than that of existing methods. We demonstrate the performance of the proposed aGE tests using Pancreatic Cancer Case-Control Consortium Exome Chip data. An R package “aGE” is available at <http://github.com/ytzhong/projects/>.

Keywords

Data-adaptive hypothesis testing; Rare variant; Gene-environment interaction; Model misspecification

*Correspondence to: Peng Wei, Department of Biostatistics, The University of Texas MD Anderson Cancer Center, 1400 Pressler St, Pickens Tower, FCT4.6044, Houston, Texas 77030, USA. pwei2@mdanderson.org.

1. Introduction

Complex diseases are likely to be the consequence of the interplay between genetic and environmental factors. Identification of gene-by-environment interaction (GxE) not only contributes to finding novel genetic loci, but also improves our understanding of disease mechanisms. A large number of genome-wide association studies (GWAS) have identified thousands of genetic loci that are associated with complex diseases; however, heritability for a given complex disease cannot yet be fully explained by those loci. The broad-sense heritability model supports that gene-by-gene interactions, GxE, and epigenetic factors contribute to the so-called “missing heritability” problem.^{1–3} Indeed, novel loci have been identified with different complex diseases through GxE analysis, underscoring its important role in disease/trait gene discovery.^{4,5} In addition, environmental exposure, as a modifiable risk factor, can significantly impact the relationship between genetic variants and disease risk.⁶ GxE analysis serves as a starting point to understand the underlying biological mechanisms of complex diseases, which can be helpful for prevention and early detection of disease.^{7–11} One GxE example in alcohol-related cancer risk has wide acceptance among researchers and robust replication of findings: Wu et al. reported that single nucleotide polymorphisms (SNPs) in the alcohol metabolism pathway, including the encoding aldehyde dehydrogenase-2 (ALDH2) gene and alcohol dehydrogenase gene cluster, interact with alcohol intake to promote the risk of esophageal cancer.^{4,7,12,13} Brooks et al. suggested that a great number of esophageal cancer cases might be prevented if ALDH2-deficient heavy drinkers reduced alcohol consumption.¹²

As next-generation sequencing (NGS) data, including whole-exome and whole-genome sequencing data, become increasingly available in large genetic epidemiology research consortia, more researchers have shifted their attention from common variants (CVs) with minor allele frequency (MAF) larger than 5% in GWAS to rare variants (RVs) with MAF less than 5% or 1%.^{2,14,15} RV analysis is complementary to CV analysis in terms of identifying novel genes and locating causal variants.^{16,17} Single-variant-based association tests are commonly used in CV analysis of GWAS, yet severe power loss occurs in RV analysis due to the low MAF. To date, many statistical methods have been proposed to increase the statistical power by aggregating RVs in a set, for example, grouping RVs in a gene, pathway, or any functional region. The power gain of such set-based methods is mainly owing to the decrease in the multiple testing burden and the genetic architecture, where multiple RVs in the set are associated with the disease/trait of interest. Nevertheless, the power gain of these methods usually relies on the true yet rarely known relationship between each RV and the given disease/trait. A well-known category of set-based methods is the burden test. Burden tests collapse RVs into genetic scores, which are most powerful when all RVs in the set are causal and have effects in the same direction.^{18,19} Another category of RV tests is variance-component tests, which assume that the genetic effects follow a distribution with both positive and negative directions. They are more powerful than the burden tests in the presence of RVs with opposite effects or non-associated neutral variants.^{20–22} In addition, researchers have developed tests that combine burden and variance-component tests to gain power across broader association patterns.²³ The adaptive sum of powered score (aSPU) test is one of such combination tests that has shown to be

powerful and adaptive to varying genetic architecture, particularly when there is a large proportion of neutral variants.²⁴ All of the aforementioned set-based methods have focused on the genetic main effects rather than the GxE effects; however, limited work has been done on testing for rare-variant-based GxE analysis.

GxE analysis involves unique challenges. First, the genetic main effects need to be estimated in the model under the null hypothesis of no GxE. While the score test for the main effect analysis only entails fitting a null model with a small number of covariates, for which the maximum likelihood estimation (MLE) method suffices, MLE has been shown to lead to poor estimation of the genetic main effects in the null model for GxE, resulting in an inflated type I error rate.^{25,26} In addition, as NGS technology captures genetic variants in essentially single-base resolution, more RVs are likely to be included in the test sets, which further increases the difficulty in model estimation. rareGE and iSKAT are two state-of-the-art methods that deal with this problem, under the variance component test framework.^{25,27} They utilize a bias-variance trade-off to obtain more stable estimates of the genetic main effects in the null model, but shrink the parameter estimates differently. rareGE uses a generalized linear mixed model (GLMM) and treats the genetic main effects as random effects, whereas iSKAT uses weighted ridge regression coupled with a generalized cross-validation method to select the tuning parameter. Recently, Coombes et al. proposed a CV-by-environment interaction test that employs ridge regression and is similar to iSKAT.²⁸ Second, there could be substantial power loss if a large number of neutral RVs are included in GxE analysis. This problem also exists in tests of the genetic main effects, but it is even more challenging in the GxE analysis. There might be far fewer RVs interacting with the environmental variable than associated with the disease/trait of interest only through the main effects. Therefore, it is worth evaluating how the existing methods perform when the RV set is large and includes a large number of neutral RVs. Third, the inclusion of environmental variables into genetic epidemiology studies raises additional challenges in parametric modeling, which features parsimonious model assumptions. For example, a linear relationship between the environmental variable and the phenotype is frequently assumed. However, when the true relationship is beyond linear, model misspecification is likely to occur. Strangely behaving quantile-quantile (QQ) plots have been reported for testing single-variant-by-environment interaction due to misspecification of the main effect of the environmental variable.^{29–31} Set-based GxE tests are likely to encounter similar problems. In our motivating data example (detailed later), pack-years of cigarette smoking have a nonlinear relationship with the risk of developing pancreatic cancer. Previous studies categorized subjects into non-smokers, light smokers, or heavy smokers based on pack-years of cigarette smoking. However, we found that the relationship between pack-years of smoking and cancer risk is more complicated than these three categories can capture, thus such simple categorization may lose much information in smoking exposure and may lead to loss of power. Therefore, we should fully utilize the continuous environmental variable of interest, e.g., pack-years, and at the same time, avoid model misspecification to maintain the type I rate. In summary, powerful and robust methods for testing GxE tailored for RVs are in need.

To fill the gaps, we propose a family of tests called adaptive gene-environment interaction (aGE) tests, with a focus on case-control studies. Our method is developed under the

framework of the aSPU test, which maintains high power when testing the genetic main effects in the presence of a large number of neutral RVs.²⁴ We show that inflated type I error rates can arise when the model is misspecified and provide a spline-based approach to tackle this problem. In addition, we have developed an aGE joint test to gain extra power across a broader set of penetrance models.³²

The rest of the manuscript is organized as follows. We review the aSPU test and present the new aGE method in Section 2. We present a simulation study to demonstrate the operating characteristics of the proposed method in Section 3, followed by application to a case-control study of pancreatic cancer to identify biological pathway-by-smoking interactions in Section 4. We conclude with a short discussion and future directions in Section 5.

