

RESEARCH ARTICLE

Quick assessment for systematic test statistic inflation/deflation due to null model misspecifications in genome-wide environment interaction studies

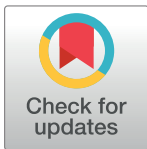
Masao Ueki^{1,3*}, Masahiro Fujii^{2‡}, Gen Tamiya^{1,3}, for Alzheimer's Disease Neuroimaging Initiative and the Alzheimer's Disease Metabolomics Consortium[¶]

1 Statistical Genetics Team, RIKEN Center for Advanced Intelligence Project, Chuo-Ku, Tokyo, Japan, **2** Graduate School of Medicine, Kurume University, Kurume, Fukuoka, Japan, **3** Tohoku Medical Megabank Organization, Tohoku University, Aoba-Ku, Sendai, Japan

✉ Current address: Alfresa Pharma Corporation, Chuo-Ku, Osaka, Japan

¶ Membership of the Alzheimer's Disease Neuroimaging Initiative and Alzheimer's Disease Metabolomics Consortium is provided in the Acknowledgments

* uekimrsd@nifty.com



OPEN ACCESS

Citation: Ueki M, Fujii M, Tamiya G, for Alzheimer's Disease Neuroimaging Initiative and the Alzheimer's Disease Metabolomics Consortium (2019) Quick assessment for systematic test statistic inflation/deflation due to null model misspecifications in genome-wide environment interaction studies. PLoS ONE 14(7): e0219825. <https://doi.org/10.1371/journal.pone.0219825>

Editor: Andre Scherag, University Hospital Jena, GERMANY

Received: December 2, 2018

Accepted: July 2, 2019

Published: July 18, 2019

Copyright: © 2019 Ueki et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The authors do not own data used in the manuscript. Data obtained were collected and owned by the Alzheimer's Disease Neuroimaging Initiative (ADNI). Researchers may request and access the data through the ADNI website (<http://adni.loni.usc.edu/>). The authors had no special access privileges to this data.

Abstract

Gene-environment (GxE) interaction is one potential explanation for the missing heritability problem. A popular approach to genome-wide environment interaction studies (GWEIS) is based on regression models involving interactions between genetic variants and environment variables. Unfortunately, GWEIS encounters systematically inflated (or deflated) test statistics more frequently than a marginal association study. The problematic behavior may occur due to poor specification of the null model (i.e. the model without genetic effect) in GWEIS. Improved null model specification may resolve the problem, but the investigation requires many time-consuming analyses of genome-wide scans, e.g. by trying out several transformations of the phenotype. It is therefore helpful if we can predict such problematic behavior beforehand. We present a simple closed-form formula to assess problematic behavior of GWEIS under the null hypothesis of no genetic effects. It requires only phenotype, environment variables, and covariates, enabling quick identification of systematic test statistic inflation or deflation. Applied to real data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), our formula identified problematic studies from among hundreds of GWEIS considering each metabolite as the environment variable in GxE interaction. Our formula is useful to quickly identify problematic GWEIS without requiring a genome-wide scan.

Introduction

Gene-environment (GxE) interaction is one potential way to unravel the missing heritability problem [1–3]. Genome-wide GxE interaction studies are becoming popular, as evidenced

Funding: This work was supported by Japan Society for the promotion of science (<http://www.jsps.go.jp/english/>), grant numbers JP16K00064 (received author is M.U.), JP16K08638 (received authors are M.U. and G.T.). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: We have the following interests: Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company;

by recent reviews [4–9]. In large-scale genomic analyses, such as genome-wide association studies (GWAS) and genome-wide environment interaction studies (GWEIS), hypothesis tests based on regression models are widely used to discover genetic susceptibility variants. Each genetic variant is marginally examined with a univariate regression model in a GWAS, and the analysis may be adjusted for covariates—such as sex and age. GWEIS is similarly conducted in a regression model that involves GxE interaction. Due to low power for testing interactions between genetic variants and environment variables (statistical interaction), however, a joint test for the presence of genetic effects allowing GxE interaction [10] is frequently used [11–14]. Its simplicity is an advantage over other existing methods, and we focus on the joint test throughout this paper. With GWAS data, up to a million genetic variants can be tested, and the multiplicity of hypotheses can be accounted for. For valid discovery of genetic susceptibility variants, it is necessary that the type I error is properly controlled. Systematic inflation or deflation in test statistics over all loci means that the study is problematic; population stratification and cryptic relatedness are typical reasons. A poor specification of null model (i.e. the model which includes no genetic effect but may include covariate/environment variables) may also cause problematic behavior of the test statistics unexpected under the null hypothesis, because then the null hypothesis is false for all loci regardless of the presence of genetic effects. Problematic behavior is observed in GWEIS more frequently than in GWAS [15, 16].

The severity of problematic behavior can be quantified by the median or mean of the chi-squared statistics for association from a genome-wide scan. For example, the genomic inflation factor constant λ is estimated by the median [17] or mean [18, 19] of genome-wide chi-squared statistics. Currently, numerous feature variables tend to be collected from a large number of participants in cohort or biobank studies [20]. Researchers often have many candidate variables that can be used as covariates and environment variables in GxE interaction analyses. Severe systematic inflation (or deflation) in genome-wide test statistics indicates that the study is problematic. However, the computational load is high when one needs numerous genome-wide scans with large samples, as in recent biobank studies, in which a large number of candidate environment variables needs to be evaluated. Moreover, with genome-wide imputed data [21, 22] or whole-genome sequencing data, a larger number of variants is tested for association than with SNPs (single nucleotide polymorphisms) in GWAS. It would therefore be helpful if problematic behavior could be found before conducting a computationally expensive genome-wide scan.

By assuming no genetic effect, we present a novel closed-form approximation to the mean of the genome-wide joint GxE interaction test statistics, which can be used to assess problematic behavior due to null model misspecification. The formula disregards characteristics of genetic loci, thereby enabling preliminary use before computing chi-squared statistics by genome-wide scan. We show that the approximation agrees well with the mean of the chi-squared statistics from various joint GxE interaction tests for GWAS data from the ADNI. Our approximation is also valid for marginal association tests. The approximation formula reveals that the joint GxE interaction test is sensitive to null model misspecification, whereas the marginal association test is not.

The remainder of the paper is organized as follows. In the Materials and Methods section, we provide the proposed approximation formula, describe the real data application using SNP-GWAS data from ADNI for joint GxE interaction and marginal association analyses, and provide the setup of simulation studies to evaluate the performance of the approximation. In the Results section, we give the results from the real data application and simulation studies. In the Discussion section, we summarize the results and discuss about the proposed approach both theoretically and empirically.

CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann–La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). Data collection and sharing for this project was also funded by the Alzheimer’s Disease Metabolomics Consortium (National Institute on Aging R01AG046171, R01AG051550 and 3U01AG024904-09S4). There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.

Materials and methods

The approximation formula

Suppose that n samples are observed with phenotypic value (binary, numeric value, or a factor) denoted by y_1, \dots, y_n , and L genetic variants, $\mathbf{g}_l = (g_{l,1}, \dots, g_{l,n})^T$ for $l = 1, \dots, L$, are to be tested for association with the phenotype. We introduce p variables, $\mathbf{w}_{li}^T = (w_{li,1}, \dots, w_{li,p})$ for sample i ($i = 1, \dots, n$) at the l th locus to be tested for association (i.e. H_{0l}), such as \mathbf{g}_l itself or an interaction between \mathbf{g}_l and an environment variable. Let $\mathbf{z}_i^T = (z_{i,1}, \dots, z_{i,q})$ denote q covariates (e.g. sex or age) of sample i to be adjusted in common for all L tests. We consider L hypothesis tests of the null hypothesis $H_{0l}: \beta_l = \mathbf{0}$ under the following regression model for the conditional mean of y_i with transformation,

$$\eta_i = \eta\{E(y_i | \mathbf{w}_{li}^T, \mathbf{z}_i^T)\} = \mathbf{w}_{li}^T \beta_l + \mathbf{z}_i^T \gamma_l, \tag{1}$$

for $i = 1, \dots, n$, where η is a monotone increasing function, and $\beta_l^T = (\beta_{l,1}, \dots, \beta_{l,p})$ and $\gamma_l^T = (\gamma_{l,1}, \dots, \gamma_{l,q})$ are the regression coefficients. The above model reduces to the ordinary linear regression model if η is the identity function and y_i follows a Gaussian distribution. The model reduces to the logistic regression model if η is the logit function and y_i follows a Bernoulli distribution. The model under H_{0l} (i.e. $\eta_i = \mathbf{z}_i^T \gamma_l$) is referred to as the null model throughout the paper.

In what follows, we show that the above test includes the joint GxE interaction test as well as the marginal association test. Let \mathbf{x}_i for the i th sample be an environment variable to be tested for interaction with $g_{l,i}$. Here we allow multivariate environment variables for \mathbf{x}_i . Then, if $\mathbf{w}_{li} = g_{l,i} \mathbf{x}_i$, the above test turns out to be the joint GxE interaction test of Kraft et al. [10] with \mathbf{x}_i as environment variables, in which the first element is one for all i . If $\mathbf{x}_i = 1$ for all i and $p = 1$, the test reduces to the marginal association test.

We study the chi-squared statistic for the score test of the above regression model (1). According to [23], the test statistic for testing $H_{0l}: \beta_l = \mathbf{0}$ can be expressed as follows:

$$t_l = \mathbf{u}^T (\mathbf{Q}_z \widetilde{\mathbf{W}}_l) (\widetilde{\mathbf{W}}_l^T \mathbf{Q}_z \widetilde{\mathbf{W}}_l)^{-1} (\mathbf{Q}_z \widetilde{\mathbf{W}}_l)^T \mathbf{u}, \tag{2}$$

where $\mathbf{Q}_z = \mathbf{I} - \mathbf{P}_z$, $\mathbf{P}_z = \widetilde{\mathbf{Z}} (\widetilde{\mathbf{Z}}^T \widetilde{\mathbf{Z}})^{-1} \widetilde{\mathbf{Z}}^T$, $\widetilde{\mathbf{W}}_l = \Omega^{1/2} \mathbf{W}_l$, $\widetilde{\mathbf{Z}} = \Omega^{1/2} \mathbf{Z}$, $\Omega = \text{diag}(\omega_1, \dots, \omega_n)$, $\mathbf{W}_l^T = (\mathbf{w}_{l,1}^T, \dots, \mathbf{w}_{l,n}^T)$, $\mathbf{Z}^T = (\mathbf{z}_1^T, \dots, \mathbf{z}_n^T)$, the ω_i s are positive values specific to the regression model, and $\mathbf{u}^T = (u_1, \dots, u_n)$ depends on y_1, \dots, y_n . The above t_l in (2) is just another representation of the standard score test statistic, and hence, the null distribution is asymptotically chi-squared with p degrees of freedom (pdf).

For example, t_l reduces to the score statistic for a logistic regression model by setting $u_i = (y_i - \hat{\mu}_i) / \sqrt{\omega_i}$ with $\omega_i = \hat{\mu}_i (1 - \hat{\mu}_i)$, $\hat{\mu}_i = 1 / \{1 + \exp(-\mathbf{z}_i^T \hat{\gamma})\}$, and $\hat{\gamma}$ is the maximum likelihood estimator under the null hypothesis $\beta_l = \mathbf{0}$. More generally, t_l reduces to the score statistic for a more general regression model having loglikelihood function $\ell = \ell(\eta)$ by setting $u_i = (\partial / \partial \eta_i) \ell / \sqrt{\omega_i}$ with $\omega_i = -(\partial^2 / \partial \eta_i^2) \ell$ evaluated at the null hypothesis $\beta_l = \mathbf{0}$. Furthermore, if $\omega_i = 1$ for all i , $\hat{\gamma} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}$, and $\mathbf{u} = \mathbf{y}$, then

$$T_l = \frac{t_l}{(\|\mathbf{Q}_z \mathbf{y}\|^2 - t_l) / n} \tag{3}$$

is approximately the t-test statistic in a Gaussian linear regression model. For the above representation of the score test statistic, see [23] for mathematical details.

We specifically provide the form of (2) for joint GxE interaction test. By defining $\tilde{\mathbf{X}} = \Omega^{1/2}\mathbf{X}$ where $\mathbf{X}^T = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)$, the relation $\mathbf{w}_{l,i} = g_{l,i}\mathbf{x}_i$ can be written in matrix form as $\mathbf{W}_l = \mathbf{G}_l\mathbf{X}$, and also as $\tilde{\mathbf{W}}_l = \mathbf{G}_l\tilde{\mathbf{X}}$, where $\mathbf{G}_l = \text{diag}(g_{l,1}, \dots, g_{l,n})$. Then, (2) is expressed as

$$t_l = \mathbf{u}^T \{ \mathbf{Q}_Z(\mathbf{G}_l\tilde{\mathbf{X}}) \} \{ (\mathbf{G}_l\tilde{\mathbf{X}})^T \mathbf{Q}_Z(\mathbf{G}_l\tilde{\mathbf{X}}) \}^{-1} \{ \mathbf{Q}_Z(\mathbf{G}_l\tilde{\mathbf{X}}) \}^T \mathbf{u}. \tag{4}$$

The Kraft's 2df test is obtained if $\mathbf{x}_i = (1, E_i)$ with an environment variable E_i considered for GxE interaction. For marginal association test, letting $\tilde{\mathbf{g}}_l = \Omega^{1/2}\mathbf{g}_l$, (2) is given by

$$t_l = \{ (\mathbf{Q}_Z\tilde{\mathbf{g}}_l)^T \mathbf{u} \}^2 / \tilde{\mathbf{g}}_l^T \mathbf{Q}_Z \tilde{\mathbf{g}}_l. \tag{5}$$

Systematic inflation (or deflation) of test statistics can be quantified by the overall behavior of genome-wide test statistics. Under the presence of population stratification, empirical distribution from test statistics is inflated from χ_1^2 to $\lambda\chi_1^2$ [24]. This inflation factor λ can be estimated from test statistics, t_1, \dots, t_L . Two estimators for λ are the median of t_1, \dots, t_L divided by the theoretical median of χ_1^2 distribution [17] and the mean [18]. Deviation of the estimated λ from one suggests that test statistics are problematic, e.g. due to the presence of population stratification or cryptic relatedness.

