

Inferring Gene-by-Environment Interactions with a Bayesian Whole-Genome Regression Model

Matthew Kerin¹ and Jonathan Marchini^{2,*}

Summary

The contribution of gene-by-environment (GxE) interactions for many human traits and diseases is poorly characterized. We propose a Bayesian whole-genome regression model for joint modeling of main genetic effects and GxE interactions in large-scale datasets, such as the UK Biobank, where many environmental variables have been measured. The method is called LEMMA (Linear Environment Mixed Model Analysis) and estimates a linear combination of environmental variables, called an environmental score (ES), that interacts with genetic markers throughout the genome. The ES provides a readily interpretable way to examine the combined effect of many environmental variables. The ES can be used both to estimate the proportion of phenotypic variance attributable to GxE effects and to test for GxE effects at genetic variants across the genome. GxE effects can induce heteroskedasticity in quantitative traits, and LEMMA accounts for this by using robust standard error estimates when testing for GxE effects. When applied to body mass index, systolic blood pressure, diastolic blood pressure, and pulse pressure in the UK Biobank, we estimate that 9.3%, 3.9%, 1.6%, and 12.5%, respectively, of phenotypic variance is explained by GxE interactions and that low-frequency variants explain most of this variance. We also identify three loci that interact with the estimated environmental scores ($-\log_{10}p > 7.3$).

Introduction

Despite long standing interest in gene-by-environment (GxE) interactions,¹ this facet of genetic architecture remains poorly characterized in humans. Detection of GxE interactions is inherently more difficult than finding additive genetics in genome-wide association studies (GWASs). One difficulty is that of sample size: a commonly cited rule of thumb suggests that detection of interaction effects requires a sample size at least four times larger than that required to detect a main effect of comparable effect size.² Another difficulty is that an individual's environment, which evolves through time, is very hard to measure in a comprehensive way and is inherently high dimensional. Also, there are many environmental variables that could plausibly interact with the genome and many ways to combine them, and typically these factors are not all present in the same dataset. The recently released UK Biobank dataset, a large population cohort study with deep genotyping and sequencing and extensive phenotyping³ offers a unique opportunity for the exploration of GxE effects.^{4–10}

It can be challenging to interpret statistical inference of interactions, which should not be interpreted as biological interaction.^{11,12} Specifically, the choice of scale for a quantitative phenotype can influence the extent to which interactions are detected.¹³ However, for discovery of associated loci, including interactions can increase power¹⁴ and explicitly modeling them genome wide, as we do in this paper, can be valuable in pointing the way to improving prediction models.

Models that consider environmental variables jointly can be advantageous, particularly if several environmental

variables drive interactions at individual loci or if an unobserved environment driving interactions is better reflected by a combination of observed environments. StructLMM⁷ models the environmental similarity between individuals (over multiple environments) as a random effect and then tests each SNP independently for GxE interactions. However, StructLMM is not a whole-genome regression (WGR) model, so it does not account for the genome-wide contribution of all other variants, which is often a major component of phenotypic variance.

Advances in methods applied for detecting genetic main effects in standard GWASs have shown that linear mixed models (LMMs) can reduce false positive associations due to population structure and improve power by implicitly conditioning on other loci across the genome.^{15–17} Often these methods model the unobserved polygenic contribution as a multivariate Gaussian with covariance structure proportional to a genetic relationship matrix (GRM).^{18–20} This approach is mathematically equivalent to a WGR model with a Gaussian prior over SNP effect sizes.¹⁵ More flexible approaches that would allow for different prior distributions that better capture SNPs of small and large effects have been proposed in both the animal breeding^{21,22} and human literature.^{23–25} The BOLT-LMM method¹⁷ uses a mixture of Gaussians (MoG) prior and shows this can increase power for detecting associated loci in some (but not all) complex traits.

Here, we propose a method called Linear Environment Mixed Model Analysis (LEMMA), which aims to combine the advantages of WGR and modeling GxE with multiple environments and is applicable to large datasets with hundreds of thousands of individuals, such as the UK Biobank.

¹Wellcome Trust Center for Human Genetics, Oxford, OX3 7BN, UK; ²Regeneron Genetics Center, Tarrytown, NY 10591, USA

*Correspondence: jonathan.marchini@regeneron.com

<https://doi.org/10.1016/j.ajhg.2020.08.009>

© 2020 The Authors. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Instead of assuming that the GxE effect over multiple environments is independent at each variant, as StructLMM does, we learn a single linear combination of environmental variables (which we call an environmental score [ES]) that has a common role in interaction effects genome wide. The ES is estimated within a Bayesian WGR model that uses two separate MoG priors on main genetic effects and GxE effects. We use variational inference to fit the model that is tractable for GxE analyses of biobank scale datasets with tens of environmental variables.

Estimating the ES satisfies one of the primary motivations of this work by providing a readily interpretable way to examine the combined effect of many environmental variables and how they might interact with genotype. A motivating example is the investigation of how modern obesogenic environments might accentuate the genetic risk of obesity. Tyrell et al.²⁶ studied environments one at a time for their interaction with a body mass index (BMI) genetic risk score (GRS) and found several significant interactions. Our method allows joint analysis of environments that might plausibly better represent an obesogenic environment, negating the need to model each environment one at a time. Our other motivations when developing LEMMA were to develop a powerful method to detect GxE interactions and to estimate the proportion of variance that could be attributable to GxE interactions.

A LEMMA analysis has several distinct steps. First, the model is fitted with a large set of SNPs genome wide (e.g., all the SNPs that have been directly assayed on a genotyping chip). The estimated ES is then used to estimate the proportion of phenotypic variability that is explained by interactions with this ES (GxE heritability) via randomized Haseman-Elston (RHE) regression.^{27,28} This heritability analysis can be run on genotyped or imputed SNPs and can be stratified by minor allele frequency (MAF) and linkage disequilibrium (LD) for better interrogation of the genetic architecture of GxE interactions. The ES is also used for testing for GxE interactions one variant at a time, typically at a larger set of imputed SNPs in the dataset. We use “robust” standard errors when testing each variant for a GxE interaction, which helps control for the conditional heteroskedasticity caused by GxE interactions. We also suggest checks and solutions for the situation where environmental variables are themselves heritable and have a non-linear relationship to the trait of interest.

We compared LEMMA to existing approaches, such as StructLMM and F-tests, by using simulated data and applied the approach to UK Biobank data for BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), and pulse pressure (PP).

Material and Methods

Linear Environment Mixed Model Analysis (LEMMA)

The standard LMM used in genome wide association studies is written as

$$y = C\alpha + u + \epsilon, \quad (\text{Equation 1})$$

where y is the centered and scaled $N \times 1$ vector of phenotypes, C is an $N \times L$ matrix of covariates with $L \times 1$ fixed effects vector α , and u and ϵ are $N \times 1$ vectors of unobserved polygenic and residual effects vectors, respectively. Typically, u is modeled as a Gaussian with a mean of zero and covariance matrix $\sigma_u^2 K$. Specification of the $N \times N$ kinship matrix K is an area of active research,^{29–32} but the simplest approach is to let $K = XX^T/M$, where X is the $N \times M$ genotype matrix and columns of X (which usually correspond to SNPs) are normalized to have a mean of zero and variance one. This can equivalently be written as a Bayesian WGR model,

$$y = C\alpha + X\beta + \epsilon, \quad (\text{Equation 2})$$

where

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2 / M). \quad (\text{Equation 3})$$

Here β is an $M \times 1$ vector modeling the random effect of each SNP. This form corresponds to the so-called infinitesimal model where every SNP is allowed to have a small but non-zero effect on a given trait. To generalize the model to a non-infinitesimal genetic architecture, we model SNP effects with a mixture of Gaussian priors. This approach has been applied previously in human genetics^{17,25} and by the “Bayesian alphabet” of genomic prediction methods in the animal breeding literature.^{21,22,33}

We extend this setup to model GxE interactions genome wide with a linear combination of multiple environmental variables by using

$$Y = C\alpha + X\beta + Z\gamma + \epsilon, \quad (\text{Equation 4})$$

where

$$Z = \eta \odot X, \quad (\text{Equation 5})$$

$$\eta = Ew, \quad (\text{Equation 6})$$

$$w \sim \mathcal{N}(0, I_L), \quad (\text{Equation 7})$$

where E is an $N \times L$ matrix of environmental variables that could potentially be involved in GxE interactions and w is an $L \times 1$ vector of weights. Together they define the $N \times 1$ vector η , which is the linear combination of environments that we refer to as the ES. This ES is learned in tandem with SNP effects. We note that all environmental variables contained in E must also be contained in C , so $L \leq L$. We chose to model the interaction weights w with a Gaussian prior, but in theory, one could consider sparser priors, such as a spike and slab. We set the variance of the prior on w to the identity matrix I_L . Setting the prior variance of w to a parameter would be unidentifiable because any change in scale would be absorbed by the prior variances on the interaction effects γ (see $\sigma_{\gamma,1}^2$ and $\sigma_{\gamma,2}^2$ in Equations 8 and 9).

The $N \times M$ matrix Z contains all of the multiplicative interaction terms of the ES η with all of the genetic variants. We use the notation $\eta \odot X$ for the element-wise product of η with each column of X . In other words, $\eta \odot X = \text{diag}(\eta)X$, where $\text{diag}(\eta)$ is an $N \times N$ diagonal matrix with η as the diagonal. The vector of interaction effect sizes γ has dimension $M \times 1$.

