

A scalable and robust variance components method reveals insights into the architecture of gene-environment interactions underlying complex traits

Authors

Ali Pazokitoroudi, Zhengtong Liu, Andrew Dahl,
Noah Zaitlen, Saharon Rosset,
Sriram Sankararaman

Correspondence

alipazoki@cs.ucla.edu (A.P.),
sriram@cs.ucla.edu (S.S.)

Pazokitoroudi et al. introduced a scalable method to estimate trait variance explained by gene-environment interactions across the genome and within specific genomic annotations. Application of the method in the UK Biobank uncovered significant genome-wide GxE heritability and enrichment of GxE heritability within population and functional genomic annotations.



A scalable and robust variance components method reveals insights into the architecture of gene-environment interactions underlying complex traits

Ali Pazokitoroudi,^{1,7,8,*} Zhengtong Liu,¹ Andrew Dahl,⁵ Noah Zaitlen,^{2,3,4} Saharon Rosset,⁶ and Sriram Sankararaman^{1,2,3,*}

Summary

Understanding the contribution of gene-environment interactions (GxE) to complex trait variation can provide insights into disease mechanisms, explain sources of heritability, and improve genetic risk prediction. While large biobanks with genetic and deep phenotypic data hold promise for obtaining novel insights into GxE, our understanding of GxE architecture in complex traits remains limited. We introduce a method to estimate the proportion of trait variance explained by GxE (GxE heritability) and additive genetic effects (additive heritability) across the genome and within specific genomic annotations. We show that our method is accurate in simulations and computationally efficient for biobank-scale datasets.

We applied our method to common array SNPs (MAF $\geq 1\%$), fifty quantitative traits, and four environmental variables (smoking, sex, age, and statin usage) in unrelated white British individuals in the UK Biobank. We found 68 trait-E pairs with significant genome-wide GxE heritability ($p < 0.05/200$) with a ratio of GxE to additive heritability of $\approx 6.8\%$ on average. Analyzing ≈ 8 million imputed SNPs (MAF $\geq 0.1\%$), we documented an approximate 28% increase in genome-wide GxE heritability compared to array SNPs. We partitioned GxE heritability across minor allele frequency (MAF) and local linkage disequilibrium (LD) values, revealing that, like additive allelic effects, GxE allelic effects tend to increase with decreasing MAF and LD. Analyzing GxE heritability near genes highly expressed in specific tissues, we find significant brain-specific enrichment for body mass index (BMI) and basal metabolic rate in the context of smoking and adipose-specific enrichment for waist-hip ratio (WHR) in the context of sex.

Introduction

Variation in a complex trait is modulated by an interplay between genetic and environmental factors. Characterizing the effects of gene-environment interactions (GxE) on complex trait variation has the potential to shed light on biological mechanisms underlying the trait,^{1–3} inform public health measures,⁴ identify sources of missing heritability,⁵ and improve the accuracy and portability of trait prediction.^{6,7} The growth of biobanks that collect genetic and deep phenotypic data (that span disease outcomes, clinical labs, lifestyle factors, and environmental exposures) across large numbers of individuals offers the possibility to gain novel insights into GxE.^{3,8} Nevertheless, characterizing GxE has proved challenging due, in part, to the small effect sizes of individual genetic variants.^{9,10}

A potentially powerful methodological approach aims to quantify GxE effects aggregated across a set of variants without needing to pinpoint individual variants. In this approach, the proportion of trait variation explained by GxE (GxE heritability or h_{gxe}^2) is estimated by fitting a class of variance components models where the model param-

eters, i.e., the variance components, are informative of h_{gxe}^2 . Methods for estimating h_{gxe}^2 using this approach include GCTA-GxE,¹¹ multitrait GREML (MV-GREML),⁵ random regression GREML (RR-GREML),^{5,12} and whole-genome reaction norm model (RNM) and its multitrait version (MRNM).¹³ All of these methods (except RNM) are able to account for differences in the noise or residual variance across environments (noise heterogeneity), which is important to mitigate biases in GxE heritability estimates.^{13,14} However, these methods work with discrete-valued environmental variables, with RNM and MRNM further restricted to fit bivariate and univariate environments, respectively. A more recent general framework, GxEMM,¹⁴ can be applied to both discrete and continuous environmental variables while modeling noise heterogeneity. However, none of these methods are practical for biobank-scale datasets with sample sizes in the hundreds of thousands and genetic variants in the millions. Two recent methods, GPLEMMA¹⁵ and MEMMA,¹⁶ attempt to scale GxE heritability estimation to large-scale datasets but do not model noise heterogeneity. A more recent method, MonsterLM,¹⁷ has been shown to be feasible for biobank-scale datasets and to produce

¹Department of Computer Science, UCLA, Los Angeles, CA, USA; ²Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA; ³Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA, USA; ⁴Department of Neurology, UCLA, Los Angeles, CA, USA; ⁵Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA; ⁶Department of Statistics, Tel-Aviv University, Tel-Aviv, Israel; ⁷Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA; ⁸Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

*Correspondence: alipazoki@cs.ucla.edu (A.P.), sriram@cs.ucla.edu (S.S.)

<https://doi.org/10.1016/j.ajhg.2024.05.015>.

© 2024 The Authors. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



unbiased estimates in many scenarios. However, MonsterLM requires SNPs to be filtered to common variants with low levels of linkage disequilibrium (LD), which may limit its application to discover GxE. As a result, current methods for estimating GxE heritability either do not scale to the biobank setting or are susceptible to biased estimates. Additional insights into the architecture of GxE can be gleaned if we can move beyond genome-wide estimates of GxE heritability and estimate GxE heritability across specific genomic annotations such as minor allele frequency (MAF), LD, and functional genomic annotations.

We propose a scalable and robust method, GENIE (gene-environment interaction estimator) that can estimate the proportion of trait variance explained by GxE and additive genetic effects (additive heritability). Using extensive simulations and real data analysis, we show that GENIE accurately estimates h_{gxe}^2 and provides calibrated tests of h_{gxe}^2 due to its ability to account for noise that is heterogeneous across environments. Importantly, GENIE is scalable: able to estimate GxE on datasets with hundreds of thousands of individuals, millions of SNPs, and tens of environmental variables in several hours. The ability of GENIE to be applied to large-scale datasets is important for power: we show that GENIE has adequate power to detect h_{gxe}^2 as low as 2% across a sample of $\approx 300,000$ unrelated individuals. Finally, GENIE is versatile: able to handle multiple environmental variables (discrete or continuous) and to estimate not only genome-wide h_{gxe}^2 but also partition h_{gxe}^2 across genomic annotations (both overlapping and non-overlapping).

To demonstrate its utility, we first applied GENIE to estimate the genome-wide h_{gxe}^2 on common SNPs ($M = 454,207$ SNPs with $MAF > 1\%$) and four environmental variables (smoking, sex, age, and statin usage) for fifty quantitative phenotypes measured across 291,273 unrelated white British individuals in the UK Biobank (UKB). Second, we leveraged the scalability of GENIE to partition h_{gxe}^2 across common and low-frequency imputed SNPs ($M = 7,774,235$ with $MAF > 0.1\%$) in UKB. We partitioned h_{gxe}^2 into genomic annotations based on the MAF and local LD score of each SNP to investigate the variation in GxE effects with population genetic features and to estimate genome-wide h_{gxe}^2 that includes the contribution of both common and low-frequency SNPs. Finally, we applied GENIE to assess whether h_{gxe}^2 shows tissue-specific enrichment by analyzing each of 53 tissue-specific gene sets identified from the GTEx dataset.¹⁸

Material and methods

Generalized GxE linear mixed model

Let \mathbf{X} denote a $N \times M$ genotype matrix, \mathbf{E} denote a $N \times L$ matrix of environmental variables, \mathbf{C} denote a $N \times P$ matrix of fixed-effect covariates, and \mathbf{y} denote an N -vector of phenotypes. We assume the following linear mixed model:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \sum_{l=1}^L (\mathbf{X} \odot \mathbf{E}_{\cdot l}) \boldsymbol{\alpha}_l + \sum_{l=1}^L (\mathbf{I}_N \odot \mathbf{E}_{\cdot l}) \boldsymbol{\delta}_l + \mathbf{C}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \boldsymbol{\beta} &\sim D\left(\mathbf{0}, \frac{\sigma_g^2}{M} \mathbf{I}_M\right) \\ \boldsymbol{\alpha}_l &\sim D\left(\mathbf{0}, \frac{\sigma_{gxe,l}^2}{M} \mathbf{I}_M\right) \\ \boldsymbol{\delta}_l &\sim D\left(\mathbf{0}, \sigma_{nxe,l}^2 \mathbf{I}_N\right) \\ \boldsymbol{\epsilon} &\sim D\left(\mathbf{0}, \sigma_e^2 \mathbf{I}_N\right) \end{aligned} \quad (\text{Equation 1})$$

Here, $D(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes an arbitrary distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\mathbf{E}_{\cdot l}$ denotes l -th column of \mathbf{E} , and \odot denotes row-wise Kronecker product. $\boldsymbol{\beta}$ denotes the M -vector of SNP effect sizes, $\boldsymbol{\gamma}$ denotes the P -vector of fixed effects, $\boldsymbol{\alpha}_l$ denotes the M -vector of genetic effect sizes in the context of environment l (GxE effects) while $\boldsymbol{\delta}_l$ denotes the N -vector of noise-by-environment effect sizes for environment l , and $\boldsymbol{\epsilon}$ denotes the N -vector of noise. σ_e^2 , σ_g^2 , $\sigma_{gxe,l}^2$, and $\sigma_{nxe,l}^2$ denote the residual variance, additive genetic, gene-by-environment, and noise-by-environment variance components, respectively. These variance components can then be transformed into the additive heritability or the proportion of variance explained by additive effects (h_g^2 associated with σ_g^2) and the GxE heritability or the proportion of variance explained by interactions of genetics with a given environment ($h_{gxe,l}^2$ associated with $\sigma_{gxe,l}^2$). The noise-by-environment matrix for environment l is obtained as the row-wise Kronecker product between the $N \times N$ identity matrix \mathbf{I}_N and the environment vector $\mathbf{E}_{\cdot l}$ so that the vector of environment-specific noise for each individual i (due to environment l) will be given by $E_{il} \boldsymbol{\delta}_l$. In the simplest case of a binary environment that is coded as $\{0, 1\}$, the phenotype of an individual whose environmental variable is set to value 1 will have an additional contribution of noise ($\boldsymbol{\delta}_l$) relative to an individual whose environmental variable is set to 0. Further, all individuals whose environmental variable takes the value 1 will have an additional term that contributes to their phenotypic variance, quantified by $\sigma_{nxe,l}^2$ relative to individuals with environmental variable 0. This formulation generalizes to settings where the environment is coded as categorical (but with values different from $\{0, 1\}$) and to continuous-valued environments. We now refer to the noise-by-environment (or heterogeneous noise) component as the NxE component and the variance σ_{nxe}^2 as the NxE variance in the following sections.

Estimation in the GxE linear mixed model

We assume without loss of generality that \mathbf{y} is centered, and the columns of \mathbf{X} and \mathbf{E} are standardized. To estimate the variance components of our linear mixed model (LMM), we use a method-of-moments (MoM) estimator that searches for parameter values so that the population moments are close to the sample moments. Since $E[\mathbf{y}] = 0$, we derived the MoM estimates by equating the population covariance to the empirical covariance. For simplicity, we exclude the matrix of covariates \mathbf{C} from the model in the following derivation as the covariates can be efficiently projected out of the phenotype, genotypes, and interaction terms with minimal additional cost (Note S1).

For compactness, we denote $\mathbf{Z}_0 = \mathbf{X}$, $\mathbf{Z}_l = \mathbf{X} \odot \mathbf{E}_{\cdot l}$ for $l = 1, \dots, L$, $\mathbf{Z}_l = \mathbf{I}_N \odot \mathbf{E}_{\cdot l}$ for $l = L + 1, \dots, 2L$, and $\mathbf{Z}_{2L+1} = \mathbf{I}_N$. The population covariance is given by

$$\text{cov}(\mathbf{y}) = E[\mathbf{y}\mathbf{y}^T] - E[\mathbf{y}]E[\mathbf{y}^T] = \sum_{l=0}^{2L+1} \sigma_l^2 \mathbf{K}_l \quad (\text{Equation 2})$$

where

$$\mathbf{K}_l = \begin{cases} \frac{\mathbf{Z}_l \mathbf{Z}_l^T}{M}, & l = 0, \dots, L \\ \mathbf{Z}_l \mathbf{Z}_l^T, & l = L + 1, \dots, 2L + 1 \end{cases}$$

and

$$\sigma_l^2 = \begin{cases} \sigma_s^2, & l = 0 \\ \sigma_{gxe,l}^2, & l = 1, \dots, L \\ \sigma_{nxe,l}^2, & l = L + 1, \dots, 2L \\ \sigma_e^2, & l = 2L + 1 \end{cases}$$

Using $\mathbf{y}\mathbf{y}^T$ as our estimate of the empirical covariance, we need to solve the following least squares problem to find the variance components.

$$\hat{\sigma}^2 = \underset{\sigma^2}{\operatorname{argmin}} \|\mathbf{y}\mathbf{y}^T - \sum_{l=0}^{2L+1} \sigma_l^2 \mathbf{K}_l\|_F^2 \quad (\text{Equation 3})$$

The MoM estimator satisfies the following normal equations:

$$\mathbf{T}\sigma^2 = \mathbf{q} \quad (\text{Equation 4})$$

where \mathbf{T} is matrix with entries $T_{ij} = \operatorname{tr}(\mathbf{K}_i \mathbf{K}_j)$, $i, j \in \{0, \dots, 2L + 1\}$, and \mathbf{q} and σ^2 are vectors with entries $c_l = \mathbf{y}^T \mathbf{K}_l \mathbf{y}$ and σ_l^2 , respectively, for $l \in \{0, \dots, 2L + 1\}$.