We study the mean of genome-wide test statistics analytically. Specifically, we approximate the expectation of the mean of test statistics for the L loci, $t_{mean} = \frac{1}{L} \sum_{l=1}^L t_l$,

$$E_{\mathbf{g}_1, \dots, \mathbf{g}_L}(t_{mean}) = \frac{1}{L} \sum_{l=1}^L E_{\mathbf{g}_l}(t_l) \tag{6}$$

by

$$\frac{1}{L} \sum_{l=1}^L \text{tr} \left[\{ E_{\mathbf{g}_l}(\tilde{\mathbf{W}}_l^T \mathbf{Q}_Z \tilde{\mathbf{W}}_l) \}^{-1} E_{\mathbf{g}_l} \{ (\mathbf{Q}_Z \tilde{\mathbf{W}}_l)^T \mathbf{u} \mathbf{u}^T (\mathbf{Q}_Z \tilde{\mathbf{W}}_l) \} \right],$$

where $E_{\mathbf{g}_1, \dots, \mathbf{g}_L}$ and $E_{\mathbf{g}_l}$ denote the expectations with respect to the joint distribution of $\mathbf{g}_1, \dots, \mathbf{g}_L$ and to the marginal distribution of \mathbf{g}_l , respectively. If $q = 1$, the matrix inverse is just the reciprocal. Hence, the proposed approximation is a multi-dimensional extension of the approximation of the mean of a ratio by the ratio of means. In order to derive the above approximation, we impose the following assumption for each tested variant independently: $g_{l,1}, \dots, g_{l,n}$ are independently and identically distributed whose all moments are finite, with mean and variance denoted by μ_l and σ_l^2 , respectively (e.g. with a binomial distribution of size 2 and success probability being the minor allele frequency (MAF), which is the law under Hardy–Weinberg equilibrium (HWE)). We also assume that the \mathbf{g}_l are independent of \mathbf{u} , \mathbf{Z} , and \mathbf{X} . Since the approximation applies separately for each l , each variant may have a different genotype distribution (i.e. different MAF), and variant frequencies may be correlated due to linkage disequilibrium. The above assumptions exclude the case where the tested variant \mathbf{g}_l itself causes systematically inflated test statistics, e.g., due to population stratification, cryptic relatedness, sample difference of genotyping efficiency, or a batch effect. In addition, we assume that $p \leq q$ and \mathbf{Z} contains \mathbf{X} . Without loss of generality, the first p columns of \mathbf{Z} correspond to \mathbf{X} , or $\mathbf{Z} = (\mathbf{Z}_{1:p}, \mathbf{Z}_{(p+1):q}) = (\mathbf{X}, \mathbf{Z}_{(p+1):q})$, where $\mathbf{Z}_{1:p}$ represents the first p columns of \mathbf{Z} and $\mathbf{Z}_{(p+1):q}$ are the remaining columns. Now we study the expectation of t_l with respect to \mathbf{g}_l . Our approximation formula for the expectation of t_l with

respect to \mathbf{g}_l is:

$$\begin{aligned}
 E_{\mathbf{g}_l}(t_l) &\approx \text{tr} \left[\{E_{\mathbf{g}_l}(\widetilde{\mathbf{W}}_l^T \mathbf{Q}_Z \widetilde{\mathbf{W}}_l)\}^{-1} E_{\mathbf{g}_l} \{(\mathbf{Q}_Z \widetilde{\mathbf{W}}_l)^T \mathbf{u} \mathbf{u}^T (\mathbf{Q}_Z \widetilde{\mathbf{W}}_l)\} \right] \\
 &= \text{tr} \left[\left\{ \sum_{i=1}^n \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^T (\mathbf{Q}_Z)_{ii} \right\}^{-1} \left\{ \sum_{i=1}^n \widetilde{\mathbf{x}}_i \widetilde{\mathbf{x}}_i^T (\mathbf{Q}_Z \mathbf{u})_i^2 \right\} \right] \\
 &= t_{approx},
 \end{aligned} \tag{7}$$

where the approximation holds by ignoring $O(n^{-1})$ terms. The derivation is given in [S1 Appendix](#), in which the proof is based on an induction and asymptotic expansion. In the above formula, $(\mathbf{Q}_Z)_{ii}$ is 1 minus the leverage score of the i th datum, $1 - (\mathbf{P}_Z)_{ii}$, while $(\mathbf{Q}_Z \mathbf{u})_i$ is the i th residual from a regression of \mathbf{u} on $\widetilde{\mathbf{Z}}$. Notably, the formula (7) no longer contains characteristics of \mathbf{g}_l . As a result, (6) is approximated by (7). The formula (7) can be used to investigate the overall behavior of t_l without requiring a genome-wide scan. In [S1 Appendix](#), we show that, if the null model is correct, the above t_{approx} is close to p , the expected value of t_{mean} . A large difference between t_{approx} and p indicates problematic null model specification because we assume that the \mathbf{g}_s do not cause a problem. Analogous to the genomic inflation factor, we consider the scaled version of t_{approx} ,

$$l_{approx} = t_{approx}/p,$$

and, similarly, $l_{mean} = t_{mean}/p$. The approximation formula for the Gaussian linear regression model (3) is

$$T_{approx} = \frac{t_{approx}}{(\|\mathbf{Q}_Z \mathbf{y}\|^2 - t_{approx})/n}. \tag{8}$$

Similarly, the scaled versions are $l_{approx} = T_{approx}/p$ and $l_{mean} = T_{mean}/p$, in which $T_{mean} = \frac{1}{L} \sum_{l=1}^L T_l$.

The case where l_{approx} is close to one suggests that the null model is appropriate (or at least has no problematic behavior), and it is expected that the test statistics behave properly unless genetic variants cause problems in the test statistics. Checking systematic inflation or deflation by marginal association scan allows to check whether genetic variants cause problems. On the other hand, a large discrepancy from one suggests null model misspecification, in which case systematically inflated test statistics will be observed after genome-wide scan and the test is unreliable.

Real data application

We illustrate the validity and usefulness of our approximation through application to a real GWAS dataset obtained from the publicly available Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early Alzheimer’s disease. For up-to-date information, see www.adni-info.org. ADNI is an ongoing, longitudinal study with primary purpose being to explore the genetic and neuroimaging information associated with late-onset Alzheimer’s disease (LOAD). The study investigators recruited elderly subjects older than 65 years. The cohort comprised about 400 subjects with mild cognitive impairment, about 200 subjects with Alzheimer’s disease, and about 200 healthy controls.

Each subject was followed for at least 3 years. During the study period, the subjects were assessed with magnetic resonance imaging (MRI) measures and psychiatric evaluation to determine cognitive status at each time point. Study subjects gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed questionnaires approved by Institutional Review Board (IRB) of each participating sites (http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf). We obtained approval from the ADNI Data Sharing and Publications Committee for use of the data and analyzed the data anonymously.

The ADNI-GWAS data were obtained from 818 DNA samples of ADNI1 participants using the Illumina Human 610-Quad genotyping array [25]. The genotype data we used is in PLINK format available from ADNI website (<http://adni.loni.usc.edu/>) which includes 620,901 SNPs for 757 individuals. We applied a quality control procedure by excluding SNPs with missing genotype rate > 0.05 , HWE test $P < 10^{-10}$, and MAF $< 5\%$; the total number of remaining SNPs was 521,203.

The dataset with 757 samples is comprised of multiple ethnic groups. We computed principal components (PCs) using the EIGENSOFT package [26, 27]. The first and second PCs are given in S1 Fig, which indicates the presence of population stratification. Our approximation imposes a stringent assumption that all samples follow the same distribution for each locus. It does not cover the case where population stratification exists, and the approximation is not guaranteed under the presence of population stratification due to association between phenotype and genotypes. To see the performance on data without population stratification, we created another dataset by extracting 684 non-Hispanic Caucasian samples from 757 samples after excluding one individual from pairs showing cryptic relatedness (revealed by the PLINK [28] pairwise $\hat{\pi}$ statistic being greater than 0.125), and we excluded subjects whose reported sex did not match the sex inferred from X chromosome SNPs. We used two datasets, one with the 684 samples and another with the 757 samples. Since population stratification is absent in the former dataset, it is expected that l_{approx} well approximates l_{mean} while it is not guaranteed that the approximation is well for the latter dataset. To make the approximation workable, we considered adjusting for PCs as covariates [26]. Although an appropriate number of PCs for adjustment depends on the population structure and the sample size, we included top 10 PCs, which is the number generally accepted within the psychiatric genetics community [29]. We also considered top 3 and 5 PCs for covariates in order to evaluate the influence of the number of PCs for adjustment.

To check the accuracy of our approximation, we compared l_{approx} with l_{mean} computed from the genome-wide test statistics from the joint GxE interaction and marginal association analyses. We obtained environment variables, phenotypes, and covariates from the R package ADNIMERGE provided by ADNI. We chose five phenotypes: height (HEIGHT), body mass index (BMI), whole brain volume (WholeBrain), intracranial volume (ICV), and mini mental state examination (MMSE), which are all quantitative. For environment variables in GxE interaction analyses, we used 142 metabolite variables stored in the `admdukep180fia` object in ADNIMERGE gathered by the Alzheimer's disease Metabolomics Consortium. By setting values coded as " $< LOD$ " as missing, and excluding metabolites showing missing rate greater than 20%, we had 117 metabolites for analysis. Because conducting genome-wide scans for all 117 variables is time-consuming, we used only nine variables—`lysoPC.a.C16.0`, `PC.ae.C38.2`, `PC.ae.C40.3`, `C10`, `PC.aa.C40.5`, `PC.ae.C36.3`, `SM..OH..C14.1`, `SM..OH..C22.1`, and `SM.C24.0`—as environment variables for genome-wide joint GxE interaction analyses using linear models for the five phenotypes. Among the nine variables, two variables were chosen based on application of the proposed approximation formula to the joint GxE interaction test with each of the five phenotypes and each of 117 metabolite variables as the environment variable. First,

we computed l_{approx} for joint GxE interaction test with respect to the five phenotypes and 117 metabolites by adjusting for sex and age. We stored the results in [S1 Table](#). The scatter plots of the phenotype-environment pairs showing $l_{approx} > 1.5$ are given in [S2 Fig](#). There seemed to be roughly two groups: the first group exhibited quadratic relationship between phenotype and environment variable rather than linear, and the second group included seeming outliers. Therefore, the large discrepancy of l_{approx} from one was caused by the null model specification. For the first group, we randomly chose the BMI—PC.ae.C38.2 pair. For the second group, we randomly chose the BMI—PC.ae.C40.3 pair. The remaining seven metabolites were chosen randomly.

In the analyses for the nine metabolites and five phenotypes, we attempted to automatically resolve the systematic inflation by using a Box–Cox transformation [30] of the phenotype (actually, the transformation was applied to the phenotype subtracted the minimum value and added 1 to make the values positive). First, we applied the standard Box–Cox transformation based on normality (i.e. making the transformed phenotype distribution close to normal) using the “boxCox” function implemented in the `car` package in R. Next, we optimized the Box–Cox transformation in terms of the closeness of l_{approx} to one, which addresses the systematic inflation issue directly.

Our proposed approximation can also be applied to the score test for generalized linear models. We dichotomized the five quantitative phenotypes by whether the phenotype value is greater than its mean, and applied the score test for a logistic regression model to joint GxE interaction and marginal association tests.

We considered the impact on the approximation when the real genotype data is replaced by simulated genotype data. We repeated the same analyses for the nine metabolites and five phenotypes on the ADNI 684 samples described above with the simulated genotype data rather than with the real genotype data, while phenotype, environment, and covariates were fixed. In the above simulation procedure, 10,000 unlinked loci were simulated using PLINK `--simulate` option, where MAFs were randomly generated from a uniform distribution in [0.05,0.5]. We also repeated the above analyses using the artificial genotype data for the ADNI data with the 757 samples.

To see the impact of null model specification, we investigated the BMI—PC.ae.C38.2 and BMI—PC.ae.C40.3 pairs in detail. We considered sophisticating null modeling by applying a quadratic model rather than the linear model or removing outliers. Voorman et al. [15], Tchetgen Tchetgen and Kraft [31] and Almli et al. [16] proposed a robust test using the Huber–White robust variance for the GxE interaction test to account for null model misspecification. We applied the robust joint GxE interaction test using the Almli’s robust joint interaction program available from “<http://genetics.emory.edu/labs/epstein/software/robust-joint-interaction/>”. Another solution to the systematic inflation of test statistics is to use the genomic control-adjusted p-values, i.e. chi-squared test statistics divided by a constant so that the median matches the expected value of one. We applied the genomic control-adjustment for the two examples.

Simulation studies

To study the behavior of l_{approx} for joint GxE interaction and marginal association tests, we carried out extensive simulation experiments. Simulation scenarios are described in [Table 1](#). The aims of each scenario are as follows.

- Baseline scenario. This is a baseline for comparison. Other scenarios are a slight modification of the baseline scenario.

