European Journal of Human Genetics (2018) 26:1679–1686
https://doi.org/10.1038/s41431-018-0236-x

ARTICLE

Application of the parametric bootstrap for gene-set analysis of gene–environment interactions

Brandon J. Coombes[1] · Joanna M. Biernacka[1,2]

Received: 28 December 2017 / Revised: 2 July 2018 / Accepted: 19 July 2018 / Published online: 8 August 2018
© European Society of Human Genetics 2018

Abstract

Testing for gene–environment (GE) interactions in a gene-set defined by a biological pathway can help us understand the interplay between genes and environments and provide insight into disease etiology. A self-contained gene-set analysis can be performed by combining gene-level *p*-values using approaches such as the Gamma Method. In a gene-set analysis of genetic main effects, permutation approaches are commonly used to avoid inflated probability of a type 1 error caused by correlation of genes within the same pathway. However, when testing interaction effects, it is typically not possible to construct an exact permutation test. We therefore propose using a parametric bootstrap. For testing an interaction term, this approach requires fitting the null model, which only contains main effects; however, for a gene-set GE interaction model, the number of main effects can be large and therefore they may not be estimable. To estimate the main effects of SNPs in a gene-set, we propose modeling them as random effects. We then repetitively simulate null data from this model and analyze it to generate the null distribution of gene-set GE *p*-values, allowing for an empirical assessment of significance of the global GE effect in the gene-set of interest. Through simulation, we demonstrate that this approach maintains correct type I error, and is well powered to detect GE interactions. We apply our method to test whether the association of obesity with bipolar disorder (BD) is modified by genetic variation in the Wnt signaling pathway.

Introduction

Studying how environmental factors interact with single nucleotide polymorphisms (SNPs) to moderate or modify their effects on disease risk can advance our understanding of disease etiology. The genome–environment-wide interaction study (GEWIS) strategy has been used to agnostically search for SNP × environment interactions over the entire genome [1]. However, this strategy suffers from a large multiple testing burden. Additionally, to maintain comparable power, the detection of interactions generally requires sample sizes around four times greater than those required to test main effects [2]. Thus, most studies are underpowered to perform genome-wide scans of SNP × environment interactions.

Methods have been proposed to combine gene–environment (GE) interactions across a gene to test for interaction of an environment with a gene, rather than SNP, of interest [3–5]. This approach not only reduces multiple testing, but can also increases power by aggregating multiple small SNP-level GE effects. Gene-level interaction tests can be further combined into gene-set (or pathway) tests to evaluate whether genetic variation in a given gene-set/pathway interacts with the environment to influence the outcome.

Gene-set analyses have been used extensively to study the main effects of genes, with methods for testing two different types of hypotheses: competitive and self-contained [6]. In a competitive gene-set analysis (GSA), data across the entire genome are used to assess whether observed associations cluster in the gene-set of interest more than in the rest of the genome. In contrast, self-contained GSAs only consider data for a given pathway of interest to test whether genetic variants in the gene-set are associated with the outcome. In the context of self-contained analysis of genetic main effects, two-step

✉ Joanna M. Biernacka
Coombes.brandon@mayo.edu

[1] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

[2] Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA

approaches were shown to be more powerful than one-step approaches [6]. In a two-step approach, gene-level p-values are first computed, and then the gene-level p-values within a pathway are combined using strategies such as Fisher's Method or the Gamma Method [7]. A key assumption of these p-value combination methods is that the gene-level p-values are independent. However, because gene-level p-values may be correlated due to SNPs in genes within the same pathway being in linkage disequilibrium (LD), this assumption may be incorrect resulting in inflated probability of a type I error for the pathway-level test. To avoid this problem, permutation techniques are typically used to control type I error in pathway-level tests of genetic main effects

There has been a lot of progress in developing methods to study genetic main effects in a gene-set [6–8]. One popular pathway tool, MAGMA [9], can also jointly test genetic effects and GE interactions. However, no methods have been proposed for solely assessing GE interactions in a gene-set. In this paper, we address the challenges of assessing GE interactions in a two-step self-contained GSA (GE-GSA). We consider the analysis of pre-defined gene-sets, for example, a set of genes within a specific biological pathway defined in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/pathway.html) or the Gene Ontology (GO) Consortium database (http://www.geneontology.org). We apply a two-step approach in which gene-level p-values are combined using the Gamma Method [7]. While permutation approaches can be used to obtain a valid pathway-level p-value in a GSA of genetic main effects, in a GE interaction analysis, it is typically not possible to construct an exact permutation test. Bužková et al. [10] showed that a parametric bootstrap, originally proposed by Efron [11], can be used instead when testing GE interactions at the gene level. Here we extend this approach to develop a test for GE interactions for a gene-set, and perform simulations to evaluate the type I error and power of our method.