We chose to use MoG priors on both the main genetic effects (β) and the interaction effects (γ) because this prior is very flexible

and spans the range of genetic architectures from polygenic to a very sparse model. The priors are

$$\beta_j | \sigma_e^2, \lambda_\beta, \sigma_{\beta,1}^2, \sigma_{\beta,2}^2 \sim \lambda_\beta \mathcal{N}(0, \sigma_e^2 \sigma_{\beta,1}^2) + (1 - \lambda_\beta) \mathcal{N}(0, \sigma_e^2 \sigma_{\beta,2}^2), \quad (\text{Equation 8})$$

$$\gamma_j | \sigma_e^2, \lambda_\gamma, \sigma_{\gamma,1}^2, \sigma_{\gamma,2}^2 \sim \lambda_\gamma \mathcal{N}(0, \sigma_e^2 \sigma_{\gamma,1}^2) + (1 - \lambda_\gamma) \mathcal{N}(0, \sigma_e^2 \sigma_{\gamma,2}^2). \quad (\text{Equation 9})$$

We use standard Gaussian priors on the covariate and error terms.

$$\alpha | \sigma_\alpha^2 \sim \mathcal{N}(0, \sigma_\alpha^2) \quad (\text{Equation 10})$$

$$\epsilon | \sigma_\epsilon^2 \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (\text{Equation 11})$$

Variational Inference

For notational convenience, we define $\theta = \{\alpha, \beta, \gamma, w\}$ as the set of latent variables, $D := \{X, E\}$ the genetic and environmental data, and ϕ as the set of hyper parameters. Then the posterior $p(\theta | Y, D, \phi)$ is given by

$$p(\theta | Y, D, \phi) \propto p(Y | \theta, D, \phi) \prod_c p(\alpha_c | \phi) \prod_i p(w_i) \prod_j p(\beta_j, u_j | \phi) \prod_j p(\gamma_j, v_j | \phi). \quad (\text{Equation 12})$$

To evaluate the posterior, we use the variational inference framework, approximating the true posterior $p(\theta | Y, D, \phi)$ with a tractable alternative distribution $q(\theta; v)$ governed by (variational) parameters v . To make inference tractable, we use the standard mean-field assumption so that $q(\theta; v)$ factorizes

$$q(\theta; v) = \prod_c q(\alpha_c) \prod_i q(w_i) \prod_j q(\beta_j, u_j) \prod_j q(\gamma_j, v_j). \quad (\text{Equation 13})$$

To make $q(\theta; v)$ a close approximation of the true posterior, we minimize the Kullback-Leibler (KL) divergence between $q(\theta; v)$ and $p(\theta | Y, D, \phi)$ with respect to variational parameters v . In this manner, the problem has been transformed from one of computing posterior distributions into one of optimization. We can show that minimizing the KL divergence is equivalent to maximizing a lower bound on the marginal log likelihood by observing

$$KL(q \| p) = -\mathbb{E}_q \left[\log \frac{p(\theta | Y, D, \phi)}{q(\theta; v)} \right], \quad (\text{Equation 14})$$

$$= -\mathbb{E}_q \left[\log \frac{p(\theta, Y | D, \phi)}{q(\theta; v)} \right] + \mathbb{E}_q [\log p(Y | D, \phi)], \quad (\text{Equation 15})$$

$$= -\mathbb{E}_q \left[\log \frac{p(\theta, Y | D, \phi)}{q(\theta; v)} \right] + \log p(Y | D, \phi). \quad (\text{Equation 16})$$

Thus, we can write

$$F(v; \phi) := \mathbb{E}_q \left[\log \frac{p(\theta, Y | D, \phi)}{q(\theta; v)} \right] \leq \log p(Y | D, \phi). \quad (\text{Equation 17})$$

Here, $F(v; \phi)$ is commonly referred to as the evidence lower bound (ELBO). As a result of the factorized form of Equation 13, we can cyclically update the approximate distribution for each latent variable in turn until we reach convergence.

Our model depends on a set of eight hyper-parameters $\phi = \{\sigma_e^2, \{\sigma_{\beta,i}^2\}_{i=1}^2, \{\sigma_{\gamma,i}^2\}_{i=1}^2, \lambda_\beta, \lambda_\gamma, \sigma_\alpha^2\}$. We set σ_α^2 to a large constant to create a flat prior on the covariates, leaving seven unknowns. Similar methods have performed a grid search over hyper-parameter values (with either cross validation¹⁷ or the in-sample ELBO to identify the optimum²⁴). For LEMMA, a grid search would be computationally demanding because the set of hyper-parameters is larger and we cannot efficiently perform multiple runs in parallel as done by Loh et al.¹⁷ Instead, we maximize a lower bound on the approximate log likelihood (the so-called ELBO) with respect to the hyper-parameters. In this manner, our approach can be viewed as a variational expectation maximization algorithm.^{34,35}

Similar to the EM algorithm, the hyper-parameter maximization step can lead to slow exploration of the hyper-parameter space and thus to slow convergence of the LEMMA algorithm. We use an accelerator, SQUAREM,³⁶ to speed up convergence. Given two estimates of the hyper-parameters ϕ_{t-2} and ϕ_{t-1} , we can adjust the maximized estimate ϕ_t with

$$\tilde{\phi}_t(v_t) = \phi_{t-2} - 2v_t \Delta \phi_{t-1} + v_t^2 \Delta^2 \phi_t, \quad (\text{Equation 18})$$

where $\Delta \phi_{t-1} = \phi_{t-1} - \phi_{t-2}$ and $\Delta^2 \phi_t = \phi_t - 2\phi_{t-1} + \phi_{t-2}$. Thus, the new adjusted estimate $\tilde{\phi}_t(v_t)$ is a continuous function of the step size v_t , which yields the original estimate ϕ_t for $v_t = -1$. As recommended by Varadhan et al.,³⁶ we set $v_t = \min(-1, -\|\Delta \phi_{t-1}\|_2^2 / \|\Delta^2 \phi_t\|_2^2)$. Occasionally this yields an estimate that is either outside of the domain of ϕ or leads to a state with a worse ELBO than the previous state. For the first issue, we use a simple backtracking method of halving the distance between v_t and -1 , and for the second, we simply judge model convergence when the absolute change in the ELBO drops below a given threshold. We use the same convergence criterion as the BOLT-LMM method,¹⁷ namely that a full pass through all latent variables yields an absolute change of less than 0.01 in the approximate log likelihood (ELBO). Figure S17 shows the evolution of the ES parameter estimates for the four UK Biobank traits we analyzed and illustrates that, at the point of convergence, the parameters appear stable.

Identifying GxE-Associated Loci

After convergence of the LEMMA variational inference algorithm, we obtain posterior mean estimates of $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\eta} = E\tilde{w}$. From these, we construct residualized phenotypes by following a leave-one-chromosome-out (LOCO) scheme:

$$Y_{\text{resid-LOCO}} = Y - C\hat{\alpha} - X_{\text{LOCO}}\hat{\beta}_{\text{LOCO}} - \hat{\eta} \odot X_{\text{LOCO}}\hat{\gamma}_{\text{LOCO}}. \quad (\text{Equation 19})$$

X_{LOCO} denotes the genotype matrix excluding SNPs on the same chromosome of the test SNP, and β_{LOCO} and γ_{LOCO} are constructed similarly. Using a LOCO scheme has been shown to increase power in LMMs because the effect of the test SNP is conditioned on the effects on a large proportion of the rest of the genome.^{16,19}

For each imputed SNP, we then perform hypothesis tests $\beta_{\text{test}} \neq 0$ and $\gamma_{\text{test}} \neq 0$ by using the linear model

$$Y_{\text{resid-LOCO}} = x_{\text{test}}\beta_{\text{test}} + (\hat{\eta} \odot x_{\text{test}})\gamma_{\text{test}} + \epsilon, \quad (\text{Equation 20})$$

$$= H\tau + \epsilon. \quad (\text{Equation 21})$$

Here, H is the $N \times 2$ design matrix with first and second columns containing x_{test} and $\hat{\eta} \odot x_{\text{test}}$, respectively, and τ is the 2×1 vector

containing parameters β_{test} and γ_{test} , which are the main genetic effect and interaction effect of the SNP being tested.

Assuming that ϵ has a mean of zero and covariance matrix Ω , we can use the standard ordinary least squares (OLS) estimator

$$\hat{\tau} = (H^T H)^{-1} H^T y, \quad (\text{Equation 22})$$

which (under certain regularity conditions) is asymptotically normally distributed with mean τ and variance $\text{Var}(\hat{\tau})$. By assuming the residual phenotype is homoskedastic, that is that $\Omega = \hat{\sigma}_e^2 I$, we can obtain the usual variance estimator given by

$$\text{Var}(\hat{\tau}) = \hat{\sigma}_e^2 (H^T H)^{-1}. \quad (\text{Equation 23})$$

It has previously been observed that GxE interaction tests are likely to suffer from conditional heteroskedasticity,³⁷ and hence, the homoskedastic variance estimator is likely to underestimate the true variance.³⁸ We explain this phenomenon in detail in the [Supplemental Notes](#).

To overcome this, we use robust standard errors, alternatively called Huber-White, sandwich, or “heteroskedastic consistent” errors,^{39,40} that are standard tools in economics⁴¹ and have previously been proposed for use in GxE interaction studies.^{37,42,43} We further include a small adjustment that reduces bias in small samples.⁴⁴ This yields the variance estimator

$$\text{Var}(\hat{\tau}) = (H^T H)^{-1} H^T \hat{\Sigma} H (H^T H)^{-1}, \quad (\text{Equation 24})$$

where $\hat{\Sigma}$ is a diagonal matrix with $\hat{\Sigma}_{ii} = (\hat{\epsilon}_i^2 / (1 - h_{ii})^2)$, where $\hat{\epsilon} = y - H\hat{\tau}$ and $h = H(H^T H)^{-1} H^T$. Hence, our GxE test statistic is given by

$$\frac{\hat{\gamma}_{\text{test}}^2}{\text{Var}(\hat{\gamma}_{\text{test}})} \quad (\text{Equation 25})$$

and, under the null hypothesis, is asymptotically distributed as χ_1^2 . Because main effects tests are not sensitive to assumptions of heteroskedasticity in the same way that GxE tests are,³⁷ we use a simple t test to test the hypothesis $\beta_{\text{test}} \neq 0$.

Heritability Estimation

Previous GWR methods^{24,45,46} have shown that it is possible to rearrange the model hyper-parameters to gain an estimate of trait heritability. We find that in our variational framework, this approach underestimates trait heritability because of the tendency of mean-field variational inference to underestimate the posterior variance of each parameter. Instead, we treat the posterior mean $\hat{\eta}_{\text{LEMMA}}$ as a fixed effect and use RHE regression^{27,28,47} to estimate heritability with a single SNP component²⁷ (RHE-SC) and multiple SNP components²⁸ (RHE-LDMS). With the multi-component model, SNPs are stratified into a total of 20 bins: 5 MAF bins (≤ 0.1 , $0.1 < \text{MAF} \leq 0.2$, $0.2 < \text{MAF} \leq 0.3$, $0.3 < \text{MAF} \leq 0.4$, and $0.4 < \text{MAF} \leq 0.5$) and 4 LD score quantiles.

