# UCSF
**UC San Francisco Previously Published Works**

**Title**
A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits

**Permalink**
https://escholarship.org/uc/item/3nb608mj

**Journal**
American Journal of Human Genetics, 106(1)

**ISSN**
0002-9297

**Authors**
Dahl, Andy
Nguyen, Khiem
Cai, Na
et al.

**Publication Date**
2020

**DOI**
10.1016/j.ajhg.2019.11.015

Peer reviewed

# A Robust Method Uncovers Significant Context-Specific Heritability in Diverse Complex Traits

Andy Dahl,[1,2,*] Khiem Nguyen,[2] Na Cai,[3,4] Michael J. Gandal,[5] Jonathan Flint,[6] and Noah Zaitlen[1,2,*]

Gene-environment interactions (GxE) can be fundamental in applications ranging from functional genomics to precision medicine and is a conjectured source of substantial heritability. However, unbiased methods to profile GxE genome-wide are nascent and, as we show, cannot accommodate general environment variables, modest sample sizes, heterogeneous noise, and binary traits. To address this gap, we propose a simple, unifying mixed model for gene-environment interaction (GxEMM). In simulations and theory, we show that GxEMM can dramatically improve estimates and eliminate false positives when the assumptions of existing methods fail. We apply GxEMM to a range of human and model organism datasets and find broad evidence of context-specific genetic effects, including GxSex, GxAdversity, and GxDisease interactions across thousands of clinical and molecular phenotypes. Overall, GxEMM is broadly applicable for testing and quantifying polygenic interactions, which can be useful for explaining heritability and invaluable for determining biologically relevant environments.

## Introduction

Many examples of gene-environment interaction (GxE) have been documented in humans at the level of individual genetic variants. In functional genomics, variants can have effects on expression that depend on external context,[1–5] age,[6] tissue,[7] or cell type.[8,9] In complex traits, some genetic variants are known to interact with air pollution,[10] microbe exposure,[11] or sex.[12–15] Variants can also interact with medical interventions and can render certain treatments ineffective.[16–18]

GxE tests can explain novel biology along two distinct, complementary axes. First, GxE can identify unappreciated genetic effects that elude linear models, which can increase genome-wide association study (GWAS) power and has recently received attention as a partial answer to the missing heritability question.[19–21] Second, GxE tests can demonstrate that an environmental measurement is biologically trait relevant and quantify its impact, which can be important for public health and can illuminate intrinsic trait biology. More generally, genetic interaction tests can be used to assess the biological significance of any sample stratification, including (putative) phenotypic subtypes, quantitative covariates like age, or even other genetic variants.

However, investigation of GxE in nominally unrelated humans has been limited by the fact that individual variant effects are typically small. In the additive model context, genome-wide Genomic Restricted Maximum Likelihood (GREML) has been an invaluable complement to variant-based tests, powerfully characterizing aggregate polygenic effect sizes without attempting to resolve causal variants.[22] Here, we propose an analogous mixed model

for GxE, GxE Mixed Model (GxEMM), to characterize the aggregate polygenic contributions of GxE. Relative to single-variant GxE tests, GxEMM has higher power (for polygenic traits) but lower resolution, a tradeoff that makes it particularly useful for characterizing the biological relevance of environmental measures.

While complex in some settings, GxEMM is easily visualized for discrete environments (Figure 1). GxEMM incorporates three key, interpretable models. First, the Hom model is equivalent to standard GREML and models only the mean effects of the environment. We call the second model IID, as it allows environment-specific genetic variance and noise but assumes that these values are constant across every environment. We call the final model Free, because it allows the genetic and noise levels to freely vary between environments. We propose these three core models as a parsimonious, biologically informative representation of possible GxE models.

Although GxEMM is simplest in the case of two discrete environments, GxEMM extends naturally to accommodate arbitrary environmental covariates. This is important because previous approaches require discrete environments (e.g., MV-GREML[20]) or univariate environments (e.g., MRNM[23]). A second significant component of GxEMM is the ability to model binary traits, like case/control disease studies. Third, GxEMM accommodates modest sample sizes, which can be important for applications like functional genomics. To obtain this flexibility, we implement three methods to estimate GxEMM parameters, each extending an existing approach to GREML: REML, for large sample sizes and continuous traits; phenotype-correlation-genotype-correlation (PCGC), for binary
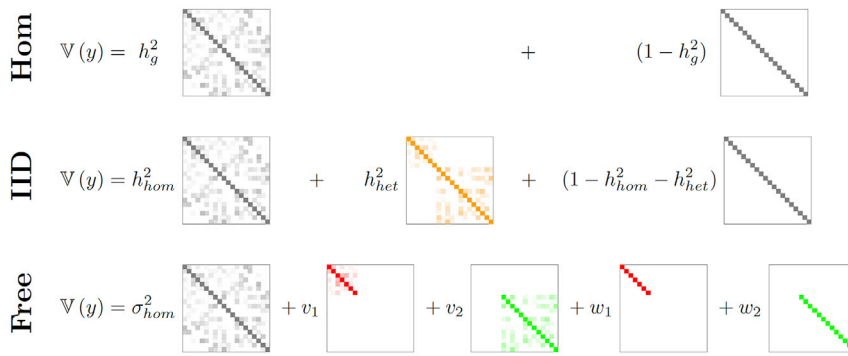
**Figure 1. The Three Key GxEMM Models in the Case of Two Discrete Environments**
Entry shading indicates absolute value. Hom (top) fits homogeneous genetic ($h_g^2$) and noise ($\sigma_e^2$) levels and is equivalent to GREML. IID (middle) adds a common effect for environment-specific heritability ($h_{het}^2$, orange), pictured as a block-diagonal matrix with each block corresponding to samples from one environment. Free (bottom) allows environment-specific genetic ($v_k$) and noise ($w_k$) levels.

traits;[24,25] and Haseman-Elston regression (HE), for modest sample sizes, e.g., hundreds of samples.[26]

We also make three important theoretical contributions. First, we demonstrate that noise heterogeneity must be modeled to avoid false positive GxE (Appendix A). Second, we provide a formally identified model description, which is essential for clear inference and future extensions (Appendix B). Third, we correct a misconception that G-E correlation generally causes GxE bias in the polygenic setting (Appendix C).

In this paper, we first provide a broad overview of GxEMM and the concept of polygenic interaction and then describe the details of the GxEMM method, our simulation study, and the three real datasets we analyze. Next, we perform a series of simulations and theoretical analyses that demonstrate that GxEMM is broadly reliable. We then apply GxEMM to study sex-specific heritability in 115 outbred rat traits and find strong, widespread signals of genetic and non-genetic heterogeneity, especially in bone and glucose-tolerance traits. Next, we find significant polygenic interactions with several stress indices and quantitative environments for major depression. Finally, we analyze RNA-seq data from postmortem brain and find strong evidence for bipolar- and schizophrenia-specific *cis*-heritability, both on average over the transcriptome and in nine known SCZ-associated genes. We conclude with a discussion on important caveats as well as future applications and extensions.

## Material and Methods

### Overview of GxEMM
GxEMM builds on GREML, an additive polygenic model that estimates heritability distributed across the genome. GxEMM additionally captures the GxE-based heritability due to polygenic gene-environment interactions. Assuming discrete environments, the model for phenotype $y_i$ in environment $k$ (i.e., $z_i = k$) is:

In this model, $X$ are covariates with fixed effects α, like age or genetic PCs. $G$ is the genotype matrix, with additive effects β. As in GREML, we assume that β and the noise, ε, are i.i.d. standard normal, and the heritability is determined by the genetic and noise variances, $\sigma_{hom}^2$ and $\sigma_e^2$. GxEMM additionally captures SNP-environment interaction effects, γ, which are also assumed i.i.d. standard normal. Further, GxEMM allows environment-specific genetic ($v_k$) and noise ($w_k$) variances. Although we have assumed that $z$ is discrete to simplify Equation 1 and Figure 1, GxEMM extends to general $z$, e.g., quantitative environments or proportional membership across discrete environments (see LMMs for Polygenic Interaction Effects (GxEMM)).

To unpack the model, imagine studying genetic effects on height across males and females. A SNP $s$ that equally increases height in both sexes has a homogeneous effect ($\beta_s > 0$) but has no sex-specific effects ($\gamma_{sf} = \gamma_{sm} = 0$), so $s$ contributes to $\sigma_{hom}^2$ but not $v_f$ or $v_m$. Conversely, a SNP $s'$ that increases height only in females has $\beta_{s'} = 0$ and $\gamma_{s'f} > \gamma_{s'm} = 0$, so $s'$ contributes to $v_f$ but not $\sigma_{hom}^2$ or $v_m$. Finally, $w_f > w_m$ means that females have higher non-genetic height variance.

We consider three distinct models for $v$: Hom, where all $v_k = 0$; IID, where $v_k = h_{het}^2$ for all $k$; and Free, where $v$ is unconstrained. We use the same models for $w$, although the IID model is not identified in the case of discrete $z$. In the sex-height example, the IID model allows sex-specific genetic effects but assumes that, on balance, male- and female-specific sizes are identical; the Free model eliminates this assumption. In this paper, we assume that the genetic effects are exchangeable across environments, i.e., we assume that $\gamma_{sk}$ and $\gamma_{sk'}$ are independent for $k \neq k'$. We also theoretically describe a Full model allowing arbitrary correlations between environment-specific effects, which is richer but statistically and computationally challenging (see LMMs for Polygenic Interaction Effects (GxEMM)).

Of the many possible statistical tests available for GxEMM, we focus on a parsimonious set of biologically meaningful tests to characterize polygenic GxE:

- Hom versus Null (i.e.: $h_g^2 = 0$?) tests whether there is any heritability.

$$y_i\big|z_i = k \quad \sim \quad \underbrace{\sum_q X_{iq}\alpha_q}_{\text{Covariate}} \quad + \quad \underbrace{\sigma_{hom}\sum_s G_{is}\beta_s}_{\text{Hom G}} \quad + \quad \underbrace{\sqrt{v_k}\sum_s G_{is}\gamma_{sk}}_{\text{GxE}} \quad + \quad \underbrace{\sqrt{w_k + \sigma_\epsilon^2}\,\epsilon_i}_{\text{Noise}} \qquad \text{(Equation 1)}$$

**Table 1. Recent LMMs for GxE Heritability**

| Publication | GxE | IxE | Binary $y$ | Allowed $Z$ | Notes |
|---|---|---|---|---|---|
| GCTA[27] | IID | unbiased | no | discrete | – |
| iSet[28] | full | unbiased | no | discrete | design-specific |
| MetaSex[29]. | free | unbiased | no | discrete | not MLE |
| MV-GREML[20] | full | unbiased | no | discrete | – |
| StructLMM[21] | IID | unbiased | no | general | $L = 1$ SNP |
| RNM[23] | free | biased | no | bivariate | models G-E corr. |
| MRNM[23] | IID | unbiased | no | univariate | models G-E corr. |
| GxEMM (us) | free | unbiased | yes | general | – |

GxE gives the richest fitted genetic heterogeneity model, and IxE indicates whether the method is biased under noise heterogeneity (Appendix A). Methods without specific notes are displayed with dash (–).

- IID versus Hom (i.e.: $h_{het}^2 = 0$?) tests polygenic GxE under homoscedasticity.
- Free versus Hom (i.e.: $v_k = 0$?) tests polygenic GxE allowing heteroscedasticity.

We implement both LR and Wald tests.

In the Free model with discrete $z$, each environment can have different levels of heritability and total variance. Assuming the kinship matrix is normalized, the heritability in environment $k$ is:

$$h_k^2 := \frac{\sigma_{hom}^2 + v_k}{\sigma_{hom}^2 + v_k + w_k + \sigma_e^2} \qquad \text{(Equation 2)}$$

Most previous approaches for polygenic GxE in unrelated samples assume discrete environments (Table 1). The GCTA GxE model fits an IID $v$.[27] The MV-GREML model fits Full $v$ and Free $w$ (and low-rank, polynomial sub-models, called RR-GREML).[20] iSet fits a similar model and can even fit Full $w$ in the case that all samples are measured in all environments.[28] MetaSex fits the Hom model independently in each of two discrete environments, which implicitly resembles Free $v$ and $w$.[29] Finally, StructLMM allows general (and potentially high-dimensional) environments, but only fits the IID model and a single SNP.[21] Complementary to GxEMM, StructLMM powerfully identifies specific SNPs with environment-dependent effects but has low resolution to determine the biological significance of specific environments.

Most recently, Ni et al.[23] developed a Free model for $v$ and $w$ that allows gene-environment (G-E) correlation. In practice, Ni et al.[23] implement simpler submodels, and particularly focus on RNM and MRNM. RNM can fit up to two environments but ignores G-E correlation and noise heterogeneity (Appendix A). MRNM can fit only one environment, where the IID and Free models coincide. These constraints resolve computational and statistical limitations common to all GREML-based approaches, including GxEMM. The constraints also resolve non-identification issues specific to the model formulation in Ni et al.[23] By contrast, the formal model underlying GxEMM is well identified and naturally accommodates multivariate environments and genetic correlations between environments (Appendix B). For completeness, we extend the GxEMM model to accommodate G-E correlation in theory (see LMMs for Polygenic Interaction Effects (GxEMM)). Nonetheless, we theoretically demonstrate that this is unlikely to cause bias for GxEMM (Appendix C); moreover, no evidence to date has shown that polygenic G-E correlation causes GxE bias or test inflation in practice, in contrast to the single-variant case.[30]

### Overview of GxEMM with Binary Traits

Continuous trait models are inappropriate for binary traits like disease status. To address this, generalized linear models assume the binary trait is driven by an underlying quantitative liability. Binary trait heritability is then naturally defined as the heritability of this liability. To our knowledge, no other method has been developed to estimate polygenic GxE for binary traits.

One prominent approach to binary trait heritability estimation is to treat the 0/1 binary label as continuous in REML and then rescale post hoc. This transformation is exact without covariates[31] and approximately extends to mild case ascertainment.[24,32,33] We attempt to extend this idea to GxE and evaluate the variance components both with and without rescaling (see GxEMM for Binary Traits).

