Supplementary Information

List of Supplementary Figures

- S1 **Distribution of the number of SNPs in protein-coding genes**. (A) The distribution of the number of array SNPs; the red line indicates the 50 SNP cutoff used for real data testing. (B) The distribution of the number of imputed SNPs. In both (A) and (B), we limited the display of gene sizes to the 99th percentile for visualization purposes.

- S2 **Calibration of QuadKAST.** (A) We simulated phenotypes under four genetic architectures with additive effects only (N = 50K individuals; details on the simulations are provided in the main text). We then applied QuadKAST to each of $9,515$ protein-coding genes while employing a quadratic kernel that includes self-interactions. (B) We applied QuadKAST using a kernel that does not consider self-interactions. (C) We repeated our test of calibration on 200 phenotypes simulated under a linear additive architecture (ALL, Causal ratio = 0.001) to account for higher p-value precision with a kernel that (C) does and (D) does not include self-interactions respectively.

- S3 **Calibration of QuadKAST under model misspecification.** (A) We considered a setting of heteroskedastic noise contributing to the phenotype. We assume that there are two latent groups of individuals across which the environmental noise has differing variances. We simulated phenotypes under a linear additive architecture (ALL, Causal ratio = 0.001) to which we added additional environmental noise that depends on the group membership. We set the additive variance component as $\sigma^2_g = 0.3$ and the environmental variances in each of the two groups as 0.3 and 0.7. (B) *large effect missing variant*. In this setting, we simulated phenotypes assuming a linear additive architecture (ALL, Causal ratio = 0.001). We additionally assign a large heritability (denoted σ^2_{hid}) to about 50 imputed SNPs. We then applied QuadKAST to SNPs typed on the UKB array so that on average, over 90% of the large effect causal variants are unobserved from the analyses. We set the parameters as ($\sigma^2_g = 0.3$, $\sigma^2_{hom} = 0.5$, $\sigma^2_{hid} = 0.2$). For each setting, we randomly simulated 5 traits and ran QuadKAST across $9,515$ protein-encoding genes in the UKB unrelated White-British individuals ($N \approx 300,000$).

- S4 **Variance component estimation on UKB data**. We vary each of the following parameters: the number of individuals (N), SNPs (M), and the quadratic heritability (σ^2_{quad}), while keeping other quantities fixed. The default values are ($N = 291,273$, $M = 25$, $\sigma^2_{quad} = 0.01$). For each setting, we performed the analysis with 100 replicates. The analysis of varying quadratic heritability (σ^2_{quad}) is conducted under a fixed region.

- S5 **Accuracy of variance component estimation** The simulation and estimation of the quadratic variance component σ^2_{quad} is performed under a kernel that (A) does and (B) does not include self-interaction of the variants in the target set. (C) and (D) report the relative error $\left(\frac{\sigma^2_{quad} - \hat{\sigma}^2_{quad}}{\sigma^2_{quad}}\right)$. Each setting is repeated 100 times, with the input of each replication conducted using a set of SNPs from the genotype array data, selected from a randomly chosen gene region of protein-encoding genes.

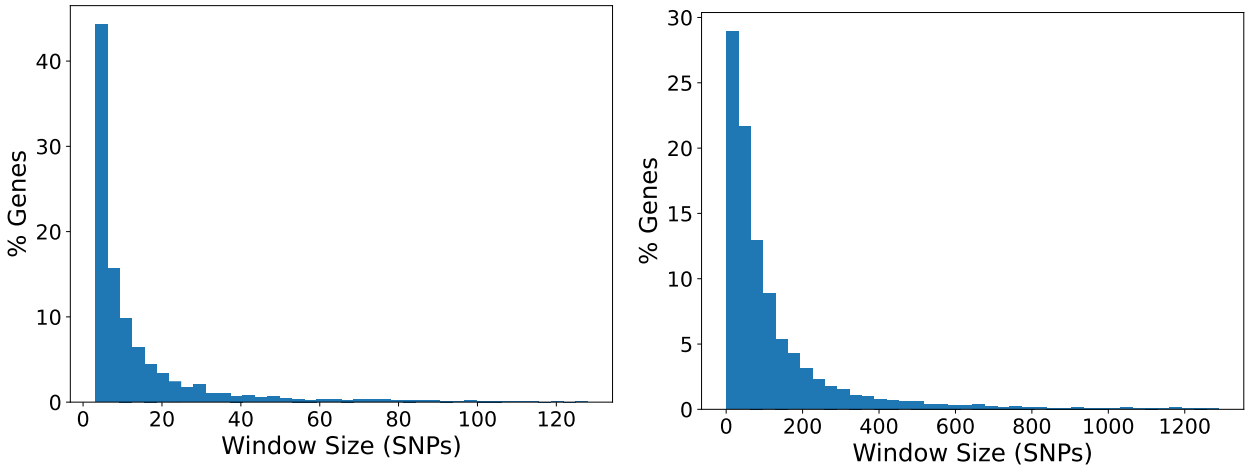
- S6 **Power analysis of QuadKAST on simulated data.** We randomly selected 5K individuals and $1,000$ protein-coding genes from the UKB to simulate phenotypes with (A) and without self-interaction effects (B). For each set, we varied the quadratic heritability and computed the power of QuadKAST at different p-value thresholds.

- S7 **Runtime and memory analysis.** Runtime (A) and memory (B) of QuadKAST and SKAT Quadratic kernel option (SKAT_QUAD) as we vary the number of individuals while analyzing a set containing 100 SNPs. (C) Runtime of QuadKAST as we vary the number of SNPs within a set. The number of individuals was fixed at 50K. All results were obtained by averaging across 10 replicates.

- S8 **Analysis of the contribution of each interaction for significant trait-gene pairs**. The axes denotes the SNP index within a set. The color indicates the strength of the interaction as quantified

by its posterior mean. The cells with interaction importance score passing the Bonferroni correction threshold $0.05/D$ were denoted with a yellow star, where D indicates the set of interactive features in a given gene.

Supplementary figures



A. UKBB array data

B. UKBB imputed data

Figure S1: **Distribution of the number of SNPs in protein-coding genes**. (A) The distribution of the number of array SNPs; the red line indicates the 50 SNP cutoff used for real data testing. (B) The distribution of the number of imputed SNPs. In both (A) and (B), we limited the display of gene sizes to the 99th percentile for visualization purposes.