We demonstrate the proposed method by testing whether the risk of obesity associated with bipolar disorder (BD) is modified by genetic variation in the Wnt signaling pathway using data from the Genetic Association Information Network (GAIN) study of BD. A previous analysis of the GAIN data revealed a genome-wide significant association of BD with the interaction between body mass index (BMI) and an SNP in the gene *TCF7L2* [12]. *TCF7L2* codes for the transcription factor TCF/LF, which plays a role in the Wnt signaling pathway [13]. The Wnt signaling pathway has crucial implications in neurodevelopment, neurogenesis, and neuroplasticity, and is involved in mechanisms of action of medications used to treat BD [14, 15]. Therefore, here, we extend the prior findings of the association of BD with the *TCF7L2* SNP × BMI interaction, to investigate interaction effects with other genetic variants in the Wnt signaling pathway. We run the GE-GSA including and excluding *TCF7L2* from the gene-set to study interactions involving genetic variation in the pathway and to assess whether observed associations at the pathway level are due solely to the previously described *TCF7L2* interaction.

Methods

Assume n unrelated subjects are genotyped for a set of variants in a gene-set or pathway consisting of K genes with p_k SNPs in the Kth gene and $P = \sum_{k=1}^{K} p_k$total SNPs. Let the genotypes (coded, for example, as the number of copies of the minor allele) at the p_k SNPs for the kth gene for the ith subject be denoted as $\boldsymbol{G}_i^{(k)} = \left(G_{i1}^{(k)}, \ldots, G_{ip_k}^{(k)}\right)$. Denote the phenotype (e.g., disease or quantitative trait), covariates, and environment for the ith subject as Y_i, X, and E_i, respectively. We model the GE interactions for both continuous and binary outcomes in a generalized linear model (GLM) framework. In particular, we model continuous traits using linear regression and binary (disease) traits using logistic regression. Thus, we consider the model:

$$(\mu_i) = \alpha_0 + \boldsymbol{\alpha}_1 \boldsymbol{X}_i + \alpha_2 E_i + \sum_{k=1}^{K} \boldsymbol{\alpha}_3^{(k)} \boldsymbol{G}_i^{(k)} + \sum_{k=1}^{K} \boldsymbol{\beta}^{(k)} (\boldsymbol{G}_i^{(k)} \times E_i), \tag{1}$$

where $g(.)$ be the canonical link function for the mean, $\mu_i = E\left(Y_i \middle| \boldsymbol{X}_i, E_i, \boldsymbol{G}_i^{(1)}, \ldots, \boldsymbol{G}_i^{(K)}\right)$, of the phenotype, and $(\boldsymbol{G}_i^{(k)} \times E_i) = (G_{i1}^{(k)} E_i, \ldots, G_{ip_k}^{(k)} E_i)$ is the vector of p_k GE interactions. We are interested in testing the null hypothesis H_0:$\boldsymbol{\beta}^{(k)} = 0$ for all $k = 1,\ldots,K$.

A two-step GE-GSA begins by testing for GE interaction in each gene within the gene-set. There are many options for performing a gene-level GE interaction test [5]. Here we use principal components (PCs) to reduce the dimensionality of SNP data within a gene: for the Kth gene, we calculate PCs using all p_k SNPs within the gene and retain the first q_k components ($q_k < p_k$) that explain 80% of the variance within the gene. We then test the GE interactions for the Kth gene by performing a score test of the $\boldsymbol{\beta}^{(k)}$ term in the following model:

$$g(\mu_i) = \alpha_0^{(k)} + \boldsymbol{\alpha}_1^{(k)} \boldsymbol{X}_i + \alpha_2^{(k)} E_i + \boldsymbol{\alpha}_3^{(k)} \boldsymbol{PC}_i^{(k)} + \boldsymbol{\beta}^{(k)} \left(\boldsymbol{PC}_i^{(k)} \times E_i\right), \tag{2}$$

where $\boldsymbol{PC}_i^{(k)}$ is the set of p_k PCs in the Kth gene for the ith person and $(\boldsymbol{PC}_i^{(k)} \times E_i) = (PC_{i1}^{(k)} E_i, \ldots, PC_{iq_k}^{(k)} E_i)$ is the vector of $q^k PC \times E$ interactions. By using PCs to model the effects of SNPs in a gene, the collinearity of correlated

SNPs within a gene is removed, and the number of parameters in the model is reduced by $2(p_k–q_k)$.

