The American Journal of Human Genetics, Volume *111*

Supplemental information

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

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

Supplemental Notes

Note S1

We can extend each of our models to include covariates as follows:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\alpha} + \sum_k \boldsymbol{Z}_k\boldsymbol{\beta}_k + \boldsymbol{\epsilon} \tag{1}$$

Here $\boldsymbol{W}$ is a $N \times C$ matrix of covariates while $\boldsymbol{\alpha}$ is a vector of fixed effects of length C. Matrix $\boldsymbol{Z}_k$ can represent genotype or GxE matrices. In this setting, we need to solve the following normal equations to estimate the variance components.

$$\begin{bmatrix} \boldsymbol{T} & \boldsymbol{b} \\ \boldsymbol{b}^T & N-C \end{bmatrix} \begin{bmatrix} \sigma^2_1 \\ \vdots \\ \sigma^2_K \\ \sigma^2_e \end{bmatrix} = \begin{bmatrix} \boldsymbol{c} \\ \boldsymbol{y}^T\boldsymbol{V}\boldsymbol{y} \end{bmatrix} \tag{2}$$

Here $\boldsymbol{V} = \boldsymbol{I}_N - \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T$ and $\boldsymbol{T}$ is a $K \times K$ matrix where $T_{k,l} = tr(\boldsymbol{K}_k\boldsymbol{V}\boldsymbol{K}_l\boldsymbol{V})$, and b is a vector of length K where $b_k = tr(\boldsymbol{V}\boldsymbol{K}_k)$, and $\boldsymbol{c}$ is a vector of length K where $c_k = \boldsymbol{y}^T\boldsymbol{V}\boldsymbol{K}_k\boldsymbol{V}\boldsymbol{y}$, $\boldsymbol{K}_k = \frac{\boldsymbol{Z}_k\boldsymbol{Z}_k^T}{M}$ where M is the number of column in Z_k. Commonly, the number of covariates C is small (tens to hundreds) so including covariates does not significantly affect the computational cost. The cost of computing the elements of the normal equations 2 includes the cost of inverting $\boldsymbol{W}^T\boldsymbol{W}$ which is a $C \times C$ matrix and multiplying $\boldsymbol{W}$ by a real-valued vector of length N can be computed in $\mathcal{O}(C^3 + NC)$.

Note S2

To simulate MAF and LD-dependent architectures, we simulated phenotypes from genotypes using the following model that extends prior models of additive genetic architecture[1,2] to include GxE effects:

$$\begin{aligned} \sigma^2_{g,m} &= Sc_m w_m^b [f_m(1-f_m)]^a \\ (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, .., \boldsymbol{\beta}_m)^T &\sim \mathcal{N}(\boldsymbol{0}, diag(\sigma^2_{g,1}, \sigma^2_{g,2}, ..., \sigma^2_{g,m})) \\ \sigma^2_{gxe,m} &= S^{'} c^{'}_m w_m^b [f_m(1-f_m)]^a \\ (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, .., \boldsymbol{\alpha}_m)^T &\sim \mathcal{N}(\boldsymbol{0}, diag(\sigma^2_{gxe,1}, \sigma^2_{gxe,2}, ..., \sigma^2_{gxe,m})) \\ \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha} &\sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X} \odot \boldsymbol{E})\boldsymbol{\alpha}, (1 - h^2_g - h^2_{gxe} - h^2_{nxe})\boldsymbol{I}_N + h^2_{nxe}(\boldsymbol{I}_N \odot \boldsymbol{E})) \end{aligned} \tag{3}$$

where $h^2_g, h^2_{gxe}, h^2_{nxe} \in [0,1]$, $a \in \{0, 0.75\}$, $b \in \{0,1\}$. Here S and $S^{'}$ are normalizing constants chosen so that $\sum_{m=1}^M \sigma^2_{g,m} = h^2_g$, $\sum_{m=1}^M \sigma^2_{gxe,m} = h^2_{gxe}$. Additive and GxE effect sizes are denoted by $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ respectively. f_m and w_m are the minor allele frequency and LD score of m^{th} SNP respectively. In this model, $c_m, c^{'}_m \in \{0,1\}$ are indicator variables for the causal status of SNP m ($c_m = 1$ and $c^{'}_m = 1$ for all SNPs). The LD score of a SNP is defined to be the sum of the squared correlation of the SNP with all other SNPs that lie within a specific distance. Setting $a = 0, b = 0$ for additive effects results in the GCTA model where the per-standardized genotypic effect sizes at a SNP do not vary with MAF or LD score[3,4] while setting $a = 0.75, b = 1$ results in the LDAK model[5]. Our simulations assume that the GxE effects follow the same coupling as the additive effects. We used phenotypes from $N = 40,000$ individuals genotyped at $M = 454,207$ SNPs on the UKB array.

Note S3

We described simulation details of several scenarios in the section **Robustness of GENIE in simulations**. To simulate Y correlated with E, we first generated $\boldsymbol{y'} = \boldsymbol{X\beta} + (\boldsymbol{X} \odot \boldsymbol{E})\boldsymbol{\alpha} + (\boldsymbol{I}_N \odot \boldsymbol{E})\boldsymbol{\delta} + \boldsymbol{\epsilon}$ as in the GENIE model and then let $\boldsymbol{y} = \frac{r}{1-r^2}\boldsymbol{E} + \boldsymbol{y'}$ be the simulated phenotype, where r is the simulated correlation. Since the environment variables are also included in the covariates as a fixed effect, we expected that the correlation between Y and E would not bias the estimates. The heritable E was simulated as $\boldsymbol{E} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \sigma_g^2\boldsymbol{I})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, (1-\sigma_g^2)\boldsymbol{I})$, with σ_g^2 denoting the variance component associated with additive genetic effects, and M the number of SNPs. We simulated the collider bias scenario: the phenotypes were simulated to be correlated with a heritable environmental variable through an unobserved confounder [6]. Finally, we drew the environment noise component ($\boldsymbol{\epsilon}$ in the model) from the Student's t-distribution with a degree of freedom = 4 to simulate the heavy-tailed noise.

Note S4

In our additional investigation of GxE heritability of phenotypes likely affected by medication, we repeated our previous analysis using GENIE after performing heuristic adjustments of SBP, DBP [7], LDL, and cholesterol [8] for relevant medication usage. Our results were broadly consistent with and without the heuristic adjustments though there are notable exceptions. Specifically, h^2_{gxe} estimates for all four exposures remained significant using the adjusted SBP/DBP values (Figure S37) while h^2_{gxe} estimates for LDL and cholesterol were about the same for smoking, sex, and statin usage. However, h^2_{gxAge} estimates decreased substantially after the adjustment for statin although the estimates for cholesterol remained statistically significant (Figure S38). The differences in h^2_{gxAge} estimates for cholesterol also appeared in the downstream tissue-specific enrichment analysis. While the additive genetic effects were concordant with previous results, h^2_{gxAge} estimates were weakly correlated after the adjustment (Figure S39, S40). After stating adjustment, we found that none of the tissue-specific gene sets were enriched for h^2_{gxAge} at FDR < 0.1 (consistent with the attenuation of the genome-wide estimates). To directly account for the effects of statin usage on baseline LDL/cholesterol levels, we also estimated $h^2_{gxStatin}$ for HbA1c while also including GxLDL and GxTC in the model (with these cholesterol traits treated as environment exposures). We observed that $h^2_{gxStatin}$ estimates remained significant and consistent with estimates from our previous analyses. Further, there was no substantial difference in the total h^2_{gxe} estimates across the three exposures (Figure S41).

Supplemental Figures

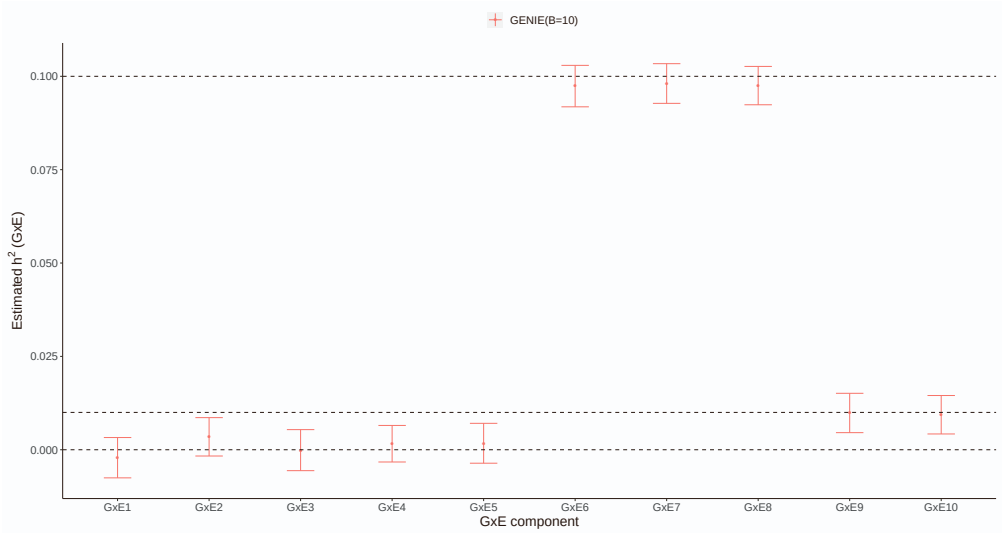

Figure S1: **Accuracy of GENIE when applied to multiple environmental variables**. We evaluated h^2_{gxe} estimates of GENIE for different values of σ^2_{gxe}. We simulated phenotypes with 10 environmental variables where $\sigma^2_g = 0.2$, $\sigma^2_{ge1} = \sigma^2_{ge2} = \sigma^2_{ge3} = \sigma^2_{ge4} = \sigma^2_{ge5} = 0$, $\sigma^2_{ge6} = \sigma^2_{ge7} = \sigma^2_{ge8} = 0.10$ and $\sigma^2_{ge9} = \sigma^2_{ge10} = 0.01$. Points and error bars represent the mean and ± 2 SE. Mean and SE are computed from 100 replicates. Here B is the number of random vectors used by GENIE with $B = 10$ the value that we use as default (we reported values of means, SEs, and p-values of a test of the null hypothesis of no bias in the estimates of variance components in Table S1).

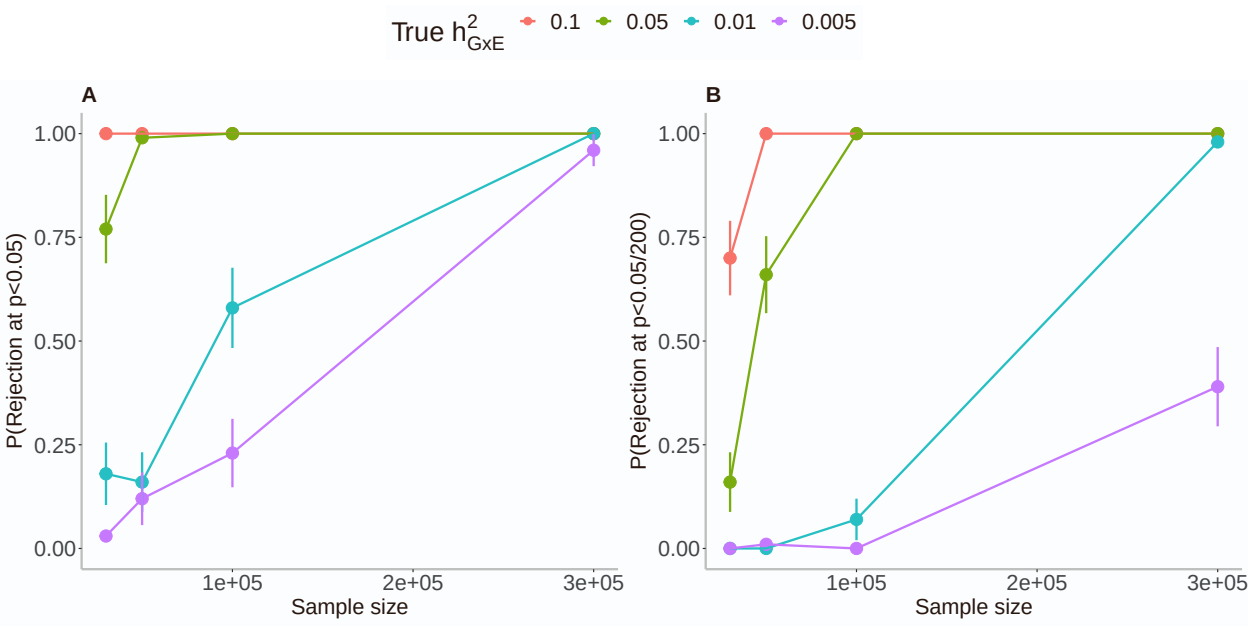

Figure S2: **The power of GENIE as a function of sample size in the absence of noise heterogeneity.** The power of GENIE under different GxE heritability as a function of sample size in the absence of noise heterogeneity at the significance level of (**A**) 0.05 and (**B**) 0.05/200. We evaluated the power of GENIE to detect GxE heritability across 0.005, 0.01, 0.05, and 0.1 on 100 simulated phenotypes. The phenotypes were simulated with varying GxE effects on $M = 454,207$ SNPs and $N = 30K$, $50K$, $100K$, or $290K$ individuals. The additive genetic heritability was fixed to be 0.25 across all simulations. The power calculated as P(Rejection at $p < t$) for $t \in \{0.05, 0.05/200\}$ along with the 95% binomial confidence interval using the normal approximation was reported for each simulation.

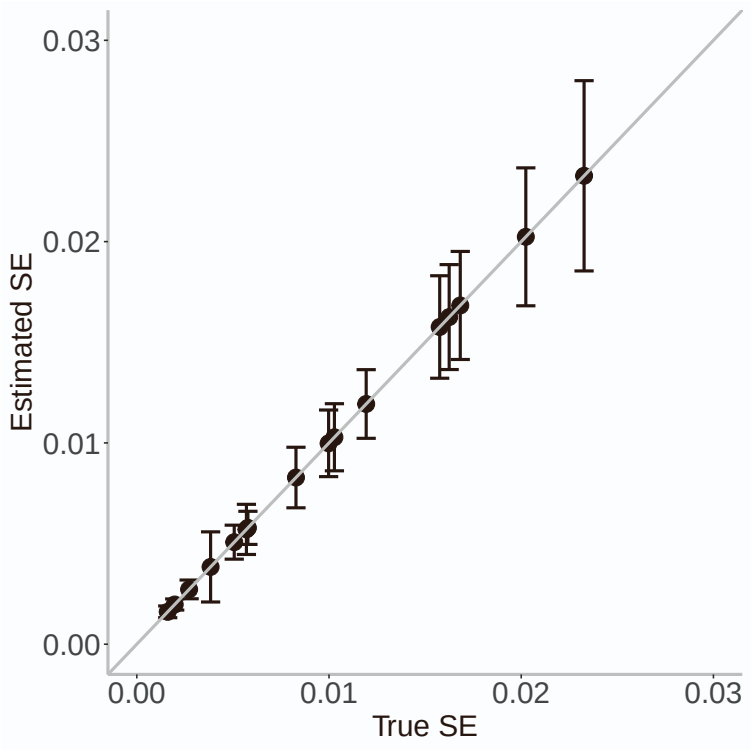

Figure S3: **Comparison between the estimated standard errors with the true standard errors for GxE heritability estimates in simulations.** Comparison of the estimated SEs using the block jackknife method against the true SEs of the point estimates for varying magnitudes of sample sizes under different true GxE heritability levels. The sample size ranged across $30K$, $50K$, $100K$, and $290K$, and true GxE heritability varied across 0.005, 0.01, 0.05, and 0.1. The average of estimated SE with 2 standard errors was shown, and the true SE was calculated based on the GxE heritability point estimates across 100 replicates in each simulation.

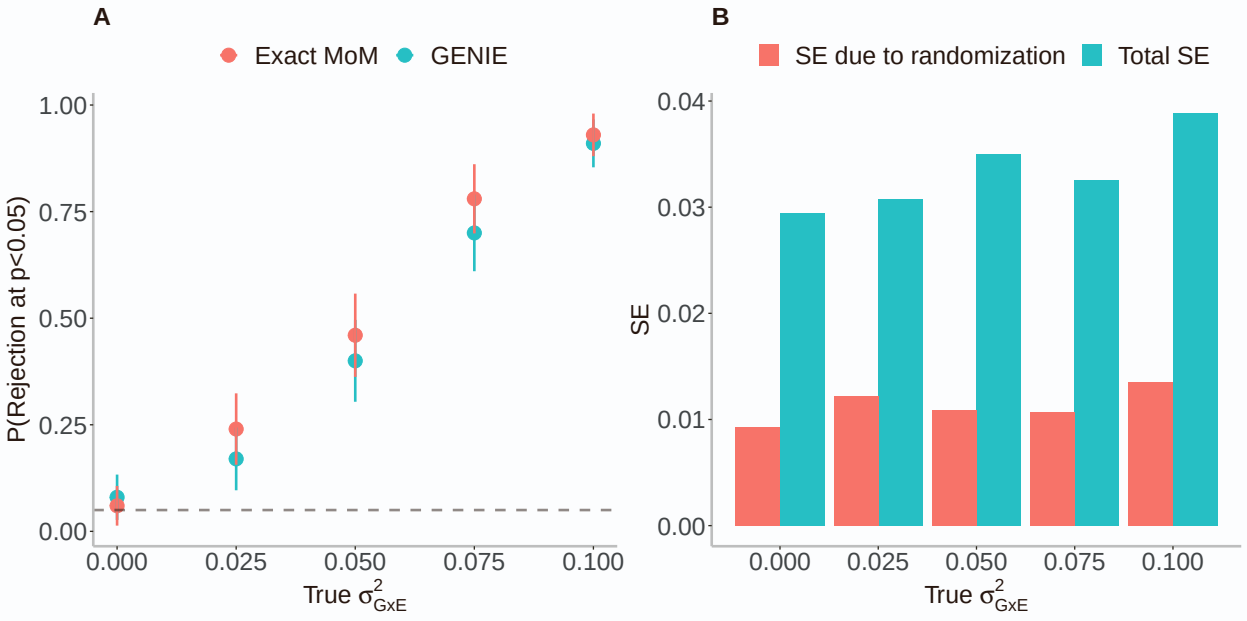

Figure S4: **Investigation of the power decrease of GENIE due to randomization on small-scale datasets in simulations.** (**A**) Comparison of the power and false positive rate (FPR) between exact MoM (as implemented in GCTA-HE) and GENIE. The FPR/Power is defined as the fraction of rejections of the null: $\sigma^2_{GxE} = 0$ at $p < 0.05$. The error bars correspond to the estimated 95% CI of the rejection rate. (**B**) Comparison of total SE and the SE due to randomization in simulations with nonzero σ^2_{GxE}. The total SE was calculated as the standard deviation of estimated GxE heritability over 100 replicates while the SE due to randomization was derived as the standard deviation of estimated GxE over 100 runs on a specific replicate.