2. Methods

2.1. Notations

Let y_i be the phenotype of individual i ; $y = [y_1, y_2, \dots, y_n]^T$, and n is the number of subjects. Here, we assume y_i is binary in the case-control study design, though continuous phenotypes can be equally accommodated within the same framework. $\mathbf{X}_i$ is the vector of covariates for individual i with length $p+2$, including the intercept and the environmental variable E_i ; $\mathbf{X}_i = [1, E_i, X_{i1}, X_{i2}, \dots, X_{ip}]$. $\mathbf{G}_i$ is the genotype vector of length q for individual i : $\mathbf{G}_i = [G_{i1}, G_{i2}, \dots, G_{iq}]$, where G_{ik} is the number of minor alleles of the k^{th} RV in a set for the i^{th} individual. $\mathbf{S}_i$ is a vector of the product of $\mathbf{G}_i$ and E_i : $\mathbf{S}_i = [G_{i1} E_i, G_{i2} E_i, \dots, G_{iq} E_i]$. In addition, we assume $\mathbf{X}, \mathbf{G}$ and $\mathbf{S}$ are $n \times (p+2)$, $n \times q$ and $n \times q$ matrices that contain all the subjects, and $\mathbf{E}$ is the vector of E_i of all the individuals: For example, $\mathbf{G} = [\mathbf{G}_1^T, \mathbf{G}_2^T, \dots, \mathbf{G}_n^T]^T$ and $\mathbf{S} = [\mathbf{S}_1^T, \mathbf{S}_2^T, \dots, \mathbf{S}_n^T]^T$. In this paper, we assume the subjects are independent. We denote parameters without hat as true values and parameters with hat as estimated values.

2.2. Review of the aSPU test

The aSPU test has been shown to have superior performance in testing the genetic main effects under a wide range of association patterns, including varying association directions, effect sizes, and the presence of a large number of neutral variants. Consider the following generalized linear model:

$$g(\mu_i) = \mathbf{X}_i \boldsymbol{\beta}_0 + \mathbf{G}_i \boldsymbol{\beta}_1,$$

where $g(\cdot)$ is the link function, μ_i is the mean of y_i , $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ ($\boldsymbol{\beta}_1 = [\beta_{11}, \dots, \beta_{1q}]^T$) are the coefficients of the covariates and the genetic main effects, respectively. Since the aSPU test focuses on testing the genetic main effects, the corresponding null hypothesis is $H_0: \beta_{1k} = 0$ for $k=1, \dots, q$. The alternative hypothesis is H_1 : at least one element of $\boldsymbol{\beta}_1$ is not equal to 0.

Briefly, the aSPU test chooses the most powerful test among a family of sum of powered score (SPU) tests, including the sum (burden) test and the sequence kernel machine

association test (SKAT) test/sum of squared test (SSU).^{20,24} The test statistic of an SPU test of positive integer power $\gamma \geq 1$ (SPU(γ)) can be written as

$$T_{SPU(\gamma)} = \sum_{k=1}^q U_k^\gamma$$

with q RVs in the set. $U_k = G_k^T(y - \mu)$ is the score function of the k^{th} RV. When $\gamma=1$, the SPU(1) test uses constant **1** as a weight and sums the individual score function component U_k of all the RVs in the set of interest, equivalent to the Sum test or burden test; when $\gamma=2$, the SPU(2) test uses U as the weight to itself and is equivalent to variance-component test SKAT/SSU; when γ keeps increasing, the SPU(γ) test puts higher weights on the k^{th} RV with larger $|U_k|$, while gradually decreasing the weights of other RVs with smaller $|U_k|$. As a large value of $|U_k|$ indicates strong association strength of the k^{th} RV with the phenotype, and a small value of $|U_k|$ indicates weak or no association of the k^{th} RV with the phenotype, a larger γ tends to put more weight on the informative RVs. According to the definition of p-norm, we have $\|U\|_\gamma = (\sum_{k=1}^q |U_k|^\gamma)^{\frac{1}{\gamma}}$, which is a monotone increasing function of $T_{SPU(\gamma)}$ for even γ . As $\gamma \rightarrow \infty$,

$$\|U\|_\infty = \max_{k=1}^q |U_k|.$$

Thus, SPU(∞) test is closely related to the minimum p-value test, which only considers the RV with the most extreme test statistic in the set, except that the variance of each score component is approximated by a constant. Since there is no uniformly most powerful test, different γ may have varying performance. By adaptively choosing the most powerful test in the family of SPU tests, the aSPU test obtains high power across a wide range of scenarios. Of note, the aSPU test gains its adaptiveness at the cost of some power loss due to looking at multiple γ 's, compared to the most powerful test under a given scenario; however, the power loss is usually minor. Practically, γ taking values in $\Gamma = \{1, 2, 3, \dots, \infty\}$ suffice to achieve optimal power in many applications. The multiple p-values are combined as follows:

$$T_{aSPU} = \min_{\gamma \in \Gamma} P_{SPU(\gamma)}$$

where $P_{SPU(\gamma)}$ is the p-value of $T_{SPU(\gamma)}$. Since T_{aSPU} is no longer a genuine p-value, permutation or parametric bootstrap is needed to obtain the final p-value for the aSPU test.

2.3. New method: aGE interaction test

To perform RV-set-based GxE analysis, we model the environmental exposure, RVs and their interactions via the following full model:

$$g(\mu_i) = X_i\beta_0 + G_i\beta_1 + S_i\beta_2, \quad (1)$$

where β_0 , β_1 , and β_2 ($\beta_2 = [\beta_{21}, \dots, \beta_{2q}]^T$) are the coefficients of $\mathbf{X}$, $\mathbf{G}$, and $\mathbf{S}$, $g(\cdot)$ is the logit link function, and $\mu_i = E(y_i)$. The null hypothesis for the interaction test is $H_0: \beta_{2k} = 0$ for $k = 1, \dots, q$, and the alternative hypothesis is H_1 : at least one element of β_2 not equal to 0. The model under H_0 or the null model is:

$$g(\mu_i) = X_i\beta_0 + G_i\beta_1, \quad (2)$$

where μ_i is estimated in the GLMM framework under model (2) by treating covariate effects β_0 as fixed effects and the genetic main effects β_1 as independent random effects that share a common variance. Namely, $\beta_1 \sim N(0, \tau\mathbf{I})$ with $\mathbf{I}$ as a $q \times q$ identity matrix. The random effects approach has been used for modeling high-dimensional genetic variants.^{25,33} Based on model (2), we derive the asymptotic distribution of our score function. Briefly, let $\mu = [\mu_1, \mu_2, \dots, \mu_n]^T$, the score function of the covariates $U_X = \mathbf{X}^T(y - \mu)$, the score function of the interaction term $U_S = \mathbf{S}^T(y - \mu)$, and

$$\mathbf{A} = \frac{1}{n} \begin{pmatrix} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{V}^{-1} \mathbf{S} \\ \mathbf{S}^T \mathbf{V}^{-1} \mathbf{X} & \mathbf{S}^T \mathbf{V}^{-1} \mathbf{S} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

with $\mathbf{V} = \tau\mathbf{G}\mathbf{G}^T + \mathbf{W}^{-1}$ and $\mathbf{W} = \text{diag}(\mu_1(1 - \mu_1), \dots, \mu_n(1 - \mu_n))$. By the central limit theorem and using Taylor expansion, it can be shown that

$$\frac{1}{\sqrt{n}} U_S \xrightarrow{d} N(0, \Sigma_A), \quad \text{where } \Sigma_A = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}.$$