Gamma Method

Once the K gene-level GE interaction p-values ($pval_1$, … $pval_K$) are computed, we propose aggregating them to the gene-set using the Gamma Method. The test statistic for the Gamma Method can be written as

$$T = \sum_{k=1}^{K} G_{\omega,1}^{-1}(1 - pval_k), \tag{3}$$

where $G_{\omega,1}^{-1}(\cdot)$ is the inverse gamma distribution with shape parameter ω and scale parameter 1. More emphasis can be given to p-values below a particular threshold, referred to as a soft truncation threshold (STT), by varying $\omega = \mathrm{G}_{\omega,1}^{-1}(1 - STT)$ [6]. The uncorrected p-value $Pval$ for a pathway can be calculated for the test statistic T using its null distribution; when the p-values are independent and have a standard uniform distribution, the null distribution of T is a Gamma distribution with shape parameter ω and scale parameter 1. When $\omega = 1$, the null distribution of T is a χ^2 distribution with degrees-of-freedom (df) equal to $2K$. This specific case of the Gamma Method, which corresponds to an STT of $1/e$, was first described by Fisher [16] and is known as Fisher's method. Because the optimal STT is unknown for a given dataset, it may be beneficial to search over multiple STTs and use the STT that leads to the smallest p-value. However, the minimum p-value resulting from this search (minGamma) will have inflated probability of a type I error unless the multiple testing is correctly taken into account. Additionally, the Gamma Method assumes that gene-level GE interaction p-values are independent. Violation of this assumption can lead to inflated probability of a type I error. Typically, permutations are used to control type I error when evaluating genetic main effects at a pathway level, but this approach is not applicable when assessing interactions [10]. Thus, we propose using a parametric bootstrap approach to control the type I error.

Parametric bootstrap

Bužková et al. [10] proposed a parametric bootstrap to test for GE interaction at the gene-level. For a GE-GSA, the parametric bootstrap can be applied to estimate p-values using the following steps:

1. Compute the uncorrected gene-set p-value ($Pval$) for the original data as described above in the section on the Gamma Method.
2. Obtain parameter estimates of SNP and environment effects from the original data by fitting the gene-set model under the null hypothesis H_0:$\boldsymbol{\beta}^{(K)} = 0$ for all k.
3. For all subjects, simulate responses $Y^{(b)}$ from the model obtained in Step 2.
4. Using the data simulated in Step 3, perform a two-step GE-GSA using the same PC-Gamma gene-set level GE test that was used in the original data, i.e., compute gene-level GE interaction p-values using the simulated response, and then combine the gene-level p-values in the gene-set using the Gamma Method with a chosen STT (or the minGamma method) to obtain the simulated gene-set p-value $Pval^{(b)}$.
5. Repeat Steps 3–4 B times to approximate the distribution of the p-values under the null hypothesis.
6. Compute the corrected (parametric bootstrap) p-value by comparing the uncorrected p-value from Step 1 with the simulated null p-values: $Pval_{PB} = \frac{1}{B}\sum_{b=1}^{B} I(Pval < Pval^{(b)})$ where $I(\cdot)$ is the indicator function. Standard error for $Pval_{PB}$ can be calculated using the binomial distribution: $\sqrt{Pval_{PB}(1-Pval_{PB})/B}$.