The single component model is given by

$$y \sim \mathcal{N}(E\alpha, \sigma_\beta^2 K + \sigma_\gamma^2 \hat{V} + \sigma_e^2 I), \quad (\text{Equation 26})$$

where $K = XX^T/M$, $\hat{V} = Z(\hat{\eta})Z(\hat{\eta})^T/M$, and $Z(\hat{\eta}) = \text{diag}(\hat{\eta})X$. HE regression is a method of moments estimator that fits the variance components ($\sigma_\beta^2, \sigma_\gamma^2, \sigma_e^2$) to minimize the difference between the empirical and expected covariances. This is mathematically equivalent to solving the following linear system:

$$\begin{pmatrix} \text{tr}(K^2) & \text{tr}(KV) & \text{tr}(K) \\ \text{tr}(KV) & \text{tr}(V^2) & \text{tr}(V) \\ \text{tr}(K) & \text{tr}(V) & N \end{pmatrix} \begin{pmatrix} \sigma_\beta^2 \\ \sigma_\gamma^2 \\ \sigma_e^2 \end{pmatrix} = \begin{pmatrix} y^T Ky \\ y^T Vy \\ y^T y \end{pmatrix}. \quad (\text{Equation 27})$$

Wu et al.²⁷ showed that this system can be solved in $O(NMB)$ time (for small B) without ever forming the kinship matrices K and V with Hutchinson's estimator and that covariates can be efficiently projected out of the phenotype, genotypes, and interaction matrix Z with minimal additional cost. Pazokitoroudi et al.²⁸ give an extension to multiple components and show that variance estimates can be obtained with the block jackknife.

Speed et al.⁴⁸ show that the usual form for h_G^2 , the proportion of trait variance explained by additive genetic effects, given by

$$\hat{h}_G^2 = \frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 + \hat{\sigma}_e^2}, \quad (\text{Equation 28})$$

holds only when genotype matrix X is standardized to have a column mean of zero and column variance of one. Although this is true in expectation for \hat{Z} (assuming that $\text{Cov}(\hat{\eta}_j, X_j) = 0$, $\forall j \in \{1, M\}$), this is not guaranteed. To obtain a column mean of zero, we include an intercept of ones among the covariates that are projected out of the phenotype, genotypes, and interaction matrix. To account for columns' having variance not equal to one, we use a more general form of the heritability estimator (see Speed et al.⁴⁸ for details)

$$\hat{h}_{\text{GxE}}^2 = \frac{\hat{\sigma}_\gamma^2 \text{tr}(\hat{V})/N}{\hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 \text{tr}(\hat{V})/N + \hat{\sigma}_e^2}. \quad (\text{Equation 29})$$

Computational Efficiency

We implement a number of steps to improve computational and memory efficiency, including vectorization using SIMD extensions, compressed data formats, pre-computing quantities, parallel computing with OpenMPI, and the use of the well-optimized Intel Math Kernel Library. Full details are given in the [Supplemental Notes](#).

Detecting Squared Environmental Dependence

By default, each of the L environmental variables is tested against the phenotype for significant squared effects. To do this, LEMMA tests the hypothesis $\beta_l \neq 0$ by using the following linear model:

$$y = 1\alpha_0 + C\alpha + E_l^2 \beta_l + \epsilon. \quad (\text{Equation 30})$$

The squared effect of any environmental variables with a p value less than 0.01 (Bonferroni correction for L multiple tests) are added to the matrix of covariates C .

Controlling for Covariates

Unlike in BOLT-LMM,¹⁷ it is not possible to efficiently project covariates out of the model (y, X, Z) because the multiplicative interaction matrix Z changes after each pass through the data. Instead, the LEMMA software package can either regress covariates out of the phenotype or model the covariates as random effects in the variational framework. For our analyses of the UK Biobank, we included all covariates within the variational model.

Comparison to Existing GxE Methods

We compare LEMMA to three other single SNP methods that jointly model interactions with multiple environments. The first comparison method, StructLMM,⁷ is a method that uses a random effects term, u , to model environmental similarity instead of genetic similarity. Specifically, StructLMM uses the model

$$y \sim \mathcal{N}(C\alpha + x_{\text{test}}\beta, \sigma_{\text{GxE}}^2 \text{diag}(x_{\text{test}})\Sigma \text{diag}(x_{\text{test}}) + \sigma_e^2 \Sigma + \sigma_n^2 I) \quad (\text{Equation 31})$$

to test the hypothesis $\sigma_{\text{GxE}}^2 \neq 0$. Here, C is the matrix of covariates with fixed effects α , x_{test} is the focal variant, and $\Sigma = EE^T$ is the environmental similarity matrix (where E is an $N \times L$ matrix of environmental variables). Although StructLMM provides both an interaction test and a joint test that looks for non-zero main and interaction effects at each SNP, we use only the interaction test in our comparisons. Finally, we note that StructLMM recommends “gaussianizing” the phenotype as a pre-processing step; however, we just center and scale the phenotype for consistency with our other methods.

Our second and third comparison methods use equivalent information to StructLMM in a fixed effects framework. Consider the linear model

$$y = C\alpha + E\alpha + x_{\text{test}}\beta_{\text{test}} + x_{\text{test}} \odot E\gamma + \epsilon \quad (\text{Equation 32})$$

$$= H\tau + \epsilon, \quad (\text{Equation 33})$$

where H is formed from column-wise concatenation of $[C, E, x_{\text{test}}, \text{diag}(x_{\text{test}})E]$ and τ is the corresponding vector of fixed effects. Let R be the indicator matrix such that $R\tau = \gamma$. We wish to test the null hypothesis $H_0: \gamma = 0$. Assuming that ϵ has a mean of zero and covariance matrix Ω , we can use the standard OLS estimator $\hat{\tau} = (H^T H)^{-1} H^T y$, which (under certain regularity conditions) is asymptotically distributed as normal with mean τ and variance given by $\text{Var}(\hat{\tau}) = (H^T H)^{-1} H^T \Omega H (H^T H)^{-1}$. Assuming homoskedasticity yields the standard F test statistic,

$$F_{\text{test}} = \frac{(R\hat{\tau})^T (R(H^T H)^{-1} R^T)^{-1} (R\hat{\tau}) / L}{\hat{\sigma}_e^2}, \quad (\text{Equation 34})$$

which follows an $F_{d_1-d_0, N-d_1}$ distribution under the null hypothesis, where d_1 is the column rank of H and d_0 is the column rank of H under the null hypothesis. Alternatively, we can use the same robust standard error used in the LEMMA test statistic

$$F_{\text{robust}} = (R\hat{\tau})^T (R(H^T H)^{-1} H^T \hat{\Omega} H (H^T H)^{-1} R^T)^{-1} (R\hat{\tau}), \quad (\text{Equation 35})$$

where $\hat{\Omega}$ is a diagonal matrix with $\hat{\Omega}_{ii} = \frac{\hat{\epsilon}_i^2}{(1-\hat{h}_{ii})^2}$, $\hat{\epsilon} = y - H\hat{\tau}$ and $h = H(H^T H)^{-1} H$. Then F_{robust} is asymptotically distributed as $\chi_{d_3}^2$, where d_3 is the rank of HR^T . In our simulations, we refer to this as the robust F-test.

SNP-Specific Interaction Profile

The SNP-specific interaction profile is defined as $\eta_{LS} = Ew_{LS}$, where w_{LS} is the least-squares parameter estimate of w in the single SNP model

$$y = C\alpha + x_{\text{test}}\beta_{\text{test}} + x_{\text{test}} \odot Ew + \epsilon \quad (\text{Equation 36})$$

and y, C , and E are the data matrices defined below. The correlation between η_{LS} for a given SNP and the ES estimated by LEMMA can

be viewed as a proxy for how well LEMMA captures the GxE interactions at that locus.

UK Biobank Analysis

We used real genotype and phenotype data from the UK Biobank, which is a large prospective cohort study of approximately 500,000 individuals living in the UK.³ To account for potential confounding effects of population structure, we first subset down to the white British subset of 344,068 individuals used by Bycroft et al.³ in a GWAS on human height. This represents unrelated individuals who self-report white British ethnicity and whose genetic data projected onto principal components lies within the white British cluster.³ After sub-setting down to individuals who had complete data across the phenotype, covariates, and environmental factors (see below), we were left with approximately 280,000 samples per trait (Table S1). Finally, we filtered genetic data on the basis of MAF (≥ 0.01) and IMPUTE-info score (≥ 0.3), leaving approximately 642,000 genotyped variants (Table S1) and 10,295,038 imputed variants per trait. For each trait, we included age³, age² \times gender, age³ \times gender, a binary indicator for the genotype chip, and the top 20 genetic principal components as additional covariates.

BMI was derived from height and weight measurements made during the first assessment visit (instance “0” of field 21001), and readings more than six standard deviations from the population mean were set to missing. logBMI and INT(BMI) refer to BMI after applying a log transformation and an inverse normal transformation (applied separately to males and females), respectively.

After calculating the mean SBP and DBP by using automated blood pressure readings from the first assessment visit (fields 4080 and 4079), we adjusted for medication usage by adding 15 mmHg and 10 mmHg to SBP and DBP, respectively.⁴⁹ Data from manual measurements (fields 93 and 94) were used in the rare instance that no automated reading was available. Blood pressure readings more than four standard deviations from the mean were set to missing. PP was then calculated as SBP minus DBP.

For our GxE analyses, we made use of 42 environmental variables from the UK Biobank, similar to those used in previous GxE analyses of BMI in the UK Biobank.^{7,50} From the data provided by the UK Biobank, we included seven continuous environmental variables (“age when attended assessment centre,” “sleep duration,” “time spent watching television,” “number of days/week walked 10+ minutes,” “number of days/week of moderate physical activity 10+ minutes,” “number of days/week of vigorous physical activity 10+ minutes,” and “Townsend deprivation index at recruitment”), one ordinal environmental variable (“alcohol intake frequency”), nine dietary ordinal variables (“salt added to food,” “oily fish intake,” “non-oily fish intake,” “processed meat intake,” “poultry intake,” “beef intake,” “lamb intake,” “pork intake,” and “cheese intake”) and two dietary continuous variables (“tea intake” and “cooked vegetable intake”). We further derived one categorical variable (“is current smoker” from the responses given in the UK Biobank field “smoking status”) and one continuous variable (“sleep sd,” the number of standard deviations from the population mean sleep duration). For analyses of blood pressure, we additionally included one further continuous variable, “waist circumference.” This left 11 dietary variables and ten non-dietary variables (11 for blood pressure traits). In addition, we included multiplicative interactions between participants’

age and gender with all non-dietary variables and included the main effect of gender, giving the data matrix E a total of 42 columns (45 for blood pressure traits). Before running LEMMA, each column was standardized as

$$E_{ij} = \frac{E_{ij} - \text{mean}(E_{:,j})}{\text{sd}(E_{:,j})}. \quad (\text{Equation 37})$$

In all cases where participants responded with “prefer not to answer,” “do not know,” or “none of the above,” we set the value to missing. For three continuous variables (“time spent watching television,” “tea intake,” and “cooked vegetable intake”), we removed the 99th percentile, and for “sleep duration,” we removed both the 1st and 99th percentiles.