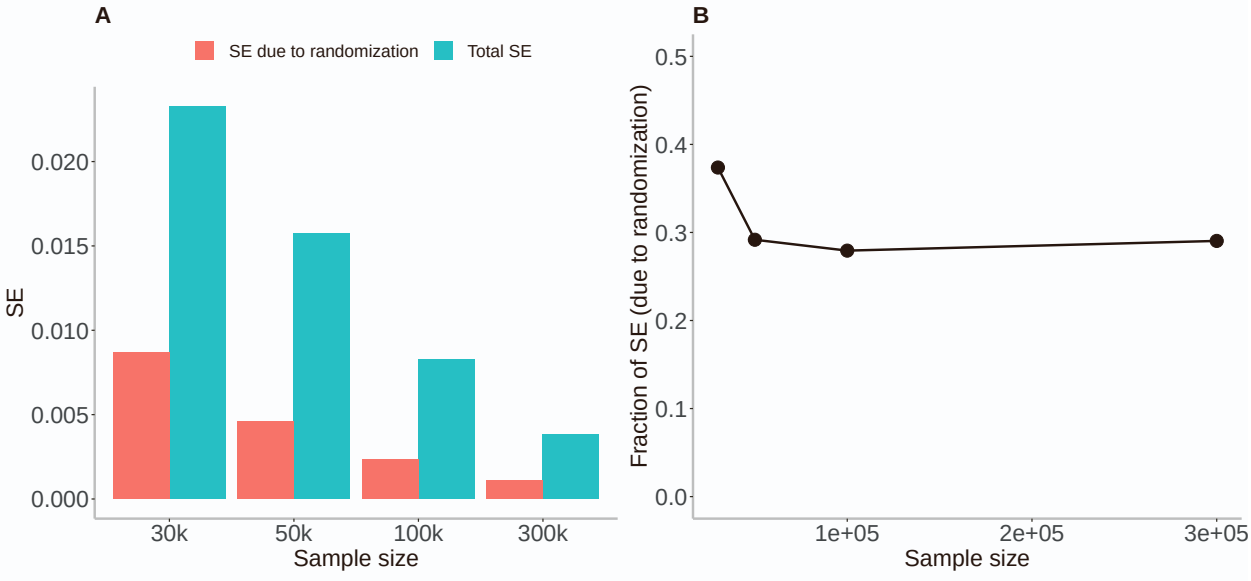

Figure S5: **Contribution of randomization to the standard error of GxE estimates.** (**A**) The mean of SEs due to randomization and total SE and (**B**) the fraction of SE due to randomization across different sample sizes. The SE due to randomization was calculated as the mean of SEs over 100 runs with different random seeds on a specific replicate, and the total SE was the true SE calculated over 100 replicates. The fraction of SE due to randomization is around 30% to 40% across different sample sizes.

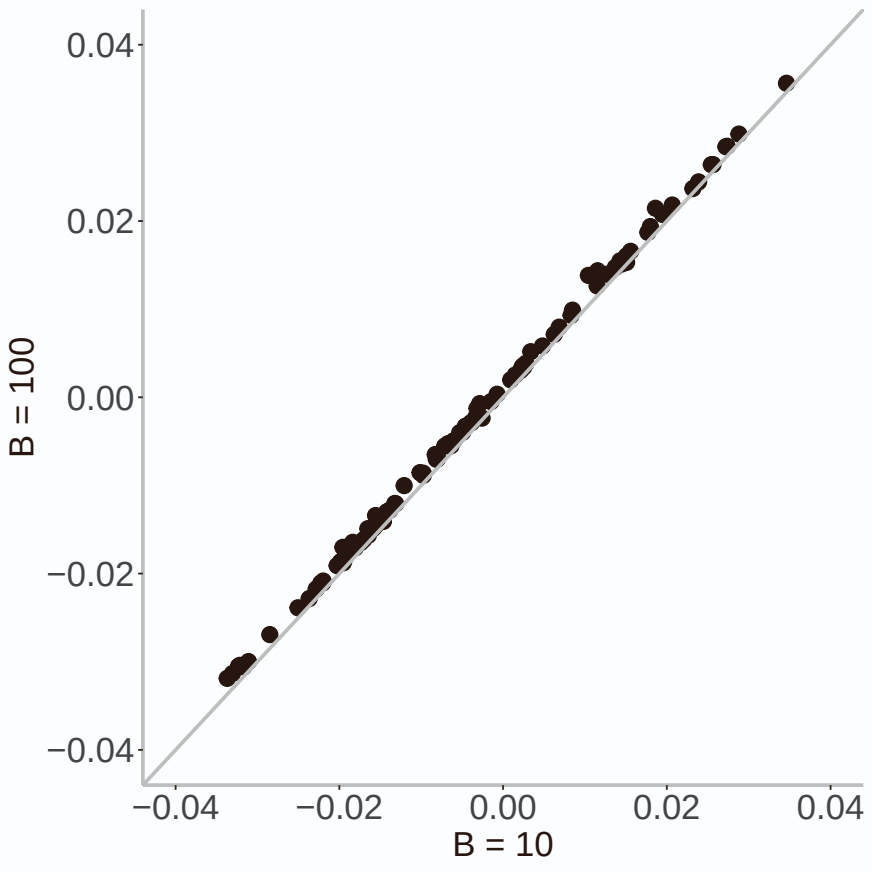

Figure S6: **Comparison of h^2_{gxe} estimates with $B = 10$ and $B = 100$.** We simulated phenotypes from $M = 60,000$ array SNPs and $N = 10,000$ individuals where $h^2_g = 0.1$, $h^2_{gxe} = 0$ and the causal ratio is 10% .

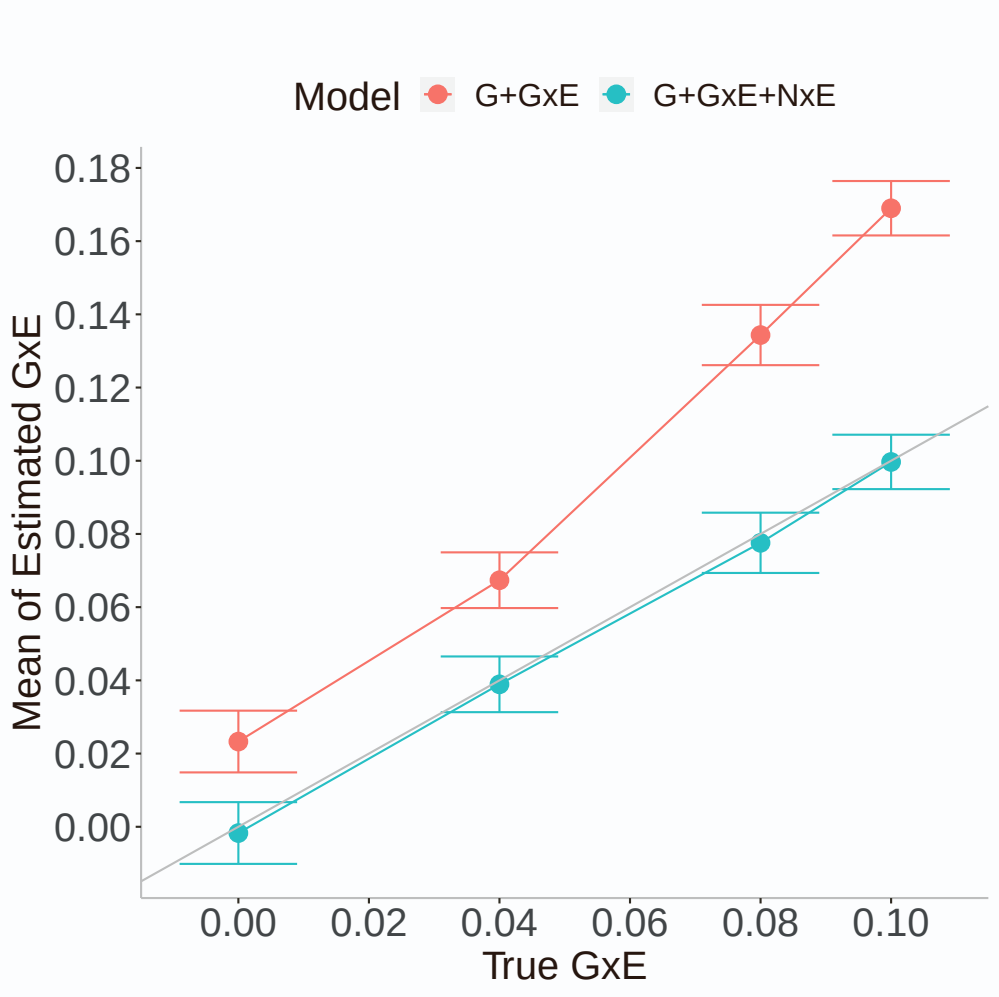

Figure S7: **Effect of noise heterogeneity (NxE) on the accuracy of estimates of GxE heritability in simulations**. Comparison of GxE heritability estimates from GENIE under a G+GxE model to those from a G+GxE+NxE model. Model G+GxE refers to a model with additive and gene-by-environment interaction components. Model G+GxE+NxE refers to a model with additive, gene-by-environment interaction, and noise heterogeneity (noise-by-environment interaction) components. We simulated phenotypes with NxE effects and GxE effects across $N = 291,273$ individuals genotyped at $M = 454,207$ SNPs. The x-axis and y-axis correspond to the true GxE and the mean of the estimated GxE (from 100 replicates), respectively. Points and error bars represent the mean and $\pm$ SE, respectively.

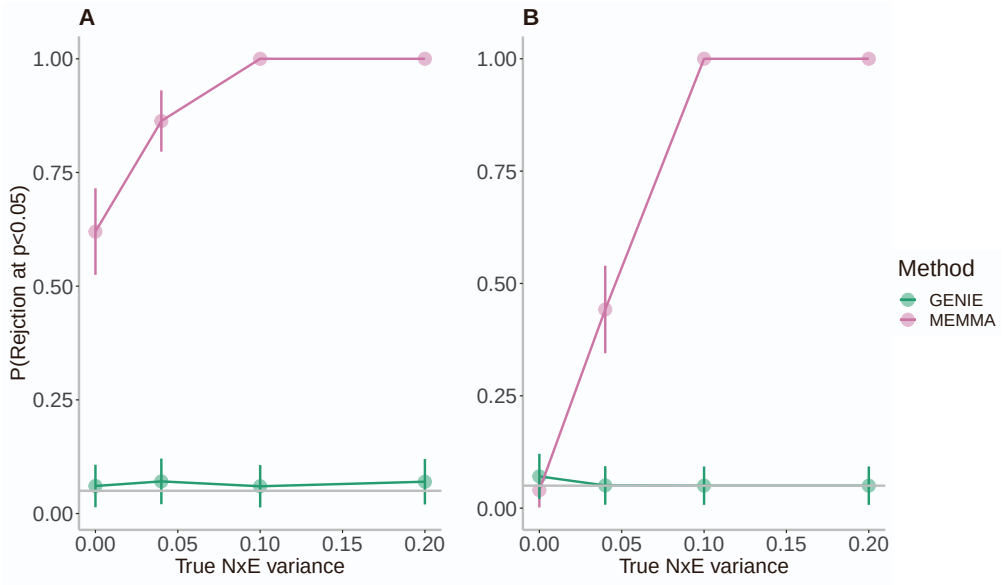

Figure S8: **Comparisons with MEMMA with the presence of noise heterogeneity in simulations**. We assessed the calibration of GENIE and MEMMA using (**A**) estimated SEs and (**B**) true SEs in simulations. We performed simulations with no GxE heritability but with varying magnitudes of the variance of the NxE effect. We computed the false positive rate as the fraction of rejections P(p-value of a test of the null hypothesis of zero GxE heritability < 0.05) over 100 replicates of phenotypes. The phenotypes were simulated from $N = 40,000$ individuals and $M = 454,207$ SNPs. MEMMA has biased estimates of SE, and the false positive rate is calibrated in the absence of a NxE effect using the true SEs. Error bars correspond to the estimated 95% CI of the rejection rate.

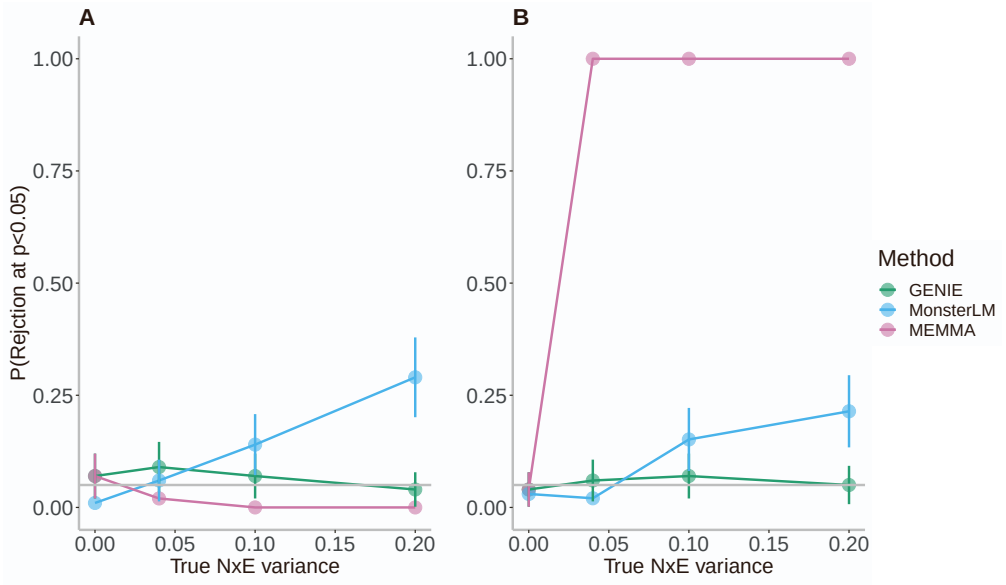

Figure S9: **Effects of estimated standard errors on controlling the false positive rate.** We assessed the calibration of GENIE and existing methods using their true SE instead of the estimation of SE in simulations with (**A**) continuous and (**B**) discrete environment exposures. The phenotypes were simulated from $N = 40,000$ individuals and $M = 223,591$ SNPs filtered from $M = 454,207$ SNPs with the genotype QC steps in MonsterLM. Error bars correspond to the estimated 95% CI of the rejection rate.

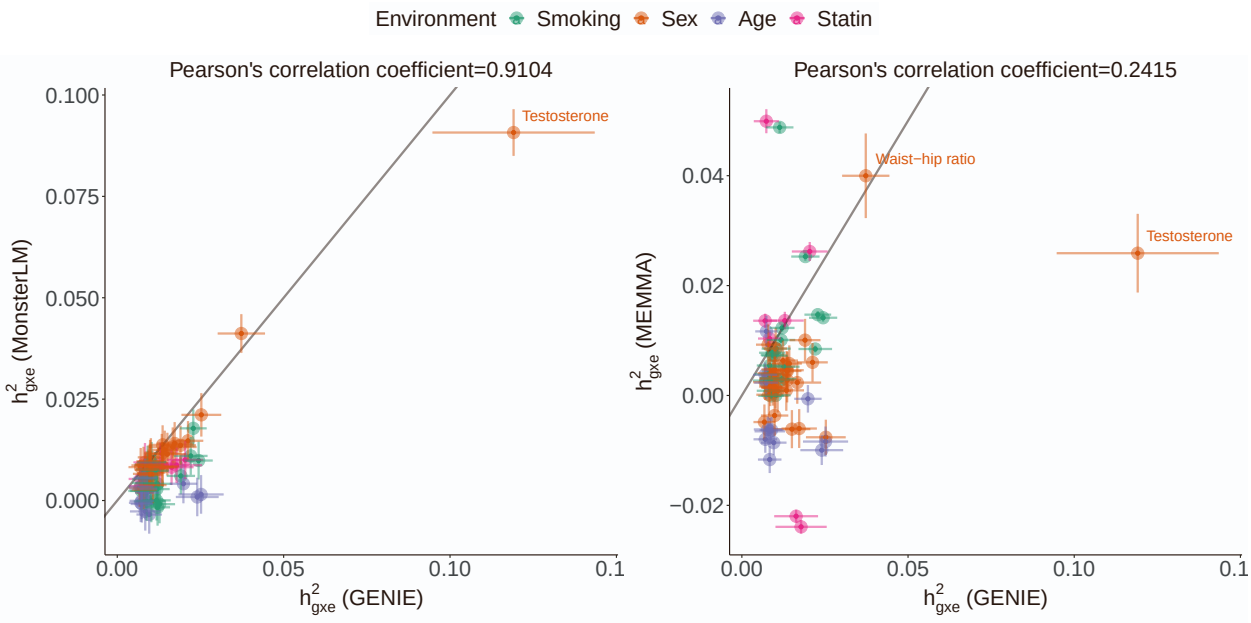

Figure S10: **Estimates of GxE heritability from GENIE, MonsterLM, and MEMMA for 68 trait-environment pairs in the UK Biobank.** We applied GENIE and MEMMA to a cohort of $N = 291,273$ unrelated white British individuals genotyped at $M = 454,207$ SNPs. Subsequent to required genotype QC procedures, which involved excluding SNPs with LD $r^2 > 0.9$ and SNPs with MAF < 0.05, a total of $223,584$ SNPs for $M = 291,273$ individuals were retained for estimating GxE heritability using MonsterLM. We included age, sex, age^2, age$\times$sex, age$^2\times$ sex, and the top genetic PCs as covariates for each trait-E pair in all methods. The dots and error bars represent the point estimates and standard errors, and the diagonal line denotes the $y = x$ curve for each plot. We labeled the traits' names for the most extreme outliers (Cook's distance ten times larger than the mean).

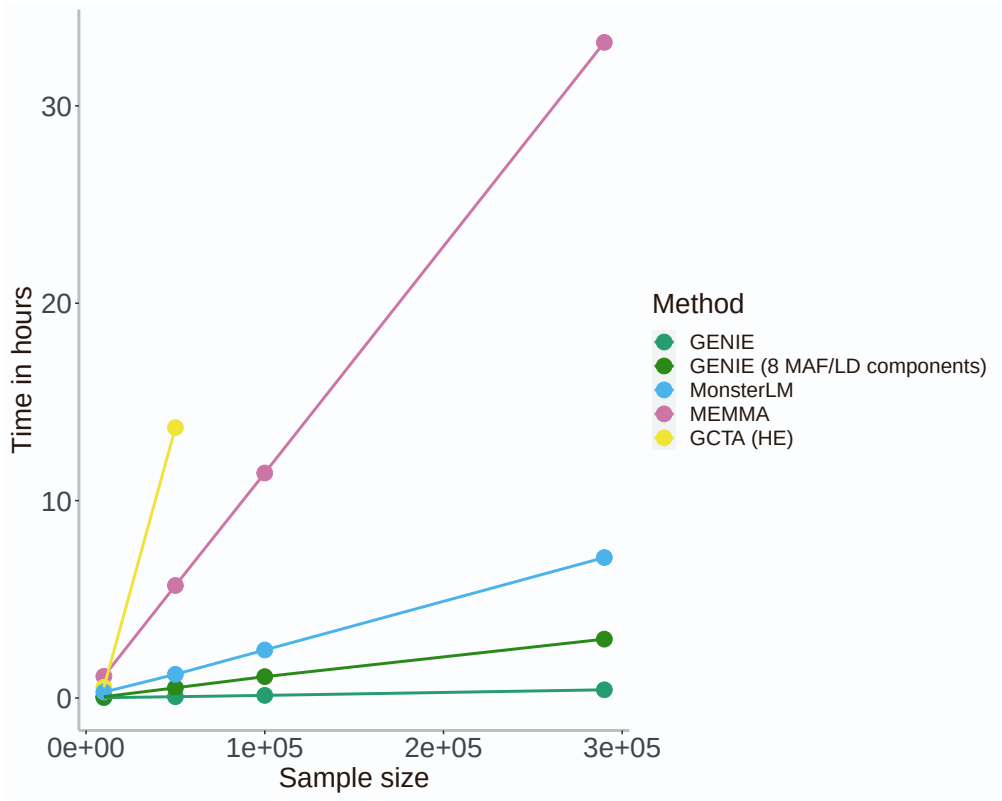

Figure S11: **Comparison of runtimes.** We evaluated the runtime of GENIE, MonsterLM, MEMMA, and GCTA (HE) with increasing sample size N (for a fixed number of SNPs $M = 454,207$ and single environmental variable). All methods were run on an Intel(R) Xeon(R) Gold 6140 CPU 2.30 GHz with 187 GB RAM. Ten random vectors are used by GENIE and MEMMA. For GENIE, runtime measurements were obtained for both the single component and eight MAF/LD components (labeled as "GENIE" and "GENIE (8 MAF/LD annotations)", respectively). All other methods fit a single G and GxE variance component. The runtime of GCTA (HE) includes the computation of the GRM matrix. Our comparison used the CPU implementation of MonsterLM, with runtime calculations excluding the preprocessing step for genotype filtering required by MonsterLM.

