PLOS GENETICS

METHODS

A robust and adaptive framework for interaction testing in quantitative traits between multiple genetic loci and exposure variables

Julian Hecker[1,2]*, **Dmitry Prokopenko**[2,3], **Matthew Moll**[1,2,4], **Sanghun Lee**[5,6], **Wonji Kim**[1,2], **Dandi Qiao**[1,2], **Kirsten Voorhies**[2,7], **Woori Kim**[2,8,9], **Stijn Vansteelandt**[10,11], **Brian D. Hobbs**[1,2,4], **Michael H. Cho**[1,2,4], **Edwin K. Silverman**[1,2,4], **Sharon M. Lutz**[2,6,7], **Dawn L. DeMeo**[1,2], **Scott T. Weiss**[1,2], **Christoph Lange**[1,2,6]

1 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **2** Harvard Medical School, Boston, Massachusetts, United States of America, **3** Genetics and Aging Unit and McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **4** Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **5** Department of Medical Consilience, Division of Medicine, Graduate School, Dankook University, Yongin, South Korea, **6** Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, **7** Department of Population Medicine, PRecisiOn Medicine Translational Research (PROMoTeR) Center, Harvard Pilgrim Health Care, Boston, Massachusetts, United States of America, **8** Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **9** Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **10** Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium, **11** Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

* rejhe@channing.harvard.edu

Abstract

The identification and understanding of gene-environment interactions can provide insights into the pathways and mechanisms underlying complex diseases. However, testing for gene-environment interaction remains a challenge since a.) statistical power is often limited and b.) modeling of environmental effects is nontrivial and such model misspecifications can lead to false positive interaction findings. To address the lack of statistical power, recent methods aim to identify interactions on an aggregated level using, for example, polygenic risk scores. While this strategy can increase the power to detect interactions, identifying contributing genes and pathways is difficult based on these relatively global results. Here, we propose RITSS (Robust Interaction Testing using Sample Splitting), a gene-environment interaction testing framework for quantitative traits that is based on sample splitting and robust test statistics. RITSS can incorporate sets of genetic variants and/or multiple environmental factors. Based on the user's choice of statistical/machine learning approaches, a screening step selects and combines potential interactions into scores with improved interpretability. In the testing step, the application of robust statistics minimizes the susceptibility to main effect misspecifications. Using extensive simulation studies, we demonstrate

OPEN ACCESS

Citation: Hecker J, Prokopenko D, Moll M, Lee S, Kim W, Qiao D, et al. (2022) A robust and adaptive framework for interaction testing in quantitative traits between multiple genetic loci and exposure variables. PLoS Genet 18(11): e1010464. https://doi.org/10.1371/journal.pgen.1010464

Editor: Heather J. Cordell, Newcastle University, UNITED KINGDOM

Received: January 12, 2022

Accepted: October 4, 2022

Published: November 16, 2022

Data Availability Statement: UK Biobank data can be accessed following successful application at https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. The R package RITSS as well as the code for the simulation studies is available at https://github.com/julianhecker/RITSS.

Funding: This research was supported by the National Heart, Lung, and Blood Institute [U01HL089856, U01HL089897, P01HL120839, P01HL132825] and National Human Genome Research Institute [R01HG008976, 2U01HG008685]. MHC was supported by R01HL149861, R01HL135142, R01HL137927, and R01HL147148. MM was supported by T32HL007427 and K08HL159318. DQ was supported by K01HL129039. SML was supported by R01MH129337. BDH was supported by K08HL136928. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: MHC has received grant funding from GSK and Bayer, and speaking or consulting fees from AstraZeneca, Illumina, and Genentech. EKS has received grant support from GSK and Bayer. WK is a current employee of Eisai Inc. MM received grant support from Bayer. STW reports personal fees from UpToDate, outside the submitted work; STW is on the Scientific Board of Histolix.

that RITSS controls the type 1 error rate in a wide range of scenarios, and we show how the screening strategy influences statistical power. In an application to lung function phenotypes and human height in the UK Biobank, RITSS identified highly significant interactions based on subcomponents of genetic risk scores. While the contributing single variant interaction signals are weak, our results indicate interaction patterns that result in strong aggregated effects, providing potential insights into underlying gene-environment interaction mechanisms.

Author summary

The understanding of gene-environment interactions provides potential insights into the pathways and mechanisms underlying complex diseases, but they are hard to detect since effect sizes are expected to be small. To facilitate the detection of such interactions in quantitative traits, we propose a robust and flexible approach called RITSS that can incorporate sets of genetic variants and multiple environmental factors. RITSS can utilize any suitable machine/statistical learning approach to screen for interactions and rigorously tests these aggregated signals using sample splitting and robust test statistics. We demonstrate the validity and power of our approach in extensive simulation studies. Furthermore, in an application to lung function and height data in the UK Biobank, RITSS discovers highly significant interactions based on subcomponents of genetic risk scores.

Introduction

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex diseases and traits [1]. However, the effect of a genetic variant on a complex phenotype or disease can be modified by environmental exposures [2,3]. The knowledge about interactions between environmental exposures (e.g., tobacco smoke or occupational exposure) and genetic variants can provide insights into the underlying pathways and biological disease mechanisms [4].

Many of the methodological approaches for gene-environment interaction testing have focused on scenarios in which a single genetic variant and a single environmental variable are tested for potential interaction. Methods have been developed for case-control data, case-only data, and quantitative trait studies, summarized in detail by Gauderman et al. [5].

Since most gene-environment interaction effect sizes are expected to be small and statistical power is consequently limited [6], power-increasing approaches were proposed. This includes so-called screening statistics that prioritize genetic variants to reduce the multiple testing burden [6–14]. Also, researchers proposed to aggregate genetic information in a genomic region in set-based tests [15–23]. Recent approaches utilize mixed models, reaction norm models, or incorporate multiple environmental variables to derive a combination of environmental factors that potentially modifies the genetic effect [24–28].

Current data analyses observed that genetic risk scores, (weighted) sums of genetic risk variants, interact with smoking exposure in Coronary heart disease [29] and lung function [30,31]. While detecting significant interactions on this aggregated level is encouraging, identifying individual genes and pathways underlying these effects is challenging based on such results.

Besides the power limitation, another caveat of gene-environment interaction testing is that the misspecification of the marginal main effect models can lead to false positive findings in interaction testing [32,33]. This is especially problematic for continuous environmental factors where the implicit linearity assumption that is commonly used in the model might not be correct. One example is pack-years of smoking, where it is not straightforward to assume that every additional pack-year has the same, constant effect on the outcome.

In this communication, we propose RITSS, a robust and flexible framework for gene-environment interaction testing in quantitative traits. The general idea is to derive an interaction score comprised of the (weighted) sum of individual genetic variant/environmental factor product pairs, such that the combination of these signals increases the power while maintaining biological interpretability. Our approach utilizes a sample splitting strategy and test statistics that are robust against misspecifications of the main effects. The general form of RITSS allows the incorporation of user-specified screening/learning approaches in constructing candidate interaction scores while providing valid statistical inference without restrictive assumptions. In extensive simulation studies, we demonstrate the robustness and power of RITSS in various scenarios. We also applied RITSS to lung function phenotypes and human height in the UK Biobank, incorporating sex information and smoking exposure.

Description of the method

Ethics statement

This research was conducted by using the UK Biobank resource under application number 20915. All participants provided written informed consent, and study protocols were approved by the North West Multi-centre Research Ethics Committee and ethical procedures were controlled by the UK Biobank Ethics Advisory Committee.

RITSS

In this section, we describe the RITSS framework and the corresponding algorithm.

Set up

For sample i, we denote the quantitative trait by Y_i, the m-dimensional genotype information by X_i, and the d-dimensional environmental information by E_i. Also, we define an additional p-dimensional covariate vector by Z_i that includes, for example, age, study indicators, and genetic ancestry principal components [34]. The genotype information X_i represents a preselected set of m genetic variants. We assume the following model

$$Y_i = \mu(E_i, Z_i) + \sum_{j=1}^{m} \pi_j(E_i, Z_i)X_{ij} + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{E}[\boldsymbol{\varepsilon}_i|X_i, Z_i, E_i] = 0$. The unknown function μ describes the main effect of the environmental factors E_i and other covariates Z_i. The genetic contribution of variant j is modeled by $\pi_j(E_i, Z_i)X_{ij}$, where π_j is an unknown function. The null hypothesis of no gene-environment interaction is described by $\boldsymbol{E}[Y_i|E_i, Z_i, X_i] = \mu(E_i, Z_i) + \sum_{j=1}^{m} \pi_{0j}(Z_i)X_{ij}$, where π_{0j} is an unknown function depending on Z_i only. Therefore, we implicitly assume the absence of gene-gene-interactions but, in general, allow for interactions between the genetic variants and the covariates Z_i under the null hypothesis.

The rationale of RITSS is to incorporate sets of genetic variants and/or multiple environmental factors to increase power and detect interactions based on aggregated signals. This step is motivated by the observation that single variant approaches are often underpowered. There are two critical aspects that RITSS aims to address: 1.) testing of aggregated interaction signals can increase power, but the results are more difficult to interpret since individual factor contributions are unobserved, and 2.) recent results demonstrated that the misspecification of main effects in interaction analyses can lead to false positive interaction findings [32,33].

