

S1. Narrow-sense heritability estimates are inflated by higher-order genetic interactions to a greater degree in close than distant relatives

For a relative pair with trait values Y_1 and Y_2 , the covariance between Y_1 and Y_2 follows:¹

$$Cov(Y_1, Y_2) = rV_A + k_2V_D + r^2V_{AA} + rk_2V_{AD} + (k_2)^2V_{DD} + r^3V_{AAA} + \dots \quad (\text{Eq. S1})$$

where r is the coefficient of relatedness (genome fraction of the pair shared identical by descent (IBD)) and k_2 is the Cotterman coefficient (genome fraction where the pair share both alleles by IBD). V_A and V_D are the additive and dominance genetic variance of the trait, respectively. V_{AA} and V_{AAA} represent the two-way and three-way additive interaction variance of the traits, while V_{AD} and V_{DD} represent the two-way additive-dominance and dominance-dominance variance of the traits. Note that additional higher order multi-way additive, additive-dominance, and dominance-dominance variance effects also exist but are not explicitly stated in (Eq. S1) for brevity.

Estimation of narrow-sense heritability requires estimation of the additive genetic variance V_A shown in Eq. S1. Suppose we estimate this quantity using parent-offspring pairs (close relatives), such that $r = 0.5$ and $k_2 = 0$. In this situation, studies often ignore interaction terms and estimate the additive genetic variance as twice the estimated covariance between the trait value of a parent Y_P and its offspring Y_O , that is $\hat{V}_A = \widehat{Cov}(Y_P, Y_O)$. Comparing this quantity to the formula in (Eq. S1), we observe that \hat{V}_A inflates the true V_A by $\frac{1}{2}V_{AA} + \frac{1}{4}V_{AAA} + \frac{1}{8}V_{AAAA} + \dots$. Now, suppose we instead estimate V_A using GWAS data from ‘distant’ relatives (using GCTA or analogous method). In this situation, the samples are putatively unrelated such that r is likely on the order of approximately $\frac{1}{100}$ and k_2 is nearly 0. In this case, based on (Eq. S1), \hat{V}_A inflates the true V_A by approximately $\frac{1}{100}V_{AA} + \left(\frac{1}{100}\right)^2V_{AA} + \left(\frac{1}{100}\right)^3V_{AAA} + \left(\frac{1}{100}\right)^4V_{AAAA} \dots$.

By comparing the \hat{V}_A inflation for parent-offspring pairs (close relatives) to that using GCTA (distant relatives), it becomes apparent that the inflation among the former group will be substantially greater than the latter group when two-way, three-way, and higher-order additive interaction variance exist. This will also lead to greater inflation in narrow-sense heritability estimates when using close relatives for inference compared to distant relatives. Thus, this result suggests that the empirical discrepancy in narrow-sense heritability estimates using close relatives versus distant relatives can possibly be due to interaction effects.

S2. Detailed derivation of Eq. (1)

Given

$$Y_{i,1} = \alpha_1 + \beta_1 G_i + \gamma_1 W_i + \delta_1 G_i W_i + \epsilon_{i,1}$$

and

$$Y_{i,2} = \alpha_2 + \beta_2 G_i + \gamma_2 W_i + \delta_2 G_i W_i + \epsilon_{i,2}.$$

$$\begin{aligned} & \text{cov}(Y_{i,1}, Y_{i,2} | G_i = g_i) \\ &= \text{cov}(\alpha_1 + \beta_1 g_i + \gamma_1 W_i + \delta_1 g_i W_i + \epsilon_{i,1}, \alpha_2 + \beta_2 g_i + \gamma_2 W_i + \delta_2 g_i W_i + \epsilon_{i,2}) \\ &= \text{cov}(\alpha_1, \alpha_2) + \text{cov}(\alpha_1, \beta_2 g_i) + \text{cov}(\alpha_1, \gamma_2 W_i) + \text{cov}(\alpha_1, \delta_2 g_i W_i) \\ &+ \text{cov}(\alpha_1, \epsilon_{i,2}) + \text{cov}(\beta_1 g_i, \alpha_2) + \text{cov}(\beta_1 g_i, \beta_2 g_i) + \text{cov}(\beta_1 g_i, \gamma_2 W_i) \\ &+ \text{cov}(\beta_1 g_i, \delta_2 g_i W_i) + \text{cov}(\beta_1 g_i, \epsilon_{i,2}) + \text{cov}(\gamma_1 W_i, \alpha_2) + \text{cov}(\gamma_1 W_i, \beta_2 g_i) \\ &+ \text{cov}(\gamma_1 W_i, \gamma_2 W_i) + \text{cov}(\gamma_1 W_i, \delta_2 g_i W_i) + \text{cov}(\gamma_1 W_i, \epsilon_{i,2}) + \text{cov}(\delta_1 g_i W_i, \alpha_2) \\ &+ \text{cov}(\delta_1 g_i W_i, \beta_2 g_i) + \text{cov}(\delta_1 g_i W_i, \gamma_2 W_i) + \text{cov}(\delta_1 g_i W_i, \delta_2 g_i W_i) \\ &+ \text{cov}(\delta_1 g_i W_i, \epsilon_{i,2}) + \text{cov}(\epsilon_{i,1}, \alpha_2) + \text{cov}(\epsilon_{i,1}, \beta_2 g_i) + \text{cov}(\epsilon_{i,1}, \gamma_2 W_i) \\ &+ \text{cov}(\epsilon_{i,1}, \delta_2 g_i W_i) + \text{cov}(\epsilon_{i,1}, \epsilon_{i,2}) \\ &= \text{cov}(\gamma_1 W_i, \gamma_2 W_i) + \text{cov}(\gamma_1 W_i, \delta_2 g_i W_i) + \text{cov}(\delta_1 g_i W_i, \gamma_2 W_i) \\ &+ \text{cov}(\delta_1 g_i W_i, \delta_2 g_i W_i) \\ &= \gamma_1 \gamma_2 \text{cov}(W_i, W_i) + \gamma_1 \delta_2 g_i \text{cov}(W_i, W_i) + \gamma_2 \delta_1 g_i \text{cov}(W_i, W_i) \\ &+ \delta_1 \delta_2 g_i^2 \text{cov}(W_i, W_i) \\ &= \gamma_1 \gamma_2 + (\gamma_1 \delta_2 + \gamma_2 \delta_1) g_i + \delta_1 \delta_2 g_i^2 \end{aligned}$$

and

$$\begin{aligned}
\text{var}(Y_{i,1}|G_i = g_i) &= \text{var}(\alpha_1 + \beta_1 g_i + \gamma_1 W_i + \delta_1 g_i W_i + \epsilon_{i,1}) \\
&= \gamma_1^2 \text{var}(W_i) + \delta_1^2 g_i^2 \text{var}(W_i) + 2\gamma_1 \delta_1 g_i \text{cov}(W_i, W_i) + 1 \\
&= \gamma_1^2 + \delta_1^2 g_i^2 + 2\gamma_1 \delta_1 g_i + 1 \\
&= (\gamma_1 + \delta_1 g_i)^2 + 1
\end{aligned}$$

As $\text{cor}(Y_{i,1}, Y_{i,2}|G_i = g_i) = \frac{\text{cov}(Y_{i,1}, Y_{i,2}|G_i = g_i)}{\sqrt{\text{var}(Y_{i,1}|G_i = g_i)}\sqrt{\text{var}(Y_{i,2}|G_i = g_i)}}$, we can use our formulas above to

show that

$$\text{cor}(Y_{i,1}, Y_{i,2}|G_i = g_i) = \frac{\gamma_1 \gamma_2 + (\delta_1 \gamma_2 + \delta_2 \gamma_1) g_i + \delta_1 \delta_2 g_i^2}{\sqrt{(\gamma_1 + \delta_1 g_i)^2 + 1} \times \sqrt{(\gamma_2 + \delta_2 g_i)^2 + 1}}$$