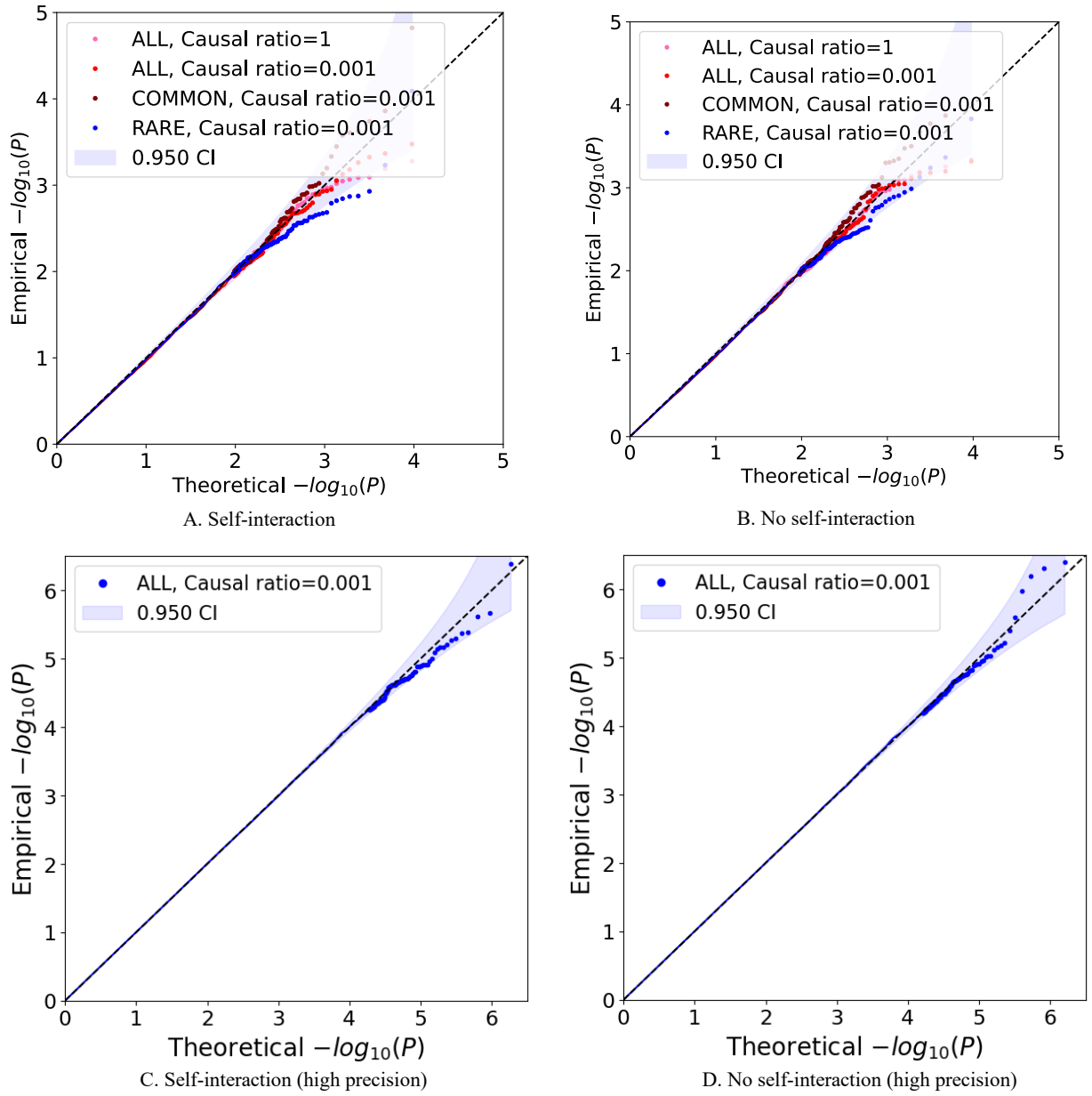


Figure S2: **Calibration of QuadKAST.** (A) We simulated phenotypes under four genetic architectures with additive effects only (N = 50K individuals; details on the simulations are provided in the main text). We then applied QuadKAST to each of 9,515 protein-coding genes while employing a quadratic kernel that includes self-interactions. (B) We applied QuadKAST using a kernel that does not consider self-interactions. (C) We repeated our test of calibration on 200 phenotypes simulated under a linear additive architecture (ALL, Causal ratio = 0.001) to account for higher p-value precision with a kernel that (C) does and (D) does not include self-interactions respectively.

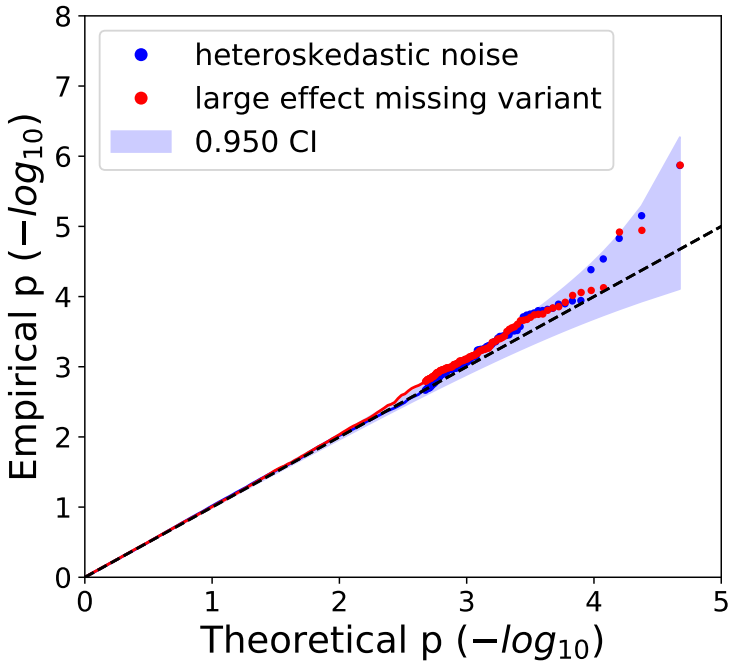
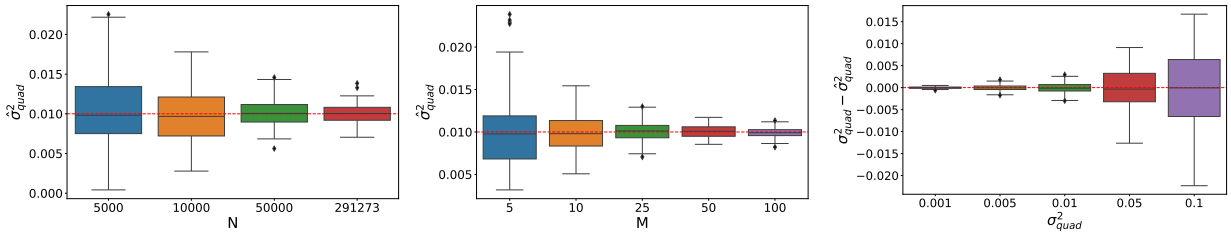