Using sample splitting to screen for interaction signals

Given the input set of genetic and environmental factors, RITSS derives interaction scores of the form $U_i = \sum_j \sum_l \pi_{jl} X_{ij} E_{il}$ that combine signals adaptively while maintaining biological interpretability by keeping the number of involved factors of moderate size. RITSS uses sample splitting since screening and testing, in general, cannot be performed using the same data.

Testing of interaction scores using a robust approach

Standard interaction tests are based on a product between the candidate interaction score and the phenotype adjusted for main effects $\hat{Y}_i^{resid} = Y_i - \hat{\mu}(E_i, Z_i) - \sum_{j=1}^{m} \hat{\pi}_{0j}(Z_i) X_{ij}$. Zhang et al. showed that, for a quantitative trait (identical link function), main effects misspecifications can lead to invalid interaction tests, especially if genetic variants and environmental factors are not independent [33]. To avoid main effect misspecifications, they proposed to apply generalized additive models (GAMs) to model the main effects flexibly. RITSS follows this idea and incorporates rich models for the main effects, including GAM-based solutions. However, while Zhang et al. considered scenarios of single variant and single environmental factors, the dimensions of X_i, E_i, and Z_i, that is m, d, and p, in our setting are of substantial size. The potential high dimensionality of the models leads to slower convergence rates so that standard results might not apply here. Furthermore, considering RITSS as a general framework, we aim to derive an algorithm that can incorporate any other suitable future approach. Therefore, we utilize a concept based on recent theoretical results in causal inference [35]. Specifically, instead of considering the standard test statistic $\sum_i U_i \hat{Y}_i^{resid}$, RITSS is based on $\sum_i \hat{U}_i' \hat{Y}_i^{resid}$, where $\hat{U}_i'$ is an estimate of a transformation of U_i that is orthogonal to (i.e., has covariance zero with) the environmental and genetic main effects. This orthogonality improves the robustness against main effect model misspecifications since it reduces the impact of residual main effects in $\hat{Y}_i^{resid}$ on the product test statistic. Furthermore, we extend the sample splitting and estimate the main effect models as well as the transformation in non-overlapping subsamples of the data that also do not overlap with the subsamples used for the screening step and the subsamples used for evaluating the test statistic. This enables to apply any suitable approach for the screening step and the estimation of the main effect models/transformations, without the need to characterize the asymptotic behavior. It reduces the necessary assumptions to suitable convergence rates [36].

Algorithm

We here describe the algorithm underlying RITSS. The algorithm is also visualized in Fig 1.

Input: data X_i, E_i, Z_i, Y_i, screening function *Screen*(S) including the parameter S (number of candidate scores), parameter K for the number of splits, splitting fractions $c = (c_1, c_2, c_3)$.

Algorithm:

- Split the data (Y_i, X_i, E_i, Z_i) randomly into K non-overlapping subsamples I_k, $k = 1,\ldots,K$, of approximately equal size.
- For each subsample I_k:

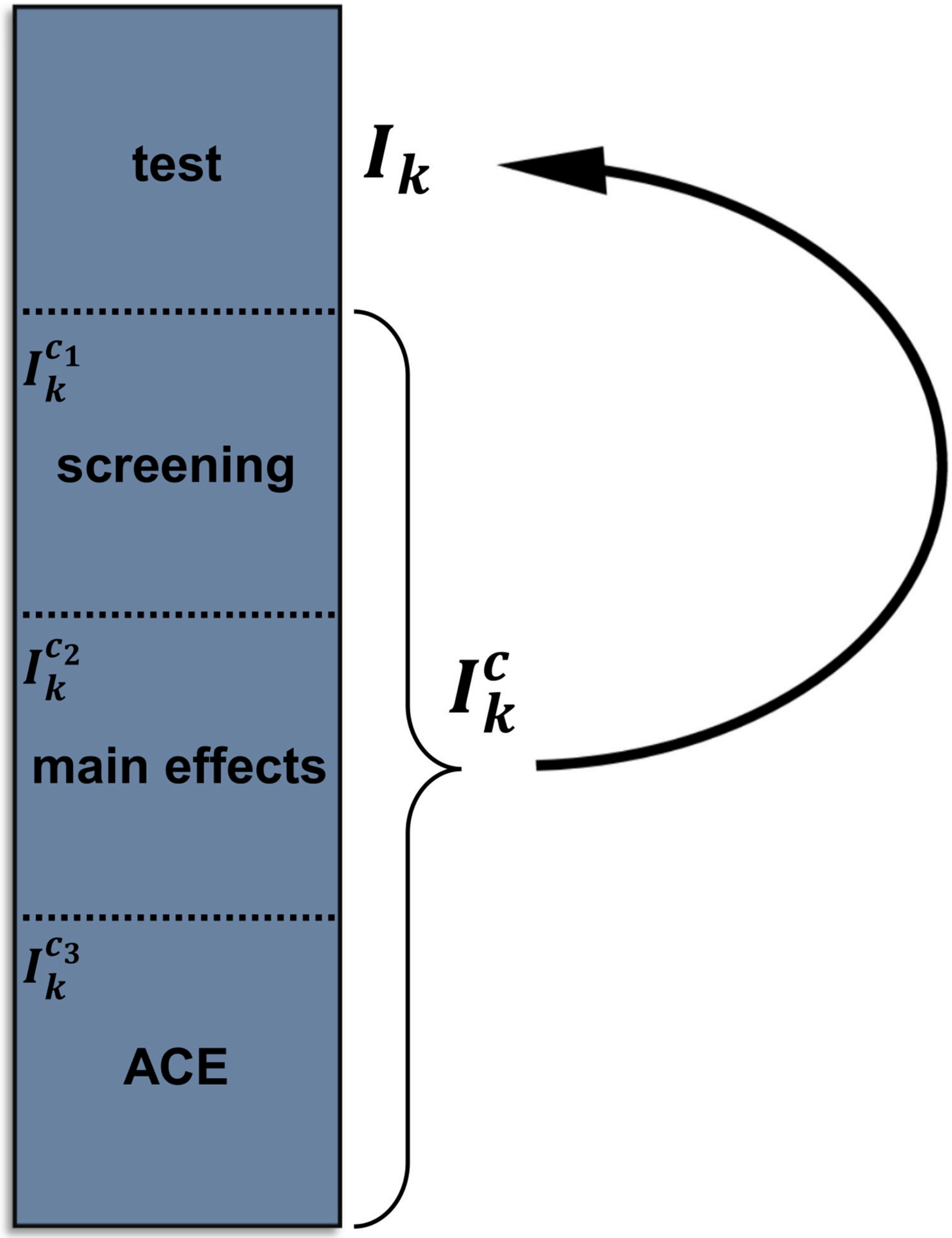

Fig 1. Visualization of the RITSS algorithm.

https://doi.org/10.1371/journal.pgen.1010464.g001

a. Denote all samples not in I_k by I_k^c and split I_k^c randomly into 3 non-overlapping parts I_k^{c1}, I_k^{c2}, and I_k^{c3}, according to the splitting fractions *c*.

b. Perform interaction signal screening based on I_k^{c1} using the screening function *Screen*(*S*) to obtain *S* candidate interaction scores $U_{is} = \sum_{j_s}\sum_{l_s}\pi^s_{j_s l_s}X_{ij_s}E_{il_s}$, *s* = 1,. . .,*S*.

c. In I_k^{c2}, fit the interaction scores U_{is}, *s* = 1,. . .,*S*, as predictors in a linear model while incorporating flexible models for the main effects $\hat{\mu}(E_i, Z_i)$ and $\sum_{j=1}^{m}\hat{\pi}_{0j}(Z_i)X_{ij}$. Select all or a subset of scores based on user-specified criteria and combine these scores into a single interaction score U_i.

d. Using the alternating conditional expectation (ACE) algorithm [37], estimate/derive the orthogonalized transformation U_i' of U_i in I_k^{c3} (S1 Appendix), denoted by $\hat{U}_i'$.

e. Based on the predictions of the trained models, compute $T_k = \sum_{i\epsilon I_k} \hat{U}_i' \hat{Y}_i^{resid}$ and $\hat{\sigma}_k^2 = \sum_{i\epsilon I_k} [\hat{U}_i' \hat{Y}_i^{resid}]^2$ in I_k.

- Compute the overall statistic $T = \sum_k T_k,\ \hat{\sigma}_T^2 = \sum_k \hat{\sigma}_k^2$, and the z-score $z = \frac{T}{\sqrt{\hat{\sigma}_T^2}}$.