S3: Approximating the Pearson correlation coefficient for two variables

Suppose we have two variables, \mathbf{V}_1 and \mathbf{V}_2 , that are each a column vector with N elements. The correlation coefficient between \mathbf{V}_1 and \mathbf{V}_2 can be estimated using the sample Pearson correlation formula. Define the row-wise pairs as $[\mathbf{V}_1, \mathbf{V}_2] = [(v_{1,1}, v_{2,1}), \dots, (v_{1,n}, v_{2,n})]^T$. The sample Pearson correlation coefficient of \mathbf{V}_1 and \mathbf{V}_2 is expressed as:

$$r_{\mathbf{V}_1\mathbf{V}_2} = \frac{\sum_{i=1}^n (v_{1i} - \bar{v}_1)(v_{2i} - \bar{v}_2)}{\sqrt{\sum_{i=1}^n (v_{1i} - \bar{v}_1)^2} \times \sqrt{\sum_{i=1}^n (v_{2i} - \bar{v}_2)^2}} \quad (\text{Eq. S2})$$

Define s_{n,\mathbf{V}_1} and s_{n,\mathbf{V}_2} as the sample standard deviation of the two variables, i.e., $s_{n,\mathbf{V}_1} = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{1i} - \bar{v}_1)^2}$ and $s_{n,\mathbf{V}_2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{2i} - \bar{v}_2)^2}$, respectively. We want to point out that the standard deviation (SD) of each variable is the average, not the Bessel-corrected average using $n - 1$. This decision is because the typical sample size is large in biobank-scale data that we consider in this work and, as such, the average and Bessel-corrected average are asymptotically equivalent. Using the SD formulas, we can rewrite (S2) as

$$\frac{\sum_{i=1}^n (v_{1i} - \bar{v}_1)(v_{2i} - \bar{v}_2)}{\sqrt{\sum_{i=1}^n (v_{1i} - \bar{v}_1)^2} \times \sqrt{\sum_{i=1}^n (v_{2i} - \bar{v}_2)^2}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{v_{1i} - \bar{v}_1}{s_{n,\mathbf{V}_1}} \right) \left(\frac{v_{2i} - \bar{v}_2}{s_{n,\mathbf{V}_2}} \right) = \frac{1}{n} \sum_{i=1}^n \tilde{v}_{1i} \tilde{v}_{2i}$$

where we can interpret $\tilde{v}_{1i} = \frac{v_{1i} - \bar{v}_1}{s_{n,\mathbf{V}_1}}$ and $\tilde{v}_{2i} = \frac{v_{2i} - \bar{v}_2}{s_{n,\mathbf{V}_2}}$ as the standardized samples of v_{1i} and v_{2i} from \mathbf{V}_1 and \mathbf{V}_2 , respectively. This step transforms \mathbf{V}_1 and \mathbf{V}_2 into $\tilde{\mathbf{V}}_1$ and $\tilde{\mathbf{V}}_2$, such that each has a mean of zero and a standard deviation of one. We can then approximate the sample Pearson correlation coefficient, $r_{\mathbf{V}_1\mathbf{V}_2}$ in (S2) as the average of these pairwise products:

$$r_{V_1V_2} = \frac{1}{n} \sum_{i=1}^n \left(\frac{v_{1i} - \bar{v}_1}{s_{n,v_1}} \right) \left(\frac{v_{2i} - \bar{v}_2}{s_{n,v_2}} \right) = \frac{1}{n} \sum_{i=1}^n \tilde{v}_{1i} \tilde{v}_{2i}$$

We note that the pairwise products of these standardized variables can also be conveniently denoted as $\tilde{V}_1 \odot \tilde{V}_2$, which is a $N \times 1$ dimensional column vector. Then, we can express the sample average of $\tilde{V}_1 \odot \tilde{V}_2$ as $\frac{1}{N} J_1^n (\tilde{V}_1 \odot \tilde{V}_2)$, where J_1^n is a $1 \times n$ vector with all elements equal to 1. Using (S2), it then follows that we can express the sample Pearson correlation coefficient as

$$r_{V_1V_2} = \frac{1}{N} J_1^n (\tilde{V}_1 \odot \tilde{V}_2).$$

To determine if the correlation coefficient differs by genotype category, $\mathbf{G} \in \{0, 1, 2\}$, we can examine whether $r_{V_1V_2}$ varies across genotype category by comparing $r_{V_1V_2, \mathbf{G}=0}$, $r_{V_1V_2, \mathbf{G}=1}$ and $r_{V_1V_2, \mathbf{G}=2}$, which can help to evaluate the presence of interaction effects in at least one of the two variables under study.

S4: Derivation of Levene's test for a single trait

Assume a sample of size N divided into k subgroups, where N_i is the sample size of the i^{th} subgroup. To test whether variance of a trait Y differs by subgroup, we define the Levene's test statistic as:

$$W = \frac{(N - k) \sum_{i=1}^k N_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - \bar{Z}_i)^2}$$

- $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$: Y_{ij} the value of Y for the j th observation of the i th subgroup
- $\bar{Y}_{i.}$: the mean of the i th subgroup
- \bar{Z}_i : the subgroup means of the Z_{ij}
- $\bar{Z}_{..}$: overall means of the Z_{ij}

The Levene test rejects the hypothesis that the variances are equal if $W > F_{(\alpha, k-1, N-k)}$ where $F_{(\alpha, k-1, N-k)}$ is the upper critical value of the F distribution with $k - 1$ and $N - k$ degrees of freedom at a significance level of α .

S5: Significance threshold for evaluating type-I error

To explain why $\alpha = 0.001$ is an adequate choice for our nominal type 1 error α level when there are 100,000 replicates, we are given $X_1, \dots, X_m \sim_{i.i.d} \text{Binom}(n, \alpha)$ by C.L.T, then $\sqrt{m}(\bar{X}_m - \alpha) \xrightarrow{d} \mathbb{N}(0, \alpha(1 - \alpha))$, equivalently, $\frac{\bar{X}_m - \alpha}{\sqrt{\frac{\alpha(1-\alpha)}{m}}} \xrightarrow{d} \mathbb{N}(0,1)$. Thus, the asymptotic mean (μ) and asymptotic variance (σ^2) of \bar{X}_m are α and $\frac{\alpha(1-\alpha)}{m}$.

In our simulation setting, we have fixed the number of simulated replicates as $M = 100,000$, and we let α be the pre-specified threshold. This table provides the corresponding pre-specified significance level (col “ α ”), asymptotic mean (col “ $\mu = \alpha$ ”), asymptotic standard error (col “ $\sigma = \sqrt{\frac{\alpha(1-\alpha)}{M}}$ ”), and 95% single-side CI bandwidth (col “95% CI bandwidth 1.96* $SE(\bar{X}_n)$ ”) when the pre-specified thresholds vary at the given fixed $M = 100,000$.

In general, the bandwidth of the 95% CI for a proportion such as α should not overlap but be at least one order of magnitude smaller than the asymptotic mean.² In this table, when $\alpha = 0.01$ or $\alpha = 0.001$, their 95% CI bandwidths are smaller than their corresponding asymptotic mean at least at one order of magnitude. When $\alpha = 0.0001$ or smaller, their asymptotic mean and 95% CI bandwidth are very close in value and do not differ by an order of magnitude.

