Title

Identifying latent genetic interactions in genome-wide association studies using multiple traits

Permalink

https://escholarship.org/uc/item/44d3p04q

Journal

Genome Medicine, 16(1)

ISSN

1756-994X

Authors

Bass, Andrew J

Bian, Shijia

Wingo, Aliza P

et al.

Publication Date

2024-04-01

DOI

10.1186/s13073-024-01329-0

Peer reviewed

METHOD

Identifying latent genetic interactions in genome-wide association studies using multiple traits

Andrew J. Bass[1*] , Shijia Bian[2], Aliza P. Wingo[3], Thomas S. Wingo[1,4], David J. Cutler[1] and Michael P. Epstein[1*]

Abstract

The "missing" heritability of complex traits may be partly explained by genetic variants interacting with other genes or environments that are difficult to specify, observe, and detect. We propose a new kernel-based method called Latent Interaction Testing (LIT) to screen for genetic interactions that leverages pleiotropy from multiple related traits without requiring the interacting variable to be specified or observed. Using simulated data, we demonstrate that LIT increases power to detect latent genetic interactions compared to univariate methods. We then apply LIT to obesity-related traits in the UK Biobank and detect variants with interactive effects near known obesity-related genes (URL: https://CRAN.R-project.org/package=lit).

Background

There are many large genome-wide association studies (GWAS) available to help facilitate gene discovery and improve our molecular understanding of complex traits and diseases. In particular, recent biobank-sized GWAS collect massive sample sizes and a broad range of phenotypic data to enhance complex trait mapping and augment knowledge of gene functionality. There are two important patterns that have emerged among the multitude of GWAS analyses that have been performed to date. First, the genetic variation of complex traits often involves many thousands of loci [1]. Second, for many traits studied, family-based estimates of heritability (i.e.,

the proportion of trait variance explained by genetic factors) tend to be substantially greater than corresponding heritability estimates from GWAS single nucleotide polymorphisms (SNP) data [2]. For example, the heritability estimates for body mass index (BMI) from GWAS-based studies are $22 - 30\%$ [3–6] compared to $40 - 70\%$ [7, 8] in family-based studies. While this "missing" heritability may be due to small sample sizes, structural variants, and/or rare variants [6, 9–11], these sources may not fully explain the difference in some traits [12]. Another possible explanation is that family-based estimates of broad-sense heritability are capturing within-family sharing of genetic variants with interactive effects (e.g., gene-by-gene and gene-by-environment interactions) which are omitted from GWAS estimates derived from nearly unrelated individuals who only generally share the additive effects of alleles (i.e., narrow-sense heritability) [12–14]. Given the evidence of such interactions [15–18], discovering genetic variants with interactive effects may explain missing heritability and broaden our understanding of the genetic architecture of complex traits.

There are many statistical challenges to discovering genetic variants with interactive effects in GWAS [19]. In particular, studies are typically underpowered to detect

*Correspondence:
Andrew J. Bass
ajbass@emory.edu
Michael P. Epstein
mpepste@emory.edu
[1] Department of Human Genetics, Emory University, Atlanta, GA 30322, USA
[2] Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA
[3] Department of Psychiatry, Emory University, Atlanta, GA 30322, USA
[4] Department of Neurology, Emory University, Atlanta, GA 30322, USA

Bass *et al. Genome Medicine*   (2024) 16:62

Page 2 of 17

interactions due to small effect sizes, a large multiple testing burden, and unknown interactive variables (e.g., other variants or environmental factors). Consequently, it can be difficult to design a study to identify and accurately observe interacting variables. Furthermore, even when interacting variables are known, the mismeasurement of such variables can lead to power loss [20]. A strategy to circumvent some of these issues is to employ a genome-wide screening procedure to identify a reduced set of variants with potential interactive effects that are then subjected to further study using conventional (regression) methods [21]. One such popular screening strategy is to identify variance quantitative trait loci (vQTL) using variance-based testing procedures, which do not require the interactive variable(s) to be observed [17, 22–25]. Intuitively, such procedures model and detect any unequal residual trait variation among genotype categories at a specific SNP (i.e., heteroskedasticity), which can indicate complex non-additive biological signals (including interactions of the SNP with other genes and/or environment). Therefore, a vQTL provides a straightforward screen for a complex biological signal, which can then facilitate the discovery of SNPs with genetic interactions. Previous work has found that variance-based testing procedures can help identify latent genetic interactions on complex traits, including inflammatory markers [23] and obesity-related traits [17, 21, 24, 25].

When there are multiple related traits measured in a study, researchers often apply variance-based procedures on a trait-by-trait (or univariate) basis to screen for evidence of latent genetic interactions. However, a univariate strategy ignores any biological pleiotropy among traits despite theoretical [26] and empirical [27, 28] support for this phenomenon. Furthermore, in the presence of pleiotropy, many studies have demonstrated that joint statistical modeling of related traits (i.e., a multivariate procedure) outperforms univariate procedures for gene mapping [29, 30]. This observation, coupled with the potential existence of pleiotropic interactive effects, suggests that analyzing multiple traits simultaneously in a statistical procedure will increase power to detect latent genetic interactions. To this end, we propose a novel statistical framework, called Latent Interaction Testing (LIT), that leverages multiple related traits to increase the power to screen for latent genetic interactions. LIT is motivated by the observation that latent genetic interactions induce not only a differential variance pattern (i.e., heteroskedasticity) as previously reported but also a differential covariance pattern between traits which is classified as a covariance QTL (covQTL or correlation QTL [31]). We can harness the differential covariance patterns to increase the power to screen for latent genetic

interactions compared to variance-based strategies. Similar to variance-based strategies, LIT does not require the interactive partner(s) to be observed or specified.

The manuscript is outlined as follows. We first introduce the LIT framework for detecting latent genetic interactions and then evaluate the performance using simulated biobank-sized datasets. We also compare LIT to univariate testing procedures and observe that LIT provides significant power gains to detect interactive effects in GWAS. Finally, we demonstrate LIT using four obesity-related traits in the UK Biobank with over 6 million single nucleotide polymorphisms from 330,868 genotyped individuals. Our screening procedure identified multiple loci with potential interaction effects near known obesity genes; many of which were subsequently confirmed using regression procedures.

Methods
Motivation
Consider the trait Y_{jk} for $j = 1, 2, \ldots, n$ unrelated individuals with $k = 1, 2, \ldots, r$ measurable traits. Suppose Y_{jk} depends on a biallelic locus with genotype X_j denoting the number of minor alleles for the jth individual, an unobserved (or latent) variable M_j, and a latent interaction $X_j M_j$. The latent variable M_j can represent any type of interacting partner: a genotype at another locus, sex, age, or an environmental factor. For clarity, we will discuss M_j as an environmental factor and so the latent interaction is a genotype-by-environment (GxE) interaction.

We assume that these components contribute to trait expression additively in the following regression model:

$$Y_{jk} = \beta_k X_j + \phi_k M_j + \gamma_k X_j M_j + \epsilon_{jk}, \tag{1}$$

where β_k is the effect size of the minor allele, ϕ_k is the effect size of the environmental variable, γ_k is the effect size of the GxE interaction, and ϵ_{jk} is an independent and identically distributed random error with mean zero and variance σ_k^2. In this simplified setting, our goal is to detect the latent GxE interaction without observing the interacting variable M_j.

Under the above model assumptions, the latent GxE interaction will induce differential trait variance and covariance patterns that differ by genotype. Without loss of generality, assume the environmental variable has mean zero with unit variance. In Supplementary methods (Additional file 1), we show that the individual-specific trait variance (ITV) of the kth trait conditional on genotype is

$$\mathrm{Var}\left[Y_{jk} | X_j\right] = a_k + b_k X_j + c_k X_j^2, \tag{2}$$

Bass *et al. Genome Medicine*   (2024) 16:62

Page 3 of 17

where $a_k = \phi_k^2 + \sigma_k^2$, $b_k = 2\phi_k\gamma_k$, and $c_k = \gamma_k^2$. We also show that the individual-specific covariance (ITC) between the kth and k'th trait conditional on genotype is

$$\text{Cov}[Y_{jk}, Y_{jk'}|X_j] = \widetilde{a}_{kk'} + \widetilde{b}_{kk'}X_j + \widetilde{c}_{kk'}X_j^2, \quad (3)$$

where $\widetilde{a}_{kk'} = \phi_k\phi_{k'}$, $\widetilde{b}_{kk'} = \phi_k\gamma_{k'} + \phi_{k'}\gamma_k$, and $\widetilde{c}_{kk'} = \gamma_k\gamma_{k'}$. It is evident that a latent GxE interaction in trait k ($\gamma_k \neq 0$) not only induces a variance pattern that depends on genotype (Eq. 2; vQTL) but can also induce a covariance pattern between traits k and k' from either a shared interaction ($\gamma_{k'} \neq 0$) or a shared environment involved in the interaction ($\phi_{k'} \neq 0$; Eq. 3; covQTL). These results suggest that we can test for loci with latent interactive effects by assessing whether the individual-specific trait variances (ITV) and covariances (ITC) differ by genotype without specifying or directly modeling the interacting variable M_j.

Latent Interaction Testing (LIT) framework

Our strategy builds from the above observations and estimates the ITV and ITC to detect non-additive biological signals that can be induced by latent genetic interactions. To derive estimates of these quantities, we first remove the additive genetic effect from the traits to ensure that any variance and covariance effects are not due to the additive effect. Let us denote the trait residuals as $e_{jk} = Y_{jk} - \beta_k X_j$ where we assume the effect size is known for simplicity. We can then express the ITV and ITC as a function of these residuals: the ITV of trait k and the ITC between traits k and k' is defined as $\text{Var}[Y_{jk}|X_j] = E[e_{jk}^2|X_j]$ and $\text{Cov}[Y_{jk}, Y_{jk'}|X_j] = E[e_{jk}e_{jk'}|X_j]$, respectively (Additional file 1). Thus, we can estimate the ITV by squaring the residuals, e_{jk}^2, and estimate the ITC between traits k and k' by the pairwise product of the residuals (i.e., the cross products), $e_{jk}e_{jk'}$. Aggregating the ITV and ITC estimates across all individuals, we denote the cross product (CP) terms in the $n \times s$ matrix $\mathbf{Z}^{\text{CP}}$ where the jth row vector is $\mathbf{Z}_j^{\text{CP}} = [e_{j1}e_{j2}, e_{j1}e_{j3}, \ldots, e_{j,r-1}e_{jr}]$, and the squared residual (SQ) terms in the $n \times r$ matrix $\mathbf{Z}^{\text{SQ}}$ where $Z_{jk}^{\text{SQ}} = e_{jk}^2$.