Table 1. Description of simulation scenarios.

Scenario	Description
Baseline scenario	
Base	No association among environment, covariate variable and genotypes
	Correctly specified null model
	One covariate and environment variables are normally distributed (continuous)
	Genotypes are in linkage equilibrium with uniformly distributed minor allele frequencies
Association among environment, covariate variables and/or genotypes	
1a	Genotypes are associated with covariate
1b	Genotypes are associated with environment variable
1c	Genotypes are associated with covariate and environment variables
1d	Environment variable is associated with covariate.
Misspecified null model	
2a	Covariate associated with genotypes is missed
2b	Covariate associated with genotypes and environment variable is missed
2c	Linear null model is incorrectly specified
2d	One outlier is included
2e	Ten outliers are included
Environment/covariate variable distribution	
3a	Five covariates and environment variables are normally distributed (continuous)
3b	One covariate and environment variables are uniformly distributed (continuous)
3c	One covariate and environment variables are binary variable (binary categorical)
3d	One covariate and environment variables are ordinal variable (ordinal categorical)
Genotype distribution	
4a	Genotypes are in linkage disequilibrium with uniformly distributed minor allele frequencies
4b	Genotypes are in linkage equilibrium with Beta distributed minor allele frequencies
4c	Genotypes are in linkage disequilibrium with Beta distributed minor allele frequencies

<https://doi.org/10.1371/journal.pone.0219825.t001>

- Association among environment, covariate variables and/or genotypes (Scenarios 1a, 1b, 1c, and 1d). The aim is to assess the influence on the proposed approximation of association among environment, covariate variables and/or genotypes. Scenarios 1a, 1b, and 1c consider the association between covariate/environment variables and genotypes, which may arise due to population stratification or genetic architecture. Scenario 1d considers the association between environment and covariate variables.
- Misspecified null model (Scenarios 2a, 2b, 2c, 2d, and 2e). The aim is to assess the impact of the misspecified null model on the approximation. It is expected that the misspecification deviates the value of l_{mean} from one. Scenarios 2a and 2b consider that the null model misses the covariate associated with genotypes, for example, adjustment for population stratification is not applied or inadequate. Scenario 2c considers that the null model misspecifies the functional form of the environment variable. Scenarios 2d and 2e consider the presence of outlier(s).
- Environment/covariate variable distribution (Scenarios 3a, 3b, 3c, and 3d). This scenario is intended to investigate the approximation performance under several kinds of environment and covariate variables. We consider four scenarios, considering continuous (e.g. age), binary categorical (e.g. sex), and ordinal categorical (e.g. questionnaire score) variables.
- Genotype distribution (Scenarios 4a, 4b, and 4c). This scenario is intended to evaluate the impact of the distribution of the genetic variants. Scenario 4a considers the correlation

between genetic variants due to linkage disequilibrium, while scenarios 4b and 4c consider the different allele frequency spectrum.

We considered four effect size scenarios. Let b_G , b_Z , and b_{GE} denote the parameters of genotype, covariate and GxE interaction effects on the phenotype. Then, the four scenarios are given as triplets (b_G, b_Z, b_{GE}) : (0, 0, 0) (no effect of genotype, covariates and GxE interaction), (1, 0, 0) (genotype effect, and no covariates and GxE interaction effects), (0, 1, 0) (covariate effect, and no genotype and GxE interaction effects) and (0, 0, 1) (GxE interaction effect, and no genotype and covariate effects). For the scenarios under the presence of any genotypic effect, $(b_G, b_Z, b_{GE}) = (1, 0, 0)$ and $(0, 0, 1)$, we considered three genotype codings, additive, recessive, and dominant. We repeated the simulations 200 times to compare l_{approx} with l_{mean} . Two sample sizes, $n = 1000$ and 10000 , were considered. To evaluate the discrepancy between l_{approx} and l_{mean} , we summarized mean and standard deviation in the 200 simulation runs. In some scenarios, the magnitude of l_{mean} varied with sample size. Thus, we considered the ratio l_{approx}/l_{mean} , which is useful to quantify how well l_{approx} approximated l_{mean} while excluding the impact of the sample size. [S1 Appendix](#) describes the technical details of simulation studies. The program code is provided in [S2 Appendix](#).

Results

Real data application: Comparison between l_{approx} and l_{mean}

[Fig 1](#) gives comparisons between l_{mean} and l_{approx} for joint GxE interaction and marginal association tests for each of five quantitative phenotypes and nine metabolites set as an environment variable on the ADNI dataset with 684 non-Hispanic Caucasian samples showing no population stratification. The top left and right panels in [Fig 1](#) give the results from joint GxE interaction and marginal association tests, respectively, which show that l_{approx} approximated l_{mean} well as seen by that all points were gathered around the diagonal line. In the top left panel, there were cases where l_{mean} showed a large discrepancy from one, and two of them are further investigated below. On the other hand, all points in the top right panel were concentrated around one, that is, the means of the chi-squared test statistics from marginal association test were all made closer to one. This behavior differs from that of the joint GxE interaction test, in which the target environment variable x_i plays an important role in determining the test statistic distribution, and dependence of l_{approx} on x_i cannot be ignored. In [S1 Appendix](#), we show that l_{approx} is close to one if $n \rightarrow \infty$ when $x_i = 1$ for all i with $p = 1$ (i.e. the model is reduced to the model for the marginal association test and the corresponding null distribution is 1df chi-squared), which in turn implies that the mean of the chi-squared statistics is approximately one irrespective of what the null model is used.

The case with large l_{mean} observed in the top left panel means the presence of systematic inflation in genome-wide test statistics. Middle left panel of [Fig 1](#) gives the comparison between l_{approx} and l_{mean} for the phenotypes with Box–Cox transformation based on normality. However, systematic inflation still appeared. On the other hand, the Box–Cox transformation optimized in terms of the closeness of l_{approx} mitigated the systematic inflation as seen in the middle right panel of [Fig 1](#).

The bottom left and right panels in [Fig 1](#) give the results for binary phenotype, created by dichotomizing quantitative phenotype, from joint GxE interaction and marginal association tests for quantitative phenotypes, respectively. The approximation looks well for the 684 non-Hispanic Caucasian samples, exhibiting a similar tendency of the results for quantitative phenotypes. Note that l_{approx} will not always be close to one for models other than Gaussian unless

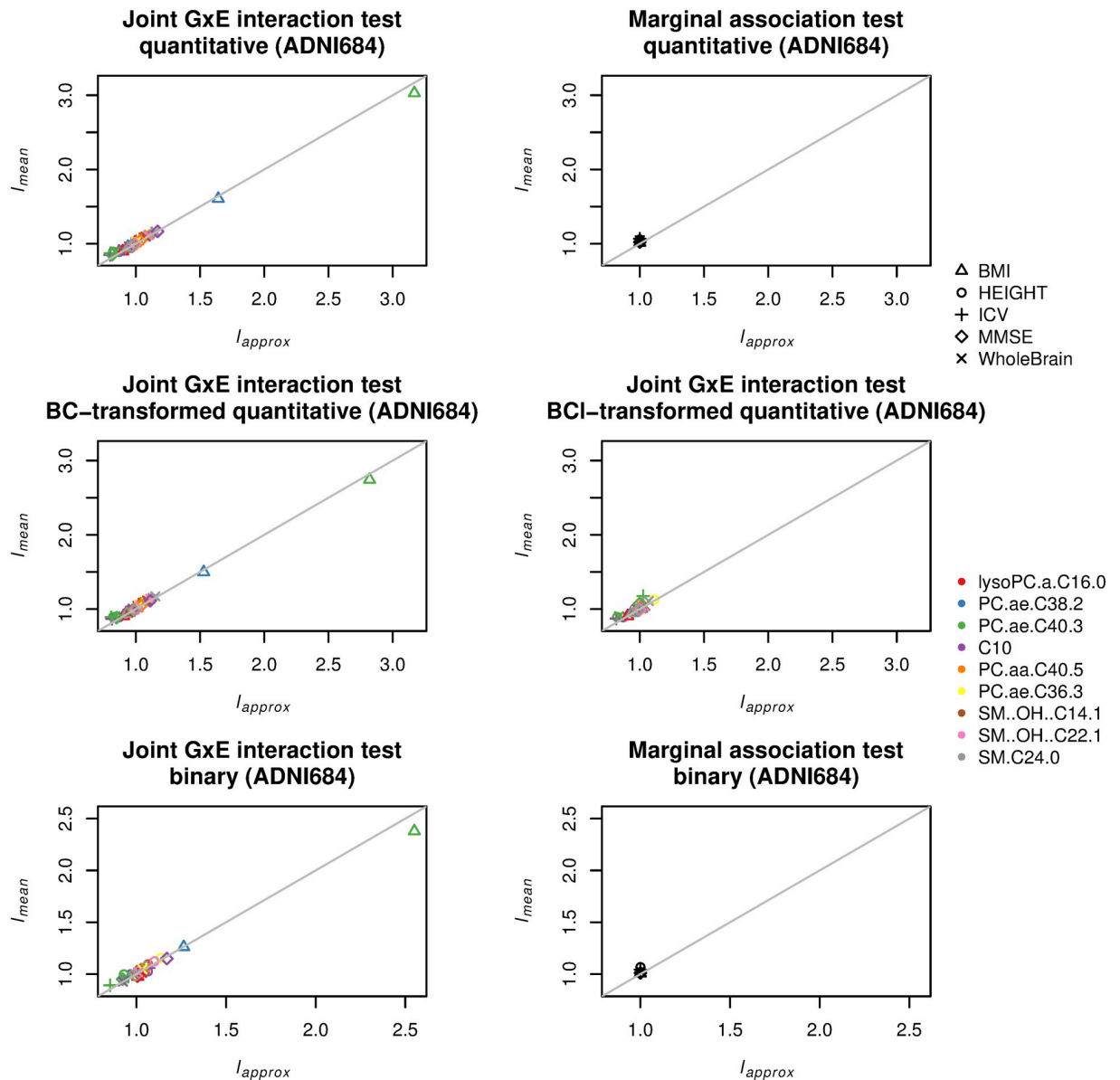


Fig 1. Comparison between l_{approx} and l_{mean} for joint GxE interaction and marginal association tests on ADNI-GWAS data for 684 non-Hispanic Caucasian samples (ADNI684). Five phenotypes—height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE)—and nine environment variables: lysoPC.a.C16.0, PC.ae.C38.2, PC.ae.C40.3, C10, PC.aa.C40.5, PC.ae.C36.3, SM.OH..C14.1, SM.OH..C22.1, and SM.C24.0. Top left: joint GxE interaction test (quantitative phenotype). Top right: marginal association test (quantitative phenotype). Middle left: joint GxE interaction test (quantitative phenotype after Box-Cox transformation optimized in terms of normality). Middle right: joint GxE interaction test (quantitative phenotype after Box-Cox transformation optimized in terms of l_{approx}). Bottom left: joint GxE interaction test (binary phenotype created by dichotomizing quantitative phenotype). Bottom right: marginal association test (binary phenotype created by dichotomizing quantitative phenotype).

<https://doi.org/10.1371/journal.pone.0219825.g001>

the null model is correctly specified. Nevertheless, l_{approx} for marginal association tests was close to one in a real GWAS data application.

The proposed l_{approx} was derived under the assumption that each of genetic variants is an independently and identically distributed random variable, and also is independent of phenotype, environment, and covariates. Therefore, the scaled mean of test statistics computed from genotype data, simulated under the assumptions above, with other variables (phenotype,

environment, and covariates) being fixed, must be close to l_{approx} . The results with the artificial genotype data for the 684 samples are given in [S3 Fig](#), and the plots were very similar to [Fig 1](#), verifying our approximation.

[Fig 2](#) gives comparisons between l_{mean} and l_{approx} for joint GxE interaction and marginal association tests with the five phenotypes and nine metabolites as the environment variable on the ADNI dataset with 757 samples showing population stratification as seen in [S1 Fig](#). The top left and right panels in [Fig 2](#) are the results from joint GxE interaction and marginal association tests for quantitative phenotypes without PC adjustment, respectively. Unlike the case with 684 samples (the top left and right panels in [Fig 1](#)), there were points deviated from diagonal line in both panels (l_{approx} often underestimated l_{mean} for lower l_{mean}). Results for binary phenotypes were in panels in the third row, and the similar deviations were observed. The left and right panels in the second row of [Fig 2](#) are the results from joint GxE interaction and marginal association tests for quantitative phenotypes with adjustment for top 10 PCs, respectively. In this case, all points were distributed around the diagonal line, meaning that the adjustment by top 10 PCs could resolve the l_{approx} 's underestimation of l_{mean} . For binary phenotypes given in the bottom panels, l_{approx} 's underestimation was resolved similarly. [S4 Fig](#) provides the results with adjustment for top 3 and top 5 PCs. No deviation was seen, implying that the PC adjustment was still successful.

The results with the artificial genotype data for the 757 samples are given in [S5 Fig](#). Unlike [Fig 2](#), there was no deviation between l_{mean} and l_{approx} which is the expected behavior since the simulated genotype data was generated under the assumption that l_{approx} is derived. It in turn implies that some of the assumptions for l_{approx} were violated in the cases of [Fig 2](#) (the top panels and the panels in the third row) where deviation was observed, and population stratification could be one possible reason because PC adjustment resolved the deviation.