Figure S3: **Calibration of QuadKAST under model misspecification.** A. We considered a setting of heteroskedastic noise contributing to the phenotype. We assume that there are two latent groups of individuals across which the environmental noise has differing variances. We simulated phenotypes under a linear additive architecture (ALL, Causal ratio = 0.001) to which we added additional environmental noise that depends on the group membership. We set the additive variance component as $\sigma_g^2 = 0.3$ and the environmental variances in each of the two groups as 0.3 and 0.7. B. *large effect missing variant.* In this setting, we simulated phenotypes assuming a linear additive architecture (ALL, Causal ratio = 0.001). We additionally assign a large heritability (denoted σ_{hid}^2) to about 50 imputed SNPs. We then applied QuadKAST to SNPs typed on the UKB array so that on average, over 90% of the large effect causal variants are unobserved from the analyses. We set the parameters as ($\sigma_g^2 = 0.3$, $\sigma_{hom}^2 = 0.5$, $\sigma_{hid}^2 = 0.2$). For each setting, we randomly simulated 5 traits and ran QuadKAST across $9,515$ protein-encoding genes in the UKB unrelated White-British individuals ($N \approx 300,000$).



A. Number of individuals

B. Number of SNPs

C. Quadratic heritability

Figure S4: **Variance component estimation on UKB data**. We vary each of the following parameters: the number of individuals (N), SNPs (M), and the quadratic heritability (σ^2_{quad}), while keeping other quantities fixed. The default values are ($N = 291,273$, $M = 25$, $\sigma^2_{quad} = 0.01$). For each setting, we performed the analysis with 100 replicates. The analysis of varying quadratic heritability (σ^2_{quad}) is conducted under a fixed region.

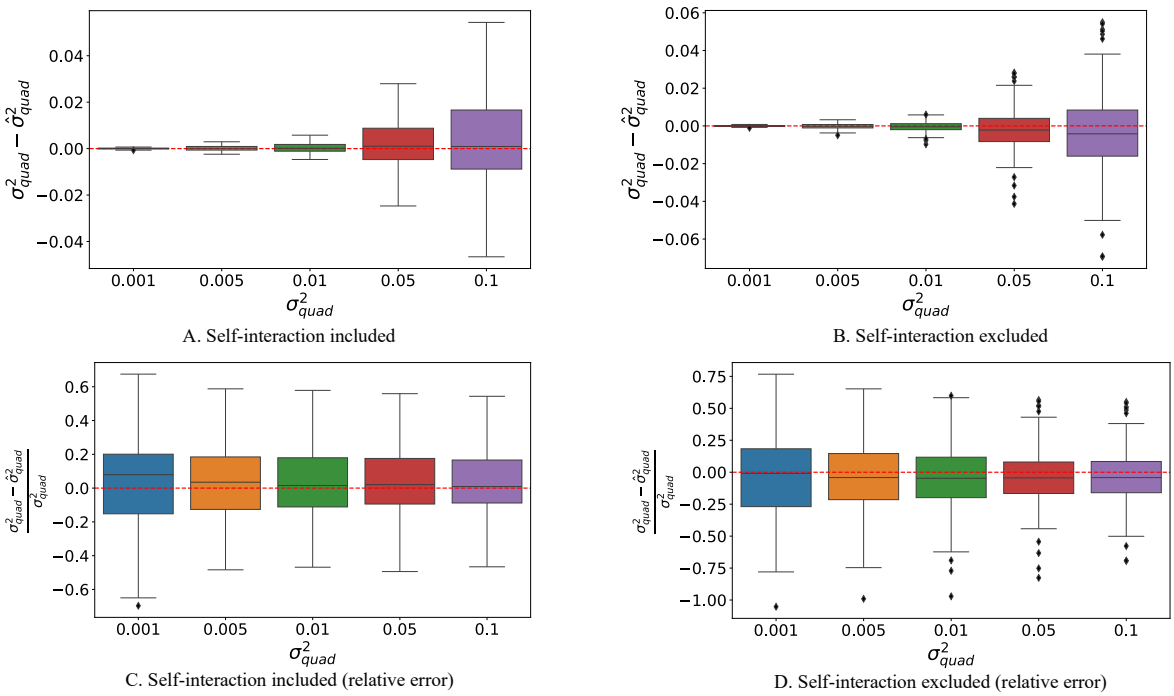
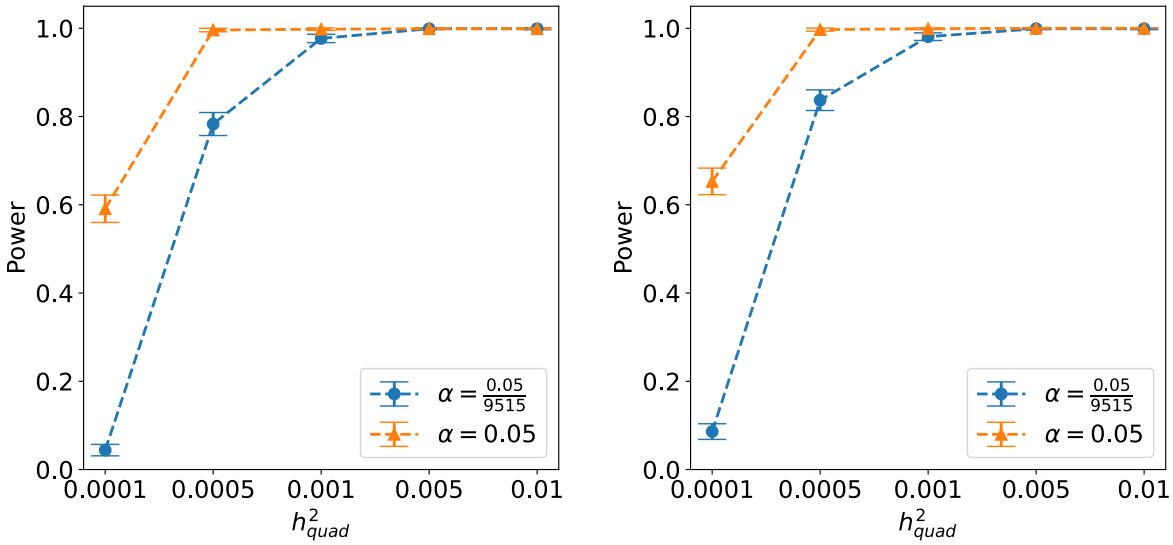