The heritability associated with component i for a component that represents additive genetic or GxE effects (equivalently, the proportion of variance explained by component i) is defined as follows:

$$h_i^2 = \frac{\sigma_i^2 \operatorname{tr}(\mathbf{K}_i)}{\sum_k \sigma_k^2 \operatorname{tr}(\mathbf{K}_k)} \quad (\text{Equation 5})$$

The aforementioned definition of heritability holds when the columns of each of the \mathbf{Z} matrices have zero means and N is large. To explicitly ensure that the columns of GxE matrices also have zero means, a column consisting of all ones is included in the covariate matrix. Consequently, when the covariates are projected out of the GxE matrices (Note S1), it guarantees that all columns have zero means.

Computational challenges

Computing the coefficients of the system of linear Equation 4 presents computational challenges. The main computational bottleneck is the evaluation of the quantities T_{ij} for $i, j \in \{0, \dots, 2L + 1\}$, which requires $\mathcal{O}(N^2 ML)$. Therefore, the total time complexity for exact MoM is $\mathcal{O}(N^2 ML + L^3)$, imposing challenging memory or computation requirements for Biobank-scale data (N in the hundreds of thousands, M in the millions, and L in the hundreds or thousands).

Scalable estimation

Instead of computing the exact value of T_{ij} , GENIE uses a randomized estimator of the trace.¹⁹ This estimator uses the fact that for a given $N \times N$ matrix \mathbf{C} , $\mathbf{w}^T \mathbf{C} \mathbf{w}$ is an unbiased estimator of $\operatorname{tr}(\mathbf{C})$ ($E[\mathbf{w}^T \mathbf{C} \mathbf{w}] = \operatorname{tr}(\mathbf{C})$ where \mathbf{w} is a random vector with mean zero and covariance \mathbf{I}_N). Hence, we can estimate the values T_{ij} , $i, j \in \{0, \dots, 2L + 1\}$ as follows:

$$T_{ij} = \operatorname{tr}(\mathbf{Z}_i \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{Z}_j^T) \approx \widehat{T}_{ij} = \frac{1}{B} \sum_b \mathbf{w}_b^T \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{Z}_j^T \mathbf{w}_b \quad (\text{Equation 6})$$

Here, $\mathbf{w}_1, \dots, \mathbf{w}_B$ are B independent random vectors with zero mean and covariance \mathbf{I}_N . In GENIE, we draw these random vectors independently from a standard normal distribution. Note that computing T_{ij} by using the above estimator involves matrix-vector multiplications, which are repeated B times. Therefore, the total running time is $\mathcal{O}(LNMB)$.

Moreover, we can leverage the structure of the genotype matrix, which only contains entries in $\{0, 1, 2\}$. For a fixed genotype matrix \mathbf{X}_k , we can improve the per iteration time complexity of matrix-vector multiplication from $\mathcal{O}(NM)$ to $\mathcal{O}\left(\frac{NM}{\max(\log_3(N), \log_3(M))}\right)$ by using the Mailman algorithm.²⁰ Solving the normal equations takes $\mathcal{O}(L^3)$ time so that for a small number of components (L), the overall time complexity of our algorithm is $\mathcal{O}\left(\frac{LNMB}{\max(\log_3(N), \log_3(M))} + L^2(NB + L)\right)$.

Standard errors of the estimates

We used a computationally efficient block jackknife²¹ to compute standard errors of the estimates, which does not require any assumptions on the distribution of the effect sizes. Each jackknife subsample was created by removing a block of the genotype matrix, and we approximated the true SE by the jackknife estimate. Specifically, if we partition the genotype \mathbf{X} into J non-overlapping blocks $[\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(J)}]$, $\widehat{SE} = \sqrt{\frac{J-1}{J} \sum_j (\widehat{h}_{(j)}^2 - \widehat{h}_{jack}^2)^2}$, where $\widehat{h}_{(j)}^2$ is the heritability estimate based on $\mathbf{X}^{(-j)}$ (removing $\mathbf{X}^{(j)}$ from \mathbf{X}), and \widehat{h}_{jack}^2 is the mean of estimates across J jackknife subsamples. The jackknife estimator was implemented efficiently in GENIE to compute the estimate in time $\mathcal{O}\left(\frac{LNMB}{\max(\log_3(N), \log_3(M))} + JL^2(NB + L)\right)$. In our analysis, we used $J = 100$ blocks defined over SNPs to compute the standard errors of the estimates.

Partitioning GxE heritability across the genome

Although the model defined in Equation 1 is beneficial in quantifying genome-wide GxE effects for a given E, it is interesting to identify and interpret the interaction of E with specific regions of the genome, such as SNPs with a particular range of minor allele frequencies or SNPs that lie within genes expressed specifically in a tissue. Following our previous work,²¹ the genotype component \mathbf{X} can be assigned to T (potentially overlapping) components with respect to a set of annotations (such as MAF/LD or functional annotations). Thus, we extend our model as follows:

$$\begin{aligned} \mathbf{y} &= \sum_{t=1}^T \mathbf{X}_t \beta_t + \sum_{t=1}^T \sum_{l=1}^L (\mathbf{X}_t \odot \mathbf{E}_{:l}) \alpha_{tl} + \sum_{l=1}^L (\mathbf{I}_N \odot \mathbf{E}_{:l}) \delta_l + \mathbf{C}\gamma + \epsilon \\ \beta_t &\sim \mathcal{D}\left(\mathbf{0}, \frac{\sigma_{s,t}^2}{M_t} \mathbf{I}_{M_t}\right) \\ \alpha_{tl} &\sim \mathcal{D}\left(\mathbf{0}, \frac{\sigma_{gxe,tl}^2}{M_t} \mathbf{I}_{M_t}\right) \\ \delta_l &\sim \mathcal{D}\left(\mathbf{0}, \sigma_{nxe,l}^2 \mathbf{I}_N\right) \\ \epsilon &\sim \mathcal{D}(\mathbf{0}, \sigma_e^2 \mathbf{I}_N) \end{aligned} \quad (\text{Equation 7})$$

Here, \mathbf{X}_t is the genotype of annotation t with M_t SNPs, and α_{tl} refers to the effect sizes of SNPs in annotation t in the context of environment l . Analogously, $\sigma_{gxe,tl}^2$ refers to the variance component for SNPs in annotation t in the context of environment l .

while $h_{gxe,t}^2$ refers to the GxE heritability associated with annotation t in the context of environment l .

Given estimated GxE heritabilities under the above model, we define the enrichment of genetic effects in annotation t in the context of environment l (also termed GxE enrichment) as follows:

$$Enrichment(gxe, t, l) = \frac{h_{gxe,t}^2 / \sum_{t=1}^T h_{gxe,t}^2}{M_t/M}, t \in \{1, \dots, T\}, l \in \{1, \dots, L\} \quad (\text{Equation 8})$$

Estimating GxE in the UK Biobank

We applied GENIE to the UKB⁸ where we considered environmental variables such as smoking status, sex, age, and statin medication. The analyses utilized the UKB Resource under application 331277, with participants' informed consents verified by the UKB.²² For every environmental variable, we applied GENIE to estimate additive heritability (h_g^2) and GxE heritability (h_{gxe}^2) across 50 quantitative phenotypes (in a model that included the environmental variable as a main effect and accounted for noise heterogeneity) (Table S2). In this study, we restricted our analysis to SNPs that were present in the UKB Axiom array used to genotype the UK Biobank. SNPs with greater than 1% missingness and MAF smaller than 1% were removed. Moreover, SNPs that failed the Hardy-Weinberg test at significance threshold 10^{-7} were removed. We restricted our study to self-reported British white ancestry individuals who are $> 3^{rd}$ degree relatives that are defined as pairs of individuals with kinship coefficient $< 1/2^{(9/2)}$.⁸ Furthermore, we removed individuals who are outliers for genotype heterozygosity and/or missingness. Finally, we obtained a set of $N = 291,273$ individuals and $M = 454,207$ SNPs for real data analyses. No LD pruning or filtering was required by GENIE subsequently.

We included age, sex, age², age \times sex, age² \times sex, and the top 20 genetic principal components (PCs) as covariates in our analysis for all traits. We always include the environmental variable as a covariate in these analyses. We used PCs precomputed by the UKB from a superset of 488,295 individuals. Additional covariates were used for waist-to-hip ratio (adjusted for body mass index [BMI]) and diastolic/systolic blood pressure (adjusted for cholesterol-lowering medication, blood pressure medication, insulin, hormone replacement therapy, and oral contraceptives). We standardized environmental variables in our primary analyses. The standardized coding for binary environmental variables has an invariant property in the sense that the covariance matrix would be the same regardless of flipping the 0/1 coding. We also considered the binary coding of environmental variables to be relevant. Statin usage is defined as a binary environmental variable based on C10AA (the American Therapeutic Chemical [ATC] code of statin), which corresponds to taking any subtype of statin medications. Smoking status is defined as a categorical variable with three possible values (never, previous, and current).

We considered an additional analysis of genotypes at high-quality imputed SNPs (with a hard call threshold of 0.2 and an INFO score ≥ 0.8) with MAF $\geq 0.1\%$ in the $N = 291,273$ unrelated white British individuals. We further restricted our analyses to SNPs that are under Hardy-Weinberg equilibrium ($p < 10^{-7}$) and are confidently imputed in more than 99% of the individuals. Additionally, we excluded SNPs in the MHC region, resulting in a total of $M = 7,774,235$ SNPs.

In our analysis of heritability partitioned based on MAF-LD annotations (primarily for the imputed SNPs), we divided SNPs into

eight annotations based on quartiles of the LD scores (computed in-sample using GCTA) and two MAF bins (MAF $< 5\%$ and MAF $\geq 5\%$). In our analyses of heritability partitioned based on tissue-specific gene expression annotations, we used the annotations for the 53 tissue-specific genes generated by Finucane et al.¹⁸ using a matrix of normalized gene expression values from the Genotype-Tissue Expression (GTEx) database, which included samples from various tissues, including the focal tissue. The authors calculated a t statistic for each gene to determine its specific expression in the focal tissue and ranked all genes based on their t -statistics. They defined the top 10% of genes with the highest t statistic as the set of specifically expressed genes for the focal tissue. To improve the accuracy of the gene set construction, 100-kb windows are added on either side of the transcribed region of each gene in the set of specifically expressed genes to generate a genome annotation that corresponds to the focal tissue.

Results

Calibration and power

We assessed the false positive rate of tests of GxE heritability based on GENIE in simulations under different genetic architectures with no GxE heritability. For each architecture, we simulated 100 phenotype replicates across $N = 291,273$ unrelated white British individuals in the UKB and $M = 454,207$ SNPs with MAF $> 1\%$ genotyped on the UKB genotyping array. We chose statin usage in the UKB as the environmental variable. We varied the percentage of causal SNPs while fixing the additive heritability at $h_g^2 = 0.25$. We ran GENIE with $B = 10$ random vectors (see the following section on the choice of the number of random vectors).

Across all simulations, the false positive rate of rejecting the null hypothesis of no GxE heritability is controlled at levels 0.05 and 0.05/200 (we consider this threshold, which controls for the number of trait-environmental variable [trait-E] pairs that we test in UKB): the average $P(\text{rejection at } p < t)$ is 7.5% and 0% for $t = 0.05$ and $t = 0.05/200$, respectively (Figure 1A).

To measure the power of GENIE to detect GxE heritability, we simulated phenotypes with a non-zero GxE heritability. Across genetic architectures, we varied the GxE heritability with no noise heterogeneity while fixing the additive heritability at 0.25 and the percentage of causal SNPs at 10% (these are default values of additive heritability and causal ratio across our simulations unless otherwise specified). We also tested GENIE by varying the sample size from 30,000 to 300,000. We simulated 100 replicates for every genetic architecture. Let $h_{gxe}^2(i)$ be the estimate of h_{gxe}^2 and SE_i be the jackknife estimate of the standard error on the i -th replicate for $i \in \{1, \dots, 100\}$. We computed the p value of a test of the null hypothesis of no h_{gxe}^2 on the i -th replicate from the Z score defined as $h_{gxe}^2(i)/SE_i$ for $i \in \{1, \dots, 100\}$. We reported the percentage of replicates with p value $< t$ as the power of GENIE on a given genetic architecture for a p value threshold of t .