Variance parameters are estimated using the standard restricted maximum likelihood (REML) method. Details of the derivation are provided in Supplemental Materials Section A.1. Then we construct the aGE test based on the score function by defining

$$T_{GE(\gamma)} = \sum_{k=1}^q U_{Sk}^\gamma, \quad (3)$$

where $\gamma - 1$ is a positive integer, and U_{Sk} is the score function of the interaction term for the k^{th} RV. The test statistic of $GE(1)$ equals $\sum_{k=1}^q U_{Sk}$; the test statistic of $GE(2)$ equals $\sum_{k=1}^q U_{Sk}^2$, which has a close relationship with rareGE. Although rareGE and $GE(2)$ are derived from two different perspectives, both of them use GLMM to estimate the null model, thus rareGE and $GE(2)$ have the same asymptotic distribution under the null, i.e., a mixture of chi-squared distribution.²⁵ The test statistic of rareGE is calculated by taking derivatives of log-quasilikelihood expansion with respect to the variance component; however, the test statistic of $GE(\gamma)$ is drawn from the derivative of the log of the marginal likelihood regarding β_2 under model (1). Our aGE interaction test statistic is based on combining $T_{GE(\gamma)}$ by either taking the minimum of the p-values (minP) across Γ or using Fisher's meta-analysis-like approach:

$$T_{aGE} = \min_{\gamma \in \Gamma} P_{GE(\gamma)},$$

$$T_{aGE_{Fisher}} = \prod_{\gamma \in \Gamma} P_{GE(\gamma)},$$

where $T_{GE(\gamma)}$ is the p-value of the $GE(\gamma)$ test. The procedure for calculating the p-value for aGE is provided in Section 2.5. We search γ in the set of $\Gamma = \{1, 2, 3, 4, 5, 6\}$ for both the aGE interaction and joint tests. We decided not to include infinity in Γ because it was found to lead to inflated type I error rates under certain scenarios.³⁴

2.4. New method: aGE joint test

When both the genetic main effect and interaction effect exist, a joint test may have higher power than either the interaction effect test or the genetic main effect test alone.³² In addition, it is computationally simpler than the interaction test. The full model remains the same as in the interaction test (model (1)). The null model is

$$g(\mu_i) = X_i \beta_0. \quad (4)$$

Accordingly, the null hypothesis for the joint test is $H_0 : \beta_1 = \beta_2 = 0$ and the alternative hypothesis is H_1 : at least one element of β_1 or β_2 is not 0. The corresponding score vector is $U_{JGS} = [G, S]^T (y - \mu)$ where μ is estimated through the MLE because the null model (4) no longer consists of the sparse and potentially large genotype matrix. Using derivations similar to those for the genetic main effect test,²⁴ we conclude that

$$\frac{1}{\sqrt{n}} U_{JGS} \xrightarrow{d} N(0, \Sigma_B),$$

where

$$\Sigma_B = B_{22} - B_{21} B_{11}^{-1} B_{12},$$

and

$$B = \frac{1}{n} \begin{pmatrix} X^T V^{-1} X & X^T V^{-1} [G, S] \\ [G, S]^T V^{-1} X & [G, S]^T V^{-1} [G, S] \end{pmatrix} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

with $V^{-1} = W = \text{diag}(\mu_1(1 - \mu_1), \dots, \mu_n(1 - \mu_n))$. We separate the sum of the powered score test by the genetic main effect test and interaction effect test: let $U_{JS} = S^T (y - \mu)$ be the score vector of the interaction effects and $U_{JG} = G^T (y - \mu)$ be the score vector of the main effects. Since U_{JS} and U_{JG} can be on different scales, the aGE joint test statistic is combined on the p-value scale in the following two ways:

$$T_{aGEjoint} = \min \left[\min_{\gamma \in \Gamma} \left(P_{U_{JS}(\gamma)}, P_{U_{JG}(\gamma)} \right) \right], \quad (5)$$

$$T_{aGEjoint_{Fisher}} = \prod \min_{\gamma \in \Gamma} \left(P_{U_{JS}(\gamma)}, P_{U_{JG}(\gamma)} \right), \quad (6)$$

where $P_{U_{JS}(\gamma)}$ and $P_{U_{JG}(\gamma)}$ are the p-values for T_{JS} and T_{JG} respectively.

2.5. P-value calculation

Since the minimum p-value and the combined p-value by Fisher's method are no longer genuine p-values, we employ Monte Carlo simulation method to obtain the null distribution and calculate the p-values for the aGE interaction and joint tests. The specific steps for the interaction test are as follows.

(1) Estimate null model (2) by the GLMM and simulate $U_S^d \sim N(0, \Sigma_A)$ for $d = 1, 2, \dots, D$ times;

(2) Calculate $T_{GE(\gamma)}^d$ following equation (3);

(3) Calculate P_{aGE} :

(3a) Calculate $P_{GE(\gamma)}$ for each γ :

$$P_{GE(\gamma)}^d = \sum_{d_1 \neq d}^D \frac{I \left(\left| T_{GE(\gamma)}^d \right| \leq \left| T_{GE(\gamma)}^{d_1} \right| \right) + 1}{D}.$$

(3b) Calculate T_{aGE} and T_{aGE}^d :

$$T_{aGE}^d = \min_{\gamma \in \Gamma} P_{GE(\gamma)}^d$$

$$T_{aGE_{Fisher}} = \prod_{\gamma \in \Gamma} P_{GE(\gamma)}^d$$

(3c) Calculate P_{aGE} :

$$P_{aGE} = \sum_{d=1}^D \frac{I \left(T_{aGE}^d \leq T_{aGE} \right) + 1}{D + 1}.$$

The p-value of the joint test is calculated as follows.

(A) Estimate null model (4) by MLE and simulate U_{JS}^d and U_{JG}^d for $d = 1, 2, \dots, D$ times;

(B) Following steps (2) and (3a), calculate $P_{U_{JS}(y)}$ and $P_{U_{JG}(y)}$; (C) Calculate

$T_{aGEjoint}^d, T_{aGEjoint}^d, T_{aGEjoint_{Fisher}}^d$ and $T_{aGEjoint_{Fisher}}^d$ following equation (5) and equation (6); then calculate $P_{aGEjoint}$ as in step (3c).

D is chosen to be 1,000 initially. It can be increased to a larger number when needed, for example, to achieve the genome-wide significance level in a real data application.