M	α	$\mu = \alpha$	$\sigma = \sqrt{\frac{\alpha(1-\alpha)}{M}}$	95% CI bandwidth 1.96* $SE(\bar{X}_n)$
1E5	1E-2	1E-2	3.15E-4	6.17E-4
	1E-3	1E-3	9.99E-5	1.96E-4
	1E-4	1E-4	3.16E-5	6.20E-5
	1E-5	1E-5	1.00E-5	1.96E-5

Figure S1

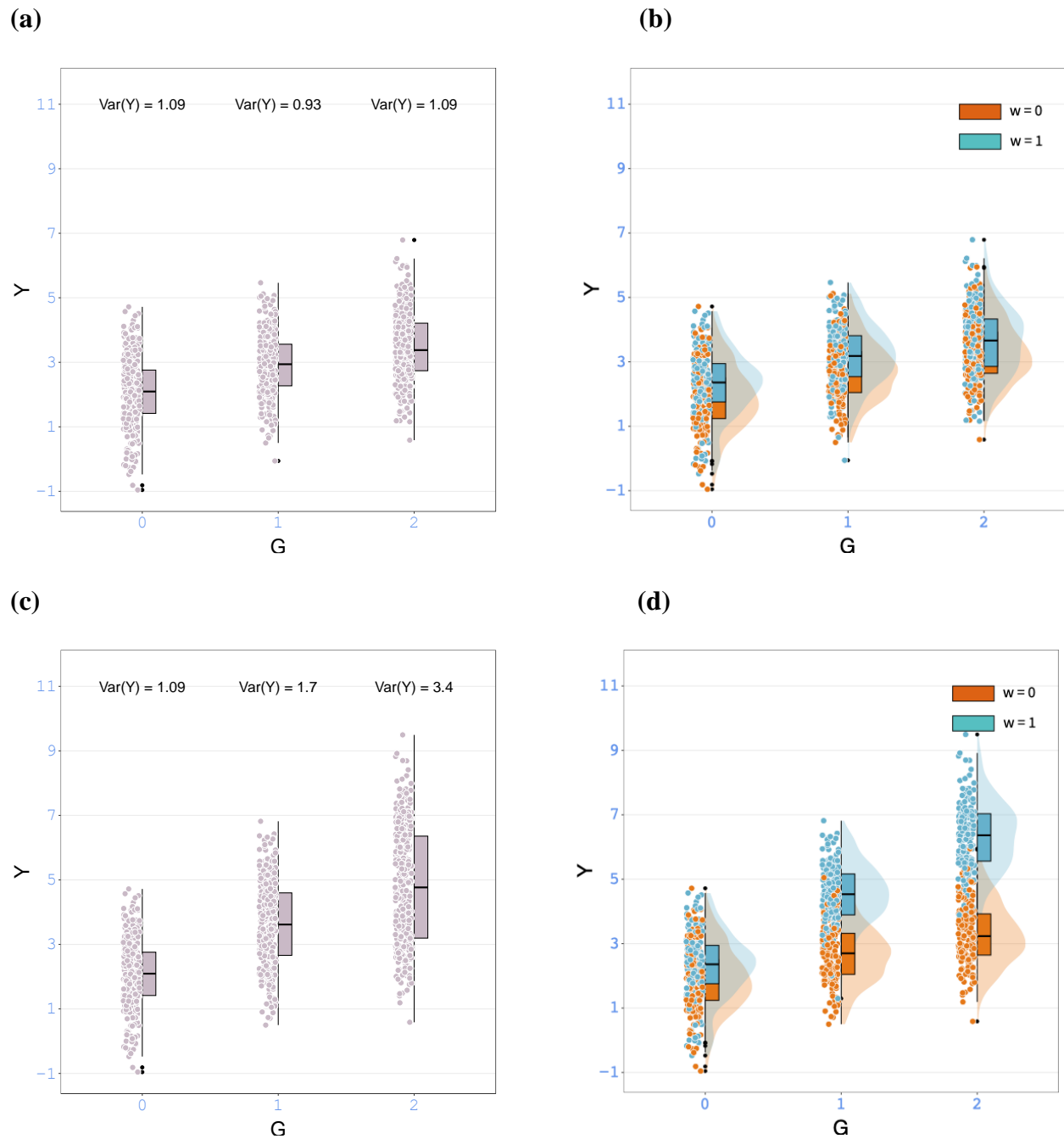


Figure S1. Interaction effect induce variance heteroscedasticity in a phenotypic trait.

Sub-figures (a) and (c) depict simulations of one phenotypic trait samples from $Y_i = \alpha_1 + \beta_1 G_i + \gamma_1 W_i + \epsilon_i$ (without interaction effect) and $Y_i = \alpha_1 + \beta_1 G_i + \gamma_1 W_i + \delta_1 G_i W_i + \epsilon_i$ (with interaction effect), respectively. In these figures, the interacting covariate W_i is simplified as a

binary factor for clarity. The genotype categories are plotted on the x-axis, and the simulated outcomes Y are on the y-axis, with variances for each genotype category annotated. The distributions of Y differ across the three genotype categories as shown in the sub-figure (c), but not observed in the sub-figure (a). The sub-figures (b) and (d) mirror (a) and (c), respectively, with the addition of labels by W using the color orange ($W_i = 0$) and blue ($W_i = 1$). By comparing the sub-figures (b) and (d), the variation in Y distributions across genotype categories, as observed in sub-figure (c), highlights the variance heterogeneity driven by the interacting effect, GW .

Figure S2.

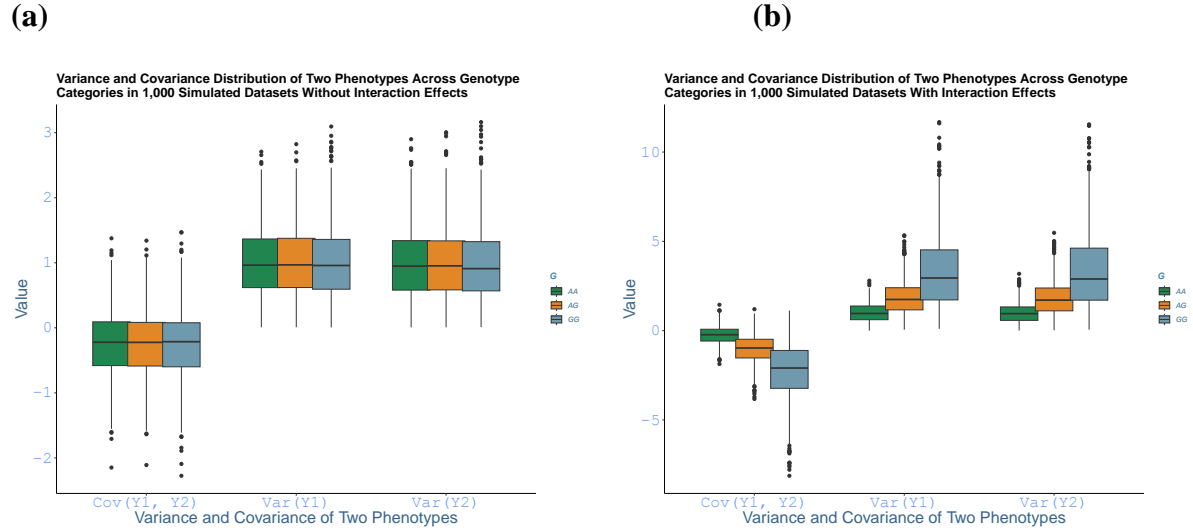


Figure S2. Interaction effect induces both variance and covariance heteroscedasticity in two phenotypic traits.

Sub-figures (a) and (b) depict the distribution of three covariance elements on the x-axis, including $cov(\mathbf{Y}_1, \mathbf{Y}_2)$, $var(\mathbf{Y}_1)$ and $var(\mathbf{Y}_2)$, across different genotypes, marked by varying colors within each covariance element. These measures are plotted using 1,000 datasets, and each dataset contains 1,000 samples of \mathbf{Y}_1 and 1,000 samples of \mathbf{Y}_2 , which were simulated from multivariate normal distributions with and without gene-environment interaction terms:

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} \sim MVN \left(\begin{matrix} \alpha_1 + \beta_1 G_i + \gamma_1 W_i \\ \alpha_2 + \beta_2 G_i + \gamma_2 W_i \end{matrix}, \Sigma = \begin{bmatrix} Unif(0,1) & Unif(0,1) \\ Unif(0,1) & Unif(0,1) \end{bmatrix} \right) \text{ and}$$

$$\begin{pmatrix} Y_{1,i} \\ Y_{2,i} \end{pmatrix} \sim MVN \left(\begin{matrix} \alpha_1 + \beta_1 G_i + \gamma_1 W_i + \delta_1 G_i W_i \\ \alpha_2 + \beta_2 G_i + \gamma_2 W_i + \delta_2 G_i W_i \end{matrix}, \Sigma = \begin{bmatrix} Unif(0,1) & Unif(0,1) \\ Unif(0,1) & Unif(0,1) \end{bmatrix} \right), \text{ as shown in sub-figures (a) and}$$

(b), respectively. We then can calculate the three covariance/variance components, $cov(\mathbf{Y}_1, \mathbf{Y}_2)$, $var(\mathbf{Y}_1)$ and $var(\mathbf{Y}_2)$, for each dataset. The two sub-figures draw the distribution of the covariance/variance components of the 1,000 datasets. Sub-figure (a) shows no visual distinction in the distribution of these statistical measures across genotypes when interaction is absent. In

contrast, sub-figure (b) demonstrates noticeable differences in $cov(\mathbf{Y}_1, \mathbf{Y}_2)$, $var(\mathbf{Y}_1)$ and $var(\mathbf{Y}_2)$ across genotypes when interaction effects are present.