Real data application: Detailed analysis of two datasets that showed large systematic inflation

Large discrepancy of l_{approx} from one suggests that the specified null model is problematic. Here, we investigated two analyses that showed large discrepancy of l_{approx} from one as described in the Materials and Methods section. [Fig 3](#) gives the first example with PC.ae.C38.2 as the environment variable in the ADNI data with 684 samples. The value of l_{mean} was 1.61. A scatter plot of BMI versus the environment variable ([Fig 3](#), top left) shows that a linear model for the environment variable is inadequate and suggests a quadratic model for better fit to the data. Given this consideration, we modified the null model by including a quadratic term, and indeed this mitigated the problematic behavior as shown in the middle left panel of [Fig 3](#); l_{mean} was reduced to 1.23. Alternatively, the plot suggests that there may be outliers. We calculated the Cook's distance [32] on linear regression model for BMI with explanatory variables, age, sex, and PC.ae.C38.2, and a single observation of index 164 had Cook's distance larger than the mean plus $4 \times sd$ (standard deviation). See also [S6 Fig](#) for regression diagnostics plot. We emphasized the influential observation in the panel by showing the index 164. When this sample was removed, the null model with linear effect mitigated the systematic inflation behavior as shown in the middle right panel of [Fig 3](#); l_{mean} was reduced to 1.15.

The result of the robust joint GxE interaction test using the Alml's program is given in the bottom left panel of [Fig 3](#). The systematic inflation still remained even when the robust test was applied. The result of the genomic control adjustment is given in the bottom right panel of [Fig 3](#). The severe problematic behavior could not be resolved.

[Fig 4](#) gives the second example that uses PC.ae.C40.3 as the environment variable for BMI as phenotype in the ADNI data with 684 samples. The value of l_{mean} was 3.03. A scatter plot

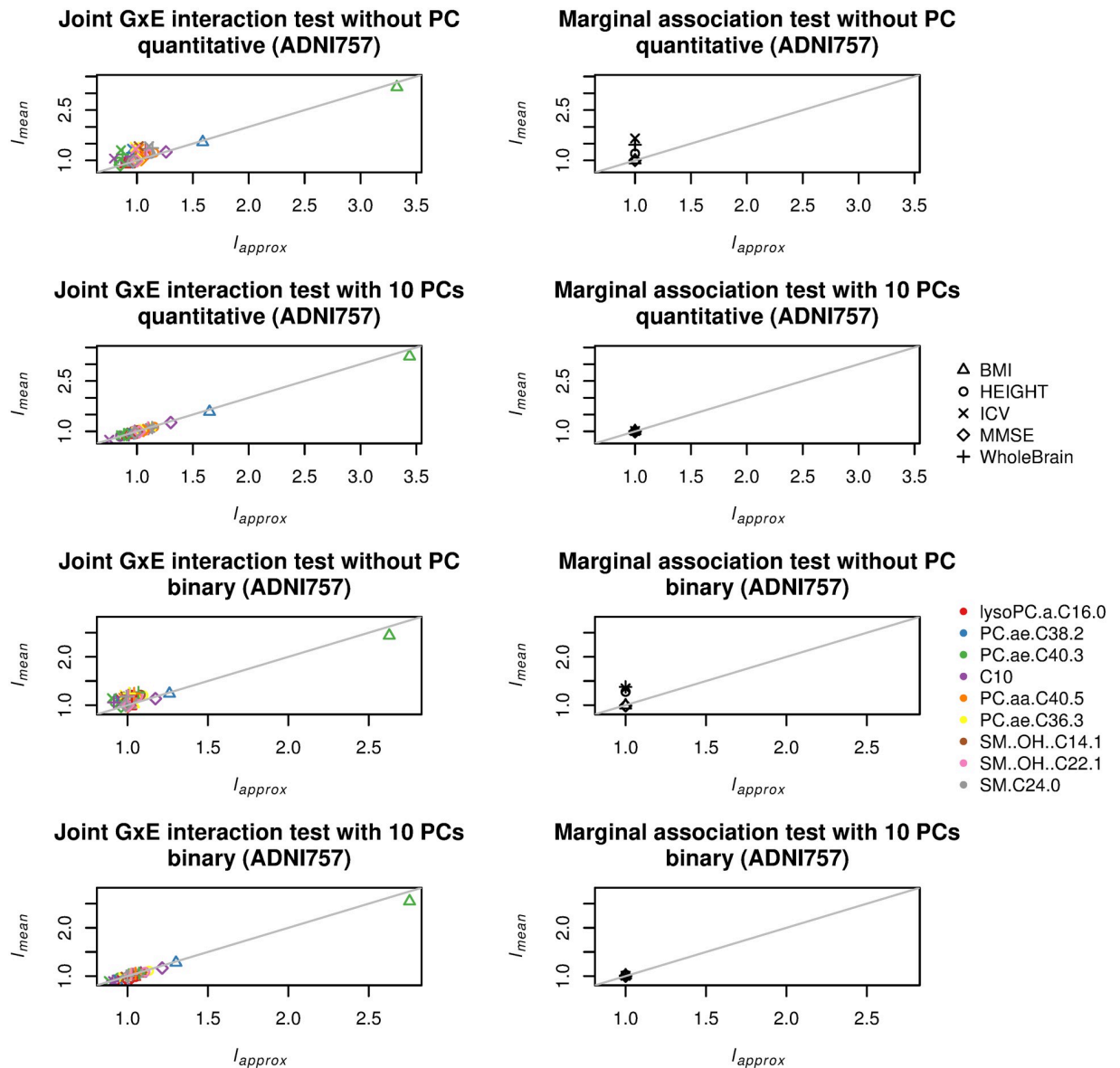


Fig 2. Comparison between I_{approx} and I_{mean} for joint GxE interaction and marginal association tests on ADNI-GWAS data for 757 samples which showed population stratification (ADNI757). Five phenotypes—height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE)—and nine metabolite variables: lysoPC.a.C16.0, PC.ae.C38.2, PC.ae.C40.3, C10, PC.aa.C40.5, PC.ae.C36.3, SM.OH.C14.1, SM.OH.C22.1, and SM.C24.0. Results are shown without and with adjustment for top 10 principal components (PCs). Top left: joint GxE interaction test without PC adjustment (quantitative phenotype). Top right: marginal association test without PC adjustment (quantitative phenotype). Left in the second row: joint GxE interaction test with adjustment for top 10 PCs (quantitative phenotype). Right in the second row: marginal association test with adjustment for top 10 PCs (quantitative phenotype). Left in the third row: joint GxE interaction test without PC adjustment (binary phenotype created by dichotomizing quantitative phenotype). Right in the third row: marginal association test without PC adjustment (binary phenotype created by dichotomizing quantitative phenotype). Bottom left: joint GxE interaction test with adjustment for top 10 PCs (binary phenotype created by dichotomizing quantitative phenotype). Bottom right: marginal association test with adjustment for top 10 PCs (binary phenotype created by dichotomizing quantitative phenotype).

<https://doi.org/10.1371/journal.pone.0219825.g002>

(Fig 4, top left) of BMI versus the environment variable suggests that two points labeled as 164 and 324 on the right side could be outliers. We calculated the Cook’s distance on linear regression model for BMI with explanatory variables, age, sex, and PC.ae.C40.3, and observations 164 and 324 had Cook’s distance larger than the mean plus $4 \times \text{sd}$ (standard deviation), where

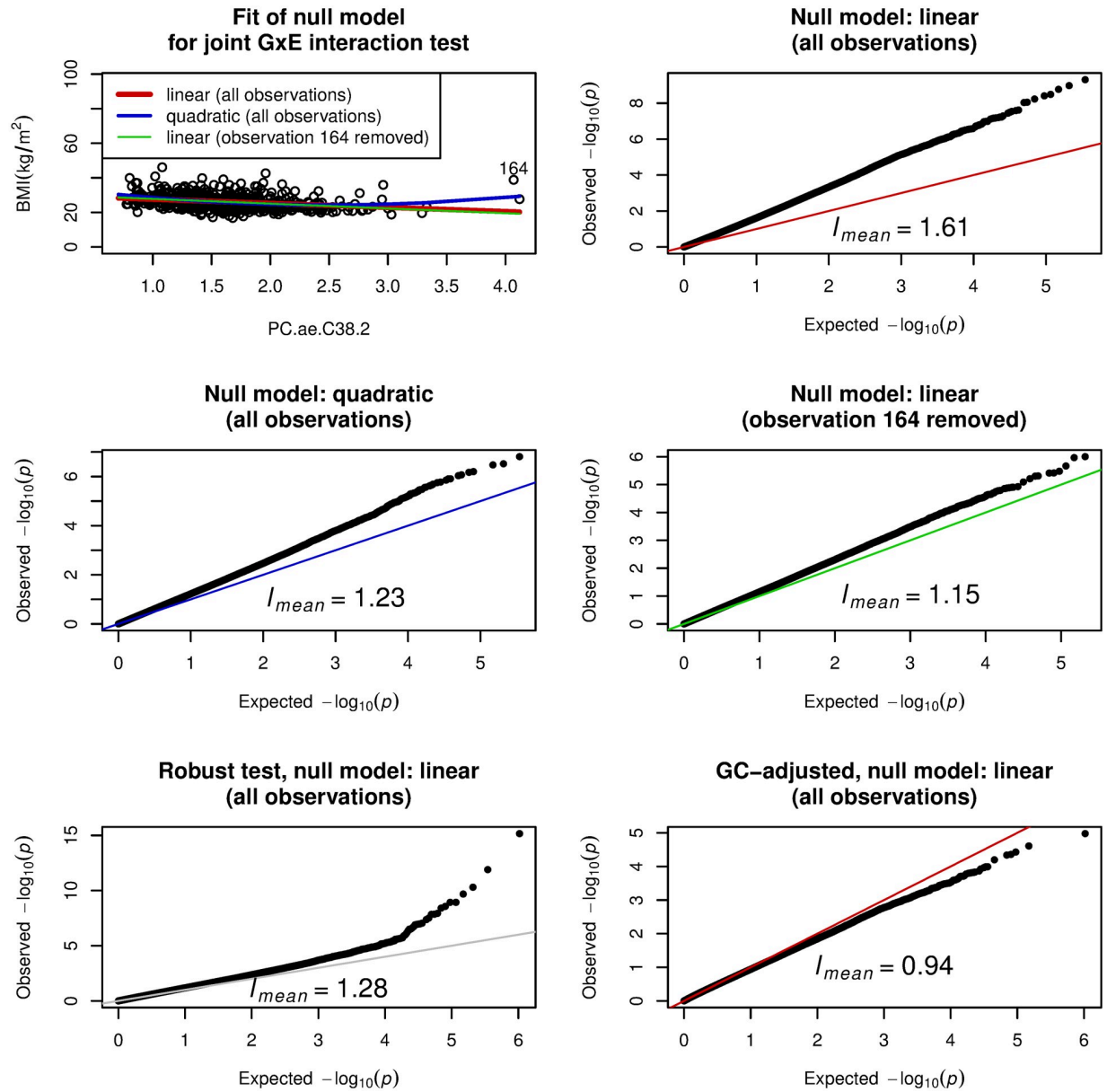


Fig 3. Systematic inflation in gene-PC.ae.C38.2 (a metabolite) interaction test for body mass index (BMI) on ADNI-GWAS data with 684 samples. Joint GxE interaction test applies to BMI as a quantitative phenotype and PC.ae.C38.2 as the environment variable by adjusting sex and age as covariates. Top left: scatter plot for BMI and PC.ae.C38.2 with emphasis on observation 164, which was the influential observation having Cook's distance larger than the mean plus $4 \times \text{sd}$ (standard deviation). Three null models are given: linear (all observations), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C38.2} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C38.2}$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , quadratic (all observations), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C38.2} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C38.2} + \gamma_5 \text{PC.ae.C38.2}^2$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , linear (observation 164 removed), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C38.2} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C38.2}$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , where the influential observation (index 164) is removed. Top right: quantile-quantile (QQ) plot from joint GxE interaction test with the linear null model on all observations. Middle left: QQ plot from joint GxE interaction test with the quadratic null model for all observations. Middle right: QQ plot from joint GxE interaction test with the quadratic null model for which observation 164 is removed. Bottom left: QQ plot from the Almlri's robust joint GxE interaction test [16] with the linear null model for all observations. Bottom right: QQ plot from genomic control (GC) adjusted joint GxE interaction test with the linear null model for all observations. I_{mean} , the scaled mean of genome-wide test statistics, is one if the null model is correctly specified.