Our inference goal is to assess whether the SNP, $\mathbf{X}_{n\times1} = [X_1, X_2, \ldots, X_n]^T$, is independent of the squared residuals and cross products,

$$\begin{aligned} \mathbf{Z}_{\cdot q}^{\text{CP}} \perp\!\!\!\perp \mathbf{X} & \quad \text{for} \quad q = 1, 2, \ldots, s, \quad \text{and} \\ \mathbf{Z}_{\cdot k}^{\text{SQ}} \perp\!\!\!\perp \mathbf{X} & \quad \text{for} \quad k = 1, 2, \ldots, r, \end{aligned} \quad (4)$$

where "·" denotes all the rows (or individuals) and "$\perp\!\!\!\perp$" denotes statistical independence. In the above regression model, this corresponds to testing the global null hypothesis $H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_r = 0$ versus the alternative hypothesis $H_1 : \gamma_k \neq 0$ for at least one of the $k = 1, 2, \ldots, r$ traits. While a regression model can be directly applied to the squared residuals and cross products to test the global null hypothesis (see Additional file 1 for mathematical details), a univariate model approach does not adequately leverage pleiotropy and requires a multiple testing correction which reduces power.

To address these issues, we develop a new multivariate kernel-based framework, Latent Interaction Testing (LIT), that captures pleiotropy across the ITV and ITC terms to increase power for detecting latent interactions. There are three key steps in the LIT framework (Fig. 1):

1. Regress out the additive genetic effects and any other covariates from the traits. Additionally, adjust the traits and genotypes for population structure.
2. Calculate estimates of the ITV and ITC for each individual using the squared residuals and the cross products of the residuals, respectively.
3. Test the global null hypothesis of no latent interaction by comparing the adjusted genotype(s) to the ITV and ITC estimates.

We expand on the above steps in detail below.

Step 1: In the first step, LIT standardizes the traits and then regresses out the additive genetic effects, population structure, and any other covariates. This ensures that any differential variance and/or covariance patterns are not due to additive genetic effects or population structure. Suppose there are l_1 measured covariates and l_2 principal components to control for structure. We denote these $l = l_1 + l_2$ variables in the $n \times l$ matrix $\mathbf{H}$. After regressing out these variables and the additive genetic effects, the $n \times r$ matrix of residuals is $\mathbf{e} = \widetilde{\mathbf{Y}} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{H}\widehat{\mathbf{A}}$, where $\widetilde{\mathbf{Y}}$ is the standardized trait matrix, $\widehat{\boldsymbol{\beta}}$ is a $1 \times r$ matrix of effect sizes, and $\widehat{\mathbf{A}}$ is a $l \times r$ matrix of coefficients estimated using least squares. We also regress out population structure from the genotypes which we denote by $\widetilde{\mathbf{X}}$.

The above approach only removes the mean effects and does not correct for variance effects from population structure which can impact type I error rate control [32]. A strategy to adjust for the variance effects is to standardize the genotypes with the estimated individual-specific allele frequencies (IAF), i.e., the allele frequencies given the genetic ancestry of an individual. However, it is computationally costly to standardize the genotypes for biobank-sized datasets as it requires estimating the IAFs of all SNPs using a generalized linear model [33, 34]. Therefore, in this work, we remove the mean effects from structure and then adjust the test statistics with the genomic inflation factor to be conservative. Our software includes an implementation to standardize the genotypes using the IAFs for smaller datasets.
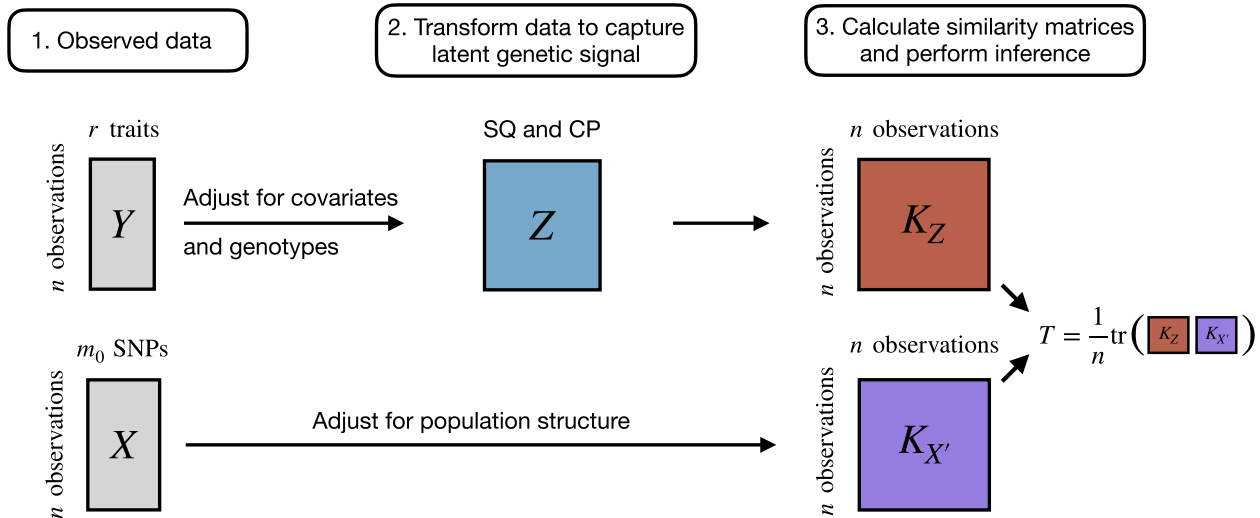
Bass *et al. Genome Medicine* (2024) 16:62

Page 4 of 17

Fig. 1 Overview of the Latent Interaction Testing (LIT) framework. Given a set of *r* traits, **Y**, and m_0 SNPs, **X**, the goal is to detect a latent genetic interaction involving the SNPs. The trait squared residuals (SQ) and cross products (CP), **Z**, are calculated while adjusting for linear effects from the genotypes and any other covariates. The traits and genotypes are also adjusted for population structure. A similarity matrix for the genotypes, $K_{X'}$, and the SQ and CP, K_Z, are calculated to construct a test statistic, *T*, which measures the overlap between the two matrices. Large values of *T* are evidence of a latent genetic interaction

Step 2: The second step uses the residuals, **e**, to reveal any non-additive biological signals by constructing estimates of the ITV and ITC. For the *j*th individual's set of trait residuals, the ITVs are estimated by squaring the trait residuals while the ITCs are estimated by calculating the cross products of the trait residuals. We express the squared residuals as $Z_j^{\text{SQ}} = \left[e_{j1}^2, e_{j2}^2, \ldots, e_{jr}^2 \right]$, and the $s = \binom{r}{2}$ pairwise cross products as $Z_j^{\text{CP}} = \left[e_{j1}e_{j2}, e_{j1}e_{j3}, \ldots, e_{j,k-1}e_{jr} \right]$. Importantly, when the studentized residuals are used, then Z_j^{SQ} and Z_j^{CP} represent an unbiased estimate of the ITVs and ITCs, respectively. We aggregate these terms across all individuals into the $n \times (r + s)$ matrix $Z = \left[Z^{\text{SQ}} \ Z^{\text{CP}} \right]$.

Step 3: In the last step, we test for association between the adjusted SNP and the squared residuals and cross products (SQ/CP) using a kernel-based distance covariance framework [35–37]. Specifically, we apply a kernel-based independence test called the Hilbert-Schmidt independence criterion (HSIC), which has been previously used for GWAS data (see, e.g., [38–41]). The HSIC generalizes many well-known testing procedures in statistics; for example, depending on the kernel function choice, the RV coefficient [42], distance covariance [43], and multivariate distance matrix regression (MDMR) [44] can be expressed as special cases of the HSIC. Due to such flexibility, it is implemented in many testing procedures in genetics (e.g., SKAT [40]). The HSIC constructs two $n \times n$ similarity matrices between individuals using the SQ/CP matrix and genotype matrix, then calculates a test statistic that measures any shared signal

between these similarity matrices. To estimate the similarity matrix, a kernel function is specified that captures the similitude between the *j*th and *j*'th individual.

Since our primary application is biobank-sized data, we use a linear kernel so that LIT is computationally efficient. The linear similarity matrix is defined as $K_{jj'} := k\left(\widetilde{X}_j, \widetilde{X}_{j'} \right) = \widetilde{X}_j \widetilde{X}_{j'}$ for the genotype matrix and $L_{jj'} := k\left(Z_j, Z_{j'} \right) = Z_j Z_{j'}^T$ for the SQ/CP matrix. The linear kernel is a scaled version of the covariance matrix and, for this special case, the HSIC is related to the RV coefficient. While our theoretical results indicate that the variance (Eq. 2) and covariance (Eq. 3) models include a quadratic term for the genotypes, the expected effect size of an interaction in a GWAS suggests that the linear term will dominant the variance and/or covariance signal compared to the quadratic term. Therefore, we only consider the linear term in this work. We note that one can choose other options for a kernel function, such as a polynomial kernel, projection kernel, and a Gaussian radial-basis function that can capture nonlinear relationships [41, 45].

Once the similarity matrices **K** and **L** are constructed, we can express the HSIC test statistic as

$$T = \frac{1}{n}\text{tr}(KL), \tag{5}$$

which follows a weighted sum of chi-squared random variables under the null hypothesis, i.e., $T \mid H_0 \sim \sum_{i,j}^n \frac{1}{n} \lambda_{K,i} \lambda_{L,j} v_{ij}^2$, where $\lambda_{K,i}$ and $\lambda_{L,j}$ are the

ordered non-zero eigenvalues of the respective matrices and $v_{ij} \sim \text{Normal}(0, 1)$. Intuitively, the test statistic measures the "overlap" between two random matrices where large values of T imply the two matrices are similar (i.e., a latent genetic interactive effect) while small values of T imply no evidence of similarity (i.e., no latent genetic interactive effects). We can approximate the null distribution of T using Davies' method, which is computationally fast and accurate for large T [40, 41, 46].

For the linear kernel considered here, we implement a simple strategy to substantially improve the computational speed of LIT. We first calculate the eigenvectors and eigenvalues of the SQ/CP and genotype matrices to construct the test statistic. Since the number of traits, r, is much smaller than the sample size, n, we can perform a singular value decomposition to estimate the subset of eigenvectors and eigenvalues in a computationally efficient manner [47–49]. This allows us to circumvent direct calculation and storage of large $n \times n$ similarity matrices. Let $\boldsymbol{L} = \boldsymbol{V}_L \boldsymbol{D}_L \boldsymbol{V}_L^T$ and $\boldsymbol{K} = \boldsymbol{V}_K \boldsymbol{D}_K \boldsymbol{V}_K^T$ be the singular value decomposition (SVD) of the similarity matrices where the matrix $\boldsymbol{D}$ is a diagonal matrix of eigenvalues and $\boldsymbol{V}$ is a matrix of eigenvectors of the respective kernel matrices. We can then express the test statistic in terms of the SVD components as $T = \frac{1}{n} \text{tr}(\boldsymbol{D}_K \boldsymbol{R} \boldsymbol{D}_L \boldsymbol{R}^T)$, where $\boldsymbol{R} = \boldsymbol{V}_K^T \boldsymbol{V}_L$ is the outer product between the two eigenvectors. Thus, for a single SNP, the test statistic is $T = \frac{1}{n} \text{tr}\left(\boldsymbol{D}_K \boldsymbol{R}_{d_1 \times d_2} \boldsymbol{D}_L \boldsymbol{R}_{d_2 \times d_1}^T\right)$, where $d_1 = r + s$ is the rank of the SQ/CP matrix, $d_2 = 1$ is the rank of the genotype matrix, and $d_1, d_2 \ll n$.