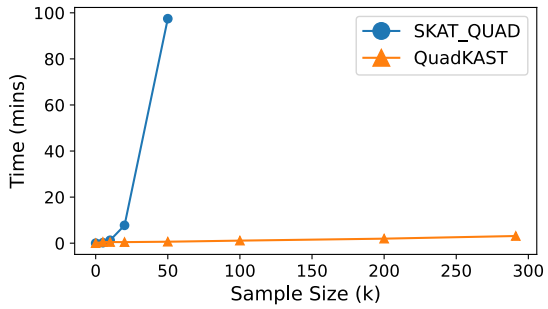
Figure S5: **Accuracy of variance component estimation** The simulation and estimation of the quadratic variance component σ^2_{quad} is performed under a kernel that (A) does and (B) does not include self-interaction of the variants in the target set. (C) and (D) report the relative error $\left(\frac{\sigma^2_{quad} - \hat{\sigma}^2_{quad}}{\sigma^2_{quad}}\right)$. Each setting is repeated 100 times, with the input of each replication conducted using a set of SNPs from the genotype array data, selected from a randomly chosen gene region of protein-encoding genes.



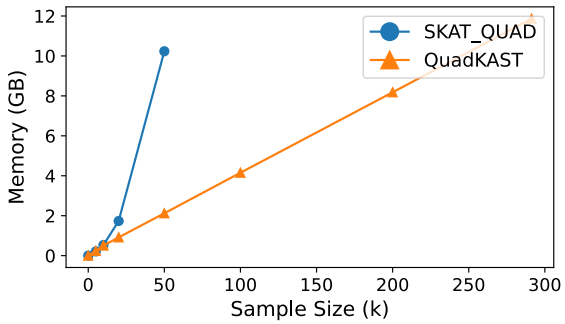
A. Self-interaction

B. No self-interaction

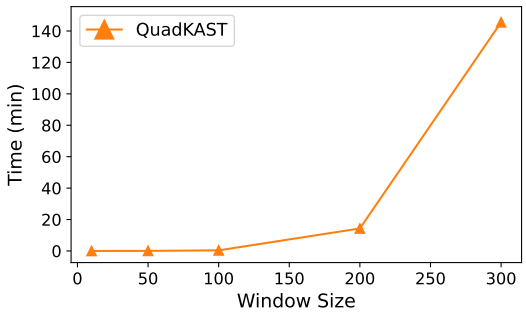
Figure S6: **Power analysis of QuadKAST on simulated data.** We randomly selected 5K individuals and 1,000 protein-coding genes from the UKB to simulate phenotypes with (A) and without self-interaction effects (B). For each set, we varied the quadratic heritability and computed the power of QuadKAST at different p-value thresholds.



A. Scaling with number of individuals



B. Scaling with memory



C. Scaling with number of SNPs

Figure S7: **Runtime and memory analysis.** Runtime (A) and memory (B) of QuadKAST and SKAT Quadratic kernel option (SKAT_QUAD) as we vary the number of individuals while analyzing a set containing 100 SNPs. (C). Runtime of QuadKAST as we vary the number of SNPs within a set. The number of individuals was fixed at 50K. All results were obtained by averaging across 10 replicates.

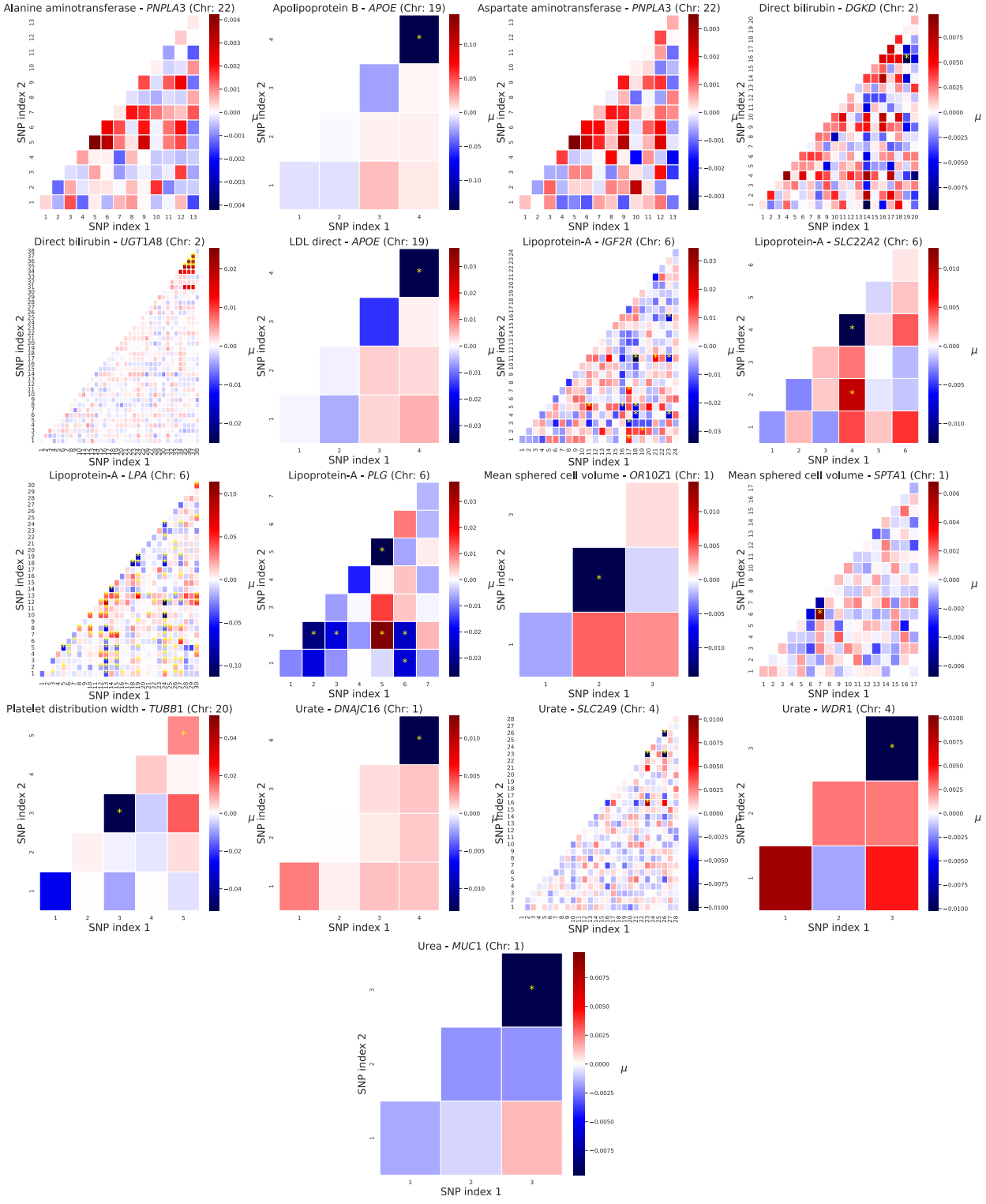


Figure S8: **Analysis of the contribution of each interaction for significant trait-gene pairs**. The axes denotes the SNP index within a set. The color indicates the strength of the interaction as quantified by its posterior mean. The cells with interaction importance score passing the Bonferroni correction threshold $0.05/D$ were denoted with a yellow star, where D indicates the set of interactive features in a given gene.