<https://doi.org/10.1371/journal.pone.0219825.g003>

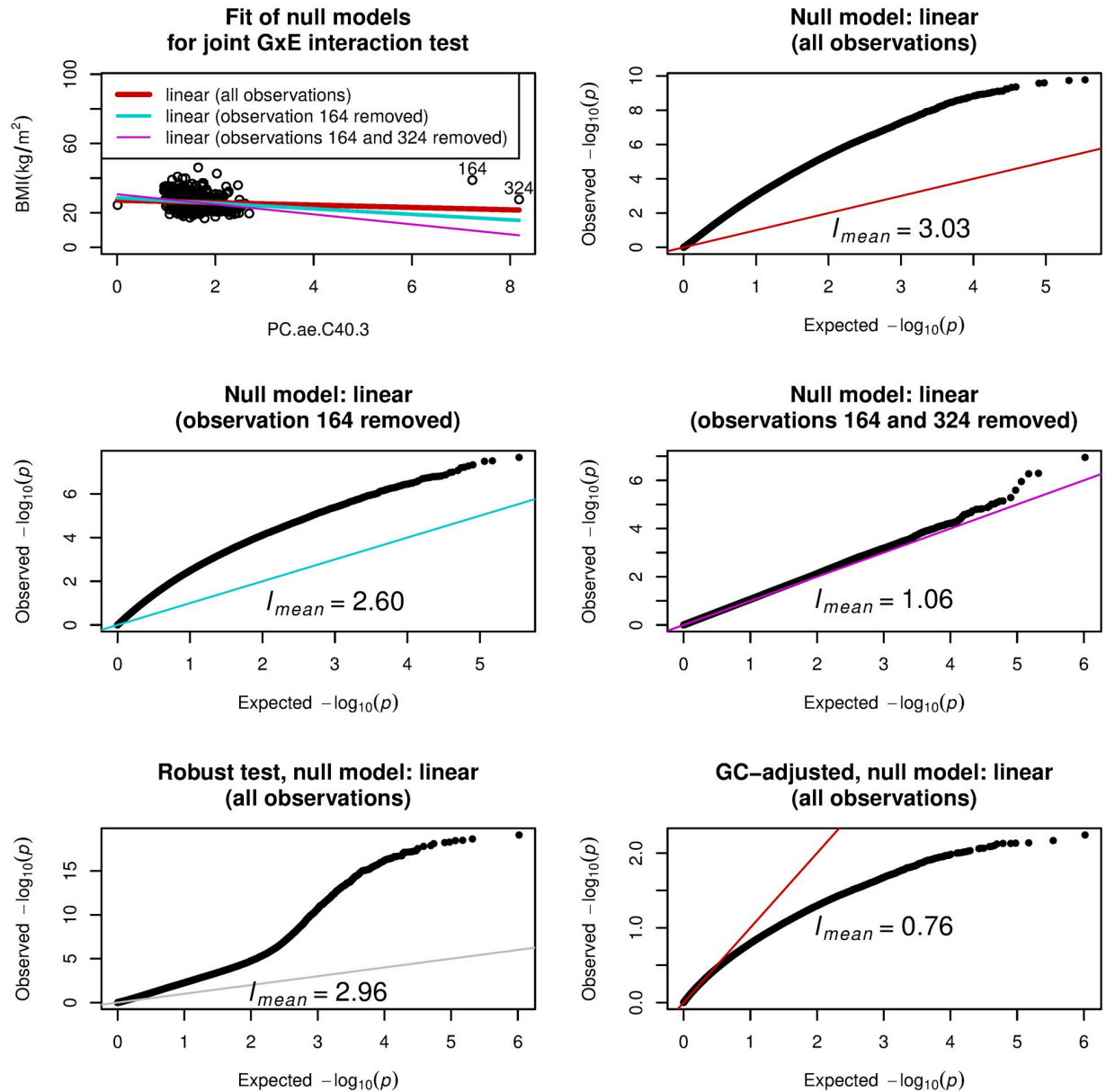


Fig 4. Systematic inflation in gene-PC.ae.C40.3 (a metabolite) interaction test for body mass index (BMI) on ADNI-GWAS data with 684 samples. Joint GxE interaction test is applied with BMI as a quantitative phenotype and PC.ae.C40.3 as the environment variable with sex and age are adjusted. Top left: scatter plot for BMI and PC.ae.C40.3 with emphasis on observations 164 and 324, which were the influential observations having Cook's distance larger than the mean plus $4 \times \text{std}$ (standard deviation), where observation 164 has larger Cook's distance than observation 324. Three null models are given: linear (all observations), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C40.3} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C40.3}$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , linear (observation 164 removed), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C40.3} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C40.3}$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , where the most influential observation (index 164) is removed, linear (observations 164 and 324 removed), $\text{BMI} = \beta_1 G_l + \beta_2 G_l \times \text{PC.ae.C40.3} + \gamma_1 + \gamma_2 \text{sex} + \gamma_3 \text{age} + \gamma_4 \text{PC.ae.C40.3}$ with $\beta_1 = \beta_2 = 0$ for l th genetic variant G_l , where two most influential observations (164 and 324) are removed. Top right: quantile-quantile (QQ) plot from joint GxE interaction test with the linear null model for all observations. Middle left: quantile-quantile (QQ) plot from joint GxE interaction test with the linear null model for which observation 164 is removed. Middle right: quantile-quantile (QQ) plot from joint GxE interaction test with the linear null model for which observations 164 and 324 are removed. Bottom left: QQ plot from the Almlri's robust joint GxE interaction test [16] with the linear null model for all observations. Bottom right: QQ plot from genomic control (GC) adjusted joint GxE interaction test with the linear null model for all observations. I_{mean} , the scaled mean of genome-wide test statistics, is one if the null model is correctly specified.

<https://doi.org/10.1371/journal.pone.0219825.g004>

observation 164 had larger Cook's distance than observation 324. See also [S7 Fig](#) for regression diagnostics plot. When observation 164 was removed, the null model with linear effect mitigated the systematic inflation behavior as shown in the middle left panel of [Fig 3](#); l_{mean} was reduced to 2.6. When observations 164 and 324 were removed, the null model with linear effect further mitigated the systematic inflation behavior as shown in the middle right panel of [Fig 3](#); l_{mean} was reduced to 1.06.

[Figs 5 and 6](#) provide results of the BMI after Box–Cox transformation for the data with the 684 and 757 samples, respectively. The Box–Cox transformation of BMI based on normality was insufficient to reduce the systematic inflation, as seen in the top right panels of [Figs 5 and 6](#). On the other hand, the Box–Cox transformation of BMI based on l_{mean} reduced the systematic inflation, as seen in the bottom right panels of [Figs 5 and 6](#).

Simulation studies

Here, we provide the results of joint GxE interaction and marginal association tests from various simulation studies. [Tables 2 and 3](#) include particularly noteworthy results while [S2 and S3](#) Tables include the remainings. In derivation of l_{approx} , we have assumed that genotypes are not associated with phenotypes, environment and covariate variables. Simulation scenarios $(b_G, b_Z, b_{GE}) = (0, 0, 0)$ and $(0, 1, 0)$ satisfy the above assumptions except in scenarios 1a, 1b, 1c, 2a, and 2b. In such cases, the approximation of l_{approx} to l_{mean} was well as the ratio l_{approx}/l_{mean} was close to one. On the other hand, in scenarios $(b_G, b_Z, b_{GE}) = (1, 0, 0)$ and $(0, 0, 1)$, phenotypes are associated with genotypes, and hence the assumptions are not satisfied. In such cases, there is no guarantee that l_{approx} approaches to l_{mean} even if n is increased. Actually, there were several scenarios showing the ratio l_{approx}/l_{mean} for $n = 10000$ that was more different from one than that for $n = 1000$. We provide brief summaries not mentioned above.

- Association among environment, covariate variables and/or genotypes (Scenarios 1a, 1b, 1c and 1d). In scenarios $(b_G, b_Z, b_{GE}) = (0, 0, 0)$ and $(0, 1, 0)$ under the presence of the association between covariate and environment variables, the assumption for the derivation of l_{approx} holds, that is, genotypes are not associated with phenotype, covariate/environment variables. Indeed, l_{approx} approximated l_{mean} well. On the other hand, in scenario $(b_G, b_Z, b_{GE}) = (0, 1, 0)$ under the presence of the association between genotype and covariate/environment variables, the assumption for the derivation of l_{approx} does not hold. Nevertheless, l_{approx} well approximated l_{mean} . Scenarios $(b_G, b_Z, b_{GE}) = (1, 0, 0)$ and $(0, 0, 1)$ correspond to the presence of the genetic effect, meaning that the approximation is not guaranteed. Actually, l_{approx} was deviated from l_{mean} in many cases. The magnitude of deviation differed depending on the setup of environment/covariate variables. The results of scenario 1d in [S2 and S3](#) Tables showed that the presence of the association between the covariate and environment variables did not make much difference from the baseline scenario.
- Misspecified null model (Scenarios 2a, 2b, 2c, 2d, and 2e). In some scenarios, l_{mean} took an extremely large value, and then, l_{approx} had a large value correspondingly. In scenarios 2a and 2b (i.e. covariate associated with genotypes was missed in the null model), no deviation between l_{approx} and l_{mean} was observed when no genotypic effect exist (i.e. $(b_G, b_Z, b_{GE}) = (0, 0, 0)$ and $(0, 1, 0)$), but l_{approx} was close to one while l_{mean} deviated from one in other cases (i.e. $(b_G, b_Z, b_{GE}) = (1, 0, 0)$ and $(0, 0, 1)$). The misspecified functional form of the null model (i.e. scenario 2c) gave larger l_{mean} , and l_{approx} could approximate the inflated l_{mean} . The existence of outlier(s) tended to give large l_{mean} just as in [Figs 3 and 4](#). l_{mean} in scenario 2d (i.e. one outlier) was larger than that in scenario 2e (i.e. ten outliers).

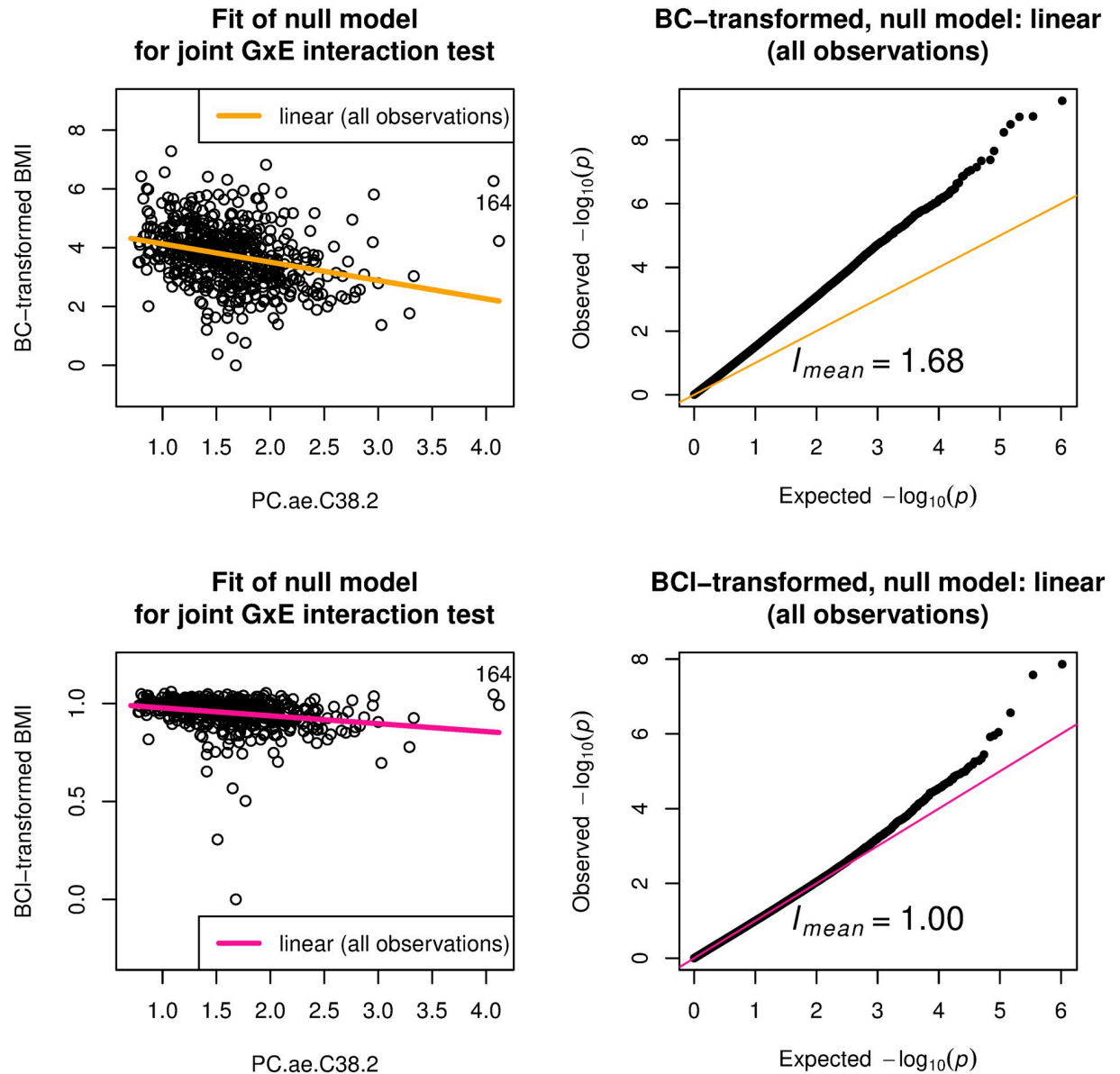


Fig 5. Gene-PC.ae.C38.2 (a metabolite) interaction test for Box-Cox transformed body mass index (BMI) on ADNI-GWAS data with 684 samples. Joint GxE interaction test applies to BMI as a quantitative phenotype and PC.ae.C38.2 as the environment variable with adjustment for sex and age. Top left: scatter plot for Box-Cox transformed BMI (BMI subtracted the minimum and then added 1) optimized in terms of normality (BC-transformed) and PC.ae.C38.2 with emphasis on observation 164, which was the influential observation having Cook's distance (based on BMI without transformation) larger than the mean plus 4×sd (standard deviation). Linear null model is given, $BC_{normality}(BMI - \min BMI + 1) = \gamma_1 + \gamma_2sex + \gamma_3age + \gamma_4PC.ae.C38.2$, where $BC_{normality}$ denotes the Box-Cox transformation with the optimal parameter determined by the closeness to normality of the transformed BMI. Top right: quantile-quantile (QQ) plot from joint GxE interaction test for BC-transformed BMI (BMI subtracted the minimum and then added 1) with the linear null model on all observations. Bottom left: scatter plot for Box-Cox transformed BMI (BMI subtracted the minimum and then added 1) optimized in terms of I_{approx} (BCI-transformed) and PC.ae.C38.2 with emphasis on observation 164, which was the influential observation having Cook's distance (based on BMI without transformation) larger than the mean plus 4×sd (standard deviation). Linear null model is given, $BC_{I_{approx}}(BMI - \min BMI + 1) = \gamma_1 + \gamma_2sex + \gamma_3age + \gamma_4PC.ae.C38.2$, where $BC_{I_{approx}}$ denotes the Box-Cox transformation with the optimal parameter determined by the closeness of I_{approx} to one. Bottom right: QQ plot from joint GxE interaction test for BCI-transformed BMI with the linear null model on all observations. I_{mean} , the scaled mean of genome-wide test statistics, is one if the null model is correctly specified.