Aggregating different LIT implementations using the Cauchy combination test

We explore an important aspect of the test statistic in Eq. 5, namely, the role of the eigenvalues in determining statistical significance. The above equations suggest that the eigenvalues of the kernel matrices are emphasizing the eigenvectors that explain the most variation in the test statistic. While this may be reasonable in some settings, the interaction signal can be captured by eigenvectors that explain the least variation and this can be very difficult to ascertain beforehand [50]. In this case, the testing procedure will be underpowered. Thus, we also consider weighting the eigenvectors equally in LIT, i.e., $T = \frac{1}{n} \text{tr}(\boldsymbol{R} \boldsymbol{R}^T) = \frac{1}{n} \sum_{i=1}^{n} D_{R,i}^2$, where D_R are the eigenvalues of the outer product matrix. In this work, we implement a linear kernel (scaled covariance matrix) and so, in this special case, weighting the eigenvectors equally is equivalent to the projection kernel.

In summary, there are two implementations of the LIT framework. The residuals are first transformed to calculate the SQ and CP to reveal any latent interactive effects.

We then calculate the weighted and unweighted eigenvectors in the test statistic which we refer to as weighted LIT (wLIT) and unweighted LIT (uLIT), respectively. We also apply a Cauchy combination test (CCT) [51] to combine the p-values from the LIT implementations to maximize the number of discoveries and hedge for various (unknown) settings where one implementation may outperform the other. More specifically, let p_c denote the p-value for the $c = 1, 2$ implementations. In this case, the CCT statistic is $T' = \frac{1}{2} \sum_{c=1}^{2} \tan\{(0.5 - p_c)\pi\}$, where $\pi \approx 3.14$ is a mathematical constant. A corresponding p-value is then calculated using the standard Cauchy distribution. Importantly, when applying genome-wide significance levels, the CCT controls the type I error rate under arbitrary dependence structures. We refer to the CCT p-value as aggregate LIT (aLIT).

Incorporating multiple loci in LIT

We can extend LIT to assess latent interactions within a genetic region (e.g., a gene) consisting of multiple SNPs. In the first step, we regress out the joint additive effects from the multiple SNPs along with any other covariates and population structure. In the second step, we calculate the squared residuals and cross products using the corresponding residual matrix. Finally, in the last step, we construct the similarity matrices and perform inference using the HSIC: the linear similarity matrix for the $n \times m_0$ genotype matrix $\widetilde{\boldsymbol{X}}$ is $K_{jj'} = k\left(\widetilde{\boldsymbol{X}}_j, \widetilde{\boldsymbol{X}}_{j'}\right) = \widetilde{\boldsymbol{X}}_j \widetilde{\boldsymbol{X}}_{j'}^T$ and our test statistic is $T = \frac{1}{n} \text{tr}\left(\boldsymbol{D}_K \boldsymbol{R}_{d_1 \times d_2} \boldsymbol{D}_L \boldsymbol{R}_{d_2 \times d_1}^T\right)$ where $d_2 = m_0$ is the rank of the genotype matrix.

Compared to the previous section, this extended version of LIT is a region-based test for interactive effects instead of a SNP-by-SNP test. A region-based test is advantageous to reduce the number of tests compared to a SNP-by-SNP approach and enable testing of rare variants [41]. However, in this work, we perform a SNP-by-SNP genome-wide scan with LIT to demonstrate the scalability.

Simulation study

We evaluated the performance of LIT using simulated data with the following assumptions. Let the individual-specific minor allele frequencies of $t = 1, 2, \ldots, m$ biallelic genotypes be denoted by π_{jt}. Of the m SNPs, $m - 1$ SNPs had no interacting partner and a minor allele frequency drawn from a $\text{Uniform}(0.1, 0.4)$. The SNP with an interacting partner had a minor allele frequency of 0.25. We fixed this MAF to remove stochastic variation in the observed power induced by simulations differing only by the MAF of the interacting SNP. The genotypes were then drawn from a Binomial distribution with parameter π_{jt}, i.e., $X_{jt} \sim Binomial(2, \pi_{jt})$. In total, there were

Bass *et al. Genome Medicine* (2024) 16:62

Page 6 of 17

$n = 300{,}000$ individuals simulated to reflect biobank-sized GWAS.

We simulated the trait expression value Y_{jk} for $k = 1, 2, \ldots, r$ traits under the polygenic trait model with two risk environmental variables M_j and W_j. Specifically, there were $r = 5, 10$ traits and $m = 100$ genotypes simulated with an additive genetic, environmental, and GxE components:

$$Y_{jk} = \alpha_k + \beta_{1k}X_{j1} + \phi_k M_j + \gamma_k M_j X_{j1} + \sum_{t=2}^{m} \beta_{tk}X_{jt} + W_j + \epsilon_{jk},$$

(6)

where the intercept, α_k, follows a normal distribution with a standard deviation of 5; the effect sizes of the GxE interaction, γ_k, interacting environment, ϕ_k, and additive genetic component, β_{tk}, follow a normal distribution with mean zero and standard deviation of 0.01; the two environmental variables were generated from a standard normal distribution where only one interacts with the risk allele; and the error term was generated from a standard normal distribution. Using the above model, we considered different types of pleiotropy. First, we assigned the effect size direction of the additive genetic component, interacting environment, and the GxE interaction to be the same in each trait. We then considered cases where the effect size for the shared GxE interaction is in the same direction (i.e., $|\gamma_k|$) and random directions across traits. These settings represent positive pleiotropy and a mixture of positive and negative pleiotropy, respectively. We also considered a variation of the above settings where the direction of the effect size for the GxE interaction is opposite of the interacting environment.

We transformed the components in the model using the function $f(x) = \dfrac{x - \hat{\mu}_x}{\hat{\sigma}_x}$, which takes a vector x and standardizes it by the estimated mean and standard deviation. We scaled each component to set the baseline correlation between traits (ignoring the risk factor, interactive environment, and GxE interaction) as 0.25, 0.50, and 0.75. In particular, the percent variance explained of the non-interactive environment was 15% and the additive genetic component (minus the risk factor) was 10%, 35%, and 60%, which represents a 0.25, 0.50, and 0.75 baseline correlation between traits, respectively. We then assigned the percent variance explained for the additive genetic risk factor as 0.2%, the interactive environment as a uniformly drawn value from 0.5 to 2.0%, the GxE interaction as a uniformly drawn value from 0.1 to 0.15%, and the remaining variation as noise.

In our simulation study, we also varied the proportion of traits with an interaction term. For r traits, let τ_r denote the proportion of traits with a shared GxE interaction signal. We varied this proportion as $\tau_r = \frac{1}{r}, \frac{2}{r}, \ldots, 1$. At each combination of baseline trait correlation, number of traits, and proportion of null traits, we generated data from the above polygenic trait model 500 times for each pleiotropy setting. We calculated the empirical power by averaging the total number of times the p-values were below a significance threshold of $\alpha = 5 \times 10^{-8}$. Under the null hypothesis of no GxE interaction, we assessed the type I error rate at $\alpha = 1 \times 10^{-3}$ using 50 simulated datasets with 10,000 SNPs where the traits do not have a GxE interaction. We also considered cases where the random error follows a Chi-squared distribution with five degrees of freedom and a t-distribution with three degrees of freedom under the null hypothesis.

UK Biobank

The UK Biobank is a collaborative research effort to gather environmental and genetic information from half a million volunteers 40–69 years old in the UK. The data was collected across 22 assessment centers from 2006 to 2010 where participants were given a general lifestyle and health questionnaire, a physical examination, and a blood test that provided genetic data [52, 53]. See ref. [54, 55] for detailed information on the study design.

We applied LIT to four obesity-related traits, namely, waist circumference, hip circumference, body mass index, and body fat percentage. We restricted our analysis to unrelated individuals with British ancestry and removed any individuals with a sex chromosome aneuploidy. Using the imputed genotypes (autosomes only), SNPs were filtered in PLINK [56] with the following thresholds: a MAF of >0.05, a genotype missingness rate of <0.05, Hardy-Weinberg equilibrium (defined as $>10^{-5}$), and an INFO score of >0.9. The traits were adjusted for age and the top 20 principal components provided by the UK Biobank to account for ancestry. We removed individuals with measurements that were four standard deviations above the average and then standardized the traits by sex. After filtering, there were 329,146 individuals and 6,186,503 SNPs in our analysis. To calculate the genomic inflation factor of our LIT analyses, we identified a subset of 34,643 "independent" SNPs using the argument —index-pairwise 500 5 0.05 in PLINK.

To demonstrate our procedure controls the type I error rate using the UK Biobank data, we implemented the double Kolmogorov-Smirnov (KS) testing framework as proposed by Leek and Storey (2011) [57] as a diagnostic tool. The double KS test is implemented as follows. For 100 permutations of the phenotypes, we apply the LIT framework to calculate the p-values for all of the "independent" SNPs (used to estimate the genomic inflation factor) under the null hypothesis. We then evaluate the joint behavior of the set of p-values within each permuted

dataset by applying a KS test to assess how "close" the observed *p*-values are to a uniform distribution. Under the null hypothesis for a well-behaved testing procedure, the *p*-values of this first KS test follow a uniform distribution across the 100 permutations. The next step is to implement a second KS test on the 100 KS test *p*-values to test how "close" the set of 100 KS test *p*-values are to a uniform distribution. Under the null hypothesis, the joint behavior of the KS test *p*-values across the permutations will also follow a uniform distribution. The double KS test simultaneously assesses the joint behavior of the *p*-values within each permutation and the marginal *p*-values across permutations.

Results

Overview

As illustrated in Fig. S1 (Additional file 1), a SNP with a latent interaction induces a genotype effect on the trait variances (vQTL) and on the covariance between traits (covQTL). We can assess this interaction effect by relating an individual's genotype to individual-specific trait variances and covariances (Additional file 1: Fig. S2). We estimate individual-specific trait variances using squared residuals (SQ) for each trait after adjusting for additive (and possibly dominance) effects and likewise estimate individual-specific covariances by multiplying the residuals of different pairs of traits together to form cross products (CP; see ref. [31]). Using a kernel-based distance covariance (KDC) statistic (Fig. 1) [35–37], we then assess evidence of a latent genetic interaction by testing whether the elements of a matrix comprised of pairwise similarity of SQ/CP terms in the sample is independent of the elements of a second matrix comprised of pairwise genotype similarity. To measure the similarity between variables, we apply a user-defined kernel function such as a linear kernel (analogous to scaled covariance) or a projection kernel [41, 45]. We show later that the optimal kernel choice depends on the complexity of the interaction signal. Researchers have previously applied variations of the KDC statistic, which yields a *p*-value testing the global null of no association between the elements of two matrices, in genetic analyses for studies of both common [38, 39, 45] and rare [40, 41, 58] variation.

The traditional KDC statistic utilizes the corresponding eigenvectors (directions of maximal variation) and eigenvalues (weights emphasizing eigenvectors) derived from the SQ/CP similarity matrix for inference. In the process, the traditional KDC statistic emphasizes signals explaining the most variation in this matrix. While we show this emphasis is suitable under certain pleiotropy settings, there are other settings where the interaction signal is not captured by the top eigenvectors of the similarity matrix and so the test may not be optimal [30, 50].