Alternatively, the parametric bootstrap procedure can be used to calculate the p-value for the interaction test. Specifically, we fit null model (2) and obtain $\hat{\mu}$. Second, we simulate a new set of traits $Y_i^d \sim \text{Bernoulli}(\hat{\mu}_i)$ D times and fit null model (2) based on each set of simulated Y^d to calculate the score function U_S^d . Third, we follow steps (2)-(3) to calculate the corresponding p-values. Additionally, the parametric bootstrap and residual permutation are valid methods for obtaining p-values for the joint test. For the specific procedures, we refer to Pan et al.²⁴ Since the parametric bootstrap requires fitting the null model multiple times, it is computationally intensive, especially for the interaction test. On the other hand, residual permutation, which requires fitting the data only once, can control the type I error rate for the joint test, not for the interaction test according to our preliminary simulation results (not shown). Therefore, we used Monte Carlo simulation method throughout as a unified and computationally efficient method. The accuracy of this method depends on the asymptotic distribution of the score function, for which we show its validity later.

2.6. Model the environmental main effect using splines

In many GxE applications, the environmental exposure is collected as a continuous variable, such as body mass index or pack-years of cigarette smoking in the application herein. Simple dichotomization of these continuous variables, e.g., as obese/non-obese and ever/never smokers, while allowing for parsimonious statistical modeling, may lose critical information in the exposure and thus lead to loss of power. On the other hand, it is known that inadequately modeling the main effect of a continuous exposure can result in inflated type I error rates in single-CV-based GxE testing.^{29–31} As to be illustrated in our simulations and real data application, misspecification of the environmental main effect in GxE testing not only leads to inflated type I error rates for the individual CV, but also for the RV set. Here, we propose to employ splines to allow for flexible modeling of the environmental main effect in the aGE test. Commonly used splines include cubic splines and linear splines. Given that the knot is known, the splines can be expressed as

$$f(E) = \sum_{k=1}^h \delta_k e_k(t) + \sum_{l=1}^r a_l R(t, t_l^0),$$

where t represents the knot, usually set at the quantiles of the distribution of the environmental variable. $e_1(t) = E$ and $R(t, t_j)$ can be regarded as residual terms.

We adopt a simple basis following Wahba and Gu for cubic splines:^{36,37}

$$R(t, t_l^0) = \left[(t_l^0 - 1/2)^2 - 1/12 \right] \left[(t - 1/2)^2 - 1/12 \right] / 4 - \left[(|t - t_l^0| - 1/2)^4 - 1/2 (|t - t_l^0| - 1/2)^2 + 7/240 \right] / 24$$

We assume that the true model has a flexible form: $\text{logit}(\mu_i) = \mathbf{X}_i \boldsymbol{\beta}_{01} + f(E_i) \boldsymbol{\beta}_{02} + \mathbf{G}_i \boldsymbol{\beta}_1 + \mathbf{G}_i g(E_i) \boldsymbol{\beta}_2$, where E_i is centered to avoid an identifiability problem and $g(E_i)$ can be of any functional form. The null hypothesis of interest is $H_0 : \boldsymbol{\beta}_2 = 0$ for the interaction test and $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 0$ for the joint test. For higher power, we use a working model with a parsimonious interaction term: $\text{logit}(\mu_i) = \mathbf{X}_i \tilde{\boldsymbol{\beta}}_{01} + f(E_i) \tilde{\boldsymbol{\beta}}_{02} + \mathbf{G}_i \tilde{\boldsymbol{\beta}}_1 + \mathbf{G}_i E_i \tilde{\boldsymbol{\beta}}_2$. Type I error rate can be controlled for both the interaction test and joint test by using parsimonious modeling because if $\boldsymbol{\beta}_2 = 0$ then $\tilde{\boldsymbol{\beta}}_2 = 0$, and if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 0$ then $\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_2 = 0$. If $g(E)$ is far from linear, then the power could be compromised at the cost of over-simplified modeling. We could further penalize $R(t_i, t_j)$ to obtain smoother splines by assuming $a_l \sim \mathcal{N}(0, \phi_2 \mathbf{I})$ and model with two random components under the null. However, we found that this approach could have an identifiability issue if G and E are dependent. Based on our preliminary experiments, we chose to fit splines with no penalization. We show later that overfitting the null model does not influence the power of the GxE test. In fact, He et al. modeled the continuous environmental variable using a similar nonparametric method (the method of “sieve”) in GxE analysis, but in the context of longitudinal CV-based analysis.³⁸

3. Simulation Study

3.1. Simulation settings

We simulated case-control studies with case-control ratio 1:1. Genotypes were simulated as in Wang and Elston, with two linkage disequilibrium (LD) blocks for each replication.³⁹ The first block consisted of 8 causal RVs (csRVs). The second block was independent of the first block, consisting of different numbers of neutral RVs (0, 50, 100, 200). The RVs within each block had an AR1 correlation structure with $\rho = 0.5$. The MAF of RVs ranged from 0.1% to 5%. We also simulated three covariates: age $\sim \mathcal{N}(55, 5.7^2)$, sex $\sim \text{Bernoulli}(0.5)$, and height $\sim \mathcal{N}(169, 9.4^2)$.

The type I error rate was evaluated by 10,000 replications with no GxE interactions. Simulated data followed the model below for evaluation of the type I error rate:

$$\text{logit}(\mu) = 0.5 \times \text{sex} + 0.05 \times \text{age} + 0.1 \times \text{height} + f(E) + \sum_{k=1}^8 \beta_{1k} \text{csRV}_k.$$

The genetic main effects had odds ratios (ORs) randomly drawn from a uniform distribution, i.e., $OR_1 = \exp(\beta_1), \exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(0, 2)$. The four scenarios listed below were assessed for 3,500 cases and 3,500 controls.

Setting 1: E is correctly specified and G is independent of E ($G \perp E$): $E \sim \mathcal{N}(0, 1)$, $f(E) = 0.25E$;

Setting 2: E is misspecified and G is independent of E ($G \perp E$): $E \sim N(0, 1)$, $f(E) = 0.25E + 0.5E^2$;

Setting 3: E is correctly specified and $G \not\perp E$: $E = 0.5 \sum_{k=1}^q csRV + u$, $u \sim N(0, 1)$, $f(E) = 0.25E$;

Setting 4: E is misspecified and $G \not\perp E$: $E \sim N(0, 1)$, $f(E) = 0.25E + 0.5E^2$.

In addition, type I error rates for 500 cases and 500 controls were assessed under these four settings because the model performance relies on the asymptotic distribution of the score function for binary outcomes and our application example has a total sample size of 1,000. Additional simulation using a symmetric main effect distribution ($\beta_{1k} \stackrel{i.i.d.}{\sim} N(0, 1)$) was performed for the interaction test in setting 1 to evaluate the performance of different estimation methods, i.e., GLMM vs ridge regression.

Power for the interaction tests was analyzed under setting 1 and setting 2 for 1,000 replications. The model for power evaluation was

$$\text{logit}(\mu) = 0.5 \times \text{sex} + 0.05 \times \text{age} + 0.1 \times \text{height} + f(E) + \sum_{k=1}^8 \beta_{1k} csRV + \sum_{k=1}^8 \beta_{2k} csRV \times f(E). \quad (7)$$

We evaluated the power when the interaction effects were in different directions by setting $OR_2 = \exp(\beta_2)$ and $\exp(\beta_2) = [\exp(\beta_{21}), \dots, \exp(\beta_{28})]^T = [3, 1/3, 3, 3, 2, 1/3, 1/3, 1/2]^T$ and in the same direction by drawing $\exp(\beta_{2k}) \stackrel{i.i.d.}{\sim} U(1/3, 3)$ and then taking absolute value of β_{2k} .