Figure S3.

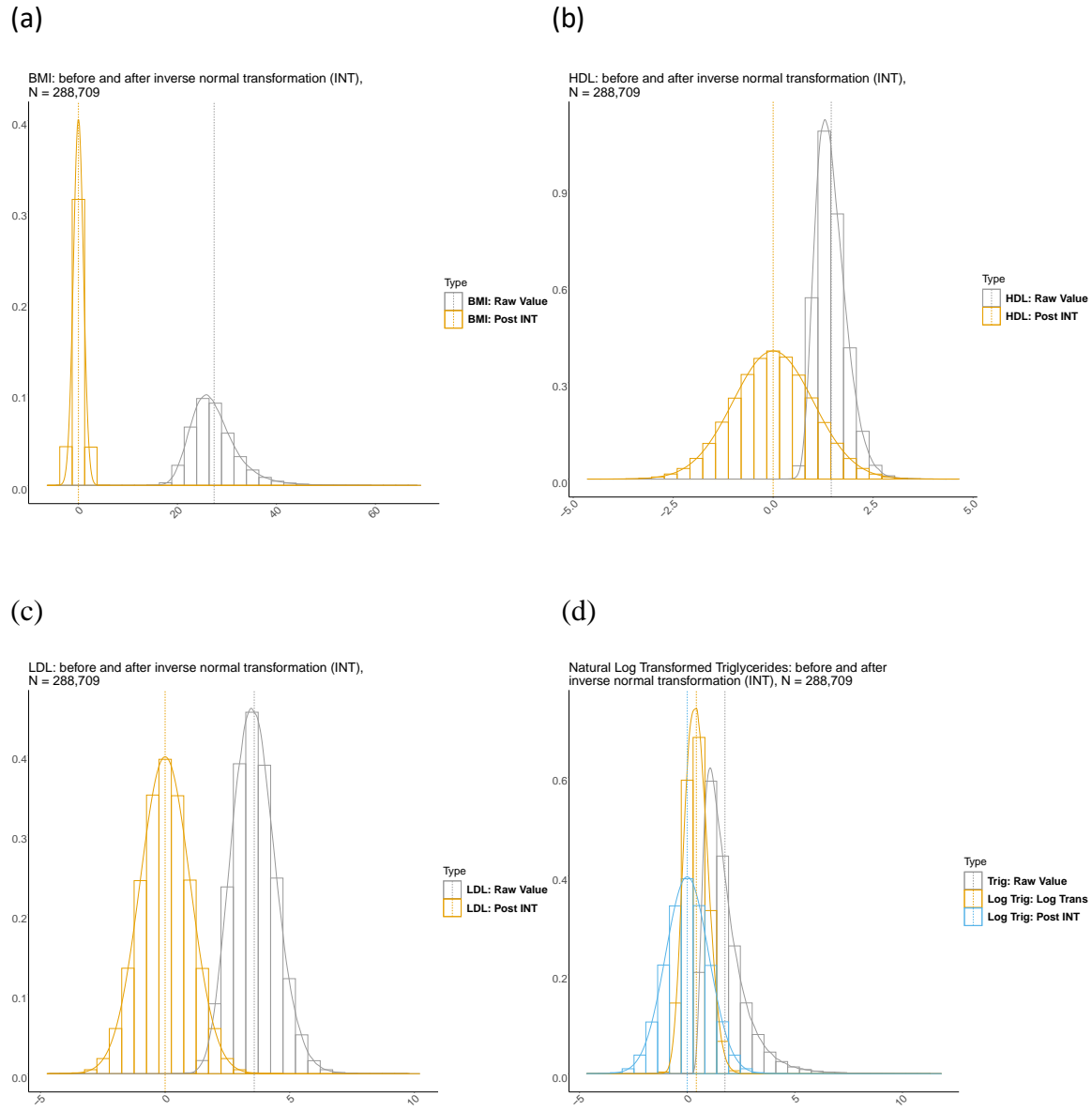


Figure S3. Distribution of the four lipids panel-related traits before and post the inverse normal transformation (INT) in UKBB analysis

In our UK Biobank (UKBB) analysis of 288,709 individuals, we investigated four traits: high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), triglycerides (TGs), and body mass index (BMI). Grey denotes the distribution prior to intervention (INT), and yellow indicates the distribution post-INT. For TGs, we transformed the

raw data using the natural logarithm for further analysis, as depicted in sub-figure (d). The dashed vertical line marks the mean of each distribution. We observed that the distributions of these traits are skewed, as evidenced by the means not aligning with the peaks of the histograms.

Figure S4.

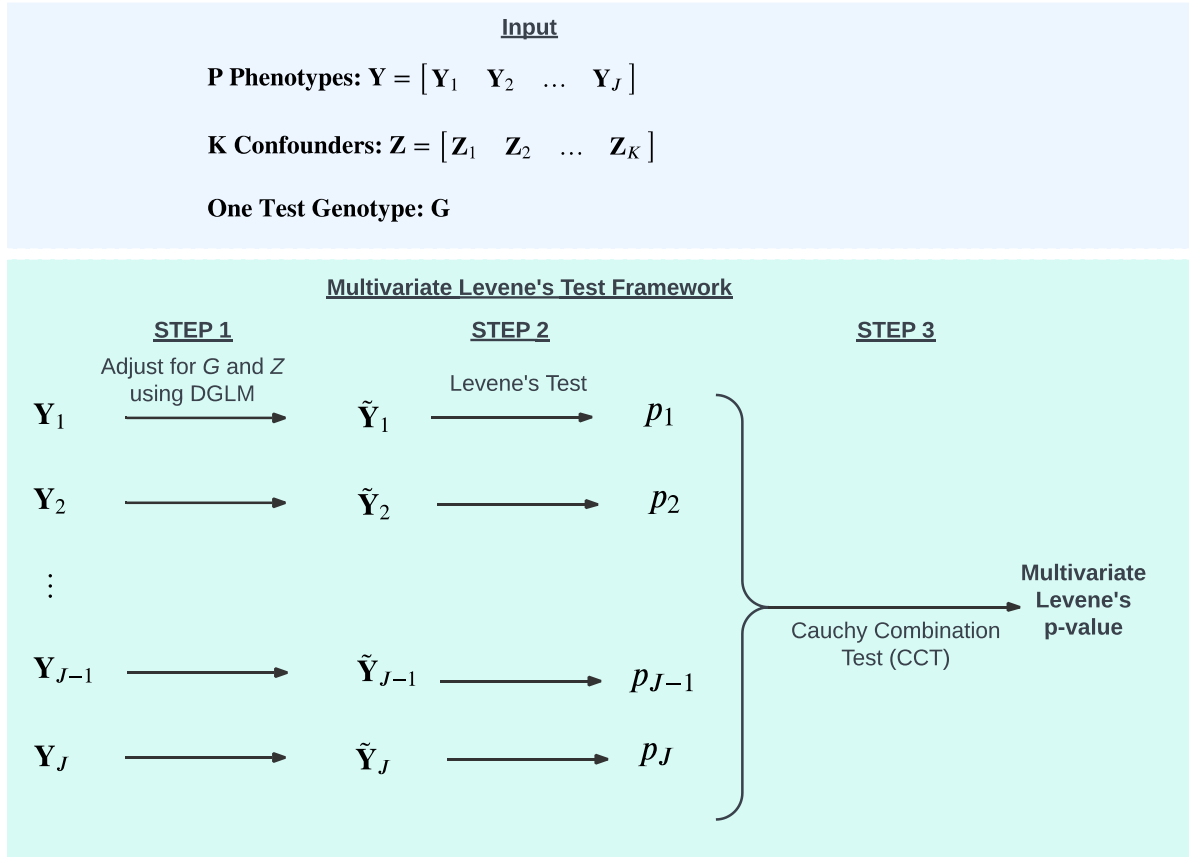


Figure S4. Illustration of the multivariate Levene’s test framework.