Therefore, we also consider weighting eigenvectors equally in our test statistic to increase power to detect interaction signals captured by the lower eigenvectors of the similarity matrix. We refer to the implementation that weights eigenvectors by corresponding eigenvalues (i.e., the traditional KDC framework) as weighted LIT (wLIT) and refer to the implementation that weights eigenvectors equally as unweighted LIT (uLIT). Since the pleiotropic genetic architecture of a trait is unknown a priori, we maximize the performance of LIT by aggregating the *p*-values from wLIT and uLIT using the Cauchy combination test (CCT) [51], which has proven valuable in a variety of genetic settings [59]. We refer to the CCT of the wLIT and uLIT *p*-values as aggregate LIT (aLIT). For simplicity, we primarily focus on implementing wLIT, uLIT, and aLIT on a SNP-by-SNP basis to test for interactive effects but discuss extensions to handling multiple SNPs simultaneously within the "Methods" section (see the "Incorporating multiple loci in LIT" section). Finally, to improve computational efficiency for biobank-sized datasets, we apply a linear kernel in wLIT and so uLIT is equivalent to using a projection kernel.

Power and type I error rate control

We simulated $r = 5, 10$ related traits for 300,000 observations (reflecting sample sizes for biobank datasets) under the polygenic trait model with additive genetic, environmental, and GxE interaction components. The baseline correlation between traits was either 0.25, 0.50, or 0.75 which represents different correlation strengths from shared genetic and environmental effects. We then simulated a genetic risk factor, an environmental factor, and a GxE interaction that explains 0.2%, a randomly drawn value from 0.5 to 2.0%, and a randomly drawn value from 0.1 to 0.15% of the trait variation, respectively. To assess the performance of LIT under different sparsity settings, we varied the proportion of traits with a shared GxE interaction as $\frac{1}{r}, \frac{2}{r}, \ldots, 1$. We considered three types of pleiotropy in our study, namely, the GxE interaction effect size is positive across traits (positive pleiotropy), a mixture of positive and negative across traits (positive and negative pleiotropy), and a variation of these two settings where the direction is opposite of the interacting environment.

We found that the LIT implementations provide type I error rate control at significance level 10^{-3}, including when the trait distribution is skewed or heavy trailed (Additional file 1: Figs. S3, S4). We also found that LIT controls the type I error rate when applied to multiple SNPs (5 total; Additional file 1: Fig. S5). We then compared the power across various configurations of number of traits, baseline correlation, proportion of traits with shared interaction effects, and direction of the interaction effect (Fig. 2A, B). In comparing wLIT with uLIT,
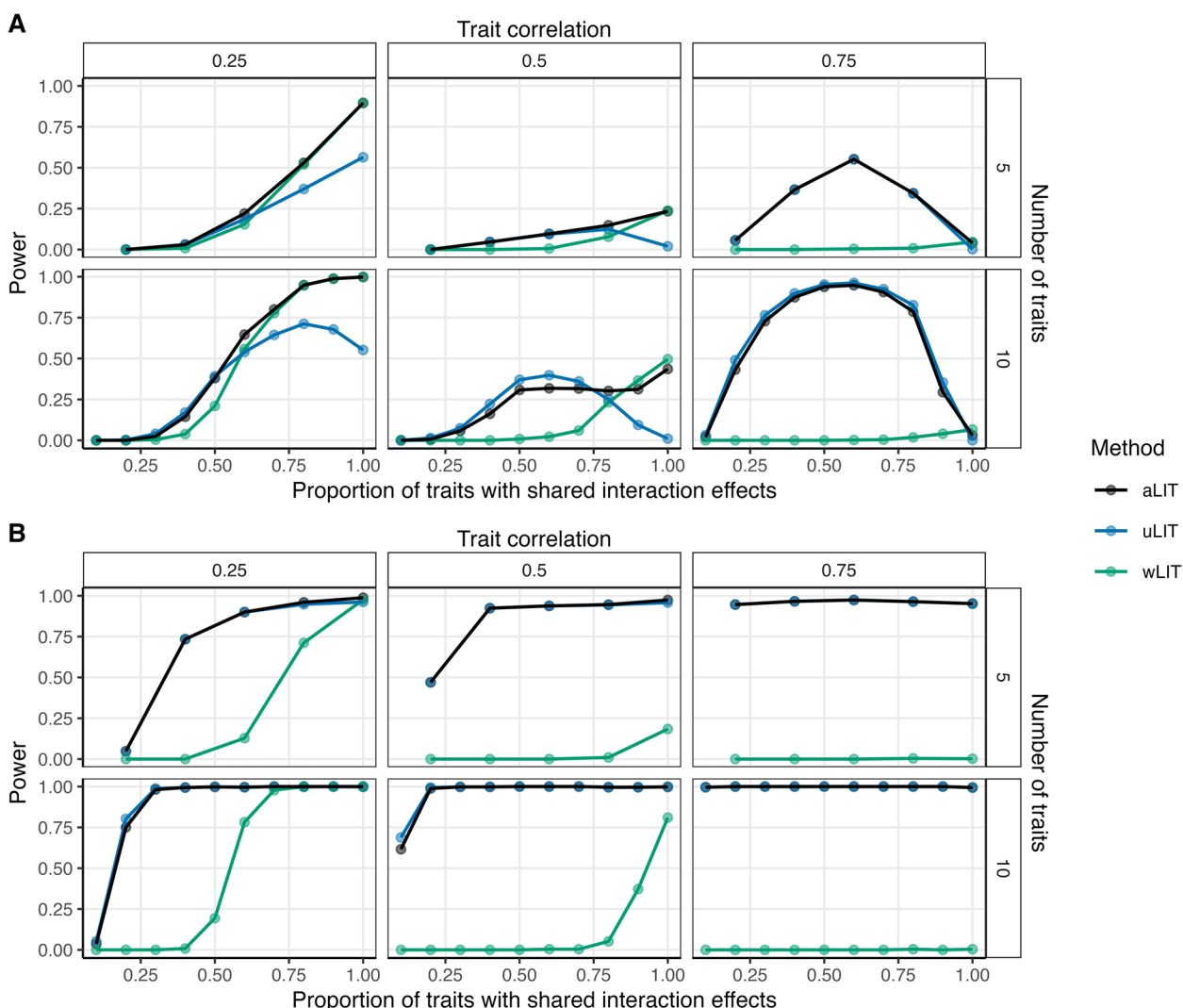
Bass *et al. Genome Medicine*   (2024) 16:62

Page 8 of 17

Fig. 2 Power comparisons of aggregate LIT (aLIT; black), unweighted LIT (uLIT; blue), and weighted LIT (wLIT; green) under **A** positive pleiotropy and **B** a mixture of positive and negative pleiotropy. The simulation study varied the correlation between traits (columns), the number of traits (rows), and the proportion of traits with an interaction term (*x*-axis) at a sample size of 300,000. The points represent the average across 500 simulations with a significance threshold of 5×10^{-8}

neither method is optimal across all settings as expected. As mentioned in the "Overview" section, wLIT emphasizes the high-variance (i.e., large eigenvalues) eigenvectors of the SQ/CP kernel matrix while uLIT weights them equally. Under a simulation model where the signal would reside on the top eigenvector, we expect wLIT to outperform uLIT. Conversely, we expect uLIT to outperform wLIT when the signal resides on the lower-variance eigenvectors of the SQ/CP kernel matrix.

To illustrate how the interaction signal can reside on different eigenvectors of the SQ/CP kernel matrix, we performed an association test between the eigenvectors and genotype under the positive pleiotropy setting

with 10 traits (Additional file 1: Fig. S6). We find that the power to detect the latent interaction signal at each eigenvector depends on the proportion of traits with shared interaction effects (sparsity level), baseline trait correlation, and the proportion of variation explained by the genotype (denoted as R^2). More specifically, for small baseline correlations, the high-variance eigenvector generally captures the signal for most sparsity settings (which explains why wLIT outperforms uLIT in these situations). As the baseline correlation increases, the power of the high-variance eigenvector can decrease rapidly (even if the proportion of traits with shared interaction effects is high) due to the reduction in R^2 (see top

right panels of Additional file 1: Fig. S6). On the other hand, an increase in baseline correlation coupled with a decrease in the proportion of traits with shared interaction effects (i.e., increase in sparsity), results in the interaction signal being separated out from the high-variance eigenvector and becoming detectable in the low-variance eigenvectors.

Given the above insights, we can delineate the performance between uLIT and wLIT by assessing which eigenvectors capture the interaction signal. In the positive pleiotropy setting with 10 traits and a baseline correlation of 0.5 (bottom center panel of Fig. 2A), as the proportion of traits with shared interaction effects increases from 0 to 0.6, the power of uLIT increases whereas the power of wLIT is constantly negligible. When the proportion of traits with shared interaction effects increases from 0.6 to 1, the power of uLIT decreases whereas the power of wLIT increases and overtakes uLIT when this proportion exceeds 0.8. These power trends are due to the low-variance eigenvectors capturing the interaction signal when the proportion of traits with shared interaction effects is small (which favors uLIT) to the high-variance eigenvectors when this is high (which favors wLIT; Additional file 1: Fig. S6). Furthermore, when the baseline correlation increases from 0.5 to 0.75 (bottom right panel of Fig. 2A), uLIT follows a similar power curve while the power of wLIT now remains negligible across all sparsity settings. In this case, the R^2 is low in the high-variance eigenvectors when the baseline correlation is high (Additional file 1: Fig. S6) and so wLIT has little power in these situations. In general, we find that wLIT tends to outperform uLIT when the baseline correlation is modest (i.e., 0.25) and the proportion of traits with shared interaction effects is high, otherwise uLIT is the optimal method.

In the setting where there is a mixture of positive and negative pleiotropy, uLIT outperforms wLIT across all settings (Fig. 2B). Intuitively, in our simulations, the high-variance eigenvector is the weighted sum of the squared residuals and cross products where the weights have the same sign. When the effect sizes are in different directions (positive and negative pleiotropy), the high-variance eigenvector may dampen the interaction signal, and thus it will also be captured by the low-variance eigenvectors. Since uLIT weights the eigenvectors equally, we observe a large increase in power compared to wLIT. We also considered a variation of the above two pleiotropy scenarios where the effect size for the GxE interaction is opposite of the interacting environment. While we find similar results, the overall power is reduced for all methods (Additional file 1: Fig. S7).