Another method for binary trait GREML, phenotype-correlation-genotype-correlation (PCGC), uses moment-matching. When covariates are absent, PCGC amounts to comparing phenotypic and genotypic correlations, similar to HE regression for quantitative traits. PCGC extends this to incorporate covariates and preferential case ascertainment using an analytic approximation.[24,25] We can directly fit GxEMM with PCGC as it allows multiple relatedness matrices.

### LMMs for Polygenic Effects (GREML)

The standard linear mixed model (LMM) for GREML assumes a quantitative trait measured on $N$ samples, $y \in \mathbb{R}^N$. We allow $Q$ background covariates in a matrix $X \in \mathbb{R}^{N \times Q}$ and $L$ SNPs that we collect into the genotype matrix $G \in \mathbb{R}^{N \times L}$. The homogeneous LMM is:

$$y = X\alpha + G\beta + \epsilon \qquad \text{(Equation 3)}$$

$$\beta_j \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_g^2 / L\right); \; \epsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_e^2\right) \qquad \text{(Equation 4)}$$

We assume $Q \ll N$ and estimate $\alpha$ as a fixed effect. In contrast, we allow $L \gg N$ and model $\beta$ as a random effect. This can be motivated as a genuine prior that all SNPs have small, nonzero, i.i.d. effects, but GREML accurately estimates heritability under more realistic architectures.[34,35] Columns of $G$ are demeaned and scaled based on some assumed relationship between MAF and effect size.[36,37]

Marginalizing out $\beta$ gives a simpler and equivalent (or almost equivalent, according to Steinsaltz et al.[38]) formulation of GREML:

$$y \sim \mathcal{N}\left(X\alpha, h_g^2 K + \left(1 - h_g^2\right)I_N\right) \qquad \text{(Equation 5)}$$

defining $I_N$ as the $N$-dimensional identity matrix and $K := \frac{1}{L}GG^T$ as the $N \times N$ kinship matrix, a natural estimator of genetic similarity. Equation 5, and most expressions in this paper, assume the total phenotypic variance is 1, so that $\sigma_g^2 = h_g^2$ and $\sigma_e^2 = 1 - h_g^2$.

The GREML model in Equation 5 is commonly fit by restricted maximum likelihood (REML), which projects out $X$ from $K$ and $y$ and then fits variance components by maximum likelihood. However, moment-matching methods are increasingly common, especially LD score regression[39] for very large N or meta-analyses—where implementing REML becomes challenging—or HE regression[40] for small N—where REML is biased[35] and computationally unstable.[26]

## LMMs for Polygenic Interaction Effects (GxEMM)

We now assume a matrix of $P$ environmental variables, $Z \in \mathbb{R}^{N \times P}$. As for $X$, we allow arbitrary binary and/or continuous variables in $Z$ and assume $P \ll N$. The GxE mixed model (GxEMM) adds environment main effects, polygenic interactions, and environment-specific noise to GREML:

$$
y_i = \underbrace{\sum_q X_{iq}\alpha_q}_{\substack{\text{Fixed}\\\text{Effects}}} + \underbrace{\sum_s G_{is}\beta_s}_{\substack{\text{Hom.}\\\text{Genetics}}} + \underbrace{\sum_{s,p} G_{is}Z_{ip}\gamma_{sp}}_{\substack{\text{Gene}-\text{x}-\text{E}\\\text{Interaction}}} + \underbrace{\sum_p Z_{ip}\delta_{ip}}_{\substack{\text{Noise}-\text{x}-\text{E}\\\text{Interaction}}} + \underbrace{\epsilon_i}_{\substack{\text{Hom.}\\\text{Noise}}} \quad \text{(Equation 6)}
$$

$X$ should include the main effect of $Z$. We define $\gamma_{sp}$ as the interaction effect of SNP $s$ and environment $p$, and $\delta_{ip}$ is the noise contributed by environment $p$ to person $i$.

We assume $\beta$ and $\epsilon$ are random, as in Equation 4, and also model $\gamma$ and $\delta$ as random effects:

$$
\gamma_{s,} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \frac{1}{L}V\right); \quad \delta_{i,} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \frac{1}{N}W\right) \quad \text{(Equation 7)}
$$

The interactions are independent between SNPs, as in $\beta$. For $\beta$, this means entries are independent, and for $\gamma$ it means rows are independent. However, $V$ allows the interaction effects to correlate across environments. Intuitively, $\sigma_{hom}^2$ captures the homogeneous effects ($\beta$) and $V_{pp}$ captures the environment $p$-specific effects ($\gamma_{.p}$). Off-diagonal terms, $V_{pp'}$, account for genetic effects shared between environments $p$ and $p'$ in excess of the homogeneous sharing across all environments, $\sigma_{hom}^2$.

$W$ is interpreted analogously, where $W_{pp}$ indicates environment $p$-specific noise and $W_{pp'}$ indicates covariance between the noise contributed by environments $p$ and $p'$. In discrete environments, $W_{pp'}$ is not identified and the $W_{pp}$ can be assumed zero mean WLOG (Appendix B).

The model in Equation 6 can be simplified using *, the column-wise Khatri-Rao product:

$$
y = X\alpha + G\beta + (G * Z)\gamma + (I_N * Z)\delta + \epsilon \quad \text{(Equation 8)}
$$

Just as the random $\beta$ and $\epsilon$ can be marginalized for additive LMMs, giving Equation 5 from Equation 4, the random $\beta$, $\gamma$, $\delta$, and $\epsilon$ in the GxEMM model in Equation 8 can be marginalized:

$$
y \sim \mathcal{N}\left(X\alpha, \sigma_{hom}^2 K + \left(ZVZ^T\right) \circ K + \left(ZWZ^T\right) \circ I_N + \sigma_e^2 I_N\right) \quad \text{(Equation 9)}
$$

The GxE variance component nicely decomposes into the Hadamard product ($\circ$) of environmental similarity ($ZVZ^T$) and genetic similarity ($K$).

Equation 9 defines a complex likelihood. But commuting the matrix multiplications in $ZVZ^T$ and $ZWZ^T$ through the (linear) Hadamard product simplifies the expression:

$$
\begin{aligned}
y \sim \mathcal{N}\Bigg( &X\alpha, \sigma_{hom}^2 K + \sum_{p,q} V_{pq}\left[\left(Z_{.p}Z_{.q}^T\right) \circ K\right] \\
&+ \sum_{p,q} W_{pq}\left[\left(Z_{.p}Z_{.q}^T\right) \circ I_N\right] + \sigma_e^2 I_N \Bigg)
\end{aligned} \quad \text{(Equation 10)}
$$

This phenotypic covariance matrix is easily visualized for discrete environments (Figure 1, with $v_p = V_{pp}$ and $w_p = W_{pp}$). Since $K$ and $Z$ are known, Equation 10 is a standard variance component estimation problem. We typically fit the model with REML using LDAK.[37] For small sample sizes we also use HE regression, which is computationally stabler and unbiased.

Interestingly, this marginalization from Equations 8, 9, and 10 holds if and only if $\gamma$ has the covariance in Equation 7 (under large N or random $G$, Appendix D). This matters because it links the simple visual intuition (Figure 1) underlying the variance component model in Equation 10 to the concrete linear model in Equation 8; *prima facie*, distinct linear models could give equivalent variance component models. This bridge also enables GxEMM to naturally generalize to arbitrary environments, as they are seamlessly accommodated by the linear regression model in Equation 3. Conversely, the pictorial representation in Figure 1 does not easily adapt to general environments, nor do models that directly start from variance components rather than the underlying linear model.

We use Wald and LR tests for GxEMM, with asymptotic standard errors derived from the information matrix and the delta method. We allow negative variance component estimates to reduce bias—which is important for aggregating estimates across traits[38]—and because non-negative total heritability does not imply either that $\sigma_{hom}^2 \geq 0$ or that $V$ is positive definite. We test individual HE model fits using permutations. In psychENCODE, we permute within-disease to mitigate violations due to non-exchangeability across individuals by preserving any noise heterogeneity across the disease groups.[41,42]

In this paper, we do not fit GxEMM with a general $V$ or $W$ (which we call the Full model) because it has $O(P^2)$ parameters, which is computationally and statistically difficult for the range of N where basic REML methods are computationally feasible. Instead, we consider several restricted models for $V$ (and $W$): Hom, where $V = 0$; IID, where $V = h_{het}^2 I_P$; and Free, where $V = \text{diag}(v_1, \ldots, v_P)$. Both IID and Free ignore correlation between genetic effects across environments (beyond the homogeneous sharing captured in $\sigma_{hom}^2$), which will perform well when environments are approximately exchangeable rather than structured, e.g., city or hospital

indicators, but not discretized height. IID further assumes that genetics explain equal variance in each environment.

Ni et al.[23] develop two models similar to Free GxEMM, RNM and MRNM. For computational and statistical reasons, they are implemented only for special cases. RNM only fits bivariate environments and assumes Hom noise (which is unreliable, Appendix A). MRNM fits only univariate environments, where the Free and IID models coincide (Appendix B). A distinct strength of MRNM is modeling gene-environment correlation. In our model, this can be phrased as learning correlations between the genetic main and interaction effects ($\rho_p := \mathrm{Cov}(\gamma_{ps}, \beta_s)$ for each SNP $s$). In the future, we will investigate G-E correlation with GxEMM by adding terms like $\rho_p K \circ (1_N Z_p^T)$ to our model in Equation 10.

Nonetheless, G-E correlation simulations in Ni et al.[23] suggest the bias is not likely to be severe in practice for GxEMM. Specifically, GxE estimates were unbiased in simulations so long as $Z$ was adjusted as a fixed effect—which is well known, and GxEMM always includes $Z$ as a fixed effect. Ni et al.[23] do observe that adjusting for $Z$ changes the homogeneous variance components estimates, but this is entirely expected—*cf.* the distinction between "Marginal" and "Conditional" variances in Weissbrod et al.[25] Overall, generally fitting Full GxEMM with G-E correlation would be useful in large datasets, but this remains computationally and statistically unrealized in unrelated samples. Randomized methods may address this gap in the future.[43,44]

## GxEMM for Binary Traits

We extend two binary trait heritability estimation methods from the GREML context to GxEMM. The first uses REML and treats the 0/1 disease label as a quantititative trait, and then rescales the REML estimates post hoc:

$$\widehat{h}^2 := \frac{K^2(1-K)^2}{P(1-P)[\phi(\Phi(1-K))]^2}\tilde{h}^2 \qquad \text{(Equation 11)}$$

where $\tilde{h}^2 \in \{\tilde{h}_g^2, \tilde{h}_{hom}^2, \tilde{h}_{het}^2, \tilde{h}_k^2\}$ is any of the REML-based GxEMM heritability estimates; $K$ is the disease prevalence in the population and $P$ is the sample prevalence; and $\phi$ and $\Phi$ are the standard Gaussian density and distribution functions. This scaling approach is well established for the Hom model[31] and works well when covariates and ascertainment are modest,[32] but otherwise can fail badly.[24]

The second approach to fit GREML for binary traits, PCGC, directly models the binary nature of the trait and estimates parameters with the method of moments. PCGC incorporates covariates and ascertainment (approximately), making it far more robust than the REML rescaling approach for the Hom model.[24] We directly fit PCGC-based GxEMM using its LDAK implementation. We estimate standard errors using resampling[25] and the delta method.

The LDAK implementation discards on-diagonal entries in the variance component moment estimating equation. Although this is efficient for standard genetic similarity matrices, which have near-constant diagonals, this ignores important information for Free noise. A subtler and more significant problem, though, is that this off-diagonal approach treats the homogeneous noise variance component asymmetrically, as it captures both $\sigma_e^2$ and the residual moment estimating error. This caused estimates of $w$ based on including $K-1$ noise heterogeneity matrices to behave strangely, with the implicit component (the $K$-th, say) be-

ing upward biased relative to the other $K-1$ as it combines both $w_K$ and the overall PCGC approximation error. To mitigate this issue and enforce symmetry across environments, we use all $K$ environmental noise similarity matrices in PCGC, which has a unique method-of-moments solution, enforces symmetry, and provides identified estimates of $w_K$ by formally defining $\sigma_e^2 = 0$. We note this issue is relevant only for discrete environments (Appendix B).

## Simulation Details

We tested GxEMM with simulations using real genotypes from CONVERGE (see CONVERGE Data). For each simulation, we randomly choose $S = 1,000$ SNPs to have both additive and interaction effects. Because causal SNPs are chosen uniformly at random, we use the standard GRM (the Gram matrix of $G$ after column centering and scaling) inside GxEMM in these simulations. We do not use the causal relatedness matrix because it is unknown in practice. We independently draw 200 datasets per parameter set, except for the $N = 1,000$ simulations where we draw 500 datasets. We note that REML failed for a small number of simulations.

We define $Z$ by assigning each sample to one of two discrete environments uniformly at random. We define $X$ to include $Z$ and the CONVERGE fixed effects (see CONVERGE Data) and draw $\alpha_q$ i.i.d. Gaussian with mean zero and variance such that $X$ explains 10% of the variation in the raw phenotype (note, however, that $\alpha$ is irrelevant after residualizing in $X$ in HE and REML). We scale columns of $X$ and $G$ to mean zero, variance one so that the effect sizes are easily interpretable.

We vary $\sigma_{hom}^2$, $v$, and $w$ depending on the simulation setting, with different choices corresponding to either the Hom, IID, or Free GxEMM models. We always set the homogeneous noise level so that the phenotype has (residual) variance 1:

$$\sigma_e^2 := \mathbb{V}(\epsilon) = 1 - \sigma_{hom}^2 - \sum_k \left(\frac{v_k}{N}\|Z_{,k}\|^2\right) - \sum_k \left(\frac{w_k}{N}\|Z_{,k}\|^2\right)$$

Given these parameter and data choices, we draw traits from the GxEMM model in Equation 6 (Figure 2).

To simulate binary traits, we treat the generated quantitative trait, $y^{(0)}$, as a disease liability and then threshold to generate the disease label:

$$y_i = \begin{cases} 1 & \text{if} \quad y^{(0)} \geq \tau_{80} \\ 0 & \text{if} \quad y^{(0)} < \tau_{80} \end{cases}$$

where $\tau_{80}$ is the 80th percentile of $y^{(0)}$ so that $y$ represents a disease with prevalence 20%.