<https://doi.org/10.1371/journal.pone.0219825.g005>

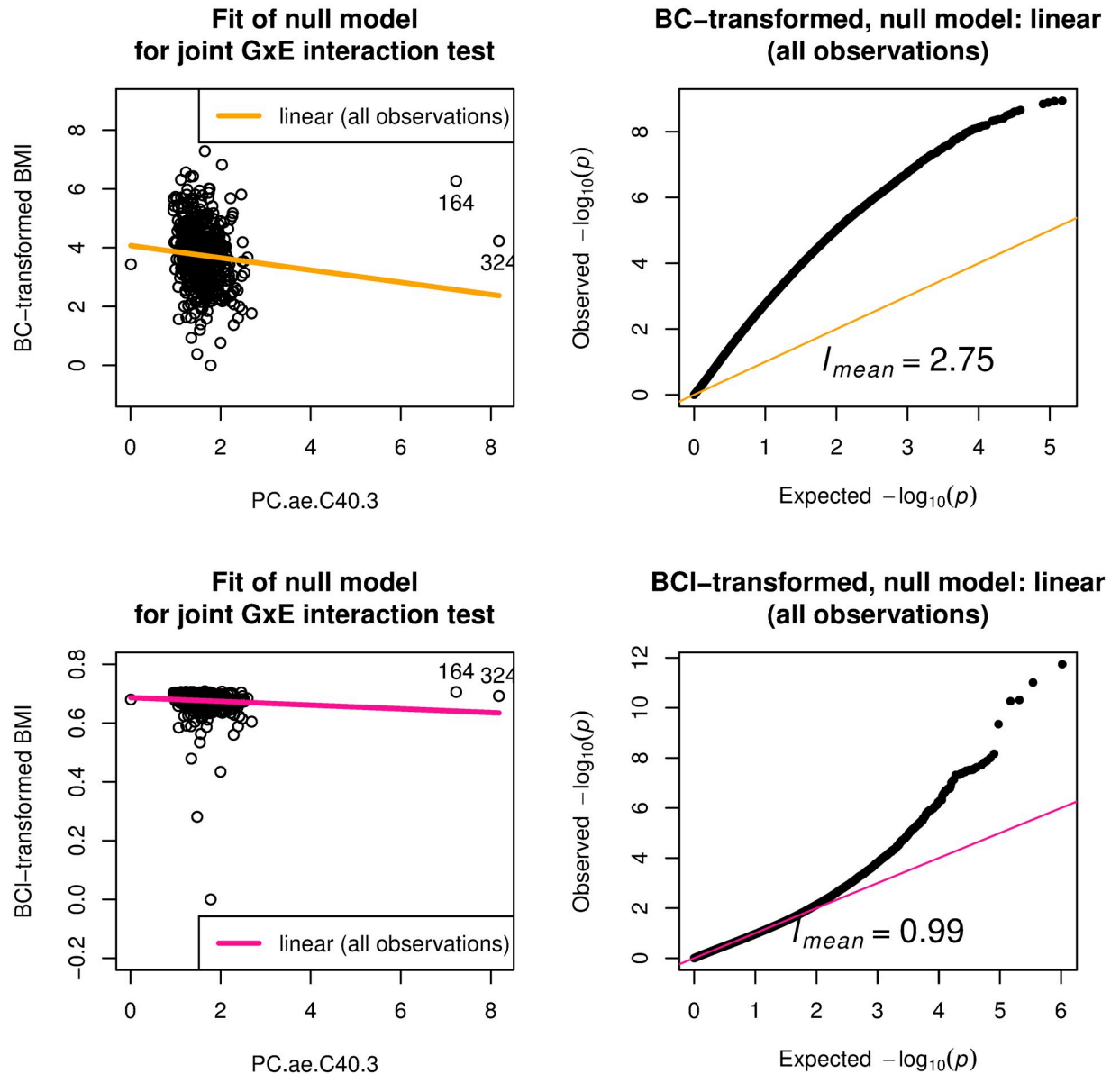


Fig 6. Gene-PC.ae.C40.3 (a metabolite) interaction test for Box-Cox transformed body mass index (BMI) on ADNI-GWAS data with 684 samples. Joint GxE interaction test applies to BMI as a quantitative phenotype and PC.ae.C40.3 as the environment variable with adjustment for sex and age. Top left: scatter plot for Box-Cox transformed BMI (BMI subtracted the minimum and then added 1) optimized in terms of normality (BC-transformed) and PC.ae.C40.3 with emphasis on observations 164 and 324, which were the influential observations having Cook's distance (based on BMI without transformation) larger than the mean plus 4×sd (standard deviation). Linear null model is given, $BC_{normality}(BMI - \min BMI + 1) = \gamma_1 + \gamma_2sex + \gamma_3age + \gamma_4PC.ae.C40.3$, where $BC_{normality}$ denotes the Box-Cox transformation with the optimal parameter determined by the closeness to normality of the transformed BMI. Top right: quantile-quantile (QQ) plot from joint GxE interaction test for BC-transformed BMI with the linear null model on all observations. Bottom left: scatter plot for Box-Cox transformed BMI (BMI subtracted the minimum and then added 1) optimized in terms of I_{approx} (BCI-transformed) and PC.ae.C40.3 with emphasis on observations 164 and 324, which were the influential observations having Cook's distance (based on BMI without transformation) larger than the mean plus 4×sd (standard deviation). Linear null model is given, $BC_{I_{approx}}(BMI - \min BMI + 1) = \gamma_1 + \gamma_2sex + \gamma_3age + \gamma_4PC.ae.C40.3$, where $BC_{I_{approx}}$ denotes the Box-Cox transformation with the optimal parameter determined by the closeness of I_{approx} to one. Bottom right: QQ plot from joint GxE interaction test for BCI-transformed BMI with the linear null model on all observations. I_{mean} , the scaled mean of genome-wide test statistics, is one if the null model is correctly specified.

<https://doi.org/10.1371/journal.pone.0219825.g006>

Table 2. Comparison between l_{approx} and l_{mean} in quantitative phenotype simulations.

Scenario (b_G, b_Z, b_{GE})		GxE ($n = 1000$)			Marginal ($n = 1000$)			GxE ($n = 10000$)			Marginal ($n = 10000$)		
		l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$
Base	(0,0,0)	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
Base	(1,0,0)	1.01	1.12	0.90	1.00	1.23	0.82	1.00	2.02	0.49	1.00	3.04	0.33
		(0.03)	(0.04)	(0.02)	(0.00)	(0.04)	(0.03)	(0.01)	(0.04)	(0.01)	(0.00)	(0.08)	(0.01)
Base	(0,1,0)	1.01	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.04)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
Base	(0,0,1)	1.05	1.06	0.99	1.00	1.01	0.99	1.31	2.41	0.54	1.00	1.00	1.00
		(0.04)	(0.05)	(0.02)	(0.00)	(0.03)	(0.03)	(0.02)	(0.07)	(0.01)	(0.00)	(0.03)	(0.03)
1a	(0,0,0)	1.01	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(1,0,0)	1.01	1.07	0.94	1.00	1.13	0.89	1.00	1.92	0.52	1.00	2.84	0.35
		(0.03)	(0.04)	(0.02)	(0.00)	(0.04)	(0.03)	(0.01)	(0.04)	(0.01)	(0.00)	(0.07)	(0.01)
1a	(0,1,0)	1.00	1.01	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(0,0,1)	1.29	1.39	0.93	1.00	1.01	0.99	1.32	2.47	0.53	1.00	1.00	1.00
		(0.08)	(0.09)	(0.02)	(0.00)	(0.03)	(0.03)	(0.02)	(0.07)	(0.01)	(0.00)	(0.03)	(0.03)
2a	(0,0,0)	1.01	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(1,0,0)	1.00	1.13	0.89	1.00	1.25	0.80	1.00	2.26	0.44	1.00	3.51	0.29
		(0.03)	(0.04)	(0.02)	(0.00)	(0.05)	(0.03)	(0.01)	(0.05)	(0.01)	(0.00)	(0.09)	(0.01)
2a	(0,1,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(0,0,1)	1.34	1.46	0.92	1.00	1.00	1.00	1.23	1.99	0.62	1.00	1.00	1.00
		(0.07)	(0.08)	(0.02)	(0.00)	(0.03)	(0.03)	(0.02)	(0.06)	(0.01)	(0.00)	(0.03)	(0.03)
2c	(0,0,0)	2.25	2.24	1.01	1.00	1.00	1.00	2.34	2.34	1.00	1.00	1.00	1.00
		(0.24)	(0.24)	(0.03)	(0.00)	(0.03)	(0.03)	(0.08)	(0.11)	(0.03)	(0.00)	(0.03)	(0.03)
2c	(1,0,0)	2.16	2.20	0.98	1.00	1.09	0.93	2.27	2.39	0.95	1.00	1.23	0.81
		(0.25)	(0.26)	(0.03)	(0.00)	(0.04)	(0.03)	(0.08)	(0.11)	(0.02)	(0.00)	(0.04)	(0.02)
2c	(0,1,0)	2.30	2.29	1.00	1.00	1.00	1.00	2.33	2.34	1.00	1.00	1.00	1.00
		(0.23)	(0.23)	(0.03)	(0.00)	(0.03)	(0.03)	(0.09)	(0.11)	(0.03)	(0.00)	(0.03)	(0.03)
2c	(0,0,1)	2.30	2.33	0.99	1.00	1.00	1.00	2.31	2.63	0.88	1.00	1.00	1.00
		(0.23)	(0.23)	(0.03)	(0.00)	(0.03)	(0.03)	(0.10)	(0.13)	(0.02)	(0.00)	(0.03)	(0.03)
2d	(0,0,0)	1.00	1.00	0.99	1.00	1.00	1.00	0.99	1.00	0.99	1.00	1.00	1.00
		(0.30)	(0.22)	(0.07)	(0.00)	(0.03)	(0.03)	(0.23)	(0.17)	(0.05)	(0.00)	(0.03)	(0.03)
2d	(1,0,0)	0.99	1.11	0.87	1.00	1.24	0.81	0.98	1.29	0.76	1.00	1.60	0.63
		(0.36)	(0.26)	(0.08)	(0.00)	(0.04)	(0.03)	(0.20)	(0.16)	(0.06)	(0.00)	(0.05)	(0.02)
2d	(0,1,0)	49.69	35.91	1.38	1.00	1.00	1.00	190.26	139.98	1.36	1.00	1.00	1.00
		(6.77)	(4.93)	(0.02)	(0.00)	(0.03)	(0.03)	(10.68)	(8.48)	(0.03)	(0.00)	(0.03)	(0.03)
2d	(0,0,1)	13.70	10.09	1.36	1.00	1.01	1.00	24.05	18.15	1.32	1.00	1.00	1.00
		(3.75)	(2.68)	(0.03)	(0.00)	(0.03)	(0.03)	(4.29)	(3.10)	(0.03)	(0.00)	(0.03)	(0.03)

Simulation results from 200 replicates under scenarios Base, 1a, 2a, 2c, and 2d with four effect size scenarios given as triplets (b_G, b_Z, b_{GE}): (0, 0, 0), (1, 0, 0), (0, 1, 0), and (0, 0, 1) with additive genotype coding. The values are the means and standard errors (in parentheses) of the proposed approximation (l_{approx}), the scaled mean test statistics (l_{mean}) and the ratio (l_{approx}/l_{mean}) for joint GxE interaction and marginal association tests. $b_G, b_Z,$ and b_{GE} are parameters of genotype, covariate and GxE interaction effects, respectively (0 corresponds to no effect); n denotes the sample size.

<https://doi.org/10.1371/journal.pone.0219825.t002>

Table 3. Comparison between l_{approx} and l_{mean} in binary phenotype simulations.