Properties and discussion

Under the null hypothesis of no gene-environment interaction, the overall z-score z is asymptotically normal, that is $z \rightarrow N(0,1)$ [35,38] and an overall RITSS interaction p-value can be computed. This theoretical result applies if, besides standard regularity conditions, the main effects $\hat{\mu}(E_i, Z_i)+\sum_{j=1}^{m} \hat{\pi}_{0j}(Z_i)X_{ij}$ and the transformations $\hat{U}\prime_i$ are estimated with sufficiently fast convergence rates (as a function of sample size). These assumptions are closely related to the corresponding dimensions of genetic information, environmental factors, and covariates (m, d, and p, respectively), since the increasing dimensions increase the complexity of estimating these models sufficiently precise. This is where the advantage of the robust approach is apparent over, for example, solely application of generalized additive models (GAMs). The robust approach reduces the required assumptions to the statement that the specific products of convergence rates (main effects and transformation) are fast enough instead of requiring fast convergence rates for the main effects directly. As a remark, if we assume the absence of interactions between genetic variants and the covariates Z_i and an additive genetic model, that is $\pi_{0j}(Z_i) = \pi_{0j}$, the genetic main effect model becomes less complex, and faster convergence rates can be achieved. Nevertheless, since general models become complex with growing dimensions rapidly, we consider sets of m genetic variants with m in the magnitude of 10^2–10^3 and $d \leq 10$ and $p \leq 10$ as suitable input dimensions for RITSS.

We emphasize that it is therefore only required to estimate the transformation U_i' of a given interaction score U_i accurately, but the interaction score U_i itself can be constructed using any suitable method in the screening step, including machine learning or other advanced statistical learning procedures. Under the null hypothesis, the test is ensured to be valid.

On the other hand, the power of RITSS depends on the ability of the screening strategy to detect signals and capture them in the interaction score U_i. Therefore, although the screening step can be implemented flexibly, it is crucial for statistical power. We discuss and investigate two different screening strategies in the next section. It is important to note that the K different interaction scores U_i, in general, will include different interaction factors since the corresponding screening was performed in different parts of the data. Investigating the overlap between the interaction scores provides further fine mapping of the results.

Specific implementation

The screening strategy influences the statistical power of RITSS. There is no uniformly optimal strategy since the optimal screening depends on the interaction signal architecture. Specific implementations of RITSS differ in the realization of steps b.) and c.); steps a.), d.), and e.) are fixed throughout different implementations. Here, we describe two different implementations; we will refer to them as RITSS1 and RITSS2. These two implementations will be investigated in the simulation studies in the next section.

Screening strategy 1: interaction between single environmental factor and subcomponent of genetic risk score

The first screening strategy aims to identify subcomponents of genetic risk scores that interact with a single environmental factor E_{it}, where t is the index of the selected factor. In step b.), we first fit the model $Y_i = \mu_0 + E_i^T\mu_1 + Z_i^T\mu_2 + X_i^T\pi_0 + \varepsilon_i$ in I_k^{c1} to obtain estimated genetic main effects $\hat{\pi}_{0j}$. Then, we perform approximate best subset selection [39] to identify sets J_s, $s = 1,2$, that capture interaction scores of the form $U_{is} = \sum_{j\in J_s} \hat{\pi}_{0j} X_{ij} E_{it}$, describing subcomponents of genetic risk scores multiplied with the environmental factor E_{it}. The best subset selection is performed using a 2-sample split of I_k^{c1} and a step size of 5 between 10 and the maximum number of variants available after additional filtering. J_1 contains all variants that intersect between the best subsets, J_2 contains all variants that were included exactly once. More details are described in S1 Appendix.

In step c.), we construct the overall interaction score $U_i = U_{i1}+c_2U_{i2}$, where c_2 is set to 1, if the corresponding regression coefficient p-value is below 0.05, otherwise $c_2 = 0$. We denote the set of genetic variants in the interaction score U_i that is tested in I_k by $m(I_k)$.

Screening strategy 2: aggregated single environmental factor interactions based on single variant testing

In step b.), we fit the models $Y_i = \mu_0 + E_i^T\mu_1 + Z_i^T\mu_2 + X_{ij}\pi_{0j} + E_{it}X_{ij}\pi_{jt} + \varepsilon_i$ and select variants together with the estimated $\hat{\pi}_{jt}$ into the at most two interaction scores based on False Discovery Rate (FDR) q-values to correct for multiple testing. Step c.) is implemented as in screening strategy 1. We refer to S1 Appendix for more details. As for the screening strategy 1, we denote the set of genetic variants in the interaction score U_i that is tested in I_k by $m(I_k)$.

Verification and comparison

In this section, we describe the simulation studies and the application of RITSS to lung function and human height data in the UK Biobank.

Simulation studies

We performed extensive type 1 error and power simulations in which we investigated the performance of RITSS and compared our approach with alternative approaches. We emphasize that the power of RITSS depends on the combination of the underlying interaction signal architecture and the implemented screening strategy. Various screening strategies can be used, and different screening approaches are suitable for different signal architectures. In all simulation studies, we used the two specific implementations RITSS1 and RITSS2 described above. Both strategies considered the single environmental factor E_{it} with $t = 1$ for interaction scores. Moreover, we applied RITSS1 and RITSS2 with $K = 3, 4$ and the splitting fractions $c = (0.5, 0.25, 0.25)$ as well as $c = (1/3, 1/3, 1/3)$, to investigate the impact of these choices. Also, the implemented main effect models assume no gene-covariate interactions and an additive genetic model under the null hypothesis.

Alternative approaches

We included two alternative gene-environment interaction testing approaches in our simulation studies. The rationale is to demonstrate the differences and highlight the advantages of RITSS compared to existing and established methods.

1. We included GESAT, a variance component based test [17]. For GESAT, we considered E_{it} with $t = 1$ as the potential interaction factor, and Z_i as well as $E_{ij}, j = 2,...,d$ as covariates. We note that GESAT tests all m genetic variants jointly for interaction with E_{i1}; GESAT does not perform sample splitting and does not aim to identify the subset of variants that is involved in the interaction.
2. We implemented a single variant-based approach in which the environmental main effects are modelled using GAMs. Specifically, for each genetic variant, a GAM is fitted with the genetic variant and the interaction term as standard covariates, together with the flexible spline-based part for the environmental main effect (see S1 Appendix for more details). The m interaction p-values are summarized by the minimum p-value after Bonferroni correction for m tests. This approach will be denoted by GAMsv.

Furthermore, besides GESAT and GAMsv, we also included a modified version of RITSS1 and RITSS2 in which the test statistic utilizes the interaction scores U_i directly, instead of the robust version based on the transformed U_i'. The rationale is to demonstrate the necessity of this step in the high-dimensional setting; this approach will be denoted by D1 and D2, respectively.

Simulation settings: type 1 error

In all scenarios, we simulated a sample size of $n = 30{,}000$, $m = 100$ SNPs, $d = 5$ environmental factors E_i, and $p = 2$ covariates Z_i. The phenotype Y_i was simulated based on $Y_i = \mu(E_i, Z_i) + X_i^T\pi_0 + \varepsilon_i(E_i, Z_i)$ with $\boldsymbol{E}[\varepsilon_i(E_i, Z_i)|X_i, Z_i, E_i] = 0$.

Under the null hypothesis of no gene-environment interaction, the overall RITSS z-score is asymptotical normal with $z \rightarrow N(0,1)$. The type 1 error simulations aim to demonstrate the accuracy of this asymptotic approximation in realistic scenarios.

As described in the methods section, the screening strategies of RITSS1 and RITSS2 assume linear environmental main effects. We included scenarios in which the true environmental main effect is non-linear and gene-environment correlation is present. In these scenarios, the screening strategies, therefore, construct interaction scores that are reflecting false positive interaction findings [32,33]. Besides, we included scenarios with non-normal and/or heteroscedastic random errors in the phenotype model. In total, we simulated five different scenarios: (1): no misspecifications, normal homoscedastic errors, (2): mis specified environmental main effects in the screening strategies, (3): mis specified environmental main effects in the screening strategies in combination with non-normal heteroscedastic errors, (4): scenario 2 including gene-environment correlations, and (5): scenario 3 including gene-environment correlations.

In addition, we investigated if the type 1 error rate is inflated in the case where the genetic variants for the analysis were selected based on GWAS results in the same dataset, i.e., based on their marginal association p-values (scenario notated as SELECT:yes/no). This is of particular interest since many reported genetic associations were identified based on the large-scale datasets that are suitable for gene-environment interaction testing. In total, this resulted in 10 different type 1 error scenarios (scenarios 1–5, SELECT:yes/no).

The specific details of the implementations of these scenarios are described in S1 Appendix. The type 1 error simulations are based on 10,000 replicates.

Simulation settings: power

In addition to the type I error simulations, we also included a set of power simulations. In the power simulations, we considered scenario 1 with SELECT:no from the type 1 error simulations but incorporated a gene-environment interaction component. The interaction contribution is controlled by the mean μ_{XE} and standard deviation σ_{XE} of interaction effects, as well as the density p_{XE} of these interaction effects. The interaction model incorporates the first component E_{i1} only. The detailed implementation is described in S1 Appendix. The power simulations are based on 1,000 replicates.