Power for the joint tests was evaluated under the same model as that for the interaction tests. The interaction effects were held across all the scenarios, i.e., $(\exp(\beta_{2k}) \stackrel{i.i.d.}{\sim} U(1/3, 3))$, while the magnitude of the genetic main effect sizes varied, i.e., no main effects (all β_{1k} equal to 0), moderate main effects $(\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/2, 2))$, or strong main effects $(\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/3, 3))$. Similarly, 3,500 cases and 3,500 controls were simulated for each replication. We further examined the power of strong main effects for 500 cases/controls and 5,000 cases/controls in Supplementary Materials A.2.1.

We considered rareGE interaction and joint, iSKAT, aGE interaction and joint tests, as well as the multivariate score test for joint effects, in the comparison. We used cubic splines to model the environmental variable for the aGE tests under different settings, denoted as $\cdot_{sp}$, such as aGE_{sp} and aGEjoint_{sp}. Without subscript sp , the tests assumed a linear relationship.

Moreover, we denoted $\cdot_{true}$ as the tests using the correct functional form of the environmental variable. Throughout the simulation study, we fixed the test significance level at $\alpha = 0.05$.

3.2. Results of the simulation study

3.2.1. Type I error rate—Table 1 shows the type I error rates for the interaction tests and joint tests under the four different settings. For simplicity, we only show the aGE tests using Fisher's method to combine p-values. Both minP and Fisher's approach of combining p-values had correct type I error rates. The p-values of the two combination methods were close when no neutral RV was presented; therefore, our conclusion remained qualitatively the same using either method. All tests (aGE, rareGE, and iSKAT) maintained the type I error rates well when the model was correctly specified with a sample size of 7,000. When the environmental main effect was misspecified, all set-based tests with oversimplified modeling of the environmental variable suffered from inflated type I error rates. Furthermore, the gene-environment dependence considerably deteriorated such inflation. This inflation was mitigated after using the cubic splines to handle the nonlinear effect. When the sample size decreased to 1,000, the type I error rates for aGE and rareGE were still under control, but iSKAT had a slight inflation (empirical type I error rate was 0.061 with 10,000 replications). It also turned out that iSKAT was sensitive to the inclusion of neutral variants. With 200 neutral RVs included in the set and a sample size of 7,000, the empirical type I error rate was 0.083. Therefore, we did not include iSKAT for power comparison in the following sections. Last, the conclusion remained the same when the genetic main effects (β_1) followed a symmetric normal distribution (result not shown). In summary, the maintenance of the type I error rate depends on the sample size, number of RVs, and correct model specification. All the tests under consideration were hardly sensitive to the skewness of the distribution of the genetic main effects under our simulation settings.

3.2.2. Power

Interaction tests: Fig 1 shows that the aGE interaction test outperformed the other tests in the presence of neutral RVs: for example, with 200 neutral RVs and interaction effects having different directions in Fig 1B, the power of the aGE test was 0.326 under setting 1, which was higher than that of the rareGE (0.265). In general, the power of the interaction test decreased as the number of neutral RVs increased; however, the rate of power loss was different. The aGE test that combines a broad set of tests, including but beyond the GE(2) test, was more flexible and adaptive. When all the variants in the set were causal and the effects were in different directions, rareGE showed better performance than the aGE test because of the trade-off between model adaptivity and power. When all the variants were causal and the effects were in the same direction, the aGE test has slightly better performance than rareGE because GE(1) propelled the power.

With increasing numbers of neutral RVs, GE(γ) with higher power γ became more powerful and aGE always approached the most powerful GE(γ) test in the family. Therefore, the most powerful GE test could be different from GE(1) and GE(2), and aGE gained power from a higher-order GE(γ) test. As expected, odd γ performed better when the interaction effects (β_2) were in the same direction, while even γ performed better when the β_2 's had different

directions. However, we found that γ with the highest power depended more on the number of neutral variants than the direction of the effects of the causal RVs. Table 2 shows that GE(4) and GE(6) had the highest power when the number of neutral RVs was more than 50, regardless of the direction of the interaction effects. This indicates that the noise level played a bigger role in deciding the γ with the highest power and $\Gamma = \{1, 2, 3, 4, 5, 6\}$ suffices in a wide range of scenarios where the number of neutral variants could be as large as 400. We also observed that Fisher's method was slightly more powerful than the minP method when the number of neutral RVs was large, while the minP method was more powerful than Fisher's method when the number of neutral RVs was small. Furthermore, Table 2 shows that the p-values of GE(1) and GE(2) were close to those based on asymptotic distributions. Therefore, we are assured of the validity of using Monte Carlo simulation method to calculate the p-values. We also demonstrated the effectiveness of using cubic splines to saturate the environmental main effect model: the power of the saturated main effect model was almost as high as that obtained when using the correct function of the main effect model (Fig 1). For example, in setting 2 with no neutral variant, the saturated model had an empirical power of 0.362, whereas the empirical power for the correct model specification was 0.381. In Supplemental Materials Section A.2.2, we demonstrate numerically that the power loss when using splines was minor with more scenarios considered, including when G and E were dependent.

Joint tests: The findings of the joint tests were similar to those of the interaction tests. We observed that the rareGE joint test had inflated type I error rates when the model was not correctly specified (Table 1). However, the aGEjoint_{sp} test controlled the type I error rates and maintained high power, as shown in Fig 2. We further showed that the rareGE joint test was powerful at different relative scales of the ratios of the main-interaction effects. While our proposed aGE joint test was as powerful as the rareGE joint test when there was no neutral variant, it became more powerful than the latter test when the number of neutral RVs was large, as shown in Table 3. Compared to the rareGE and aGE joint tests, the multivariate score test lost a large amount of power as the number of neutral RVs increased. In addition, the power of the joint tests was as high as that of the interaction tests when the genetic main effects were absent, but it was much higher when the genetic main effects were moderate or strong.

4. Real Data Example: Pathway-by-smoking Interaction Analysis for Pancreatic Cancer

We further applied our proposed aGE tests, together with rareGE and iSKAT, to the data from a case-control study of pancreatic cancer based at MD Anderson Cancer Center as part of the Pancreatic Cancer Case-Control Consortium (PanC4).⁴⁰ Cases were defined as patients diagnosed with primary adenocarcinoma of the pancreas; controls were matched to cases according to birth year, sex and self-reported race, and were free of pancreatic cancer at the time of recruitment. Pancreatic cancer is the third leading cause of cancer-related death in the USA.⁴¹ Smoking is one of the most important modifiable risk factors for pancreatic cancer. Pathway analysis has shown promise in providing biological insights into disease mechanisms. Previous GxE analysis of GWAS data found that genes that interact