After running LEMMA, we found it convenient to interpret weights corresponding to a re-scaled data matrix E_1 . Assuming the column space of E and E_1 is the same, weights w_1 that correspond to E_1 can be extracted from the ES via least-squares

$$w_1 = (E_1^T E_1)^{-1} E_1^T \hat{\eta}_{\text{LEMMA}}, \quad (\text{Equation 38})$$

where $\hat{\eta}_{\text{LEMMA}}$ represents the ES. We note that, although multivariate linear regression is invariant to a re-scaling of the design matrix, ridge regression is not because of the penalization place on the magnitude of the learned parameters. However, because the magnitude of the weights from our UK Biobank analysis is typically small (less than 0.2) compared to the standard deviation of our Gaussian prior (1), in this case, the re-scaling makes minimal difference.

Re-coded data matrix E_1 was formed with one column for each of the 11 dietary variables (normalized to have a mean of zero and variance of one) and three columns for each of the ten (11 for blood pressure traits) non-dietary variables; the first column was augmented by a binary male indicator vector, the second by a binary female indicator vector, and the third by a continuous vector of participant age. Columns augmented by male and female binary indicator vectors were normalized to have a mean of zero and variance one (not including zeros due to augmentation), apart from age (scaled to represent the number of decades aged past 40 years). Columns augmented by age were normalized first and then multiplied by age on the per-decade scale. We further included indicator columns for men and women, which can be interpreted as gender-specific intercepts and is equivalent to including an intercept and a binary column for only one gender (men or women). We note this leaves 43 (46) columns where the extra column comes from including an intercept within the column space of E_1 and is necessary because some columns have a mean not equal to zero. Thus, the column space of E_1 is equivalent to E under the constraint that the ES has a mean of zero.

Simulation Studies

Genetic data was sub-sampled from the UK Biobank by default with $N = 25,000$ unrelated individuals of mixed ancestry and $M = 100,000$ genotyped SNPs. Environmental variables were simulated from a standard Gaussian distribution. By default, we constructed phenotypes with 2,500 causal main effects and 1,250 causal interaction effects explaining 20% and 5% of trait variance, respectively. For each phenotype, we constructed a weighted average of the environmental variables, which we used to simulate multiplicative interaction effects. Environments with a non-zero weight are referred to as active. All non-zero effects

were drawn from SNPs in the first half of each chromosome, allowing us to test the calibration of each method on “null” SNPs from the second half of each chromosome. To allow for direct power comparisons across different scenarios, we included an additional 60 SNPs with standardized effect sizes that together accounted for 1% of trait variance with their main effects and 1% of trait variance with their interaction effects. Finally, a further 1% of trait variance was modeled via the first genetic principal component (PC). For all methods, we included the first genetic PC as a covariate. For each method, we calculated power as the proportion of the SNPs of standardized effect identified at a threshold of $p < 0.01$.

In simulations used to test RHE regression, phenotypes were constructed with 10,000 causal main effects, explaining 20% of trait variance, and in simulations with non-zero GxE heritability, with 10,000 causal SNPs with interaction effects.

Model Misspecification

We simulated a scenario where a disease trait Y depends non-linearly on a heritable environmental factor S . More explicitly, suppose that X is the centered and scaled genotype matrix so that columns have a mean of zero and variance of one, that S is modeled as

$$S = X\tau + \epsilon_s, \quad (\text{Equation 39})$$

where $\epsilon_s \sim \mathcal{N}(0, (1 - h_s^2)I)$, τ models random SNP effects for S , and trait Y is given by

$$Y|a = aS^2 + X\beta + \epsilon. \quad (\text{Equation 40})$$

Here, a is a constant that we use to control the strength of the contribution of S^2 to Y , $\epsilon \sim \mathcal{N}(0, (1 - h_\beta^2))$ and β is the random SNP effects for Y . For simulation, we suppose that τ and β have spike and slab priors

$$\tau_j \left| v_j \sim v_j \mathcal{N}\left(0, \frac{h_\tau^2}{P\lambda_\tau}\right) + (1 - v_j)\delta_0(\tau_j), \quad (\text{Equation 41})$$

$$\beta_j \left| u_j \sim u_j \mathcal{N}\left(0, \frac{h_\beta^2}{P\lambda_\beta}\right) + (1 - u_j)\delta_0(\beta_j), \quad (\text{Equation 42})$$

$$v_j \sim \text{Ber}(\lambda_\tau), \quad (\text{Equation 43})$$

$$u_j \sim \text{Ber}(\lambda_\beta). \quad (\text{Equation 44})$$

Results

Performance on Simulated Data

Figure 1 compares the ability of different methods to detect GxE interactions at SNPs in simulations where a single true ES interacts with SNPs across the genome. **Figure S1** shows the false positive rate (FPR) to detect main effects. We compared our default version of LEMMA, which uses robust standard errors, StructLMM, a simple F-test of interaction, and an F-test that uses robust standard errors (see **Material and Methods**). The simulations vary GxE heritability, the total number of environmental variables, and sample size. When sample size is large ($N = 100,000$), all

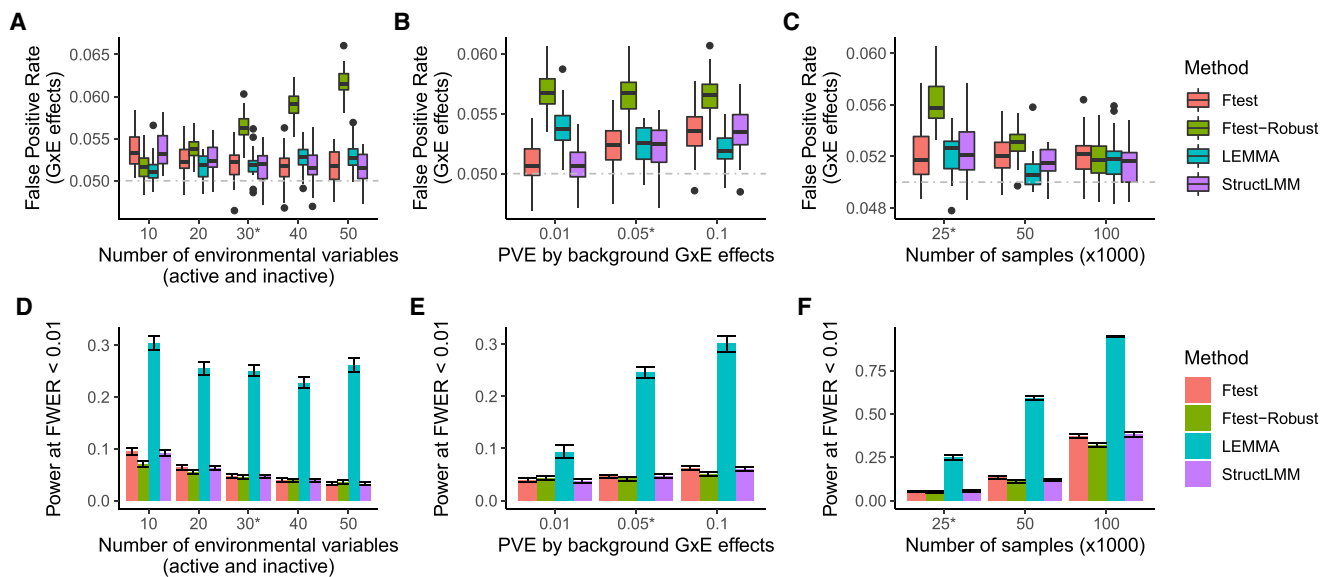


Figure 1. Type I Error and Power of Tests to Detect GxE Effects in Simulation

(A–C) Comparison of false positive rate as the number of environments increases (A), as phenotype variance explained by GxE effects increases (B), or as the number of samples increases (C). (D–F) Analogous comparison of the power to detect GxE interactions. Simulations used genotypes subsampled from the UK Biobank and by default contained $N = 25,000$ samples, $M = 100,000$ SNPs, six environmental variables that contributed to the ES, and 24 that did not (default parameters denoted by stars). Error bars extend to mean $\pm 1 \times$ standard error of the mean. We assess power (at family-wise error rate [FWER] < 0.01) to detect 60 causal SNPs whose GxE effect each explained 0.00016% of trait variance. See [Materials and Methods](#) for full details of phenotype construction.

the methods have reasonable control of FPR and LEMMA controls FPR at least as well as other methods across the range of simulations. When sample size is smaller ($N = 25,000$), the robust F-test performs less well as the number of environments grows (Figure 1A) and the F-test and StructLMM perform less well as the amount of GxE variance increases (Figure 1B). When we increase the sample size to $N = 200,000$, we still find that LEMMA has a slightly inflated type I error rate (see Figure S2).

It is interesting that all the methods we tested have a slightly inflated type I error, and this is most likely due to a number of different reasons. StructLMM and the F-test fit a model at each variant and ignore GxE effects at other loci, which can induce heteroskedasticity that can inflate type I error.^{37,43} We used robust standard errors for the robust F-test, but it seems that this approximation works best when the number of environmental traits is small. LEMMA does account for GxE effects at other loci and also uses robust standard errors, but it still has a slightly inflated type I error that gets worse as the number of environments increases (Figures 1A and S2). In parallel simulations (see Figure S3), we find that our model slightly over-estimates GxE heritability as the number of environments increases. Because our simulations test for GxE effects at SNPs used to estimate the ES, we suspect that the type I error inflation is due to this two-stage approach.

When there is a single true ES involved in GxE interactions, we found that LEMMA provided a substantial power increase (Figures 1 and S4). StructLMM and F-tests have

very similar power in these simulations, although previous work suggests that StructLMM may outperform the F-test in small samples.⁷

When estimating the GxE heritability of the LEMMA ES by using RHE regression with a single SNP component (RHE-SC), we observed some upward bias as the number of environments increases. This effect is ameliorated by increasing sample size (see Figure S3), suggesting that the influence of over-fitting in our Biobank analyses is mild. In 20 simulations with $L = 30$ environmental variables, $N = 100,000$ samples, and true GxE heritability of 5%, we observed a mean GxE heritability of 5.2%. Figure S5 further illustrates the ES estimation accuracy of LEMMA.