Results: the choice of *K* and the splitting fractions *c*

RITSS can be applied using different choices for *K* and the splitting fractions *c*. While the asymptotic theory, that is the asymptotic standard normality of the RITSS z-score under the null hypothesis, applies for any fixed choice, the finite sample performance in practice as well as the power can be influenced by these choices. Choosing these parameters faces specific tradeoffs. For a fixed *K*, splitting fractions *c* that allocate more sample size to the screening strategy are expected to have higher power, while the decreased sample size for the main effect and transformation estimation might lead to less accurate model fits and therefore inflated type 1 error rates. Since every sample contributes exactly once to the test statistic for each fixed *K*, increasing *K* (moderately) is expected to increase power since a bigger fraction of the overall sample is used for screening. At the same time, a larger *K* can lead to increased computational burden, depending on how the screening and main effect model estimation tasks scale with sample size. We considered the four combinations of $K = 3, 4$ and the splitting fractions $c = (0.5, 0.25, 0.25)$ and $c = (1/3, 1/3, 1/3)$. In Table B in S1 Appendix, we report initial type 1 error results for scenarios 1–5 with SELECT:no for all four combinations based on 1,000 replicates. Table C in S1 Appendix summarizes power results for selected scenarios based on 1,000 replicates.

The results in Table B in S1 Appendix show that, although all four combinations provide approximately uniformly distributed p-values under these null hypothesis scenarios, the splitting fraction $c = (1/3, 1/3, 1/3)$ leads to the most stable results; this is in line with the intuition described above. Furthermore, D1 and D2 do not control the type 1 error rates, demonstrating that the robust test statistics are needed to provide valid results. The power results in Table C in S1 Appendix demonstrate that $K = 4$ has slightly more power than $K = 3$, as intuitively described above, but the overall power differences between the four combinations are small. For the full type 1 error and power simulations, we considered the combination of $K = 4$ and $c = (1/3, 1/3, 1/3)$.

Results: type 1 error

We report the quantile-quantile-plots (qq-plots) for scenarios 1–5 with SELECT:no in Fig 2, as well as the qq-plots for scenarios 1–5 with SELECT:yes in Fig 3.

Overall, the simulations demonstrate that the results for SELECT:no and SELECT:yes are similar. RITSS1 and RITSS2 provide controlled type 1 error rates across all five scenarios. GESAT demonstrates valid results for scenario 1 but is highly inflated in the scenarios 2–5. We note that we cut off the GESAT p-values at 10^{-10} in the plots for better visualization. The inflated results are expected since GESAT does not account for heteroscedastic errors (scenarios 3 and 5) and the linear main effect model implies a mis specification in scenarios 2–5. GAMsv, which models the environmental main effects similar to RITSS, shows valid type 1 error rates for scenarios 1, 2, 4, and 5. In scenario 3, the heteroscedastic errors lead to inflation. In summary, our type 1 error simulations demonstrate the necessity of flexible main effect models to ensure robustness and that RITSS provides valid type 1 error rates in all investigated scenarios.

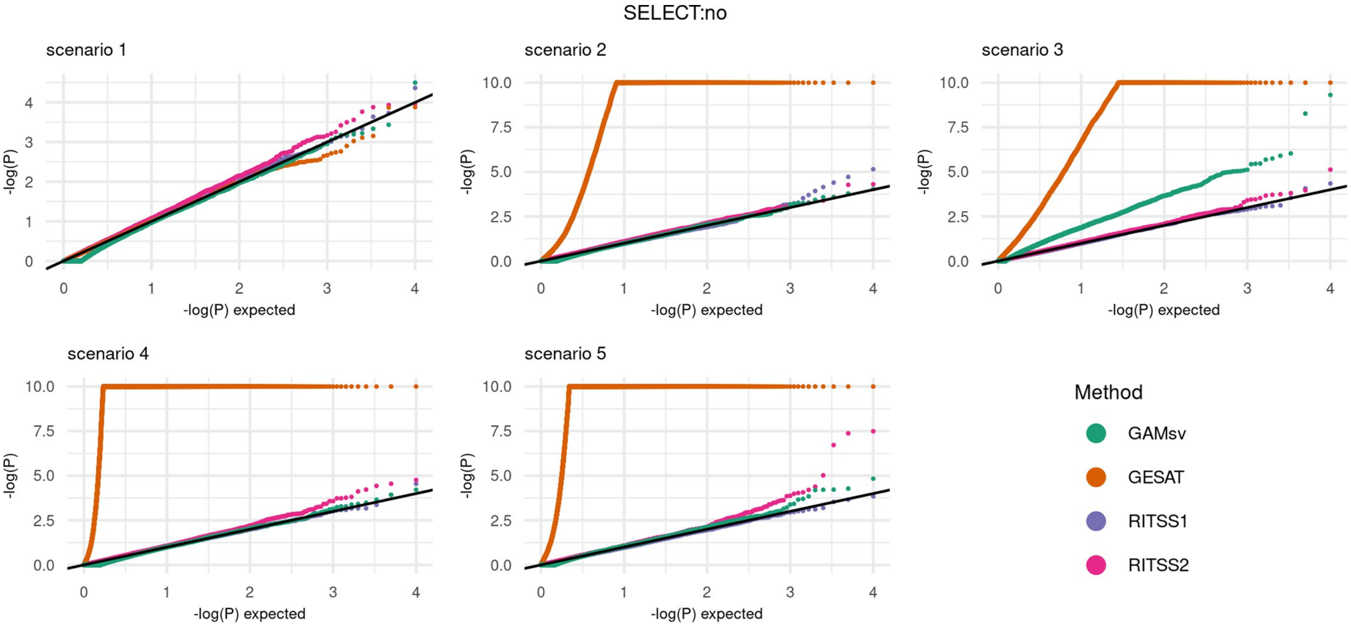

Fig 2. Type 1 error: Quantile-quantile-plots for RITSS1, RITSS2, GESAT, and GAMsv for scenarios 1–5 with SELECT:no. All results based on 10,000 replicates. P-values with $p<10^{-10}$ were set to $p = 10^{-10}$. SELECT:no refers to the scenario where all simulated genetic variants are included in the analysis.

https://doi.org/10.1371/journal.pgen.1010464.g002

Results: power

The results of the power simulations are reported in Figs 4 and 5, where the power is displayed as a curve over increasing signal density p_{XE}, for different effect size distributions. The significance level was set to $\alpha = 0.005$.

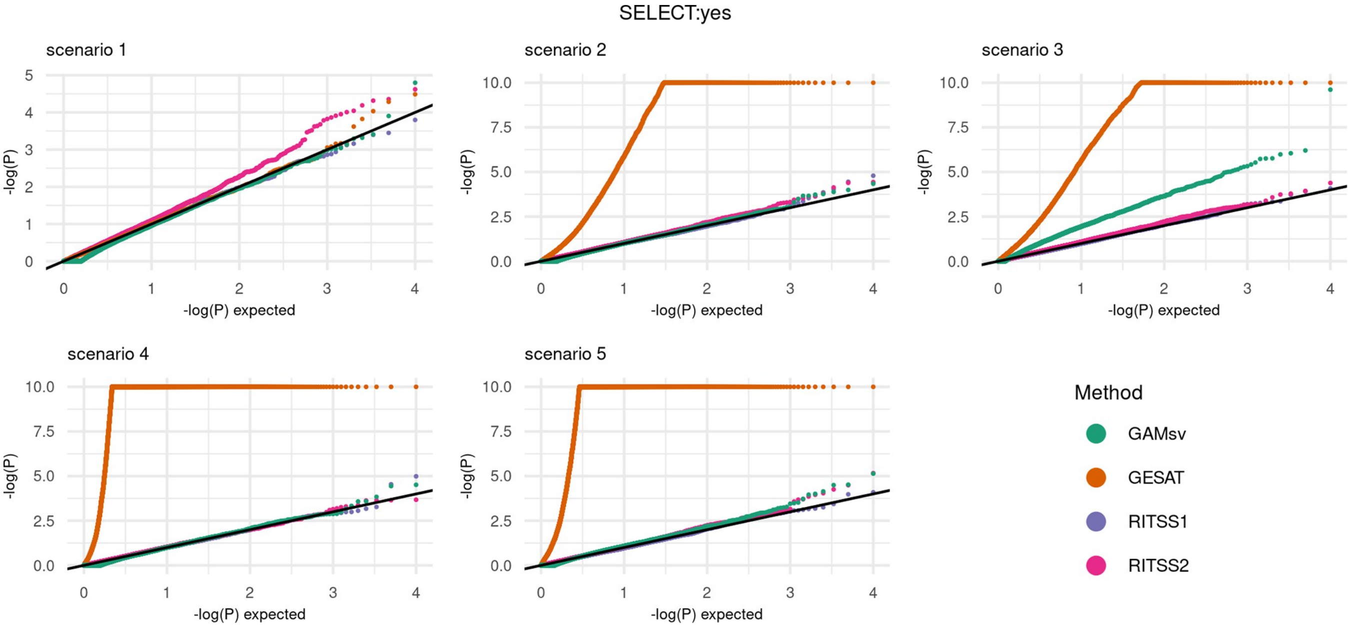

Fig 3. Type 1 error: Quantile-quantile-plots for RITSS1, RITSS2, GESAT, and GAMsv for scenarios 1–5 with SELECT:yes. All results based on 10,000 replicates. P-values with $p<10^{-10}$ were set to $p = 10^{-10}$. SELECT:yes refers to the scenario where the simulated genetic variants are selected into the analysis based on marginal association p-values.