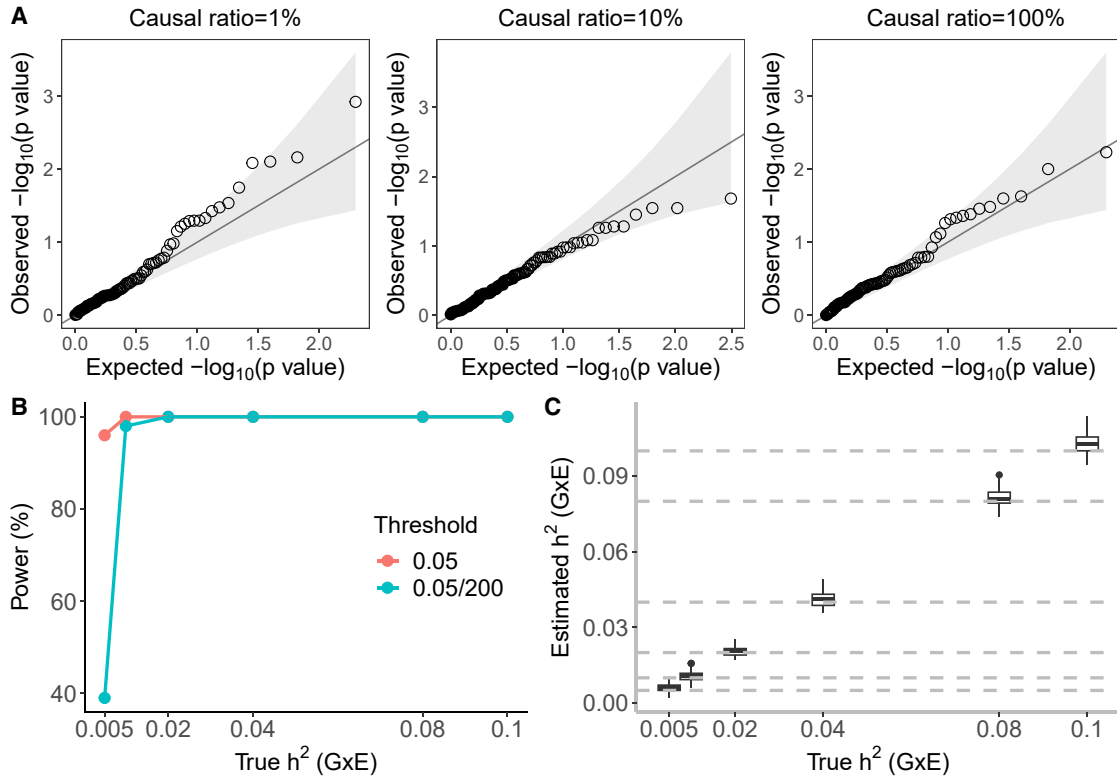


Figure 1. Calibration and power of GENIE in simulations ($N = 291,273$ unrelated individuals, $M = 454,207$ SNPs)

(A) Q-Q plot of p values (of a test of the null hypothesis of zero GxE heritability) when GENIE is applied to phenotypes simulated in the absence of GxE effects. Each panel contains 100 replicates of phenotypes simulated with additive heritability $h_g^2 = 0.25$ and varying proportions of causal variants. The causal ratios are the same for the G and GxE components (10%), and the causal SNPs for the GxE component are independently sampled to those for the additive genetic component. Across all architectures, the mean of $P(\text{rejection at } p < t)$ is 7.5% and 0% for $t = 0.05$ and $t = \frac{0.05}{200}$, respectively (7.5% is not significantly different from the nominal rate of 5%).

(B) The power of GENIE across genetic architectures as a function of GxE heritability. We report power for p value thresholds of $t \in \{0.05, \frac{0.05}{200}\}$.

(C) The accuracy of h_{gxe}^2 estimates obtained by GENIE. Across all simulations, statin usage in UKB was used as the environmental variable.

GENIE has adequate power to detect GxE effects with $h_{gxe}^2 \geq 0.005$ in a sample of 300,000 unrelated individuals at $p < 0.05$ (Figure 1B). The power increases from around 20% to 100% as the sample size grows from 30,000 to 300,000 when $h_{gxe}^2 = 0.01$ at $p < 0.05$ and remains almost 100% for $h_{gxe}^2 \geq 0.05$ as the sample size reaches 50,000 (Figure S2A). Additionally, GENIE yields unbiased estimates of GxE heritability (Figure 1C), and the SEs estimated by GENIE were concordant with the true SEs (Figure S3).

Next, we assessed the accuracy of GENIE in a setting with multiple environmental variables. We simulated phenotypes from a sub-sampled set of UKB genotypes, choosing a subset of $N = 10,000$ individuals and 20,000 SNPs on chromosome 1 of the UKB Axiom array. We considered a setting with $L = 10$ environmental variables with $\sigma_g^2 = 0.2$, five environmental variables with $\sigma_{gxe}^2 = 0$, three environmental variables with $\sigma_{gxe}^2 = 0.1$, and two with $\sigma_{gxe}^2 = 0.01$. We generated 100 replicates of simulated phenotypes for each set of parameters. We find that GENIE ob-

tains estimates of h_{gxe}^2 that are accurate across the environmental variables (Figure S1; Table S1).

Impact of randomization on GxE estimates

We investigated the impact of randomization on the estimates obtained by GENIE by comparing it to the exact MoM. Since exact MoM is computationally infeasible for large sample sizes, we choose to experiment on a small-scale dataset consisting of $N = 10,000$ unrelated white British individuals and $M = 60,000$ SNPs selected from the UK Biobank array SNPs on chromosome 1. We generated 100 replicates of phenotypes with no noise heterogeneity, $h_g^2 = 0.1$, and varying h_{gxe}^2 with standardized smoking status as the environment variable. We ran GENIE using the G + GxE + Nx E model with $B = 10$ random vectors and compared the estimated G and GxE heritability with the results from GCTA-HE regression¹¹ (exact MoM) on G and GxE GRM matrices. We see that exact MoM has a slightly higher statistical power than GENIE (with an increase in power of 2% to 8% across the values tested; Figure S4A). Further, the relative contribution of randomization to the

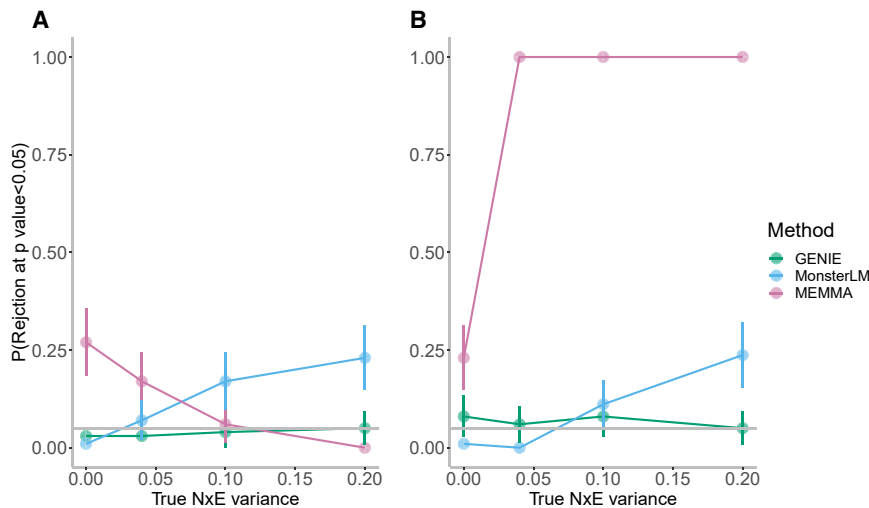


Figure 2. Comparisons of false positive rates with existing methods with the presence of noise heterogeneity

False positive rates for GxE heritability across GENIE, MEMMA, and MonsterLM using (A) continuous and (B) discrete environment exposures. We performed simulations with no GxE heritability but with varying magnitudes of the variance of the Nx/E effect. We computed the false positive rate as the fraction of rejections (p value of a test of the null hypothesis of zero GxE heritability < 0.05) over 100 replicates of phenotypes. The phenotypes were simulated from $N = 40,000$ individuals and $M = 223,591$ SNPs filtered from $M = 454,207$ SNPs with the genotype QC steps in MonsterLM: SNPs that failed the Hardy-Weinberg test at the significance threshold 10^{-10} were excluded, and highly correlated SNPs with LD $r^2 > 0.9$ and SNPs with MAF < 0.05 were removed. Error bars correspond to the estimated 95% CI of the rejection rate.

SE of GENIE remains around 30% despite the variation of power difference across simulations (Figure S4B).

Confirming that randomization makes a modest difference on the power of GENIE, we quantified the effect of the number of random vectors. We explored the choice of the number of random vectors in two ways. First, we quantified the contribution of randomization to the SE of the GxE estimator in GENIE. We simulated 100 phenotypes where $h_{gxe}^2 = 0$. We compared the SE of GxE estimates with $B = 10$ random vectors run 100 times over one of the replicates (the contribution of the randomization to the SE) to the SE of GxE estimates across 100 replicates to determine that, with $B = 10$, randomization contributes to about 30% of the total SE across various sample sizes (Figure S5). Second, we verified that our GxE estimates are highly correlated for the choice of random vectors $B = 10$ vs. $B = 100$ (Pearson's correlation $r = 0.99$; Figure S6). These results lead us to conclude that $B = 10$ random vectors provide stable estimates, and we use this setting in our remaining analyses.

Noise heterogeneity

Previous studies have shown that accounting for noise heterogeneity (Nx/E component) is essential to avoid false positives and inflation in estimates of GxE effects.^{13,14,23} To demonstrate the importance of modeling Nx/E, we simulated phenotypes in the presence of Nx/E effect such that $h_{gxe}^2 = \sigma_{nxe}^2 \in \{0, 0.04, 0.08, 0.10\}$ (we set σ_{nxe}^2 to 0.04 when $h_{gxe}^2 = 0$). We ran GENIE, in turn, with and without the Nx/E component. Across all simulations, the model that does not account for the Nx/E component ($G + GxE$) yields statistically significant upward bias in its GxE estimates (relative bias ranges from 2.5% to 69% across genetic architectures) while the model that fits a noise heterogeneity component ($G + GxE + Nx/E$) achieves unbiased estimates of GxE (Figure S7).

Comparison with existing methods in simulations

We compared the calibration of tests of GxE from GENIE with MEMMA¹⁶ and MonsterLM.¹⁷ GPLEMMA¹⁵ was excluded due to its focus on multiple environmental variables. We conducted the benchmark experiments on $M = 454,207$ SNPs from a subset of $N = 40,000$ unrelated white British individuals. To ensure a fair comparison with MonsterLM, which requires genotype QC steps, we filtered SNPs by removing those with high LD ($r^2 > 0.9$) and low MAF (MAF < 0.05), resulting in 223,591 SNPs (we report results for GENIE and MEMMA on unfiltered SNPs in Figure S8). We then simulated phenotypes with both continuous (cystatin-C) and discrete (statin usage) environmental variables on the filtered SNPs. In simulations with no GxE or Nx/E effects, MEMMA had inflated false positive rates while GENIE and MonsterLM were calibrated (Figure 2). The inflated false positive rate for MEMMA in the absence of the Nx/E effect can be explained by a bias in their estimates of the SE of the variance components (Figure S9). Under scenarios with noise heterogeneity, GENIE remained calibrated while MonsterLM displayed inflation in its false positive rate with increasing Nx/E variance for both continuous and discrete environment variables. MEMMA showed elevated false positive rates with discrete environment variables, and lower but still inflated false positives with continuous environmental variables (Figure 2).