Multivariate Levene’s test framework is a three-step process that includes adjustment for genotype and confounders, applying univariate Levene’s test for each one of the J traits, and then combines the J p-values using Cauchy Combination Test (CCT) to compute the final Multivariate Levene’s p-value to determine overall significance.

Figure S5.

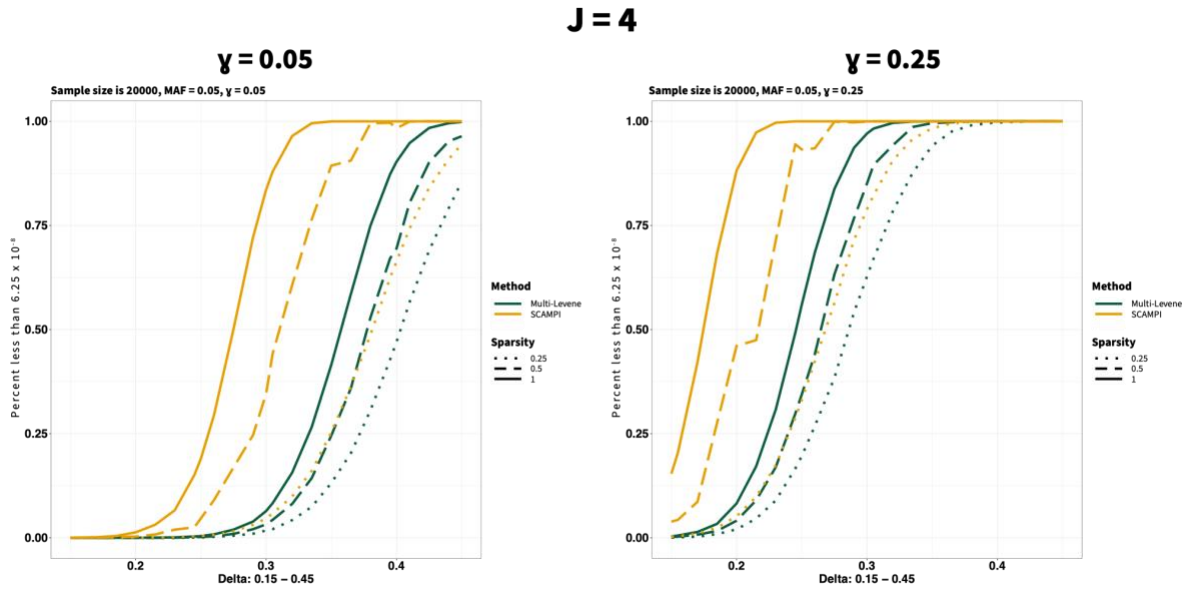
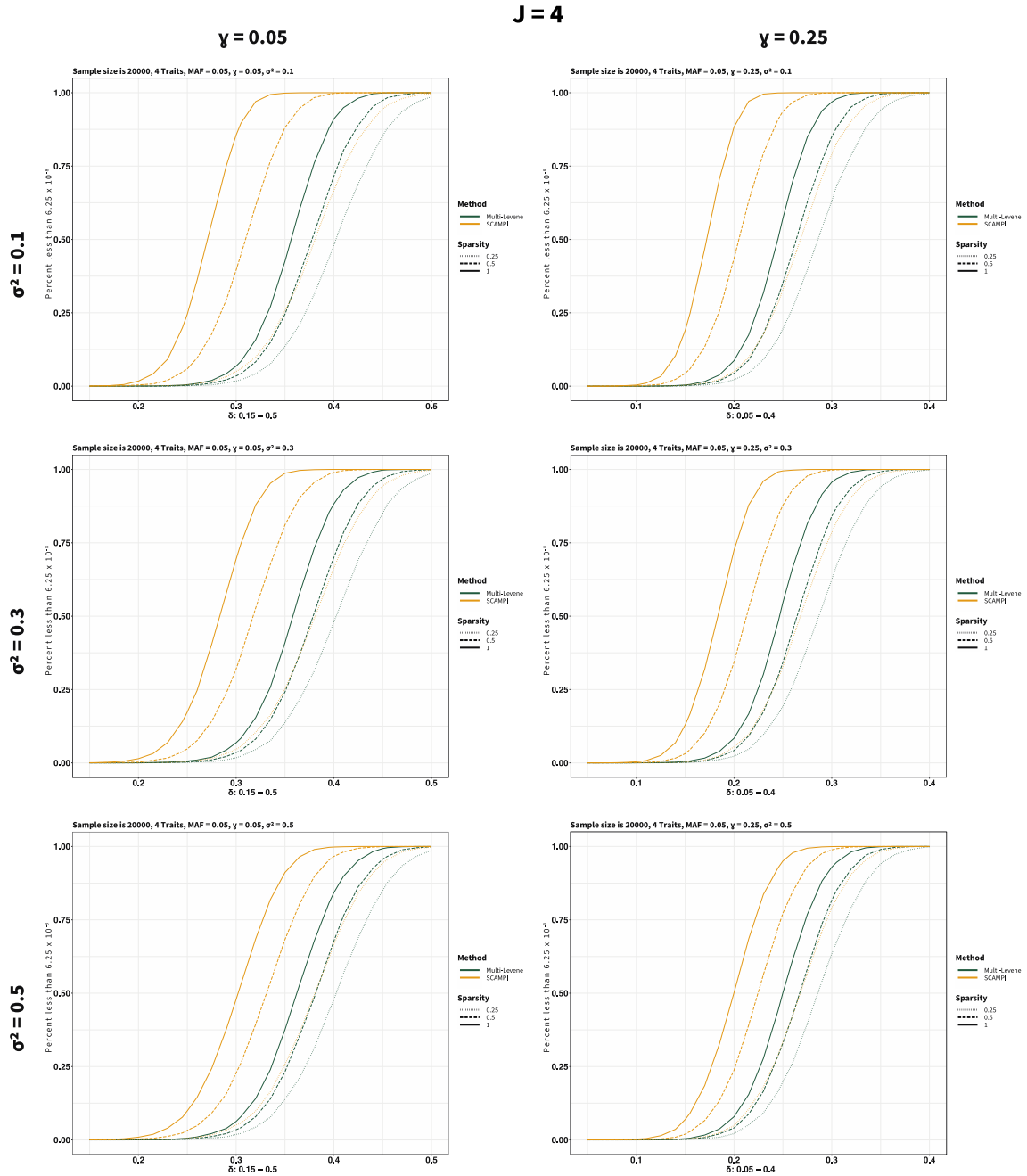


Figure S5. Power results for UKBB inspired simulations of SCAMPI and Multivariate Levene's test varying proportions of traits with interactions.