Finally, we ran LEMMA on two sets of simulated datasets ($N = 25,000$) with causal SNPs chosen either randomly or to be low frequency ($MAF < 0.1$). We used the ES estimated from each simulated dataset to estimate h_G^2 and h_{GxE}^2 by using RHE with SNPs stratified by MAF and LD (RHE-LDMS) and then without any stratification (RHE-SC). Previous studies have established that estimating heritability with a single SNP component makes assumptions about the relationship between MAF, LD, and trait architecture that may not hold up in practice,^{48,51} whereas stratifying SNPs into bins according to MAF and LDscore (the LDMS approach) is relatively unbiased.^{51–53} Figure S6 confirms that stratifying by MAF and LD results in accurate heritability estimates irrespective of the MAF distribution of causal SNPs and suggests that this method can be used to interrogate the MAF distribution of GxE components of a trait via LEMMA. However, when causal SNPs are low frequency,

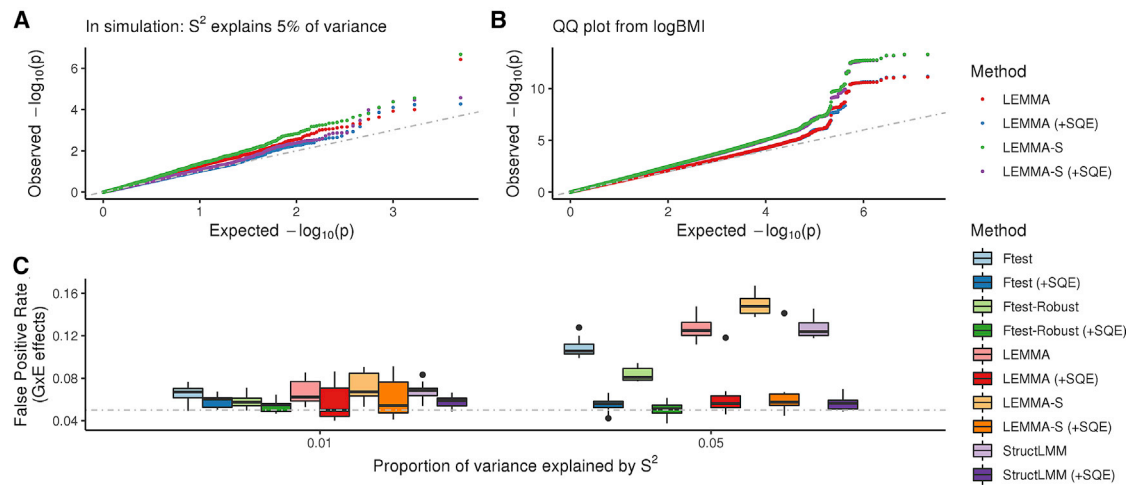


Figure 2. Bias from Model Misspecification of a Heritable Environmental Variable

(A) Comparison of GxE association test statistics from a single simulation where non-linear dependence on the confounder explains 5% of trait variance. FPR at heritable sites of the misspecified environment only. (B) Comparison of GxE association test statistics from an analysis of \log BMI in 281,149 participants from the UK Biobank. (C) FPR at heritable sites of the misspecified environment while the strength of squared dependence varies. 20 repeats per scenario. Abbreviations are as follows: LEMMA-S, LEMMA with non-robust variances used to compute test statistics; (+SQE), significant squared environmental variables (Bonferroni correction) included as additional covariates.

not stratifying by MAF and LD results in underestimation of h_G^2 .

Controlling for Heritable Environmental Variables

Previous work by Tchetgen et al.⁴² has shown that misspecifying the functional form of an environmental variable can induce heteroskedasticity into tests for GxE interactions. The authors further show that use of robust standard errors will control for this heteroskedasticity but only if the environment is independent of the variant being tested. Independence between genotypes and the misspecified environment is important because it means that the (least-squares) mean estimator is still unbiased.

However, environmental variables themselves often have a genetic basis. We therefore performed simulations where the phenotype depended on the non-linear (squared) effect of a heritable environmental variable. In simulation (Figures 2A and 2C), we observed that misspecification of the environmental variable can cause substantial inflation in GxE test statistics at heritable sites of the confounding environment. Relatively smooth non-linearities, such as squared effects, are easily detected by regression modeling before using LEMMA (see Materials and Methods) and can then be included as covariates [indicated by (+SQE) in Figure 2]. This procedure produced well-calibrated test statistics for all methods in simulation (Figure 2C).

In Figure 2B we compare the GxE association test statistics from our analysis of \log BMI in the UK Biobank with and without adjusting for detected squared effects. Although we detected squared effects for 30 of the 42 environmental variables (significance level 0.01; Bonferroni correction for multiple testing), the ES obtained from the two analyses was almost identical (Pearson $r^2 > 0.999$).

Because the additional variance explained collectively by the squared effects was negligible (incremental $R^2 < 0.00001$), it would be surprising if this was not the case. Negative $\log_{10}(p)$ values from the two analyses were also highly correlated (Pearson $r^2 = 0.961$), although there were small changes in the p values at the *FOXO3* locus (which remained genome-wide significant in both analyses) and at the *SNAP25* locus [which was genome-wide significant in the (–SQE) analysis only]. We therefore conclude that the influence from this form of confounding in our analysis of \log BMI was minor. However, because the cost to this procedure is small, LEMMA uses the (+SQE) strategy by default for all analyses of UK Biobank traits.

GxE Interaction Analysis in the UK Biobank

We applied LEMMA to characterize GxE interactions in BMI (\log BMI), SBP, DBP, and PP by using a set of 42 environmental variables similar to those used in previous analyses,^{7,8,50} including data on smoking, hours of TV watched, Townsend index, physical exercise, and alcohol consumption (see Materials and Methods and Table S1).

We analyzed GxE heritability due to multiplicative effects with the ES by using both $M = 639,005$ genotyped SNPs and $M = 10,270,052$ common imputed SNPs (MAF ≥ 0.01 in the full UK Biobank cohort) stratified by MAF and LDscore into 20 components. Using imputed SNPs, we estimated GxE heritability of 9.3%, 12.5%, 3.9%, and 1.6% for \log BMI, PP, SBP, and DBP, respectively (see Table 1). On genotyped SNPs, the GxE heritability estimates were slightly lower for \log BMI and PP ($h_{GxE}^2 = 8.6\%$ and $h_{GxE}^2 = 11.1\%$, respectively) and almost identical for SBP and DBP (see Table S2). For all traits, the heritability of additive SNP effects was slightly higher on imputed data, consistent with previous results.⁵²

Table 1. Partitioned Heritability Estimates for Four Quantitative Traits in the UK Biobank

Trait	h^2_G (SE)	$h^2_{G \times E}$ (SE)
logBMI	0.274 (0.056)	0.093 (0.028)
INT (BMI)	0.278 (0.056)	0.059 (0.024)
BMI	0.268 (0.055)	0.137 (0.031)
PP	0.228 (0.051)	0.125 (0.028)
SBP	0.251 (0.05)	0.039 (0.023)
DBP	0.254 (0.05)	0.016 (0.02)

Heritability estimates obtained with common imputed SNPs (MAF > 0.01 in the full UK Biobank cohort) with RHE-LDMS. GxE heritability estimates were obtained via the ES from each model fit. All analyses controlled for the same covariates used in the WGR analysis (including the top 20 principal components). Abbreviations are as follows: SE, standard error estimated with the block jackknife (see [Materials and Methods](#)); h^2_G , heritability due to additive genetic effects; $h^2_{G \times E}$, heritability due to multiplicative GxE effects; RHE, randomized HE regression;^{27,28} LDMS, SNPs stratified by minor allele frequency and LDscore (20 components); INT, inverse normal transform applied to males and females separately.

When working with quantitative traits, it can be hard to choose an optimal transformation or scale for each trait. Tyrell et al.²⁶ analyzed BMI by using the raw scale and then also by transforming to a standard normal distribution. They observed larger interaction effects on the raw scale and suggested that this was due to larger variance in BMI in individuals in the high-risk environment groups, which causes heteroskedasticity and inflates effect estimates. In addition to our main analysis, which used log BMI, we re-ran LEMMA by using the raw BMI measurement and then also by transforming to a standard normal distribution in females and males separately. These results are presented in [Table 1](#) and agree with the results of Tyrell et al.²⁶ estimates of GxE heritability on the raw, log, and inverse normal scales were 13.7%, 9.3%, and 5.9%, respectively.

Previous work on models of natural selection has suggested that the variance explained by additive SNP effects should be uniformly distributed as a function of MAF in a neutral evolutionary setting⁵⁴ and that enrichment of the variance explained by low-frequency SNPs is evidence for negative selection. For all four traits, we found that variance explained by the additive genetic effects of low-frequency SNPs (MAF < 0.1) was slightly elevated, consistent with previous observations of negative selection⁴⁶ ([Figure 3](#)). Additionally, the distribution of additive genetic effects by MAF for logBMI was qualitatively similar to that found by GREML-LDMS in a previous study.⁵² In contrast, we found that variance explained by GxE effects was overwhelmingly attributed to low-frequency SNPs (MAF < 0.01), especially those with low LD. However, we are not aware of any evolutionary theory that has been extended to model the MAF distribution of GxE effects.

For logBMI, we estimated an ES that put high weight on alcohol intake frequency, Townsend index, and physical activity measures ([Figure 4C](#)). Almost all of the non-dietary

environmental exposures had a higher effect in women than in men; smoking status was the one exception. This is reflected in the facts that (A) the ES has much higher variance in women and (B) those with a negative ES were almost all female (97%) (see [Figure 4B](#)). When comparing the characteristics of those in the bottom 5% of the ES to the whole cohort (by using the mean for continuous variables and the mode for categorical), we found that those in the bottom 5% were predominantly female (100% versus 53%), younger (51 years versus 56 years), had a higher Townsend deprivation index (0.91 versus -1.74), drank less often (“special occasions” versus “once or twice a week”), and watched more TV (3.28 h versus 2.69 h of TV daily) ([Table S3](#)). We note that positive values of the Townsend index indicate material deprivation, whereas negative values indicate relative affluence.

Previous cross-sectional studies have reported GxE interactions between a linear predictor formed from BMI-associated SNPs and alcohol intake frequency,⁵⁵ Townsend index,^{26,55} physical activity measures,^{5,26,55,56} and time watching TV,^{26,55,56} all of which had high relative weight in the logBMI ES. An alternative approach from Robinson et al.⁶ binned samples according to their environmental exposure (e.g., age) and tested for significant differences in SNP heritability by using a likelihood ratio test. They reported strong interaction effects with age in a cohort of 43,407 individuals whose ages spanned 18–80 years but only reported significant interactions with smoking in the UK Biobank interim release. This suggests that we might expect age to play a more dominant role in the logBMI ES in a cohort that included younger individuals. Finally, one category that is notably down weighted is the contribution from dietary variables. Although significant interactions with fried food consumption⁵⁷ and sugar sweetened drinks⁵⁸ have previously been reported in a cohort of US health professionals, these dietary variables were not included in the diet questionnaire used by the UK Biobank.

The ES for PP was dominated by the effects of age and gender (age, age², age-x-gender, and gender together explained 94.9% of variance in the ES). The magnitude of the ES was strongly associated with increased age,⁵⁹ whereas the sign of the ES was strongly associated with gender, implying that GxE effects were stronger in the elderly but acted in the opposite direction in men and women ([Figure S7](#)).