List of Supplementary Tables

- S1 **Trait category information.** This table presents the trait and the corresponding category information for all the 52 quantitative traits.
- S2 **Calibration analysis** (Self-interaction included). False positive rate for varying p-value thresholds corresponding to Supplementary Figure S2C. After filtering out genes with SNPs number less than three, we obtained 1,872,415 p-values from the simulation setting (ALL, Causal ratio = 0.001) from which we report the FPR. The p-value row describes whether the reported FPR significantly differs from the expected FPR for the given threshold. None of the p-values are significant after correcting for multiple testing suggesting that the FPR is calibrated across significance thresholds.
- S3 **Calibration analysis** (Self-interaction excluded). False positive rate for varying p-value thresholds corresponding to Supplementary Figure S2D. After filtering out genes with SNPs number less than three, we obtained 1,613,833 p-values from the simulation setting (ALL, Causal ratio = 0.001) from which we report the FPR. The p-value row describes whether the reported FPR significantly differs from the expected FPR for the given threshold. None of the p-values are significant after correcting for multiple testing suggesting that the FPR is calibrated across significance thresholds.
- S4 **Runtime as a function of the number of individuals.** Comparison of the runtime QuadKAST and SKAT Quadratic kernel option (SKAT_QUAD) with varying number of individuals. We varied the number of individuals for a set containing 100 SNPs on the UKB genotype data. We report the average runtime $\pm$ the standard deviation across 10 replicates.
- S5 **Runtime as a function of the number of SNPs.** We applied QuadKAST to a subsample of 50K individuals on the UKB genotype data. At each setting, we report the average runtime of 10 replicates.
- S6 **Epistatic and additive variance component comparison of the significant trait-gene pairs**. We quantified the variance explained by the additive effect as σ^2_g, and the variance explained by the quadratic effect after excluding the additive effect as σ^2_{quad}, and the corresponding standard error (termed $SE(\sigma^2_g)$ and $SE(\sigma^2_{quad})$) We reported the ratio of σ^2_{quad} over σ^2_g as *Ratio*
- S7 **Variance components comparison across different settings**. We tested the data with QuadKAST using three settings: default settings, using 40 PCs to control for population structure, and using an imputed dataset. We compared the estimated variance component obtained from each setting.

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

Table S1: **Trait category information.** This table presents the trait and the corresponding category information for all the 52 quantitative traits.

Threshold (α)	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
FPR ($\times\alpha$)	1.004	1.010	0.988	1.020	0.694	0.534
p-value	0.045	0.182	0.604	0.784	0.186	0.524

Table S2: **Calibration analysis** (Self-interaction included). False positive rate for varying p-value thresholds corresponding to Supplementary Figure S2C. After filtering out genes with SNPs number less than three, we obtained 1,872,415 p-values from the simulation setting (ALL, Causal ratio = 0.001) from which we report the FPR. The p-value row describes whether the reported FPR significantly differs from the expected FPR for the given threshold. None of the p-values are significant after correcting for multiple testing suggesting that the FPR is calibrated across significance thresholds.

Threshold (α)	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
FPR ($\times\alpha$)	1.005	1.008	1.004	0.979	0.806	1.859
P-value	0.026	0.287	0.878	0.790	0.435	0.275

Table S3: **Calibration analysis** (Self-interaction excluded). False positive rate for varying p-value thresholds corresponding to Supplementary Figure S2D. After filtering out genes with SNPs number less than three, we obtained 1,613,833 p-values from the simulation setting (ALL, Causal ratio = 0.001) from which we report the FPR. The p-value row describes whether the reported FPR significantly differs from the expected FPR for the given threshold. None of the p-values are significant after correcting for multiple testing suggesting that the FPR is calibrated across significance thresholds.

Kernel	Sample Size	Time (mins)
QuadKAST	5000	0.5
	10000	0.5
	20000	0.5
	50000	0.6
	100000	1.1
	200000	2.0
	291273	3.1
SKAT_QUAD	5000	0.2
	10000	1.3
	20000	7.7
	50000	97.5

Table S4: **Runtime as a function of the number of individuals.** Comparison of the runtime QuadKAST and SKAT Quadratic kernel option (SKAT_QUAD) with varying number of individuals. We varied the number of individuals for a set containing 100 SNPs on the UKB genotype data. We report the average runtime across 10 replicates.

Window Size	Time (mins)
10	≤ 0.1
50	≤ 0.1
100	0.40
200	14.3
300	146
500	489

Table S5: **Runtime as a function of the number of SNPs.** We applied QuadKAST to a subsample of 50K individuals on the UKB genotype data. At each setting, we report the average runtime of 10 replicates.

Table S6: **Epistatic and additive variance component comparison of the significant trait-gene pairs**. We quantified the variance explained by the additive effect as σ_g^2, and the variance explained by the quadratic effect after excluding the additive effect as σ_{quad}^2, and the corresponding standard error (termed $SE(\sigma_g^2)$ and $SE(\sigma_{quad}^2)$) We reported the ratio of σ_{quad}^2 over σ_g^2 as *Ratio*