Fitting the gene-set model under the null hypothesis (i.e., Step 2 above) requires estimating effects for the environment, covariates, and all SNPs in the gene-set. Since the number of SNPs in a gene-set is typically very large ($P > n$), this model usually cannot be fit using a standard GLM. Even if the parametric bootstrap was based on modeling the effects of PCs (rather than SNPs) under the null hypothesis of no GE effects, problems with estimation would still likely to be encountered unless very few PCs were used, which may not appropriately model the main effect. Instead, we propose fitting the null model by treating the genetic main effects as random effects in the following model:

$$g(\mu_i) = \alpha_0 + \boldsymbol{\alpha}_1 \boldsymbol{X}_i + \alpha_2 E_i + \gamma, \tag{4}$$

where $\gamma \sim \mathrm{MVN}(0, \sigma_G^2 \mathbf{G}\mathbf{G}^T)$ and

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1^{(1)} & \mathbf{G}_1^{(2)} & \cdots & \mathbf{G}_1^{(K)} \\ \mathbf{G}_2^{(1)} & \mathbf{G}_2^{(2)} & \cdots & \mathbf{G}_2^{(K)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{G}_n^{(1)} & \mathbf{G}_n^{(2)} & \cdots & \mathbf{G}_n^{(K)} \end{bmatrix},$$

which is an $n \times P$ matrix. This model is similar to that used in Genome-wide Complex Trait Analysis (GCTA) but is confined to only the SNPs in the gene-set of interest [17]. This formulation of our model avoids estimating the proportion of variance explained by each SNP and instead estimates the total proportion of variation explained by all of the SNPs in the gene-set. It can be shown that the above model is equivalent to using a ridge penalty to penalize the genetic main effects [18, 19]. With this parameterization,

the approximate null gene-set model requires fitting only $3 + L$ parameters, where L is the number of covariates. This model can be fit using the R package GMMAT [20]. Responses can then be simulated from this null model (Step 3 above) using the best linear unbiased prediction. Code for performing the described parametric bootstrap approach can be found at https://github.com/bcoombes/Parametric_Bootstrap.

Simulations

To illustrate the proposed approach, we analyzed Wnt signaling pathway data in a subset of 388 BD cases and 1020 controls from the GAIN study. We considered models with BD as the outcome predicted by gene–BMI interactions, and with BMI as the outcome predicted by gene–BD interactions. Before analyzing the BD data, we performed simulations to study type I error and power of the proposed approach. For the simulations, we generated data with properties similar to the GAIN-BD Wnt signaling pathway data.

To simulate each replicate dataset, we first sampled genotypes in the Wnt signaling pathway from the GAIN-BD data without replacement $N = 1000$ times. Sex was independently sampled from a Bernoulli distribution with $Pr(Male) = 0.5$. We used the GAIN Wnt signaling pathway data to estimate each SNP effect and SNP × BD interaction effect on BMI, using regression models with scaled BMI as the outcome and with sex as a covariate. The directionality of the main and interaction effects for the top SNP in each gene was recorded. We randomly selected J different "causal" SNPs ($J = 2$, 20, or 100) from this list to have a main effect and an interaction in the simulations. We then simulated a quantitative trait conditional on sex, genotypes from the gene set, and BD status sampled from the real data using a linear model with a standard normal error term. The variation in the quantitative trait explained by the chosen causal SNPs and their interactions was held fixed to explain 30% of the variation throughout the simulations so that under models with more causal SNPs, the per-SNP effects would be smaller. We also analyzed a binary trait derived by dichotomizing the quantitative outcome at the median, to study the methods' properties when the outcome is binary.

We calculated gene-level p-values using the PC method described above using the simulated quantitative or dichotomized quantitative variable as the outcome variable in the model and testing for the presence of G × BD interactions. Sex was included as a covariate in both analyses. We combined the gene-level results using the Gamma Method with STT equal to 0.01, 0.05, 0.10, 0.15, 0.2, or $1/e$ (Fisher's method) as well as the minGamma approach, and computed uncorrected as well as parametric bootstrap gene-set level p-values (i.e., $Pval$ and $Pval_{PB}$, respectively) for each simulated dataset, in order to evaluate the performance of the proposed method in the absence or presence of GE interactions.

For type I error estimation, we simulated data where the top J SNPs had a main effect and no GE interactions. To estimate power, for the same J SNPs with main effects, we included interaction terms with BD. Type 1 error was estimated at the $\alpha = 0.05$ and 0.01 levels based on 10,000 null replicate datasets, while power was estimated at the $\alpha = 0.05$ level based on 1000 replicate datasets under each scenario.