We also perform a specific simulation study of large-effect covariates on binary traits (Figure 3, left). Rather than choose $\alpha$ randomly, we set all its entries to 0 except for the main effect of environment 1, i.e., the entry in $\alpha$ corresponding to $Z_{,1}$. This term, $\mu_1$, controls the relative prevalence of the binary trait between environments. We also assess a variant where the disease has low population prevalence but a 50/50 case/control cohort is ascertained by preferentially measuring diseased samples (Figure 3, right), setting $\mu_1 = .4$. In these simulations, we simulate a population of size 500,000 and then ascertain 5,000 cases and controls; this strategy is computationally limited to modest population prevalences, e.g., $\approx 1\%$. We draw random SNPs i.i.d. with 50% frequency in these ascertainment simulations.

Finally, we performed simulations specifically designed to compare GxEMM to RNM (Figure S1). To mirror the simulations in Ni et al.,[23] we assume a continuous, centered, and scaled
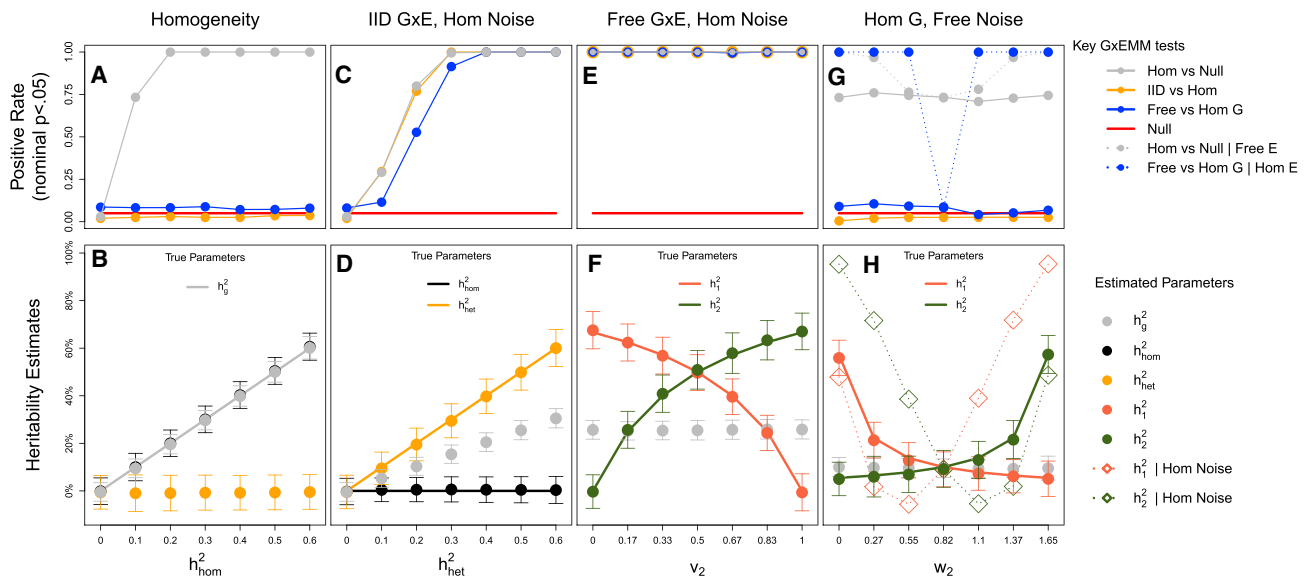
**Figure 2. GxEMM Simulations**

GxEMM simulations under (A and B) homogeneity, (C and D) IID heterogeneity, (E and F) Free genetic heterogeneity, or (G and H) Free noise heterogeneity. Results are shown for Hom, IID, and Free GxEMM and, in (G) and (H), the Free GxE model with Hom noise. In the top panels, the (true or false) positive rates are shown for Wald tests at nominal $p < .05$. In the bottom panels, mean GxEMM estimates are shown as points ($\pm 1$ SD), with true generative parameters as background lines.

univariate environment and exclude other fixed effects. We then define the liability by drawing a 35% heritable trait with variance 1 and then adding the main environmental effect. We vary the main environment effect, $\beta_z$, to evaluate a range of realistic settings. We note that large environmental main effects are plausible in practice (e.g., Peterson et al.[19]), especially because candidate environments for GxE are generally proposed based on their direct trait relevance.

## Rat Data

We studied 1,407 rats with genotype and partial phenotype information. We use the same 115 traits that we previously used,[45] which were deemed suitable for mixed model analysis in the original study.[46] We excluded one wound-healing trait where REML struggled to converge, and 24 traits with <1,000 observed samples.

The studied rats are an outbred mixture of eight inbred strains, a strategy designed to increase genetic association mapping power. The outbreeding strategy is not simple to describe, but it is designed to maintain the frequencies of the eight founder alleles at modest frequencies, avoid extreme inbreeding, and ensure mixing between founder strains.

We used the same kinship matrix, covariates, and trait transformations as in Baud et al.,[46] which we have summarized in Table S1. In particular, this involves adjusting for trait-relevant covariates—always including sex—and using trait-specific transformations.

To be conservative, we do not impute the phenotypes, though genetics-unaware imputation seems unlikely to substantially bias GxEMM.

## CONVERGE Data

We studied 9,303 samples with genotype and covariate information and pairwise kinship <.05, following Peterson et al.[19] We studied 14 binary stressors and 10 quantitative measurements as environmental covariates (Table S2). Two stress questionnaire items were very rare and excluded.

For each choice of environment variable (E), we used an intercept, age, ten genetic PCs (from an LDAK relatedness matrix[37]), and E as fixed effects. We also always include the interactions between these terms and E, which can be important for reducing bias in GxE testing in large samples.[47]

When fitting mixed models, we use the GCTA-based genetic relatedness matrix (GRM) in simulations and the LDAK-based GRM in real data. We previously found the difference between the LDAK and ordinary GRMs to be qualitatively minimal.[19]

To analyze a single quantitative covariate $z$, we linearly scaled it to have minimum 0 and maximum 1, and then defined $Z = (z|1-z) \in \mathbb{R}^{N \times 2}$. This is analogous to the discrete environment $Z$ matrix: they exactly coincide when $z$ covariate takes only two values, and in general $z$ can be interpreted as proportional membership between two stylistic groups, the min and max values of $z$. In turn, $v_1$ ($w_1$) describes genetic (non-genetic) variance specific to low $z$, $v_2$ ($w_2$) describes the genetic (non-genetic) variance specific to high $z$, and $\sigma^2_{hom}$ describes heritability that is independent of $z$. This construction of $z$ also allows naturally extending our definition of discrete environment-specific heritabilities (Equation 2) to quantitative environments: $h_1^2$ ($h_2^2$) is the heritability for samples with the minimum (maximum) observed environmental value. We feel this coding of the environmental effect makes the interaction random effects more interpretable, but we emphasize that the simpler model which evaluates only $z$ (instead of $Z$) is perfectly valid when its assumptions are met.

For higher-dimensional environments, the per-environment $Z$ matrices could be concatenated by $Z = (z_1|1-z_1|\cdots|z_P|1-z_P)$. The environment-specific heritabilities from Equation 2 are not easily extensible to this setting, however.

## psychENCODE Data

We analyzed the processed genotype and prefrontal cortex gene expression data from Gandal et al.[48] and Wang et al.[49] These data have already been adjusted for nonlinearity in raw expression

measurements and large, non-genetic covariate effects, including biological factors like age, technical factors like RIN, and latent confounders as estimated by expression Surrogate Variables.[50] We restricted only to measurements from the pre-frontal cortex, as cerebullum and temporal cortex had much lower sample sizes.

We restricted only to autosomal SNPs and genes, and we only used SNPs with MAF > 5%, missingness < 10%, and Hardy-Weinberg p > .001. We then filtered our samples such that no pair had genome-wide relatedness above 0.05.[22] After these filters, we obtained a sample size of $N = 931$ with both genotype and gene expression data.

For each autosomal gene, we extracted *cis* SNPs, which we define here as SNPs within 1 Mb of the gene transcription start or end site. This yielded 24,905 total genes. The median number of *cis*-SNPs per gene was 2,189. We then constructed *cis* kinship matrices for each gene by centering and scaling each SNP.

For each gene, we performed Hom, IID, and Free GxEMM with HE, due to the modest sample size. For comparison, we also fit two REML tools: REML-based GxEMM, and the Hom and IID models as implemented in GCTA.[27] As expected, GCTA and GxEMM with REML obtained essentially identical estimates for both Hom and IID (Figure S2). And, as expected, the HE estimates were also highly correlated ($\rho > .64$) but noisier. We note that GCTA had substantially higher average $h_g^2$ than REML-based GxEMM (5.8% versus 4.2%), but this was driven almost entirely by GCTA's lower rate of convergence: when restricting to genes where both methods converge, the methods gave similar results (5.8% versus 5.7%). This is redolent of the bias induced by restricting to genes with positive heritability estimates.[38]

We assess significance in HE by using 10,000 permutations of samples within-disease class (Table S3). Even under noise heterogeneity, this approach is an exact permutation test when genetics are completely null because samples remain exchangeable within-disease class. More generally, our permutations are in line with approximate permutation tests used widely (e.g., for FEATHER[42] or HE regression[26]).

## Results

### Quantitative Trait Simulations

We simulate data from increasingly complex polygenic interaction models to assess GxEMM and to compare the Hom, IID, and Free models. We first simulate from the purely additive Hom model, equivalent to GREML, by varying $h_{hom}^2$ and fixing the heterogeneity terms to 0, i.e., $v_k = w_k = 0$. As expected, all GxEMM models performed well, as illustrated by the roughly unbiased Hom estimates of $h_g^2$ (Figures 2A and 2B, gray points). IID and Free also gave unbiased estimates for the total heritability, and their heterogeneity tests were appropriately null (Figure 2A, orange and blue lines).

Second, we draw from the IID model by varying the single heterogeneity parameter $h_{het}^2$ and fixing $h_{hom}^2 = 0$. As expected, fitting the IID model provides roughly unbiased estimates of both $h_{het}^2$ and $h_{hom}^2$ (Figures 2C and 2D), as does Free GxEMM. However, the Hom model underestimates total heritability (gray points, Figure 2D) and also gives

the false impression that genetic factors are shared between environments.

Third, we simulate from the Free model by specifying different levels of genetic variance in each environment, varying $v_1 \in [0, 1]$ and then setting $v_2 = 1 - v_1$; we keep $\sigma_{hom}^2 = 0$ and Hom noise ($w_1 = w_2 = \sigma_e^2$). For all $v_1$, Free GxEMM accurately estimates environment-specific heritabilities (dark orange and green lines in Figures 2E and 2F), but Hom GxEMM underestimates total heritability.

Our fourth setting again draws from the Free model, except now genetics are homogeneous ($\sigma_{hom}^2 = .1$, and $v_1 = v_2 = 0$) and we instead vary the distribution of noise per environment, taking $w_1 \in [0, 1.65]$ and setting $w_2 = 1.65 - w_1$. Free GxEMM again performs well], correctly avoiding GxE false positives (Figure 2G, solid blue) and providing unbiased estimates (Figure 2H). Also, note that allowing Free noise increases Hom power (dotted gray, Figure 2G), showing that GxEMM can be useful even in the absence of GxE.

In this specific simulation, we also evaluate a variant of the Free model allowing Free GxE but assuming Hom noise. Related models have been used before.[23,51] However, this approach is susceptible to dramatic, replicating false positives under noise heterogeneity (Figure 2G, dotted blue lines) as well as severe bias: for example, consistently estimating a 10% heritable trait to have 100% heritability (dashed dark orange and green lines with diamonds, Figure 2H). We provide a theoretical characterization of this bias in Appendix A. Such GxE models that do not allow noise heterogeneity should almost never be used in practice.

We used Wald tests in these simulations, but results were similar when using LRT (Figure S3). Also, although we used two discrete environments here for simplicity, simulations with bivariate quantitative environments gave similar results (Figure S4). Finally, we also found that the GxEMM standard error estimates were accurate or slightly conservative (Figure S5).

We report the runtimes for the three GxEMM models on this simulated dataset, other simulated datasets, and the three real datasets in Table S4. Broadly, GxEMM runs in tens of minutes for thousands of samples, with higher costs for richer models and larger sample sizes.

### Binary Trait Simulations

We next examine binary traits by simulating a 20% prevalence disease with $h_{hom}^2 = .35\%$ and a discrete, binary environment with a 25%/75% split (Material and Methods), roughly based on the CONVERGE data (below). We focus on simulations from the Hom model to assess which approaches yield calibrated genetic heterogeneity tests. We compare fitting GxEMM either with REML, using the standard liability scale adjustment, or with PCGC, which directly models the binary nature of the trait.