Scenario (b_G, b_Z, b_{GE})		GxE ($n = 1000$)			Marginal ($n = 1000$)			GxE ($n = 10000$)			Marginal ($n = 10000$)		
		l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$	l_{approx}	l_{mean}	$\frac{l_{approx}}{l_{mean}}$
Base	(0,0,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
Base	(1,0,0)	1.00	1.02	0.99	1.00	1.03	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
Base	(0,1,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
Base	(0,0,1)	1.03	1.05	0.98	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(0,0,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(1,0,0)	1.00	1.01	0.99	1.00	1.02	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(0,1,0)	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.04)	(0.04)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
1a	(0,0,1)	1.03	1.05	0.98	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(0,0,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(1,0,0)	1.00	1.02	0.98	1.00	1.04	0.97	1.00	1.00	1.00	1.00	1.00	1.00
		(0.01)	(0.03)	(0.02)	(0.00)	(0.04)	(0.04)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(0,1,0)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2a	(0,0,1)	1.03	1.05	0.98	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.00
		(0.03)	(0.04)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.04)
2c	(0,0,0)	0.86	0.86	1.00	1.00	1.00	1.00	0.85	0.85	1.00	1.00	1.00	1.00
		(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2c	(1,0,0)	0.92	0.94	0.98	1.00	1.03	0.97	0.88	0.89	1.00	1.00	1.00	1.00
		(0.02)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)	(0.00)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2c	(0,1,0)	0.89	0.90	1.00	1.01	1.00	1.00	0.89	0.89	1.00	1.00	1.00	1.00
		(0.02)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2c	(0,0,1)	0.86	0.86	1.00	1.00	1.01	1.00	0.95	0.95	1.00	1.01	1.03	0.98
		(0.01)	(0.03)	(0.02)	(0.00)	(0.03)	(0.03)	(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.03)
2d	(0,0,0)	0.99	1.00	0.99	1.00	1.01	1.00	1.02	1.01	1.00	1.00	1.00	1.01
		(0.19)	(0.14)	(0.05)	(0.00)	(0.03)	(0.03)	(0.16)	(0.12)	(0.04)	(0.00)	(0.03)	(0.03)
2d	(1,0,0)	1.03	1.03	0.99	1.00	1.03	0.97	1.03	1.02	1.01	1.00	1.00	1.00
		(0.26)	(0.18)	(0.06)	(0.00)	(0.03)	(0.03)	(0.17)	(0.13)	(0.04)	(0.00)	(0.03)	(0.03)
2d	(0,1,0)	9.15	5.86	1.30	1.00	1.00	1.00	2.15	2.13	1.01	1.00	1.00	1.00
		(8.73)	(4.97)	(0.34)	(0.01)	(0.03)	(0.03)	(1.17)	(1.15)	(0.03)	(0.00)	(0.03)	(0.03)
2d	(0,0,1)	11.34	6.65	1.41	0.99	0.99	1.00	2.11	2.08	1.01	0.99	0.99	1.00
		(9.50)	(5.09)	(0.41)	(0.01)	(0.03)	(0.03)	(1.18)	(1.15)	(0.03)	(0.01)	(0.03)	(0.03)

Simulation results from 200 replicates under scenarios Base, 1a, 2a, 2c, and 2d with four effect size scenarios given as triplets (b_G, b_Z, b_{GE}): (0, 0, 0), (1, 0, 0), (0, 1, 0), and (0, 0, 1) with additive genotype coding. The values are the means and standard errors (in parentheses) of the proposed approximation (l_{approx}), the scaled mean test statistics (l_{mean}) and the ratio (l_{approx}/l_{mean}) for joint GxE interaction and marginal association tests. $b_G, b_Z,$ and b_{GE} are parameters of genotype, covariate and GxE interaction effects, respectively (0 corresponds to no effect); n denotes the sample size.

<https://doi.org/10.1371/journal.pone.0219825.t003>

- Environment/covariate variable distribution (Scenarios 3a, 3b, 3c, and 3d). Under the absence of the genetic effect (i.e. $(b_G, b_Z, b_{GE}) = (0, 0, 0)$ and $(0, 1, 0)$), the approximation of l_{approx} to l_{mean} looked well. Under the presence of the genetic effect (i.e. $(b_G, b_Z, b_{GE}) = (1, 0, 0)$ and $(0, 0, 1)$), where the approximation is not guaranteed, l_{approx} was deviated from l_{mean} as expected. The magnitude of deviation differed depending on the setup of environment/covariate variables.
- Genotype distribution (Scenarios 4a, 4b, and 4c). The difference in MAF distribution gave no much difference in the approximation as our derivation does not require specific MAF distribution. Correlation between genotypes did not alter the approximation in terms of mean values of l_{approx} and l_{mean} , but the result under the presence of correlation was more variable than the result under the absence of correlation. This is perhaps due to that the correlation between genetic variants reduced the effective number of independent loci.

For binary phenotype simulations (i.e. [Table 3](#) and [S3 Table](#)), l_{mean} s were less deviated from one compared with the results on quantitative phenotypes (i.e. [Table 2](#) and [S2 Table](#)), probably due to the fact that binary phenotype has lower variation than quantitative phenotype, but the overall tendency was similar to the quantitative simulations. Regarding genotype coding, the additive coding gave larger discrepancy of l_{approx} from l_{mean} than recessive and dominant codings, and the recessive coding resulted in smallest discrepancy, which is due to the assignment of effect size 1 regardless of genotype coding.

Discussion

In this paper, we presented a novel closed-form approximation to the mean of the chi-squared statistics for genome-wide joint GxE interaction tests by assuming that the null model is well specified. Interestingly, characteristics of genetic loci no longer appear in the approximation formula. It allows quick assessment of systematic inflation/deflation due to null model misspecification without requiring a genome-wide scan. To the best of our knowledge, there are no approach comparable to our proposed method. The approximation formula is particularly useful when many null models for GxE interaction analyses must be handled, e.g. with hundreds of environment variables such as the metabolites considered in this paper. For example, our examples in [Figs 3](#) and [4](#) showing systematic inflation were identified by applying the proposed statistic; it would have been laborious work if all null models with 117 metabolites for large number of phenotypes were exhaustively investigated by genome-wide scan as well as by detailed examination of null models, because the procedure involves various visual inspections.

Once problematic null models have been detected, detailed investigation of adequacy of the null model must be made, for example by standard regression diagnostics, e.g. modeling of covariate effects, presence of outliers and influential samples, or heteroscedasticity. We also showed that existing methods—the robust test and genomic control-adjustment—may not always work. Note that the robust test differs from regression methods in robust statistics [[33](#), [34](#)] in which the focus is on issues owing to outliers. The null model misspecification may arise not only by outliers but also by misspecification of functional form or statistical model. The robust test only accounts for the impact of misspecification on the variance, whereas no correction is made for bias, which in general depends on the true model and thus is difficult to correct without additional modeling assumption. If bias is not negligible, the robust test may fail to resolve the problematic behavior. Rendering the null models more sophisticated, such as by better specifying the environment variable effect or by removing outliers, could reduce

systematic inflation or deflation. Manual removal of outliers is not straightforward, in particular, when a large number of covariates are used. In this case, robust linear [33, 34] or logistic regression [35] are attractive approaches. If the misspecification of the functional form of the environment variable is the main concern, it is worth to apply the method recently developed by [36].

Theoretically, the test statistic for the marginal association test is inflated from χ_1^2 to $\lambda\chi_1^2$ under the presence of population stratification [17, 24, 37, 38]. In this situation, the test statistics divided by the inflation factor λ follow χ_1^2 distribution, which is the key idea of the genomic control method [17]. In contrast, in the case of problematic behavior for the joint GxE interaction tests due to null model misspecification, it is unclear what the distribution of test statistics is. The examples above demonstrate that genomic control-adjusted p-values do not always work. The above additional analyses also suggest that a more sophisticated null model or suitable statistical method could resolve the problematic behavior.

We also showed that the standard Box–Cox transformation based on normality may not always resolve the problematic behavior, whereas the Box–Cox transformation based on l_{approx} can directly resolve systematic inflation. Although we recommend detailed investigation of problematic null models, the Box–Cox transformation based on l_{approx} can provide a quick solution.

The proposed approximation is derived by assuming that the genotype data at each locus are independently and identically distributed across subjects (but the distribution across loci can differ and be correlated due to linkage disequilibrium). This simplification enabled us to derive a closed-form formula. On the one hand, our theory does not cover the presence of population stratification, which is the limitation of our approach. Indeed, we observed situations where the approximation failed in the presence of population stratification (Fig 2). Interestingly, the failure disappeared by adjusting for PCs estimated from genome-wide SNP data. The extensibility of the phenomenon to other datasets is an interesting future topic.

The statistic l_{approx} quantifies problematic behavior based on the extent of the discrepancy from one, as with the genomic inflation factor. For example, we can roughly judge that null models having l_{approx} within the range [0.9, 1.1] or [0.95, 1.05] are not a problem, so that a genome-wide scan can be safely conducted.

In derivation of the approximation formula, we imposed the technical condition of no genetic effect on phenotype, covariates, and environment variables. This assumption may not necessarily hold in real GWAS data. However, as seen in the real ADNI data application, the approximation worked well in most of the cases. The simulation studies also showed that the approximation worked under the scenarios with no genetic effect on phenotype (i.e. $(b_G, b_Z, b_{GE}) = (0, 0, 0)$ and $(0, 1, 0)$). In real GWAS, we sometimes encounter the situation where the marginal association test gives a very limited number of loci being genome-wide significant while other loci follow the expected null distribution (as checked by the quantile–quantile plot of genome-wide p-values), implying that the majority of loci have approximately no genetic effect on phenotype. In such cases, we expect that the approximation works well.

In contrast, if many loci have effect on phenotype, covariates, and environment variables, the above assumption does not hold, e.g. under the polygenic architecture [39] or in the candidate gene studies. As seen in the simulation studies as well as in the ADNI data application with 757 samples (c.f. Fig 2), the presence of genetic effect on phenotype yielded discrepancy between l_{approx} and l_{mean} . In the cases with l_{approx} far from one, l_{mean} was also far from one, meaning that l_{approx} far from one indicates systematic inflation or deflation. In contrast, l_{approx} close to one does not guarantee the absence of systematic inflation or deflation. Simulation scenarios 2a and 2b correspond to the above phenomenon. Specifically, ignorance of covariates

associated with genotypes and phenotype made a deviation of l_{mean} from one but l_{approx} was still close to one. In this case, l_{approx} was unable to detect the systematic inflation, which means that the proposed approximation fails if confounding factors (or the surrogate variables) are unavailable and the genetic variants caused the problem, e.g. due to population stratification as in the ADNI data with 757 samples (c.f. Fig 2).

One might consider that traditional goodness-of-fit tests may be used just like l_{approx} . However, this would reject most of the null models when the sample size is very large because the test requires correct specification of the true model, which is rarely achieved with real data [40]. On the other hand, the criterion based on l_{approx} allows one to explore null models that give acceptably less systematic departure of the test statistics distribution. The requirement is less strict than the correct model specification needed for goodness-of-fit tests. Our approximation is for the score statistics. For the Wald and likelihood ratio tests, l_{approx} is still useful, at least for the purpose of identifying null model misspecification, because the Wald and likelihood ratio tests are asymptotically the same as the score test.

We conclude that our proposed approximation is useful to quickly assess systematic inflation/deflation due to null model misspecification without requiring a genome-wide scan. It helps researchers to reconsider and improve null model specification. The benefit should be great when many covariates and environment variables are considered.

Supporting information

S1 Fig. 1st and 2nd PCs (principal components) plot for 757 individuals in ADNI data.

PCs were computed by the EIGENSOFT package using GWAS data with 757 ADNI samples. The first and second PCs for the 757 samples were provided. Ethnic group label for each individual is taken from PTRACCAT object in ADNIMERGE package: American Indian or Alaskan Native (Am Indian/Alaskan), Asian (Asian), Hawaiian/Other PI (Hawaiian or Other Pacific Islander), Black or African American (Black), White (White), More than One Reported (More than one), Unknown or Not Reported (Unknown).
(EPS)

S2 Fig. Scatter plots of the phenotype–environment pairs showing $l_{approx} > 1.5$ from five phenotypes and 117 metabolite variables. $l_{approx} > 1.5$ was computed for the ADNI 684 non-Hispanic Caucasian samples. Phenotypes: HEIGHT (height, cm), BMI (body mass index, kg/m^2), WholeBrain (whole-brain volume, cm^3), and MMSE (mini mental state examination, score); Metabolites as environment variables: C10.2, C10.2, C4, C5, PC.aa.C42.6, PC.ae.C38.2, PC.ae.C40.3, PC.ae.C44.3, C10.2, C5.DC..C6.OH., C5.DC..C6.OH.
(EPS)

S3 Fig. Comparison between l_{approx} and l_{mean} for joint GxE interaction and marginal association tests on ADNI-GWAS data for 684 non-Hispanic Caucasian samples with simulated genotype data. Joint GxE interaction and marginal association tests are carried out on ADNI-GWAS data for 684 non-Hispanic Caucasian samples where only the real genotype data was replaced by simulation with 10000 loci independently generated under linkage equilibrium, where MAFs are set by uniform distribution on [0.05, 0.5] (ADNI684sim) Phenotypes, covariates and environment variables in real data are fixed. Five phenotypes—height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE)—and nine environment variables: lysoPC.a.C16.0, PC.ae.C38.2, PC.ae.C40.3, C10, PC.aa.C40.5, PC.ae.C36.3, SM..OH..C14.1, SM..OH..C22.1, and SM.C24.0. Top left: joint GxE interaction test (quantitative phenotype). Top right: marginal association test (quantitative phenotype). Middle left: joint GxE interaction test (quantitative

phenotype after Box–Cox transformation optimized in terms of normality). Middle right: joint GxE interaction test (quantitative phenotype after Box–Cox transformation optimized in terms of l_{approx}). Bottom left: joint GxE interaction test (binary phenotype created by dichotomizing quantitative phenotype). Bottom right: marginal association test (binary phenotype created by dichotomizing quantitative phenotype). (EPS)