Robustness of GENIE in simulations

We tested the robustness of GENIE by varying the correlation between the phenotype (Y) and the environment (E), simulating heritable E , imposing that the causal SNPs are the same for the G and GxE components, simulating Y that has the same causal SNPs with the heritable E , and simulating a collider bias scenario. In addition, we also considered a scenario where the environment noise is

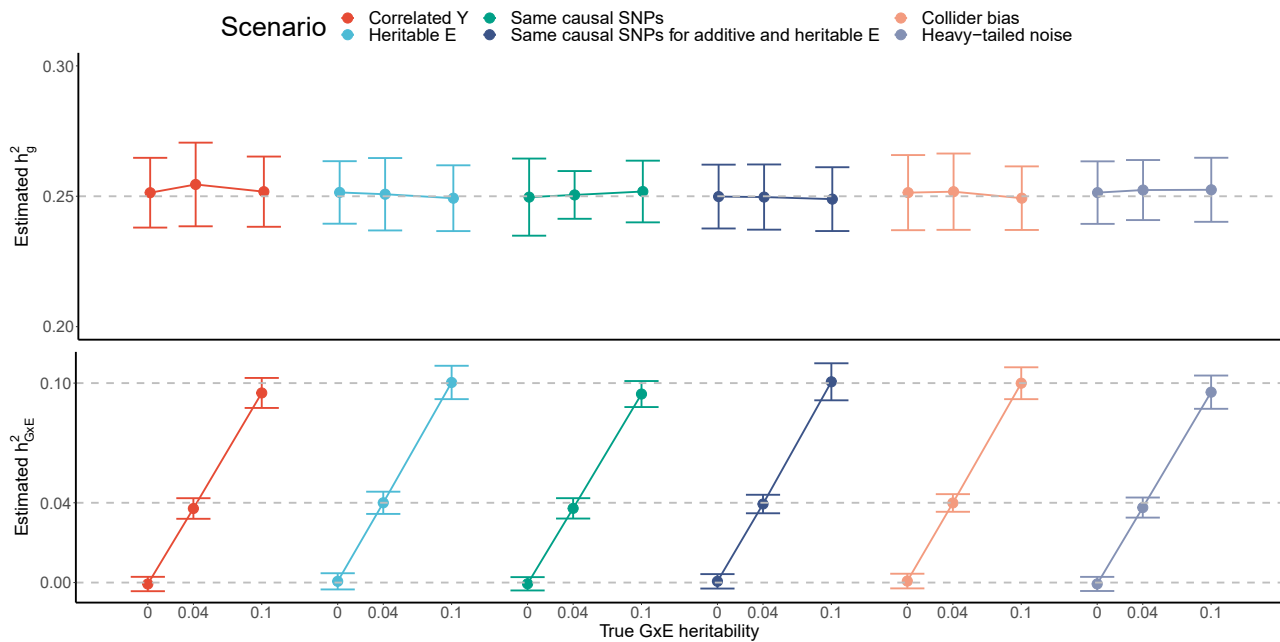


Figure 3. Estimation of G and GxE heritability in six simulated scenarios

We investigated the performance of GENIE in estimating G and GxE heritability under six simulated scenarios. (1) Correlated Y: the phenotypes were correlated with the continuous environment exposure, with Pearson's correlation $r = 0.5$; (2) heritable E: the environment exposure E was simulated from the same set of genotype data as in the phenotype simulation, with an additive genetic heritability of 0.1; (3) same causal SNPs: additive genetic causal SNPs completely overlap with GxE causal SNPs; (4) same causal SNPs for additive and heritable E: additive genetic causal SNPs completely overlap with the causal SNPs explaining heritability in E, where E is the same as in scenario (2); (5) collider bias: the phenotype Y and environment exposure E are correlated through an unobserved confounder; we simulated a heritable environment variable with a genetic heritability of 0.1. The phenotypes were then generated to have a Pearson's correlation $r = 0.2$ with the heritable E. We assumed that the correlation was due to an unobserved confounder.¹⁷ (6) Heavy-tailed noise: we drew the environment noise component from the Student's t-distribution with degrees of freedom = 4. In all scenarios, we simulated 100 replicates of phenotypes with Nx E and varying magnitude of GxE effects across $N = 291,273$ individuals genotyped at 454,207 SNPs. The ground truth GxE heritability was 0, 0.04, and 0.1, with corresponding Nx E variance of 0.04, 0.04, and 0.1. The additive genetic heritability was fixed at 0.25. The x and y axes denote the true GxE heritability and the estimated G and GxE heritability. Points and error bars represent the mean and estimated 95% CI, respectively. Across all simulations where there is no GxE, the mean of $P(\text{rejection at } p < t)$ are 5.5% and 0% for $t = 0.05$ and $t = 0.05/200$, respectively (5.5% is not significantly different from the nominal rate of 5%).

drawn from a heavy-tailed distribution (see [Note S3](#) for details). In these simulations, we use a continuous environmental exposure (to complement our previous set of simulations that used a discrete environmental exposure, i.e., statin usage). In scenarios where the environmental exposure is heritable, we simulated continuous environmental exposure with specific genetic architecture. In simulations where the environment exposure is not heritable, we use a continuous exposure measured in UKB (cystatin-C). In all simulations, we simulated phenotypes with Nx E and varying GxE effects across $N = 291,273$ individuals genotyped at 454,207 SNPs for 100 replicates. The results summarized in [Figure 3](#) indicate that GENIE obtains accurate estimates across these scenarios.

Computational efficiency

We evaluated the runtime of GENIE, MonsterLM, MEMMA, and GCTA(HE) (which implements an exact MoM estimator) with increasing sample size ($N \in \{10000, 50000, 100000, 290000\}$) for a fixed number of SNPs ($M = 454,207$) and a single environmental variable. All methods were run on an Intel(R) Xeon(R) Gold 6140 CPU 2.30GHz,

with 187GB RAM. Ten random vectors are used by GENIE and MEMMA. For GENIE, runtime measurements were obtained for the single component and eight MAF/LD components. All other methods fit a single G and GxE variance component. The runtime of GCTA(HE) includes the computation of the GRM matrix. Our comparison used the CPU implementation of MonsterLM, with runtime calculations excluding the preprocessing step for genotype filtering required by MonsterLM. GENIE is highly scalable and can estimate GxE on about 300,000 individuals and roughly 500,000 SNPs within an hour, with the eight-component model nearly as efficient as the single-component model ([Figure S11](#)).

Estimating GxE in the UKB

We applied GENIE to estimate additive heritability (h_g^2) and GxE heritability (h_{GxE}^2) for 50 quantitative phenotypes measured in UKB across unrelated white British individuals. These 50 phenotypes fall into eight broader phenotypic categories (blood biochemistry, kidney biomarkers, anthropometry, lipid metabolism biomarkers, blood pressure, liver biomarkers, lung, and glucose metabolism

biomarkers) that have been analyzed in prior works.^{24–26} Following these studies, we applied a rank-based inverse normal transformation to all phenotypes. For certain phenotypes affected by medication usage (systolic/diastolic blood pressure, LDL direct, and total cholesterol), we adopted heuristic adjustments for medication variables.^{24,27} We then reevaluated the GxE heritability estimates using GENIE (see [Note S4](#) for details). We considered, in turn, smoking status, sex, age, and statin usage as environmental variables. We included each environmental variable as a fixed effect in the relevant analyses. First, we explored the importance of modeling NxE in real data (building on our simulation results). We then analyzed, in turn, common SNPs genotyped on the UKB array (MAF > 1%) and then common and low-frequency imputed SNPs (MAF ≥ 0.1%). For selected combinations of phenotypes and environmental variables, we also applied GENIE to partition GxE heritability across functional annotations to estimate GxE heritability in genes expressed in specific tissues.

We note that individuals with missing environmental or phenotype data were removed in the implementation of GENIE instead of being imputed by the mean value. We observed that the application of mean imputation to the phenotype results in underestimation of h_g^2 and h_{gxe}^2 while mean imputation of the environment variables affected the estimation of h_{gxe}^2 but not h_g^2 ([Figure S12](#)). We therefore recommend that users leave missing exposure and outcome data as it is when applying GENIE in their analysis based on the simulation results.

Robustness of GENIE in the UKB

We first assessed the robustness of GENIE by estimating h_g^2 under three different models: G, G + GxE, and G + GxE + NxE, where each model is named by the set of variance components fitted jointly. The additive heritability estimates were highly correlated across the models (Pearson's correlation $r \geq 0.98$ for every pair of models), leading us to conclude that GENIE provides robust estimates of additive heritability across different models ([Figure S13](#)). We observed a significant difference in h_g^2 for a handful of trait-E pairs when estimated with G + GxE and G + GxE + NxE that include alcohol frequency intake and overall health with smoking status, sex, or age as the environmental variable. In previous work,²¹ we compared the additive h_g^2 estimates from RHE with S-LDSC,²⁸ GRE,²⁹ SumHer,³⁰ and LDSC³¹ to find that RHE estimates of additive heritability for 22 complex traits are consistent with the existing methods. We additionally compared the additive heritability estimates from GENIE with those obtained using LDSC (run with in-sample LD scores estimated from a subset of 50 K unrelated white British individuals in UKB). The estimates of additive h_g^2 from LDSC were compared against those from GENIE with environmental exposures of smoking status, sex, age, and statin.

The estimates across 50 traits were consistently correlated for the two methods, with Pearson's correlations ranging from 0.87 to 0.93 ([Figure S14](#)).

Our simulations in the previous section revealed the importance of modeling noise heterogeneity ([Figure S7](#)). To investigate the consequences of modeling NxE in real data, we fitted, in turn, models without and with NxE (in addition to G and GxE components). The number of trait-E pairs with significant h_{gxe}^2 ($p < 0.05/200$) decreased from 135 under the G + GxE model to 68 under the G + GxE + NxE model: changing from 40 to 21 for smoking ([Figure 4B](#)), 27 to 28 for sex ([Figure S15B](#)), 28 to 12 for age ([Figure S16B](#)), and 40 to 7 for statin usage ([Figure S17B](#)). For traits with significant h_{gxe}^2 , the magnitudes of the estimates varied across the two models: the ratios of h_{gxe}^2 estimates under the G + GxE + NxE to the G + GxE model were 137% on average (range: 43% – 350%), 110% (70% – 224%), 131% (99% – 166%), and 42% (21% – 72%) for smoking ([Figure 4A](#)), sex ([Figure S15A](#)), age ([Figure S16A](#)), and statin ([Figure S17A](#)), respectively. The magnitude of noise heterogeneity across trait-E pairs can be substantial: 0.05%, 164%, 10%, and 14% of the additive heritability on average for smoking, sex, age, and statin, respectively ([Figures S18–S21](#)). To further investigate the effect of modeling NxE, we performed permutation analyses by randomly shuffling the genotypes while preserving the trait-E relationship (a setting where there is expected to be no GxE by construction while the relationship between phenotype and E is preserved). We applied GENIE under the G + GxE and G + GxE + NxE models to each trait-E pair. The false positive rate of rejecting the null hypothesis of no GxE across the trait-E pairs is substantially inflated under the G + GxE model while being controlled under the G + GxE + NxE model ([Figures 4C, S15C, S16C, and S17C](#) for smoking, sex, age, and statin respectively). These results indicate that modeling NxE is critical to avoid spurious findings of GxE.

Gene-by-smoking interaction

We applied GENIE to estimate the proportion of phenotypic variance explained by gene-by-smoking interactions ($h_{gxSmoking}^2$) for 50 quantitative phenotypes. We find 21 traits showing statistically significant evidence for $h_{gxSmoking}^2$ ($p < 0.05/200$) with $h_{gxSmoking}^2$ about 6.1% of h_g^2 on average ([Figures 5A and 6A](#)). Two of the traits with the largest $h_{gxSmoking}^2$ were basal metabolic rate and BMI with estimates of 2.4% and 2.3%, respectively (estimates remained significant when we used the binary coding of the smoking status variable obtained by merging the categories of never and previous; [Figures S25 and S28C](#)). Our estimates are consistent with a previous study that analyzed BMI and lifestyle factors in the UKB to find significant GxE for smoking behavior.⁵ The $h_{gxSmoking}^2$ estimates for basal metabolic rate and BMI are about 11% and 7% of their respective h_g^2 estimates.

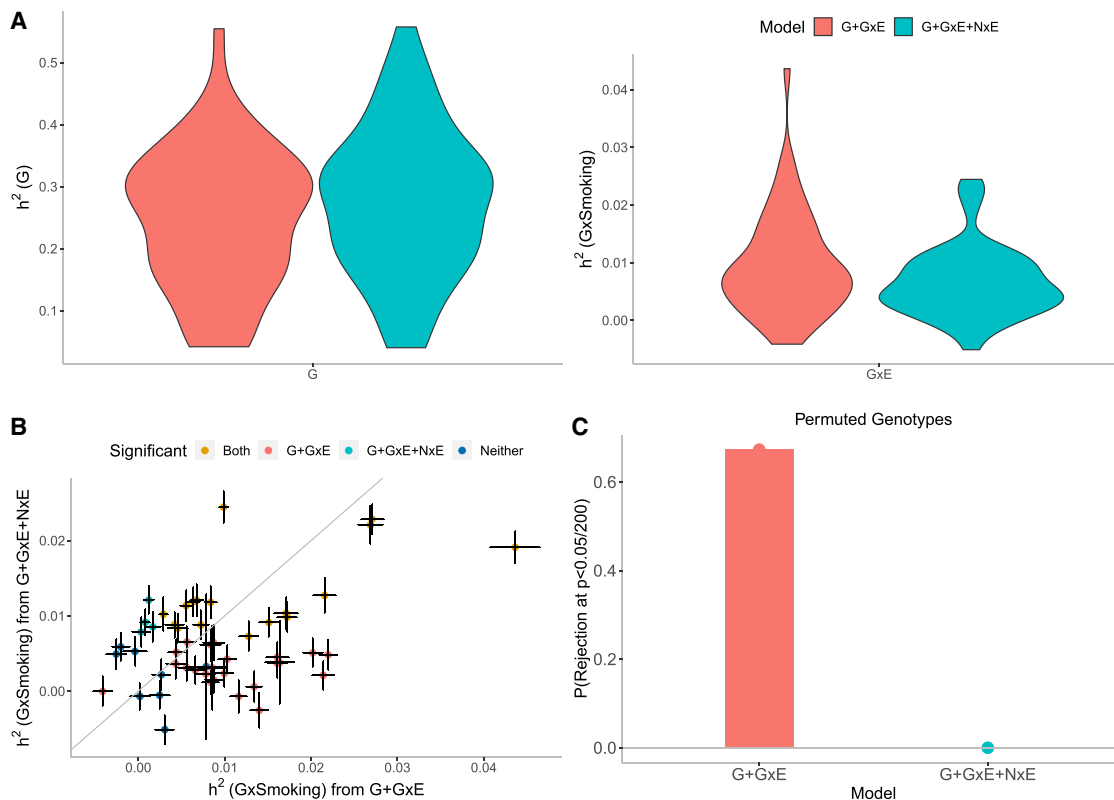


Figure 4. Effect of noise heterogeneity (Nx) on estimates of heritability associated with GxSmoking across 50 quantitative phenotypes in UKB

Model G + GxE refers to a model with additive and gene-by-environment interaction components where the environmental variable is smoking status. Model G + GxE + Nx refers to a model with additive, gene-by-environment interaction, and noise heterogeneity (noise-by-environment interaction) components.