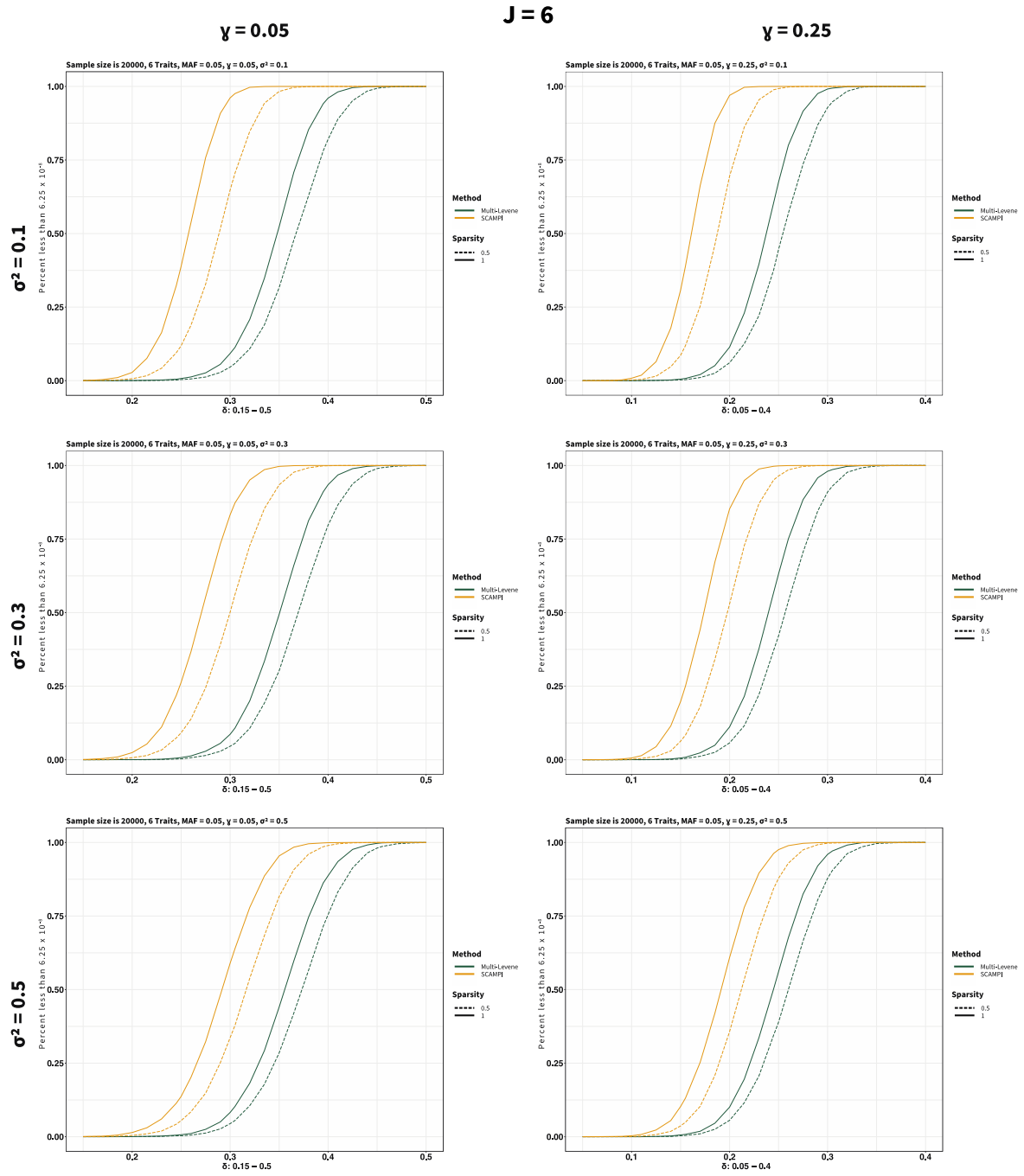
The power for UKBB inspired simulation of SCAMPI and Multivariate Levene's test from two scenarios is depicted by overlaying different sparsity levels in two sub-figures. Each sub-figure corresponds to the two magnitudes of the main effects from interacting covariates ($\gamma = 0.05$ for mild and $\gamma = 0.25$ for strong effects). For each of the two scenarios, we assume a sample size of 200,000 and the test SNP has a MAF of 0.05. We plot the power of SCAMPI (denoted as color yellow) and Multivariate Levene's test (denoted as color green) with different proportions of traits with interactions (25%, 50%, and 100%, denoted as different line types) across values of the interaction effect δ .

Figure S6.

(a)



(b)



(c)

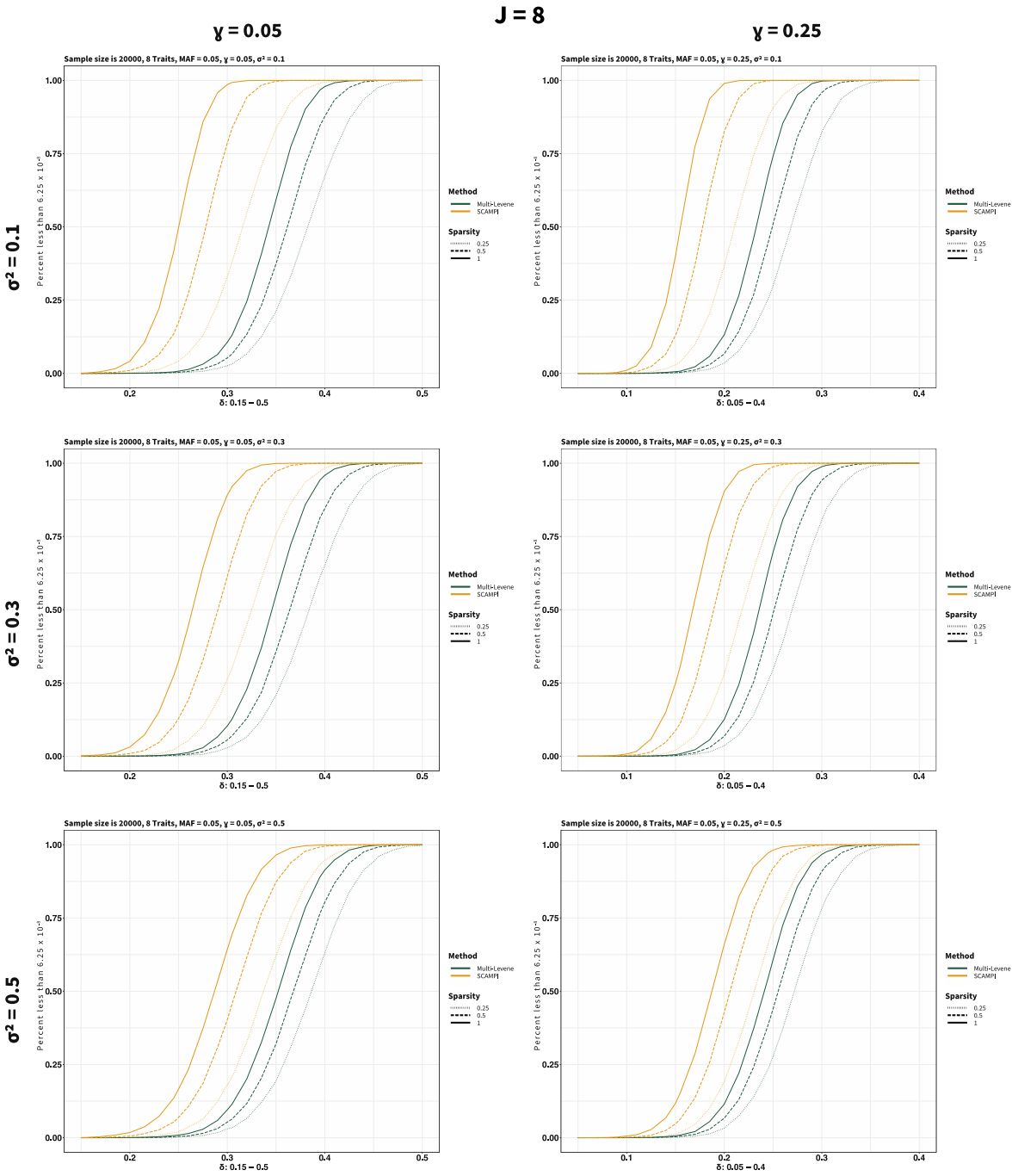


Figure S6. Additional power results for SCAMPI and Multivariate Levene’s test varying proportions of traits with interactions.

The power of SCAMPI and Multivariate Levene’s test across eighteen scenarios is depicted by overlaying different sparsity levels in eighteen sub-figures. Each scenario is defined by a distinct combination of variables: number of phenotypes ($J = 4, 6, 8$), magnitude of main effects from interacting covariates ($\gamma = 0.05$ for mild and $\gamma = 0.25$ for strong effects), and correlation strength among traits ($\sigma^2 = 0.1$ for mild, $\sigma^2 = 0.3$ for medium, and $\sigma^2 = 0.5$ for strong). The results are organized by number of phenotypes ($J = 4, 6, 8$) in sub-figure (a), (b) and (c). For each sub-figure, we assume 200,000 samples and that the test SNP has a MAF of 0.05. Within each sub-figure, we plot the power of SCAMPI (denoted as color yellow) and Multivariate Levene’s test (denoted as color green) with different proportions of traits with interactions (25%, 50%, and 100%, denoted as different line types) across values of the interaction effect δ .