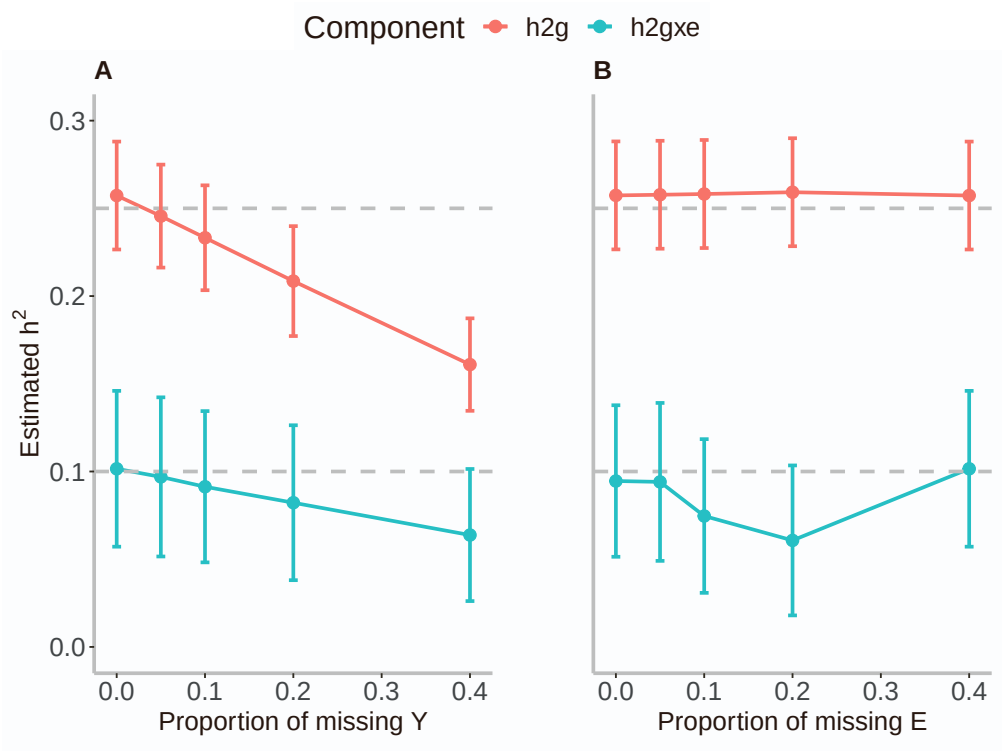

Figure S12: **Impact of mean imputation on the estimation of G and GxE heritability in simulations.** (**A**) Missing phenotypes and (**B**) missing environment variables were imputed by mean imputation. We simulated phenotypes with GxE and NxE effects with variance explained by the G, GxE, and NxE components to be 0.25, 0.1, and 0.1, respectively. We varied the proportion of missing phenotypes across 0, 0.05, 0.1, 0.2, and 0.4 over 100 replicates of simulated phenotypes of $N = 40,000$ individuals genotyped at $454,207$ SNPs. The randomly masked phenotypes were replaced with the mean of those unmasked values. Statin usage was used as the environmental variable in the experiment. To test the performance of the missing environment, we randomly masked and mean-imputed E in a similar manner while keeping the phenotypes as all observed. The x-axis shows the proportion of missing and mean-imputed Y or E, and the y-axis represents the estimated heritability by GENIE. The dots and error bars indicate the mean of estimated G and GxE heritability and 95% CI.

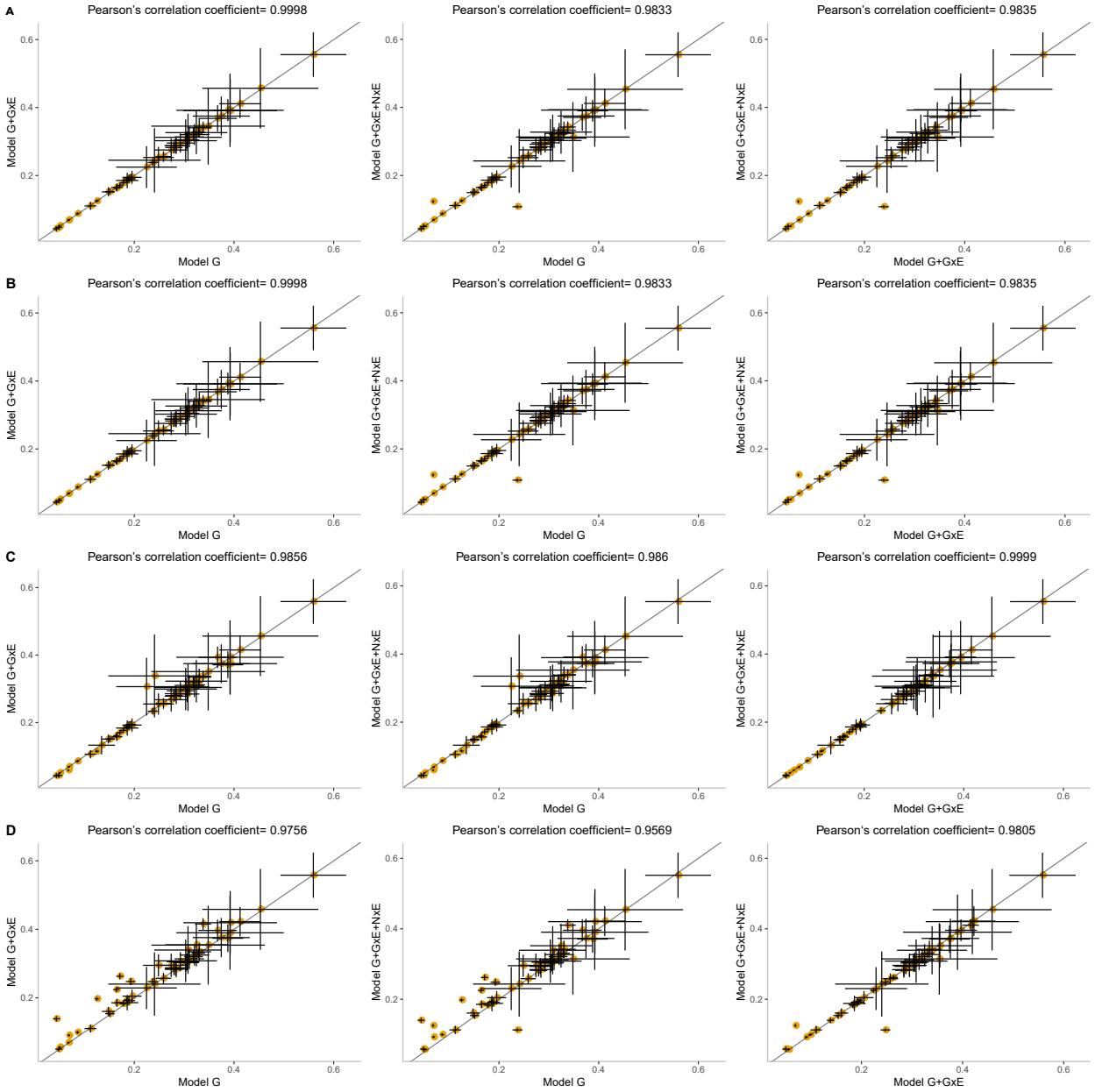

Figure S13: **Estimated additive heritability from three different models.** In this figure, G, GxE, and NxE refer to additive, gene-by-environment, and noise-by-environment components, respectively. Every model is named by a set of variance components fitted jointly under that model. Estimates of additive components under the three models where the environmental variable is (**A**) smoking status, (**B**) sex, (**C**) statin usage, and (**D**) age. The estimates of additive heritability obtained by GENIE are consistent under these three models across environmental variables.

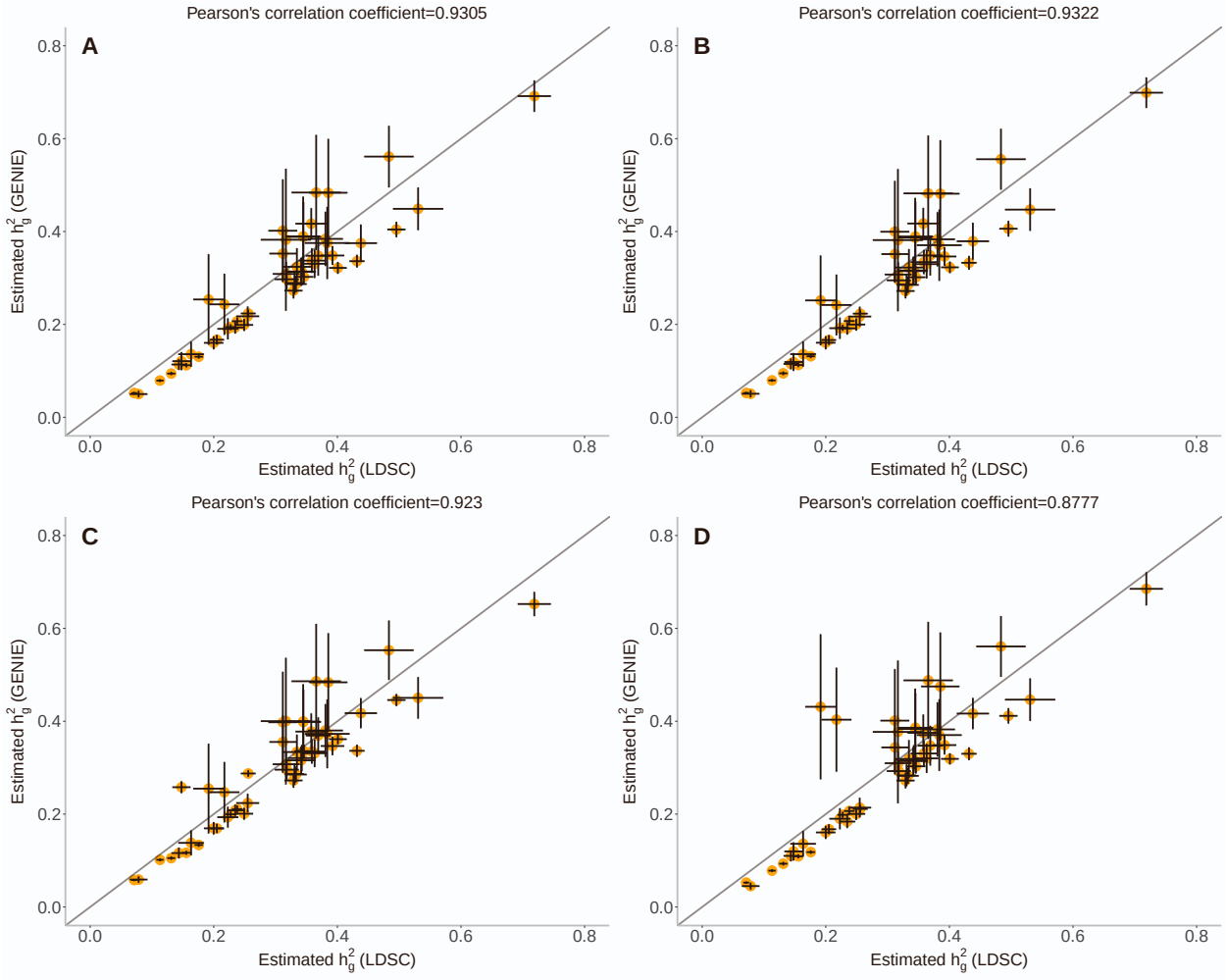

Figure S14: **Comparison of estimated additive heritability with LDSC.** Estimates of additive h_g^2 by LDSC compared with those of GENIE with environment exposures (**A**) smoking status, (**B**) sex, (**C**) age, and (**D**) statin. Both methods estimated additive h_g^2 on $N = 291,273$ individuals and $M = 454,207$ array SNPs and 50 quantitative traits from the UK Biobank. We fitted a single component in LDSC regression to enable a direct comparison with GENIE (which was also applied with a single component in this case). The dots and error bars denote the point estimates and corresponding SEs. GENIE showed consistent and highly correlated estimates compared with LDSC estimates across four environment variables.

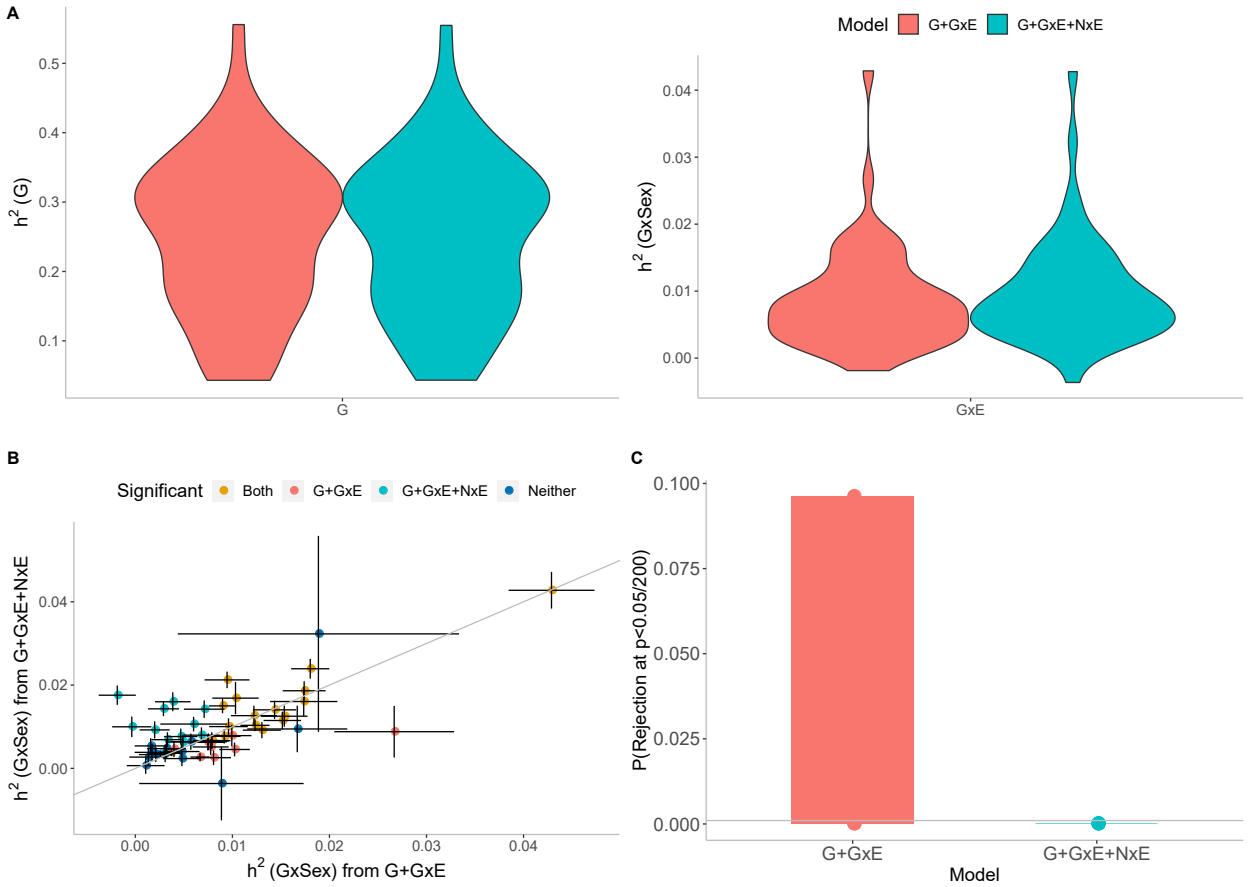

Figure S15: **Effect of Noise heterogeneity (NxE) on estimates of heritability associated with GxSex across 50 quantitative phenotypes in UKB**. Model G+GxE refers to a model with additive and gene-by-environment interaction components. Model G+GxE+NxE refers to a model with additive, gene-by-environment interaction, and noise heterogeneity (noise-by-environment interaction) components. (**A**) We ran GENIE under G+GxE and G+GxE+NxE models to assess the effect of fitting an NxE component on the GxE and additive heritability estimates. (**B**) Comparison of GxE heritability estimates obtained from GENIE under a G+GxE+NxE model (x-axis) to a G+GxE model (y-axis). Black error bars mark $\pm$ standard errors centered on the estimated GxE heritability. The color of the dots indicates whether estimates of GxE heritability are significant under each model. (**C**) We performed permutation analyses by randomly shuffling the genotypes while preserving the trait-E relationship and applied GENIE in each setting under G+GxE and G+GxE+NxE models. We report the fraction of rejections P(p-value of a test of the null hypothesis of zero GxE heritability $< \frac{0.05}{200}$ that accounts for the number of phenotypes tested) over 50 UKB phenotypes.