GE-GSA of the GAIN-BD dataset

To illustrate the proposed GE-GSA approach, we used genetic data from the Wnt signaling pathway in a subset of subjects from the GAIN-BD study with available BMI data. The dataset [21] was obtained from dbGaP (phs000017.v3.p1), and includes the same set of subjects as the paper that originally reported the genome-wide significant association between BD and the TCF7L2-BMI interaction [12]. The analyzed gene-set was the Wnt canonical pathway as defined by KEGG. Gene regions were defined as 20 kb up/downstream of the RefSeq transcription start/end sites according to the Human Genome Browser build hg18 [37]. The analysis included 143 genes in the gene-set that had more than one SNP mapping to the gene; 3767 SNPs in these genes were included in our analysis. For each gene, we computed PCs accounting for 80% of the variation, which resulted in a total of 580 PCs in the pathway. We used either BD or log BMI as the outcome testing for either G × BMI or G × BD interactions in the Wnt pathway. BMI was log-transformed to create an approximately normal distribution. Sex was included as a covariate in the model.

Results

Type I error

We first assessed the empirical type I error of the proposed GE-GSA strategies for $\alpha = 0.05$ and 0.01 (Table 1). We present results for a scenario with $J = 2$ SNPs that have a main effect but no GE interaction. Other choices of J yielded similar results (not shown). Results in Table 1 demonstrate that the uncorrected (asymptotic) Gamma Method, which assumes the gene-level tests are independent, has inflated probability of a type I error. As expected, the type I error for the minGamma method is even more inflated due to the search over multiple STTs. Using our parametric bootstrap strategy, the type I error for the Gamma and minGamma methods was controlled at the correct level, for either choice of α.

Table 1 Empirical type I error for each choice of STT for the Gamma Method

		α=0.05		α=0.01	
	Outcome	BD	BMI	BD	BMI
	STT=				
Uncorrected/ Asymptotic	0.01	0.049	**0.057**	0.011	**0.013**
	0.05	**0.059**	**0.066**	**0.017**	**0.018**
	0.1	**0.072**	**0.075**	**0.020**	**0.023**
	0.15	**0.084**	**0.084**	**0.024**	**0.027**
	0.2	**0.094**	**0.090**	**0.028**	**0.029**
	1/*e* (Fisher's)	**0.119**	**0.096**	**0.041**	**0.032**
	minGamma	**0.167**	**0.156**	**0.056**	**0.053**
	STT=				
Parametric bootstrap	0.01	0.051	0.053	0.009	0.010
	0.05	0.051	0.049	0.011	0.010
	0.1	0.052	0.047	0.009	0.011
	0.15	0.050	0.048	0.010	0.011
	0.2	0.049	0.049	0.010	0.010
	1/*e* (Fisher's)	0.050	0.044	0.010	0.009
	minGamma	0.048	0.046	0.009	0.010

Inflated type I error probabilities are denoted in bold

Empirical power

We next compared the power of the GE-GSA strategies which maintained type I error. To evaluate empirical power at $\alpha = 0.05$, we varied the number of SNPs with main effect and interaction with BD status ($J = 2$, 20, or 100). Figure 1 shows the power for the Gamma Method using either the largest or smallest choice of STT searched as well as the minGamma method across different choices of J. For $J = 2$, the smallest values of STT were most powerful and as J increased, larger values of STT performed better. The minGamma method was robust to choices of J and was always powerful. Fisher's method (STT = 1/*e*) was most powerful than other options of STT when J was very large. As expected, using a dichotomized outcome resulted in a loss of power.

GSA of the Wnt signaling pathway in GAIN-BD data

To test for GE interaction in the Wnt signaling pathway in the GAIN-BD data, we first computed gene-level p-values using the PC method. The gene-level p-values from the G × BMI or G × BD models were highly correlated ($r^2 = 0.95$). The top genes with interaction are listed in Table 2. The majority of gene-level results were more significant when testing for G × BD association with BMI rather than G × BMI association with BD; *TCF7L2* was the second most significant gene in both analyses.