Similarly, we observed that variance of the ES increased with age in both SBP and DBP, but instead of age itself being highly weighted, we found that age interactions with other environmental variables were most important for explaining variation in the ES. Specifically, for SBP, we found that age interactions with smoking, Townsend index, and alcohol frequency explained 86% of variance in the ES ([Figure S8](#)). When compared to the cohort average, we found that participants in the top 5% of the SBP ES were older (63 years versus 58 years), had a higher Townsend deprivation index (1.2 versus -1.74), and were more likely to smoke (59% versus 9%), whereas those in the bottom

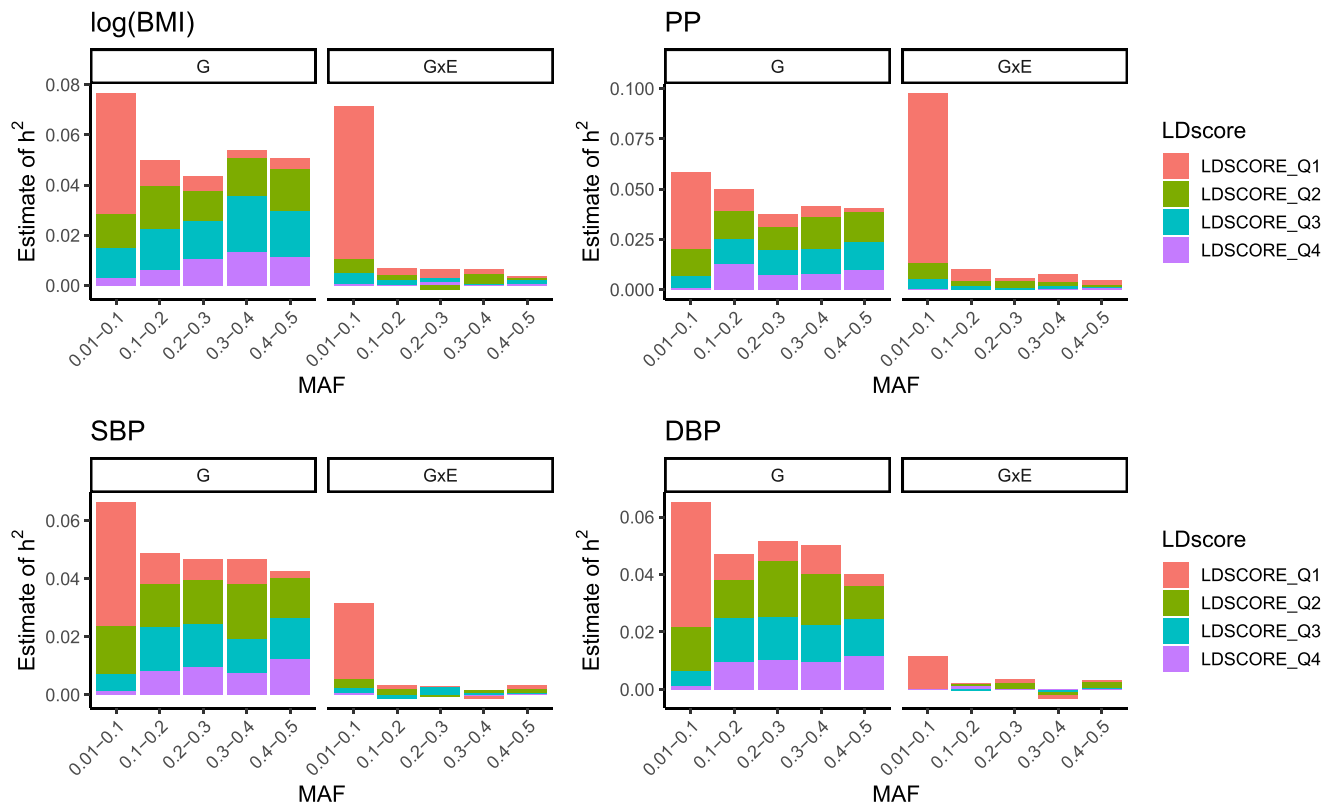


Figure 3. Partitioned Heritability Estimates for Four Quantitative Traits in the UK Biobank

Heritability estimates partitioned into additive genetic and multiplicative GxE interaction effects for four quantitative traits in the UK Biobank with approximately 280,000 unrelated white British individuals (see Table S1) and $M = 10,270,052$ common imputed SNPs ($MAF > 0.01$ in the full UK Biobank cohort). Multiplicative GxE interactions were computed with the ES from each model fit. Heritability estimation was performed via a multi-component implementation of RHE regression^{27,28} with SNPs stratified into 20 components (5 MAF bins and 4 LD score quantiles).

5% were also older (65 years versus 58 years), predominantly female (91.5%), rarely drank alcohol (43.9% drank “never”), and had a low Townsend deprivation index (−2.9 versus −1.74) (Table S4).

Finally, we observed notably higher variance in the ES for DBP among men, most of which appeared to be driven by high gender-specific weights for smoking status and alcohol frequency (Figure S9). We further observed that alcohol frequency and smoking status became increasingly influential with age. The total SNP-GxE heritability for this ES, however, was quite low.

When testing for significant GxE interactions between the estimated ESs and imputed markers across the genome, we observed that use of the robust standard errors made a noticeable difference to the calibration of LEMMA (Figure S10 and Table S5). We identified two loci for logBMI (Figure 4), one locus for DBP (Figure S9A), and zero loci for SBP and PP by using a threshold of 5×10^{-8} for genome-wide significance (Table 2). This table also includes results from a standard linear regression GWAS test at the three loci. Table S6 provides full parameter estimates of the environmental, SNPs, and SNP-ES effects.

For logBMI, LEMMA identified GxE interactions at rs2153960 ($p = 6.5 \times 10^{-9}$; Figure S11) and at rs539515

($p = 6.5 \times 10^{-12}$; Figure S12). The SNP rs2153960 is an intron in *FOXO3* and has been previously associated with insulin-like growth factor 1 (IGF-1) concentration in a cohort of 10,000 middle-aged Europeans.⁶⁰ IGF-1 is known to be a central mediator of metabolic, endocrine, and anabolic effects of growth hormones and is also involved in carbohydrate homeostasis.⁶⁰ The patterns of main effect association and GxE association show considerable overlap (Figure S11A). This SNP did not reach genome-wide levels of significance with the standard linear regression GWAS test (Table 2).

The SNP rs539515 is located 6 kb downstream of *SEC16B*. The patterns of main effect association and GxE association are very similar (Figure S12A). Multiplicative GxE interactions have been reported at *SEC16B* with multiple environmental variables in a similar analysis in the UK Biobank⁷ and with physical activity separately in Europeans⁵ ($p = 0.025$) and in Hispanics⁶¹ ($p = 8.1 \times 10^{-5}$). Highly significant variance effects ($p = 3.88 \times 10^{-17}$), which can be indicative of GxE, have also been reported at the *SEC16B* locus via $N = 456,422$ Europeans in the UK Biobank.⁶² *SEC16B* transcribes one of the two mammalian orthologs of SEC16, which has a key role in organizing endoplasmic reticulum exit sites by interacting with COPII components.⁶³ Although several GWASs have identified

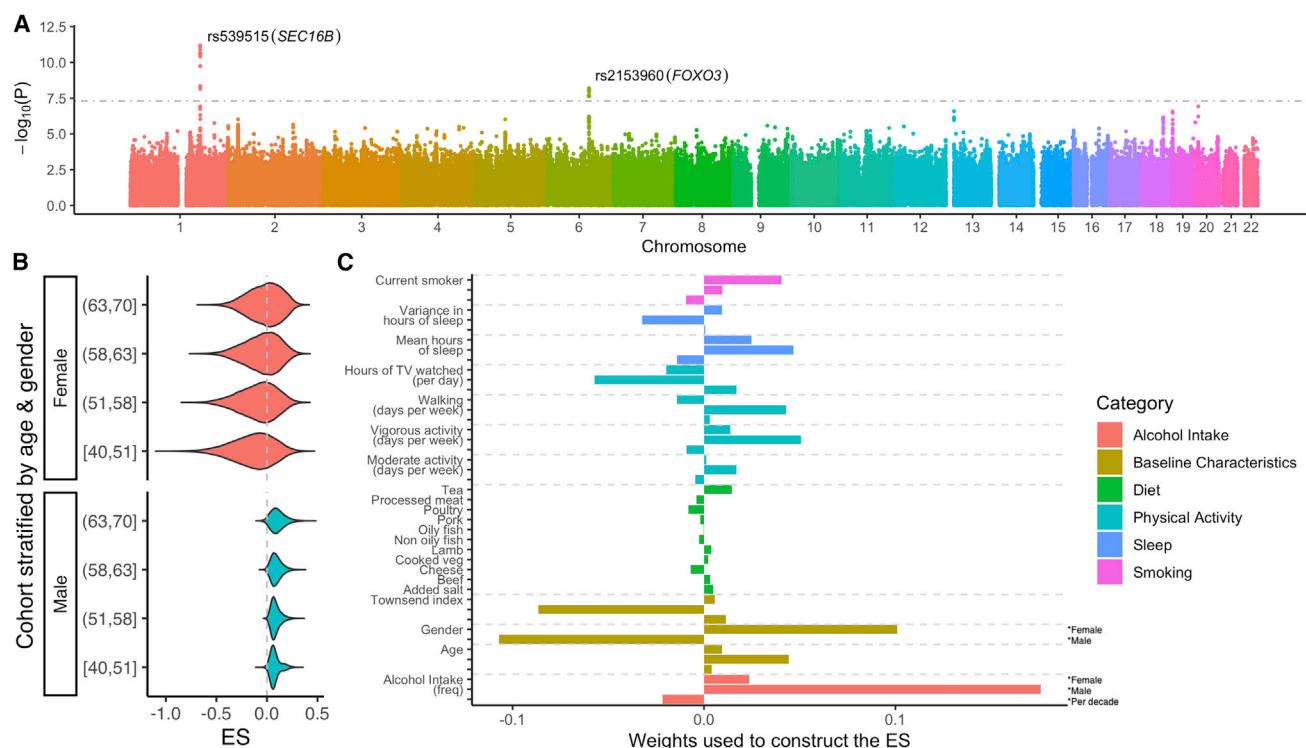


Figure 4. GxE Analysis of \log BMI in the UK Biobank

(A) LEMMA association statistics testing for multiplicative GxE interactions at each SNP. The horizontal gray line denotes ($p = 5 \times 10^{-8}$), and p values are shown on the $-\log_{10}$ scale.

(B) Distribution of the environmental score (ES), stratified by gender and age quantile.