Figure S7.

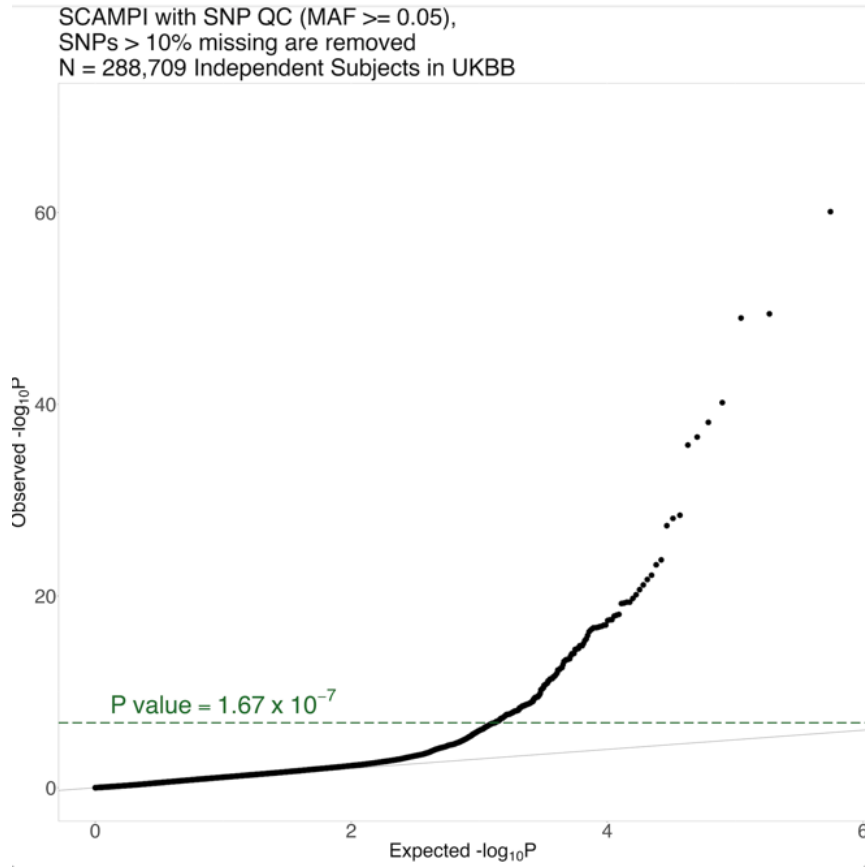
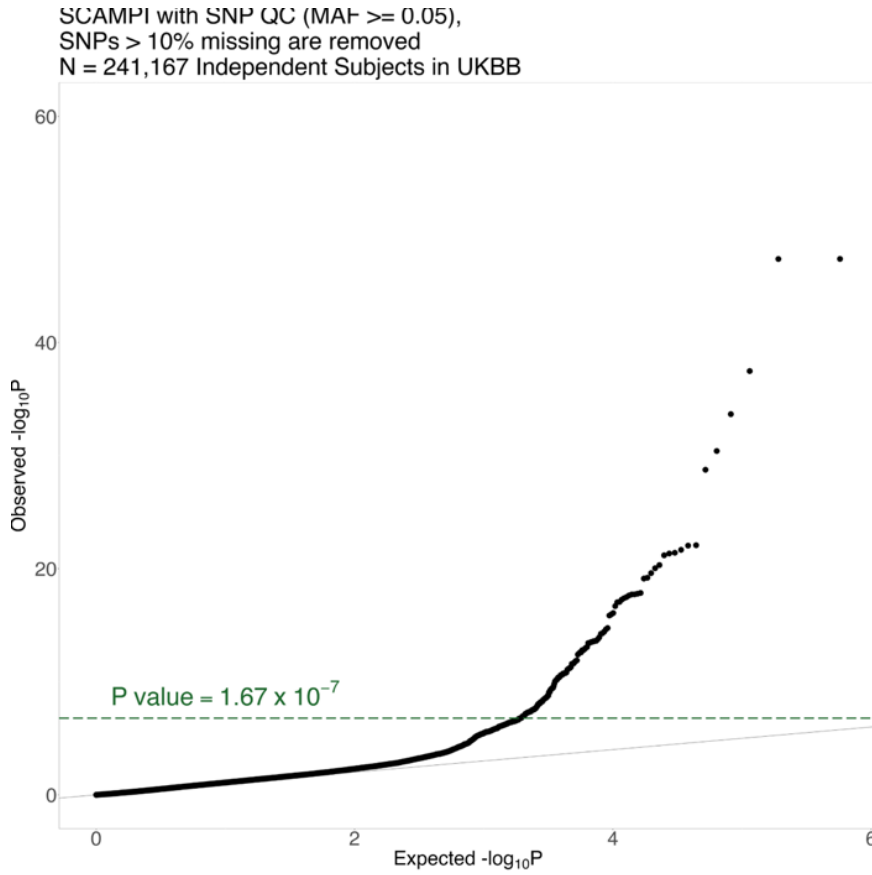


Figure S7. Q-Q plot of the genome-wide results on lipid traits in UKBB using SCAMPI

The corresponding Q-Q plot of the 210 SNPs displayed in the Manhattan plot in figure 3. The dash green line indicates the pre-specified study-wide significance ($\alpha = 1.67 \times 10^{-7}$).

Figure S8.



(b)

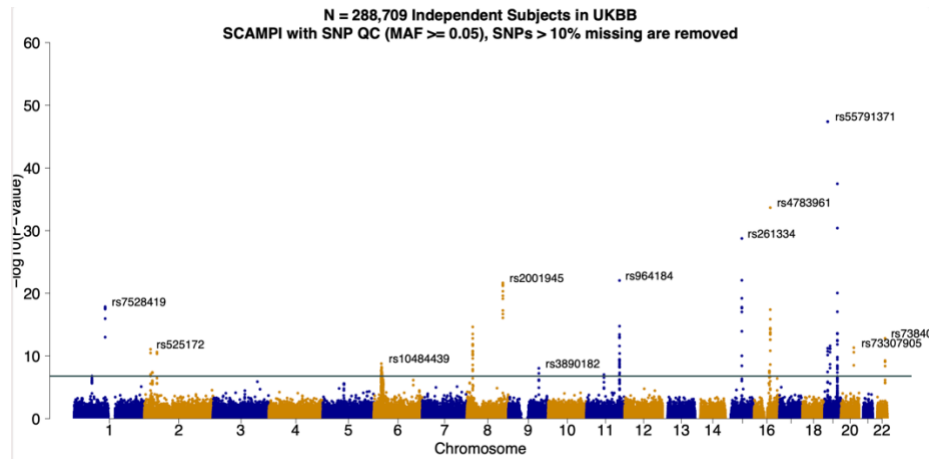


Figure S8. Conditional SCAMPI analysis adjusting for APOE SNPs

Plots (a) and (b) are QQ plot and Manhattan plot based on the SCAMPI p-values analyzing HDL-C, LDL-C, TGs and BMI in UKBB. The analysis involves 241,167 independent

participants. In this analysis, we adjust for the mean and variance effects of five APOE SNPs, rs440446, rs769449, rs769450, rs429358, and rs7412. SNPs having more than 10% missing are removed from the visualization. The p-value threshold is 1.67×10^{-7} to account for the approximately 300,000 tests using Bonferroni correction (solid horizontal line). In the Manhattan plot, the rsID of the SNP having the largest p-value within each chromosome is labeled.

Figure S9.