(A) We ran GENIE under G + GxE and G + GxE + Nx models to assess the effect of fitting an Nx component on the additive and GxE heritability estimates.

(B) Comparison of GxE heritability estimates obtained from GENIE under a G + GxE + Nx model (x axis) to a G + GxE model (y axis). Black error bars mark \pm standard errors centered on the estimated GxE heritability. The color of the dots indicates whether estimates of GxE heritability are significant under each model.

(C) We performed permutation analyses by randomly shuffling the genotypes while preserving the trait-E relationship and applied GENIE in each setting under G + GxE and G + GxE + Nx models. We report the fraction of rejections $P(p \text{ value of a test of the null hypothesis of zero GxE heritability} < \frac{0.05}{200})$ that accounts for the number of phenotypes tested) over 50 UKB phenotypes.

Gene-by-sex interaction

We find 28 traits with statistically significant h^2_{gxSex} ($p < 0.05/200$) with h^2_{gxSex}/h^2_g observed to be 8.7% on average (Figures 5B and 6B). Serum testosterone levels showed the largest h^2_{gxSex} of 11% with the h^2_{gxSex} nearly as large as h^2_g consistent with prior work showing differences in genetic associations^{32,33} and heritability³⁴ across males and females. Beyond testosterone, we observe significant h^2_{gxSex} for several anthropometric traits, such as waist-hip-ratio (WHR) adjusted for BMI ($h^2_{gxSex} = 4.3\%$ and $\frac{h^2_{gxSex}}{h^2_g} = 20\%$), and lipid measures (results consistent for binary encoding; Figures S26 and S28B) consistent with previous work documenting sex-specific differences in the genetic architecture of anthropometric traits.^{34–39} Consistent with prior GWAS that identified genetic variants with sex-dependent effects,^{40,41} our analyses of serum urate levels show substantial point estimates of

h^2_{gxSex} , although these estimates are not statistically significant.

Gene-by-age interaction

We find 12 traits with statistically significant h^2_{gxAge} ($p < 0.05/200$) with h^2_{gxAge}/h^2_g observed to be 4.3% on average (Figures 5C and 6C). Lipid and blood pressure measures show some of the largest h^2_{gxAge} (about 2.5% for LDL and total cholesterol and 1.9% for diastolic blood pressure). Previous studies have found genetic variants in *SORT1* to have age-dependent effects on LDL cholesterol⁴² and nominal evidence for age-dependent genetic effects on blood pressure regulation.⁴³ We find that BMI shows evidence for significant h^2_{gxAge} while WHR does not, expanding on prior work that identified age-dependent genetic variants for BMI but not for WHR in genome-wide association studies (GWASs).³⁶ Interestingly, we used a

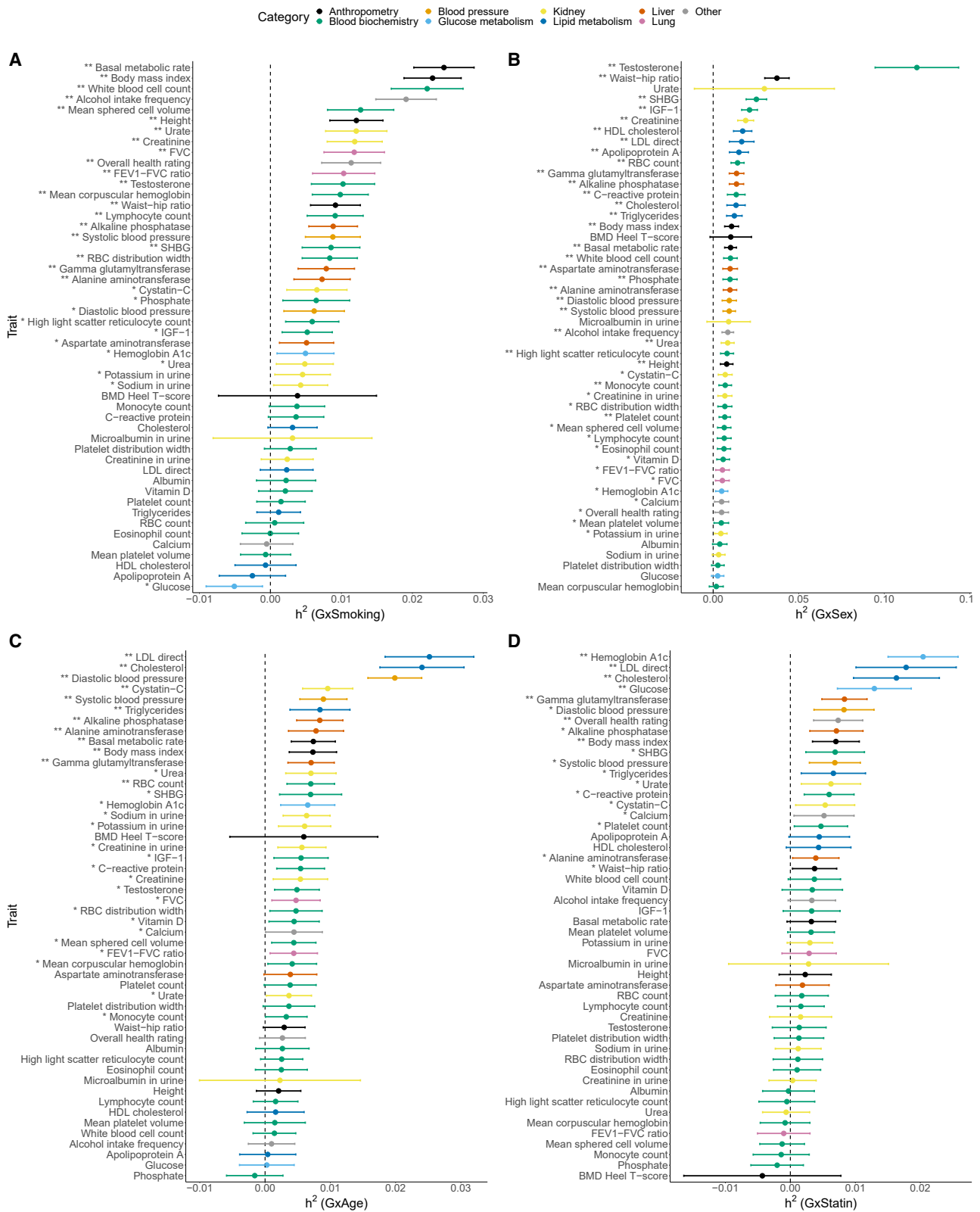


Figure 5. Estimates of GxE heritability across phenotypes in UKB

Estimates of (A) GxSmoking, (B) GxSex, (C) GxAge, and (D) GxStatin heritability across 50 UKB phenotypes. We applied GENIE to $N = 291,273$ unrelated white British individuals and $M = 454,207$ array SNPs ($MAF \geq 1\%$). Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. The environmental variable is standardized in these analyses. Error bars mark ± 2 standard errors centered on the point estimates. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

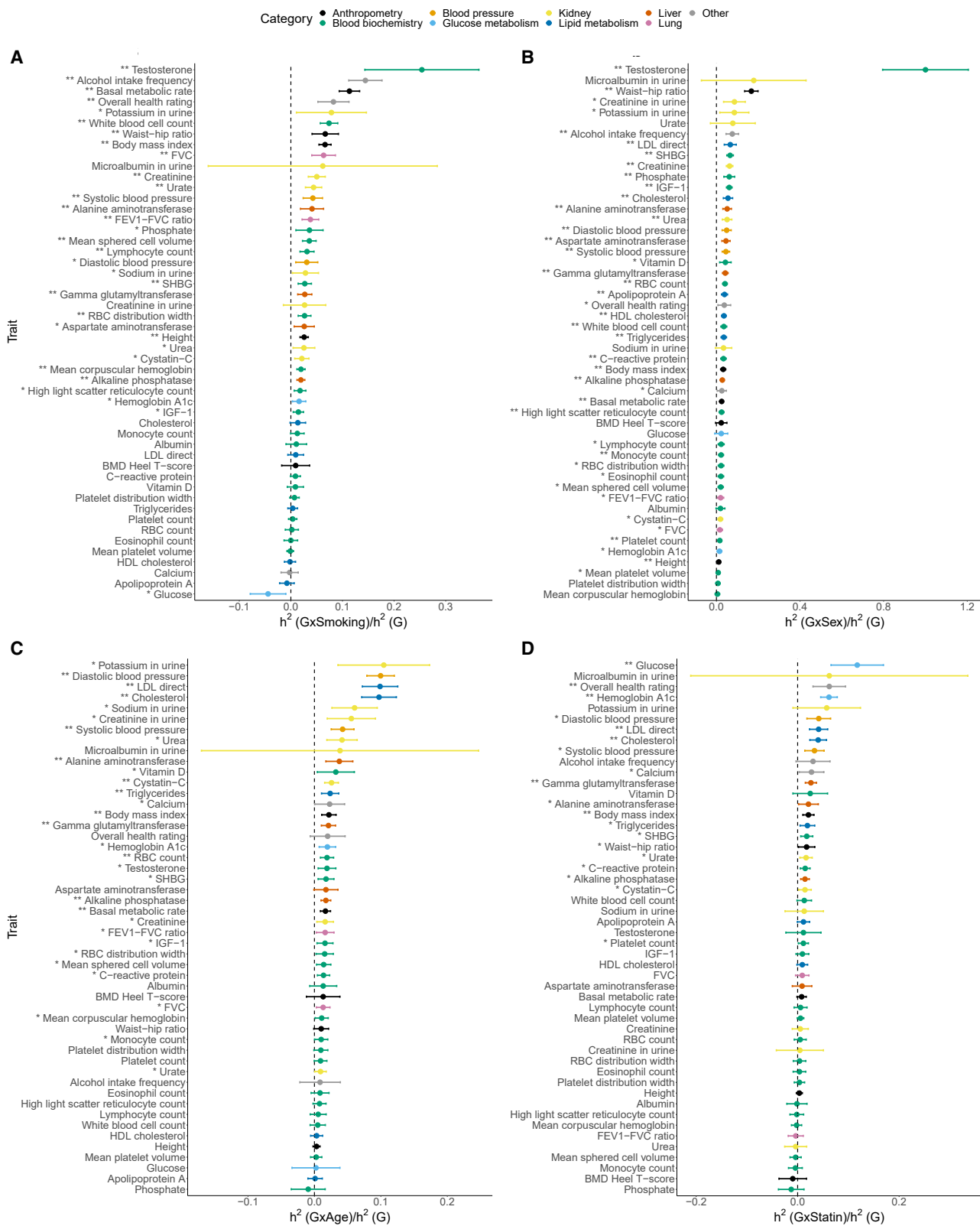


Figure 6. Estimates of the ratio of GxE to additive heritability across phenotypes in UKB

Estimates of the ratio of (A) GxSmoking, (B) GxSex, (C) GxAge, and (D) GxStatin to additive heritability across 50 UKB phenotypes. Error bars mark ± 2 standard errors centered on the point estimates. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

standardized encoding of age so that GxAge effects capture the interaction of genetic effects on the phenotype as a function of deviation from the mean age in UKB while previous studies typically focus on changes in genetic effects in bins of age. It is plausible that other codings of age, e.g., coding age to measure interactions as a function of older vs. younger individuals, could yield differing results.

Gene-by-statin interaction

We find seven traits that show statistically significant evidence for $h_{gxStatin}^2$ ($p < 0.05/200$) with an average ratio of $h_{gxStatin}^2$ to h_g^2 across traits of 5.2% (Figures S5D and S6D). We find that LDL and total cholesterol show significant $h_{gxStatin}^2$ (1.7% and 1.6% respectively) while HDL cholesterol with a point estimate of $h_{gxStatin}^2$ of 0.4% does not (results consistent for binary encoding; Figures S27 and S28A). We observe the largest estimates of $h_{gxStatin}^2$ for HbA1c and blood glucose measurements (2% and 1.2% respectively), which are interesting in light of statin usage being shown to be associated with a small increase in risk for type 2 diabetes.⁴⁴

GxE heritability estimates stratified by sex

Quantitative measurements like testosterone concentrations are strongly determined by sex, and therefore, one might be concerned with the possibility of collider bias in h_{gxe}^2 estimates on the whole population for these sex-determined traits. To address this issue, we repeated our previous analyses to estimate GxSmoking, GxAge, and GxStatin in females and males separately across the 50 traits. The results show that the sex-specific GxE heritability estimates are overall consistent with the results on all individuals (Pearson's correlations ranging from 0.67 to 0.80). By comparing GxE heritability estimates between female and male individuals, we noted Pearson's correlations of 0.50, 0.61, and 0.40 for GxSmoking, GxAge, and GxStatin, respectively (Figures S22–S24). In terms of the GxE heritability of testosterone specifically, we see that $\frac{h_{gxSmoking}^2}{h_g^2}$ is no longer significant for testosterone in female and male individuals (Figure S22) while estimates of $h_{gxSmoking}^2$ overlap with the previous results: (–0.82%, 0.97%) and (–0.71%, 1.37%) in females and males, respectively, and (0.58%, 1.47%) in the whole population. Hence, the attenuation of our estimates could be explained by the possibility of collider bias or a reduction in power. In general, the phenotypes that have the most significant GxE interactions are in the categories of anthropometry and blood biochemistry for GxSmoking, blood pressure and glucose metabolism for GxAge, and glucose metabolism and lipid metabolism for GxStatin in the sex-stratified analyses. In particular, GxSmoking estimates on BMI, basal metabolic rate, and white blood cell count remain significant for both males and females under $p < 0.05/200$. The differences in the GxE estimates between males and females

could suggest the presence of sex-specific GxE interaction effects.