Trait	Gene	σ_g^2 $\times 10^{-3}$	$SE(\sigma_g^2)$ $\times 10^{-3}$	σ_{quad}^2 $\times 10^{-3}$	$SE(\sigma_{quad}^2)$ $\times 10^{-3}$	Ratio
Alanine aminotransferase	*PNPLA3*	2.72	1.26	0.34	0.11	0.12
	SAMM50	2.34	1.18	0.17	0.08	0.07
Apolipoprotein B	*BCAM*	8.14	4.41	2.72	1.24	0.33
	APOE	64.21	45.42	22.48	10.26	0.35
Aspartate aminotransferase	*PNPLA3*	2.73	1.26	0.34	0.11	0.12
Creatinine	*DNAJC16*	0.17	0.15	0.37	0.24	2.10
Direct bilirubin	*SAG*	21.99	11.02	14.68	4.34	0.67
	DGKD	72.58	23.24	4.49	1.02	0.06
	USP40	23.30	11.28	1.12	0.35	0.05
	UGT1A8	74.98	18.98	18.18	1.87	0.24
Eosinophil count	*PRG3*	0.36	0.30	8.11	4.76	22.41
HDL cholesterol	*CETP*	17.64	6.99	0.19	0.08	0.01
Hemoglobin A1c	*HK1*	18.78	4.62	1.03	0.22	0.05
LDL direct	*APOE*	35.18	24.89	1.99	1.06	0.06
Lipoprotein-A	*IGF2R*	40.00	11.65	23.22	3.27	0.58
	SLC22A2	24.68	14.27	0.53	0.24	0.02
	SLC22A3	86.38	35.30	7.61	1.66	0.09
	LPA	546.11	141.81	183.37	14.48	0.34
	PLG	176.95	94.65	7.84	2.71	0.04
	AGPAT4	23.36	7.63	17.75	2.56	0.76
Mean corpuscular hemoglobin	*TMPRSS6*	10.15	3.51	0.70	0.23	0.07
Mean sphered cell volume	*OR10Z1*	7.66	6.26	0.31	0.21	0.04
	SPTA1	9.58	3.62	0.74	0.20	0.08
Monocyte count	*ADA2*	0.12	0.08	0.23	0.10	1.90
Platelet distribution width	*TUBB1*	21.44	13.83	5.20	2.52	0.24
	EDN3	0.85	0.37	1.17	0.29	1.37
SHBG	*TNK1*	1.53	1.50	0.28	0.20	0.18
Urate	*DNAJC16*	0.25	0.21	0.38	0.26	1.52
	SLC2A9	23.76	6.38	2.17	0.28	0.09
	WDR1	9.61	7.86	0.28	0.20	0.03
	MEPE	0.38	0.28	0.28	0.16	0.73
Urea	*MUC1*	0.82	0.68	0.17	0.12	0.20

Table S7: **Variance components comparison across different settings**. We tested the data with QuadKAST using three settings: default settings, using 40 PCs to control for population structure, and using an imputed dataset. We compared the estimated variance component obtained from each setting.

Trait	Gene	σ^2_{quad} $\times 10^{-3}$	σ^2_{quad} (40 PC) $\times 10^{-3}$	σ^2_{quad} (Imputed) $\times 10^{-3}$
Alanine aminotransferase	*PNPLA3*	0.34	0.34	0.57
	SAMM50	0.17	0.17	0.52
Apolipoprotein B	*BCAM*	2.72	2.73	21.82
	APOE	22.48	22.41	22.98
Aspartate aminotransferase	*PNPLA3*	0.34	0.34	0.48
Creatinine	*DNAJC16*	0.37	0.37	0.14
Direct bilirubin	*SAG*	14.68	14.69	9.68
	DGKD	4.49	4.51	16.43
	USP40	1.12	1.13	21.25
	UGT1A8	18.18	18.16	0.00
Eosinophil count	*PRG3*	8.11	8.11	13.98
HDL cholesterol	*CETP*	0.19	0.19	0.30
Hemoglobin A1c	*HK1*	1.03	1.03	0.55
LDL direct	*APOE*	1.99	1.98	2.10
Lipoprotein-A	*IGF2R*	23.22	23.20	15.45
	SLC22A2	0.53	0.54	31.10
	SLC22A3	7.61	7.62	70.18
	LPA	183.37	183.26	792.13
	PLG	7.84	7.86	179.78
	AGPAT4	17.75	17.77	38.29
Mean corpuscular hemoglobin	*TMPRSS6*	0.70	0.70	0.34
Mean sphered cell volume	*OR10Z1*	0.31	0.31	0.32
	SPTA1	0.74	0.74	0.89
Monocyte count	*ADA2*	0.23	0.23	1.71
Platelet distribution width	*TUBB1*	5.20	5.21	4.91
	EDN3	1.17	1.17	0.00
SHBG	*TNK1*	0.28	0.28	1.04
Urate	*DNAJC16*	0.38	0.37	0.23
	SLC2A9	2.17	2.18	4.77
	WDR1	0.28	0.28	4.77
	MEPE	0.28	0.28	0.05
Urea	*MUC1*	0.17	0.17	0.24

S1 Supplementary Notes

S1.1 Score statistics

Under the assumption that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\alpha}, \sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I})$, we can write down the log-likelihood as

$$\ell(\boldsymbol{\alpha}, \sigma^2_{quad}, \sigma^2_\epsilon) = -\frac{1}{2}[\log|\sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I}| + (\boldsymbol{y} - \boldsymbol{X\alpha})^T(\sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I})^{-1}(\boldsymbol{y} - \boldsymbol{X\alpha})] + Const$$

Under the null hypothesis $\sigma^2_{quad} = 0$, the constrained MLE of $(\boldsymbol{\alpha}, \sigma^2_\epsilon)$:

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{X^T X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$
$$\hat{\sigma}^2_\epsilon = \frac{\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}}{N-K}$$

We can further write down the partial derivative of ℓ with respect to σ^2_{quad} as

$$\frac{\partial \ell(\boldsymbol{\alpha}, \sigma^2_{quad}, \sigma^2_\epsilon)}{\sigma^2_{quad}} = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\alpha})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{K}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\alpha}) - \frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{K})$$

where $\boldsymbol{\Sigma} = \sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I}$.
Plugging in the parameter estimates under the null hypothesis:

$$\begin{aligned}
\frac{\partial \ell(\hat{\boldsymbol{\alpha}}, \sigma^2_{quad}, \hat{\sigma}^2_\epsilon)}{\partial \sigma^2_{quad}}\Big|_{\sigma^2_{quad}=0} &= \frac{1}{2\hat{\sigma}^4_\epsilon}(\boldsymbol{y} - \boldsymbol{X\hat{\alpha}})^T\boldsymbol{K}(\boldsymbol{y} - \boldsymbol{X\hat{\alpha}}) - \frac{1}{2\hat{\sigma}^2_\epsilon}tr(\boldsymbol{K}) \\
&= \frac{1}{2\hat{\sigma}^4_\epsilon}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X}^T\boldsymbol{y})^T\boldsymbol{K}(\boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X^T X})^{-1}\boldsymbol{X}^T\boldsymbol{y}) - \frac{1}{2\hat{\sigma}^2_\epsilon}tr(\boldsymbol{K}) \\
&= \frac{1}{2\hat{\sigma}^4_\epsilon}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{K}\boldsymbol{P}\boldsymbol{y} - \frac{1}{2\hat{\sigma}^2_\epsilon}tr(\boldsymbol{K})
\end{aligned}$$

This leads to the score test statistic that is used in SKAT (Wu et al. 2011): $Q = \frac{1}{\hat{\sigma}^2_\epsilon}\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{K}\boldsymbol{P}\boldsymbol{y}$. An alternative score statistic for which the exact sampling distribution under the null hypothesis can be computed has been proposed in (Zhou et al. 2016).