https://doi.org/10.1371/journal.pgen.1010464.g003

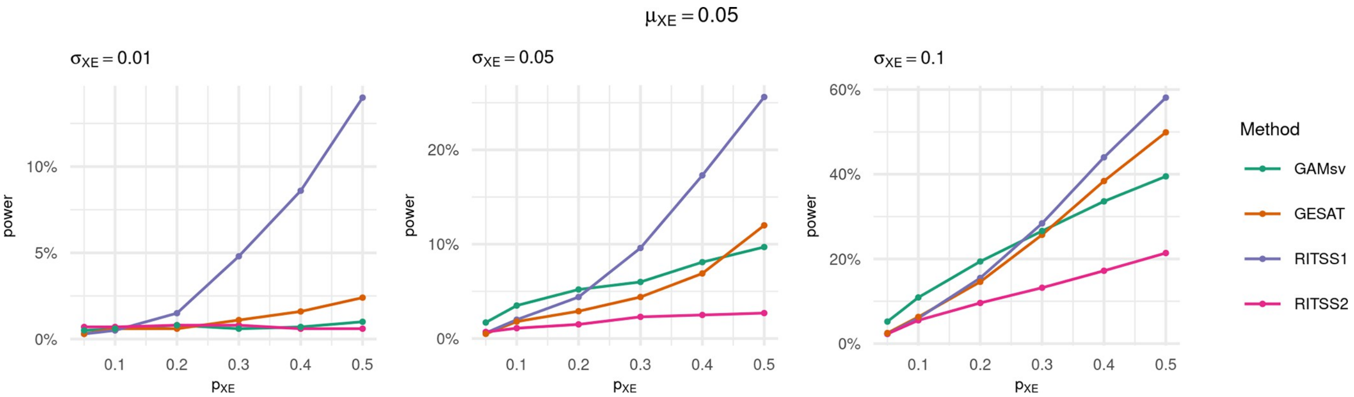

Fig 4. Power curves for RITSS1, RITSS2, GESAT, and GAMsv over increasing signal density p_{XE}. Significance level $\alpha = 0.005$, $\mu_{XE} = 0.05$, and results based on 1,000 replicates. Power was simulated for $p_{XE} = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$.

https://doi.org/10.1371/journal.pgen.1010464.g004

We consider the results for $\mu_{XE} = 0.05$ in Fig 4 first. We observe that RITSS1 is more powerful than RITSS2 with increasing signal density. This is expected since the simulated signal architecture represents interactions with subcomponents of risk scores, and the screening strategy of RITSS1 aims to detect exactly these signals.

RITSS1 outperforms GESAT and GAMsv in most scenarios, especially for $\sigma_{XE} = 0.01$ and increasing p_{XE}. The power difference is smaller for $\sigma_{XE} = 0.1$. With larger effect sizes and sparse signals, GAMsv shows advantages since the single variant tests are powerful enough to circumvent the multiple testing burden in this case.

The results in Fig 5 for $\mu_{XE} = 0.1$ are similar, but the power advantages of RITSS1 are smaller or vanish ($\sigma_{XE} = 0.1$). Overall, the power simulations demonstrate the interplay between the screening strategy implemented in RITSS and the resulting power. Especially for dense signals with small effect sizes, RITSS1 shows substantial power advantages. We hypothesize that these signal architectures approximate the signals observed in recent interaction studies based on polygenic risk scores.

Applications

We applied RITSS to UK Biobank data to analyze gene-environment interactions in lung function (measured by forced expiratory volume in 1 second (FEV$_1$), forced vital capacity (FVC),

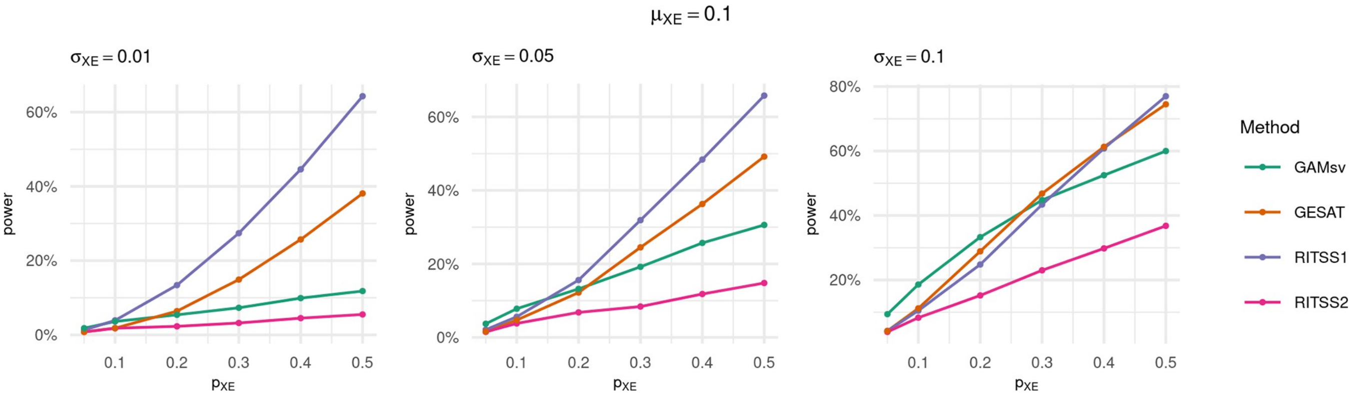

Fig 5. Power curves for RITSS1, RITSS2, GESAT, and GAMsv over increasing signal density p_{XE}. Significance level $\alpha = 0.005$, $\mu_{XE} = 0.1$, and results based on 1,000 replicates. Power was simulated for $p_{XE} = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$.

https://doi.org/10.1371/journal.pgen.1010464.g005

and the ratio FEV_1/FVC) and height. Here, for our analyses, we restricted the input to previously reported GWAS associations for the respective trait. This step reduces the dimensionality of X_i, denoted by m, to a range between 383 and 3368. We note that this choice can be too restrictive since interaction effects do not need to be accompanied by main effects. Details about the study population as well as the extraction of genetic, environmental, and phenotypic data are described in S1 Appendix.

Analysis setup

Table 1 contains the configurations of Y_i, X_i, E_i, and Z_i for the four different phenotype analyses. Age and pack-years of smoking (P-Y-S) were mean-centered before computing the squared variable. Sex was coded as male = 1 and female = 0 in this analysis. For the lung function traits, we also included height in E_i as a potentially interacting variable. We note that, given the coding of sex in this analysis, height and sex are (strongly) positively correlated. The lung function measurements were analyzed on the original scale and not transformed, as robustness of the approach against non-normal errors was demonstrated in the type 1 error simulations. After quality control, we split the resulting 254,033 samples (European ancestry) in two parts: 180,000 randomly selected samples for the main analysis using RITSS and 74,033 samples that serve as an independent dataset to validate the analysis results of RITSS. We applied RITSS using the same two different implementations described in the simulation studies (RITSS1 and RITSS2). For both RITSS1 and RITSS2, we chose $K = 4$ and $c = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Each of the factors in E_i was tested for interaction separately, resulting in a total of 32 tests. In S1 Fig, we plotted the estimated densities of the standardized residuals for each phenotype after adjusting for X_i, E_i, and Z_i using standard linear regression. The density plots are based on the 180,000 samples in the main analysis.

Results

Table 2 contains the results of our UK Biobank main analysis using 180,000 samples based on RITSS1 and RITSS2. For each of the four traits, we observed a highly significant interaction between sex and subcomponents of the genetic risk score using RITSS1. For the lung function traits, we also observed similar interaction findings with height. We investigated this in more detail in the validation analysis below.

Furthermore, in line with previous results in the literature, we observed a significant interaction with pack-years of smoking in FEV_1/FVC [30,31] that maintains significant after Bonferroni correction for $n = 32$ tests at a significance level of $\alpha = 0.05$.

Based on the $K = 4$ different interaction scores, we investigated the overlap between the genetic variants included in the sets $m(I_k)$, $k = 1,\ldots,4$. We denote all genetic variants that were

Table 1. Analysis configurations and number of genetic variants incorporated.

Y_i	X_i	E_i	Z_i
FEV_1	383 SNPs	sex, height, P-Y-S, E-S, P-Y-S^2	age, age^2, PCs 1–10, genotyping-array
FVC	638 SNPs	sex, height, P-Y-S, E-S, P-Y-S^2	age, age^2, PCs 1–10, genotyping-array
FEV_1/FVC	885 SNPs	sex, height, P-Y-S, E-S, P-Y-S^2	age, age^2, PCs 1–10, genotyping-array
height	3368 SNPs	sex	age, age^2, PCs 1–10, genotyping-array

P-Y-S: pack-years of smoking, E-S: ever-smoking, PCs: genetic ancestry principal components. FEV_1: forced expiratory volume in 1 second, FVC: forced vital capacity (FVC). Pack-years of smoking and age were centered before computing the respective squared variable.

https://doi.org/10.1371/journal.pgen.1010464.t001

Table 2. Results of the interaction testing using the two approaches RITSS1 and RITSS2 in the UK Biobank. The environmental factor tested for interaction is denoted by E_{it}. $|m|$ is the number of total SNPs in the analysis, $|m_4|$ and $|m_3|$ are the number of SNPs that are shared by all four and exactly three interaction scores, respectively. P-Y-S: pack-years of smoking, E-S: ever-smoking.