Comparison with existing methods on significant trait-E pairs

We compared GxE heritability estimates of MEMMA, MonsterLM, and GENIE on real UKB phenotypes. While the consistency of GxE estimates from methods based on different model assumptions can enhance our confidence in the results, such comparisons have inherent limitations—our simulations have revealed variations in false positive rates among different methods. With these caveats, we evaluated GxE heritability using MonsterLM and MEMMA on 68 significant trait-E pairs detected by GENIE ($p < 0.05/200$). We noted Pearson's correlation $r = 0.91$ between the point estimates of GENIE and MonsterLM and 0.24 between GENIE and MEMMA across the 68 trait-E pairs (Figure S10). The closer alignment between the point estimates by GENIE and MonsterLM can be attributed to the shared consideration of noise heterogeneity within both models.

Estimating GxE heritability from imputed SNPs

We applied GENIE to estimate $h_{gxSmoking}^2$, h_{gxSex}^2 , h_{gxAge}^2 , and $h_{gxStatin}^2$ attributable to $M = 7,774,235$ imputed SNPs with $MAF \geq 0.1\%$. Prior work has shown that analyzing common and low-frequency variants with a single variance component can result in biased estimates of additive heritability.^{45,46} A solution to this problem involves fitting multiple variance components obtained by partitioning SNPs based on their frequency and local LD scores (as quantified by the LD scores³¹ or the LDK scores⁴⁵).^{30,46–48} We follow this approach by partitioning SNPs into eight annotations based on quartiles of the LD scores and two MAF annotations ($MAF < 5\%$ and $MAF > 5\%$; material and methods).

We performed simulations to show that GENIE applied with SNPs partitioned based on MAF and LD scores can accurately estimate h_{gxe}^2 across varying MAF and LD-dependent genetic architectures while using a single component for all SNPs can lead to substantial biases (Note S2, Figure S29). We applied GENIE using MAF-LD partitions to jointly estimate h_g^2 and h_{gxe}^2 (Figures S30–S33). While estimates of h_{gxe}^2 from imputed SNPs are largely concordant with the estimates obtained from array SNPs, we identify nine trait-E pairs for which the h_{gxe}^2 estimates are significantly different ($p < 0.05/200$). In all these cases, h_{gxe}^2 estimates from imputed SNPs are higher than those from array SNPs. For example, we estimated $h_{gxSmoking}^2$ for BMI = $6.5 \pm 0.5\%$, which is larger than our estimate based on array SNPs as well as a previous estimate of $4.0 \pm 0.8\%$ based on common HapMap3 SNPs.⁵ Across all trait-E pairs, we observed that the average ratio ($\frac{h_{gxe}^2(\text{imputed})}{h_{gxe}^2(\text{array})}$) is 1.17 (1.66, 1.23, 0.71, and 1.17, respectively, for GxSmoking, GxSex,

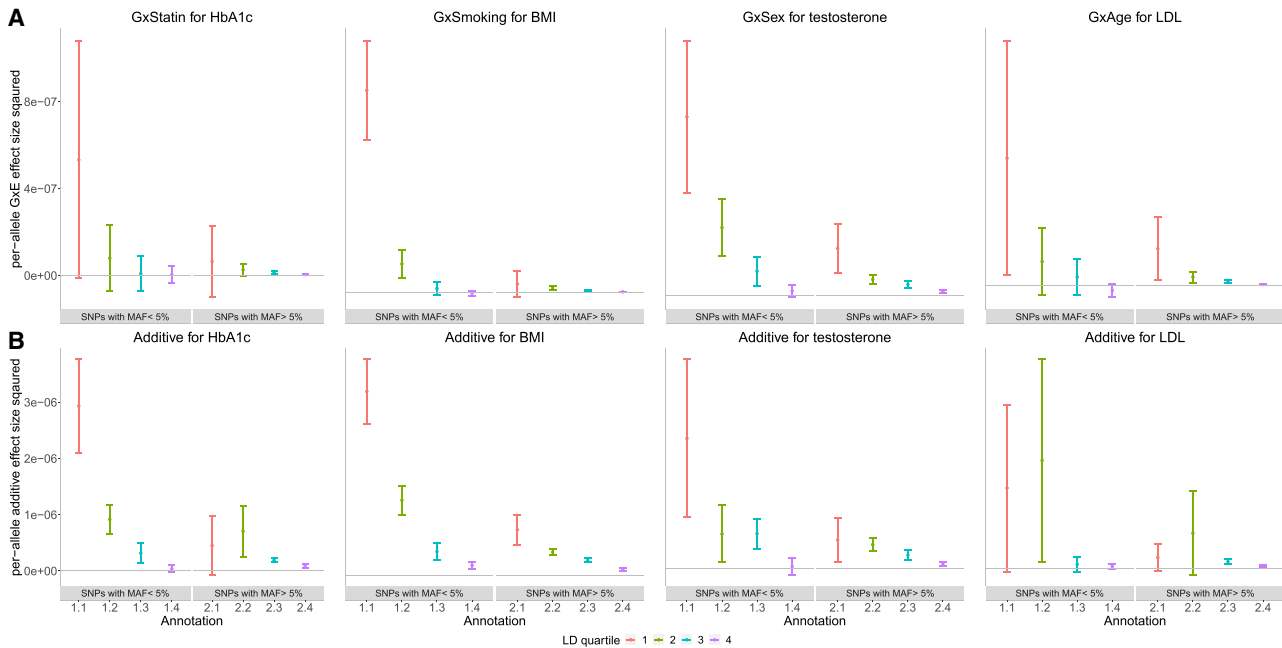


Figure 7. Per-allele squared GxE and additive effect sizes as a function of MAF and LD

(A) The squared per-allele GxE effect size for four selected pairs of trait and environment (trait-E pairs).

(B) The squared per-allele additive effect size for the same trait-E pairs. The x axis corresponds to MAF-LD annotations where annotation ij includes SNPs in MAF bin i and LD quartile j where MAF bin 1 and MAF bin 2 correspond to SNPs with $MAF \leq 5\%$ and $MAF > 5\%$, respectively, while the first quartile of LD scores correspond to SNPs with the lowest LD scores respectively). The y axis shows the per-allele GxE (or additive) effect size squared defined as $\frac{h_k^2}{2M_k f_k (1-f_k)}$ where h_k^2 is the GxE (or additive) heritability attributed to bin k , M_k is the number of SNPs in bin k , and f_k is the mean MAF in bin k . Error bars mark ± 2 standard errors centered on the estimated effect sizes.

GxAge, and GxStatin; Figure S34). Across trait-E pairs with significant h_{gxe}^2 , the average h_{gxe}^2 is 2.8% on the imputed data compared to 1.5% on array data while the ratio of $\frac{h_{gxe}^2}{h_s^2}$ is 14.3% on the imputed data compared to 6.8% on the array data (averaged across trait-E pairs, we estimated $h_{gxe}^2 = 0.9\%$ on imputed vs. 0.7% on array data).

We explored the impact of fitting multiple variance components based on MAF and LD by applying GENIE to fit a single GxE and additive variance component using smoking status as the environmental variable. While ten traits showed significant $h_{gxSmoking}^2$ in both analyses, five traits were exclusively significant in the MAF-LD model while one was exclusively significant in the single-component model. Restricting to traits with significant GxSmoking in both models, $h_{gxSmoking}^2$ estimates in the MAF-LD model were about three times those from the single-component model on average (Figure S35). We also investigated whether MAF-LD partitioning affected estimates of $h_{gxSmoking}^2$ obtained from array SNPs. We find that $h_{gxSmoking}^2$ estimates are largely concordant whether obtained from a single component or an MAF-LD partitioned model (ratio of 0.99 on average) consistent with the array SNPs being relatively common ($MAF > 1\%$). Our analysis suggests that partitioning by MAF and LD is helpful for estimating h_{gxe}^2 from both common and low-frequency SNPs and the inclusion low-frequency SNPs can increase estimates of h_{gxe}^2 for specific traits.

Partitioning GxE heritability across MAF and LD annotations

Previous studies have shown that the additive SNP effects increase with decreasing MAF and local levels of LD^{21,49–51} likely due to the effects of negative selection. Similar to previous analyses,^{15,17} we explored the MAF-LD dependence of SNP effects in the context of specific environmental factors. Our analyses in the preceding section, showing differences in the genome-wide h_{gxe}^2 estimates when partitioning by MAF and LD vs. fitting a single variance component, suggest that GxE effects are expected to vary by MAF and LD in a pattern that is distinct from what would be expected when fitting a single variance component, which assumes that the effect size at a SNP varies with its allele frequency f as $\frac{1}{f(1-f)}$ while not varying with local LD (for a fixed value of the allele frequency f). To explore the MAF-LD dependence of GxE effects, we used GENIE to partition h_{gxe}^2 across MAF and LD annotations (while simultaneously partitioning additive heritability) of $M = 7,774,235$ imputed SNPs divided into eight annotations based on quartiles of LD-scores and two MAF bins (low-frequency bins with $MAF < 5\%$ and high-frequency bins with $MAF \geq 5\%$). Within each of these eight bins, we defined the per-allele squared effect size as $\beta_k^2 = \frac{h_k^2}{2M_k f_k (1-f_k)}$ where h_k^2 is the GxE (or additive) heritability attributed to bin k , M_k is the number of SNPs in bin k , and f_k is the mean MAF in bin k .

For the sake of presentation, we selected one phenotype with high genome-wide GxE heritability for each of the four environmental variables analyzed (Figure 7; see Table S4 for results on all trait-E pairs). Across bins of MAF and LD, the magnitude of additive allelic effects tends to be larger than those of the GxE effects consistent with the genome-wide results. We observed that the per-allele squared GxE effect size β_{gxe}^2 tends to increase with lower MAF within a given quartile of LD score and to increase with lower bins of LD score for a fixed MAF bin (Figure 7A). These trends are analogous to the relationship observed for additive per-allele effect sizes (Figure 7B). Across the trait-E pairs, restricting to the lowest quartile of LD scores, low-frequency SNPs tend to have higher per-allele GxE effect sizes compared to high-frequency SNPs: the ratio of β_{gxe}^2 in low vs. high MAF bins is 8.2 ± 11.2 , 24.6 ± 19.7 , 3.4 ± 2.1 , and 3.7 ± 1.2 for HbA1c-statin, BMI-smoking, LDL-age, and testosterone-sex, respectively. In the highest quartile of LD scores, we found no statistically significant differences in β_{gxe}^2 across low and high MAF SNPs in any of the four trait-E pairs (we also plot the per-standardized genotype additive and GxE heritability, $\frac{h_g^2}{M_k}$, in Figure S36).

Partitioning GxE heritability across tissue-specific genes

The ability of GENIE to simultaneously estimate multiple, potentially overlapping, additive and GxE variance components enables us to explore how h_{gxe}^2 is localized across the genome. Specifically, we set to answer the question of whether h_{gxe}^2 is enriched in genes specifically expressed in a given tissue as a means to identify tissues that are relevant to a trait in a specific environmental context.

We applied GENIE to estimate h_g^2 and h_{gxe}^2 across each of 53 sets of genomic annotations defined as regions around genes that are highly expressed in a specific tissue in the GTEx dataset¹⁸ (Table S3). For each of the four environmental variables, we analyzed only traits with genome-wide significant h_{gxe}^2 based on our prior analyses of the array SNPs. For every set of tissue-specific genes, we followed prior work¹⁸ by jointly modeling the tissue-specific gene annotation as well as 28 genomic annotations that are part of the baseline LDSC annotations that include genic regions, enhancer regions, and conserved regions.²⁸ Specifically, our model has 29 additive variance components and 29 GxE variance components and estimates the additive and GxE heritability that can be attributed to genes specifically expressed in a tissue while controlling for the effects of the background annotations. A positive $h_{g,tissue}^2$ represents a positive contribution of genetic effects in a tissue to additive heritability.¹⁸ Analogously, a positive $h_{gxe,tissue}^2$ represents a positive contribution of genetic effects in this tissue to trait heritability in the context of the specific environment. We test estimates of $\frac{h_{gxe,tissue}^2/h_{g,tissue}^2}{M_{tissue}/M_{total}}$

$\left(\frac{h_{g,tissue}^2/h_{g,total}^2}{M_{tissue}/M_{total}}\right)$ to answer whether a tissue of interest is en-

riched for GxE (additive) heritability conditional on the remaining genomic annotations included in the model.