In summary, even though the top eigenvectors explain the largest amount of variation, it does not imply that they are the ones most correlated to genotype. The interaction signal may be captured by the high-variance eigenvectors or the low-variance eigenvectors depending on the number of traits, baseline correlation, R^2 at each eigenvector, proportion of traits with shared interaction effects, and type of pleiotropy. Since the particular eigenvectors that are most powerful can vary widely and are unknown a priori, we applied aLIT to the *p*-values from the above LIT implementations to maximize the number of discoveries. We find that aLIT controls the type I error rate (Additional file 1: Figs. S3-S4) while making more discoveries than each individual implementation (Fig. 2; Additional file 1: Fig. S7). More specifically, aLIT has similar power to wLIT when the signal is captured by the high-variance eigenvectors and similar power to uLIT when the signal is captured by the low-variance eigenvectors. Therefore, we implement aLIT in subsequent analyses.

aLIT increases power compared to marginal testing procedures

Using the same simulation configuration as in the "Power and type I error rate control" section, we considered two competing procedures for identifying latent genetic interactions using multiple traits. The first procedure performs an association test between the squared residuals and a SNP (Marginal (SQ)), while the second procedure additionally includes the cross product terms (Marginal (SQ/CP)). More specifically, Marginal (SQ) tests the squared residuals for all *r* traits and selects the minimum *p*-value from these *r* different tests. We note that Marginal (SQ) outperformed Levene's test which has been previously applied to detect vQTLs (Additional file 1: Figs. S8, S9) [23, 24]. Marginal (SQ/CP) adds tests for the $\binom{r}{2}$ cross products and selects the minimum *p*-value from the $r + \binom{r}{2} = \binom{r+1}{2}$ individual tests. Because we are testing the global null hypothesis of no latent genetic interaction, the marginal testing procedures require a Bonferroni correction for the total number of tests, i.e., $\alpha' := \alpha/K$ where α is the significance threshold and K is chosen to be the number of principal components that explains 95% of the variation. We then threshold the minimum *p*-value by α' to determine statistical significance. Across all power simulations (Fig. 3; Additional file 1: Fig. S9), we observed that Marginal (SQ/CP) was more powerful than Marginal (SQ), suggesting that the inclusion of cross products improves performance to detect latent interactions. Given these findings, we compare the performance of aLIT to Marginal (SQ/CP) for the remainder of this work.

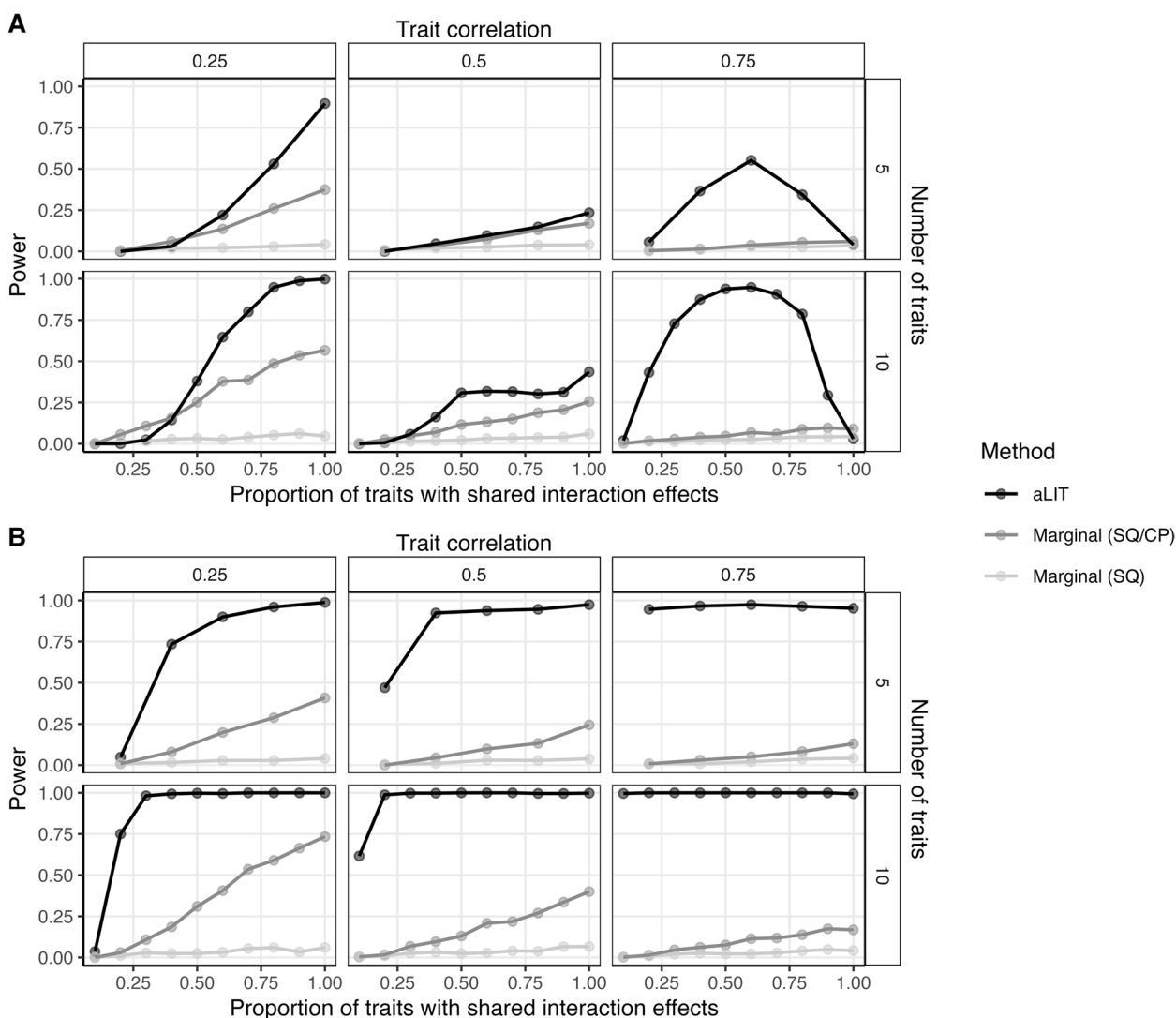In the positive pleiotropy setting with a low baseline correlation, aLIT increases the power to detect

Bass *et al. Genome Medicine*   (2024) 16:62

Page 10 of 17

Fig. 3 A comparison of aLIT (black) to a marginal testing procedure using the squared residuals (SQ; light gray) and a marginal testing procedure using the squared residuals and cross products (SQ/CP; dark gray) under **A** positive pleiotropy and **B** a mixture of positive and negative pleiotropy. The simulation study is identical to Fig. 2 where the empirical power is calculated as a function of the proportion of traits with an interaction term (*x*-axis), the number of traits (rows), and trait correlation (columns)

GxE interactions when there are a higher proportion of traits with shared interaction effects compared to Marginal (SQ/CP) (Fig. 3A). Furthermore, the difference is more pronounced as the number of traits increases. For example, when the baseline correlation is 0.25, and the proportion of traits with shared interaction effects is 0.8, the empirical power of aLIT is 53% and 94.8% for five and ten traits, respectively. On the other hand, the empirical power of Marginal (SQ/CP) is 26% and 48.6%, respectively. While aLIT provides substantial increases in power when the proportion of traits with shared interaction effects is high, Marginal (SQ/CP) can outperform aLIT when the proportion of

traits with shared interaction effects is low in the positive pleiotropy setting (see lower left panel of Fig. 3A). Intuitively, when there is little correlation between traits due to shared interactions, selecting the minimum *p*-value across traits slightly outperforms combining information between the traits.

The difference in power between aLIT and Marginal (SQ/CP) is also evident across baseline correlations. Interestingly, the improvement in power of aLIT compared to Marginal (SQ/CP) reduces when the baseline correlation increases from 0.25 to 0.50. This observation agrees with our simulation results from Fig. S6 (Additional file 1) which suggest that the

Bass *et al. Genome Medicine*   (2024) 16:62

Page 11 of 17

power to detect an interaction signal at any particular eigenvector decreases as the baseline correlation increases from 0.25 to 0.50. Alternatively, when the baseline correlation increases to 0.75, aLIT provides drastic increases in power for most sparsity settings. For example, when there are 10 traits where the proportion of traits with shared interaction effects is 0.5, aLIT's power increases from 30.8 to 93.8% at a baseline correlation of 0.50 and 0.75 while Marginal (SQ/CP) decreases from 11.6 to 4.6%, respectively. However, in this same example, Marginal (SQ/CP) slightly outperforms aLIT when all traits have an interaction.

Overall, while our results suggest that aLIT outperforms Marginal (SQ/CP) for most baseline correlations and sparsity settings under positive pleiotropy, there are some rare cases where Marginal (SQ/CP) has similar (or improved) performance. Meanwhile, under the simulation setting where there is a mixture of positive and negative pleiotropy (Fig. 3B) or the direction of GxE effect sizes are opposite of the interactive environment (Additional file 1: Fig. S9), the increase in power from aLIT over Marginal (SQ/CP) is substantial across all settings. Note that aLIT outperformed Marginal (SQ) across all power simulations.

aLIT applied to the UK Biobank data

We applied the LIT framework to screen for shared latent genetic interactions in four obesity-related traits from the UK Biobank, namely, waist circumference (WC), hip circumference (HC), body mass index (BMI), and body fat percentage (BFP). After preprocessing, there were 329,146 unrelated individuals that have measurements for all traits and 6,186,503 SNPs (the "UK Biobank" section). The correlation between traits ranged from 0.75 (BMI and BFP) to 0.87 (BMI and WC). The total computational time of LIT was approximately 3.3 days using 12 cores of a 3.2-GHz Intel Xeon W-3245 processor.

In each implementation of LIT, after filtering for LD and significant SNPs, the genomic inflation factor of wLIT and uLIT was 1.14 (Additional file 1: Fig. S10a). Additionally, we also observed inflated test statistics with the Marginal (SQ) approach (Additional file 1: Fig. S11). While the test statistics are inflated, it is difficult to distinguish the factors driving inflation, e.g., bias due to unmodeled population structure versus biological signal under polygenic inheritance. Interestingly, we found that the genomic inflation factor increases as a function of minor allele frequency which may suggest there are vQTLs/covQTLs that are still underpowered to be discovered at current sample sizes (Additional file 1: Fig. S12) [60]. To rule out bias within the LIT framework driving inflation, we implemented a double Kolmogorov-Smirnov (KS) test [57] (detailed in "UK Biobank" section)

to assess LIT's *p*-values under the null hypothesis. We found that the double KS test *p*-values are $p = 0.56$ and $p = 0.76$ for uLIT and wLIT, respectively, and thus LIT demonstrates the characteristics of a well-behaved testing procedure on the UK Biobank data (Additional file 11: Fig. S13). However, it remains plausible there are other sources of bias (see the "Discussion" section) driving covariance/variance-specific effects, and to be conservative, we adjusted the significance results of each approach by the corresponding genomic inflation factor (Additional file 1: Fig. S10b). These adjusted *p*-values were then combined in aLIT to detect latent genetic interactions (Fig. 4).