We next combined the gene-level results for each model using the strategies described in the Methods section. For the parametric bootstrap, 10,000 p-values were generated from the null model to generate an empirical p-value for the observed data. We report the gene-set GE interaction p-values for all choices of STT in Table 3. The asymptotic Gamma Method p-values are provided to show how the parametric bootstrap adjusts the p-value to control the type I error. When testing for association of BD with G × BMI interactions, significant evidence of interaction in the Wnt pathway was only obtained with Fisher's method when *TCF7L2* was included in the gene-set. When we tested for association of BMI with G × BD interaction, the Gamma with STT >0.1, and minGamma methods, provided significant evidence of interaction in the pathway regardless of whether *TCF7L2* was included or excluded. Fisher's method produced the smallest p-value throughout our analyses, which indicates that there are likely many genes with small GE interaction effects that contributed to the significant pathway-level interaction result. It is not surprising that the pathway-level results were more significant for models with BMI as the outcome, because the gene-level p-values for this model were typically smaller than the p-values for the model with BD as the outcome.

Discussion

In this paper, we proposed a variation of the parametric bootstrap to test for GE interaction in a gene-set. In a GE-GSA, the null model requires estimating a large number of genetic main effects. Instead of fitting all of the genetic main effects, we treated the effects as a random effect, which allowed us to substantially reduce the number of parameters in the model. Treating the genetic main effects as a random effect can be shown to be equivalent to penalizing the effects with a ridge penalty. The proposed parametric bootstrap technique corrected the inflated type-I error probability associated with using an asymptotic Gamma Method in a GE interaction GSA. The parametric bootstrap also maintained correct type I error when searching over multiple values of STT, which provides a GE-GSA test that is powerful across a range of different underlying situations.

In our analysis of the GAIN-BD data using the parametric bootstrap, we found statistically significant evidence that the association of BD with BMI is modified by genetic variation in the Wnt signaling pathway. For this analysis, Fisher's method produced the smallest p-value which indicates that this pathway may contain many genes with G × BD interaction effects on BMI. Our results also suggest that the previously reported relationship between BD, BMI,

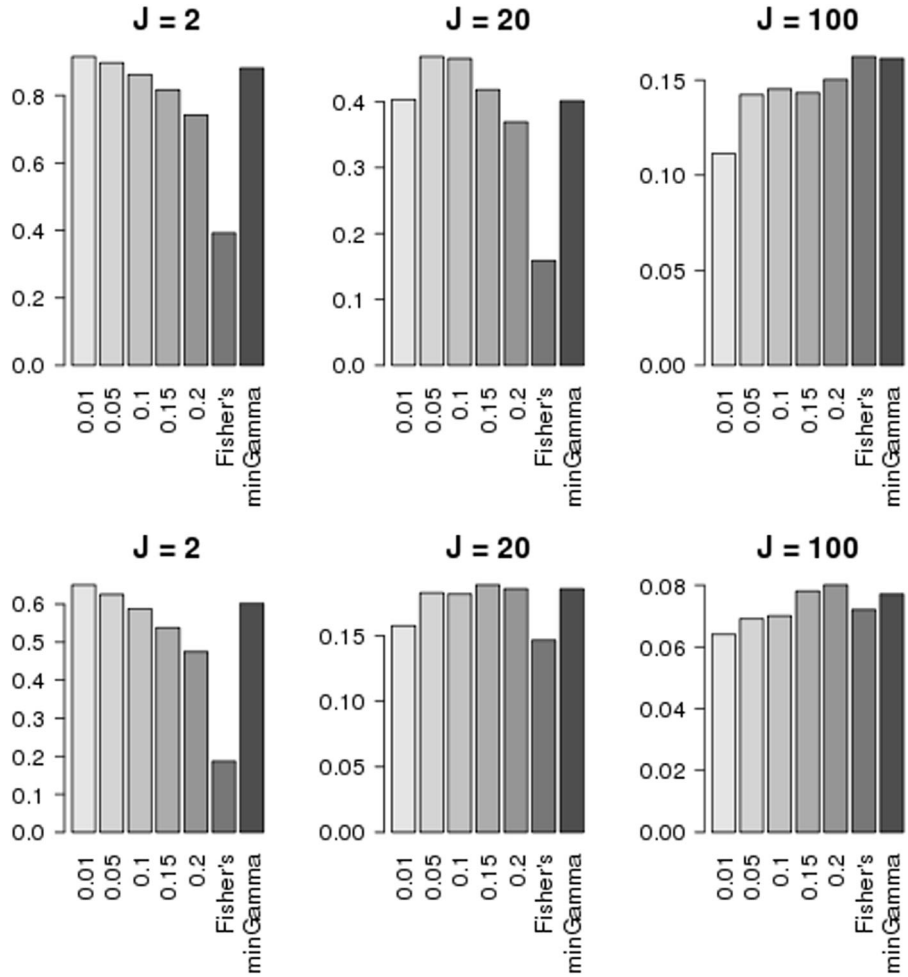

Fig. 1 Empirical power at $\alpha = 0.05$ level for different choices of STT and varying number of causal genes. Pathway analysis was performed using a quantitative trait (top row) or dichotomous trait (bottom row) as the outcome. The *y*-axis shows the empirical power