We first verified that our approach is able to detect previously reported enrichments for additive effects such as brain-specific enrichment for BMI and adipose-specific enrichment for WHR (Figure 8).¹⁸ Across 68 trait-E pairs with significant genome-wide GxE that we tested, we observed significant enrichment of $h_{gxe,tissue}^2$ (FDR < 0.10) for at least one tissue in five trait-E pairs (we plot four of these pairs in Figure 8 since the results from the fifth LDL-age are highly correlated with cholesterol-age). Across these trait-E pairs, we documented differential patterns of enrichments for GxE effects compared to additive effects. BMI exhibits brain-specific enrichment of $h_{gSmoking}^2$ and h_g^2 while WHR exhibits enrichment of h_{gSex}^2 and h_g^2 in adipose and breast tissue (in addition to the enrichment of h_g^2 in the uterus and cardiovascular tissues). The adipose-tissue-specific enrichment of h_{gSex}^2 in WHR is notable in light of known instances of genes associated with WHR in adipose tissue in a sex-dependent manner. *ADAMTS9*, a gene involved in insulin sensitivity,³⁵ is specifically expressed in adipose tissue and has been shown to be located near GWAS hits for WHR that are specific to females.^{35,36,52} The transcription factor, *KLF14*, is located near a sex-dependent GWAS variant for WHR, type 2 diabetes, and multiple other metabolic and anthropometric traits.⁵³ Further, the expression level of this gene is associated with the GWAS variant in adipose but not with other tissues.⁵³ We also found instances where tissues that are enriched for h_{gxe}^2 are distinct from those that are enriched for h_g^2 . We observed that the enrichment of h_{gSex}^2 for basal metabolic rate in brain and adipose tissues is distinct from the tissues that are enriched in h_g^2 for the same trait (cardiovascular and digestive tissues) (Figure 8). Finally, we find suggestive evidence that the liver is the most enriched tissue for $h_{gStatin}^2$ in HbA1c ($p = 0.02$) as well as for h_{gSex}^2 in testosterone ($p = 0.005$), although neither enrichment is significant at FDR of 0.10. These enrichments recapitulate known biology: the liver-specific enrichment of GxStatin effects for HbA1c reflect the tissues in which the target of statins (HMG-CoA-reductase) is expressed⁵⁴ while the liver-specific enrichment of GxSex for testosterone is consistent with previous findings implicating *CYP3A7*, a gene involved in testosterone metabolism that is specifically expressed in the liver and lies within a locus that contains one of the strongest GWAS signals for serum testosterone in females.³²

Discussion

We have described GENIE, a method that can jointly estimate the proportion of variation in a complex trait that can be attributed to GxE and additive genetic effects. GENIE can also partition GxE heritability across the

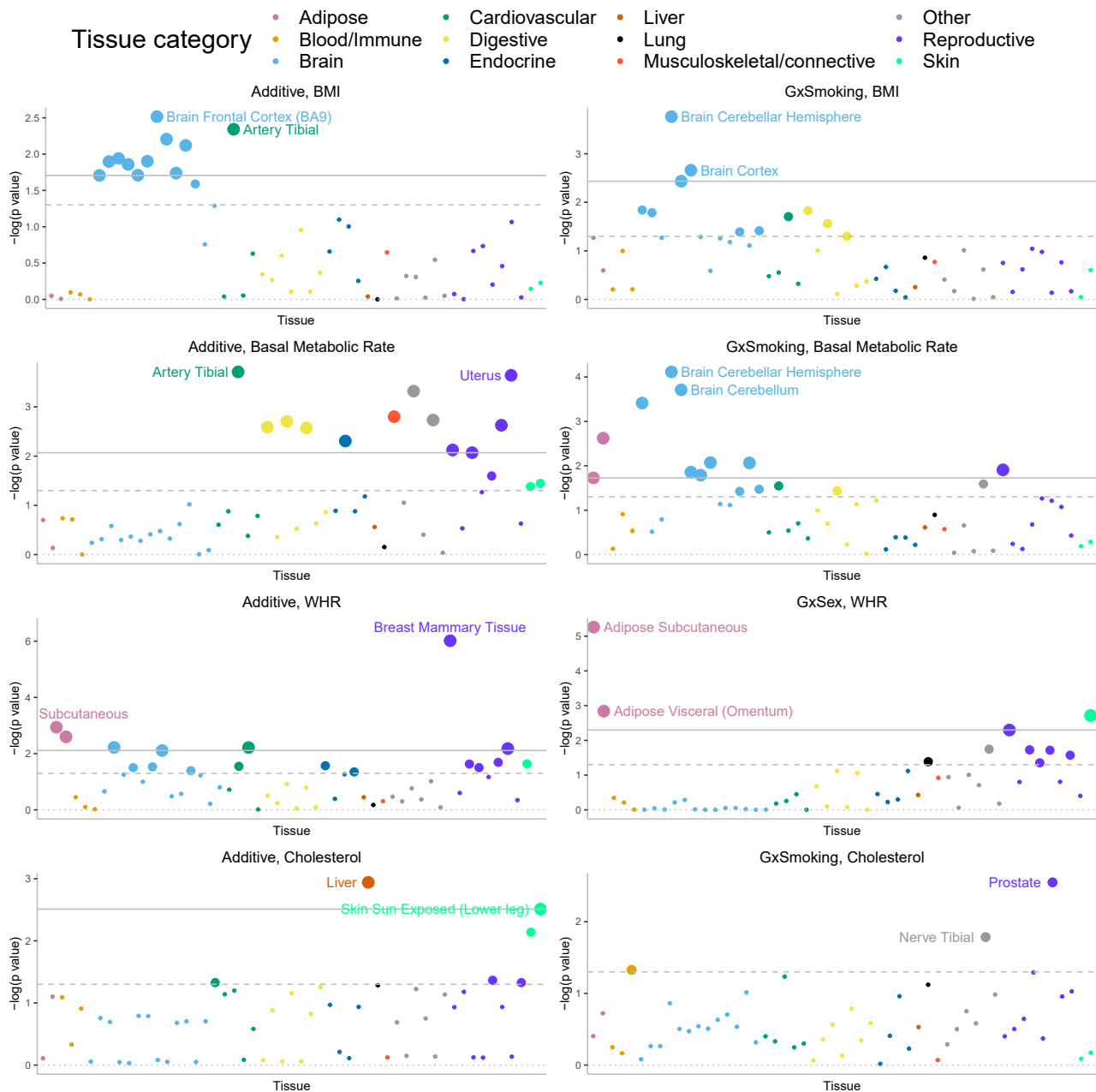


Figure 8. Partitioning GxE heritability across 53 tissue-specific genes

We plot $-\log_{10}(p)$ where p is the corresponding p value of the tissue-specific GxE enrichment defined as $\frac{h^2_{gxe,tissue}/h^2_{gxe,total}}{M_{tissue}/M_{total}}$. For every tissue-specific annotation, we use GENIE to test whether this annotation is significantly enriched for per-SNP heritability, conditional on 28 functional annotations that are part of the baseline LDSC annotations. The dashed and solid lines correspond to the nominal $p < 0.05$ and $FDR < 0.1$ threshold, respectively. We have labeled two tissues with the most significant p values for each figure.

genome with respect to annotations such as functional and tissue-specific annotations or annotations defined based on the MAF and local LD score of each SNP to localize signals of GxE. GENIE provides well-calibrated tests for the existence of a GxE effect and has high power to detect GxE effects while being scalable to large datasets.

Our simulations and real data analysis results confirm the importance of including noise heterogeneity in GxE models. Simulations comparing the calibration of GENIE to MEMMA and MonsterLM suggest that modeling Nx

does not introduce biases in scenarios without noise heterogeneity. Furthermore, it aids in controlling false positive rates when noise heterogeneity exists. In UKB data analyses, we observed that about half of trait-E pairs with significant h^2_{gxe} under the $G + GxE$ model are no longer significant under the $G + GxE + Nx$ model. Consistent with this observation, we estimated a substantial contribution of noise heterogeneity to trait variation. While our results demonstrated the importance of integrating noise heterogeneity for a more reliable and accurate estimation of GxE

heritability, alternative methods—adjusting the phenotype values of individuals in different quantile bins of the environment variable separately as proposed in Di Scipio et al.¹⁷—can prove effective under moderate levels of noise heterogeneity.

After accounting for noise heterogeneity, we observe significant genome-wide h_{gxe}^2 across more than a quarter of the trait-E pairs analyzed. Our finding has implications for understanding trait heritability by moving beyond the definition of narrow-sense heritability that only includes additive genetic effects. Based on our analyses, it is conceivable that approaches that can jointly model the hundreds of environmental variables measured in biobank-scale datasets will further increase estimates of h_{gxe}^2 . Additionally, our recovery of additional h_{gxe}^2 from low-frequency SNPs ($0.1\% \leq \text{MAF} < 1\%$) point to traits where an understanding of GxE effects can benefit from whole-exome and whole-genome studies. Our analyses of common and low-frequency SNPs lead us to recommend that SNPs should be partitioned based on MAF and LD when estimating GxE heritability (while such partitioning does not qualitatively affect results for common SNPs). Further, our results point to traits where GxE has the potential to improve genome-wide polygenic scores (GPSs) of complex traits (since h_{gxe}^2 quantifies the maximum predictive accuracy that is achievable by a linear predictor based on GxE effects). In the context of sex as an environmental variable, sex-specific GPS has been shown to provide improved accuracy over agnostic scores.^{34,39,55,56} GxE has also been recently proposed as a possible explanation for why GPS may not generalize beyond the cohort on which these predictors were trained⁶ so that modeling GxE in relevant traits could improve their transferability. Our finding that allelic effects for GxE increase with decreasing MAF and LD analogous to the relationship observed for additive allelic effects motivates an evolutionary understanding of these trends and can inform what we expect to learn from studies of rare genetic variation. Finally, our identification of sets of genes that are enriched for GxE can offer clues on trait-relevant tissues and pathways and has the potential to inform functional genomic studies.^{57,58}

We discuss the limitations of our work as well as directions for future research. First, GENIE does not explicitly model G-E correlations.¹³ While such correlations can lead to biases in estimates of GxE in the fixed-effect setting,⁵⁹ it has been shown that, in the polygenic setting, the GxE variance component estimates remain unbiased when G-E correlations are independent of the polygenic GxE effects.¹⁴ Further, our simulations suggest that GENIE is robust in the presence of G-E correlations. Nevertheless, there are plausible settings, where such correlations can lead to false positive or biased estimates of GxE, e.g., where the phenotype directly affects the environmental variable. Developing scalable methods that are accurate in these settings is an important direction for future work. Second, estimates of GxE heritability are sensitive to the scale on which traits and environmental variables are

measured and how environmental variables are encoded. In this work, we analyze quantile-normalized traits (following prior studies) and encode discrete environmental variables using a univariate parameterization (either as a 0–1 vector for each environmental variable or as a standardized version). It might be preferable to work with traits measured on their original scale and to encode each level of discrete environmental variables by a separate 0–1 covariate (leading to k environmental covariates for a k -valued environmental variable). While such choices would necessarily be guided by domain knowledge and interpretability, GENIE supports easy-to-use and rapid exploration of the consequences of these choices and can aid in assessing the robustness of these choices (we have explored a limited space of these choices here). Third, the environmental variable relevant for GxE may not be measured directly or accurately, so the environmental variable that is measured in a dataset is best viewed as a proxy for the relevant latent environmental covariate. It is essential to acknowledge that the missingness patterns of phenotypes in biobanks frequently display structure that is more intricate than random missingness.^{60,61} Consequently, removing individuals with missing data on Es can potentially affect GxE and other heritability estimates. One approach to tackle this complexity involves accurate imputation of missing data while mitigating the introduction of additional biases as observed in the mean imputation simulations (Figure S12). We view this as an important direction for future work. Fourth, the model underlying GENIE is not applicable to binary traits (either with or without ascertainment). GENIE can be extended to be applicable to binary traits (e.g., disease status) along the lines proposed in the context of additive^{62,63} and GxE estimation.¹⁴

Apart from the constraints inherent to the GENIE model, we stress the need for cautious interpretations of the results of this study due to several limitations. While GENIE can model the impact of heterogeneous noise resulting from observed environmental variables by introducing NxG components, it is important to note that the heterogeneous noise may also arise due to non-observed environmental variables. Several recent works have tried to test for GxE when the environmental variables are not observed.^{10,64} These issues along with the possibility of reverse causality, i.e., where the phenotype affects the environmental variable, warrant caution in any causal interpretation of our results (although it might be possible to overcome some of these limitations in specific analyses such as GxSex). Moreover, while the primary focus of our work is on the methodological aspects of GxE heritability estimation, our application of GENIE to medication-sensitive traits highlights the complexities arising in this setting that warrant care in interpreting the results. To explore these issues, we repeated our previous analyses after performing heuristic adjustments of phenotypes for relevant medications. Our additional analyses of GxE estimates on measurements adjusted for medication usage suggest that, while most of our results are robust to these issues

(e.g., GxE for systolic and diastolic blood pressure, GxStatin on HbA1c), some are less so (e.g., GxAge on LDL and cholesterol) (see [Note S4](#) for details). Finally, while analyses in this work were based on a cohort of self-identified white British individuals, it is valuable to investigate GxE effects using GENIE across a broader range of populations for stronger and more comprehensive results.