Y_i	E_{it}	m	RITSS1 p-value	$\|m_4\|$	$\|m_3\|$	RITSS2 p-value	$\|m_4\|$	$\|m_3\|$
FEV$_1$	sex	383	2.86E-25	44	54	7.68E-01	0	0
FEV$_1$	height	383	1.75E-13	56	52	4.81E-01	0	0
FEV$_1$	P-Y-S	383	8.81E-02	1	3	6.37E-01	0	0
FEV$_1$	E-S	383	6.08E-01	0	12	4.16E-01	0	0
FEV$_1$	P-Y-S^2	383	5.41E-02	0	2	8.77E-01	0	0
FVC	sex	638	1.27E-22	46	94	1.76E-01	0	0
FVC	height	638	1.07E-21	54	101	5.10E-01	0	0
FVC	P-Y-S	638	6.08E-01	1	1	4.51E-01	0	0
FVC	E-S	638	1.19E-01	1	4	4.02E-01	0	0
FVC	P-Y-S^2	638	6.77E-01	0	0	9.77E-01	0	0
FEV$_1$/FVC	sex	885	7.33E-22	83	106	3.91E-03	0	0
FEV$_1$/FVC	height	885	8.41E-16	74	119	9.13E-01	0	0
FEV$_1$/FVC	P-Y-S	885	2.01E-04	78	100	2.61E-01	0	0
FEV$_1$/FVC	E-S	885	6.35E-03	74	106	5.56E-04	0	0
FEV$_1$/FVC	P-Y-S^2	885	1.07E-01	0	1	2.35E-01	0	3
height	sex	3368	1.51E-20	210	341	2.78E-01	0	0

https://doi.org/10.1371/journal.pgen.1010464.t002

in all four scores by m_4 and all genetic variants included in exactly three scores by m_3. For the validation analysis, we considered all Y_i/E_{it} pairs that had a Bonferroni corrected significant interaction p-value at a significance level of $\alpha = 0.05$ and non-empty m_4 or m_3 sets; this included RITSS1 results only.

The validation analysis was based on the remaining and independent 74,033 samples and a standard regression interaction test while fitting X_i, E_i, and Z_i. We tested two interaction scores, the first based on the variants in the corresponding m_4 set and the second based on variants in m_3. Here, the genetic main effects for the genetic risk scores were estimated based on the 180,000 samples in the main analysis. The regression interaction p-values were evaluated based on model-based standard errors as well as standard errors obtained from a robust sandwich estimator.

The results of our validation analysis are described in Table 3. The results show that all findings replicate. We also note that the directions of effects were consistent between the main analysis and validation analysis.

Table 3. Results of the validation analysis in the 74,033 remaining samples in the UK Biobank. The two sets m_4 and m_3 were identified in the corresponding main analysis (Table 2) and correspond to the results of the RITSS1 analysis. The interactions between the scores (corresponding to m_4 and m_3 respectively) and the environmental factor E_{it} were tested using a linear regression, adjusting for all X_i, E_i and Z_i. The robust p-values are based on standard errors obtained by a robust sandwich variance estimator. P-Y-S: pack-years of smoking.

Y_i	E_{it}	p-value m_3	robust p-value m_3	$\|m_3\|$	p-value m_4	robust p-value m_4	$\|m_4\|$
FEV$_1$	sex	9.11E-06	2.68E-05	54	3.41E-10	1.65E-09	44
FEV$_1$	height	8.26E-07	2.30E-06	52	2.19E-09	1.81E-08	56
FVC	sex	1.48E-07	5.61E-07	94	5.29E-07	1.68E-06	46
FVC	height	1.96E-07	9.79E-07	101	5.04E-10	6.96E-09	54
FEV$_1$/FVC	sex	7.23E-04	8.08E-04	106	4.88E-07	6.80E-07	83
FEV$_1$/FVC	height	4.95E-05	6.63E-05	119	1.36E-06	1.82E-06	74
FEV$_1$/FVC	P-Y-S	1.21E-04	2.91E-03	100	2.22E-07	6.31E-05	78
height	sex	1.46E-02	1.56E-02	341	2.07E-02	2.19E-02	210

https://doi.org/10.1371/journal.pgen.1010464.t003

Table 4. Results of the validation analysis in the 74,033 remaining samples in the UK Biobank II. The two sets m_4 and m_3 were identified in the corresponding main analysis based on the RITSS1 approach (Table 2). All variants in m_4/m_3 were tested for interaction with the respective environmental factor E_{it}. In this context, gws. interaction refers to a single variant interaction p-value $p<5\times10^{-8}$. Similar, bfs. interaction corresponds to $p<0.05/(|m_4|+|m_3|)$ (Bonferroni correction). The last columns report the empirical distribution of effect directions in these tests and the corresponding 5% quantile of p-values (weighted analysis). In the weighted analysis, the genotype/env. factor product term was multiplied by the estimated genetic main effect (obtained from the independent main dataset). P-Y-S: pack-years of smoking.

Y_i	E_{it}	$\lvert m_3\rvert+\lvert m_4\rvert$	No. gws interactions	No. bfs. interactions	Distribution effect directions unweighted -/+ (in %)	Distribution effect directions weighted -/+ (in %)	5% quantile interaction p-values in weighted -/+
FEV_1	sex	98	0	1	53/47	23/77	1.636e-01/8.190e-03
FEV_1	height	108	0	1	51/49	29/71	7.963e-02/7.341e-03
FVC	sex	140	0	0	46/54	26/74	3.493e-02/1.107e-02
FVC	height	155	0	0	43/57	31/69	4.220e-02/1.103e-02
FEV_1/FVC	sex	189	0	0	51/49	37/63	1.730e-01/1.962e-02
FEV_1/FVC	height	193	0	0	48/52	39/61	1.478e-01/2.565e-02
FEV_1/FVC	P-Y-S	178	0	0	46/54	36/64	6.412e-02/1.051e-02
height	sex	551	0	0	49/51	45/55	5.655e-02/3.849e-02

https://doi.org/10.1371/journal.pgen.1010464.t004

To disentangle these findings in more detail, we examined the single variant interaction p-values for the genetic variants in the combined set m_4/m_3 with the respective environmental factor in the validation dataset. This analysis was based on a linear regression that fits X_i, E_i, and Z_i, and the respective interaction term. Table 4 lists the number of genome-wide significant single variant interactions, i.e., $p<5\times10^{-8}$, as well as the number of significant single variant interactions using a Bonferroni correction adjusting for the number of variants in the corresponding combined set m_4/m_3 at a significance level of $\alpha = 0.05$. Both numbers are either 0 or very small, for all trait/environmental factor combinations. This suggests that the interaction effects at the single variant level are weak. However, we made an interesting observation by investigating the empirical distribution of directions of effect across the single variant interaction tests. Table 4 provides these empirical distributions of directions of effect for a.) the unweighted interaction terms $X_{ij}\times E_{it}$ and b.) the weighted terms $\hat{\pi}_{0j} * X_{ij} * E_{it}$, where $\hat{\pi}_{0j}$ is the estimated genetic effect based on the main analysis dataset. In addition, we compared the interaction p-value distributions according to the directions of effect of the weighted interaction terms. As we can see by the enrichment of positive effects, the weighting by the genetic main effect, as it is applied in the computation of the genetic risk score, "aligns" the interaction effect directions across the genetic variants. Furthermore, the p-values for weighted interaction terms with positive effect show substantially higher inflation. This indicates that RITSS1 identified genetic variants whose interaction signals are weak at the single variant level but when summed up in the genetic risk score, they lead to a significant interaction. Since we used a coding of sex of male = 1 and female = 0, this suggests that the genetic effects in males and females across these variants have the same direction, but the magnitudes of effect in males are slightly larger. Related, we note that the genetic variants in m_4/m_3 between the sex and height interactions in lung function highly overlap, which is in line with the observation that height and sex are strongly positively correlated. However, if we test the height interaction separately in males and females in the validation dataset, the signal vanished or was strongly diminished. Therefore, we hypothesize that the observed interactions are a result of sex-differential effects, but further analyses are required to disentangle the exact mechanisms. For example, investigating the impact of sitting or thoracic height, instead of standing height [40]. In this context, we note that, related to recent results regarding so-called participation biases and their impact on sex-related analyses [41], the genetic variants in m_4/m_3 across the three lung function phenotypes and height were not significantly associated with sex, but replication of the interaction signal in an independent cohort is preferable. Detailed information about the selected variants in the gene-by-sex interactions for lung function and height can be found in S2 Appendix.

We also applied GESAT in the validation analysis. For each of the four traits, we applied GESAT to test for interaction with sex (using the remaining E_i and Z_i as covariates) based on 1.) all genetic variants in the analysis and 2.) all genetic variants in the corresponding combined m_4/m_3 set. For all traits, these tests were at least nominal significant ($p<0.05$), and we obtained similar p-values in the two analyses 1.)/2.) (except for FVC): FEV$_1$: $p = 1.81\times10^{-11}$/ $p = 8.92\times10^{-7}$, FVC: $p = 2.71\times10^{-14}$/$p = 2.29\times10^{-4}$, FEV$_1$/FVC: $p = 2.61\times10^{-2}$/$p = 2.29\times10^{-2}$, height: $p = 1.08\times10^{-3}$/$p = 6.19\times10^{-3}$. These results add further evidence that RITSS1 identified a fine mapped set of interacting genetic variants.