Using the aLIT *p*-values, we discovered 2252 candidate SNPs that are either a vQTL or covQTL in 11 distinct regions. Table 1 shows the most significantly associated (lead) SNP in each region. As a comparison, we also applied Marginal (SQ/CP) and detected 2099 SNPs. Of those found by Marginal (SQ/CP), aLIT's results overlapped with ≈98% of the detected SNPs and had substantially smaller *p*-values at most loci (Additional file 1: Fig. S14). Although Marginal (SQ/CP) detects a few regions that are not found by aLIT, the aLIT *p*-values are comparable in magnitude at these regions (Additional file 1: Table S1). On the other hand, there are three distinct regions found by aLIT (rs2821230, rs11030066, and rs157845) where the *p*-values are substantially smaller than the *p*-values from Marginal (SQ/CP) (Table 1). Thus, in agreement with our simulation results, depending on the type of pleiotropy at a particular locus, there are regions where the significance results of aLIT are substantially more powerful than Marginal (SQ/CP), and other regions where Marginal (SQ/CP) is slightly more or as powerful than aLIT.

There are a couple of challenges for interaction tests that do not require observing the interactive variable(s). The first is that false positives are possible due to linkage disequilibrium (LD) with a SNP that has a large additive effect (see, e.g., [24]). To address this issue, for each lead SNP, we first identified nearby SNPs (within 1 Mb and correlation >0.1) with significant additive effects using the multivariate testing procedure, GAMuT [41]. We then applied the LIT implementations to the lead SNPs while regressing out nearby significant SNPs and found all of the lead SNPs remain significant (Table 1). The second challenge is that departures from linearity due to dominance or misspecification of the trait scale will induce a variance effect [17, 25]. To be conservative and help distinguish interactive effects, we removed the non-linear genetic signal within a locus by fitting a two degree of freedom genotypic model. After removing the dominance/scaling effects, the lead SNP rs9469860 on chromosome 6 was above the genome-wide significance
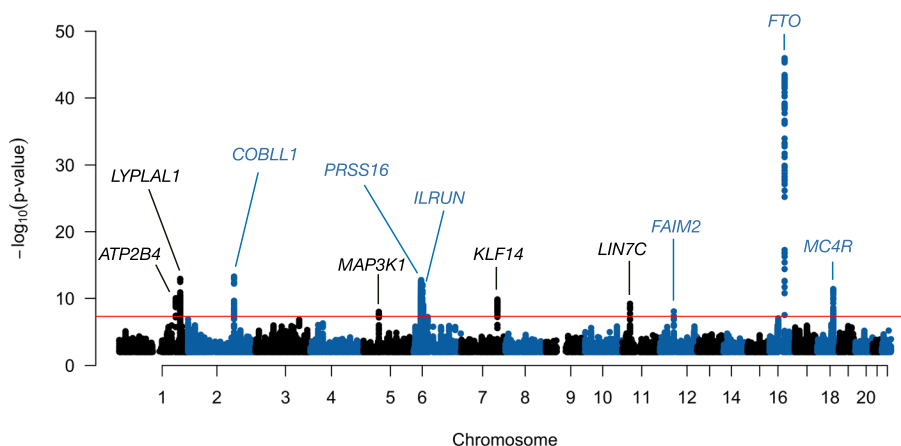
Bass *et al. Genome Medicine*   (2024) 16:62

Page 12 of 17

Fig. 4 Manhattan plot of aLIT *p*-values using obesity-related traits (waist circumference, hip circumference, body mass index, and body fat percentage) in the UK Biobank. The red line represents the genome-wide significance threshold of 5×10^{-8}. Note that *p*-values below 0.1 are removed from the plot

Table 1 aLIT and Marginal (SQ/CP) significance results of the lead SNPs from the UK Biobank analysis. There are two other *p*-values reported to help assess statistical significance in aLIT: (i) accounting for significant SNPs in linkage disequilibrium with the lead SNP (labeled "LD") and (ii) removing dominance and/or scaling effects (labeled "Dom.")

Chr.	Gene	Lead SNP	MAF	*p*-value (aLIT)	*p*-value (SQ/CP)	*p*-value (LD)	*p*-value (Dom.)
16	*FTO*	rs11642015	0.402	1.08×10^{-46}	2.73×10^{-40}	1.23×10^{-46}	1.28×10^{-46}
2	*COBLL1*	rs5835988	0.406	5.30×10^{-14}	1.08×10^{-9}	1.29×10^{-13}	5.32×10^{-14}
1	*LYPLAL1*	rs2820444	0.299	1.17×10^{-13}	2.64×10^{-13}	1.27×10^{-13}	1.19×10^{-13}
6	*PRSS16*	rs13212921	0.136	1.78×10^{-13}	3.40×10^{-12}	1.90×10^{-12}	4.85×10^{-9}
18	*MC4R*	rs35614134	0.234	3.95×10^{-12}	9.34×10^{-12}	1.39×10^{-11}	3.97×10^{-12}
1	*ATP2B4*	rs2821230	0.474	9.45×10^{-11}	1.19×10^{-5}	1.02×10^{-10}	9.38×10^{-11}
7	*KLF14*	rs972284	0.389	1.43×10^{-10}	2.43×10^{-9}	1.25×10^{-10}	1.43×10^{-10}
11	*LIN7C*	rs11030066	0.143	6.38×10^{-10}	5.03×10^{-6}	6.97×10^{-10}	6.62×10^{-10}
6	*ILRUN*	rs9469860	0.146	1.01×10^{-9}	4.47×10^{-9}	7.72×10^{-9}	3.62×10^{-7}
12	*FAIM2*	rs7132908	0.384	8.85×10^{-9}	4.22×10^{-9}	8.49×10^{-9}	8.96×10^{-9}
5	*MAP3K1*	rs157845	0.253	1.03×10^{-8}	2.39×10^{-4}	1.30×10^{-8}	1.03×10^{-8}

threshold (Table 1). In general, while the significance of a few loci were impacted (nearly all on chromosome 6 and a few on chromosome 18; Additional file 1: Fig. S15), most of the lead SNPs remained significant.

After evaluating the significant SNPs found by aLIT, we focused on ten lead SNPs that remained significant after accounting for LD and dominance/scaling issues. Using the Variants to Genes (V2G) measure on Open Targets Platform [61], we assigned the lead SNPs to the highest ranked genes (Table 1). A few of these genes are found in other GWAS of obesity-related traits. In particular, the *FTO* gene is a known obesity-related gene that is associated with type 2 diabetes (see, e.g., [62–64]). While V2G score assigned rs5935988 to *COBLL1* (198,262 bp), it was also close to *GRB14* (23,935 bp) which has been

associated with body fat distribution and may be involved in regulating insulin signaling [65–67]. Other genes that were assigned to lead SNPs are involved in regulating satiety and energy homeostasis (*MC4R*; [68]), adiposity (*LYPLAL1*; [69, 70]), and metabolic diseases such as type 2 diabetes (*KLF14*; [71, 72]).

We then searched for evidence of known interacting variables using the significant SNPs identified by aLIT. We focused our analysis on various lifestyle (alcohol frequency and smoking status) and socio-demographic (age at recruitment, sex, and household income before tax) environmental factors. We fit a multivariate regression model and found evidence of multiple genotype-by-environment interactions at a false discovery rate of 0.05 (Fig. 5; Additional file 1: Table S2). In particular, the

Bass *et al. Genome Medicine*   (2024) 16:62

Page 13 of 17

strongest genotype-by-environment signals involved sex as an interactive variable: rs5835988 ($p = 1.32 \times 10^{-51}$; *COBLL1/GRB14*), rs972284 ($p = 2.39 \times 10^{-14}$; *KLF14*), rs2820444 ($p = 7.23 \times 10^{-23}$; *LYPLAL1*), and rs35613134 ($p = 6.18 \times 10^{-5}$; *MC4R*). These results agree with previous work that has found sex-specific effects of variants in *GRB14*, *KLF14*, *LYPLAL1*, and *MC4R* [67, 72–74]. Thus, a subset of variants screened by LIT provide evidence of a genotype-by-sex interaction while also being nearby genes that have been implicated to have sex-specific effects from previous studies. It is also worth noting that there were significant interactions at the FTO locus with alcohol frequency ($p = 6.54 \times 10^{-5}$) and smoking ($p = 1.32 \times 10^{-3}$), which has been previously found in other work [24, 75].

Discussion

It is challenging to identify, observe, accurately measure, and then detect genetic interactions in a GWAS study. While there are methods to screen for interactions that do not require specifying the interactive partner(s) [17, 23–25], these approaches only consider a single trait. To increase statistical power, our proposed kernel-based framework, Latent Interaction Testing (LIT), leverages the shared genetic interaction signal from multiple related

traits (i.e., pleiotropy) while maintaining the flexibility of single trait approaches. In our simulation study, we found that the optimal implementation between wLIT and uLIT depends on the genetic architecture. We also found that combining the *p*-values from both approaches in aLIT maximized the number of discoveries while controlling the type I error rate. Furthermore, aLIT increased the power to detect latent genetic interactions compared to marginal testing procedures, and the difference was drastic for certain genetic architectures. We then applied the LIT framework to four obesity-related traits in the UK Biobank and found many loci with potential interactive effects, including evidence of vQTLs/covQTLs that are still underpowered to be discovered at current sample size.

While we emphasized the linear and projection kernels in our study, aLIT can incorporate multiple kernel choices (e.g., Gaussian) which may increase the power to detect complex interaction signals. However, including additional kernel functions will also increase the computational complexity which may be time prohibitive for biobank-sized datasets. In particular, calculating non-linear kernels such as a Gaussian kernel is computationally demanding for large sample sizes. One promising direction for implementing non-linear kernels is to project the data onto a randomized low-dimensional
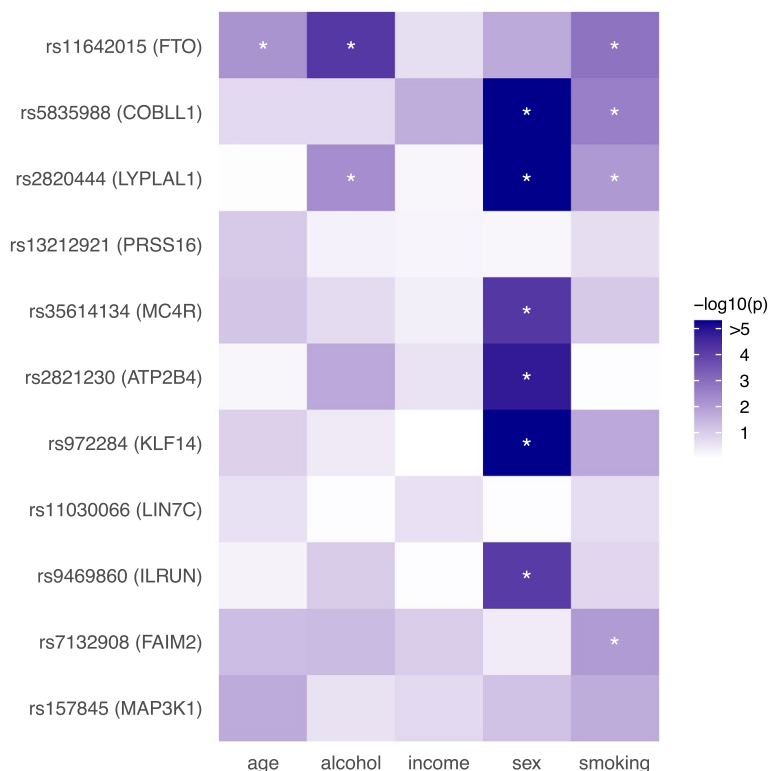

Fig. 5 SNP-by-environment interaction results in UK Biobank for lead SNPs in genes detected by LIT. Asterisk denotes tests significant at a false discovery rate threshold of 0.05

Bass *et al. Genome Medicine*   (2024) 16:62

Page 14 of 17

subspace to approximate the kernel function [76, 77]. We explored this approximation in aLIT using a Gaussian kernel and found comparable performance to the linear kernel (Additional file 1: Figs. S16-S17; see Fig. S18 for computational time comparisons). Because the accuracy of this approximation depends on the chosen dimensionality, a two-stage procedure can be implemented where a crude approximation (very low dimensionality) is applied genome-wide and then a subset of tests below some significance threshold are selected to calculate an improved kernel approximation (larger dimensionality) [77]. This two-stage strategy can substantially reduce the computational time when extending the LIT framework to non-linear kernels for biobank-sized datasets.