We first assessed the impact of differential prevalence between environments by varying $\mu_1$, the mean liability in environment 1 (Figure 3, left). For $\mu_1 \approx 0$, both estimators
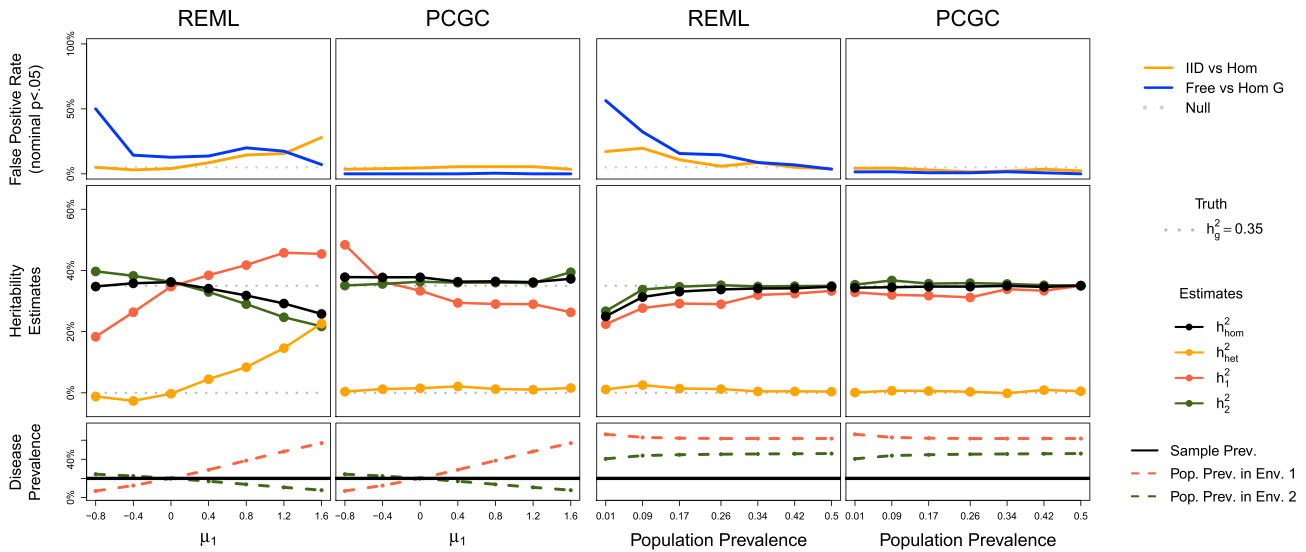
**Figure 3. REML and PCGC GxEMM in Binary Trait Simulations, with Liabilities Drawn from Hom GxEMM with $h_g^2 = 35\%$**
Left plots vary the per-environment disease prevalences through the mean liability in environment 1, $\mu_1$, while fixing $\mu_2 = 0$. Right plots vary the population disease prevalence and then ascertain a 50/50 disease cohort. The top panels test for genetic heterogeneity between environments.

(PCGC and REML) and all GxEMM models (Hom, IID, and Free) perform well. PCGC continued to perform well for all tested $\mu_1$, giving calibrated heterogeneity tests and unbiased heritability estimates for the larger group, $\widehat{h}_2^2$; however, the heritability in the smaller group, $\widehat{h}_1^2$, was conservative (or very noisy at $\mu_1 = -.8$, as basically no cases are observed in environment 1).

On the other hand, REML breaks down in several ways for $|\mu_1| > 0$. First, the $h_g^2$ estimate is downward biased, as expected.[24] Second, $h_{het}^2$ becomes upward biased, and the IID heterogeneity test becomes inflated. Third, Free GxEMM gives severely biased environment-specific heritability estimates, and often badly inflated heterogeneity tests.

To assess power, we repeated these simulations under Free genetic heterogeneity and found that IID GxEMM had power to detect genetic heterogeneity for both REML and PCGC (Figure S6, left). However, Free GxEMM had low power, reflecting the loss of information from the liability thresholding process. Second, under Hom genetics and Free noise, PCGC was unbiased or slightly conservative while REML was again upwardly biased (Figure S6, right). Third, we performed binary trait simulations mirroring the quantitative trait simulations in LMMs for Polygenic Effects (GREML) to broadly test PCGC-based GxEMM (Figure S7). PCGC performed similar to REML for quantitative traits, except power is lower, as expected, and PCGC performed poorly under extreme variance heterogeneity, likely because the underlying first-order approximation breaks down in extreme settings.

Disease studies often preferentially ascertain disease cases to increase power. As ascertainment causes bias in GREML,[24] we assessed its impact on GxEMM through simulations. We fixed $\mu_1 = .4$ (similar to CONVERGE)

and ascertained 50/50 case/control cohorts (Material and Methods). Although GxEMM with REML breaks down under ascertainment—in particular, Free/IID GxEMM tests have roughly 50%/25% false positive rate for 1% prevalence—PCGC remains calibrated (Figure 3, right).

Finally, we evaluated the RNM model, a recent REML-based polygenic GxE model implemented in the MTG2 software package.[52] To streamline comparison, we performed simplified versions of the above non-ascertained simulations with a continuous, univariate environment and no additional covariates. Consistent with our above results (Figure 3, left), we found that PCGC broadly performed well, as did REML-based GxEMM and RNM when the environment had no main effect (Figure S1). However, both RNM- and REML-based GxEMM become biased as the environment main effect grows. In particular, RNM obtains roughly 90% false positive rates when the environment explains roughly 50% of the liability-scale phenotypic variation.

In summary, repurposing standard tools for inferring polygenic GxE for quantitative traits is not viable. On the other hand, PCGC-based inference provides calibrated heterogeneity tests and approximately unbiased GxEMM parameter estimates. We caution, though, that PCGC estimator variances are large and, in particular, the Free model adds negligible value beyond the IID model for binary traits at these sample sizes.

## Phenome-wide Sex-Specific Genetics in Outbred Rats
Our first application is to sex-specific genetic effects across 115 phenotypes in 1,407 outbred rats from the Rat Genome Sequencing and Mapping Consortium[46] (Material and Methods). These samples have high genetic relatedness, which aids mixed model power. Although many traits are known to be sexually dimorphic, it is not
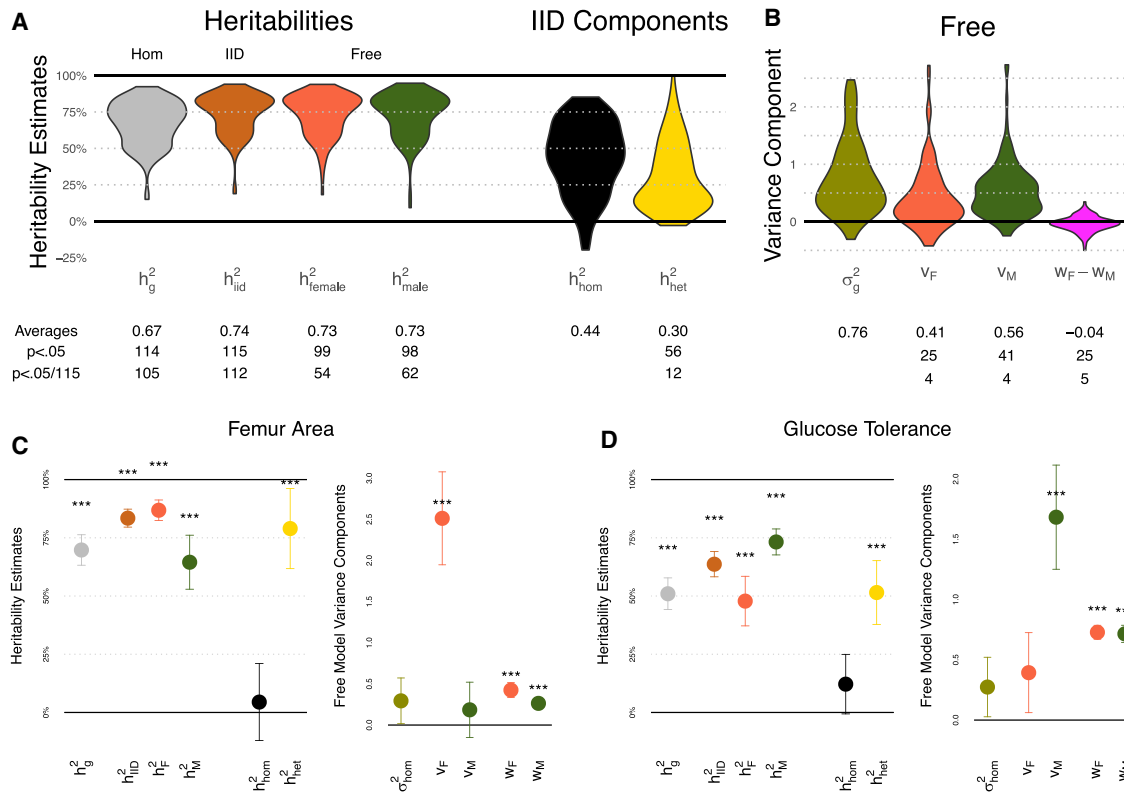
**Figure 4. GxEMM Estimates Aggregated across p = 115 Traits in Outbred Rats for the Hom, IID, and Free Models**

p values marginally test whether variance components are 0.

(A, left) Heritability estimates from each model, with labels for which violins correspond to which GxEMM model.

(A, right) Homogeneous and heterogeneous heritability components in the IID model.

(B) Genetic variance component estimates for the Free model, as well as noise heterogeneity estimates ($w_F - w_M$). Middle table: for each component of the violin plots, we provide the average across traits as well as the number of traits where the component is significantly different from zero.

(C and D) Heritability and variance component estimates from GxEMM for two specific traits, bone density and glucose tolerance. Error bars represent 1 SE; *p < .05, **p < .01, and **p < .001.

generally well known to what extent, and for which traits, sex differences are driven by autosomal genetic variation.

First, we fit Hom GxEMM and found 105/115 traits were heritable at p = .05/115, with average $\widehat{h}^2_g = 67.3\%$ (gray violin, Figure 4A). We next fit IID GxEMM, testing for GxSex interaction and found that the number of heritable traits increased to 112/115 (rust violin). The average IID heritability was 73.8%, which is 9.7% larger than the GREML heritability. The sex-specific component of the IID model explained 30.3% of trait variance on average and was Bonferroni significant for 12 traits (gold violin), including measurements of bone density, hemoglobin, platelets, serum composition, and glucose trial response (Figure S8, Table S1).

On average across all traits, the Free heritability estimates were similar, for both sexes, to the IID model ($h^2_F$ orange violin and $h^2_M$ green violin, Figure 4A). Nonetheless, the Free model uncovered many traits with sex-specific genetic variance, including eight that were Bonferroni significant (orange $v_F$ and green $v_M$ violins, Figure 4B). All eight had significant $h^2_{het}$ in the IID analysis, but the Free analysis more precisely characterizes the genetic heterogeneity. For example, 3/8 of these traits have significantly different genetic effect sizes between sexes (Wald p < .05/115; 7/8 have p < .05): female-specific genetic effects drove two bone measurements, white blood cell count, and serum chloride; male-specific genetic effects drove two platelet traits, glucose response, and serum potassium (Figure S8). The Free model also found sex-specific noise levels for 5/115 traits (p < .05/115, pink violin), including three of the genetically heterogeneous traits and an additional bone density trait.

To further explore the data and illustrate the GxEMM model, we plot results for two of the traits with sex-specific heritability. The first is a femur size measurement (distal femur cortical area, Figure 4C). In Hom GxEMM, this trait has high heritability, roughly 75%. Nonetheless, heritability is increased to roughly 85% by the IID model and, moreover, the homogeneous component of IID GxEMM ($h^2_{hom}$) is near zero. Finally, the Free model goes further by revealing that the sex-specific genetic effects are primarily active in females. This analysis shows that GxEMM can uncover dramatically non-additive genetic architecture even when ordinary GREML explains substantial heritability.
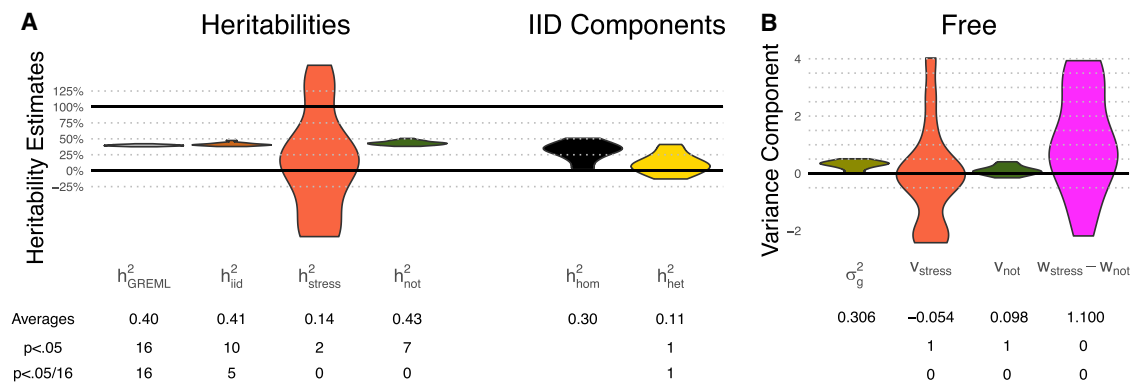
**Figure 5. GxEMM Stress-Specific Heritability Estimates for Major Depression**
p values marginally test whether variance components are 0.
Estimates are shown aggregated over 14 different choices for the binary stress measurement that is used as the environment in GxEMM.
All stress measures are correlated and have less than 50% prevalence in the dataset.
(A, left) Heritabilities from the Hom, IID, and Free models.
(A, right) IID heritability components.
(B) Free variance components.

| | $h^2_{GREML}$ | $h^2_{iid}$ | $h^2_{stress}$ | $h^2_{not}$ | $h^2_{hom}$ | $h^2_{het}$ | $\sigma^2_g$ | $v_{stress}$ | $v_{not}$ | $w_{stress} - w_{not}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Averages | 0.40 | 0.41 | 0.14 | 0.43 | 0.30 | 0.11 | 0.306 | −0.054 | 0.098 | 1.100 |
| p<.05 | 16 | 10 | 2 | 7 | | 1 | | 1 | 1 | 0 |
| p<.05/16 | 16 | 5 | 0 | 0 | | 1 | | 0 | 0 | 0 |

An approximate mirror image is obtained for a glucose tolerance trait (area under glycemia curve over baseline during intraperitoneal glucose tolerance test, Figure 4D): $h^2_g \approx 50\%$ is large, yet homogeneous heritability largely vanishes after accounting for sex-specific heritability in IID GxEMM, which is in turn revealed by the Full model to be driven largely by male-specific genetic effects. This result is consistent with a previously reported sex-APOE interaction effect for glucose.[53] This rich characterization of sex-specific architecture has clear implications for genetic association studies: power can be maximized by studying the sex with greater heritability, and associations can be interpreted in light of known average difference between sexes.

**Polygenic Stress Interaction in Major Depression**
We next apply GxEMM to major depression (MD), a moderately heritable disease that is likely genetically and environmentally heterogeneous.[54] We analyzed the CONVERGE cohort, which recruited about 10,000 Han Chinese women between 30 and 60 years old. Women were chosen because they have higher MD heritability than men (i.e., 42% versus 29%[55]), and clinically ascertained MD-affected case subjects were selected for the same reason.[56] This strategy successfully led to replicated GWAS hits for MD[54] and later yielded three SNP effects specific to people without major lifetime stress.[19] Here, we extend these SNP heterogeneity analyses to the polygenic level with GxEMM using our robust approach for binary traits based on PCGC.