Discussion

We propose RITSS, a robust and flexible interaction testing framework for quantitative traits. Our framework aims to identify aggregated interaction signals and tests these signals using sample splitting and robust test statistics. Since interactions at the single genetic variant level are hard to detect due to small effect sizes, we hypothesize that strategies that aggregate suitable signals across a limited number of factors/loci have higher statistical detection power while, at the same time, they still permit an interpretation and can reveal potential biological mechanisms.

In extensive simulations, we investigated the performance of RITSS based on two distinct implementations and demonstrated that our framework controls the type 1 error rates across different scenarios. In contrast, alternative approaches showed inflated type 1 error rates in these scenarios. Additionally, we demonstrated the interplay between the screening strategy and RITSS's power based on selected power simulations. We applied our approach to UK Biobank data and observed potential interactions between subcomponents of risk scores and sex in lung function and height. Since interactions for complex traits at the single genetic variant level were rarely detected, even in recent large-scale analyses, we hypothesize that our approach will be an important tool for the identification of genetic interactions and the underlying mechanisms.

Fawcett et al. performed a genome-wide genotype-by-sex interaction analysis of lung function in the UK Biobank, using more than 300,000 samples [40]. Although they found a small number of five genome-wide significant interactions, only one interaction signal was replicated in an independent study. In a different study, Bernabeu et al. utilized the UK Biobank to analyze genotype by sex interactions for 530 complex traits. This analysis revealed several genome-wide significant interaction findings, but heritability and genetic correlation analyses suggest that substantial proportions of the sex-differential genetic architecture are yet to be discovered and identified [42]. Our results are in line with these findings since our interaction signals are not driven by strong effects at the single variant level but aggregations across sets of variants. Furthermore, recent work by Zhu et al. hypothesizes and demonstrates that so-called amplification is a common mode of genotype-by-sex interactions [43]. Amplification describes the setting of shared genetic variants, with the same direction of effect, but different magnitudes. This is exactly what we observed in our UK Biobank sex interaction analysis of lung function and human height. However, we note that analyses based on independent datasets are required to further replicate, validate, and disentangle these potential gene-by-sex interactions.

Our approach has the following limitations. A significant interaction p-value of the aggregated score does not imply that all included pairs of genetic variants/environmental factors necessarily truly contribute to the interaction signal. Another limitation is that the set of genetic variants needs to be pre-selected and, as described in the Methods section, the size m is assumed to be of moderate magnitude in the range of 10^2–10^3 (assuming typical GWAS sample sizes). This restriction is related to the required assumption that the main effects and transformations can be estimated accurate enough given the sample size. The computational burden, depending on the specific implementation of the screening strategy, also increases with m. Therefore, an important question is how to pre-select the set of genetic variants for RITSS testing. It is possible to integrate an external screening approach. For a quantitative trait, one can screen for potential interaction signals based on marginal effects or variance-effects [10,14]. Based on the results of such a screening, one can choose sets of genetic variants for RITSS testing and introduce weighted hypotheses reflecting the results of this external screening approach. It is important to note that this screening should not be confused with the screening and testing steps within the RITSS framework. In the UK Biobank application, we restricted the input to previously reported GWAS findings for the respective trait, which can be interpreted as a special case of this prior external screening. However, interaction effects do not need to be accompanied by marginal effects and therefore our analysis provides only a limited insight.

In general, the interpretation of significant RITSS p-values requires careful consideration. RITSS aims to identify deviations from the null hypothesis model $\boldsymbol{E}[Y_i|X_i, Z_i, E_i] = \mu(E_i, Z_i) + \sum_{j=1}^{m} \pi_{0j}(Z_i)X_{ij}$, represented as significant interaction scores $U_i = \sum_j \sum_l \pi_{jl} X_{ij} E_{il}$. However, a significant contribution by U_i does not imply interaction between E_i and X_i without assumptions regarding no unmeasured confounding/omitted variables and the sampling design [44,45]. Moreover, measurement of environmental factors is often subject to noise or even bias. This can result in false positive interaction findings since RITSS assumes that environmental factors are measured without error. Furthermore, due to the computational burden, our type 1 error simulations are based on 10,000 replicates only and, therefore, the validity at small significance levels cannot be empirically investigated. Nevertheless, we emphasize that the RITSS approach is supported by theoretical results describing the asymptotic normality of the test statistic. The simulation results demonstrated the concordance with these theoretical approximations. Finally, due to the application of sample splitting and several independent estimation steps, RITSS requires datasets with sample sizes that are in the order of recent GWAS publications (30,000 samples or more).

Future directions include more detailed follow-up analyses of the observed gene-by-sex interactions, the incorporation of other screening techniques, such as, for example, LASSO [46] or RaSE [47], and the improvement of the computational efficiency. We plan to extend RITSS to incorporate dichotomous traits and implement these extensions in the R package *RITSS*. We note that RITSS can also be used to analyze gene-gene-interactions and interaction models can be based on more general, non-linear approaches. In addition, RITSS can be applied to other omics data layers, such as DNA methylation, transcriptomics, or metabolomics. The R package *RITSS* as well as the code for the simulation studies is available at https://github.com/julianhecker/RITSS.

Supporting information

S1 Appendix. ACE algorithm and RITSS main effects, Screening strategies, Implementation of GAMsv, Simulation details, Additional simulation study results, and UK Biobank data. This appendix contains additional information regarding the ACE algorithm and RITSS main effects, the screening strategies, the implementation of GAMsv, simulation details,

additional simulation study results, and the UK Biobank data.
(PDF)

S2 Appendix. Additional information regarding genetic variants tested in the UK Biobank analysis. These tables contain all genetic variants as well as the m_4/m_3 information for the sex-interaction analysis in the UK Biobank.
(XLSX)

S1 Fig. Density plots for standardized residuals for all four traits in the analysis. FEV_1: forced expiratory volume in 1 second, FVC: forced vital capacity.
(TIF)

Author Contributions

Conceptualization: Julian Hecker, Dmitry Prokopenko, Scott T. Weiss, Christoph Lange.

Data curation: Julian Hecker, Dmitry Prokopenko, Matthew Moll, Wonji Kim, Dandi Qiao, Woori Kim, Brian D. Hobbs, Michael H. Cho.

Formal analysis: Julian Hecker.

Funding acquisition: Scott T. Weiss, Christoph Lange.

Methodology: Julian Hecker, Stijn Vansteelandt, Sharon M. Lutz, Christoph Lange.

Resources: Michael H. Cho, Scott T. Weiss.

Software: Julian Hecker.

Supervision: Scott T. Weiss, Christoph Lange.

Visualization: Julian Hecker.

Writing – original draft: Julian Hecker, Dmitry Prokopenko, Scott T. Weiss, Christoph Lange.

Writing – review & editing: Matthew Moll, Sanghun Lee, Wonji Kim, Dandi Qiao, Kirsten Voorhies, Woori Kim, Stijn Vansteelandt, Brian D. Hobbs, Michael H. Cho, Edwin K. Silverman, Sharon M. Lutz, Dawn L. DeMeo.

References

1. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017; 45: D896–D901. https://doi.org/10.1093/nar/gkw1133 PMID: 27899670
2. Hunter DJ. Gene-environment interactions in human diseases. Nat Rev Genet. 2005; 6: 287–298. https://doi.org/10.1038/nrg1578 PMID: 15803198
3. Khoury MJ. Editorial: Emergence of Gene-Environment Interaction Analysis in Epidemiologic Research. Am J Epidemiol. 2017; 186: 751–752. https://doi.org/10.1093/aje/kwx226 PMID: 28978194
4. Thomas D. Gene—environment-wide association studies: emerging approaches. Nat Rev Genet. 2010; 11: 259–272. https://doi.org/10.1038/nrg2764 PMID: 20212493
5. 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. Am J Epidemiol. 2017; 186: 762–770. https://doi.org/10.1093/aje/kwx228 PMID: 28978192
6. Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ. Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genet Epidemiol. 2011; 35: 201–210. https://doi.org/10.1002/gepi.20569 PMID: 21308767
7. Dai JY, Kooperberg C, Leblanc M, Prentice RL. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. Biometrika. 2012; 99: 929–944. https://doi.org/10.1093/biomet/ass044 PMID: 23843674

8. Gauderman WJ, Thomas DC, Murcray CE, Conti D, Li D, Lewinger JP. Efficient genome-wide association testing of gene-environment interaction in case-parent trios. Am J Epidemiol. 2010; 172: 116–122. https://doi.org/10.1093/aje/kwq097 PMID: 20543031

9. Gauderman WJ, Zhang P, Morrison JL, Lewinger JP. Finding Novel Genes by Testing G×E Interactions in a Genomewide Association Study. Genet Epidemiol. 2013; 37: 603–613. https://doi.org/10.1002/gepi.21748 PMID: 23873611

10. Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C. Powerful cocktail methods for detecting genome-wide gene-environment interaction. Genet Epidemiol. 2012; 36: 183–194. https://doi.org/10.1002/gepi.21610 PMID: 22714933

11. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. Genet Epidemiol. 2008; 32: 255–263. https://doi.org/10.1002/gepi.20300 PMID: 18200600

12. Murcray CE, Lewinger JP, Gauderman WJ. Gene-Environment Interaction in Genome-Wide Association Studies. Am J Epidemiol. 2009; 169: 219–226. https://doi.org/10.1093/aje/kwn353 PMID: 19022827

13. Paré G, Cook NR, Ridker PM, Chasman DI. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. PLoS Genet. 2010; 6: e1000981. https://doi.org/10.1371/journal.pgen.1000981 PMID: 20585554

14. Zhang P, Lewinger JP, Conti D, Morrison JL, Gauderman WJ. Detecting Gene–Environment Interactions for a Quantitative Trait in a Genome-Wide Association Study. Genet Epidemiol. 2016; 40: 394–403. https://doi.org/10.1002/gepi.21977 PMID: 27230133

15. Jiao S, Hsu L, Bézieau S, Brenner H, Chan AT, Chang-Claude J, et al. SBERIA: Set-Based Gene-Environment Interaction Test for Rare and Common Variants in Complex Diseases. Genet Epidemiol. 2013; 37: 452–464. https://doi.org/10.1002/gepi.21735 PMID: 23720162

16. Liu Q, Chen LS, Nicolae DL, Pierce BL. A unified set-based test with adaptive filtering for gene–environment interaction analyses. Biometrics. 2016; 72: 629–638. https://doi.org/10.1111/biom.12428 PMID: 26496228

17. Lin X, Lee S, Christiani DC, Lin X. Test for interactions between a genetic marker set and environment in generalized linear models. Biostat Oxf Engl. 2013; 14: 667–681. https://doi.org/10.1093/biostatistics/kxt006 PMID: 23462021

18. Tzeng J-Y, Zhang D, Pongpanich M, Smith C, McCarthy MI, Sale MM, et al. Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. Am J Hum Genet. 2011; 89: 277–288. https://doi.org/10.1016/j.ajhg.2011.07.007 PMID: 21835306

19. Zhao G, Marceau R, Zhang D, Tzeng J-Y. Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. Genetics. 2015; 199: 695–710. https://doi.org/10.1534/genetics.114.171686 PMID: 25585620

20. Lin X, Lee S, Wu MC, Wang C, Chen H, Li Z, et al. Test for rare variants by environment interactions in sequencing association studies. Biometrics. 2016; 72: 156–164. https://doi.org/10.1111/biom.12368 PMID: 26229047

21. Su Y-R, Di C-Z, Hsu L, Genetics and Epidemiology of Colorectal Cancer Consortium. A unified powerful set-based test for sequencing data analysis of GxE interactions. Biostat Oxf Engl. 2017; 18: 119–131. https://doi.org/10.1093/biostatistics/kxw034

22. Jiao S, Peters U, Berndt S, Bézieau S, Brenner H, Campbell PT, et al. Powerful Set-Based Gene-Environment Interaction Testing Framework for Complex Diseases. Genet Epidemiol. 2015; 39: 609–618. https://doi.org/10.1002/gepi.21908 PMID: 26095235

23. Kim J, Ziyatdinov A, Laville V, Hu FB, Rimm E, Kraft P, et al. Joint Analysis of Multiple Interaction Parameters in Genetic Association Studies. Genetics. 2019; 211: 483–494. https://doi.org/10.1534/genetics.118.301394 PMID: 30578273

24. Moore R, Casale FP, Bonder MJ, Horta D, Franke L, Barroso I, et al. A linear mixed-model approach to study multivariate gene–environment interactions. Nat Genet. 2019; 51: 180–186. https://doi.org/10.1038/s41588-018-0271-0 PMID: 30478441

25. Ni G, Werf J van der, Zhou X, Hyppönen E, Wray NR, Lee SH. Genotype–covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. Nat Commun. 2019; 10: 1–15. https://doi.org/10.1038/s41467-019-10128-w PMID: 31110177

26. Dahl A, Nguyen K, Cai N, Gandal MJ, Flint J, Zaitlen N. A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits. Am J Hum Genet. 2020; 106: 71–91. https://doi.org/10.1016/j.ajhg.2019.11.015 PMID: 31901249

27. Kerin M, Marchini J. Inferring Gene-by-Environment Interactions with a Bayesian Whole-Genome Regression Model. Am J Hum Genet. 2020; 107: 698–713. https://doi.org/10.1016/j.ajhg.2020.08.009 PMID: 32888427

28. Wang X, Lim E, Liu C-T, Sung YJ, Rao DC, Morrison AC, et al. Efficient gene–environment interaction tests for large biobank-scale sequencing studies. Genet Epidemiol. 2020; 44: 908–923. https://doi.org/10.1002/gepi.22351 PMID: 32864785

29. Hindy G, Wiberg F, Almgren P, Melander O, Orho-Melander M. Polygenic Risk Score for Coronary Heart Disease Modifies the Elevated Risk by Cigarette Smoking for Disease Incidence. Circ Genomic Precis Med. 2018; 11: e001856. https://doi.org/10.1161/CIRCGEN.117.001856 PMID: 29874179

30. Aschard H, Tobin MD, Hancock DB, Skurnik D, Sood A, James A, et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. Int J Epidemiol. 2017; 46: 894–904. https://doi.org/10.1093/ije/dyw318 PMID: 28082375

31. Kim W, Moll M, Qiao D, Hobbs BD, Shrine N, Sakornsakolpat P, et al. Interaction of Cigarette Smoking and Polygenic Risk Score on Reduced Lung Function. JAMA Netw Open. 2021; 4: e2139525. https://doi.org/10.1001/jamanetworkopen.2021.39525 PMID: 34913977

32. Sun R, Carroll RJ, Christiani DC, Lin X. Testing for gene–environment interaction under exposure misspecification. Biometrics. 2018; 74: 653–662. https://doi.org/10.1111/biom.12813 PMID: 29120492

33. Zhang M, Yu Y, Wang S, Salvatore M, Fritsche LG, He Z, et al. Interaction analysis under misspecification of main effects: Some common mistakes and simple solutions. Stat Med. 2020; 39: 1675–1694. https://doi.org/10.1002/sim.8505 PMID: 32101638

34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38: 904–909. https://doi.org/10.1038/ng1847 PMID: 16862161

35. Vansteelandt S, VanderWeele TJ, Tchetgen EJ, Robins JM. Multiply Robust Inference for Statistical Interactions. J Am Stat Assoc. 2008; 103: 1693–1704. https://doi.org/10.1198/016214508000001084 PMID: 21603124

36. Newey WK, Robins JR. Cross-Fitting and Fast Remainder Rates for Semiparametric Estimation. 2018 [cited 16 Sep 2022]. https://doi.org/10.48550/arXiv.1801.09138

37. Breiman L, Friedman JH. Estimating Optimal Transformations for Multiple Regression and Correlation: Rejoinder. J Am Stat Assoc. 1985; 80: 614–619. https://doi.org/10.2307/2288477

38. Dukes O, Vansteelandt S, Whitney D. On doubly robust inference for double machine learning. 2021 [cited 22 Jul 2022]. https://doi.org/10.48550/ARXIV.2107.06124

39. Bertsimas D, King A, Mazumder R. Best subset selection via a modern optimization lens. Ann Stat. 2016; 44: 813–852. https://doi.org/10.1214/15-AOS1388

40. Fawcett KA, Obeidat M, Melbourne C, Shrine N, Guyatt AL, John C, et al. Variants associated with HHIP expression have sex-differential effects on lung function. Wellcome Open Res. 2020; 5: 111. https://doi.org/10.12688/wellcomeopenres.15846.2 PMID: 33728380

41. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic analyses identify widespread sex-differential participation bias. Nat Genet. 2021; 53: 663–671. https://doi.org/10.1038/s41588-021-00846-7 PMID: 33888908

42. Bernabeu E, Canela-Xandri O, Rawlik K, Talenti A, Prendergast J, Tenesa A. Sex differences in genetic architecture in the UK Biobank. Nat Genet. 2021; 53: 1283–1289. https://doi.org/10.1038/s41588-021-00912-0 PMID: 34493869

43. Zhu C, Ming MJ, Cole JM, Kirkpatrick M, Harpak A. Amplification is the Primary Mode of Gene-by-Sex Interaction in Complex Human Traits. bioRxiv; 2022. p. 2022.05.06.490973. https://doi.org/10.1101/2022.05.06.490973

44. Dudbridge F, Fletcher O. Gene-Environment Dependence Creates Spurious Gene-Environment Interaction. Am J Hum Genet. 2014; 95: 301–307. https://doi.org/10.1016/j.ajhg.2014.07.014 PMID: 25152454

45. VanderWeele TJ. On the Distinction Between Interaction and Effect Modification. Epidemiology. 2009; 20: 863–871. https://doi.org/10.1097/EDE.0b013e3181ba333c PMID: 19806059

46. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B Methodol. 1996; 58: 267–288.

47. Tian Y, Feng Y. RaSE: A Variable Screening Framework via Random Subspace Ensembles. J Am Stat Assoc. 2021; 0: 1–12. https://doi.org/10.1080/01621459.2021.1938084