Table 2 Top ten gene-level interaction *p*-values for the GAIN dataset using either BD status (G × BMI *p*-value) or log BMI (G × BD *p*-value) as the outcome

Gene name	# SNPs	# PCs	G × BMI *p*-value	G × BD *p*-value
VANGL1	39	6	0.0090	0.0020
TCF7L2	52	11	0.0073	0.0028
NFATC3	6	2	0.0067	0.0051
CHP2	14	2	0.0217	0.0089
SMAD3	73	9	0.0244	0.0097
SFRP1	23	4	0.0178	0.0143
ROCK2	24	2	0.0313	0.0179
CTBP2	87	12	0.0507	0.0226
SMAD4	5	1	0.0326	0.0254
DAAM1	72	5	0.0297	0.0296

Rows are sorted by the far-right column. Gene-level *p*-values are not corrected for multiple testing of genes across the Wnt pathway

and TCF7L2 genetic variation [12] may be reversed where obesity risk is affected by interactions between BD status and the SNPs in the Wnt pathway rather than BD risk being affected by G × BMI interactions. This is in agreement with our simulation results where we generated BMI as dependent on sex, genetics, BD, and G × BD interactions and found the model that treated BMI as the outcome performed better than the model with BD as the outcome. However, directionality of the BD-BMI association cannot be conclusively determined from these analyses, and other approaches such as Mendelian Randomization would be required to assess causality.

One limitation of the parametric bootstrap is that like all sampling-based techniques, it can be computationally intensive. Our analysis of one candidate pathway required us to analyze 10,000 samples generated from the null model. If we were interested in testing GE interactions for many gene-sets in a hypothesis-generating context, the parametric bootstrap would become computationally infeasible. In this situation, we recommend using the much faster uncorrected (asymptotic) Gamma or minGamma method, followed by analysis of the top resulting gene-sets using the parametric bootstrap to ensure proper control of type I error. Another limitation to the parametric bootstrap is that it assumes the null model is correct, and thus a severe

Table 3 GE gene-set (Wnt signaling pathway) analysis of the GAIN dataset

Including *TCF7L2*	STT	G × BMI *p*-value		G × BD *p*-value	
		Uncorrected/ Asymptotic	Parametric Bootstrap	Uncorrected/ Asymptotic	Parametric Bootstrap
Gamma Method	0.01	0.57	0.56	0.19	0.20
	0.05	0.33	0.33	0.044	0.056
	0.1	0.16	0.17	0.011	0.020
	0.15	0.081	0.10	0.0035	0.0099
	0.2	0.052	0.077	0.0015	0.0057
	Fisher's (1/*e*)	**0.015**	**0.040**	**0.00017**	**0.0022**
	minGamma	**0.015**	0.061	**0.00017**	0.0032
Excluding *TCF7L2*					
Gamma Method	0.01	0.61	0.60	0.25	0.25
	0.05	0.47	0.46	0.10	0.12
	0.1	0.26	0.27	0.032	0.047
	0.15	0.15	0.17	0.012	0.026
	0.2	0.10	0.13	0.0056	0.016
	Fisher's (1/*e*)	**0.030**	0.064	**0.00056**	**0.0035**
	minGamma	**0.030**	0.097	**0.00056**	0.0056

The *p*-values for the asymptotic Gamma Method are expected to have inflated probability of a type I error. Because a significant GE interaction was identified for *TCF7L2* in previous analyses of the GAIN data, the gene-set analysis was conducted either including (top) or excluding (bottom) *TCF7L2*. The smallest *p*-values are bolded. Standard errors for the bootstrapped *p*-values range from 0.0001 to 0.005

misspecification of the null model could result in inflated probability of a type I error. However, the parametric bootstrap is currently one of the only approaches that can be used to simulate a null distribution for GE interaction tests. Further investigation of the impact of null model misspecification and ongoing method development is therefore needed to address this issue. The current simulations also only considered the case where G and E are independent. More study is necessary to investigate GE dependence. Finally, while our simulations only used one type of gene-level GE interaction test, there are other gene-level GE interaction tests that could be implemented in this framework. In our future work, we plan to incorporate the proposed parametric bootstrap to explore which of these gene-level GE interaction tests are most powerful in a two-step pathway analysis.