We first fit Hom GxEMM with PCGC to MD and found $\widehat{h}^2_g = 38.5\%$ (SE 4.5%) assuming an 8.8% population prevalence (Figure 5).[19] Importantly, we adjusted for interactions between genetic PCs and the environment, which can be essential for avoiding bias and false positives.[47]

We then fit 14 IID GxEMM models with each of 14 different binary stress measures (CONVERGE Data). IID modestly increased heritability on average ($\bar{h}^2_{IID} = 41.2\%$, Table S2), broadly supporting polygenic stress interactions. One stress measure (divorced/separated/widowed status) was Bonferroni significant (p = 0.0020 < .05/14). This measure had relatively high prevalence and MD effect severity, which both increase interaction test power; more severe measures (e.g., "Child Abuse") were less frequent, and more frequent measures (e.g., "Natural Disaster") were less severe. Altogether, the IID model supports polygenic stress interaction for MD, but large standard errors and high correlations between measures prevent conclusions about which specific major life stresses drive the interactions.

The Free model performed reasonably for the non-stress groups, finding slightly higher heritability relative to IID (41.2% versus 43.4%) but not finding any nonstress-specific genetic variance at Bonferroni significance. More importantly, however, the Free model performed badly in the smaller stress groups, consistent with simulations (Figure S6, left). Similarly, differential noise estimates had large standard errors and none were significantly different from 0.

Although we primarily focus on stress interactions due to prior knowledge that they are relevant to MD, we also investigated polygenic interaction across each of ten quantitative environmental measures (Material and Methods). IID results were qualitatively similar to the results from the binary stress measures, with average heritability increasing to $\bar{h}^2_{IID} = 42.9\%$ (Figure S9) and one measure, "Cold Mother," that was nearly Bonferroni significant for polygenic interaction (p = 0.0079 > .05/10; "Cold Father" point estimates were similar and had p = .0198). As for the binary stress measures, the Free model was not informative. Overall, IID GxEMM had power in this dataset and

**Table 2. Transcriptome-wide Average GxEMM Heritability Estimates under the Hom ($h_g^2$), IID ($h_{IID}^2$), and Free Models ($h_{BPD}^2$, $h_{SCZ}^2$, $h_{CTRL}^2$)**

| | $h_g^2$ | $h_{IID}^2$ | $h_{BPD}^2$ | $h_{SCZ}^2$ | $h_{CTRL}^2$ |
|---|---|---|---|---|---|
| Mean (%) | 4.06 | 4.12 | 4.32 | 4.34 | 3.82 |
| SE (%) | 0.07 | 0.07 | 0.08 | 0.08 | 0.07 |
| Wald $p$ for $h^2 > 0$ | $<2 \times 10^{-16}$ | $<2 \times 10^{-16}$ | $<2 \times 10^{-16}$ | $<2 \times 10^{-16}$ | $<2 \times 10^{-16}$ |
| Wald $p$ for $h^2 > h_g^2$ | – | $1.15 \times 10^{-11}$ | $1.93 \times 10^{-9}$ | $<2 \times 10^{-16}$ | 1.00 |
| Wald $p$ for $h^2 > h_{IID}^2$ | – | – | $3.83 \times 10^{-6}$ | $<2 \times 10^{-16}$ | 1.00 |

Because $N$ is modest, we fit GxEMM with HE regression. We compare the models using Wald tests, estimating the standard errors of the transcriptome-wide averages assuming independence across genes. Dash (–) indicates that the comparison is not meaningful.

provided evidence for genetic heterogeneity, but larger sample sizes seem necessary for Free GxEMM to be useful.

### Brain Expression Heterogeneity in Psychiatric Disease

Finally, we apply GxEMM to brain gene expression data from psychENCODE[48,49] to test for differential genetic and nongenetic factors governing transcriptional regulation in psychiatric disease relative to control subjects. In this analysis, disease state plays the role of the "environment" in GxEMM. After quality control (Material and Methods), the data include 931 samples, including 356 schizophrenia-affected case subjects (SCZ), 158 bipolar disorder (BPD)-affected case subjects, and 417 control subjects (CTRL), each measured on 24,905 genes. In this analysis, we focus on fitting GxEMM with HE regression because REML is computationally and statistically unstable at smaller sample sizes, especially for richer models.[26] (We note, though, that REML gave broadly consistent results when it converged [Figure S2].) For each gene, we estimate local *cis*-heritability by using relatedness matrices built specifically from SNPs within 1 Mb of the gene's transcription start site.

We first fit Hom GxEMM and found $h_g^2 = 4.06\%$ (SE 0.07%) on average across the genome (Table 2). This is in line with previous transcriptome-wide heritability estimates from whole blood.[26,57] Here and below, we test transcriptome-wide averages with a Wald test, where standard errors are calculated assuming each gene is independent, which is an established approximation (e.g., Hernandez et al.[26]).

To investigate transcriptome-wide patterns of disease-dependent genetic architecture, we next fit the IID model. We estimated transcriptome-wide average heritability of $h_{IID}^2 = 4.12\%$, which slightly (1.5%) but statistically significantly (p $< 2 \times 10^{-16}$, paired t test) increased heritability over the homogeneous GREML model. We then fit the richer Free model, which estimated disease-specific heritabilities of 4.32% (SE 0.08%), 4.34% (SE 0.08%), and 3.82% (SE of 0.07%) for BPD, SCZ, and CTRL, respectively. For BPD and SCZ, the increase over $h_g^2$ is $\approx 7\%$ and significant (both p $< 2 \times 10^{-16}$); the increases over $h_{IID}^2$ are also statistically significant (p $= 3.83 \times 10^{-6}$ and p $< 2 \times 10^{-16}$). Conversely, $h_{CTRL}^2$ is lower than $h_g^2$, suggesting that GREML actually overestimates the heritability that is common across all disease states, consistent with simulations (Fig-

ures 2D and 2F). Overall, GxEMM shows statistically significant interaction between psychiatric disease state and genetic control of gene expression in the prefrontal cortex, and the Free model further suggests an increased role of genetics specifically active in the disease states.

To confirm that our results were not primarily driven by confounding from population structure, we performed the same analysis after restricting our dataset to samples from the "European" population, which was the most numerous population in our psychENCODE data (N = 872). We found that transcriptome-wide average estimates were essentially unchanged (Table S5).

We next tested for genetic heterogeneity at individual genes, which is less powerful but more precise than transcriptome-wide average tests. Based on simulations, we have low power for genetic heterogeneity tests across all genes at this sample size (Figure S10). Instead, we test 63 genes with known significant gene-based genetic associations for SCZ.[58] We test for significance using within-disease group permutations (Material and Methods).

IID GxEMM does not find any gene with significant polygenic-disease interaction at 25% FDR (Benjamini-Hochberg, Figure 6A, Table S3). However, Free GxEMM detects 9 associations across the three subgroup-specific genetic variances (3 are Bonferroni significant). These results are consistent with our transcriptome-wide average tests, where we found that the Free model had considerably higher power than IID.

Figures 6B and 6C show the GxEMM estimates for two of these nine genes. First, *SNX19* is an SCZ-associated protein coding gene that also has known experimentally validated functional eQTL that overlap SCZ risk loci.[59–63] This gene has high *cis*-heritability in the Hom model (roughly 75%). Next, the well-established IID model fails to find even suggestive genetic heterogeneity. However, the rich Free model finds substantial heterogeneity that is largely driven by SCZ-specific genetics. This is consistent with simulation results showing low power for the IID GxE test when most of the specific heritability is concentrated in one group.

The second significant gene we illustrate, *FURIN*, is a protease that has been robustly associated with neurodevelopment in zebrafish and human neural progenitor cell development *in vitro*.[64] *FURIN* also has a strong *cis*-eQTL
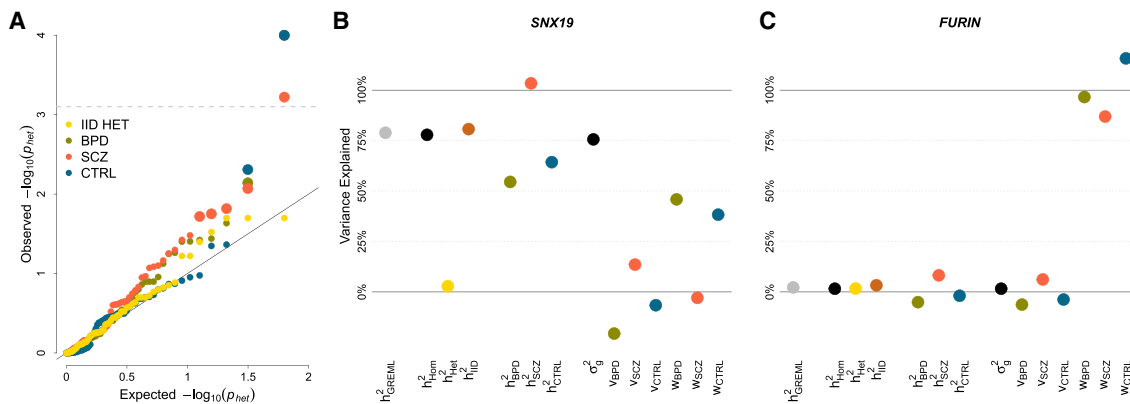
**Figure 6. Psychiatric Disease-Specific Heritability in Prefrontal Cortex Gene Expression**
(A) IID and Free genetic heterogeneity tests based on permutations across 63 known SCZ-associated genes. Large points have FDR < 25%, and the horizontal line is Bonferroni-adjusted p < .05. The maximum BPD and CTRL points overlap at p = 1/(1+10,000), the minimum possible value after 10,000 permutations.
(B and C) GxEMM results for 2/9 candidate SCZ genes with significant genetic heterogeneity, *SNX19* (B) and *FURIN* (C).

that colocalizes with GWAS-significant SCZ SNPs,[65] and this eQTL was experimentally shown to modify *FURIN* expression and, in turn, brain-derived neurotrophic factor.[66] In psychENCODE, we found that *FURIN* has essentially zero homogeneous or IID heritability, in contrast with *SNX19*. On the other hand, the Free model is able to reveal significant heritability specifically in SCZ. These stories both support the links between genotype, expression, and disease and demonstrate how GxEMM can be used to uncover genetic variants with disease-specific functional genomic effects. We also found Bonferroni-significant genetic heterogeneity for *CHRNA2*, a previously identified "high confidence" gene supported by association with SCZ and Hi-C interaction with nearby eQTL.[49]

Finally, we tested for differential expression variance between disease groups transcriptome-wide. We find evidence for pervasive differences in expression variance across disease states: for example, 2,459 genes (roughly 10%) have Levene p < .001. This heterogeneity was driven primarily by higher expression variance in SCZ: of these 2,459 genes, 97.1% had higher variance in SCZ than CTRL. We also note that 92.2% of genes had higher variance in SCZ than BPD, but this is at least partially because SCZ has higher sample size (and thus power). Adding confidence in the Levene tests, we permuted disease labels and found no evidence of test inflation (e.g., 10% of genes have Levene p <.099 after permuting), suggesting disease state truly correlates with expression variance. Although this test does not establish genetic heterogeneity, it does show that Free GxEMM is likely more powerful than IID: in simulations, variance heterogeneity increased Free power over IID regardless whether it was genetic (Figure 2E) or non-genetic (Figure 2G).

## Discussion

Gene-environment interactions (GxE) and polygenicity are separately well documented. We help bring these concepts together with a linear mixed model for polygenic GxE that we call GxEMM. GxEMM consists of three key models: Hom, IID, and Free. Each pairwise comparison adds biological interpretation: comparing Hom to IID can demonstrate the existence of GxE, and comparing IID to Free can demonstrate and appropriately adapt to heteroscedasticity across environments. Finally, comparing Hom and Free models can add power for detecting GxE over the IID model; for example, the Free model, but not the IID model, had power to show GxE for *SNX19* and *FURIN* in psychENCODE. Generally, GxEMM can be used for any covariate putatively interacting with the genome, including genetic variants,[67] study indicators for meta-analysis[68,69] or genotype quality control,[70] or phenotypic subtypes.[71]

We have introduced several methods to fit GxEMM, based on REML, HE, and PCGC; all are implemented as an R wrapper of the LDAK software package. In practice, only one method is recommended for a given dataset: HE for small sample sizes, PCGC for binary traits, and REML for large sample sizes and continuous traits. These same concepts apply to standard homogeneous heritability estimates.

GxEMM unifies, formalizes, and solves several important biases in recent approaches to estimate GxE-based heritability. In particular, only GxEMM can accommodate general environments, noise heterogeneity, modest sample size, and binary traits, which we support with theory and simulation.

There are several methodological limitations to GxEMM. First, like most LMM methods, GxEMM is computationally intensive, as it fits several variance components to individual-level data. In special cases, the problem can be rewritten to roughly reduce the computational complexity from $O(N^3)$ to $O(N)$.[21,28] In the future, we will investigate larger sample sizes using randomized method of moments to fit GxEMM[43,44] (Material and Methods).

A second limitation is our assumption of Gaussian random effects—although the central limit theorem does support

this for polygenic effects. But non-normality is not obviously more salient for GxEMM than for ordinary GREML, and we used quantile normalization in psychENCODE and trait-specific transformations for the rat traits to mitigate the impact of outliers.[26,42,72] We note that these normalizations, too, have downsides, and in particular reduce power for Levene tests for variance heterogeneity.

A third limitation to GxEMM is that we do not test or correct for gene-environment (G-E) correlation. This is a well-known source of bias in the fixed effect case.[30] Surprisingly, however, Ni et al.[23] did not observe inflation for polygenic interaction tests in simulations. In fact, we prove that this is due to a fundamental difference between the polygenic and single-SNP case (Proposition 3): absent systemic coordination between the main genetic effects on the environment and trait, the per-SNP biases average out over the genome. While such coordination would be intrinsically interesting, it has not been documented in real data and is not likely to be large in practice, and thus polygenic G-E is likely not to cause false positives for polygenic GxE tests. Nonetheless, we describe how to incorporate G-E interaction within the GxEMM model framework in the Material and Methods, for completeness.