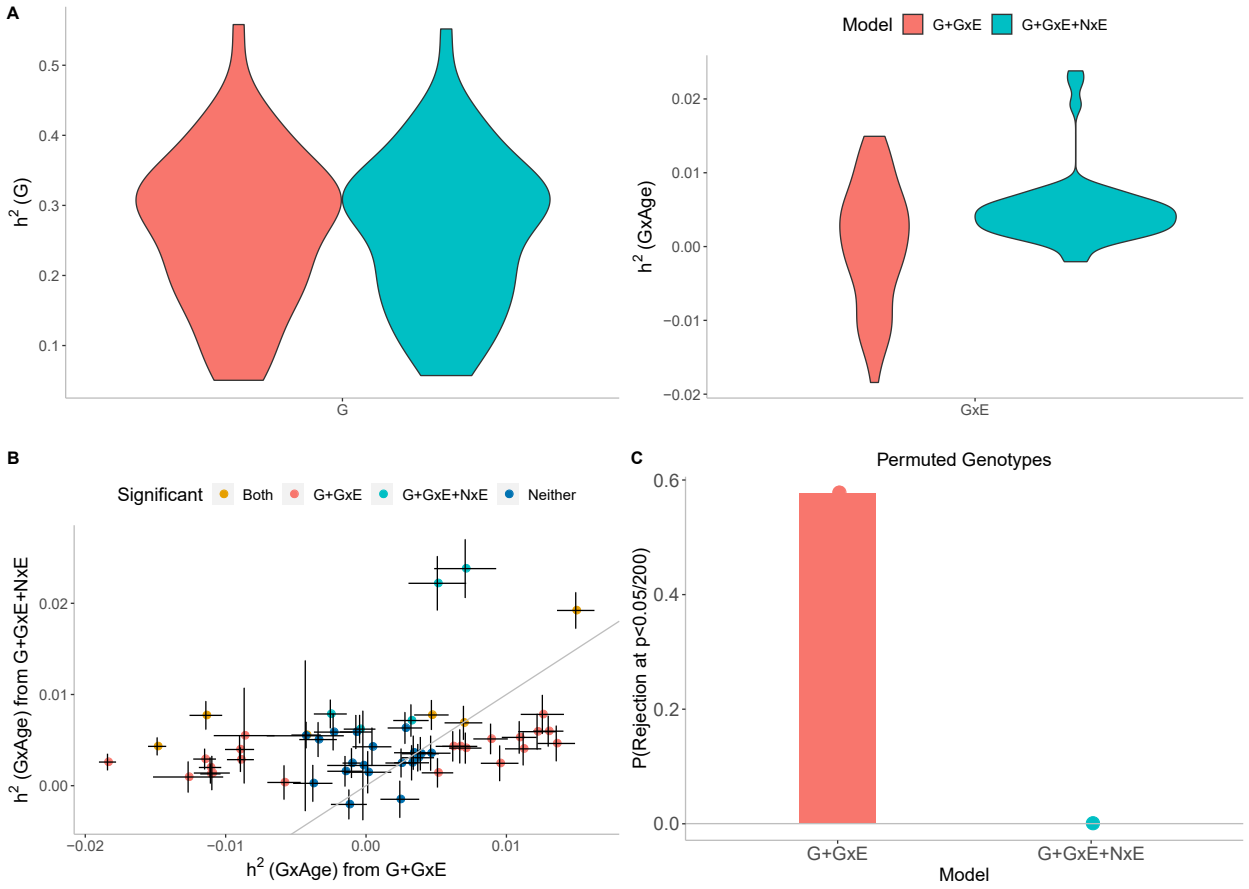

Figure S16: **Effect of Noise heterogeneity (NxE) on estimates of heritability associated with GxAge across 50 quantitative phenotypes in UKB**. Model G+GxE refers to a model with additive and gene-by-environment interaction components. Model G+GxE+NxE refers to a model with additive, gene-by-environment interaction, and noise heterogeneity (noise-by-environment interaction) components. (**A**) We ran GENIE under G+GxE and G+GxE+NxE models to assess the effect of fitting an NxE component on the GxE and additive heritability estimates. (**B**) Comparison of GxE heritability estimates obtained from GENIE under a G+GxE+NxE model (x-axis) to a G+GxE model (y-axis). Black error bars mark ± standard errors centered on the estimated GxE heritability. The color of the dots indicates whether estimates of GxE heritability are significant under each model. (**C**) We performed permutation analyses by randomly shuffling the genotypes while preserving the trait-E relationship and applied GENIE in each setting under G+GxE and G+GxE+NxE models. We report the fraction of rejections P(p-value of a test of the null hypothesis of zero GxE heritability $< \frac{0.05}{200}$ that accounts for the number of phenotypes tested) over 50 UKB phenotypes.

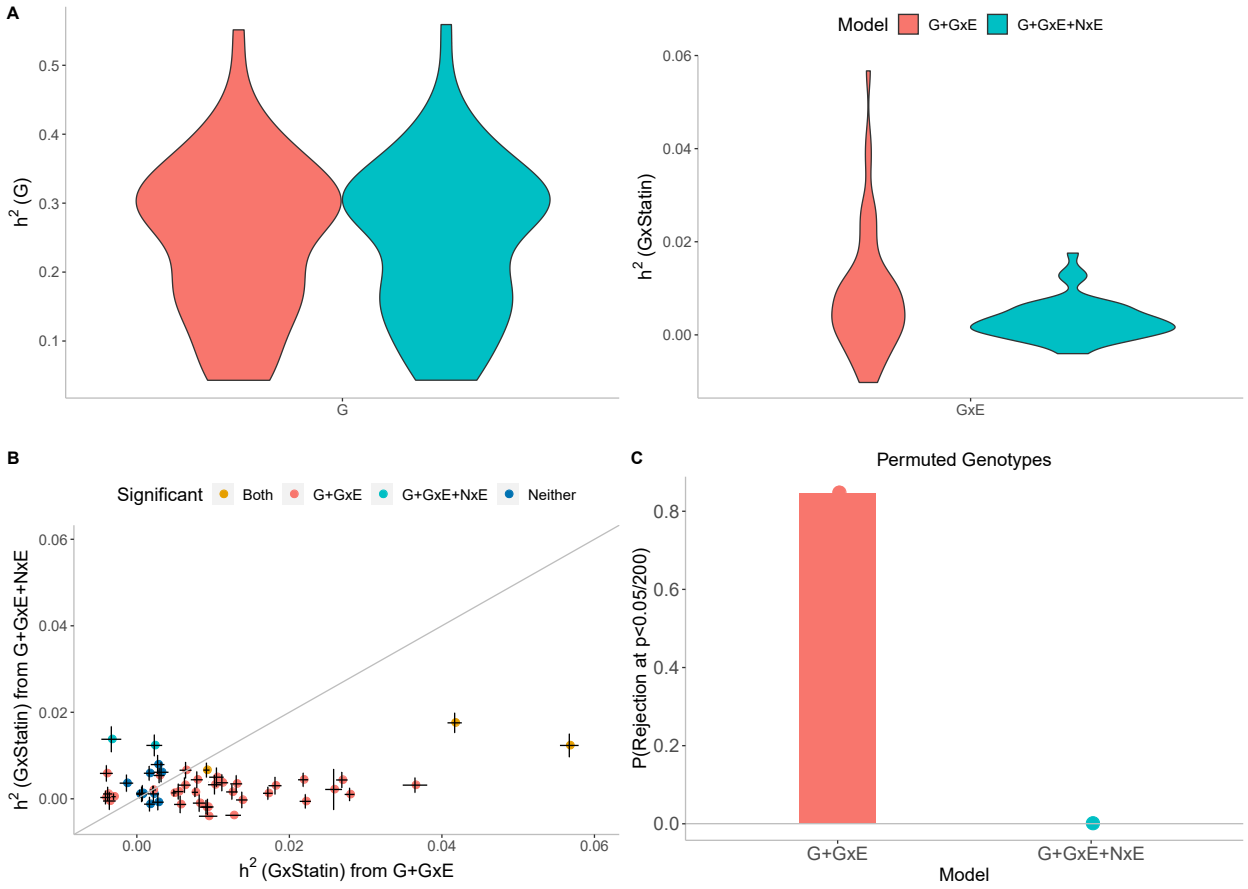

Figure S17: **Effect of Noise heterogeneity (NxE) on estimates of heritability associated with GxStatin across 50 quantitative phenotypes in UKB**. Model G+GxE refers to a model with additive and gene-by-environment interaction components. Model G+GxE+NxE refers to a model with additive, gene-by-environment interaction, and noise heterogeneity (noise-by-environment interaction) components. (**A**) We ran GENIE under G+GxE and G+GxE+NxE models to assess the effect of fitting an NxE component on the GxE and additive heritability estimates. (**B**) Comparison of GxE heritability estimates obtained from GENIE under a G+GxE+NxE model (x-axis) to a G+GxE model (y-axis). Black error bars mark ± standard errors centered on the estimated GxE heritability. The color of the dots indicates whether estimates of GxE heritability are significant under each model. (**C**) We performed permutation analyses by randomly shuffling the genotypes while preserving the trait-E relationship and applied GENIE in each setting under G+GxE and G+GxE+NxE models. We report the fraction of rejections P(p-value of a test of the null hypothesis of zero GxE heritability $< \frac{0.05}{200}$ that accounts for the number of phenotypes tested) over 50 UKB phenotypes.

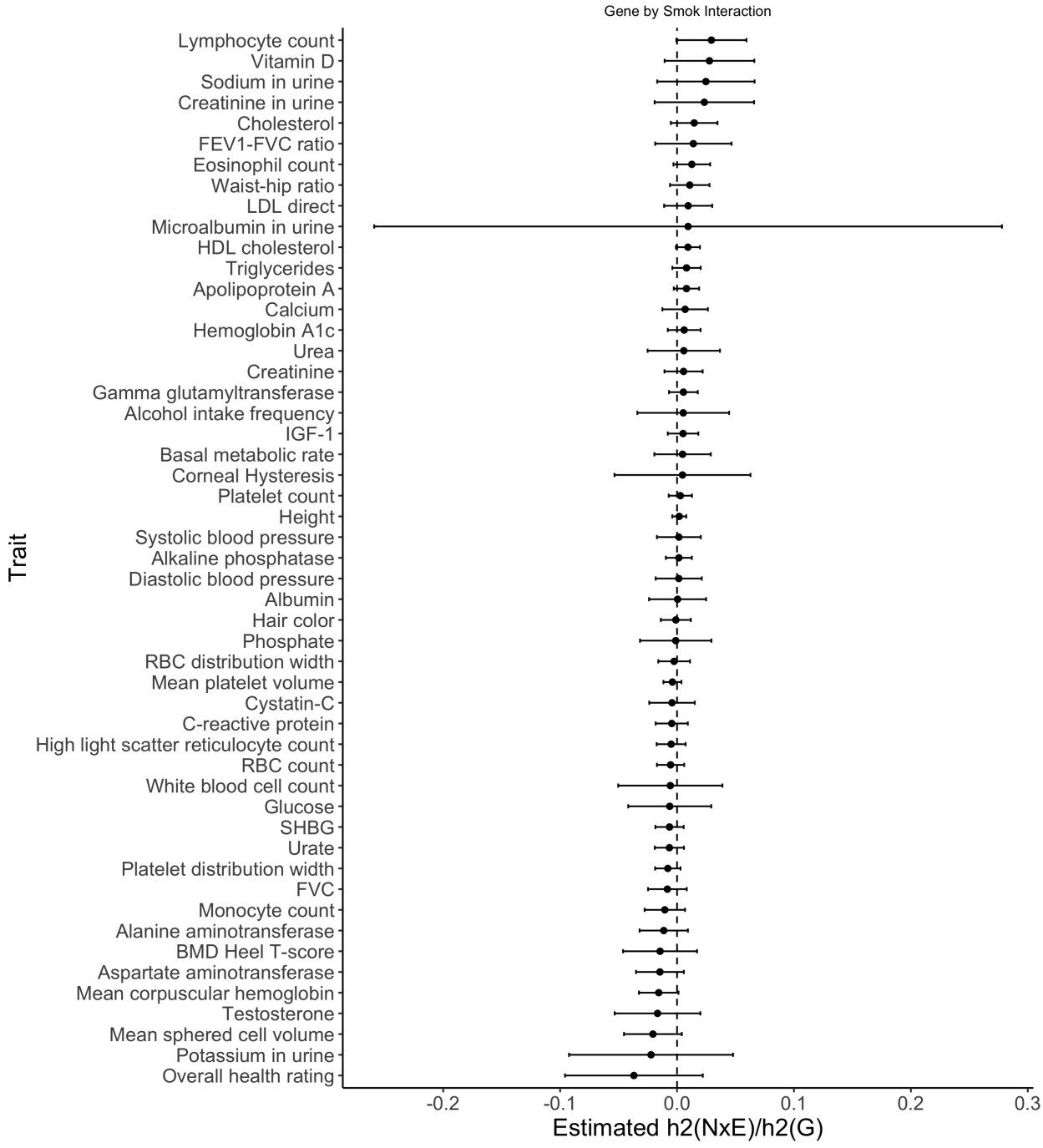

Figure S18: **Estimated ratio of variance attributed to noise heterogeneity over additive heritability for Smoking**. Black error bars mark ± 2 standard errors centered on the estimated ratio.

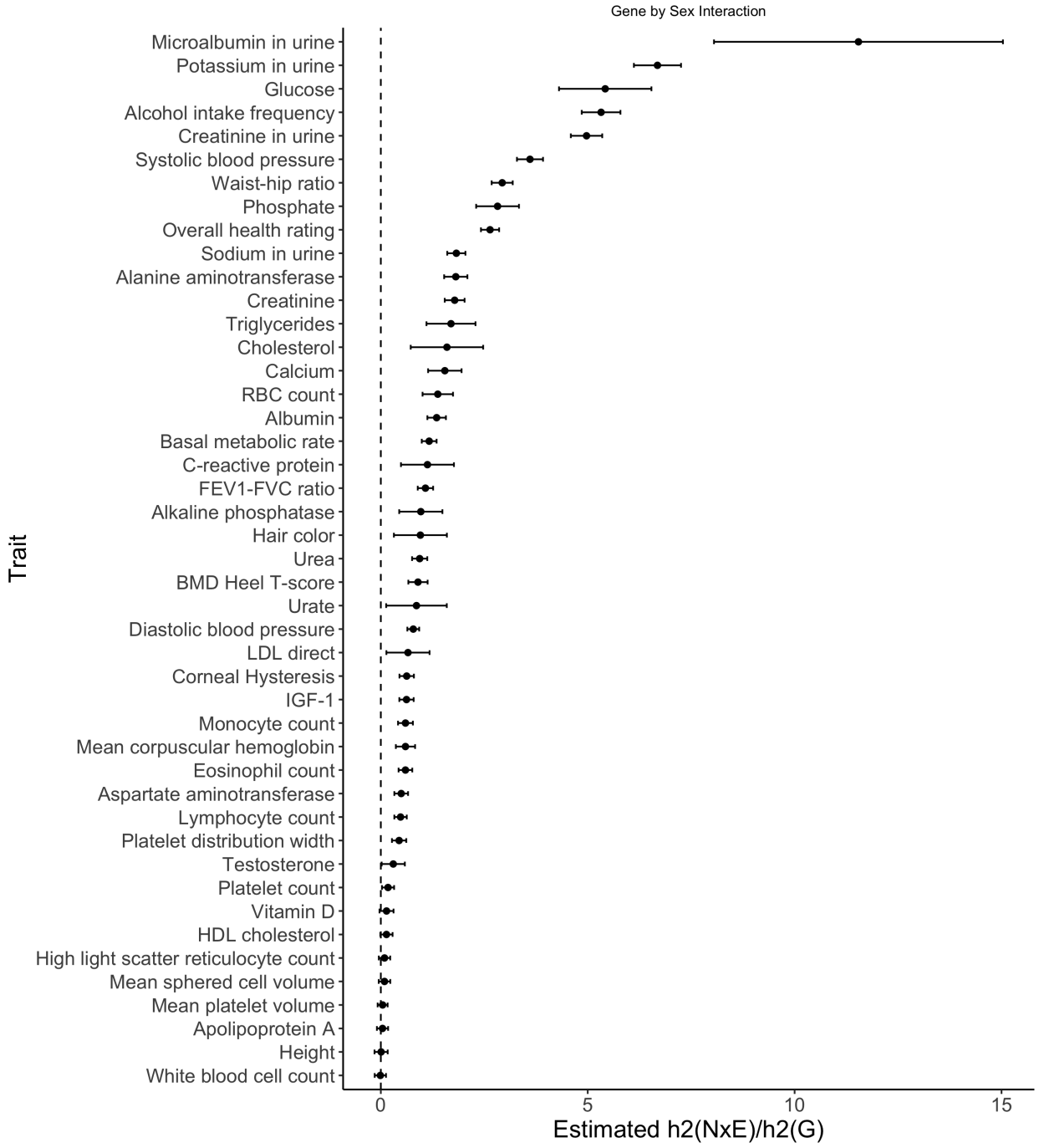

Figure S19: **Estimated ratio of variance attributed to noise heterogeneity over additive heritability for Sex**. Black error bars mark ±2 standard errors centered on the estimated ratio.

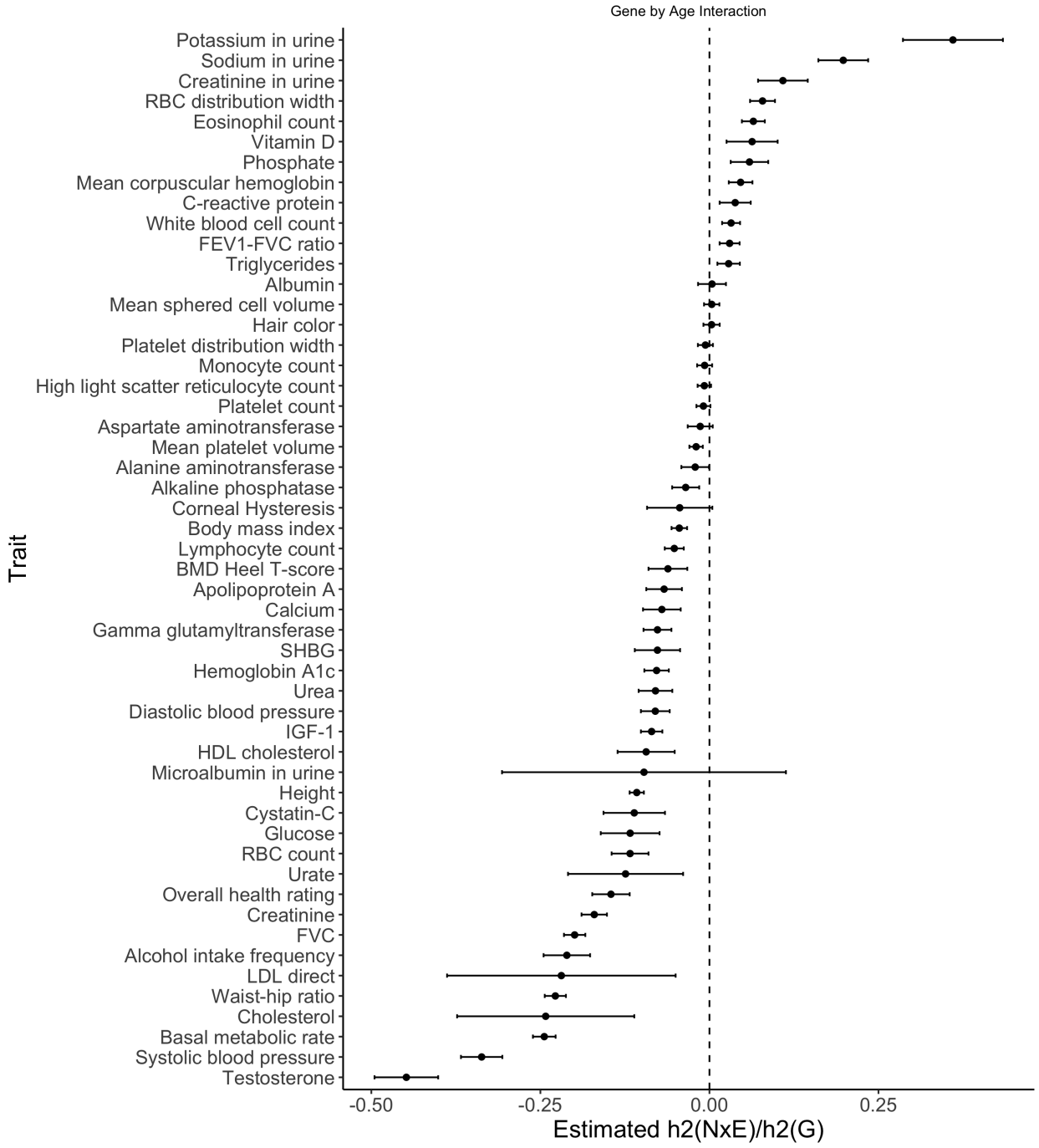

Figure S20: **Estimated ratio of variance attributed to noise heterogeneity over additive heritability for Age**. Black error bars mark ±2 standard errors centered on the estimated ratio.