with smoking showed enrichment in the axonal guidance signaling pathway and the α -adrenergic signaling pathway.⁴² Along the same line, we investigated the biological pathway-by-smoking interaction for pancreatic cancer with a focus on RVs measured using the Illumina Exome Chip array.⁴³ Pack-years of cigarette smoking as our continuous environmental variable, defined as the number of packs of cigarettes smoked per day multiplied by the number of years the person has smoked, is more informative than binary variables, e.g., former/current smoker. As shown in Supplemental Materials Figure S4, pack-years of cigarette smoking had a nonlinear relationship with the logit of the probability of having pancreatic cancer. This nonlinearity would lead to model misspecification if not adequately taken into account, such as simply modeled as a linear function. As covariates, our GxE models also included age (continuous), sex, the top five principal components capturing the population substructure, and categorized body mass index. Genotype data were obtained as part of the PanC4 ExomeChip project.⁴³ The quality control steps included removing RVs with missing rate greater than 2%, or deviating from the Hardy-Weinberg equilibrium ($P < 10^{-6}$). Individuals with identity by stated coefficient larger than 0.25 or with missing RV rate greater than 2% were further removed, resulting in 613 cases and 498 controls. RVs (MAF less than 5%) with at least two minor alleles in the sample population were grouped into pathways according to the Kyoto Encyclopedia of Genes and Genomes (KEGG).⁴⁴ We identified 197 KEGG pathways with the number of RVs ranging from 3 to 1,105 (mean = 173, median=119, standard deviation = 173), and the cumulative minor allele count (cMAC) ranging from 57 to 17,173 (mean = 2,622, median=1,840, standard deviation = 2,631). Detailed definitions of the 197 KEGG pathways were previously provided.⁴⁵

The QQ plots of the p-values were severely inflated for tests that linearly modeled pack-years of cigarette smoking, including rareGE, iSKAT and rareGE joint tests (Fig (3)). Tests that used splines to model the environmental main effect had much better performance in controlling the inflation. Consistent with our simulation results, iSKAT had an inflated type I error rate even after correcting the model misspecification for the environmental variable, indicating that iSKAT requires a larger sample size for its asymptotic property to be held and is less robust to the presence of a large number of RVs. In this study, none of the pathways reached genome-wide significance based on the Bonferroni correction ($\alpha = 0.05/197 = 0.00025$) due to the moderate sample size and thus lack of power. However, we found one particular pathway (hsa03020) related to RNA polymerase that showed nominal significance with aGE_{sp} p-value = 0.0020, which was also previously reported by a CV-based pathway-by-smoking analysis of pancreatic cancer with a much larger sample size.⁴² This pathway consisted of 35 RVs with a cMAC of 360. Interestingly, $rareGE_{sp}$ and $rareGE_{joint,sp}$ tests had p-values larger than 0.05, and $GE_{sp}(1)$ was the main driver of the small p-value within the $GE(\gamma)$ family, thus shedding light on the underlying genetic architecture. Since this pathway is related to gene transcription regulation, further replication is warranted to confirm this observation with a larger sample size. We report 11 additional pathways that interacted with smoking at a nominal level (p-value ≤ 0.05) in Supplemental Materials Table S3. Nine of the top 12 pathways had smaller aGE_{sp} p-values than that of $aGE_{joint,sp}$, demonstrating the importance of interaction tests in finding novel pathways and understanding the biological mechanism of complex diseases.

5. Discussion

We have proposed a family of powerful and data-adaptive GxE tests for RV, called aGE, which maintains high power across varying GxE interaction patterns and in the presence of a large number of neutral RVs. Using simulations and a real data application, we have demonstrated that RV-based GxE tests can be prone to inflated type I error rates due to either poor estimation of the genetic main effects under the null hypothesis of no GxE effects or model misspecification for the environmental main effect. We tackled the former problem by using the GLMM to estimate the genetic main effects of RVs while addressing the latter challenge by saturating the model for the environmental main effect using splines. As a result, the proposed aGE test can maintain high power without inflating the type I error rate.

A noticeable feature of the aGE test is that it is more resilient to power loss due to an increasing numbers of neutral RVs in a test set, and thus decreasing signal-to-noise ratio, than existing RV-based GxE tests such as rareGE. This is an appealing property for NGS-based GxE analysis. In fact, while whole-exome/genome sequencing technology allows for genotyping genetic variants at the single-base resolution, it also generates more RVs as the sample size increases. For example, in the UK10K project, whole-genome sequencing of 3,781 individuals resulted in 42 million genetic variants, more than half of which were singletons or doubles, i.e., RVs that were only observed once or twice in the entire sample.⁴⁶ Therefore, the abundance of RVs has become the norm in large-scale NGS-based genetic epidemiology consortia, such as the National Heart, Lung, and Blood Institute Trans-Omics for Precision Medicine (TOPMed) program that has already generated whole-genome sequencing data in more than 100,000 individuals.⁴⁷ Our proposed aGE test is thus uniquely positioned and well suited for GxE analysis in these large research consortia.

Due to the low MAFs of multiple RVs in a set, a unique challenge with RV-set-based GxE tests is the need to estimate the genetic main effects under the null model of no GxE effects. In the proposed aGE test, we have employed the GLMM to estimate the null model by assuming the genetic main effects as random effects and found it works well in our extensive simulations and real data application. On the other hand, iSKAT uses generalized cross-validation-based ridge regression to penalize the genetic main effects in a similar manner as GLMM; however, we found it to be more prone to inflated type I error rates and non-convergence when the sample size was relatively small or a large number of neutral RVs were present. Therefore, to estimate the main effects for RVs, we recommend using mixed model methods, i.e., GLMM, for binary phenotypes and a linear mixed model for continuous outcomes. It is natural to extend the current model in the mixed model framework by including additional random components to account for the relatedness or population stratification among individuals in family-based studies, for example, the Framingham Heart Study in the TOPMed program; see Chen et al. for a review of related methods for the genetic main effect tests.⁴⁸ As for computational speed, aGE uses Monte Carlo simulation method to obtain the null distribution of the test statistics. It is much faster than the parametric bootstrap procedure because it involves fitting the null model only once. A drawback of Monte Carlo simulation method is that its accuracy relies on the asymptotic convergence rate. In our simulation studies and real data application, a total sample size of 1,000 appeared to suffice for aGE and rareGE.

We observed that joint tests were relatively more robust to misspecification of the main effect than interaction tests. The interaction tests suffer from potential bias in the estimation of both the genetic main effects and the environmental main effect, whereas the joint tests only suffer from bias in the latter estimation. In addition to robustness, the power of joint tests was comparable or higher than that of the interaction tests when the genetic main effects were merely moderate. Furthermore, joint tests are usually faster in estimation than interaction tests because the genetic main effects are absent under the null model of the former. Therefore, for the purpose of finding novel genes, joint tests are recommended due to their robustness, high power, and fast computational speed. The interaction tests can serve as a complementary analytic tool. If the study objective is to find genetic variants that modify the effect of an environmental exposure on disease risk and to understand the biological mechanism underlying a complex disease, then our proposed aGE interaction test is powerful and versatile for searching for such RVs. The aGE test was mainly developed for RVs; however, as a complementary analysis to individual SNP-based tests, it can also be used for CVs, for which the existence of high linkage disequilibrium may be problematic for the MLE. We provide simulation results for CVs in Supplemental Materials Section A.2.3.