Fourth, we did not fit Full GxEMM, which can learn structured genetic correlations across environments. Such models are flexible but complex, with many degrees of freedom and non-trivial identification concerns. However, at biobank-scale sample sizes, there will likely be sufficient power to fit the Full model.

A fifth limitation is that GxEMM does not currently allow random effects beyond the core Hom and GxE variance components. However, in the context of family studies (e.g., Diego et al.[73]) or high-dimensional covariates (e.g., Moore et al.[21]), it can be valuable to model background confounding variables as random effects in GxE tests. In future work, GxEMM could in principle be easily extended to accommodate such additional variance components, e.g., Zaitlen et al.[74] jointly fit pedigree- and additive-relatedness matrices in the homogeneous setting. Nonetheless, in practice, we use and recommend standard filters on kinship and population structure corrections when analyzing natural populations, as is standard in GREML analyses.[22] In the case of the rat experimental data, we did not perform these steps because confounding by cryptic relatedness is very unlikely;[45,46] however, additional random effects may still be helpful, e.g., for cage effects.

The results from our rat analysis shows widespread sex-specific heritability. The Free model, specifically, has strong statistical support for several traits even at stringent thresholds. This power derived from the high levels of genetic relatedness in these model organisms. Likewise, GxEMM will be powerful in other model organism and family studies, which can help build broad intuition about causal genetic architecture that can inform human disease studies. This is particularly true for autosome-sex interactions, which may be important for many complex human diseases.[75]

There are several limitations to our analysis of the psychENCODE data. First, our LDAK-based REML approach often failed to converge, so we used HE regression, which is robust but less powerful than REML. Second, cell type proportions are known to vary between people and disease states, causing bulk differential expression across disease groups for any gene that is simply differentially expressed across cell types. In future work, we will extend GxEMM to partition expression heritability into cell type proportion effects and within-cell type effects.

GxEMM may also be useful for GWAS applications. An IID-based approach has been shown to improve power and calibration over the standard Hom model in some cases,[47] suggesting that Free GxEMM will likely yield further improvements when the IID assumption fails.

## Appendix A: Unmodelled Heteroscedasticity Causes Bias in Polygenic Interaction Estimates

In this appendix we quantify the bias under the Free GxEMM model when environmental heterogeneity is ignored. We assume discrete environments and that the modeler fits the similarity matrices $\{K_k, I\}$, where $K_k$ is the kinship matrix $K$ with entries outside of environment $k$ zeroed out. We assume that $\text{tr}(K_k) = n_k$ for all $k$. We also assume that the true variance of $y$ is:

$$\mathbb{V}(y) = \sum_k w_k E_k + \sigma^2 I \qquad \text{(Equation A1)}$$

where $E_k$ is the identity matrix with entries outside environment $k$ zeroed out—i.e., we assume there is no genetic variance but that the noise variance may differ across environments when $w_k \neq w_j$ for some $j$ and $k$. For simplicity, we also ignore the homogeneous genetic component and covariates. For identifiability, we assume $\sum_k n_k w_k = 0$; under this constraint, $\sigma^2$ is the average phenotypic variance across all samples. Finally, we assume approximate independence between genetics and environment. It is worth noting how simple this model is—pure heteroscedastic noise.

We analytically study the behavior of the HE regression, which is more tractable than the MLE.

**Proposition 1.** *Assume for all $k$ that $n_k \approx n_0 \gg 1$ and that $a \approx \mathbb{V}(diag(K_k))$ and $b \approx \mathbb{V}(off-diag(K_k))$ for all $k$. Under the pure heteroscedastic noise model (Equation A1), the bias in the Free genetic heterogeneity estimate with Hom noise is:*

$$\mathbb{E}(\widehat{v}_k) \approx \frac{1}{1 + a + n_0 b} w_k \qquad \text{(Equation A2)}$$

*Proof.* The moment estimating problem is:

$$\min_{v, \sigma^2} \left\| yy^T - \sum_k v_k K_k - \sigma^2 I \right\|_F$$
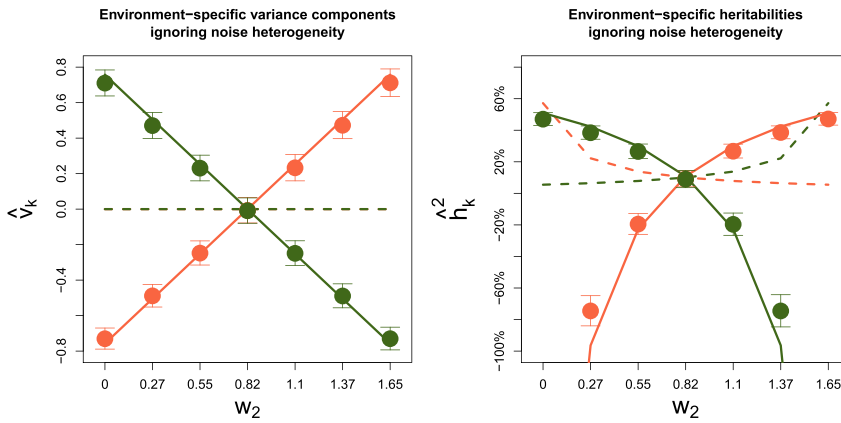
**Figure 7. Bias from HE Regression when Fitting Free GxEMM with Hom Noise**
Estimates are based on the simulations used in Figures 2G and 2H, which assume Hom genetics and Free noise. Lines provide the bias approximation derived above. The true simulation model has $v_1 = v_2 = 0$; the true values are shown as dotted lines in the background.

Note that modeling the diagonal terms, rather than regressing only on the strict lower triangular entries, is important when variance may be heterogeneous across samples.

The OLS solution for the parameters $v$ and $\sigma^2$ is given by:

$$\begin{pmatrix} D^{-1} & n \\ n^T & N \end{pmatrix} \begin{pmatrix} \widehat{v} \\ \widehat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} S \\ y^T y \end{pmatrix}$$

where $n$ is the vector of environment-specific sample sizes, $S_k := y^T K_k y$, and $D$ is diagonal (because $\langle K_k, K_j \rangle = 0$ for $k \neq j$) with $D_{kk} := d_k := (1/\mathrm{tr}(K_k^2))$. In expectation:

$$\mathbb{E}(y^T y) = \mathrm{tr}(\mathbb{V}(y)) = \sum_k n_k w_k + \sigma^2 N = \sigma^2 N$$

The last equality derives from our identification assumption on $(w, \sigma^2)$. Likewise, the expected values of the quadratic forms in $S$ are:

$$\begin{aligned} \mathbb{E}(S_k) &= \mathrm{tr}(K_k \mathbb{V}(y)) = \mathrm{tr}\left(K_k\left(\sum_k n_k w_k I_k + \sigma^2 I\right)\right) \\ &= \sum_j w_j \mathrm{tr}(K_k I_j) + \sigma^2 \mathrm{tr}(K_k) \\ &= w_k \mathrm{tr}(K_k) + \sigma^2 \mathrm{tr}(K_k) = n_k(w_k + \sigma^2) \end{aligned}$$

Let $s := \dfrac{1}{N - n^T D n} = \dfrac{1}{N - \sum_j n_j^2 d_j}$. The expected variance component estimates are (using block inversion in the second equality):

$$\mathbb{E}\left(\begin{pmatrix} \widehat{v} \\ \widehat{\sigma}^2 \end{pmatrix}\right) = \begin{pmatrix} D^{-1} & n \\ n^T & N \end{pmatrix}^{-1} \begin{pmatrix} n \circ (w + \sigma^2 1_N) \\ \sigma^2 N \end{pmatrix}$$

$$= \begin{pmatrix} D + sDnn^T D & -sDn \\ \cdot & s \end{pmatrix} \begin{pmatrix} n \circ (w + \sigma^2 1_N) \\ \sigma^2 N \end{pmatrix} \Rightarrow$$

$$\mathbb{E}(\widehat{v}_k) = d_k n_k(w_k + \sigma^2) + s d_k n_k \sum_j n_j^2 (w_j + \sigma^2) d_j - s d_k n_k \sigma^2 N$$

$$= d_k n_k \left[w_k - \frac{\sum_j n_j^2 d_j w_j}{\sum_j n_j^2 d_j - N}\right]$$

The last line follows from some algebra and nicely emphasizes that the $\widehat{v}$ are unbiased when $w_k = 0$ for all $k$, i.e., under noise variance homogeneity.

The $d_k$ terms can be simplified assuming that the environments are independent of the genotypes and that each $n_k \gg 1$, so that:

$$\begin{aligned} \mathrm{tr}(K_k^2) &= \sum_{i,j \in k} K_{ij}^2 = \sum_{i \in k} K_{ij}^2 + \sum_{i \neq j \in k} K_{ij}^2 \\ &= n_k + n_k \mathbb{V}(K_{ii \in k}) + n_k(n_k - 1)\mathbb{V}(K_{i < j \in k}) \\ &\approx n_k(1 + a + n_k b) \end{aligned}$$

The approximation in the last line uses the assumption that the same $a$ and $b$ values can be used across $k$, which is implied by our assumed gene-environment independence and large $n_k$.

To approximate $d_k$, we assume that $n_k = n_0$ is constant across all groups, giving:

$$\begin{aligned} \mathbb{E}(\widehat{v}_k) &\approx \frac{1}{1 + a + n_0 b}\left[w_k - \frac{\frac{1}{1 + a + n_0 b}\sum_j n_j w_j}{\frac{1}{1 + a + n_0 b}\sum_j n_j - N}\right] \\ &= \frac{1}{1 + a + n_0 b} w_k \end{aligned}$$

The last equality derives from our identification assumption on $(w, \sigma^2)$. This approximation could be expanded to capture modest variability across $n_j$ by Taylor approximation around $n_0$.

□

In the case of noise homogeneity, $w_k = 0$ for all $k$, hence the HE estimate is unbiased, as is well known. The bias is also small in the case where the kinship matrix is very far from the identity (either because off-diagonals are large, increasing $b$, or because on-diagonals are highly variable, increasing $a$). However, in all other settings the expected estimate for environment-$k$-specific heritability $v_k$ will be proportional to the true environment-$k$-specific noise ($w_k$), attenuated by a factor measuring the distance between $K$ and $I$.

We show that this bias approximation can be accurate in practice by evaluating the HE regression model fits from the quantitative simulations performed in the main text.

Specifically, we fit HE regression ignoring noise heterogeneity to the simulations with homogeneous genetics but heterogeneous noise (Figures 2G and 2H) and compare the estimated genetic heterogeneity with our bias approximation (Figure 7). We find that our bias approximation is accurate even though these simulations include nonzero heritability and fit a homogeneous genetic variance component, unlike our simplified theoretical analysis in this appendix—supporting the practical relevance of our bias approximation.

$W$ is also not fully identified for discrete $Z$. First, off-diagonal entries cannot be estimated, so the Full and Free models are identical. Second, $(\sigma_e^2, W) \equiv (\sigma_e^2 + \lambda, W - \lambda I)$, which eliminates a degree of freedom from the diagonal of $W$. In particular, the IID and Hom noise models coincide.

### Covariate Scale Identification

Columns of $G$ can be assumed mean zero WLOG because the restricted likelihood is invariant under the mapping $G \mapsto G + 1_N \mu^T$ for all $\mu \in \mathbb{R}^L$:

$$
\begin{aligned}
G\beta + (G * Z)\gamma \mapsto G\beta + (1_N \mu^T)\beta + (G * Z)\gamma + \left([1_N \mu^T] * Z\right)\gamma \quad &\text{(linearity of Khatri} - \text{Rao)}\\
= G\beta + 1_N(\mu^T\beta) + (G * Z)\gamma + Z(\text{mat}(\gamma)\mu) \quad &\text{(direct computation)}\\
\equiv G\beta + (G * Z)\gamma
\end{aligned}
$$

## Appendix B: Identification of $V/W$ and Standardizing $G/Z$

### Parameter Identification

The Free and IID assumptions resolve an identification problem that arises for the genetic variance parameters of the Full model, $(\sigma_{hom}^2, V)$, when $Z$ is discrete. Under the Full model, any constant $\lambda$ can be passed between $\sigma_{hom}^2$

where the equivalence is modulo the projection orthogonal to the fixed effects, which include $1_N$ and $Z$ (mat($\cdot$) stacks $\gamma$ into an $K \times L$ matrix column-wise).

$Z$, however, cannot be centered without changing the likelihood, because $G$ is not projected out as a fixed effect. Mapping $Z \mapsto Z + 1_N \mu^T$, now for $\mu \in \mathbb{R}^K$, the initial covariance parameters $(V, \sigma_z^2)$ are equivalent to the the new parameters $(\Sigma_v, \sigma_v^2)$ iff

$$
\begin{aligned}
\sigma_v^2 K + K \circ \left([Z + 1_N \mu^T]\Sigma_v[Z + 1_N \mu^T]^T\right) = \sigma_g^2 K + K \circ (ZVZ^T) &\Leftrightarrow\\
K \circ \left(\left(\sigma_v^2 - \sigma_g^2\right)J_N + 1_N \mu^T \Sigma_v Z^T + Z\Sigma_v \mu 1_N^T - Z(V - \Sigma_v)Z^T\right) = 0 &\Leftrightarrow\\
\left(\sigma_v^2 + \mu^T \Sigma_v \mu - \sigma_g^2\right)1_N 1_N^T + 1_N \mu^T \Sigma_v Z^T + Z\Sigma_v \mu 1_N^T - Z(V - \Sigma_v)Z^T = 0 &\Leftrightarrow\\
\left(\sigma_v^2 + \mu^T \Sigma_v \mu - \sigma_g^2\right)Z 1_K 1_K^T Z^T + Z 1_K \mu^T \Sigma_v Z^T + Z\Sigma_v \mu 1_K^T Z^T - Z(V - \Sigma_v)Z^T = 0 &\Leftrightarrow\\
Z\left(\left(\sigma_v^2 + \mu^T \Sigma_v \mu - \sigma_g^2\right)J_K - (V - \Sigma_v) + 1_K \mu^T \Sigma_v + \Sigma_v \mu 1_K^T\right)Z^T = 0 &\Leftrightarrow\\
\left(\sigma_v^2 + \mu^T \Sigma_v \mu - \sigma_g^2\right)J_K - (V - \Sigma_v) + 1_K \mu^T \Sigma_v + \Sigma_v \mu 1_K^T = 0 &\Leftrightarrow\\
\Sigma_v + \left(\sigma_v^2 + \mu^T \Sigma_v \mu - \sigma_g^2\right)J_K + 1_K(\Sigma_v \mu)^T + (\Sigma_v \mu)1_K^T = V
\end{aligned}
$$

and $V$ without changing the likelihood by $(\sigma_g^2, V) \mapsto (\sigma_g^2 - \lambda, V + \lambda 1_P 1_P^T)$, with $1_P$ a vector of 1 s. Conceptually, this is equivalent to the fact that the population mean and the environmental main effects are not jointly identified in linear regression models with discrete environments.