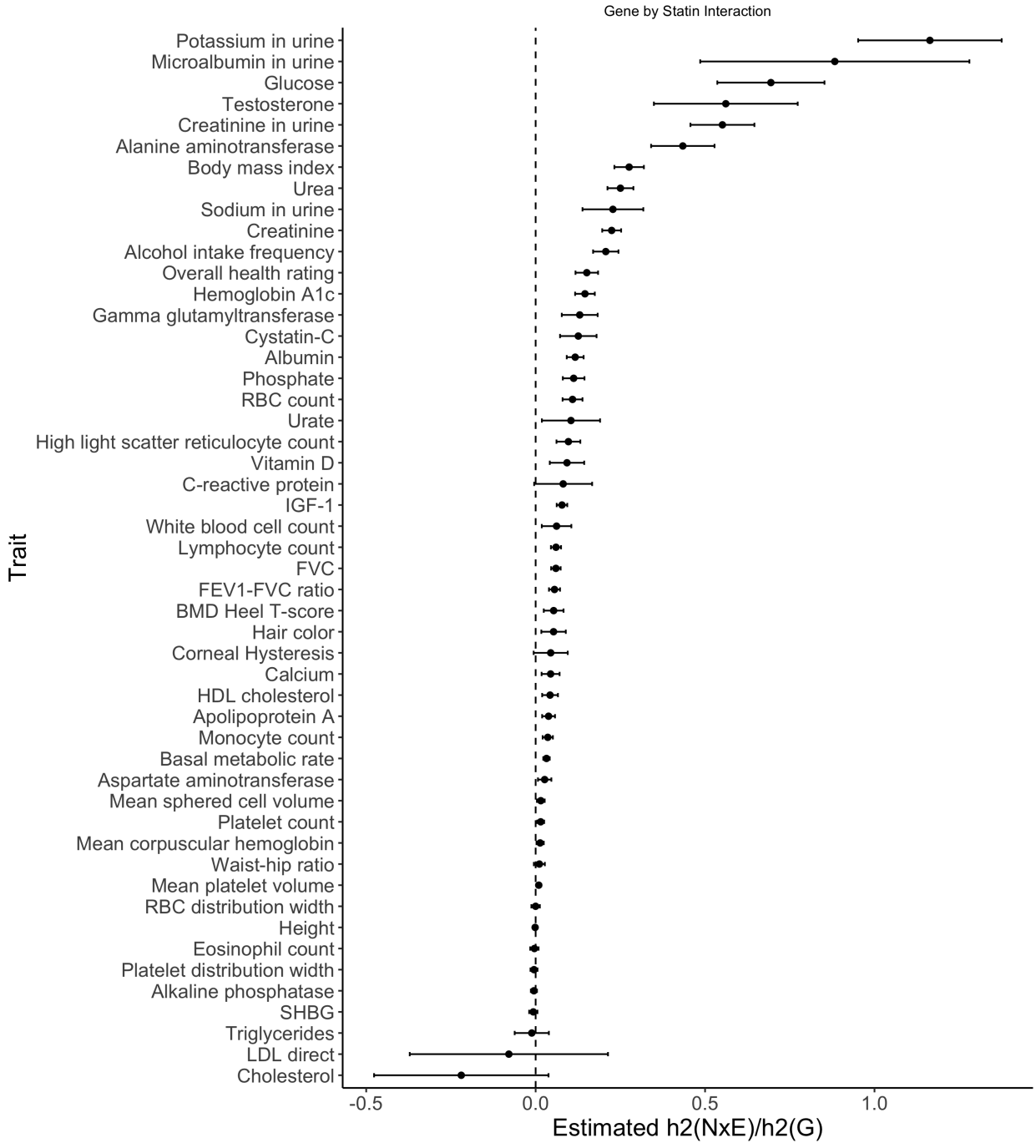

Figure S21: **Estimated ratio of variance attributed to noise heterogeneity over additive heritability for Statin usage**. Black error bars mark ±2 standard errors centered on the estimated ratio.

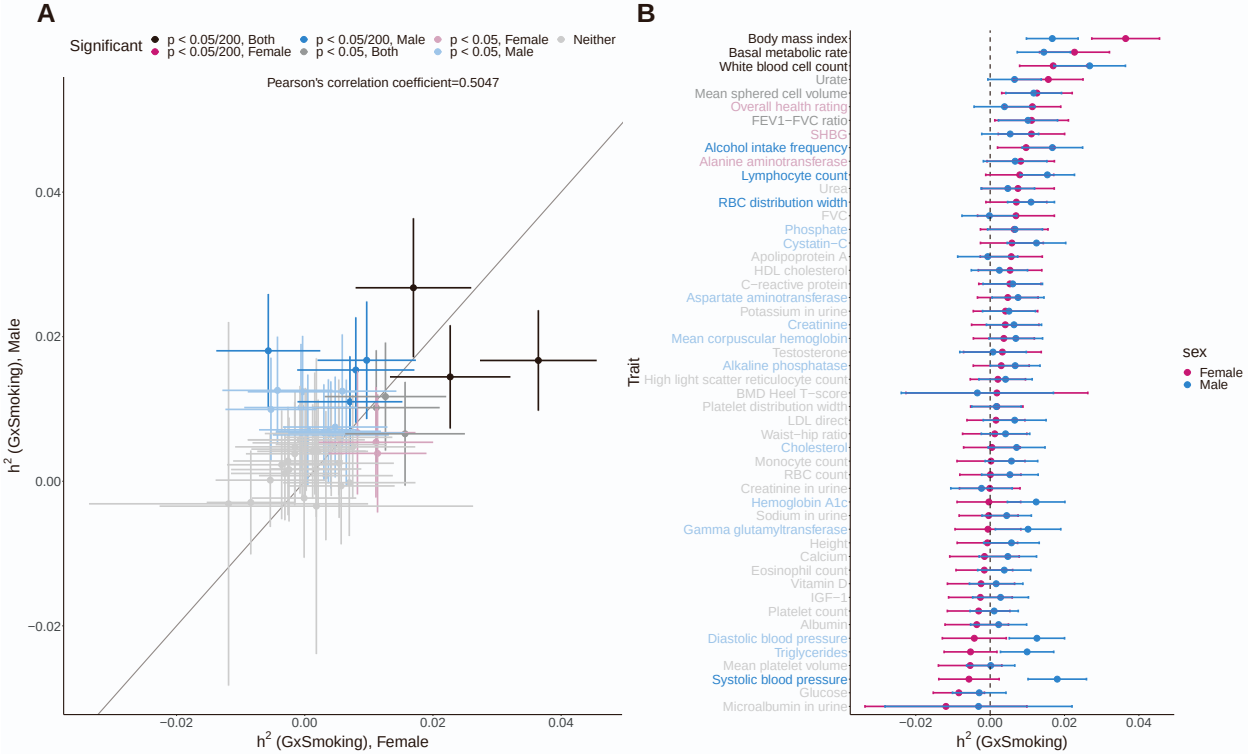

Figure S22: **Estimates of GxSmoking heritability across phenotypes in UK Biobank stratified by sex.** We applied GENIE to $N = 291,273$ unrelated white British individuals and $M = 454,207$ array SNPs, with $155,469$ male and $135,804$ female individuals. The phenotype vectors were masked such that only the measurements of female/male individuals remained. (**A**) The comparison of GxSmoking heritability estimates in female and male individuals across fifty traits. Error bars mark ± 2 standard errors centered on the point estimates. We grouped the results into seven significant levels based on the p-values in males and females. (**B**) Estimates of GxSmoking heritability across fifty traits stratified by sex. The trait labels were colored based on the defined significance levels.

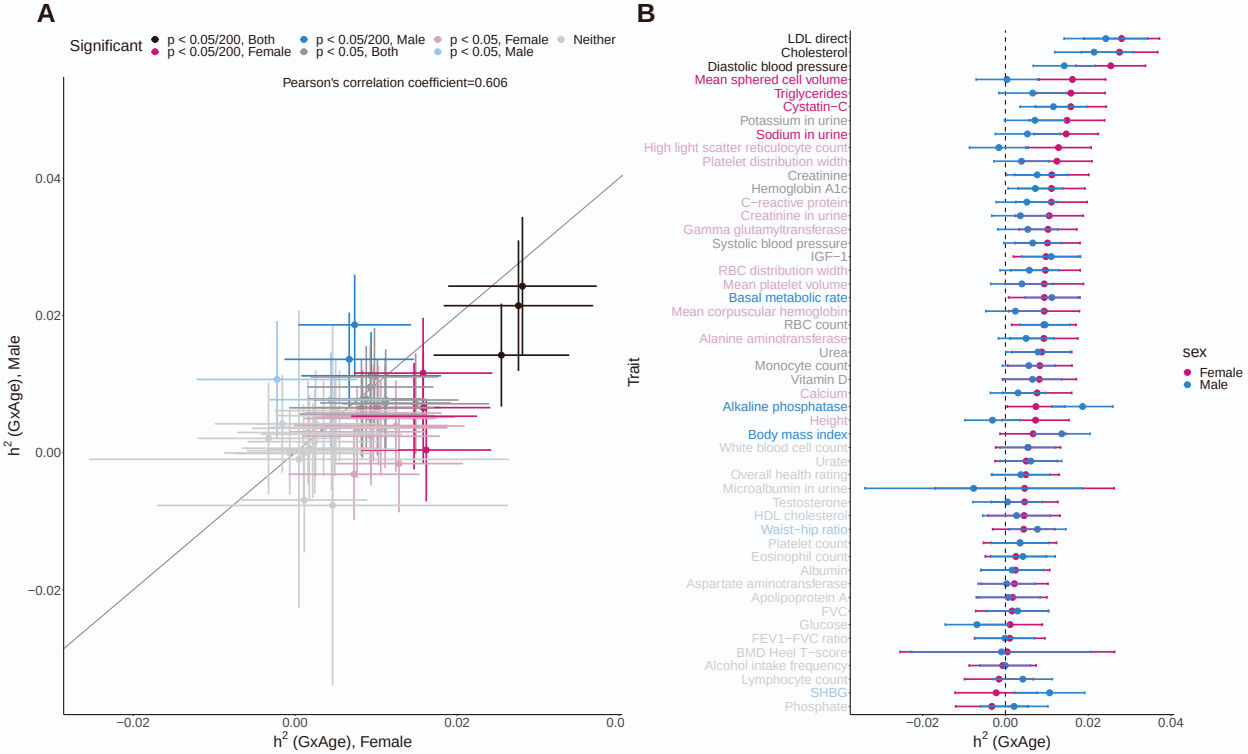

Figure S23: **Estimates of GxAge heritability across phenotypes in UK Biobank stratified by sex.** We applied GENIE to $N = 291,273$ unrelated white British individuals and $M = 454,207$ array SNPs, with $155,469$ male and $135,804$ female individuals. The phenotype vectors were masked such that only the measurements of female/male individuals remained. (**A**) The comparison of GxAge heritability estimates in female and male individuals across fifty traits. Error bars mark ± 2 standard errors centered on the point estimates. We grouped the results into seven significant levels based on the p-values in males and females. (**B**) Estimates of GxAge heritability across fifty traits stratified by sex. The trait labels were colored based on the defined significance levels.

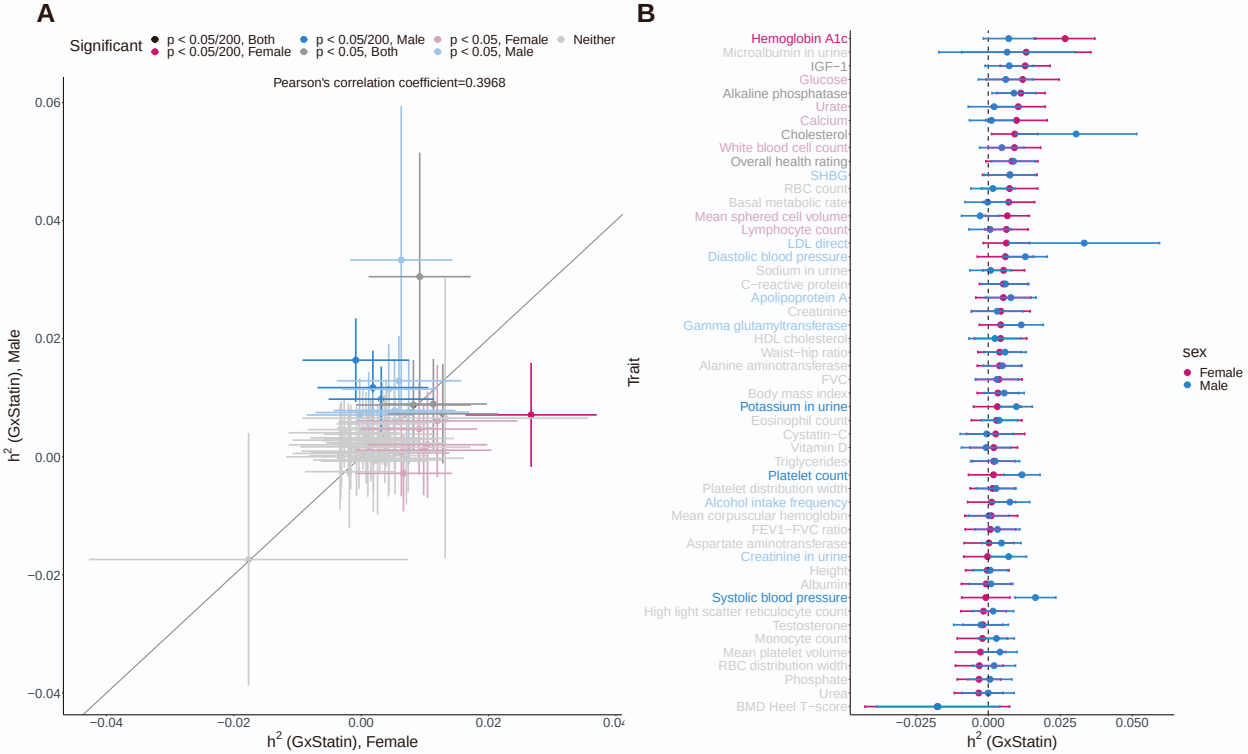

Figure S24: **Estimates of GxStatin heritability across phenotypes in UK Biobank stratified by sex.** We applied GENIE to $N = 291,273$ unrelated white British individuals and $M = 454,207$ array SNPs, with $155,469$ male and $135,804$ female individuals. The phenotype vectors were masked such that only the measurements of female/male individuals remained. (**A**) The comparison of GxStatin heritability estimates in female and male individuals across fifty traits. Error bars mark ± 2 standard errors centered on the point estimates. We grouped the results into seven significant levels based on the p-values in males and females. (**B**) Estimates of GxStatin heritability across fifty traits stratified by sex. The trait labels were colored based on the defined significance levels.

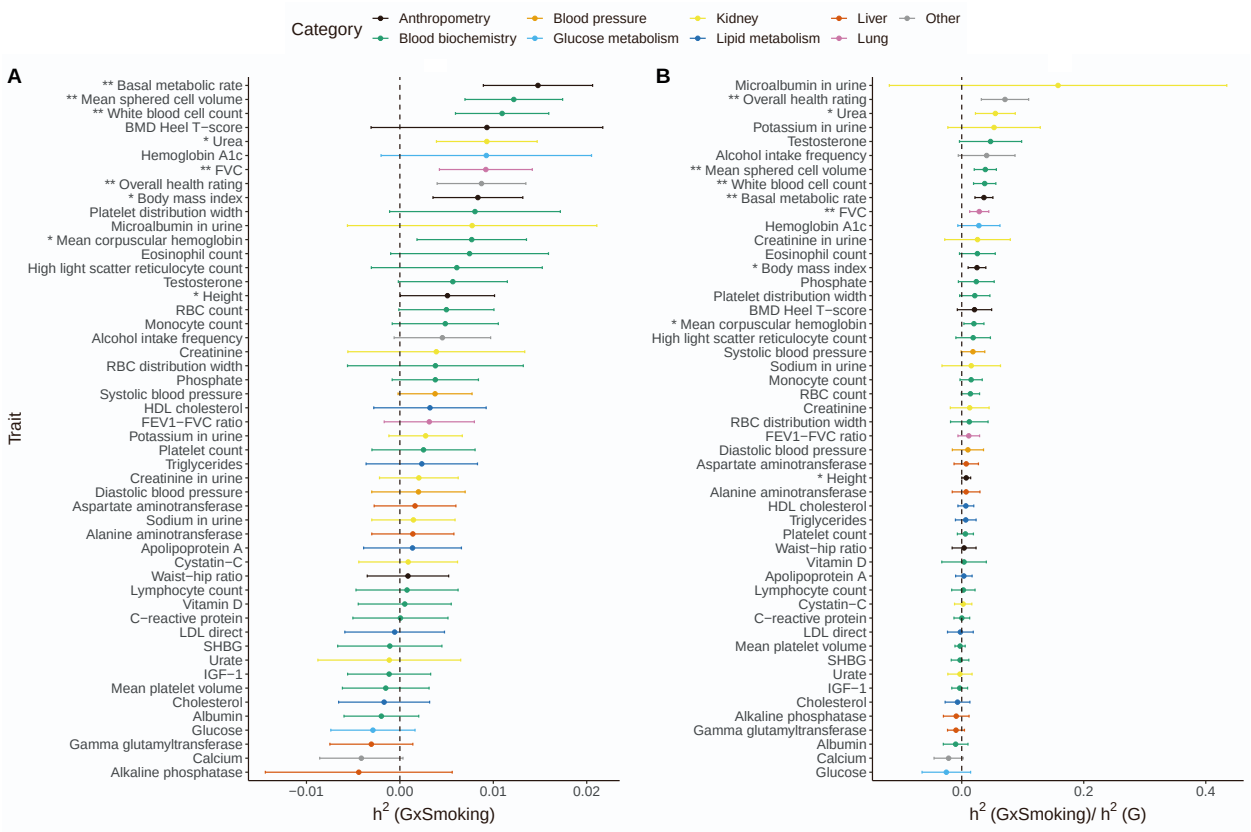

Figure S25: **GxSmoking across phenotypes in UK Biobank with the environmental variable coded as binary.** (**A**) GxSmoking heritability and (**B**) the ratio of GxSmoking to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Black error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$ respectively.