S1.1.1 Sampling distribution of the score statistic

For completeness, we derive the sampling distribution of the score statistic, following closely the treatment in (Zhou et al. 2016).

Let $\mathcal{C}()$ denote the column space of a matrix. Let R and S denote the ranks of $\boldsymbol{X}$ and $\boldsymbol{PKP}$ respectively. Let $\boldsymbol{A}_0 \in \mathbb{R}^{N\times R}$ denote an orthonormal basis of $\mathcal{C}(\boldsymbol{X})$ and $\boldsymbol{A}_1 \in \mathbb{R}^{N\times S}$ denote an orthonormal basis based on an eigendecomposition of $\boldsymbol{PKP}$. $\boldsymbol{A}_2 \in \mathbb{R}^{N\times(N-R-S)}$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{A}_0, \boldsymbol{A}_1)^\perp = \mathcal{C}(\boldsymbol{X}, \boldsymbol{A}_1)^\perp$ so that $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2] \in \mathbb{R}^{N\times(N-R)}$ is an orthonormal basis of $\mathcal{C}(\boldsymbol{X})^\perp$.
Using the uniqueness of orthogonal projections, we have $\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$. Under the null hypothesis:

$$\begin{aligned}
\boldsymbol{w} &:= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \\
&\sim \mathcal{N}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X\alpha}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{I}_N\sigma^2_\epsilon\boldsymbol{A}) \\
&\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{N-R}\sigma^2_\epsilon)
\end{aligned} \tag{7}$$

We also have:

$$\begin{aligned}
\boldsymbol{PKP} &= \boldsymbol{A}_1 diag(\lambda_1, \ldots, \lambda_S)\boldsymbol{A}_1^{\mathrm{T}} \\
&= [\boldsymbol{A}_1, \boldsymbol{A}_2] diag(\lambda_1, \ldots, \lambda_S, 0, \ldots, 0)[\boldsymbol{A}_1, \boldsymbol{A}_2]^{\mathrm{T}} \\
&= \boldsymbol{A} diag(\lambda_1, \ldots, \lambda_S, 0, \ldots, 0)\boldsymbol{A}^{\mathrm{T}}
\end{aligned} \tag{8}$$

Here $(\lambda_1, \ldots, \lambda_S)$ are the non-zero eigenvalues of $\boldsymbol{PKP}$.

Combining Equations 7 and 8:

$$\begin{aligned} \boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{K}\boldsymbol{P}\boldsymbol{y} &= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}diag(\lambda_1,\ldots,\lambda_S,0,\ldots,0)\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \\ &= \boldsymbol{w}^{\mathrm{T}}diag(\lambda_1,\ldots,\lambda_S,0,\ldots,0)\boldsymbol{w} \\ &= \sigma_\epsilon^2\sum_{i=1}^{S}\lambda_i\chi_i^2 \end{aligned} \tag{9}$$

Here χ_i^2 are independent random variables with a χ^2, 1-df distribution.
Since $\hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2$, we then have

$$Q \xrightarrow{d} \sum_{i=1}^{S}\lambda_i\chi_i^2 \tag{10}$$

S1.1.2 Computing the p-value of the score statistic

To compute the score statistic, we write the statistic as: $Q = \frac{1}{D\hat{\sigma}_\epsilon^2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{y}$. The statistic can be computed in $\mathcal{O}(ND)$ time. Computing the p-value using the sampling distribution in Equation 9 requires computing the eigenvalues of $\boldsymbol{P}\boldsymbol{K}\boldsymbol{P}$. The eigenvalues are obtained as the squared singular values of $\boldsymbol{P}\boldsymbol{\Phi} = \boldsymbol{\Phi} - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Phi}$. The matrix $\boldsymbol{P}\boldsymbol{\Phi}$ can be formed in $\mathcal{O}(NKD + NK^2 + K^3)$ times and its singular values can be computed in $\mathcal{O}(ND^2)$ time (for $D < N$).

S1.2 Variance components estimation

When no covariates are included, we have $\boldsymbol{y} \sim \mathcal{N}(0, \sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I})$. We can estimate $(\sigma^2_{quad}, \sigma^2_\epsilon)$ by maximizing the log-likelihood:

$$\ell(\sigma^2_{quad}, \sigma^2_\epsilon) = -\frac{1}{2}\left[\log|\sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I}| + \boldsymbol{y}^T(\sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I})^{-1}\boldsymbol{y}\right] + Const$$

Defining $\delta = \frac{\sigma^2_\epsilon}{\sigma^2_{quad}}$, the log-likelihood can be re-parameterized as:

$$\ell(\sigma^2_{quad}, \delta) = -\frac{1}{2}\left[\log|\sigma^2_{quad}(\boldsymbol{K} + \delta \boldsymbol{I})| + \boldsymbol{y}^T(\sigma^2_{quad}(\boldsymbol{K} + \delta \boldsymbol{I}))^{-1}\boldsymbol{y}\right] + Const$$

Computing the partial derivative w.r.t. σ^2_{quad}, we get

$$\frac{\partial \ell(\sigma^2_{quad}, \delta)}{\partial \sigma^2_{quad}} = -\frac{1}{2}[\frac{N}{\sigma^2_{quad}} - \frac{\boldsymbol{y}^T(\boldsymbol{K} + \delta \boldsymbol{I})^{-1}\boldsymbol{y}}{\sigma^4_{quad}}] = 0$$
$$\sigma^2_{quad} = \frac{\boldsymbol{y}^T(\boldsymbol{K} + \delta \boldsymbol{I})^{-1}\boldsymbol{y}}{N}$$

Substituting σ^2_{quad} into the log-likelihood function, we get the profile log-likelihood:

$$\ell_P(\delta) = -\frac{1}{2}\left[N \log \frac{1}{N}\boldsymbol{y}^T(\boldsymbol{K} + \delta \boldsymbol{I})^{-1}\boldsymbol{y} + \log|(\boldsymbol{K} + \delta \boldsymbol{I})|\right] + Const \tag{11}$$