LIT is designed to screen for vQTLs/covQTLs in genome-wide data to identify loci with complex non-additive genetic signals. While these non-additive genetic signals may be due to interactions with other genes or environment (which is our main focus), we note that other phenomena can induce non-additive genetic signals. From a biological perspective, variants demonstrating parent-of-origin effects can also present as vQTLs/covQTLs [78, 79]. In addition, non-additive genetic effects can also result from non-biological origins; for example, a vQTL/covQTL could be due to an indirect effect of the variant on phenotype mediated through an environmental factor. While LIT (and other vQTL methods) cannot distinguish the source of the non-additive genetic effect, it is a valuable tool to detect loci to use in follow-up investigation where the multiple testing burden will be reduced when considering multiple environmental (and genetic) factors. For example, in our applied analysis, we tested sex as a potential interaction partner of the lead SNPs and found loci in *GRB14*, *KLF14*, *LYPLAL1*, and *MC4R* with significant interactive effects. Additionally, the subset of screened loci can also be used for other purposes such as constructing improved estimates of polygenic risk scores (PRS) in PRS-by-environment testing [18]. It is this type of follow-up analysis that we anticipate will make LIT useful for researchers.

There are some caveats when interpreting the significance results of LIT or, more generally, any approach that does not require observing the interactive variable(s). Type I error rate control is impacted by loci with large additive effects and trait scaling issues. To address the former issue, we performed inference using all SNPs in LD with the lead SNPs and found it did not impact significance. Previous work has found that the test statistic inflation from common variants in LD with a strong additive effect is small for a single trait [24]. While it is likely that using multiple traits will increase the inflation

compared to a single trait, we demonstrate in our applied analysis that it is simple to account for a LD region with LIT. A related issue is that the true causal SNP may not be tagged but this is unlikely an issue in this work since there is dense coverage with the imputed genotypes. We also assessed model misspecification due to an incorrect trait scaling by fitting a genotypic model to flexibly capture non-linear genetic signals. Interestingly, we found that our significance results were primarily impacted at loci located on chromosome 6 (outside the MHC). While this strategy can help identify the extent of genome-wide inflation due to dominance/scaling, it cannot determine whether the latent interactive effects are an artifact of the scale, even though our simulations suggest that detections by LIT are robust to deviations from normality. When presented with traits that follow a non-normal distribution, a rank-based inverse normal transformation is typically applied so a trait "appears" as a standard normal distribution. However, for screening latent interactions, we recommend against such practice as it does not correct for the mean-variance relationship and can lead to invalid inference [25]. In general, model misspecification from an incorrect scaling is problematic for any population genetic analysis and may require other approaches such as goodness-of-fit testing to help identify an appropriate variance-stabilizing transformation.

There are also several important considerations when applying LIT to genetic data. Importantly, in this work, we assume that individuals are unrelated and traits follow a multivariate normal distribution. While LIT assumes the data follows a multivariate normal distribution, our simulation study suggests that it is robust to violations of this assumption. In general, the computational time of LIT increases as the number of traits and sample size increases (Additional file 1: Fig. S18). Therefore, in order to analyze biobank-sized datasets, LIT uses multiple cores to distribute SNPs (e.g., on the same chromosome) for interaction testing to be computationally more efficient. Because calculating the residual cross products for a large number of traits is computationally intensive ($\binom{r}{2}$ increase in computational time per SNP), LIT provides a user option to only use the squared residuals. However, as demonstrated in simulation and in the UK Biobank dataset, employing this option is likely to reduce power. Finally, while a discovery in LIT suggests evidence of a non-additive effect, LIT does not identify the trait, or subset of traits, driving that result, and, consequently, does not distinguish whether the variant is a vQTL and/or covQTL. To do so, investigators might consider running Marginal (SQ/CP) at the "lead" SNP to rank/identify individual traits with non-additive effects (squared residuals; vQTLs) and pairs

Bass *et al. Genome Medicine*   (2024) 16:62

Page 15 of 17

of traits with shared non-additive effects (cross products; covQTLs, although, see the "Methods" section).

Conclusions

For many complex traits, there is strong discrepancy between GWAS-based estimates of heritability (which explicitly assume additive effects of genetic variation) and family-based estimates (which may incorporate non-additive effects and higher-order interactions). For GWAS-based estimates of heritability, the large multiple testing burden, coupled with small interaction effect sizes, has made it very difficult to discover such effects. With recent biobank-sized datasets, we can begin to screen for loci with non-additive genetic variation that contribute to this missing heritability while understanding its role in the etiology of complex traits. As biobank-sized datasets become more prevalent, we anticipate that computationally scalable approaches that leverage information across multiple traits, such as LIT, will become increasingly important to discovering non-additive genetic loci.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01329-0.

Additional file 1. Supplementary methods. Theoretical results for latent interaction testing. **Figure S1.** General strategy to detect latent genetic interactions when there are two unobserved environments. **Figure S2.** Revealing latent interactive effects using multiple traits. **Figure S3.** False positive rate of the LIT implementations under the null hypothesis of no interaction. **Figure S4.** Q-Q plot of the LIT implementations under the null hypothesis of no interaction. **Figure S5.** False positive rate of the LIT implementations when applied to 5 SNPs. **Figure S6.** The empirical power of the principal components for the squared residual and cross product matrix at various baseline correlations. **Figure S7.** Comparing uLIT, wLIT, and aLIT when the direction of the effect size for the interaction term is opposite of the environmental variable. **Figure S8.** Comparing Levene's test to aLIT, Marginal (SQ/CP), and Marginal (SQ) when the direction of the effect size for the interaction term is the same as the environmental variable. **Figure S9.** Comparing Levene's test to aLIT, Marginal (SQ/CP), and Marginal (SQ) when the direction of the effect size for the interaction term is opposite of the environmental variable. **Figure S10.** Quantile-Quantile plot of the uLIT, wLIT, and aLIT *p*-values from the UK Biobank. **Figure S11.** Quantile-Quantile plot of the Marginal (SQ) *p*-values from the UK Biobank using the traits BFP, BMI, HC and WC. **Figure S12.** The genomic inflation factor as a function of minor allele frequency in the UK Biobank analysis. **Figure S13.** The double KS test procedure using the set of independent SNPs. **Figure S14.** Comparison of the significance results using the marginal testing procedure and aLIT. **Figure S15.** Comparison of aLIT *p*-values after adjusting for additive genetic effects and dominance/scaling effects. **Figure S16.** Simulation results using a Gaussian kernel in LIT when the direction of the effect size for the interaction term is the same as the environmental variable. **Figure S17.** Simulation results using a Gaussian kernel in LIT when the direction of the effect size for the interaction term is opposite of the environmental variable. **Figure S18.** The average computational time to run aLIT on a SNP as a function of sample size, number of traits, and kernel function. **Table S1.** Lead SNPs of significant findings from Marginal (SQ/CP) in the UK Biobank. **Table S2.** Genotype-by-environment interaction results for lifestyle (alcohol and smoking) and socio-demographic (age, sex, and income) environmental factors in the UK Biobank.

Availability of data and materials
LIT is publicly available in the R package `lit`. The package can be downloaded at https://CRAN.R-project.org/package=lit (release) [80] or https://github.com/ajbass/lit (most recent version) [81]. The code to reproduce the results in this work can be found at https://github.com/ajbass/lit_manuscript and access to the UK Biobank data can be requested at https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

Declarations

Ethics approval and consent to participate
The research in this work abides by the Helsinki Declaration. The UK Biobank is advised by its Ethics Advisory Committee (EAC) and has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB). The UK Biobank obtained informed consent from all participants. This research has been conducted using the UK Biobank Resource under Applied Number 58259.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
2. Maher B. Personal genomes: the case of the missing heritability. Nature. 2008;456(7218):18–21. https://doi.org/10.1038/456018a.
3. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206. https://doi.org/10.1038/nature14177.
4. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47(10):1114–20. https://doi.org/10.1038/ng.3390.
5. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–9. https://doi.org/10.1093/hmg/ddy271.
6. Wainschtein P, Jain D, Zheng Z, Aslibekyan S, Becker D, Bi W, et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet. 2022;54(3):263–73. https://doi.org/10.1038/s41588-021-00997-7.
7. Elks C, Den Hoed M, Zhao JH, Sharp S, Wareham N, Loos R, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. Front Endocrinol. 2012;3. https://doi.org/10.3389/fendo.2012.00029.

8. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. Nat Rev Genet. 2022;23(2):120–33. https://doi.org/10.1038/s41576-021-00414-z.

9. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53. https://doi.org/10.1038/nature08494.

10. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42(7):565–9. https://doi.org/10.1038/ng.608.

11. Gibson G. Rare and common variants: twenty arguments. Nat Rev Genet. 2012;13(2):135–45. https://doi.org/10.1038/nrg3118.

12. López-Cortegano E, Caballero A. Inferring the nature of missing heritability in human traits using data from the GWAS catalog. Genetics. 2019;212(3):891–904. https://doi.org/10.1534/genetics.119.302077.

13. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci. 2012;109(4):1193–8. https://doi.org/10.1073/pnas.1119675109.

14. Hemani G, Knott S, Haley C. An evolutionary perspective on epistasis and the missing heritability. PLoS Genet. 2013;9(2):1–11. https://doi.org/10.1371/journal.pgen.1003295.

15. Robinson MR, English G, Moser G, Lloyd-Jones LR, Triplett MA, Zhu Z, et al. Genotype-covariate interaction effects and the heritability of adult body mass index. Nat Genet. 2017;49(8):1174–81. https://doi.org/10.1038/ng.3912.

16. Tyrrell J, Wood AR, Ames RM, Yaghootkar H, Beaumont RN, Jones SE, et al. Gene-obesogenic environment interactions in the UK Biobank study. Int J Epidemiol. 2017;46(2):559–75. https://doi.org/10.1093/ije/dyw337.

17. Sulc J, Mounier N, Günther F, Winkler T, Wood AR, Frayling TM, et al. Quantification of the overall contribution of gene-environment interaction for obesity-related traits. Nat Commun. 2020;11(1):1385. https://doi.org/10.1038/s41467-020-15107-0.

18. Nagpal S, Tandon R, Gibson G. Canalization of the polygenic risk for common diseases and traits in the UK Biobank cohort. Mol Biol Evol. 2022;39(4). https://doi.org/10.1093/molbev/msac053.