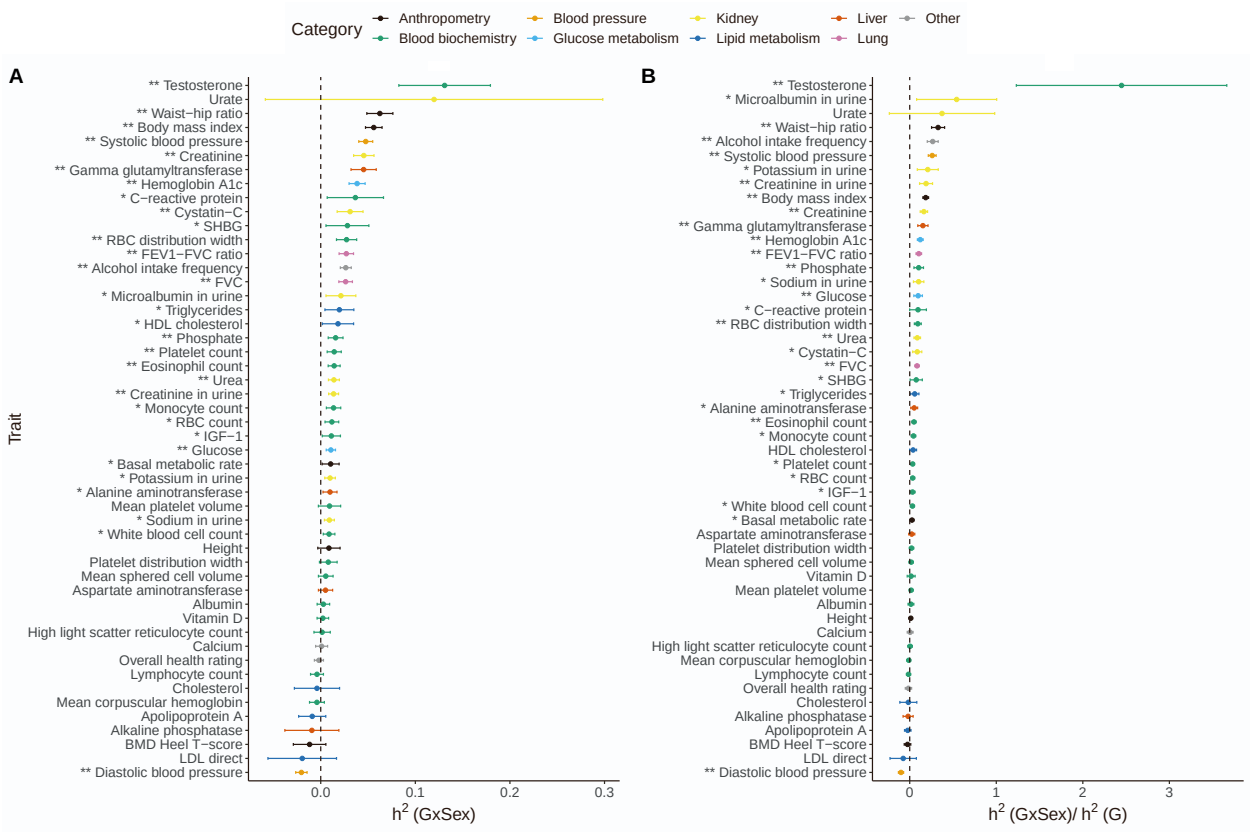

Figure S26: **GxSex across phenotypes in UK Biobank with the environmental variable coded as binary.** (**A**) GxSex heritability and (**B**) the ratio of GxSex to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Black error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/50$ respectively.

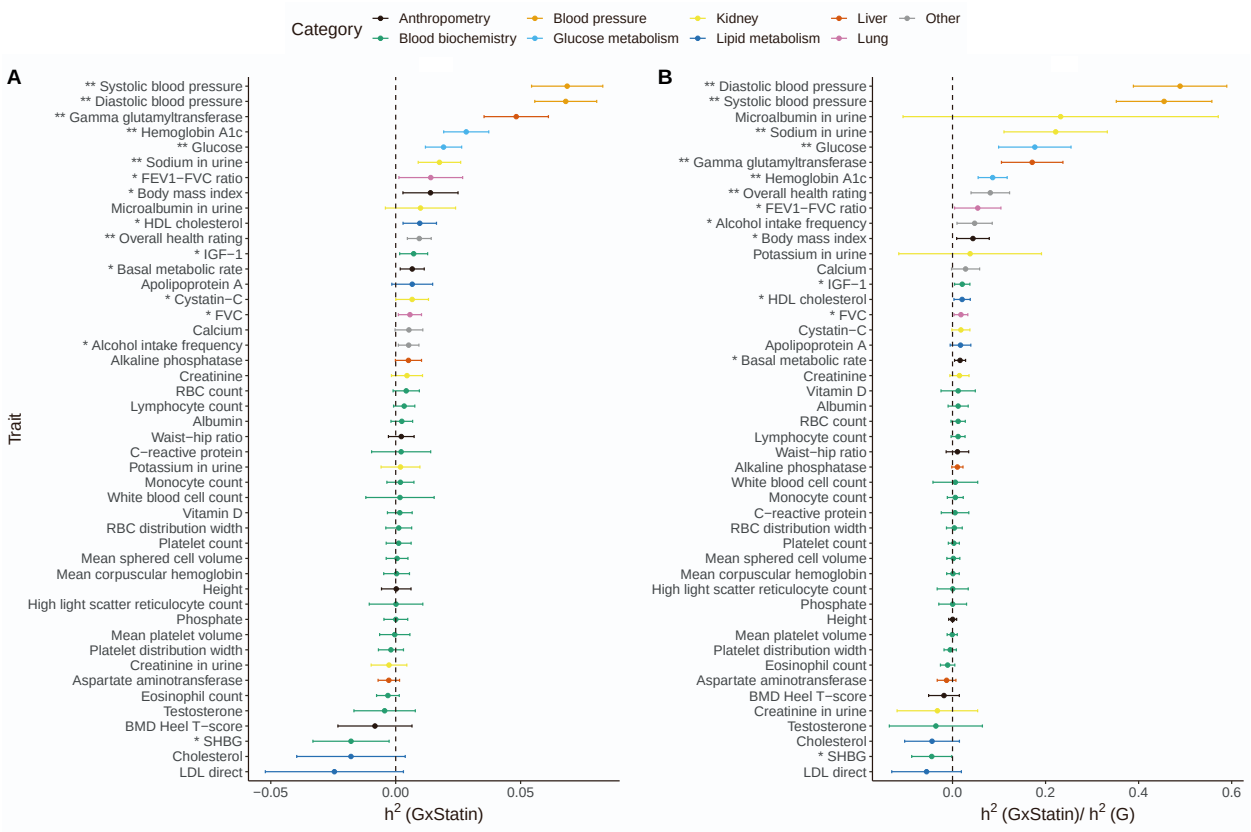

Figure S27: **GxStatin across phenotypes in UK Biobank with the environmental variable coded as binary.** (**A**) GxStatin heritability and (**B**) the ratio of GxStatin to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Black error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/50$, respectively.

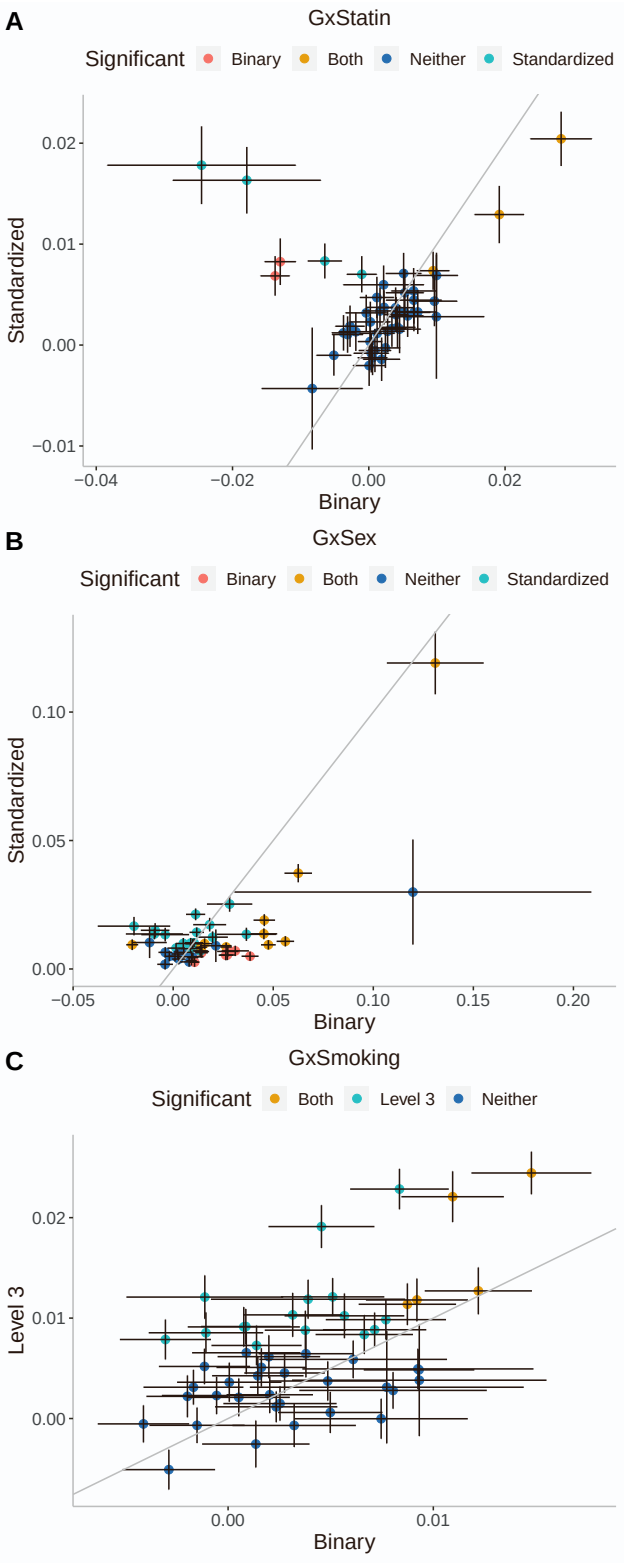

Figure S28: **Comparisons of GxE estimates using standardized or binary encodings of the environment.** We compared (**A**) GxStatin, (**B**) GxSex and (**C**) GxSmoking estimates on N = 291,273 individuals genotyped at 454,207 SNPs for 50 traits using standardized or binary encodings. The traits are significant at $p < 0.05/200$ using standardized encoding and/or binary encoding (or neither) are labeled with different colors. The standardized encodings of binary variables statin usage and sex and ternary variable smoking status (denoted as "level 3") were plotted on the y-axis. The points and error bars represent the point estimates and estimated SEs respectively, and the $y = x$ curve was drawn in gray.

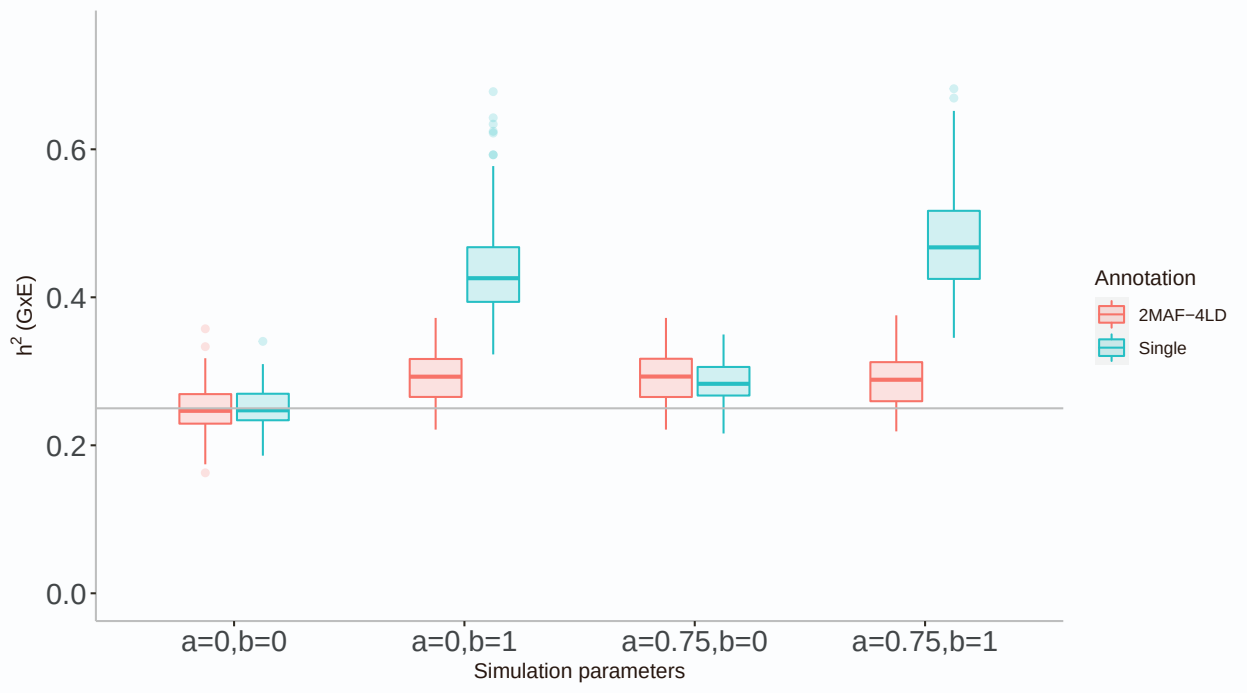

Figure S29: **Effect of MAF-LD partitioning on estimated GxE heritability in simulation.** We assessed the effect of MAF-LD partitioning on estimates of h^2_{gxe} in simulations. We ran GENIE in two settings: 1) fitting a model with a single additive and a single GxE variance component, and 2) fitting a model with eight additive and eight GxE components defined based on four LD annotations (quartiles of LD scores) and two MAF annotations. we simulated phenotypes with GxE effects and G effects from a subset of $N = 40,000$ individuals genotyped at array SNPs $M = 459,792$ by varying the coupling of MAF with effect size (a) and the effect of local LD on effect size (b). The per-SNP variance component $\sigma^2_{g,m} \propto w^b_m[f_m(1-f_m)]^a$ in simulations, where f_m and w_m are the minor allele frequency and LD score of m^{th} SNP respectively (see Note Note S2 for details). Here we have $h^2_g = h^2_{gxe} = 0.25$, $h^2_{nxe} = 0.05$, and all SNPs are causal for both additive and GxE effects. Each box plot represents estimates from 100 simulations.

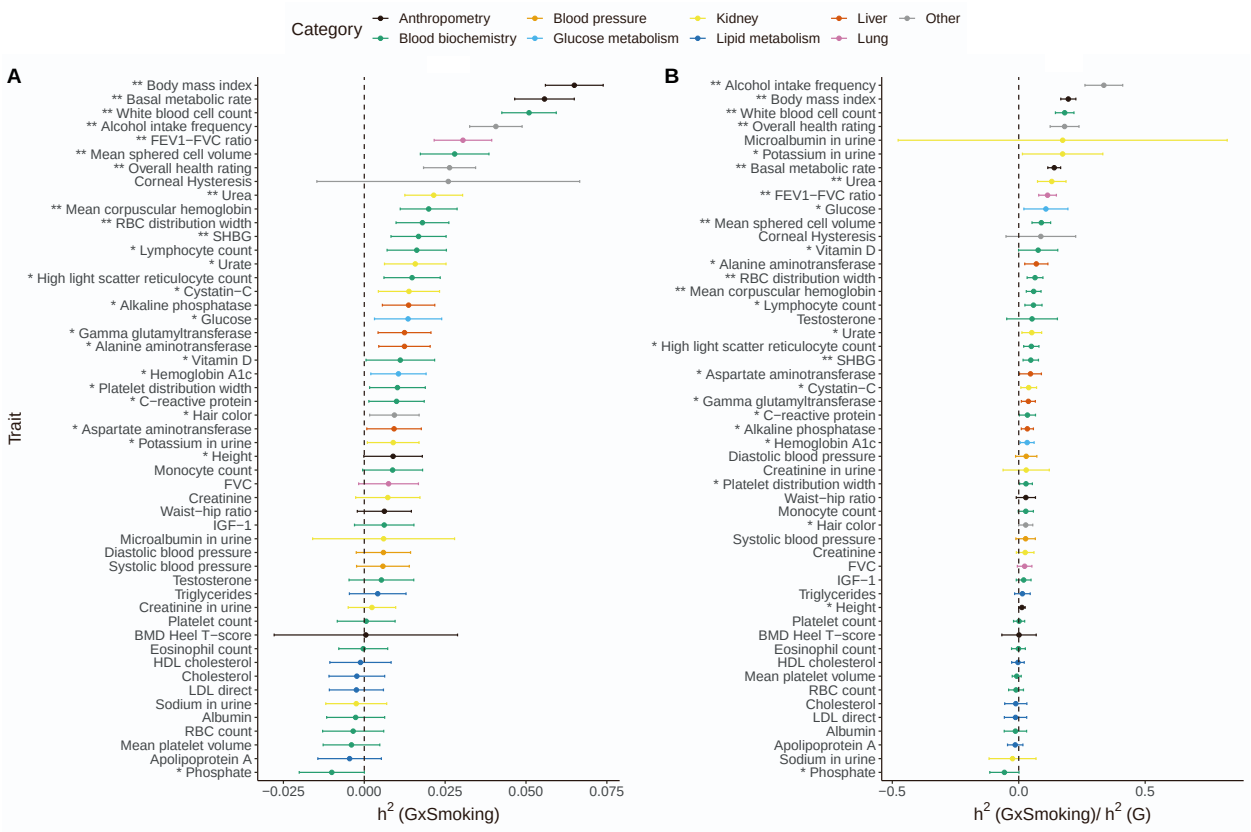

Figure S30: **GxSmoking across phenotypes from imputed SNPs in UK Biobank by MAF-LD partitioning.** (**A**) GxSmoking heritability and (**B**) the ratio of GxSmoking to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Black error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