(C) Weights used to construct the ES. Dietary variables have a single weight shown on the per-standard-deviation (SD) scale. “Gender” has two weights; a gender-specific intercept for women (first) and for men (second). Remaining non-dietary variables have three weights: (1) a per-SD effect for women only, (2) a per-SD effect for men only, and (3) a per-SD, per-decade effect, which is the same for both genders. SD for the male- and female-specific weights is computed for each gender separately. Age is computed as the number of decades aged from 40 years. See [Materials and Methods](#) for details.

associations between *SEC16B*,^{64,65} the relevance of *SEC16B* to BMI is not well characterized.⁶⁶ Some evidence exists to suggest that *SEC16B* has role in the transport of peroxisome biogenesis factors; peroxisomes are an organelle involved in the catabolism of long-chain fatty acids found ubiquitously in eukaryotic cells. Previous authors⁶⁵ have also speculated that *SEC16B* might play a role in the transport of appetite regulatory peptides; however, we are not aware of any evidence for this theory.

The DBP-associated SNP is rs8090962, but it only just passes our threshold for significance and we are least confident that this is a true GxE association for a few reasons. The SNP is located within an enhancer, approximately 100 kb downstream of *SEC11C* and 50 kb upstream of *ZNF532*. Neither gene has previously been associated with blood pressure traits. There is some evidence of a main effect close by ([Figure S13A](#)), but the pattern of main effect association does not coincide well with the pattern of GxE association. In addition, the pattern of GxE association by genotype ([Figure S13B](#)) shows a striking cross over by genotype between extremes of the ES. We have observed above that our test statistics are very slightly inflated, so this could be a false positive association.

Relationship of Genetic PCs and Environmental Scores

We regressed the estimated ESs against the PCs for each of the four UK Biobank traits, and the results are included in [Table S7](#). We found some significant associations, mostly with PC5, which seems to correlate with North-South geography in the UK.⁶⁷ To explore further, we also re-ran the heritability analysis by including interaction terms of the ES with the genetic PCs as control variables, but the results were almost unchanged (see [Table S8](#)).

Comparison of the LEMMA ES with a Marginal ES

For each trait, we used least-squares regression to compute a linear model fit using all of the non-genetic covariates used in the LEMMA analysis. We then constructed an ES (referred to as ES_{marginal}) by using the marginal environmental effects from this model fit. The correlation between the LEMMA ES and ES_{marginal} was -0.062 , -0.019 , -0.297 , and -0.088 for \log BMI, PP, SBP, and DBP, respectively, suggesting that these vectors are quite dissimilar. [Figure S14](#) shows a comparison of the interaction weights used to construct the LEMMA ES and ES_{marginal} for each of the four traits. Visually, the weights learnt through each approach look quite distinct. In particular, age, age², and

Table 2. Loci with Genome-wide Significant GxE Interaction Effects with the ES

SNP	rs539515	rs2153960	rs8090962
Trait	log(BMI)	log(BMI)	DBP
Chr	1	6	18
BP	177889025	108988184	56694404
A0	A	G	A
A1	C	A	G
AF	0.21	0.71	0.44
Nearest gene	<i>SEC16B</i>	<i>FOXO3</i>	<i>OACYLP / SEC11C</i>
Standard GWAS Tests			
β_G (SE)	0.0043 (0.0003)	0.0014 (0.0003)	-0.00001 (0.0192)
p value _G	5.7×10^{-51}	1.1×10^{-6}	$1.0 \times 10^{+0}$
LEMMA Association Tests			
β_G (SE)	0.0254 (0.0016)	0.0087 (0.0015)	0.0011 (0.0015)
$\beta_{G \times E}$ (SE)	-0.0117 (0.0017)	-0.0098 (0.0017)	0.0087 (0.0016)
p value _G	1.6×10^{-60}	1.6×10^{-8}	4.5×10^{-1}
p value _{G×E}	6.5×10^{-12}	6.5×10^{-9}	3.6×10^{-8}

Independent loci with genome-wide significant ($p < 5 \times 10^{-8}$) GxE interaction effects with the environmental score (ES). Loci at least 0.5 cM apart were judged to be independent. SNP effect sizes reported on a per-SD scale. SNP locations follow the GrCh37 human genome assembly. All loci had an IMPUTE info score >0.99 . Abbreviations are as follows: BP, base pairs; A0, reference allele; A1, alternative allele; AF, reference allele frequency; SD, standard deviation.

age \times gender have much higher relative weights in ES_{marginal} than in the LEMMA ES.

Comparison of Methods on UK Biobank Data

To compare LEMMA with existing single SNP methods, we also ran StructLMM, the F-test, and the robust F-test on *logBMI* by using the same set of environmental variables as used by LEMMA (but not including the significant squared environments as covariates). Manhattan plots are displayed in Figure S15. Test statistics from both the F-test ($\lambda_{GC} = 1.37$) and StructLMM ($\lambda_{GC} = 1.235$) were substantially inflated when compared to the robust F-test and LEMMA ($\lambda_{GC} = 1.03$ and $\lambda_{GC} = 1.062$, respectively; see Table S5), suggesting that StructLMM does not properly control for heteroskedasticity. There are clear differences between the four methods, especially among SNPs with suggestive evidence of GxE interaction results (Figure S16). LEMMA did not find the *FTO* locus, StructLMM and F-test did not find the *SEC16B* locus, and the robust F-test only found the *FTO* locus.

LEMMA relies on the assumption that all GxE interaction effects for a single trait share a common ES, and we have shown in simulation that, when this assumption holds, LEMMA achieves substantial increases in power. However, we would expect LEMMA to have little power to detect SNPs that interact with a combination of environments that are not well correlated with the genome-wide ES estimated by LEMMA. *FTO* seems to be one clear example of this. We extracted an estimate of the SNP-spe-

cific interaction profile at *FTO* by using the robust F-test (Materials and Methods), and we found that its correlation with LEMMA's ES was low (Pearson $r^2 = 0.3$). In comparison, a similar analysis at *SEC16B* and *FOXO3* yielded much higher correlations (Pearson $r^2 = 0.725$ and $r^2 = 0.713$, respectively).

Discussion

In this study, we proposed a WGR method, LEMMA, that estimates a single ES that interacts with SNPs across the genome. In simulation, we have demonstrated that the ES can be used to compute well-calibrated p values of the multiplicative interaction effect at each SNP. LEMMA is also able to quantify the trait variance attributable to MAF- and LD-stratified interaction effects of the ES.

In our analyses of four quantitative traits in the UK Biobank, we have demonstrated that GxE effects among common imputed SNPs make a non-trivial contribution to the heritability of *logBMI* and PP (9.3% and 12.5%, respectively). Our stratified heritability analysis has suggested that GxE interactions for these traits are mostly driven by low-frequency variants. Our analysis identified three loci with statistically significant GxE interaction effects. As far as we are aware, two of these loci, rs539515 (*FOXO3*) and rs8090962, are novel, and for the other, rs539515 (*SEC16B*), we show stronger evidence for statistically significant GxE interaction effects than the previous study.⁷

Robinson et al.⁶ have previously attempted to quantify the contribution of GxE interactions to the heritability of

BMI in a study performed on imputed SNPs from the interim UK Biobank release. Using the GCI-GREML model implemented in GCTA⁶⁸ and eight environmental variables that included measures of smoking, hours of TV watched, and alcohol frequency, Robinson et al.⁶ reported that only smoking had significant GxE heritability (4.0%). In contrast, the ES estimated for *log*BMI in our analysis had non-zero contributions from many environmental variables, including hours of TV watched and smoking, suggesting that multiple environmental variables can influence the genetic predisposition to BMI. Modeling these environmental variables jointly allowed LEMMA to capture a combination whose GxE interactions explained 9.3% of heritability.

We have also evaluated the performance of three existing single SNP methods (StructLMM, the F-test, and a robust F-test) both in simulation and on *log*BMI from this same dataset. In simulation with large datasets, we observed that StructLMM and the F-test had similar performance, an observation that also held in our analysis of *log*BMI. Both of these methods appeared vulnerable to heteroskedasticity, which we showed is likely to occur in traits with non-trivial GxE heritability. A simple adjustment, using “robust” or Huber-White variance estimators, solved this problem. The two F-test methods further benefit from a wealth of existing theory⁴¹ and, being theoretically simpler than StructLMM, could be easily implemented as an R-plugin with PLINK⁶⁹ (for example³⁷). In our opinion, the robust F-test is therefore the most appropriate of the three single SNP methods to model GxE effects with tens of environments in biobank-scale datasets.

Although LEMMA represents a method with increased power to detect GxE interaction effects, our approach does have some caveats. First, the gain in power is dependent on a strong assumption on the underlying genetic architecture. Although our analysis suggests that this does hold to some extent for PP and *log*BMI, this may not be the case for other traits.

In addition, LEMMA only estimates the proportion of phenotypic variance that is explained by interactions with this ES and we do not claim that this captures all the GxE heritability of a trait. If relevant GxE environments are not included in the analysis, and these environments have low correlation to the environments that are included, then LEMMA cannot account for them and will most likely underestimate the true GxE heritability. Unobserved environments can cause trait variance to depend on genotype⁸ (see Figure S19), and extending LEMMA in this direction is left for future work.

LEMMA has the requirement that none of the environmental variables have any missing values. This could lead to a reduction in samples size if many environmental variables are included. If the amount of missing data is small, it should not pose a big problem, and missing data imputation methods are also an option. If LEMMA is applied in situations where the missing data structure is related

to the phenotype of interest, then this could cause bias in the results.

Despite much effort to provide an efficient implementation, the LEMMA algorithm is still computationally demanding. Using randomized HE regression to estimate an improved initialization of the interaction’s weights may help to reduce runtime and is an avenue that we are currently pursuing.

Finally, for simplicity, LEMMA currently searches only for GxE interactions with a single linear combination of environments. Generalizing the LEMMA approach to several orthogonal linear combinations or using functional annotation to restrict the SNPs that each ES interacts with may yet yield more power to identify interactions in complex traits and explain more phenotypic variation.

Data and Code Availability

The genetic and phenotype datasets generated by UK Biobank analyzed during the current study are available via the UK Biobank data access process. The resource is available to all bona fide researchers from academic, charity, public, and commercial sectors for all types of health-related research that is in the public interest: there is no preferential or exclusive access for any person. More details are available at <http://www.ukbiobank.ac.uk/register-apply/>. Pre-compiled binaries and source code for the LEMMA method are available from <https://jmarchini.org/lemma/>.

Supplemental Data

Supplemental Data can be found online at <https://doi.org/10.1016/j.ajhg.2020.08.009>.

Acknowledgments

Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. Financial support was provided by the Wellcome Trust Core Award (203141/Z/16/Z). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. M.K. was supported by a Wellcome Trust funding grant (109106/Z/15/Z). We are grateful to Kevin Sharp, David Steinsaltz, and Helen Warren for discussions about this work. J.M. carried out this work while affiliated with the University of Oxford.