Free and IID GxEMM break this symmetry because $V_{ij}$ and $V_{ij} + \lambda$ cannot both be 0 for nonzero $\lambda$. Intuitively, the Free and IID models prioritize $(\beta, \gamma)$ pairs where the mean effect across environments ($\overline{\gamma}_s$) is absorbed into $\beta_s$ (so $\overline{\gamma}_s = 0$). We also note that some parameters in the model of Ni et al.[23] are not identified for this reason, and the problem grows as the number of environments increases.

where $J_K$ is a matrix of 1 s. So adding $\mu$ to the columns of $Z$ comes WLOG if and only if there always exists a $\sigma_v^2$ such that the implied $\Sigma_v$ above remains in the set of allowed covariance matrices. In particular, demeaning the columns of $Z$ comes WLOG only under the Full model, because the LHS in the final equation is non-diagonal for $\mu \neq 0$.

Fundamentally, this asymmetry between $Z$ and $G$ is because we treat the former's main effect as fixed ($\alpha$) and the latter's as random ($\beta$).

The scale of columns of $Z$ is irrelevant for Free GxEMM because the feasible set for $V$ is closed under conjugation by diagonal matrices—multiplying $Z_{,k}$ by $\lambda$ can be counterbalanced by multiplying row $k$ and column $k$ of $V$ by $\lambda^{-1}$.

However, the column scaling of $Z$ can be very important for IID GxEMM. By default, we do not scale or center $Z$ when it measures discrete environments in GxEMM. We do, nonetheless, require that $\|Z\|_F^2 = N$, which comes without loss of generality and permits simpler formulas

SNP ($\gamma$) must correlate with the product of its direct effect on the phenotype ($\beta$) and its effect on the environment ($\alpha$):

$$\eta := \mathbb{E}(\alpha\beta\gamma) \neq 0$$

$$\mathbb{E}(y_i y_j) = G_i \mathbb{E}(\beta\beta^T) G_j^T + (G_i \otimes G_i)\left(\mathbb{E}\left([I_L \otimes \alpha]\mathbb{E}(\gamma\gamma^T|\alpha)[I_L \otimes \alpha]^T\right)\right)(G_j \otimes G_j)^T +$$
$$\mathbb{E}\left((G_i \otimes e_i)\gamma\gamma^T(G_j \otimes e_j)^T\right) + G_i \mathbb{E}\left(\beta\gamma^T[I_L \otimes \alpha]^T\right)(G_j \otimes G_j)^T 2$$
$$= \sigma_\beta^2 G_i G_j^T + (G_i \otimes G_i)\left(I_L \otimes \mathbb{E}(\alpha\Sigma_{\gamma|\alpha}\alpha^T)\right)(G_j \otimes G_j)^T + \sum_{l_1,l_2,l_3,p} G_{il_1}G_{jl_2}G_{jl_3}\mathbb{E}(\beta_{l_1}\alpha_{l_2p}\gamma_{l_3p})2$$

for heritability. We note that Moore et al.[21] and common practice in many machine learning tasks centers and scales columns of $Z$; when environments are discrete, this corresponds to an assumption that rarer environments harbor larger specific genetic effects.

In the Full model, the columns of $Z$ become identified only up to span($Z$). This has pros and cons: a benefit is invariance to linear transformations, but a con is that individual elements of $V$ become meaningless. Conversely, individual elements of the diagonal $V$ fit by Free GxEMM are meaningful, as the explicit assumption of diagonal $V$ identifies a basis for span($Z$).

## Appendix C: Gene-Environment Correlation and GxE Bias in Polygenic Setting

It is well known that gene-environment (G-E) correlation can cause bias in gene-environment interaction (GxE) tests if uncorrected. This is a general statistical issue in fixed effect regression, and main covariate effects should almost always be adjusted when testing interactions. This has been emphasized recently in the GxE literature for testing individual genetic variants; e.g., Dudbridge and Fletcher[30] provide a worked example analyzing the simplest case of a single causal genetic variant.

Ni et al.[23] recently introduced a polygenic GxE model that incorporates G-E. The motivation for modeling G-E was by analogy to the fixed effect setting, where the bias is not in dispute. However, Ni et al.[23] provide no reason this bias would persist in the polygenic setting, nor simulations where G-E causes false-positive GxE.

Here, we provide a theoretical analysis that answers these questions. It is possible for polygenic G-E to bias GxE, but only under a very specific condition that, to our knowledge, has never been evaluated in reality. In particular, we feel that G-E correlation is unlikely to be a significant source of false GxE in the polygenic setting.

Specifically, the condition for G-E correlation to bias GxE estimates is that the GxE effects for each

For example, in the case of GxSmoking for BMI, the fact that G-Smoking correlation exists ($\alpha \neq 0$) is insufficient to cause bias. However, the polygenic GxSmoking estimates will be biased if the interaction of SNPs that both directly increase smoking likelihood ($\alpha > 0$) and BMI ($\beta > 0$) tend to have a positive interaction ($\gamma > 0$), indicating positive epistasis between main BMI effects and indirect effects mediated through smoking ($\eta > 0$). If, instead, we retain $\alpha, \beta > 0$ but assume $\gamma < 0$, then these two pathways to BMI are instead negatively epistatic, meaning the total SNP effect is less than the sum of its direct and smoking-mediated effects.

Intuitively, the bias from G-E correlation on GxE per-SNP averages out to 0 across all SNPs. For a small number of causal SNPs, however, the contributions will not perfectly cancel. In the simplest, most studied case—a single causal SNP—the problem is at its most extreme.

**Proposition 2.** *Assume the GxE model with G-E correlation*[23] *defined by:*

$$y = \mu + Z\omega + G\beta + (G*Z)\gamma + \epsilon$$

$$z = G\alpha + e$$

*Assume also that both $(e_i, \epsilon_i)$ and $(\alpha_l, \beta_l, \gamma_{(l)})$ are i.i.d., and also independent of each other, where $\gamma_{(l)} \in \mathbb{R}^P$ are the GxE effects for SNP l. Then G-E correlation causes false positive inflation in the HE regression test for polygenic GxE if and only if:*

$$\eta := \sum_p Cov\left(\alpha_{\cdot p} \circ \beta, \gamma_{(p)}\right) = \sum_{l,p} \mathbb{E}(\alpha_{lp}\beta_l\gamma_{lp}) \neq 0$$

*where $\gamma_{(p)} \in \mathbb{R}^L$ are the GxE interaction effects for environment p and $\gamma_{lp}$ is the SNP l-by-environment p interaction effect.*

*Proof.* First, note that the fixed effects $\mu$ and $\omega$ are eliminated by residualizing covariates; we do not track the impact of this finite-sample projection going forward, however, which is not quite formal but is very standard. The model simplifies to:

$$\gamma = G\beta + (G * (G\alpha))\gamma + (G * e)\gamma + \epsilon$$
$$= G\beta + (G * G)([I_L \otimes \alpha]\gamma) + (G * e)\gamma + \epsilon$$

In expectation—over the random variables $e$, $\epsilon$, $\alpha$, $\beta$, and $\gamma$—the quadratic form in $y$ for $i \neq j$ is:

defining $\Sigma_{\gamma|\alpha} := \mathbb{E}(\gamma_{(l)}\gamma_{(l)}^T | \alpha)$ and using the notation $f(i,j)^2 = f(i,j) + f(j,i)$ for any two-argument term $f$.

The final term simplifies because the expectation is zero unless $l_1 = l_2 = l_3$, giving:

$$\sum_{l_1,l_2,l_3,p} G_{il_1}G_{jl_2}G_{jl_3}\mathbb{E}\left(\beta_{l_1}\alpha_{l_2p}\gamma_{l_3p}\right) = \sum_{l,p} G_{il}G_{jl}^2\mathbb{E}\left(\beta_l\alpha_{lp}\gamma_{lp}\right)$$

$$= \left(\sum_{l,p} G_{il}G_{jl}^2\right)\mathbb{E}\left(\beta_{l'}\sum_p [\alpha_{l'p}\gamma_{l'p}]\right) \Rightarrow$$

$$\sum_{l_1,l_2,l_3,p} G_{il_1}G_{jl_2}G_{jl_3}\mathbb{E}\left(\beta_{l_1}\alpha_{l_2p}\gamma_{l_3p}\right)[2] = \eta K_{ij}^{1.5}L$$

(Equation A3)

where $l'$ is an arbitrary dummy index (these terms are i.i.d. over $l'$) and where we define $K_{ij}^{1.5} := \frac{1}{L}\sum_l G_{il}G_{jl}^2 + \frac{1}{L}\sum_l G_{il}^2 G_{jl}$.

We allow the conditional variance $\Sigma_{\gamma|\alpha}$ to depend on $\alpha$. But the GxE term $\mathbb{E}(\alpha\Sigma_{\gamma|\alpha}\alpha^T)$ is still proportional to $I_L$, because loci are assumed independent. Each diagonal entry is:

$$\mathbb{E}\left(\alpha_{l,}\Sigma_{\gamma|\alpha}\alpha_{l,}^T\right) = : \sigma_{\gamma,\alpha}^2$$

Together, these reductions give:

$$\mathbb{E}\left(y_iy_j\right) = \sigma_\beta^2 G_iG_j^T + \sigma_{\gamma,\alpha}^2\left(G_iG_j^T\right)^2 + \eta\sum_l G_{il}G_{jl}^2[2]$$

$$= L\sigma_\beta^2 K_{ij} + L\sigma_{\gamma,\alpha}^2 K_{ij}^2 + L\eta K_{ij}^{1.5}$$

(Equation A4)

This decomposes the expectation into the appropriate first two terms, for the homogeneous and GxE effects, and a third bias term. This bias is clearly zero when $\eta = 0$. The second term, for GxE, also appropriately disappears under the GxE null—if $\gamma$ is deterministically zero, then $\Sigma_{\gamma|\alpha} = 0$. $\square$

A large-$N$ argument could likely also be made that $K^{1.5} \to 0$ under certain genotype matrix structures.

$\eta$ is closely related to the concept of coordinated interaction we are developing in related work. $\eta$ conceptually measures the sum of coordinated interactions between the primary polygenic effect ($\beta$) and the indirect pathways through $Z_{\cdot p}$ ($\alpha_{\cdot p}$).

Our proposition does not claim unbiasedness in general for the estimated GxE variance components, nor does it control any overdispersion of the estimated GxE parameter due to G-E correlation. Instead, we show that the GxE estimate is unbiased when $\eta = 0$ under the null GxE hypothesis ($\gamma = 0$). This is precisely the criterion used in Dudbridge and Fletcher[30] to argue that G-E correlation *does* cause bias for GxE. Our result differs substantially, howev-

er, because we model the full spectrum of polygenic SNP effects rather than individual SNPs.

This result also relates to calculations in Sulc et al.,[76] which focuses on a related problem where the "E" is unmeasured. However, they assume (1) $\alpha \propto \beta$ and (2) that $Z \in \mathbb{R}^N$ is univariate, which together make the math more straightforward[76] and also do not link their results to the equivalent assumptions on the SNP effect sizes, which is crucial for biological interpretation. In particular, it is not clear from their analysis that significant interaction results are obtained if and only if there is coordinated interaction via $\eta \neq 0$.

In summary, we provide a necessary and sufficient condition for G-E correlation to bias GxE interaction tests in the polygenic setting. This demonstrates that the bias is likely negligible in practice, as well as provides a proper theoretical characterization of the relationship between G-E and GxE bias under polygeneicity. This was previously unknown: for example, Ni et al.[23] did not demonstrate GxE bias in simulations with G-E correlation.

## Appendix D: Marginalizing Random Interaction Coefficients

This section shows that the set of Gaussian GxE coefficients ($\gamma$) with covariance $D \otimes V$ ($V$ for the environmental covariance, $D$ for the SNPs covariance) describes essentially the same distributions as the set of linear mixed models with Hadamard kinship matrices $(GDG^T) \circ (ZVZ^T)$. The former are the natural model for polygenic GxE; the latter are intuitive (for discrete environments) estimates for environment-specific heritability and easily fit by REML. In practice, $D$ is usually taken to be the identity matrix.

**Proposition 3.** *Assume that* $\gamma \sim \mathcal{N}(0, \Sigma) \in \mathbb{R}^{P \times L}$, *that* $G \in \mathbb{R}^{N \times L}$ *has continuous, random entries, and that* $Z \in \mathbb{R}^{N \times P}$ *has rank P, with P < N. Write* $K := GDG^T$ *for some fixed D. Then*

$$\mathbb{V}((G * Z)\gamma) = K \circ \Omega \Leftrightarrow (\Sigma, \Omega) = (D \otimes W, ZWZ^T) \text{ for some } W$$

*where * is the column-wise Khatri-Rao product; $\circ$ is Hadamard; and $\otimes$ is Kronecker.*

*Proof.* First, note the identity:

$$(G_{i,} \otimes Z_{i,})(D \otimes W)(G_{j,} \otimes Z_{j,})^T = \left(G_{i,}DG_{j,}^T\right)\left(Z_{i,}WZ_{j,}^T\right) = K_{ij}\Omega_{ij}$$

Now, right to left is easy:

$$\mathbb{V}((G * Z)\gamma)_{ij} = (G_{i,} \otimes Z_{i,})\mathbb{V}(\gamma)(G_{j,} \otimes Z_{j,})^T = K_{ij}\Omega_{ij}$$

The other direction assumes decomposition of the variance into Hadamard products holds, so

$$(G_{i,} \otimes Z_{i,})(D \otimes W)(G_{j,} \otimes Z_{j,})^T = K_{ij}\Omega_{ij}$$
$$= (G_{i,} \otimes Z_{i,})\sum (G_{j,} \otimes Z_{j,})^T$$

Using the standard identity $(A \otimes B)\text{vec}(C) = \text{vec}(B^T CA)$, where $\text{vec}(\cdot)$ concatenates columns of a matrix, the above equality can be written as

$$\left(G_{j,} \otimes Z_{j,} \otimes G_{i,} \otimes Z_{i,}\right)\text{vec}(\Sigma - D \otimes W) = 0$$

By assumption, this identity holds $\forall i, j$ and almost all genotypes $G_{i,}, G_{j,} \in R^L$. The span of such pure four-way tensors is $\mathbb{R}^{L^2 P^2}$, so its kernel has dimension zero and thus $\Sigma - D \otimes W = 0$.