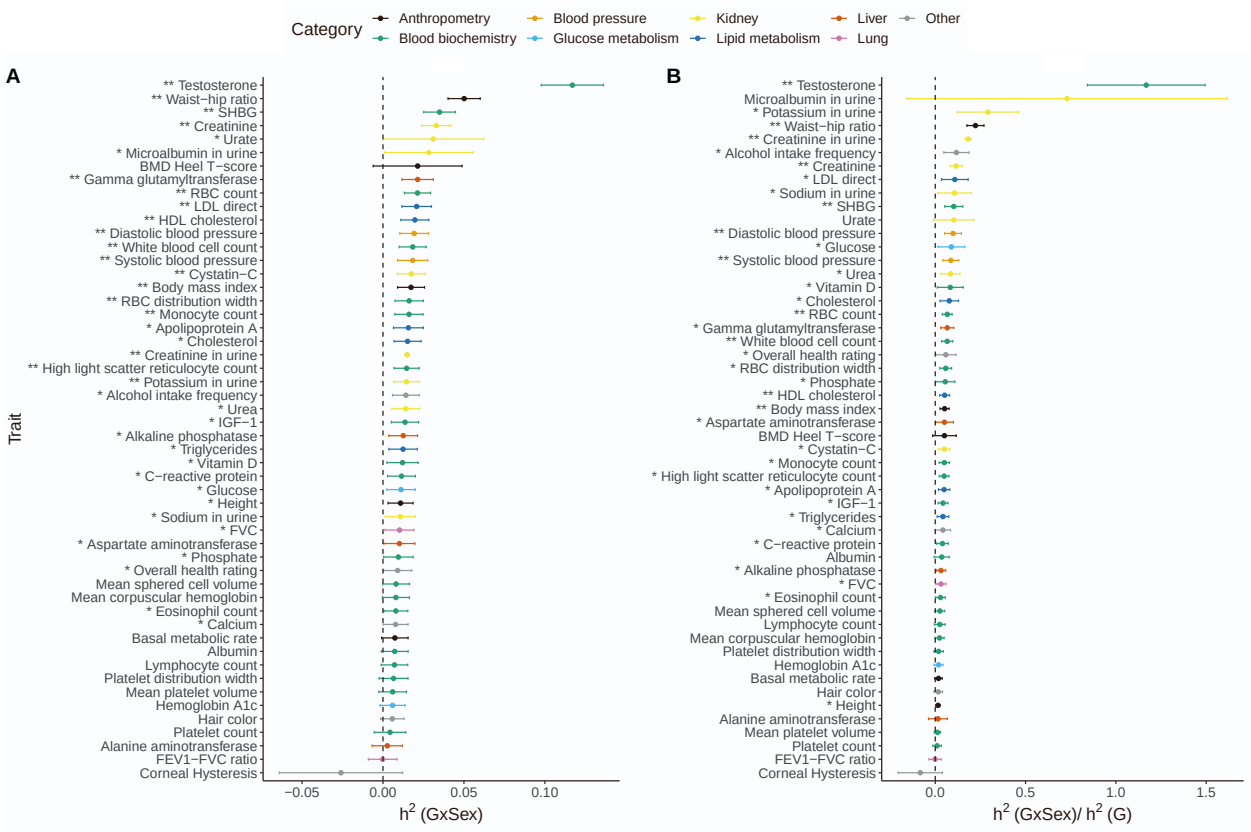

Figure S31: **GxSex across phenotypes from imputed SNPs in UK Biobank by MAF-LD partitioning.** (**A**) GxSex heritability and (**B**) the ratio of GxSex to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

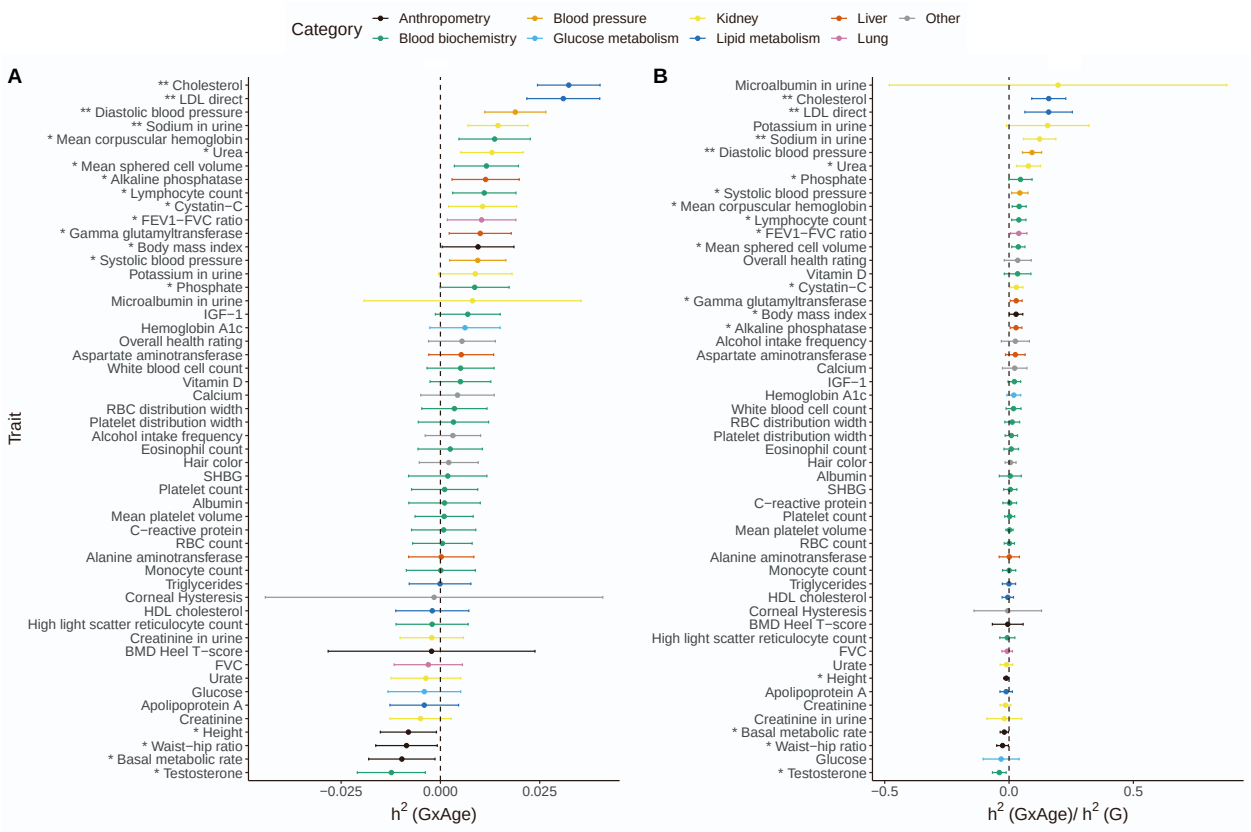

Figure S32: **GxAge across phenotypes from imputed SNPs in UK Biobank by MAF-LD partitioning.** (**A**) GxAge heritability and (**B**) the ratio of GxAge to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Error bars mark ±2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

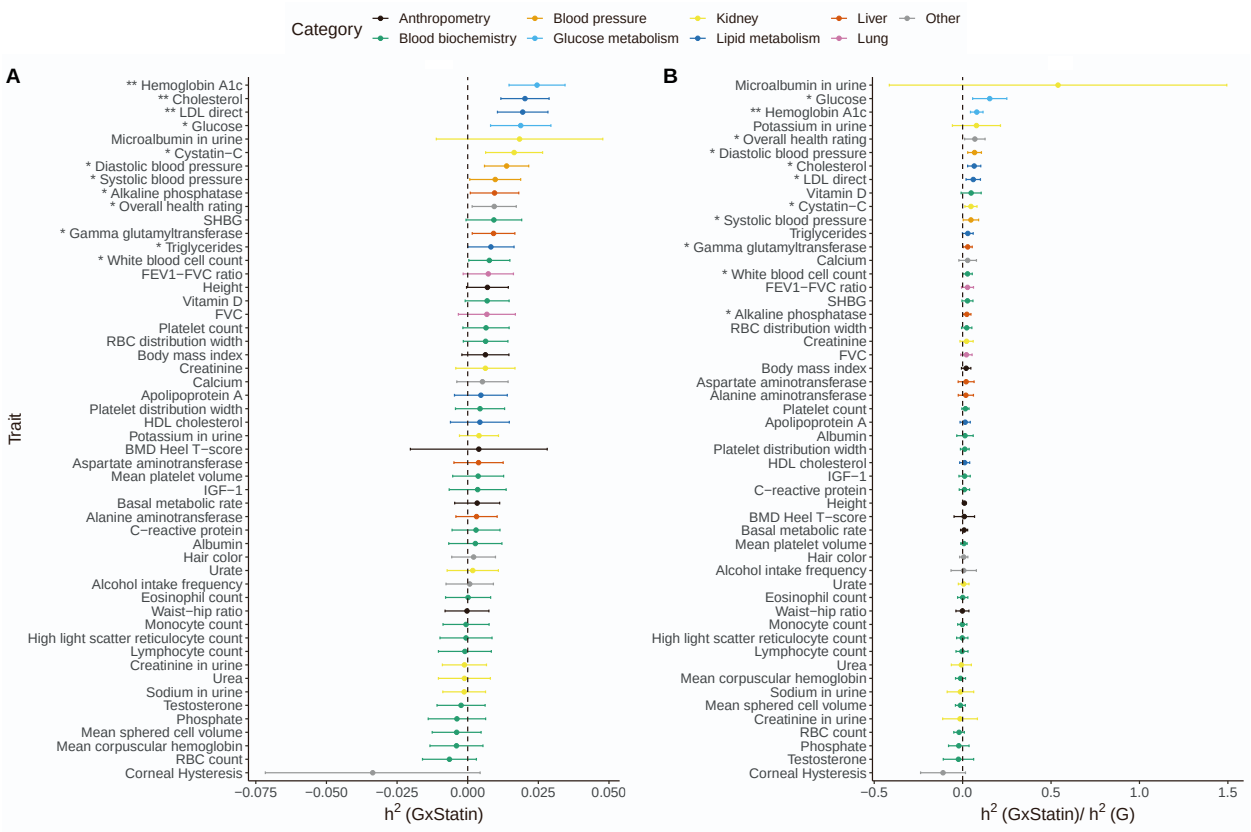

Figure S33: **GxStatin across phenotypes from imputed SNPs in UK Biobank by MAF-LD partitioning.** (**A**) GxStatin heritability and (**B**) the ratio of GxStatin to additive heritability. Our model includes the environmental variable as a fixed effect and accounts for noise heterogeneity. Error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$, respectively.

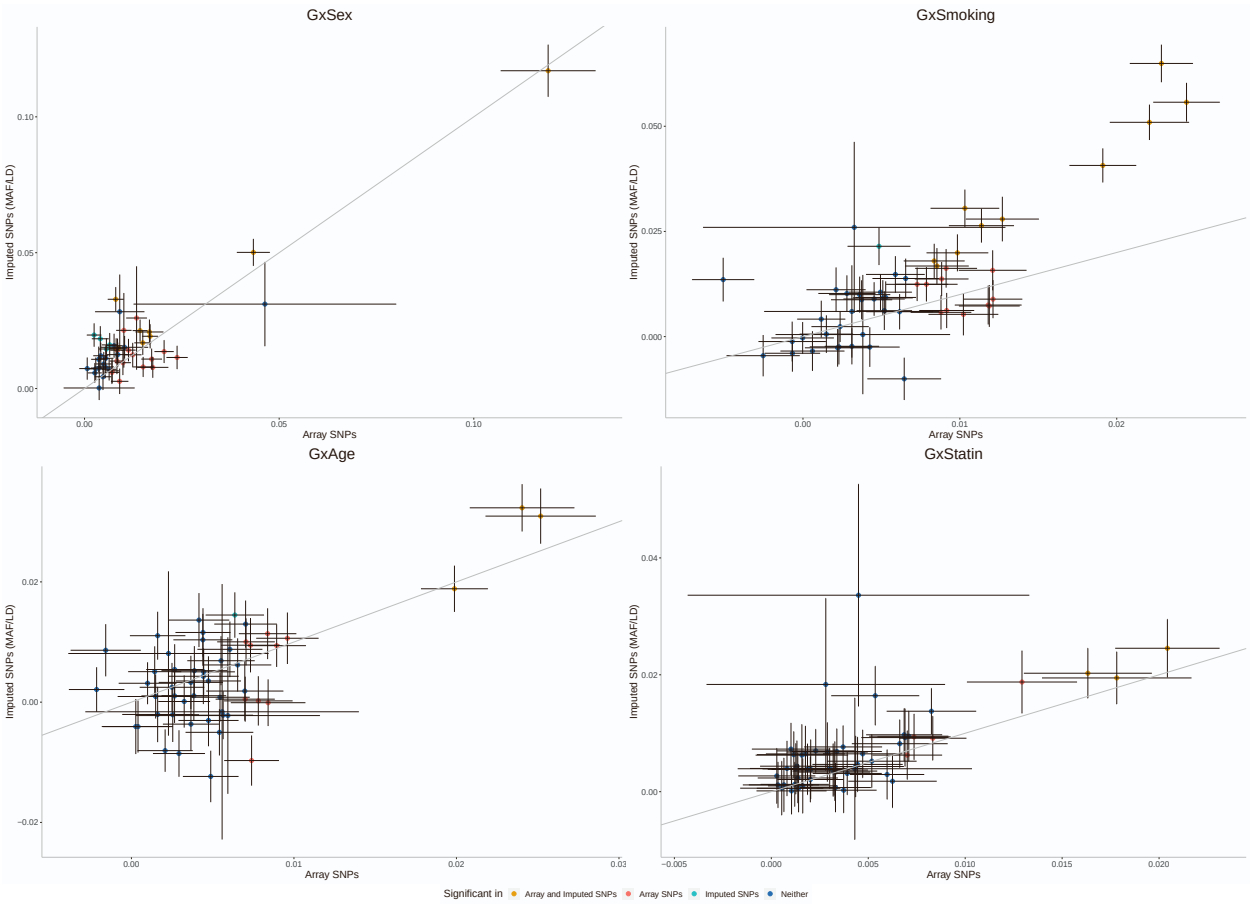

Figure S34: **Comparing GxSex, GxSmoking, GxAge, and GxStatin estimates from imputed SNPs (MAF$\geq$ 0.1%) and array SNPs (MAF$\geq$ 1%).** In this analysis, we applied GENIE to imputed SNPs with MAF/LD stratification and array SNPs with a single component. Black error bars mark ± 2 standard errors. The asterisk and double asterisk correspond to the nominal $p < 0.05$ and $p < 0.05/200$ respectively. The color of the dots indicates whether estimates of GxE heritability are significant under each model.

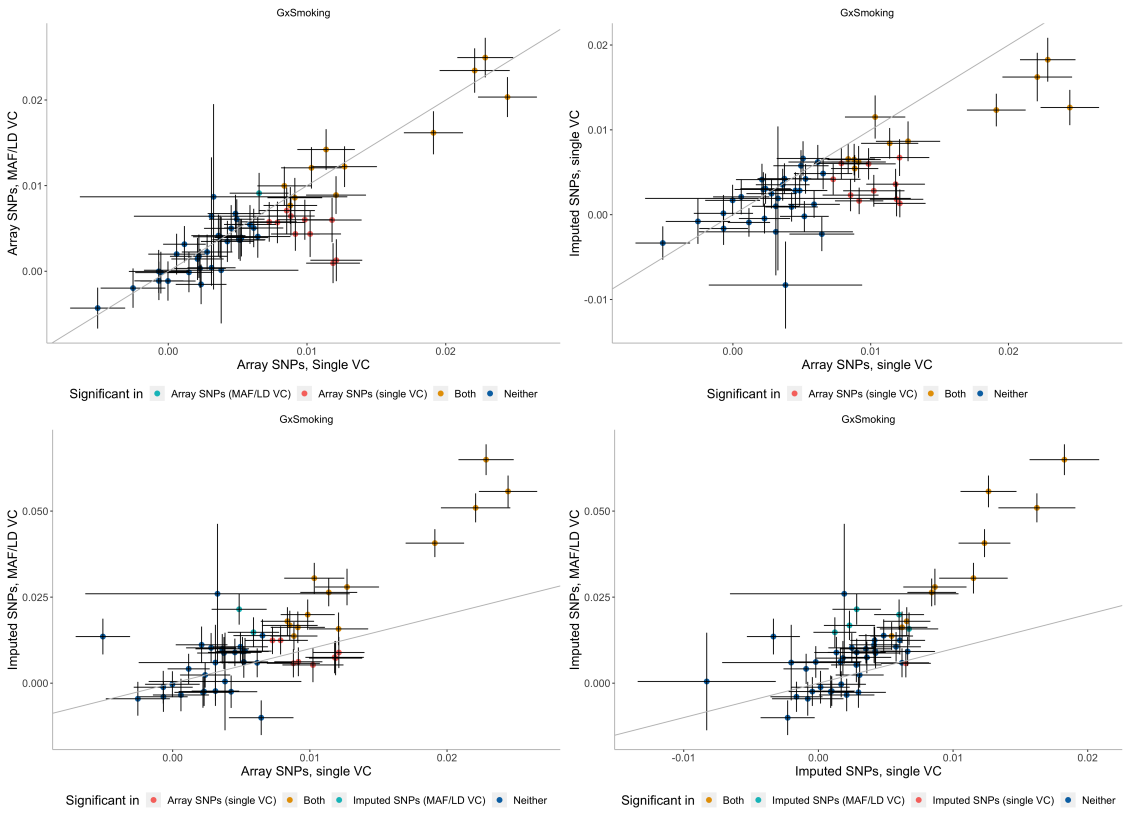

Figure S35: **Effect of MAF-LD partitioning on estimated GxE heritability.** We assessed the effect of MAF-LD partitioning on estimates of $h^2_{gxSmoking}$ from array SNPs and imputed SNPs. We ran GENIE in two settings: (1) fitting a model with a single additive and a single GxE variance component, and (2) fitting a model with eight additive and eight GxE components defined based on four LD annotations (quartiles of LD scores) and two MAF annotations. Black error bars mark ± 2 standard errors centered on the estimates of $h^2_{gxSmoking}$. The color of the dots indicates whether estimates of $h^2_{gxSmoking}$ are significant under each model.

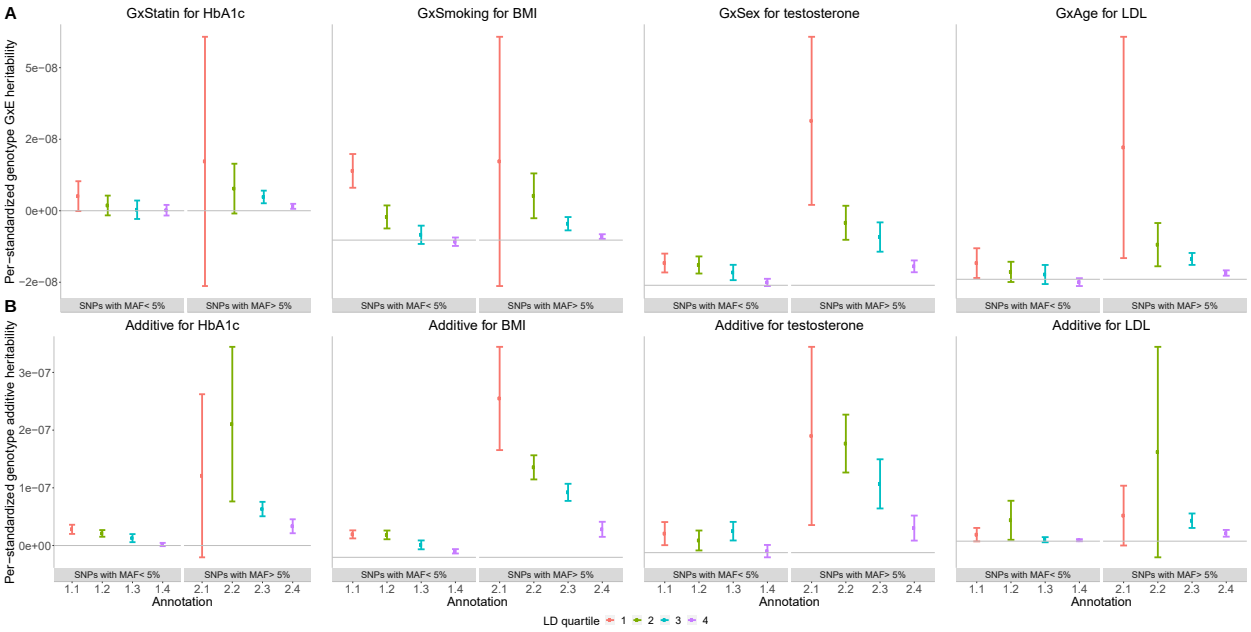

Figure S36: **Per-standardized genotype GxE and additive heritability as a function of MAF and LD.** (**A**) The per-standardized genotype GxE heritability for four selected traits and environments (trait-E pairs). (**B**) The per-allele additive heritability for the same trait-E pairs. The x-axis corresponds to MAF-LD annotations where annotation $i.j$ includes SNPs in MAF bin i and LD quartile j where MAF bin 1 and MAF bin 2 correspond to SNPs with MAF $\leq 5\%$ and MAF $> 5\%$ respectively while the first quartile of LD-scores correspond to SNPs with the lowest LD-scores respectively). The y-axis shows the per-standardized genotype GxE (or additive) heritability defined as $\frac{h_k^2}{2M_k}$ where h_k^2 is the GxE (or additive) heritability attributed to bin k, M_k is the number of SNPs in bin k. Error bars mark ± 2 standard errors centered on the estimated effect sizes.