Declaration of Interests

J.M. owns stocks and stock options in Regeneron Pharmaceuticals.

Received: October 28, 2019

Accepted: August 11, 2020

Published: September 3, 2020

Web Resources

StructLMM as implemented in LIMIX 2.0.0, <https://github.com/limix/limix>

References

- Hunter, D.J. (2005). Gene-environment interactions in human diseases. *Nat. Rev. Genet.* 6, 287–298.
- Smith, P.G., and Day, N.E. (1984). The design of case-control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.* 13, 356–365.
- 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.
- Kilpeläinen, T.O., Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C.H., et al. (2011). Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med.* 8, e1001116.
- Ahmad, S., Rukh, G., Varga, T.V., Ali, A., Kurbasic, A., Shungin, D., Ericson, U., Koivula, R.W., Chu, A.Y., Rose, L.M., et al.; InterAct Consortium; and DIRECT Consortium (2013). Gene \times physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. *PLoS Genet.* 9, e1003607.
- Robinson, M.R., English, G., Moser, G., Lloyd-Jones, L.R., Triplett, M.A., Zhu, Z., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., Esko, T., et al.; LifeLines Cohort Study (2017). Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* 49, 1174–1181.
- Moore, R., Casale, F.P., Jan Bonder, M., Horta, D., Franke, L., Barroso, I., Stegle, O.; and BIOS Consortium (2019). A linear mixed-model approach to study multivariate gene-environment interactions. *Nat. Genet.* 51, 180–186.
- 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.
- de Leeuw, C.A., Stringer, S., Dekkers, I.A., Heskes, T., and Posthuma, D. (2018). Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. *Nat. Commun.* 9, 3768.
- Burgoine, T., Sarkar, C., Webster, C.J., and Monsivais, P. (2018). Examining the interaction of fast-food outlet exposure and income on diet and obesity: Evidence from 51,361 UK Biobank participants. *Int. J. Behav. Nutr. Phys. Act.* 15, 1–12.
- Thompson, W.D. (1991). Effect modification and the limits of biological inference from epidemiologic data. *J. Clin. Epidemiol.* 44, 221–232.
- Siemiatycki, J., and Thomas, D.C. (1981). Biological Models and Statistical Interactions: an Example from Multistage Carcinogenesis. *Int. J. Epidemiol.* 10, 383–387.
- Hill, W.G., and Mackay, T.F.C.D.S. (2004). D. S. Falconer and Introduction to quantitative genetics. *Genetics* 167, 1529–1536.
- Kraft, P., Yen, Y.-C., Stram, D.O., Morrison, J., and Gauderman, W.J. (2007). Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* 63, 111–119.
- Listgarten, J., Lippert, C., Kadie, C.M., Davidson, R.I., Eskin, E., and Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nat. Methods* 9, 525–526.
- Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* 46, 100–106.
- Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* 47, 284–290.
- Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723.
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835.
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824.
- Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- de Los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., and Calus, M.P.L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345.
- Logsdon, B.A., Hoffman, G.E., and Mezey, J.G. (2010). A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC Bioinformatics* 11, 58.
- Carbonetto, P., and Stephens, M. (2012). Scalable variational inference for bayesian variable selection in regression, and its accuracy in genetic association studies. *Bayesian Anal.* 7, 73–108.
- Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9, e1003264.
- Tyrrell, J., Wood, A.R., Ames, R.M., Yaghootkar, H., Beaumont, R.N., Jones, S.E., Tuke, M.A., Ruth, K.S., Freathy, R.M., Davey Smith, G., et al. (2017). Gene-obesogenic environment interactions in the UK Biobank study. *Int. J. Epidemiol.* 46, 559–575.
- Wu, Y., and Sankararaman, S. (2018). A scalable estimator of SNP heritability for biobank-scale data. *Bioinformatics* 34, i187–i194.
- Pazokitoroudi, A., Wu, Y., Burch, K.S., Hou, K., Pasaniuc, B., and Sankararaman, S. (2019). Scalable multi-component linear mixed models with application to SNP heritability estimation. *bioRxiv*. <https://doi.org/10.1101/522003>.
- 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.; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and RACI Consortium (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235.
- Speed, D., Cai, N., Johnson, M.R., Nejentsev, S., Balding, D.J.; and UCLEB Consortium (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49, 986–992.

31. 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.
32. Speed, D., and Balding, D.J. (2019). SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat. Genet.* **51**, 277–284.
33. Erbe, M., Hayes, B.J., Matukumalli, L.K., Goswami, S., Bowman, P.J., Reich, C.M., Mason, B.A., and Goddard, M.E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129.
34. Murphy, K.P., and Bach, F. (2012). *Machine Learning: A Probabilistic Perspective* (MIT Press).
35. Hoffman, M.D., Blei, D.M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14**, 1303–1347.
36. Varadhan, R., and Roland, C. (2008). Simple and globally convergent methods for accelerating the convergence of any em algorithm. *Scand. J. Stat.* **35**, 335–353.
37. Almli, L.M., Duncan, R., Feng, H., Ghosh, D., Binder, E.B., Bradley, B., Ressler, K.J., Conneely, K.N., and Epstein, M.P. (2014). Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392–1399.
38. Wilcox, R.R. (2011). *Introduction to robust estimation and hypothesis testing* (Academic press).
39. Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 221–233.
40. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838.
41. Greene, W.H. (2003). *Econometric Analysis*, Fifth Edition (Pearson Education India).
42. Tchetgen Tchetgen, E.J., and Kraft, P. (2011). On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. *Epidemiology* **22**, 257–261.
43. Voorman, A., Lumley, T., McKnight, B., and Rice, K. (2011). Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS ONE* **6**, e19416.
44. Long, J.S., and Ervin, L.H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *Am. Stat.* **54**, 217–224.
45. Guan, Y., and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Ann. Appl. Stat.* **5**, 1780–1815.
46. 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.
47. Haseman, J.K., and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19.
48. 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.
49. Tobin, M.D., Sheehan, N.A., Scurrah, K.J., and Burton, P.R. (2005). Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935.
50. Young, A.I., Wauthier, F., and Donnelly, P. (2016). Multiple novel gene-by-environment interactions modify the effect of FTO variants on body mass index. *Nat. Commun.* **7**, 12724.
51. Evans, L.M., Tahmasbi, R., Vrieze, S.J., Abecasis, G.R., Das, S., Gazal, S., Bjelland, D.W., de Candia, T.R., Goddard, M.E., Neale, B.M., et al.; Haplotype Reference Consortium (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* **50**, 737–745.
52. Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., Robinson, M.R., Perry, J.R., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al.; LifeLines Cohort Study (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120.
53. 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.
54. Visscher, P.M., Goddard, M.E., Derks, E.M., and Wray, N.R. (2012). Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol. Psychiatry* **17**, 474–485.
55. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2017). Gene-environment interaction study for BMI reveals interactions between genetic factors and physical activity, alcohol consumption and socioeconomic status. *PLoS Genet.* **13**, e1006977.
56. Qi, Q., Li, Y., Chomistek, A.K., Kang, J.H., Curhan, G.C., Pasquale, L.R., Willett, W.C., Rimm, E.B., Hu, F.B., and Qi, L. (2012). Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation* **126**, 1821–1827.
57. Qi, Q., Chu, A.Y., Kang, J.H., Huang, J., Rose, L.M., Jensen, M.K., Liang, L., Curhan, G.C., Pasquale, L.R., Wiggs, J.L., et al. (2014). Fried food consumption, genetic risk, and body mass index: gene-diet interaction analysis in three US cohort studies. *BMJ* **348**, g1610.
58. Qi, Q., Chu, A.Y., Kang, J.H., Jensen, M.K., Curhan, G.C., Pasquale, L.R., Ridker, P.M., Hunter, D.J., Willett, W.C., Rimm, E.B., et al. (2012). Sugar-sweetened beverages and genetic risk of obesity. *N. Engl. J. Med.* **367**, 1387–1396.
59. 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.; International Consortium of Blood Pressure (ICBP) 1000G Analyses; BIOS Consortium; Lifelines Cohort Study; Understanding Society Scientific group; CHD Exome+ Consortium; ExomeBP Consortium; T2D-GENES Consortium; GoT2D-Genes Consortium; Cohorts for Heart and Ageing Research in Genome Epidemiology (CHARGE) BP Exome Consortium; International Genomics of Blood Pressure (iGEN-BP) Consortium; and UK Biobank CardioMetabolic Consortium BP working group (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403–415.
60. Kaplan, R.C., Petersen, A.K., Chen, M.H., Teumer, A., Glazer, N.L., Döring, A., Lam, C.S., Friedrich, N., Newman, A., Müller, M., et al. (2011). A genome-wide association study identifies

- novel loci associated with circulating IGF-I and IGFBP-3. *Hum. Mol. Genet.* 20, 1241–1251.
61. Richardson, A.S., North, K.E., Graff, M., Young, K.M., Mohlke, K.L., Lange, L.A., Lange, E.M., Harris, K.M., and Gordon-Larsen, P. (2014). Moderate to vigorous physical activity interactions with genetic variants and body mass index in a large US ethnically diverse cohort. *Pediatr. Obes.* 9, e35–e46.
 62. Wang, H., Zhang, F., Zeng, J., Wu, Y., Kemper, K.E., Xue, A., Zhang, M., Powell, J.E., Goddard, M.E., Wray, N.R., et al. (2019). Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank. *Sci. Adv.* 5, w3538.
 63. Bhattacharyya, D., and Glick, B.S. (2007). Two mammalian Sec16 homologues have nonredundant functions in endoplasmic reticulum (ER) export and transitional ER organization. *Mol. Biol. Cell* 18, 839–849.
 64. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadóttir, A., Styrkarsdóttir, U., Gretarsdóttir, S., Thorlacius, S., Jonsdóttir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* 41, 18–24.
 65. Hotta, K., Nakamura, M., Nakamura, T., Matsuo, T., Nakata, Y., Kamohara, S., Miyatake, N., Kotani, K., Komatsu, R., Itoh, N., et al. (2009). Association between obesity and polymorphisms in SEC16B, TMEM18, GNPDA2, BDNF, FAIM2 and MC4R in a Japanese population. *J. Hum. Genet.* 54, 727–731.
 66. Schmid, P.M., Heid, I., Buechler, C., Steege, A., Resch, M., Birner, C., Endemann, D.H., Riegger, G.A., and Luchner, A. (2012). Expression of fourteen novel obesity-related genes in Zucker diabetic fatty rats. *Cardiovasc. Diabetol.* 11, 48.
 67. Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K.E., Nivard, M.G., Veul, L., Holtz, Y., Zietsch, B.P., Frayling, T.M., Wray, N.R., et al. (2019). Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* 3, 1332–1342.
 68. 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.
 69. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.