$\square$

A similar result can be obtained if $N$ grows large with fixed $L$.

We note that the easy direction appears implicitly in many human genetics papers (e.g., Robinson et al.,[20] Ni et al.,[23] Yang et al.[27]). We do not know where to find a clear proof, however, and are unaware of any discussion of the other direction.

## Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.11.015.

## Acknowledgments

## Declaration of Interests

The authors declare no competing interests.

## Web Resources

GCTA, https://cnsgenomics.com/software/gcta

GxEMM (free R implementation), https://github.com/andywdahl/gxemm

GxEMM scripts to reproduce the simulations and rat analyses, https://github.com/andywdahl/gxemm-scripts

GxEMM to reproduce the pyschENCODE analyses, https://github.com/nguyenkhiemv/GxEMM

MTG2, https://sites.google.com/site/honglee0707/mtg2

## References

1. Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. USA *109*, 1204–1209.

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. Lee, M.N., Ye, C., Villani, A.C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., et al. (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science *343*, 1246980.

4. Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., and Knight, J.C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949.

5. Knowles, D.A., Davis, J.R., Edgington, H., Raj, A., Favé, M.J., Zhu, X., Potash, J.B., Weissman, M.M., Shi, J., Levinson, D.F., et al. (2017). Allele-specific expression reveals interactions between genetic variation and environment. Nat. Methods *14*, 699–702.

6. Glass, D., Viñuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A., Hedman, A.K., Small, K.S., Buil, A., et al.; UK Brain Expression consortium; and MuTHER consortium (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. Genome Biol. *14*, R75.

7. Battle, A., Brown, C.D., Engelhardt, B.E., Montgomery, S.B.; GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx (eGTEx) groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI; Biospecimen Collection Source Site—RPCI; Biospecimen Core Resource—VARI; Brain Bank Repository—University of Miami Brain Endowment Bank; Leidos Biomedical—Project Management; ELSI Study; Genome Browser Data Integration &Visualization—EBI; Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz; Lead analysts; Laboratory, Data Analysis &Coordinating Center (LDACC); NIH program management; Biospecimen collection; Pathology; and eQTL manuscript working group (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

8. Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J., et al. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. *49*, 139–145.

9. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94.

10. Favé, M.J., Lamaze, F.C., Soave, D., Hodgkinson, A., Gauvin, H., Bruat, V., Grenier, J.C., Gbeha, E., Skead, K., Smargiassi, A., et al. (2018). Gene-by-environment interactions in urban populations modulate risk phenotypes. Nat. Commun. *9*, 827.

11. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IIBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

12. Myers, R.A., Scott, N.M., Gauderman, W.J., Qiu, W., Mathias, R.A., Romieu, I., Levin, A.M., Pino-Yanes, M., Graves, P.E., Villarreal, A.B., et al.; GRAAD (2014). Genome-wide interaction studies reveal sex-specific asthma risk alleles. Hum. Mol. Genet. *23*, 5251–5259.

13. Mitra, I., Tsang, K., Ladd-Acosta, C., Croen, L.A., Aldinger, K.A., Hendren, R.L., Traglia, M., Lavillaureix, A., Zaitlen, N., Oldham, M.C., et al. (2016). Pleiotropic Mechanisms Indicated for Sex Differences in Autism. PLoS Genet. *12*, e1006425.

14. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638.

15. 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.

16. Exner, D.V., Dries, D.L., Domanski, M.J., and Cohn, J.N. (2001). Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. N. Engl. J. Med. *344*, 1351–1357.

17. Mega, J.L., Simon, T., Collet, J.P., Anderson, J.L., Antman, E.M., Bliden, K., Cannon, C.P., Danchin, N., Giusti, B., Gurbel, P., et al. (2010). Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. JAMA *304*, 1821–1830.

18. Riaz, N., Havel, J.J., Kendall, S.M., Makarov, V., Walsh, L.A., Desrichard, A., Weinhold, N., and Chan, T.A. (2016). Recurrent SERPINB3 and SERPINB4 mutations in patients who respond to anti-CTLA4 immunotherapy. Nat. Genet. *48*, 1327–1329.

19. Peterson, R.E., Cai, N., Dahl, A.W., Bigdeli, T.B., Edwards, A.C., Webb, B.T., Bacanu, S.A., Zaitlen, N., Flint, J., and Kendler, K.S. (2018). Molecular Genetic Analysis Subdivided by Adversity Exposure Suggests Etiologic Heterogeneity in Major Depression. Am. J. Psychiatry *175*, 545–554.

20. 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.

21. 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.

22. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

23. 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.

24. 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.

25. 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.

26. Hernandez, R.D., Uricchio, L.H., Hartman, K., Ye, C., Dahl, A., and Zaitlen, N. (2019). Ultrarare variants drive substantial cis heritability of human gene expression. Nat. Genet. *51*, 1349–1355.

27. 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.

28. Casale, F.P., Horta, D., Rakitsch, B., and Stegle, O. (2017). Joint genetic analysis using variant sets reveals polygenic gene-context interactions. PLoS Genet. *13*, e1006693.

29. Kang, E.Y., Lee, C.H., Furlotte, N.A., Joo, J.W.J., Kostem, E., Zaitlen, N., Eskin, E., and Han, B. (2018). An Association Mapping Framework To Account for Potential Sex Difference in Genetic Architectures. Genetics *209*, 685–698.

30. Dudbridge, F., and Fletcher, O. (2014). Gene-environment dependence creates spurious gene-environment interaction. Am. J. Hum. Genet. *95*, 301–307.

31. Dempster, E.R., and Lerner, I.M. (1950). Heritability of Threshold Characters. Genetics *35*, 212–236.

32. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. *88*, 294–305.

33. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. *9*, e1003264.

34. Jiang, J., Li, C., Paul, D., Yang, C., and Zhao, H. (2016). On high-dimensional misspecified mixed model analysis in genome-wide association study. Ann. Stat. *44*, 2127–2160.

35. Steinsaltz, D., Dahl, A., and Wachter, K.W. (2018). Statistical properties of simple random-effects models for genetic heritability. Electron. J. Stat. *12*, 321–356.

36. Stephens, M., and Balding, D.J. (2009). Bayesian statistical methods for genetic association studies. Nat. Rev. Genet. *10*, 681–690.

37. 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.

38. Steinsaltz, D., Dahl, A., and Wachter, K.W. (2018). On Negative Heritability and Negative Estimates of Heritability. bioRxiv. https://doi.org/10.1101/232843.

39. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.; and Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. *47*, 291–295.

40. Chen, G.B. (2014). Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. Front. Genet. *5*, 107.

41. Abney, M. (2015). Permutation testing in the presence of polygenic variation. Genet. Epidemiol. *39*, 249–258.

42. Schweiger, R., Fisher, E., Weissbrod, O., Rahmani, E., Müller-Nurasyid, M., Kunze, S., Gieger, C., Waldenberger, M., Rosset, S., and Halperin, E. (2018). Detecting heritable phenotypes without a model using fast permutation testing for heritability and set-tests. Nat. Commun. *9*, 4919.

43. Wu, Y., and Sankararaman, S. (2018). A scalable estimator of SNP heritability for biobank-scale data. Bioinformatics *34*, i187–i194.

44. 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.

45. Dahl, A., Iotchkova, V., Baud, A., Johansson, Å., Gyllensten, U., Soranzo, N., Mott, R., Kranis, A., and Marchini, J. (2016). A multiple-phenotype imputation method for genetic studies. Nat. Genet. *48*, 466–472.

46. Baud, A., Hermsen, R., Guryev, V., Stridh, P., Graham, D., McBride, M.W., Foroud, T., Calderari, S., Diez, M., Ockinger, J., et al.; Rat Genome Sequencing and Mapping Consortium (2013). Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. Nat. Genet. *45*, 767–775.

47. 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.

48. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al.; PsychENCODE Consortium (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, 9–999.

49. Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T., et al.; PsychENCODE Consortium (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*, eaat8464.

50. Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLoS Genet. *3*, 1724–1735.

51. Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. *127*, 595–607.

52. Lee, S.H., and Van der Werf, J.H.J. (2006). Using dominance relationship coefficients based on linkage disequilibrium and linkage with a general complex pedigree to increase mapping resolution. Genetics *174*, 1009–1016.

53. Hanson, A.J., Banks, W.A., Hernandez Saucedo, H., and Craft, S. (2016). Apolipoprotein E Genotype and Sex Influence Glucose Tolerance in Older Adults: A Cross-Sectional Study. Dement. Geriatr. Cogn. Disord. Extra *6*, 78–89.

54. CONVERGE consortium (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. Nature *523*, 588–591.

55. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). A Swedish national twin study of lifetime major depression. Am. J. Psychiatry *163*, 109–114.

56. McGuffin, P., Katz, R., Watkins, S., and Rutherford, J. (1996). A hospital-based twin register of the heritability of DSM-IV unipolar depression. Arch. Gen. Psychiatry *53*, 129–136.

57. Price, A.L., Patterson, N., Hancks, D.C., Myers, S., Reich, D., Cheung, V.G., and Spielman, R.S. (2008). Effects of cis and trans genetic ancestry on gene expression in African Americans. PLoS Genet. *4*, e1000294.

58. Huckins, L.M., Dobbyn, A., Ruderfer, D.M., Hoffman, G., Wang, W., Pardiñas, A.F., Rajagopal, V.M., Als, T.D., T Nguyen, H., Girdhar, K., et al.; CommonMind Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; and iPSYCH-GEMS Schizophrenia Working Group (2019). Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nat. Genet. *51*, 659–674.

59. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.

60. Fullard, J.F., Giambartolomei, C., Hauberg, M.E., Xu, K., Voloudakis, G., Shao, Z., Bare, C., Dudley, J.T., Mattheisen, M., Robakis, N.K., et al. (2017). Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. Hum. Mol. Genet. *26*, 1942–1951.

61. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat. Commun. *9*, 224.

62. Wu, Y., Zeng, J., Zhang, F., Zhu, Z., Qi, T., Zheng, Z., Lloyd-Jones, L.R., Marioni, R.E., Martin, N.G., Montgomery, G.W., et al. (2018). Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat. Commun. *9*, 918.

63. Ma, L., Semick, S.A., Chen, Q., Li, C., Tao, R., Price, A.J., Shin, J.H., Jia, Y., Brandon, N.J., Cross, A.J., et al.; BrainSeq Consortium (2019). Schizophrenia risk variants influence multiple classes of transcripts of sorting nexin 19 (SNX19). Mol. Psychiatry. Published online January 11, 2019. https://doi.org/10.1038/s41380-018-0293-0.

64. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453.

65. Dobbyn, A., Huckins, L.M., Boocock, J., Sloofman, L.G., Glicksberg, B.S., Giambartolomei, C., Hoffman, G.E., Perumal, T.M., Girdhar, K., Jiang, Y., et al.; CommonMind Consortium (2018). Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. Am. J. Hum. Genet. *102*, 1169–1184.

66. Hou, Y., Liang, W., Zhang, J., Li, Q., Ou, H., Wang, Z., Li, S., Huang, X., and Zhao, C. (2018). Schizophrenia-associated rs4702 G allele-specific downregulation of FURIN expression by miR-338-3p reduces BDNF production. Schizophr. Res. *199*, 176–180.

67. Crawford, L., Zeng, P., Mukherjee, S., and Zhou, X. (2017). Detecting epistasis with the marginal epistasis test in genetic mapping studies of quantitative traits. PLoS Genet. *13*, e1006869.

68. Han, B., and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. *88*, 586–598.

69. Wen, X., and Stephens, M. (2014). Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene–environment interactions. Ann. Appl. Stat. *8*, 176–203.

70. 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.

71. Dahl, A., Cai, N., Ko, A., Laakso, M., Pajukanta, P., Flint, J., and Zaitlen, N. (2019). Reverse GWAS: Using genetics to identify and model phenotypic subtypes. PLoS Genet. *15*, e1008009.

72. Liu, X., Mefford, J.A., Dahl, A., Subramaniam, M., Battle, A., Price, A.L., and Zaitlen, N. (2018). GBAT: a gene-based

association method for robust trans-gene regulation detection. bioRxiv. https://doi.org/10.1101/395970.

73. Diego, V.P., Rainwater, D.L., Wang, X.L., Cole, S.A., Curran, J.E., Johnson, M.P., Jowett, J.B., Dyer, T.D., Williams, J.T., Moses, E.K., et al. (2007). Genotype x adiposity interaction linkage analyses reveal a locus on chromosome 1 for lipoprotein-associated phospholipase A2, a marker of inflammation and oxidative stress. Am. J. Hum. Genet. *80*, 168–177.

74. Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., and Price, A.L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. *9*, e1003520.

75. Traglia, M., Bseiso, D., Gusev, A., Adviento, B., Park, D.S., Mefford, J.A., Zaitlen, N., and Weiss, L.A. (2017). Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. Genetics *205*, 979–992.

76. Sulc, J., Mounier, N., Felix, G., Winkler, T., Wood, A.R., Frayling, T.M., et al. (2019). Maximum likelihood method quantifies the overall contribution of gene-environment interaction to complex traits: an application to obesity traits. bioRxiv. https://doi.org/10.1101/632380.