With the completion of large-scale functional genomic projects, such as ENCODE,⁴⁹ Roadmap Epigenomic,⁵⁰ and GTEx projects,⁵¹ biological functional annotations of RVs have become available and have been incorporated as external weights in statistical tests for main effects as well as GxE interactions.^{52–54} Although we did not consider functional annotations in the proposed aGE test, in principle they can be incorporated as external weights in the score function, which could further improve the power and warrants future investigation. We have implemented our proposed tests in an R package, “aGE”, available at <https://github.com/ytzhong/projects/>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institutes of Health (NIH) grant R01CA169122; P.W. was supported by NIH grants R01HL116720 and R21HL126032; H.C. was supported by NIH grants R00HL130593 and U01HL120393. The authors are grateful to two anonymous reviewers for their many helpful and constructive comments that improved the presentation of the paper. The authors thank Ms. Lee Ann Chastain for editorial assistance. The authors acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing HPC resources that contributed to the research results reported within this paper. The authors declare no conflict of interest.

References

1. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci*. 2012;109(4): 1193–1198. [PubMed: 22223662]
2. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13(2):135. [PubMed: 22251874]
3. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747 [PubMed: 19812666]

4. Wu C, Kraft P, Zhai K, et al. Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene-environment interactions. *Nat Genet.* 2012;44(10):1090–7. [PubMed: 22960999]
5. Hancock D, Artigas MS, Gharib S, et al. Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. *PLoS Genet.* 2012;8(12):e1003098. [PubMed: 23284291]
6. Ritz BR, Chatterjee N, Garcia-Closas M, et al. Lessons learned from past gene-environment interaction successes. *Am J Epidemiol.* 2017;186(7):778–86. [PubMed: 28978190]
7. Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol.* 2013;37(7):643–57. [PubMed: 24123198]
8. Zhang Y, Ren J. ALDH2 in alcoholic heart diseases: molecular mechanism and clinical implications. *Pharmacol Ther.* 2011;132(1):86–95. [PubMed: 21664374]
9. Fleming SM. Mechanisms of Gene-Environment Interactions in Parkinson's Disease. *Curr Environ Health Rep.* 2017;4(2):192–9. [PubMed: 28417442]
10. Carreón T, Ruder AM, Schulte PA, et al. NAT2 slow acetylation and bladder cancer in workers exposed to benzidine. *Int J Cancer.* 2006;118(1):161–8. [PubMed: 16003747]
11. Wolock SL, Yates A, Petrill SA, et al. Gene $\times$ smoking interactions on human brain gene expression: finding common mechanisms in adolescents and adults *J Child Psychol Psychiatry.* 2013;54(10):1109–1119 [PubMed: 23909413]
12. Brooks PJ, Enoch MA, Goldman D, Li TK, Yokoyama A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. *PLoS Med.* 2009;6(3):e1000050.
13. Matsuo K, Hamajima N, Shinoda M, et al. Gene-environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. *Carcinogenesis.* 2001;22(6):913–6. [PubMed: 11375898]
14. Yano K, Yamamoto E, Aya K, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet.* 2016;48(8):927–934. [PubMed: 27322545]
15. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci.* 2006;103(6):1810–5. [PubMed: 16449388]
16. MacArthur DG, Manolio TA, Dimmock DP, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature.* 2014;508(7497):469 [PubMed: 24759409]
17. Auer PL, Teumer A, Schick U, et al. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nat Genet.* 2014;46(6):629–34. [PubMed: 24777453]
18. Asimit JL, Day-Williams AG, Morris AP, Zeggini E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum Hered.* 2012;73(2):84 [PubMed: 22441326]
19. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res.* 2007;615(1–2):28. [PubMed: 17101154]
20. 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;89(1):82–93. [PubMed: 21737059]
21. Pan W Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol.* 2009;33(6):497–507. [PubMed: 19170135]
22. Neale BM, Rivas MA, Voight BF, et al. Testing for an unusual distribution of rare variants. *PLoS Genet.* 2011;7(3):e1001322. [PubMed: 21408211]
23. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95(1):5–23. [PubMed: 24995866]
24. Pan W, Kim J, Zhang Y, Shen X & Wei P. A powerful and adaptive association test for rare variants. *Genetics.* 2014; 197: 1081–1095. [PubMed: 24831820]

25. Chen H, Meigs JB, Dupuis J. Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered.* 2014;78(2):81. [PubMed: 25060534]
26. Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics.* 2013;14(4):667–681.6 [PubMed: 23462021]
27. Lin X, Lee S, Wu MC, et al. Test for rare variants by environment interactions in sequencing association studies. *Biometrics.* 2016;72(1):156. [PubMed: 26229047]
28. Coombes B, and Basu S and McGue M. A combination test for detection of gene-environment interaction in cohort studies. *Genet Epidemiol.* 2017;41(5):396–412. [PubMed: 28370330]
29. Voorman A, Lumley T, McKnight B, Rice K. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PloS One.* 2011;6(5):e19416. [PubMed: 21589913]
30. Tchetgen Tchetgen EJ, Kraft P. On the Robustness of Tests of Genetic Associations Incorporating Gene-environment Interaction When the Environmental Exposure is Misspecified. *Epidemiology.* 2011;22(2):257–261. [PubMed: 21228699]
31. Cornelis MC, Tchetgen Tchetgen EJ, Liang L, et al. Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *Am J Epidemiol.* 2011;175(3):191–202. [PubMed: 22199026]
32. Kraft P, Yen Y, Stram DO, Morrison J, Gauderman WJ. Exploiting Gene-Environment Interaction to Detect Genetic Associations. *Hum Hered.* 2007;63(2):111–119. [PubMed: 17283440]
33. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet.* 2011;88(1):76–82. [PubMed: 21167468]
34. Wei P, Cao Y, Zhang Y, et al. On Robust Association Testing for Quantitative Traits and Rare Variants. *G3-Genes Genom Genet.* 2016;6(12):3941–50.
35. Davison AC, Hinkley DV. Bootstrap methods and their application. Cambridge university press; 1997.
36. Wahba G Spline models for observational data. Society for Industrial and Applied Mathematics; 1990.
37. Gu C Smoothing spline ANOVA models. Springer Science & Business Media; 2013.
38. He Z, Zhang M, Lee S, et al. Set-Based Tests for the Gene-Environment Interaction in Longitudinal Studies. *J Am Stat Assoc.* 2017;112(519):966–78. [PubMed: 29780190]
39. Wang T, Elston RC. Improved Power by Use of a Weighted Score Test for Linkage Disequilibrium Mapping. *Am J Hum Genet.* 2007;80(2):353–360. [PubMed: 17236140]
40. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat Genet.* 2009;41(9):986–90. [PubMed: 19648918]
41. Jemal A, Ward EM, Johnson CJ, et al. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Inst.* 2017;109(9).
42. Tang H, Wei P, Duell EJ, et al. Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene-and pathway-based interaction analysis of GWAS data. *Carcinogenesis.* 2014;35(5):1039–45. [PubMed: 24419231]
43. Childs EJ, Mocci E, Campa D, et al. Common variation at 2p13. 3, 3q29, 7p13 and 17q25. 1 associated with susceptibility to pancreatic cancer. *Nat Genet.* 2015;47(8):911–6. [PubMed: 26098869]
44. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 2010;38(suppl1):D355–D360. [PubMed: 19880382]
45. Wei P, Tang H, Li D. Insights into pancreatic cancer etiology from pathway analysis of genome-wide association study data. *PloS One.* 2012;7(10):e46887. [PubMed: 23056513]
46. UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature.* 2015;526(7571):82. [PubMed: 26367797]
47. Brody JA, Morrison AC, Bis JC, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet.* 2017;49(11):1560–1563. [PubMed: 29074945]