In this paper, we used a parametric bootstrap approach to derive a valid GE-GSA test with correct type 1 error. It was previously shown that the parametric bootstrap can be a useful tool when the asymptotic distribution of a test statistic is difficult or impossible to derive. However, the parametric bootstrap requires fitting the null model, which sometimes poses a challenge. The parametric bootstrap approach that we proposed for GE-GSA overcame this challenge, and may prove useful in other "big data" applications, such as tests of interactions among SNPs in a gene-set, and gene-level tests of interactions when the number of SNPs in a gene is very large.

Acknowledgements This work was supported by a generous gift from the Marriot Family and the Mayo Clinic Center for Individualized Medicine.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Fan Q, Verhoeven VJM, Wojciechowski R, et al. Meta-analysis of gene-environment-wide association scans accounting for education level identifies additional loci for refractive error. Nat Commun. 2016; 29(7):11008.
2. Thomas D. Gene-environment-wide association studies: emerging approaches. Nat Rev Genet. 2010;11:259–72.
3. Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. Biostatistics. 2013;14:667–81.
4. Lin X, Lee S, Wu MC, et al. Test for rare variants by environment interactions in sequencing association studies. Biometrics. 2016; 72(1):156–164.
5. Coombes B, Basu S, Mcgue M. A combination test fors detection of gene-environment interaction in cohort studies. Genet Epidemiol. 2017. https://doi.org/10.1002/gepi.22043.
6. Fridley BL, Biernacka JM. Gene set analysis of SNP data: benefits, challenges, and future directions. Eur J Hum Genet. 2011;19:837–43.
7. Biernacka JM, Jenkins GD, Wang L, Moyer AM, Fridley BL. Use of the gamma method for self-contained gene-set analysis of SNP data. Eur J Hum Genet. 2012. https://doi.org/10.1038/ejhg.2011.236.

8. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. Nat Rev Genet. 2016;17:353–64.
9. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:1–19.
10. Bužková P, Lumley T, Rice K. Permutation and parametric bootstrap tests for gene-gene and gene-environment interactions. Ann Hum Genet. 2011;75:36–45.
11. Efron B. Bootstrap methods: another look at the jackknife. In: Breakthroughs in statistics. Annals Stat., 7 (1979), pp. 1-26
12. Winham SJ, Cuellar-Barboza AB, Oliveros A, et al. Genome-wide association study of bipolar disorder accounting for effect of body mass index identifies a new risk allele in TCF7L2. Mol Psychiatry. 2014;19:1010–6.
13. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial--mesenchymal transition: new insights in signaling, development, and disease. J Cell Biol. 2006;172:973–81.
14. Abu-Baker A, Laganiere J, Gaudet R, et al. Lithium chloride attenuates cell death in oculopharyngeal muscular dystrophy by perturbing Wnt/$β$-catenin pathway. Cell Death Dis. 2013;4: e821.
15. Valvezan AJ, Klein PS. GSK-3 and Wnt signaling in neurogenesis and bipolar disorder. Front Mol Neurosci. 2012;5. p. 1
16. Fisher R. Statistical methods for research workers. 1925. https://doi.org/10.1056/NEJMc061160.
17. Yang J, Lee SH, Goddard ME, Visscher PM. {GCTA}: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.
18. Hodges JS. Richly parameterized linear models: additive, time series, and spatial models using random effects. CRC Press; 2013. Boca Raton, FL.
19. Mazo Lopera MA, Coombes BJ, de Andrade M. An efficient test for gene-environment interaction in generalized linear mixed models with family data. Int J Environ Res Public Health. 2017;14:1134.
20. Chen H, Wang C, Conomos M, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet. 2016;98:653–66.
21. Smith EN, Bloss CS, Badner JA, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. Mol Psychiatry. 2009;14:755–63.