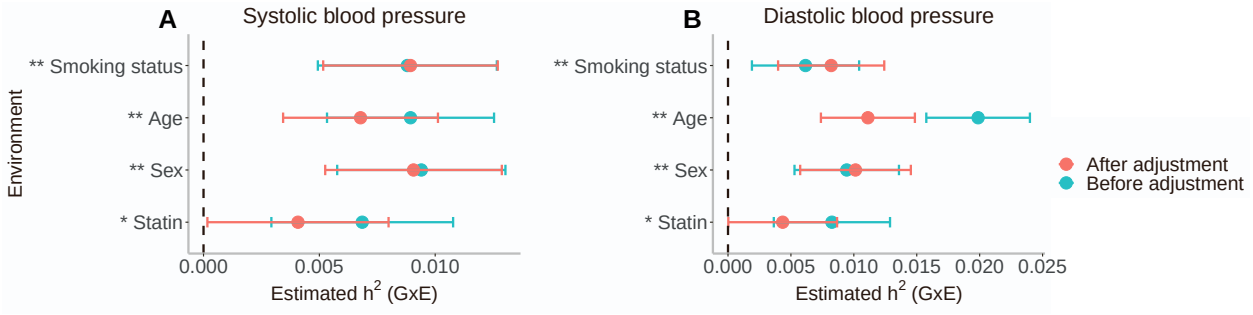

Figure S37: **Effects of adjustment for blood pressure medication on the GxE heritability estimates for systolic and diastolic blood pressure.** The estimated GxE heritability for (**A**) systolic and (**B**) diastolic blood pressure before and after the adjustment for medication. We added 15/10 mmHg to the systolic and diastolic blood pressure measurements for individuals with blood pressure medication on record in UKB, respectively [7]. Then we applied GENIE using the smoking status, age, sex, or statin as the environment exposure on $N = 291,273$ individuals genotyped at $454,207$ SNPs. The point estimates and $\pm$ 2 standard errors were shown, with single and double asterisks denoting p-value < 0.05 and $0.05/200$ for the results after the blood pressure medication adjustment.

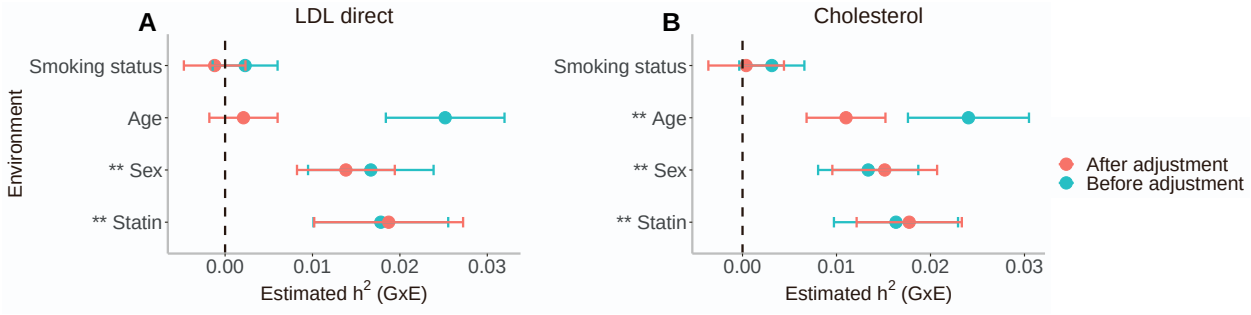

Figure S38: **Effects of adjustment for statin on the GxE heritability estimates for LDL and cholesterol.** The estimated GxE heritability for (**A**) LDL direct and (**B**) cholesterol before and after the adjustment for statin. We divided by 0.684 or 0.749 for LDL direct and cholesterol for those individuals who had taken statin medications, respectively [8]. Then we applied GENIE using the smoking status, age, sex, or statin as the environment exposure on $N = 291,273$ individuals genotyped at $454,207$ SNPs. The point estimates and ± 2 standard errors were shown, with single and double asterisks denoting p-value < 0.05 and $0.05/200$ for the results after the statin medication adjustment.

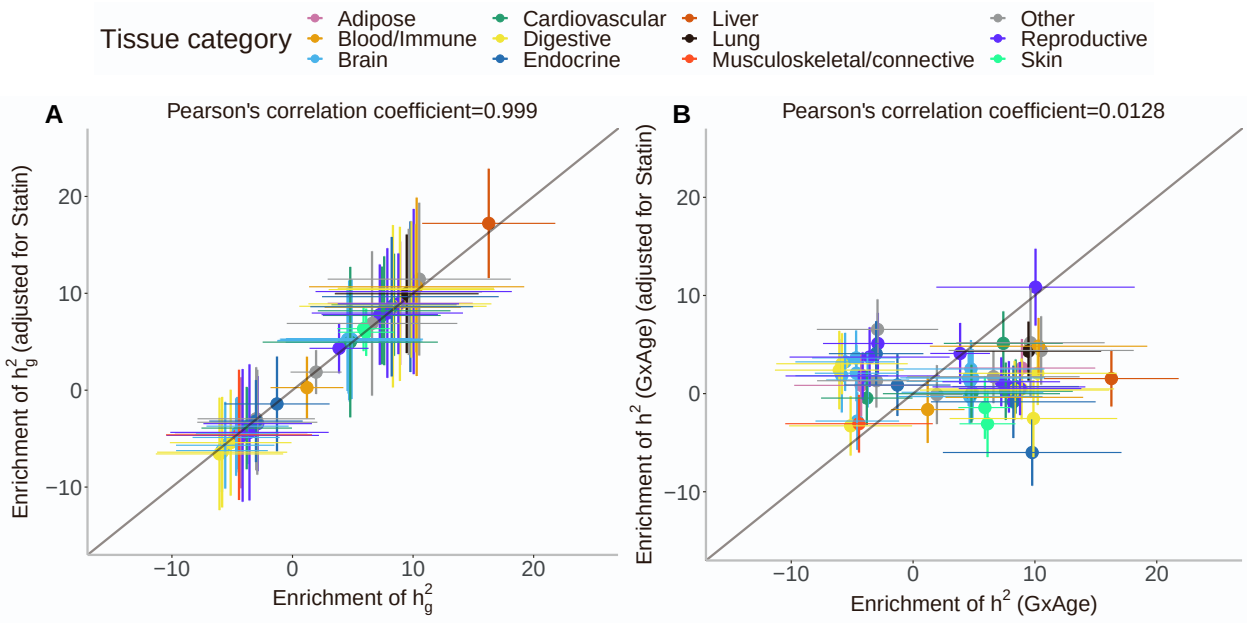

Figure S39: **Comparison of tissue-specific enrichment of G and GxAge heritability for cholesterol before and after adjustment.** The comparison of enrichment of (**A**) additive and (**B**) GxAge heritability in 53 tissues-specific genes for cholesterol before and after adjustment for statin. We divided the measured cholesterol values by 0.749 for individuals who had taken statin medications[8]. The enrichment in each tissue-specific annotation is conditional on 28 functional annotations that are part of the baseline LDSC annotations. The dot and error bars represent the point estimates and SE of the heritability enrichment in a tissue-specific annotation, respectively.

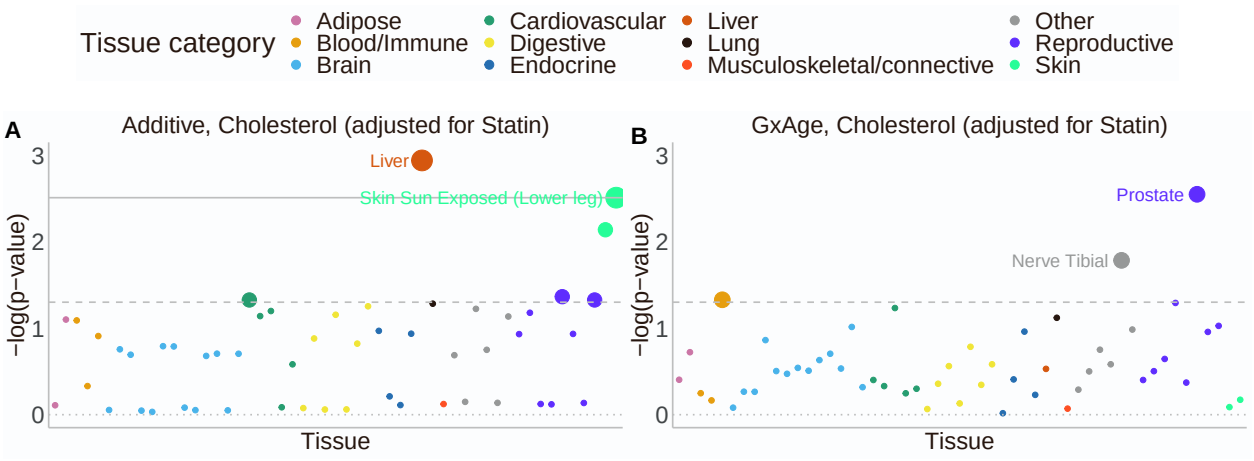

Figure S40: **Partitioning G and GxAge heritability for cholesterol (adjusted for statin) across 53 tissue-specific genes.** We plot $-\log10(p\text{-value})$ of the tissue-specific (**A**) G and (**B**) GxAge enrichment. For every tissue-specific annotation, we use GENIE to test whether this annotation is significantly enriched for per-SNP heritability, conditional on 28 functional annotations that are part of the baseline LDSC annotations. The dashed and solid lines correspond to the nominal $p < 0.05$ and $FDR < 0.1$ threshold, respectively. No tissue had a GxAge enrichment p-value below the $FDR < 0.1$ threshold. We have labeled two tissues with the most significant p-values for each figure.

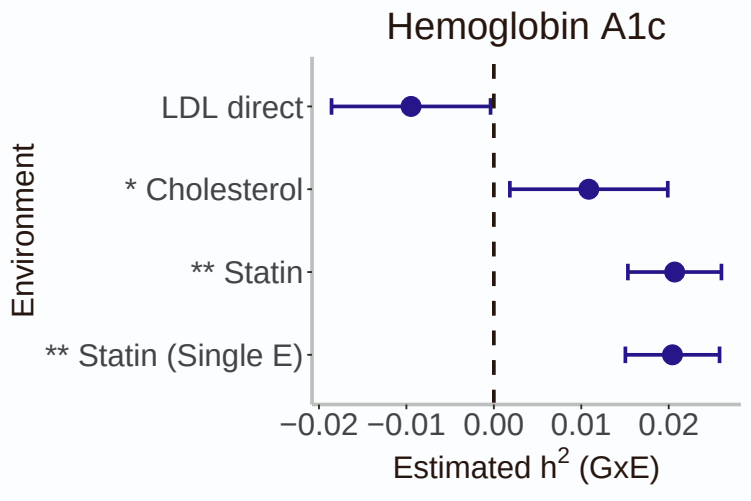

Figure S41: **Impact of LDL and cholesterol as environment exposures on the GxStatin estimates.** The GxStatin heritability for Hemoglobin A1c on $N = 291,273$ individuals genotyped at 454,207 SNPs while incorporating LDL and cholesterol as environment exposures. The GxE heritability of three environment variables was jointly estimated in one pass of GENIE. The point estimates and ± 2 standard errors were shown, with single and double asterisks denoting p-value < 0.05 and $0.05/200$ for the specific GxE heritability estimate. Here "single E" refers to the case where statin was the single environment exposure used in the experiment. The sum of the GxE heritability of LDL, cholesterol, and statin was 2.21%. The GxStatin heritability point estimate was 2.07%, compared to 2.05% where statin is the single environmental exposure.

Supplemental Tables

Method	Variance component	Mean	SE	Bias	Test of bias p-value
GENIE	G	0.2005	0.0382	5e-04	0.9006
GENIE	GxE1	-0.0021	0.027	-0.0021	0.4302
GENIE	GxE2	0.0035	0.0257	0.0035	0.1778
GENIE	GxE3	-1e-04	0.0275	-1e-04	0.9692
GENIE	GxE4	0.0016	0.0245	0.0016	0.5094
GENIE	GxE5	0.0017	0.0267	0.0017	0.5159
GENIE	GxE6	0.0974	0.0278	-0.0026	0.3449
GENIE	GxE7	0.0981	0.0266	-0.0019	0.4673
GENIE	GxE8	0.0975	0.0257	-0.0025	0.3329
GENIE	GxE9	0.0098	0.0263	-2e-04	0.9537
GENIE	GxE10	0.0094	0.0257	-6e-04	0.8082

Table S1: **Accuracy of GENIE in the setting of multiple environmental variables**: We reported the bias, and SE of GENIE under different settings with $L = 10$ environmental variables. Bias, mean and SE are computed from 100 replicates. We report the p-value of a test of the null hypothesis of no bias in the estimates of variance components.

Trait	Category
Alanine aminotransferase	Liver
Albumin	Blood biochemistry
Alcohol intake frequency	Other
Alkaline phosphatase	Liver
Apolipoprotein A	Lipid metabolism
Aspartate aminotransferase	Liver
BMD Heel T-score	Anthropometry
Basal metabolic rate	Anthropometry
Body mass index	Anthropometry
C-reactive protein	Blood biochemistry
Calcium	Other
Cholesterol	Lipid metabolism
Creatinine in urine	Kidney
Creatinine	Kidney
Cystatin-C	Kidney
Diastolic blood pressure	Blood pressure
Eosinophil count	Blood biochemistry
FEV1-FVC ratio	Lung
FVC	Lung
Gamma glutamyltransferase	Liver
Glucose	Glucose metabolism
HDL cholesterol	Lipid metabolism
Height	Anthropometry
Hemoglobin A1c	Glucose metabolism
High light scatter reticulocyte count	Blood biochemistry
IGF-1	Blood biochemistry
LDL direct	Lipid metabolism
Lymphocyte count	Blood biochemistry
Mean corpuscular hemoglobin	Blood biochemistry
Mean platelet volume	Blood biochemistry
Mean sphered cell volume	Blood biochemistry
Microalbumin in urine	Kidney
Monocyte count	Blood biochemistry
Overall health rating	Other
Phosphate	Blood biochemistry
Platelet count	Blood biochemistry
Platelet distribution width	Blood biochemistry
Potassium in urine	Kidney
RBC count	Blood biochemistry
RBC distribution width	Blood biochemistry
SHBG	Blood biochemistry
Sodium in urine	Kidney
Systolic blood pressure	Blood pressure
Testosterone	Blood biochemistry
Triglycerides	Lipid metabolism
Urate	Kidney
Urea	Kidney
Vitamin D	Blood biochemistry
Waist-hip ratio	Anthropometry
White blood cell count	Blood biochemistry

Table S2: **Mapping between the 50 UK Biobank traits and their categories.**

Tissue	Category
Adipose Subcutaneous	Adipose
Adipose Visceral (Omentum)	Adipose
Adrenal Gland	Endocrine
Artery Aorta	Cardiovascular
Artery Coronary	Cardiovascular
Artery Tibial	Cardiovascular
Bladder	Other
Brain Amygdala	Brain
Brain Anterior cingulate cortex (BA24)	Brain
Brain Caudate (basal ganglia)	Brain
Brain Cerebellar Hemisphere	Brain
Brain Cerebellum	Brain
Brain Cortex	Brain
Brain Frontal Cortex (BA9)	Brain
Brain Hippocampus	Brain
Brain Hypothalamus	Brain
Brain Nucleus accumbens (basal ganglia)	Brain
Brain Putamen (basal ganglia)	Brain
Brain Spinal cord (cervical c-1)	Brain
Brain Substantia nigra	Brain
Breast Mammary Tissue	Reproductive
Cells EBV-transformed lymphocytes	Blood/Immune
Cells Transformed fibroblasts	Other
Cervix Ectocervix	Reproductive
Cervix Endocervix	Reproductive
Colon Sigmoid	Digestive
Colon Transverse	Digestive
Esophagus Gastroesophageal Junction	Digestive
Esophagus Mucosa	Digestive
Esophagus Muscularis	Digestive
Fallopian Tube	Reproductive
Heart Atrial Appendage	Cardiovascular
Heart Left Ventricle	Cardiovascular
Kidney Cortex	Other
Liver	Liver
Lung	Lung
Minor Salivary Gland	Other
Muscle Skeletal	Musculoskeletal/connective
Nerve Tibial	Other
Ovary	Reproductive
Pancreas	Other
Pituitary	Endocrine
Prostate	Reproductive
Skin Not Sun Exposed (Suprapubic)	Skin
Skin Sun Exposed (Lower leg)	Skin
Small Intestine Terminal Ileum	Digestive
Spleen	Blood/Immune
Stomach	Digestive
Testis	Endocrine
Thyroid	Endocrine
Uterus	Reproductive
Vagina	Reproductive
Whole Blood	Blood/Immune

Table S3: **Mapping between the 53 GTEx tissues and their categories.**

References

[1] Evans, L. M., Tahmasbi, R., Vrieze, S. I., Abecasis, G. R., Das, S., Gazal, S., Bjelland, D. W., Candia, T. R., Goddard, M. E., Neale, B. M. *et al.* (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nature genetics *50*, 737.

[2] Hou, K., Burch, K. S., Majumdar, A., Shi, H., Mancuso, N., Wu, Y., Sankararaman, S., and Pasaniuc, B. (2019). Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. Nature genetics *51*, 1244–1251.

[3] Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics *88*, 76–82.

[4] 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., of the Psychiatric Genomics Consortium, S. W. G. *et al.* (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics *47*, 291.

[5] Speed, D., Hemani, G., Johnson, M. R., and Balding, D. J. (2012). Improved heritability estimation from genome-wide snps. The American Journal of Human Genetics *91*, 1011–1021.

[6] Di Scipio, M., Khan, M., Mao, S., Chong, M., Judge, C., Pathan, N., Perrot, N., Nelson, W., Lali, R., Di, S. *et al.* (2023). A versatile, fast and unbiased method for estimation of gene-by-environment interaction effects on biobank-scale datasets. Nature Communications *14*, 5196.

[7] Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., Mifsud, B., Ntalla, I., Surendran, P., Liu, C., Cook, J. P. *et al.* (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nature genetics *49*, 403–415.

[8] Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G. R., Wainberg, M., Ollila, H. M., Kiiskinen, T. *et al.* (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. Nature genetics *53*, 185–194.