48. Chen H, Malzahn D, Balliu B, Li C, Bailey JN. Testing genetic association with rare and common variants in family data. *Genet Epidemiol.* 2014 9 1;38(S1).
49. ENCODE Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414), 57–74. [PubMed: 22955616]
50. Roadmap Epigenomics. Integrative analysis of 111 reference human epigenomes. *Nature.* 2015;518(7539), 317–330. [PubMed: 25693563]
51. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675), 204–213. [PubMed: 29022597]
52. Kim T, Wei P. Incorporating ENCODE information into association analysis of whole genome sequencing data. *BMC Proc.* 2016;10(Suppl 7):257–261. [PubMed: 27980646]
53. Morrison AC, Huang Z, Yu B, et al. Practical approaches for whole-genome sequence analysis of heart-and blood-related traits. *Am J Hum Genet.* 2017;100(2):205–15. [PubMed: 28089252]
54. Su YR, Di CZ, Hsu L. A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics.* 2017;18(1):119–31. [PubMed: 27474101]

Figure 1.

Empirical power curves for interaction tests with increasing numbers of neutral variants. (aGE_{sp} : aGE with cubic splines modeling environmental main effect; aGE_{true} : aGE testing linear GxE based on using the correct linear or quadratic function to fit the environmental main effect. A: $f(E) \propto E$ and β_2 's are in the same direction; B: $f(E) \propto E$ and β_2 's are in different directions; C: $f(E) \propto E^2$ and β_2 's are in the same direction; D: $f(E) \propto E^2$ and β_2 's are in different directions. G and E are independent in panels A to D.)

Figure 2.

Empirical power curves for joint tests with increasing numbers of neutral RVs. aGEjoint_{sp}:

aGE joint test that uses cubic splines to model the environmental main effect. A: no main effects: all genetic main effects (β_{1k} equal to 0); B: moderate main effects:

$\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/2, 2)$; C: strong main effects: $\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/3, 3)$.

Figure 3.

QQ plots for pathway-by-pack-years interaction for pancreatic cancer. Subscript $_{sp}$: splines were used to fit the environmental main effect; A. aGE_{sp}: genomic control $\gamma = 0.97$; B. aGEjoint_{sp}: $\lambda = 0.85$; C. rareGE: $\lambda = 1.59$; D. rareGE_{sp}: $\lambda = 0.99$; E. rareGEjoint: $\lambda = 0.66$; F. rareGEjoint_{sp}: $\lambda = 0.92$; G. iSKAT: $\lambda = 3.69$; H. iSKAT_{sp}: $\lambda = 2.92$.

Table 1.

Empirical type I error rates at the significance level 0.05 when no neutral RV was included. aGE_{sp} : aGE interaction test that uses cubic splines to model the environmental main effect; $aGE_{joint,sp}$: aGE joint test that uses cubic splines to model the environmental main effect; $G \perp E$: genetic risk factors and environmental variable are independent; $G \not\perp E$: genetic risk factors and environmental variable are not independent; $f(E)$: the true functional form of the environmental main effect. Inflated type I error rates are in boldface.

Setting	Interaction test				Joint test		
	aGE	aGE_{sp}	rareGE	iSKAT	aGEjoint	$aGE_{joint,sp}$	rareGEjoint
3,500 cases and 3,500 controls							
$G \perp E, f(E) \propto E$	0.048	0.050	0.054	0.054	0.046	0.048	0.047
$G \perp E, f(E) \propto E^2$	0.080	0.053	0.098	0.113	0.070	0.052	0.077
$G \not\perp E, f(E) \propto E$	0.054	0.044	0.056	0.056	0.047	0.049	0.047
$G \not\perp E, f(E) \propto E^2$	1	0.046	1	1	0.930	0.049	0.933
500 cases and 500 controls							
$G \perp E, f(E) \propto E$	0.048	0.050	0.051	0.061	0.047	0.050	0.047
$G \perp E, f(E) \propto E^2$	0.067	0.053	0.078	0.101	0.058	0.048	0.064
$G \not\perp E, f(E) \propto E$	0.044	0.048	0.047	0.059	0.051	0.052	0.046
$G \not\perp E, f(E) \propto E^2$	0.666	0.046	0.659	0.885	0.741	0.056	0.700

Table 2.

Empirical power for interaction tests under setting 1 with no neutral variant (3,500 cases and 3,500 controls). The most powerful GE test in the $GE(\gamma)$ family is in boldface. β_2 's are the coefficients of the interaction effects; $aymGE(1)$ is $GE(1)$ based on asymptotic null distribution; $aymGE(2)$ is $GE(2)$ based on asymptotic null distribution, i.e., mixture of chi-squared distributions.

Tests	β_2 in the same direction					β_2 in different directions				
	0	50	100	200	400	0	50	100	200	400
$GE(1)$	0.750	0.191	0.131	0.088	0.059	0.306	0.104	0.067	0.056	0.053
$GE(2)$	0.584	0.284	0.217	0.165	0.094	0.867	0.504	0.377	0.251	0.171
$GE(3)$	0.604	0.296	0.227	0.154	0.091	0.601	0.325	0.250	0.215	0.149
$GE(4)$	0.509	0.281	0.255	0.201	0.147	0.797	0.509	0.412	0.354	0.265
$GE(5)$	0.505	0.267	0.235	0.176	0.132	0.653	0.378	0.333	0.301	0.231
$GE(6)$	0.480	0.263	0.247	0.200	0.158	0.747	0.470	0.387	0.342	0.272
aGE	0.686	0.298	0.237	0.193	0.124	0.806	0.457	0.349	0.293	0.231
aGE_{Fisher}	0.610	0.304	0.257	0.208	0.163	0.743	0.462	0.385	0.327	0.257
$aymGE(1)$	0.752	0.196	0.136	0.086	0.059	0.308	0.104	0.069	0.060	0.055
$aymGE(2)$	0.593	0.290	0.222	0.171	0.096	0.865	0.509	0.379	0.251	0.181

Table 3.

Empirical power for joint tests with 200 neutral variants and $OR_{2k} = \exp(\beta_{2k}) \stackrel{i.i.d.}{\sim} Unif(1/3, 3)$ (3,500 cases and 3,500 controls). aGEjoint_{sp}: aGE joint test that uses cubic splines to model the environmental main effect; $OR_{1k} = \exp(\beta_{1k})$ is the odds ratio of the genetic main effect for the k^{th} RV. The most powerful test is in boldface.

Setting	rareGE	rareGEjoint	aGEjoint	aGEjoint _{sp}	score
$\beta_1 = 0$	0.119	0.102	0.111	0.109	0.076
$\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/2, 2)$	0.173	0.736	0.806	0.796	0.574
$\exp(\beta_{1k}) \stackrel{i.i.d.}{\sim} U(1/3, 3)$	0.198	0.973	0.983	0.983	0.945