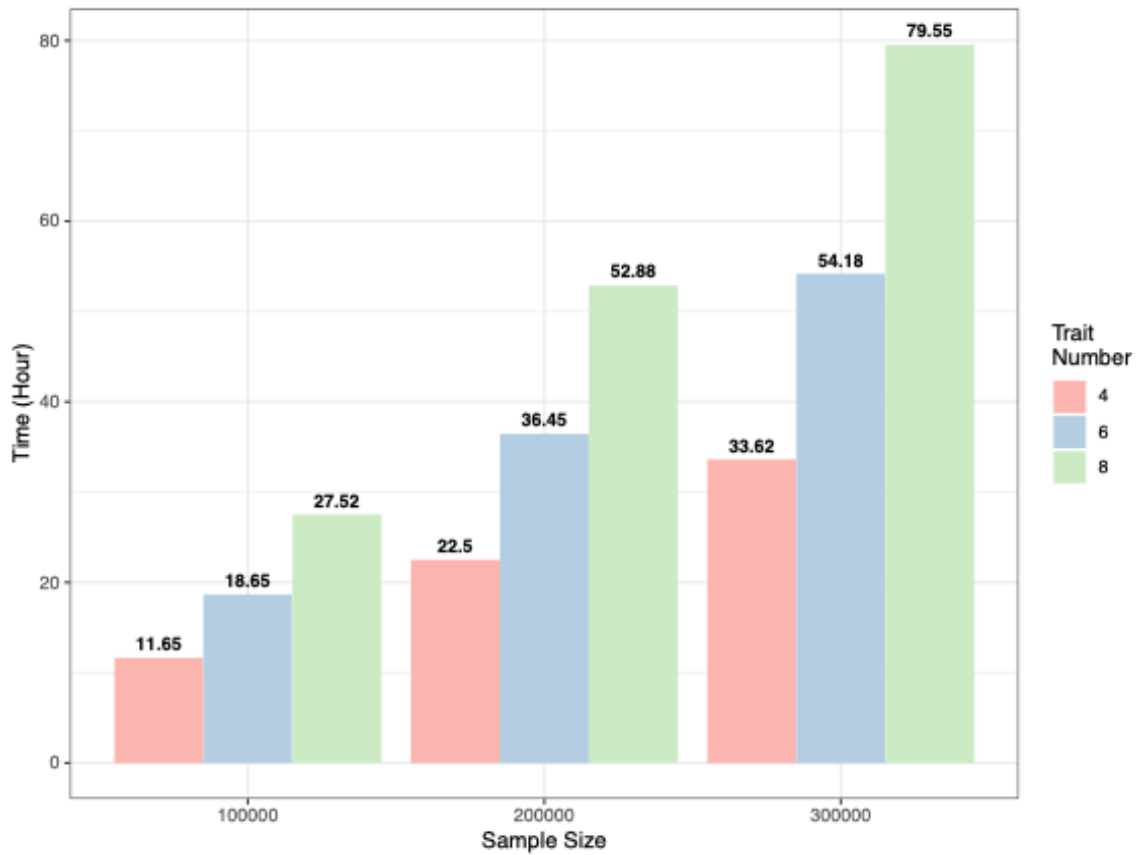
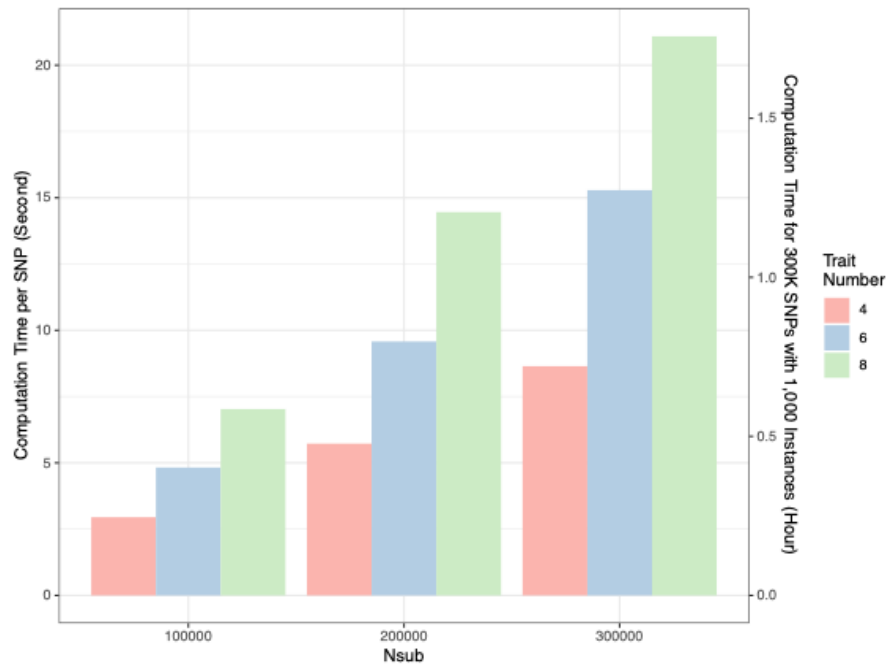


Figure S9. Computational performance of SCAMPI for analyzing 6,000,000 SNPs

Number of hours it will take SCAMPI to analyze 6,000,000 SNPs for different sample sizes and number of traits assuming 1,000 job instances by using High-Performance Computing (HPC) cluster hosted by Emory University Rollins School of Public Health (RSPH). Like Figure 4, the computational run time is benchmarked based on the average time of 1,000 simulations for different scenarios with varying trait number and sample sizes on RSPH HPC.

Figure S10.

(a)



(b)

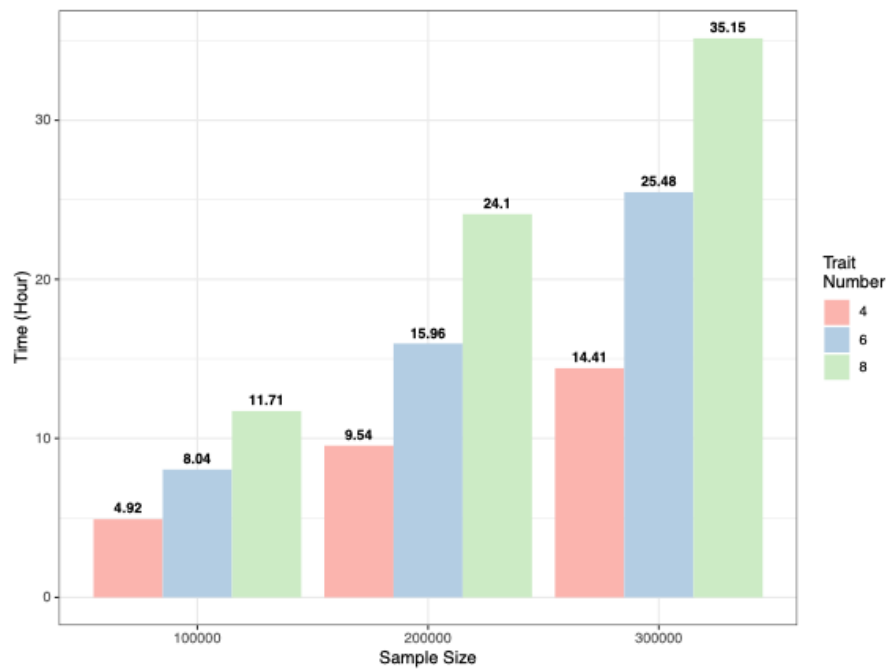


Figure S10. Computational performance of SCAMPI estimated using Apple M1 chip

Computational run time of SCAMPI for different sample sizes and number of traits using local MacBook Pro with Apple M1 chip. Computational run time is based on average of 1,000 simulations for different scenarios with varying trait number and sample size. In **Figure S10 (a)**, the first y-axis displays the time in seconds to complete SCAMPI for one SNP at different configurations. The second y-axis shows hours required to complete analyzing 300,000 SNPs assuming 1,000 job instances on a high-performance cluster. **Figure S10 (b)** displays the hours required to complete analyzing 6,000,000 SNPs assuming 1,000 job instances on a high-performance cluster.

Reference

1. Kempthorne, O., Bancroft, T.A., Gowen, J.W., and Lush, J.L. (1954). Statistics and mathematics in biology. *Statistics and mathematics in biology*.
2. Payton, M.E., Greenstone, M.H., and Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science* 3. 10.1093/jis/3.1.34.