Data and code availability

GENIE software is an open-source software freely available at <https://github.com/sriramlab/GENIE>. The software requires `g++`, `cmake`, and `make` to compile the `C++` code on a Linux machine. Please see the documentation in the GitHub repository for further information.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2024.05.015>.

Acknowledgments

This research was conducted using the UK Biobank Resource under application 331277. We thank the participants of UK Biobank for making this work possible. This work was funded by NIH grants R35GM125055 (A.P. and S.S.), HG006399 (S.S.), and NSF grant CAREER-1943497 (A.P. and S.S.).

The authors would like to thank Alkes Price and Arbel Harpak for their feedback on the manuscript. The authors would also like to acknowledge the stimulating discussions at the UCLA Computational Genomics Summer Institute (supported by NIH grants GM135043 and GM112625) and the 2018 Bertinoro workshop in Statistical and Computational Genomics that enabled this work.

Declaration of interests

The authors declare no competing interests.

Received: September 20, 2023

Accepted: May 15, 2024

Published: June 11, 2024

Web resources

GCTA (v1.94.1), <https://yanglab.westlake.edu.cn/software/gcta>

LDSC (v1.0.1), <https://github.com/bulik/ldsc>

LEMMA (v1.0.4), <https://github.com/mkerin/LEMMA>

MonsterLM (v0.1.1), <https://github.com/GMELab/MonsterLM>

PLINK (v1.90), <https://www.cog-genomics.org/plink>

References

1. Yang, J., Loos, R.J., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490, 267–272.
2. Gagneur, J., Stegle, O., Zhu, C., Jakob, P., Tekkedil, M.M., Aiyar, R.S., Schuon, A.-K., Pe'er, D., and Steinmetz, L.M. (2013). Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *PLoS Genet.* 9, e1003803.
3. Virolainen, S.J., VonHandorf, A., Viel, K.C.M.F., Weirauch, M.T., and Kottyan, L.C. (2023). Gene-environment interactions and their impact on human health. *Genes Immun.* 24, 1–11.
4. Khoury, M.J., and Wagener, D.K. (1995). Epidemiological evaluation of the use of genetics to improve the predictive value of disease risk factors. *Am. J. Hum. Genet.* 56, 835–844.
5. Robinson, M.R., English, G., Moser, G., Lloyd-Jones, L.R., Triplett, M.A., Zhu, Z., Nolte, I.M., van Vliet-Ostapchouk, J.V., Snieder, H., et al.; LifeLines Cohort Study (2017). Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* 49, 1174–1181.
6. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *Elife* 9, e48376.
7. Laville, V., Majarian, T., Sung, Y.J., Schwander, K., Feitosa, M.F., Chasman, D.I., Bentley, A.R., Rotimi, C.N., Cupples, L.A., de Vries, P.S., et al. (2022). Gene-lifestyle interactions in the genomics of human complex traits. *Eur. J. Hum. Genet.* 30, 730–739.
8. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209.
9. Moore, R., Casale, F.P., Jan Bonder, M., Horta, D., BIOS Consortium, Franke, L., Barroso, I., and Stegle, O. (2019). A linear mixed-model approach to study multivariate gene-environment interactions. *Nat. Genet.* 51, 180–186.
10. Young, A.I., Wauthier, F.L., and Donnelly, P. (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nat. Genet.* 50, 1608–1614.
11. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82.
12. Lee, S.H., and van der Werf, J.H.J. (2016). MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32, 1420–1422.
13. Ni, G., Van Der Werf, J., Zhou, X., Hyppönen, E., Wray, N.R., and Lee, S.H. (2019). Genotype-covariate correlation and interaction disentangled by a whole-genome multivariate reaction norm model. *Nat. Commun.* 10, 2239.
14. Dahl, A., Nguyen, K., Cai, N., Gandal, M.J., Flint, J., and Zaitlen, N. (2020). A robust method uncovers significant context-specific heritability in diverse complex traits. *Am. J. Hum. Genet.* 106, 71–91.
15. Kerin, M., and Marchini, J. (2020a). Inferring Gene-by-Environment Interactions with a Bayesian Whole-Genome Regression Model. *Am. J. Hum. Genet.* 107, 698–713.
16. Kerin, M., and Marchini, J. (2020b). A non-linear regression method for estimation of gene-environment heritability. *Bioinformatics* 36, 5632–5639.
17. Di Scipio, M., Khan, M., Mao, S., Chong, M., Judge, C., Pathan, N., Perrot, N., Nelson, W., Lali, R., Di, S., et al. (2023). A versatile, fast and unbiased method for estimation of gene-by-environment interaction effects on biobank-scale datasets. *Nat. Commun.* 14, 5196.
18. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Guisev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shores, H., et al. (2015). Multi-ancestry fine-mapping of complex traits. *Nat. Genet.* 47, 1243–1251.

- N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* *50*, 621–629.
19. Hutchinson, M. (1989). A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Commun. Stat. Simulat. Comput.* *18*, 1059–1076.
 20. Liberty, E., and Zucker, S.W. (2009). The mailman algorithm: A note on matrix–vector multiplication. *Inf. Process. Lett.* *109*, 179–182.
 21. Pazokitoroudi, A., Wu, Y., Burch, K.S., Hou, K., Zhou, A., Pasaniuc, B., and Sankararaman, S. (2020). Efficient variance components analysis across millions of genomes. *Nat. Commun.* *11*, 4020.
 22. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* *12*, e1001779.
 23. Sul, J.H., Bilow, M., Yang, W.-Y., Kostem, E., Furlotte, N., He, D., and Eskin, E. (2016). Accounting for population structure in gene-by-environment interactions in genome-wide association studies using mixed models. *PLoS Genet.* *12*, e1005849.
 24. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* *53*, 185–194.
 25. Pazokitoroudi, A., Chiu, A.M., Burch, K.S., Pasaniuc, B., and Sankararaman, S. (2021). Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data. *Am. J. Hum. Genet.* *108*, 799–808.
 26. Wei, X., Robles, C.R., Pazokitoroudi, A., Ganna, A., Gusev, A., Durvasula, A., Gazal, S., Loh, P.-R., Reich, D., and Sankararaman, S. (2023). The lingering effects of Neanderthal introgression on human complex traits. *Elife* *12*, e80757.
 27. Warren, H.R., Evangelou, E., Cabrera, C.P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J.P., et al. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* *49*, 403–415.
 28. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015a). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* *47*, 1228–1235.
 29. Hou, K., Burch, K.S., Majumdar, A., Shi, H., Mancuso, N., Wu, Y., Sankararaman, S., and Pasaniuc, B. (2019). Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nat. Genet.* *51*, 1244–1251.
 30. Speed, D., and Balding, D.J. (2019). Sumher better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* *51*, 277–284.
 31. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., et al. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
 32. Sinnott-Armstrong, N., Naqvi, S., Rivas, M., and Pritchard, J.K. (2021b). GWAS of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *Elife* *10*, e58615.
 33. Ruth, K.S., Day, F.R., Tyrrell, J., Thompson, D.J., Wood, A.R., Mahajan, A., Beaumont, R.N., Wittmans, L., Martin, S., Busch, A.S., et al. (2020). Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* *26*, 252–258.
 34. Zhu, C., Ming, M.J., Cole, J.M., Edge, M.D., Kirkpatrick, M., and Harpak, A. (2023). Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genom.* *3*, 100297.
 35. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpeläinen, T.O., Esko, T., Mägi, R., Li, S., et al. (2013). Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* *9*, e1003500.
 36. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpeläinen, T.O., et al. (2015). The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* *11*, e1005378.
 37. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* *28*, 166–174.
 38. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat. Commun.* *10*, 339.
 39. Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., and Tenesa, A. (2021). Sex differences in genetic architecture in the UK Biobank. *Nat. Genet.* *53*, 1283–1289.
 40. Döring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., Fischer, G., Henke, K., Klopp, N., Kronenberg, F., et al. (2008). SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat. Genet.* *40*, 430–436.
 41. Kolz, M., Johnson, T., Sanna, S., Teumer, A., Vitart, V., Perola, M., Mangino, M., Albrecht, E., Wallace, C., Farrall, M., et al. (2009). Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet.* *5*, e1000504.
 42. Shirts, B.H., Hasstedt, S.J., Hopkins, P.N., and Hunt, S.C. (2011). Evaluation of the gene–age interactions in HDL cholesterol, LDL cholesterol, and triglyceride levels: the impact of the SORT1 polymorphism on ldl cholesterol levels is age dependent. *Atherosclerosis* *217*, 139–141.
 43. Simino, J., Shi, G., Bis, J.C., Chasman, D.I., Ehret, G.B., Gu, X., Guo, X., Hwang, S.-J., Sijbrands, E., Smith, A.V., et al. (2014). Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP consortia. *Am. J. Hum. Genet.* *95*, 24–38.
 44. Sattar, N., Preiss, D., Murray, H.M., Welsh, P., Buckley, B.M., de Craen, A.J.M., Seshasai, S.R.K., McMurray, J.J., Freeman, D.J., Jukema, J.W., et al. (2010). Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* *375*, 735–742.
 45. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide snps. *Am. J. Hum. Genet.* *91*, 1011–1021.
 46. Evans, L.M., Tahmasbi, R., Vrieze, S.I., Abecasis, G.R., Das, S., Gazal, S., Bjelland, D.W., de Candia, T.R., Haplotype Reference Consortium, and Goddard, M.E., et al. (2018). Comparison of methods that use whole genome data to estimate the

- heritability and genetic architecture of complex traits. *Nat. Genet.* *50*, 737–745.
47. Speed, D., Cai, N., UCLEB Consortium, Johnson, M.R., Nejentsev, S., Balding, D.J., et al. (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* *49*, 986–992.
 48. Gazal, S., Loh, P.-R., Finucane, H.K., Ganna, A., Schoech, A., Sunyaev, S., and Price, A.L. (2018). Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* *50*, 1600–1607.
 49. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B.M., Gusev, A., and Price, A.L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* *49*, 1421–1427.
 50. Schoech, A.P., Jordan, D.M., Loh, P.-R., Gazal, S., O'Connor, L.J., Balick, D.J., Palamara, P.F., Finucane, H.K., Sunyaev, S.R., and Price, A.L. (2019). Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* *10*, 790.
 51. Zeng, J., De Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* *50*, 746–753.
 52. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* *518*, 187–196.
 53. Small, K.S., Todorčević, M., Civelek, M., El-Sayed Moustafa, J.S., Wang, X., Simon, M.M., Fernandez-Tajes, J., Mahajan, A., Horikoshi, M., Hugill, A., et al. (2018). Regulatory variants at *KLF14* influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nat. Genet.* *50*, 572–580.
 54. Stancu, C., and Sima, A. (2001). Statins: mechanism of action and effects. *J. Cell Mol. Med.* *5*, 378–387.
 55. Rawlik, K., Canela-Xandri, O., and Tenesa, A. (2016). Evidence for sex-specific genetic architectures across a spectrum of human complex traits. *Genome Biol.* *17*, 166.
 56. Flynn, E., Tanigawa, Y., Rodriguez, F., Altman, R.B., Sinnott-Armstrong, N., and Rivas, M.A. (2021). Sex-specific genetic effects across biomarkers. *Eur. J. Hum. Genet.* *29*, 154–163.
 57. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* *167*, 1853–1866.e17.
 58. Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, C.A., Alazizi, A., Pazokitoroudi, A., Sankararaman, S., Wen, X., et al. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *Elife* *10*, e67077.
 59. Dudbridge, F., and Fletcher, O. (2014). Gene-environment dependence creates spurious gene-environment interaction. *Am. J. Hum. Genet.* *95*, 301–307.
 60. Mitra, R., McGough, S.F., Chakraborti, T., Holmes, C., Copping, R., Hagenbuch, N., Biedermann, S., Noonan, J., Lehmann, B., Shenvi, A., et al. (2023). Learning from data with Structured Missingness. *Nat. Mach. Intell.* *5*, 13–23.
 61. An, U., Pazokitoroudi, A., Alvarez, M., Huang, L., Bacanu, S., Schork, A.J., Kendler, K., Pajukanta, P., Flint, J., Zaitlen, N., et al. (2023). Deep learning-based phenotype imputation on population-scale Biobank data increases genetic discoveries. *Nat. Genet.* *55*, 2269–2276.
 62. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci. USA* *111*, E5272–E5281.
 63. Weissbrod, O., Flint, J., and Rosset, S. (2018). Estimating snp-based heritability and genetic correlation in case-control studies directly and with summary statistics. *Am. J. Hum. Genet.* *103*, 89–99.
 64. Marderstein, A.R., Davenport, E.R., Kulm, S., Van Hout, C.V., Elemento, O., and Clark, A.G. (2021). Leveraging phenotypic variability to identify genetic interactions in human phenotypes. *Am. J. Hum. Genet.* *108*, 49–67.