19. Aschard H. A perspective on interaction effects in genetic association studies. Genet Epidemiol. 2016;40(8):678–88. https://doi.org/10.1002/gepi.21989.

20. Kraft P, Aschard H. Finding the missing gene-environment interactions. Eur J Epidemiol. 2015;30(5):353–5. https://doi.org/10.1007/s10654-015-0046-1.

21. Marderstein AR, Davenport ER, Kulm S, Van Hout CV, Elemento O, Clark AG. Leveraging phenotypic variability to identify genetic interactions in human phenotypes. Am J Hum Genet. 2021;108(1):49–67. https://doi.org/10.1016/j.ajhg.2020.11.016.

22. Trevor S Breusch and Adrian R Pagan. A simple test for heteroscedasticity and random coefficient variation. Econometrica. 1979;47(5):1287–94.

23. 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(6):1–10. https://doi.org/10.1371/journal.pgen.1000981.

24. Wang H, Zhang F, Zeng J, Wu Y, Kemper KE, Xue A, et al. Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. Sci Adv. 2019;5(8). https://doi.org/10.1126/sciadv.aaw3538.

25. Young AI, Wauthier FL, Donnelly P. Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. Nat Genet. 2018;50(11):1608–14. https://doi.org/10.1038/s41588-018-0225-6.

26. Barton NH, Turelli M. Evolutionary quantitative genetics: how little do we know? Annu Rev Genet. 1989;23(1):337–70. https://doi.org/10.1146/annurev.ge.23.120189.002005.

27. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet. 2011;89(5):607–18. https://doi.org/10.1016/j.ajhg.2011.10.004.

28. Gratten J, Visscher PM. Genetic pleiotropy in complex traits and diseases: implications for genomic medicine. Genome Med. 2016;8(1):78. https://doi.org/10.1186/s13073-016-0332-x.

29. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014;11(4):407–9. https://doi.org/10.1038/nmeth.2848.

30. Aschard H, Vilhjálmsson BJ, Greliche N, Morange PE, Trégouët DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. Am J Hum Genet. 2014;94(5):662–76. https://doi.org/10.1016/j.ajhg.2014.03.016.

31. Lea A, Subramaniam M, Ko A, Lehtimäki T, Raitoharju E, Kähönen M, et al. Genetic and environmental perturbations lead to regulatory decoherence. eLife. 2019;8:e40538. https://doi.org/10.7554/eLife.40538.

32. Song M, Hao W, Storey JD. Testing for genetic associations in arbitrarily structured populations. Nat Genet. 2015;47(5):550–4. https://doi.org/10.1038/ng.3244.

33. Hao W, Song M, Storey JD. Probabilistic models of genetic variation in structured populations applied to global human studies. Bioinformatics. 2015;32(5):713–21. https://doi.org/10.1093/bioinformatics/btv641.

34. Bass AJ. High-dimensional methods to model biological signal in genome-wide studies [PhD thesis]. Princeton University; 2021. http://arks.princeton.edu/ark:/88435/dsp01m326m485q.

35. Gretton A, Fukumizu K, Teo C, Song L, Schölkopf B, Smola A. A kernel statistical test of independence. In: Platt J, Koller D, Singer Y, Roweis S, editors. Advances in neural information processing systems, vol. 20. Curran Associates, Inc.; 2007. https://proceedings.neurips.cc/paper/2007/file/d5cfead94f5350c12c322b5b664544c1-Paper.pdf.

36. Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. Ann Stat. 2007;35(6):2769–94. https://doi.org/10.1214/009053607000000505.

37. Zhang K, Peters J, Janzing D, Schölkopf B. Kernel-based conditional independence test and application in causal discovery. In: Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. UAI'11. Arlington: AUAI Press; 2011. p. 804-813.

38. Kwee LC, Liu D, Lin X, Ghosh D, Epstein MP. A powerful and flexible multilocus association test for quantitative traits. Am J Hum Genet. 2008;82(2):386–97.

39. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet. 2010;86(6):929–42. https://doi.org/10.1016/j.ajhg.2010.05.002.

40. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82–93. https://doi.org/10.1016/j.ajhg.2011.05.029.

41. Broadaway KA, Cutler DJ, Duncan R, Moore JL, Ware EB, Jhun MA, et al. A statistical approach for testing cross-phenotype effects of rare variants. Am J Hum Genet. 2016;98(3):525–40. https://doi.org/10.1016/j.ajhg.2016.01.017.

42. Josse J, Holmes S. Measuring multivariate association and beyond. Stat Surv. 2016;10(none):132 – 67. https://doi.org/10.1214/16-SS116.

43. Sejdinovic D, Sriperumbudur B, Gretton A, Fukumizu K. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. Ann Stat. 2013;41(5):2263–91. https://doi.org/10.1214/13-AOS1140.

44. Hua WY, Ghosh D. Equivalence of kernel machine regression and kernel distance covariance for multidimensional phenotype association studies. Biometrics. 2015;71(3):812–20. https://doi.org/10.1111/biom.12314.

45. Schaid DJ. Genomic similarity and kernel methods II: methods for genomic information. Hum Hered. 2010;70(2):132–40.

46. Davies RB, Algorithm AS. 155: The distribution of a linear combination of χ^2 random variables. J R Stat Soc Ser C Appl Stat. 1980;29(3):323–33.

47. Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. J Res Natl Bur Stand. 1950;45:255–82.

48. Arnoldi WE. The principle of minimized iterations in the solution of the matrix eigenvalue problem. Q Appl Math. 1951;9(1):17–29.

49. Lehoucq RB, Sorensen DC. Deflation techniques for an implicitly restarted Arnoldi iteration. SIAM J Matrix Anal Appl. 1996;17(4):789–821. https://doi.org/10.1137/S0895479895281484.

50. Jolliffe IT. Principal component analysis. Springer Series in Statistics. Springer; 2002.

51. Liu Y, Xie J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J Am Stat Assoc. 2020;115(529):393–402. https://doi.org/10.1080/01621459.2018.1554485.

52. 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 Med. 2015;12(3):1–10. https://doi.org/10.1371/journal.pmed.1001779.

53. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9. https://doi.org/10.1038/s41586-018-0579-z.

54. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2016. https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf. Accessed 16 Mar 2022.

55. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource. 2015. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/genotyping_qc.pdf. Accessed 16 Mar 2022.

56. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

57. Leek JT, Storey JD. The joint null criterion for multiple hypothesis tests. Stat Appl Genet Mol Biol. 2011;10(1). https://doi.org/10.2202/1544-6115.1673.

58. Dutta D, Scott L, Boehnke M, Lee S. Multi-SKAT: General framework to test for rare-variant association with multiple phenotypes. Genet Epidemiol. 2019;43(1):4–23. https://doi.org/10.1002/gepi.22156.

59. Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: a fast and powerful p-value combination method for rare-variant analysis in sequencing studies. Am J Hum Genet. 2019;104(3):410–21. https://doi.org/10.1016/j.ajhg.2019.01.002.

60. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807–12. https://doi.org/10.1038/ejhg.2011.39.

61. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et al. Open Targets Platform: supporting systematic drug-target identification and prioritisation. Nucleic Acids Res. 2020;49(D1):D1302–10. https://doi.org/10.1093/nar/gkaa1027.

62. Dina C, Meyre D, Gallina S, Durand E, Körner A, Jacobson P, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet. 2007;39(6):724–6. https://doi.org/10.1038/ng2048.

63. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007;316(5826):889–94. https://doi.org/10.1126/science.1141634.

64. Herrera BM, Keildson S, Lindgren CM. Genetics and epigenetics of obesity. Maturitas. 2011;69(1):41–9.

65. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012;44(9):981–90. https://doi.org/10.1038/ng.2383.

66. Karaderi T, Drong AW, Lindgren CM. Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of obesity-related traits. Curr Diabetes Rep. 2015;15(10):83. https://doi.org/10.1007/s11892-015-0648-8.

67. Pulit SL, Karaderi T, Lindgren CM. Sexual dimorphisms in genetic loci linked to body fat distribution. Biosci Rep. 2017;37(1). https://doi.org/10.1042/BSR20160184.

68. Namjou B, Stanaway IB, Lingren T, Mentch FD, Benoit B, Dikilitas O, et al. Evaluation of the MC4R gene across eMERGE network identifies many unreported obesity-associated variants. Int J Obes. 2021;45(1):155–69. https://doi.org/10.1038/s41366-020-00675-4.

69. Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet. 2010;42(11):949–60. https://doi.org/10.1038/ng.685.

70. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, et al. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. PLoS Genet. 2009;5(6):1–13. https://doi.org/10.1371/journal.pgen.1000508.

71. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet. 2010;42(7):579–89. https://doi.org/10.1038/ng.609.

72. Small KS, Todorčević M, Civelek M, El-Sayed Moustafa JS, Wang X, Simon MM, et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. Nat Genet. 2018;50(4):572–80. https://doi.org/10.1038/s41588-018-0088-x.

73. Horstmann A, Kovacs P, Kabisch S, Boettcher Y, Schloegl H, Tönjes A, et al. Common genetic variation near MC4R has a sex-specific impact on human brain structure and eating behavior. PLoS ONE. 2013;8(9):1–9. https://doi.org/10.1371/journal.pone.0074362.

74. Norheim F, Hasin-Brumshtein Y, Vergnes L, Chella Krishnan K, Pan C, Seldin MM, et al. Gene-by-sex interactions in mitochondrial functions and cardio-metabolic traits. Cell Metab. 2019;29(4):932–949.e4. https://doi.org/10.1016/j.cmet.2018.12.013.

75. Young AI, Wauthier F, Donnelly P. Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. Nat Commun. 2016;7(1):12724. https://doi.org/10.1038/ncomms12724.

76. Rahimi A, Recht B. Random features for large-scale kernel machines. In: Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07. Red Hook: Curran Associates Inc.; 2007. p. 1177-1184.

77. Fu B, Pazokitoroudi A, Sudarshan M, Liu Z, Subramanian L, Sankararaman S. Fast kernel-based association testing of non-linear genetic effects for biobank-scale data. Nat Commun. 2023;14(1):4936. https://doi.org/10.1038/s41467-023-40346-2.

78. Head ST, Leslie EJ, Cutler DJ, Epstein MP. POIROT: a powerful test for parent-of-origin effects in unrelated samples leveraging multiple phenotypes. Bioinformatics. 2023;39(4):btad199. https://doi.org/10.1093/bioinformatics/btad199.

79. Hoggart CJ, Venturini G, Mangino M, Gomez F, Ascari G, Zhao JH, et al. Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. PLoS Genet. 2014;10(7):1–12. https://doi.org/10.1371/journal.pgen.1004508.

80. Bass A, Epstein M. lit: latent interaction testing for genome-wide studies. CRAN; 2023. https://CRAN.R-project.org/package=lit. Accessed 15 Aug 2023.

81. Bass A, Epstein M. lit: latent interaction testing for genome-wide studies. GitHub; 2023. https://github.com/ajbass/lit. Accessed 15 Aug 2023.

Publisher's Note