S4 Fig. Comparison between l_{approx} and l_{mean} for joint GxE interaction and marginal association tests on ADNI-GWAS data for 757 samples showing population stratification with adjustment for top 3 and 5 principal components. Comparison between l_{approx} and l_{mean} , for which population stratification exists, for joint GxE interaction and marginal association tests on ADNI-GWAS data for 757 samples showing population stratification (ADNI757). Five phenotypes—height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE)—and nine metabolite variables: lysoPC.a.C16.0, PC.ae.C38.2, PC.ae.C40.3, C10, PC.aa.C40.5, PC.ae.C36.3, SM.OH.C14.1, SM.OH.C22.1, and SM.C24.0. Results are shown without and with adjustment for top 10 principal components (PCs). Top left: joint GxE interaction test with adjustment for top 3 PCs (quantitative phenotype). Top right: marginal association test with adjustment for top 3 PCs (quantitative phenotype). Left in the second row: joint GxE interaction test with adjustment for top 5 PCs (quantitative phenotype). Right in the second row: marginal association test with adjustment for top 5 PCs (quantitative phenotype). Left in the third row: joint GxE interaction test with adjustment for top 3 PCs (binary phenotype created by dichotomizing quantitative phenotype). Right in the second row: marginal association test with adjustment for top 3 PCs (binary phenotype created by dichotomizing quantitative phenotype). Bottom left: joint GxE interaction test with adjustment for top 5 PCs (binary phenotype created by dichotomizing quantitative phenotype). Bottom right: marginal association test with adjustment for top 5 PCs (binary phenotype created by dichotomizing quantitative phenotype). (EPS)

S5 Fig. Comparison between l_{approx} and l_{mean} for joint GxE interaction and marginal association tests on ADNI-GWAS data for 757 samples with simulated genotype data. Joint GxE interaction and marginal association tests are carried out on ADNI-GWAS data for 757 samples where only the real genotype data, which showed population stratification, was replaced by simulation with 10000 loci independently generated under linkage equilibrium, where MAFs are set by uniform distribution on [0.05, 0.5] (ADNI757sim). Phenotypes, covariates, and environment variables in real data are fixed. Five phenotypes—height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE)—and nine metabolite variables: lysoPC.a.C16.0, PC.ae.C38.2, PC.ae.C40.3, C10, PC.aa.C40.5, PC.ae.C36.3, SM.OH.C14.1, SM.OH.C22.1, and SM.C24.0. Results are shown with adjustment for top 3 and 5 principal components (PCs). Top left: joint GxE interaction test without PC adjustment (quantitative phenotype). Top right: joint GxE interaction test with adjustment for top 10 PCs (quantitative phenotype). Left in the second row: marginal association test without PC adjustment (quantitative phenotype). Right in the second row: marginal association test with adjustment for top 10 PCs (quantitative phenotype). Left in the third row: joint GxE interaction test without PC adjustment (binary phenotype created by dichotomizing quantitative phenotype). Right in the third row: joint GxE interaction test with adjustment for top 10 PCs (binary phenotype created by dichotomizing quantitative phenotype). Bottom left: marginal association test without PC adjustment (binary phenotype created by dichotomizing quantitative phenotype). Bottom right: marginal association test with adjustment for top 10

PCs (binary phenotype created by dichotomizing quantitative phenotype).
(EPS)

S6 Fig. Regression diagnostics plots from linear model fit of BMI on PC.ae.C38.2 in the ADNI data with 684 samples generated by “plot” for “lm” result.
(EPS)

S7 Fig. Regression diagnostics plots from quadratic model fit of BMI on PC.ae.C40.3 in the ADNI data with 684 samples generated by “plot” for “lm” result.
(EPS)

S1 Table. l_{approx} on ADNI-GWAS data for 684 non-Hispanic Caucasian samples. l_{approx} computed for joint GxE interaction on ADNI-GWAS data for 684 non-Hispanic Caucasian samples, where sex and age are adjusted for, with respect to five quantitative phenotypes, height (HEIGHT), body mass index (BMI), whole brain (WholeBrain), intracranial volume (ICV), and mini mental state exam (MMSE), and 117 metabolite variables as environment variable.
(CSV)

S2 Table. Comparison between l_{approx} and l_{mean} in quantitative phenotype simulations. Additional quantitative phenotype simulation results.
(PDF)

S3 Table. Comparison between l_{approx} and l_{mean} in binary phenotype simulations. Additional binary phenotype simulation results.
(PDF)

S1 Appendix. Technical details. Details of theoretical results and simulation studies.
(PDF)

S2 Appendix. Program code for simulation studies. R code for simulation studies, including a function `lapprox` to compute l_{approx} using phenotype, environment and covariate variables as input.
(R)

Acknowledgments

Data collection and sharing for the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) led by Principal Investigator Michael W. Weiner, MD (Michael.Weiner@ucsf.edu), and generated by the Alzheimer’s Disease Metabolomics Consortium (ADMC), lead by Dr. Kaddurah-Daouk (kaddu001@mc.duke.edu). Data used in preparation of this article were obtained from the ADNI and the ADMC (adni.loni.usc.edu). As such, the investigators within the ADNI and the ADMC contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf, and a complete listing of ADMC investigators can be found at: <https://sites.duke.edu/adnimetab/team/>.

The authors thank Prof. Heather Cordell, Prof. Takashi Yanagawa, Prof. Tatsuyuki Kakuma, Prof. Satoshi Hattori, Dr. John Cologne, Prof. Saurabh Ghosh, an anonymous reviewer, and the associate editor for their insightful comments, and also sincerely thank Dr. Miriam Kesselmeier for careful review and for providing many incisive comments and concrete suggestions that led to significant improvement of the paper.

Author Contributions

Conceptualization: Masao Ueki.

Data curation: Masao Ueki.

Formal analysis: Masao Ueki.

Funding acquisition: Masao Ueki, Gen Tamiya.

Investigation: Masao Ueki, Masahiro Fujii, Gen Tamiya.

Methodology: Masao Ueki.

Project administration: Masao Ueki.

Resources: Masao Ueki, Gen Tamiya.

Software: Masao Ueki.

Supervision: Masao Ueki.

Validation: Masao Ueki.

Visualization: Masao Ueki, Masahiro Fujii.

Writing – original draft: Masao Ueki.

Writing – review & editing: Masao Ueki.

References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB. Finding the missing heritability of complex diseases. *Nature*. 2009; 461(7265):747–753. <https://doi.org/10.1038/nature08494> PMID: 19812666
2. Thomas D. Gene–environment-wide association studies: emerging approaches. *Nature Reviews Genetics*. 2010; 11(4):259–272. <https://doi.org/10.1038/nrg2764> PMID: 20212493
3. Ober C, Vercelli D. Gene–environment interactions in human disease: nuisance or opportunity? *Trends in Genetics*. 2011; 27(3):107–115. <https://doi.org/10.1016/j.tig.2010.12.004> PMID: 21216485
4. Kraft P, Aschard H. Finding the missing gene–environment interactions. *European Journal of Epidemiology*. 2015; 30(5):353–355. <https://doi.org/10.1007/s10654-015-0046-1> PMID: 26026724
5. Ritz BR, Chatterjee N, Garcia-Closas M, Gauderman WJ, Pierce BL, Kraft P, et al. Lessons learned from past gene-environment interaction successes. *American Journal of Epidemiology*. 2017; 186(7):778–786. <https://doi.org/10.1093/aje/kwx230> PMID: 28978190
6. Khoury MJ. Editorial: emergence of gene-environment interaction analysis in epidemiologic research. *American Journal of Epidemiology*. 2017; 186(7):751–752. <https://doi.org/10.1093/aje/kwx226> PMID: 28978194
7. Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Update on the state of the science for analytical methods for gene-environment interactions. *American Journal of Epidemiology*. 2017; 186(7):762–770. <https://doi.org/10.1093/aje/kwx228> PMID: 28978192
8. Ritchie MD, Davis JR, Aschard H, Battle A, Conti D, Du M, et al. Incorporation of biological knowledge into the study of gene-environment interactions. *American Journal of Epidemiology*. 2017; 186(7):771–777. <https://doi.org/10.1093/aje/kwx229> PMID: 28978191
9. McAllister K, Mechanic LE, Amos C, Aschard H, Blair IA, Chatterjee N, et al. Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *American Journal of Epidemiology*. 2017; 186(7):753–761. <https://doi.org/10.1093/aje/kwx227> PMID: 28978193
10. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Human Heredity*. 2007; 63(2):111–119. <https://doi.org/10.1159/000099183> PMID: 17283440
11. Hamza TH, Chen H, Hill-Burns EM, Rhodes SL, Montimurro J, Kay DM, et al. Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS Genetics*. 2011; 7(8):e1002237. <https://doi.org/10.1371/journal.pgen.1002237> PMID: 21876681

12. Sung YJ, de las Fuentes L, Schwander KL, Simino J, Rao DC. Gene–smoking interactions identify several novel blood pressure loci in the Framingham heart study. *American Journal of Hypertension*. 2014; 28(3):343–354. <https://doi.org/10.1093/ajh/hpu149> PMID: 25189868
13. Sung YJ, Winkler TW, Manning AK, Aschard H, Gudnason V, Harris TB, et al. An empirical comparison of joint and stratified frameworks for studying G x E interactions: systolic blood pressure and smoking in the CHARGE gene-lifestyle interactions working group. *Genetic Epidemiology*. 2016; 40(5):404–415. <https://doi.org/10.1002/gepi.21978> PMID: 27230302
14. Keaton JM, Gao C, Guan M, Hellwege JN, Palmer ND, Pankow JS, et al. Genome-wide interaction with the insulin secretion locus MTNR1B reveals CMIP as a novel type 2 diabetes susceptibility gene in African Americans. *Genetic Epidemiology*. 2018; 42(6):559–570. <https://doi.org/10.1002/gepi.22126> PMID: 29691896
15. 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. <https://doi.org/10.1371/journal.pone.0019416> PMID: 21589913
16. Almli LM, Duncan R, Feng H, Ghosh D, Binder EB, Bradley B, et al. Correcting systematic inflation in genetic association tests that consider interaction effects. *JAMA Psychiatry*. 2014; 71(12):1392–1399. <https://doi.org/10.1001/jamapsychiatry.2014.1339> PMID: 25354142
17. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55(4):997–1004. <https://doi.org/10.1111/j.0006-341X.1999.00997.x> PMID: 11315092
18. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genetic Epidemiology*. 2001; 20(1):4–16. [https://doi.org/10.1002/1098-2272\(200101\)20:1<4::AID-GEPI2>3.0.CO;2-T](https://doi.org/10.1002/1098-2272(200101)20:1<4::AID-GEPI2>3.0.CO;2-T) PMID: 11119293
19. Devlin B, Bacanu SA, Roeder K. Genomic control to the extreme. *Nature Genetics*. 2009; 36(11):1129–1130. <https://doi.org/10.1038/ng1104-1129>
20. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*. 2015; 12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779> PMID: 25826379
21. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*. 2012; 44(8):955–959. <https://doi.org/10.1038/ng.2354> PMID: 22820512
22. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*. 2010; 11(7):499–511. <https://doi.org/10.1038/nrg2796> PMID: 20517342
23. Ueki M. Enhancing power of score tests for regression models via Fisher transformation. *Journal of the Japanese Society of Computational Statistics*. 2018; 30(2):37–53. https://doi.org/10.5183/jjscs.1702001_234
24. Dadd T, Weale ME, Lewis SA, Bacanu CM, Devlin B, Roeder K. A critical evaluation of genomic control methods for genetic association studies. *Genetic Epidemiology*. 2009; 33(4):290–298. <https://doi.org/10.1002/gepi.20379> PMID: 19051284
25. Shen L, Thompson P, Potkin S, Bertram L, Farrer L, Foroud T, et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging and Behavior*. 2014; 8(2):183–207. <https://doi.org/10.1007/s11682-013-9262-z> PMID: 24092460
26. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 2006; 38(8):904–9. <https://doi.org/10.1038/ng1847> PMID: 16862161
27. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genetics*. 2006; 2(12):e190. <https://doi.org/10.1371/journal.pgen.0020190> PMID: 17194218
28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 2007; 81(3):559–575. <https://doi.org/10.1086/519795> PMID: 17701901
29. Marees A, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome wide association studies: quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*. 2018; 27(2):e1608. <https://doi.org/10.1002/mpr.1608> PMID: 29484742
30. Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*. 1964; 26(2):211–252.
31. Tchetgen Tchetgen EJ, Kraft P. On the robustness of tests of genetic associations incorporating gene-environment interactions when the environmental exposure is misspecified. *Epidemiology*. 2011; 22(2):257–261. <https://doi.org/10.1097/EDE.0b013e31820877c5> PMID: 21228699
32. Cook RD. Detection of influential observations in linear regression. *Technometrics*. 1977; 19(1):15–18. <https://doi.org/10.2307/1268249>

33. Huber PJ. Robust Statistics. John Wiley & Sons: New York; 1981.
34. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons: New York; 1986.
35. Kesselmeier M, Lorenzo Bermejo J. Robust logistic regression to narrow down the winner's curse for rare and recessive susceptibility variants. *Briefings in Bioinformatics*. 2017; 18(6):962–972. <https://doi.org/10.1093/bib/bbw074> PMID: 27543791
36. Sun R, Carroll RJ, Christiani DC, Lin X. Testing for gene-environment interaction under exposure mis-specification. *Biometrics*. 2018; 74(2):653–662. <https://doi.org/10.1111/biom.12813> PMID: 29120492
37. Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies. *Theoretical Population Biology*. 2001; 60(3):155–166. <https://doi.org/10.1006/tpbi.2001.1542> PMID: 11855950
38. Bacanu SA, Devlin B, Roeder K. Association studies for quantitative traits in structured populations. *Genetic Epidemiology*. 2002; 22(1):78–93. <https://doi.org/10.1002/gepi.1045> PMID: 11754475
39. Dudbridge F. Polygenic epidemiology. *Genetic Epidemiology*. 2016; 40(4):268–272. <https://doi.org/10.1002/gepi.21966> PMID: 27061411
40. Box GEP. Science and statistics. *Journal of the American Statistical Association*. 1976; 71(356):791–799. <https://doi.org/10.1080/01621459.1976.10480949>