Given the eigen-decomposition of $\boldsymbol{K} = \boldsymbol{U} diag(\rho_1, \ldots, \rho_N)\boldsymbol{U}^T$ where $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ is the matrix of eigenvectors and $(\rho_1, \ldots, \rho_N)$ are the eigenvalues of $\boldsymbol{K}$, Equation 11 can be rewritten as

$$\begin{aligned}
\ell_P(\delta) &= -\frac{1}{2}\left[N \log\left(\frac{1}{N}\boldsymbol{y}^T(\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^T + \delta \boldsymbol{I})^{-1}\boldsymbol{y}\right) + \log|(\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^T + \delta \boldsymbol{I})|\right] + Const \\
&= -\frac{1}{2}\left[N \log\left(\frac{1}{N}\boldsymbol{y}^T(\boldsymbol{U}(\boldsymbol{S} + \delta \boldsymbol{I})\boldsymbol{U}^T)^{-1}\boldsymbol{y}\right) + \log|\boldsymbol{U}(\boldsymbol{S} + \delta \boldsymbol{I})\boldsymbol{U}^T|\right] + Const \\
&= -\frac{1}{2}\left[N \log\left(\frac{1}{N}\left(\sum_{i=1}^{N}\frac{\tilde{y}_i^2}{\rho_i + \delta}\right)\right) + \sum_{i=1}^{N}\log(s_i + \delta)\right] + Const
\end{aligned} \tag{12}$$

Here $\tilde{y}_i$ is the i^{th} entry of $\tilde{\boldsymbol{y}} = \boldsymbol{U}^T\boldsymbol{y}$. Evaluating this function takes $\mathcal{O}(N)$ complexity at each iteration of the optimization algorithm.

The above derivation assumes that $\boldsymbol{K}$ is a full rank matrix. However, the rank of $\boldsymbol{K}$ can often be substantially smaller than N leading to additional efficiency in evaluating an optimizing ℓ_P. In our application, $\boldsymbol{K} = \frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^T}{D}$ where $\boldsymbol{\Phi} \in \mathbb{R}^{N \times D}$ so that the rank of $\boldsymbol{K}$: $R \leq min(D, N) = D$ when $D < N$. In this setting, we have $\rho_i = 0$ for $i > R$ and we can write $\boldsymbol{U} = [\boldsymbol{U}_1, \boldsymbol{U}_2]$, where $\boldsymbol{U}_1 \in \mathbb{R}^{N \times R}$ and $\boldsymbol{U}_2 \in \mathbb{R}^{N \times (N-R)}$ so that:

$$\tilde{y}_i = \begin{cases} [\boldsymbol{U}_1^T\boldsymbol{y}]_i, i \leq R \\ [\boldsymbol{U}_2^T\boldsymbol{y}]_i, i > R \end{cases}$$

We then can rewrite the profile log-likelihood as:

$$
\begin{aligned}
\ell_P(\delta) &= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(\sum_{i=1}^{N}\frac{\tilde{y}_i^2}{\rho_i+\delta}\right)\right)+\sum_{i=1}^{N}\log(\rho_i+\delta)\right]+Const \\
&= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(\sum_{i=1}^{R}\frac{\tilde{y}_i^2}{\rho_i+\delta}+\sum_{i=R+1}^{N}\frac{\tilde{y}_i^2}{\delta}\right)\right)+\sum_{i=1}^{P}\log(\rho_i+\delta)+(N-R)\log(\delta)\right]+Const \\
&= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(\sum_{i=1}^{R}\frac{\tilde{y}_i^2}{\rho_i+\delta}+\frac{1}{\delta}(\sum_{i=1}^{N}\tilde{y}_i^2-\sum_{i=1}^{R}\tilde{y}_i^2)\right)\right)+\sum_{i=1}^{P}\log(\rho_i+\delta)+(N-R)\log(\delta)\right]+Const \\
&= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(-\sum_{i=1}^{R}\frac{\rho_i\tilde{y}_i^2}{(\rho_i+\delta)\delta}+\frac{1}{\delta}\sum_{i=1}^{N}\tilde{y}_i^2\right)\right)+\sum_{i=1}^{P}\log(\rho_i+\delta)+(N-R)\log(\delta)\right]+Const \\
&= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(-\sum_{i=1}^{R}\frac{\rho_i\tilde{y}_i^2}{(\rho_i+\delta)\delta}+\frac{1}{\delta}\|\tilde{\boldsymbol{y}}\|_2^2\right)\right)+\sum_{i=1}^{P}\log(\rho_i+\delta)+(N-R)\log(\delta)\right]+Const \\
&= -\frac{1}{2}\left[N\log\left(\frac{1}{N}\left(-\sum_{i=1}^{R}\frac{\rho_i\tilde{y}_i^2}{(\rho_i+\delta)\delta}+\frac{1}{\delta}\|\boldsymbol{y}\|_2^2\right)\right)+\sum_{i=1}^{P}\log(\rho_i+\delta)+(N-R)\log(\delta)\right]+Const
\end{aligned}
$$

Evaluating ℓ_P in this setting requires computing $\tilde{y}_i$ and ρ_i, $i \leq R$ which can be obtained by a one-time computation of the R non-zero eigenvalues and corresponding eigenvectors of $\boldsymbol{K}$. Computation of these eigenvalues and eigenvectors can be obtained in $\mathcal{O}(ND^2)$ time from a SVD of $\boldsymbol{\Phi}$) while subsequent evaluation of ℓ_P requires $\mathcal{O}(D)$ time.

S1.3 Variance component estimation with covariates included

When covariates are included in the model, we have: $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{X\alpha}, \sigma^2_{quad}\boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I})$. Let L and S denote the ranks of $\boldsymbol{X}$ and $\boldsymbol{PKP}$ respectively. Let the full SVD of $\boldsymbol{X} = \boldsymbol{B}_0\boldsymbol{\Sigma}\boldsymbol{C}_0^T$ where $\boldsymbol{B}_0 \in \mathbb{R}^{N\times N}$, $\boldsymbol{C}_0 \in \mathbb{R}^{K\times K}$, and $\boldsymbol{\Sigma}$ is an $N \times K$ rectangular diagonal matrix such that

$$\boldsymbol{\Sigma}_{ij} = \begin{cases} s_i & \text{If } i = j \text{ and } i \leq L \\ 0 & \text{Otherwise} \end{cases}$$

Let $\boldsymbol{B}_0 = [\boldsymbol{B}_1, \boldsymbol{B}_2]$ where $\boldsymbol{B}_1 \in \mathbb{R}^{N\times L}$ and $\boldsymbol{B}_2 \in \mathbb{R}^{N\times(N-L)}$. Similarly, we define $\boldsymbol{C}_1 \in \mathbb{R}^{K\times L}$ and $\boldsymbol{C}_2 \in \mathbb{R}^{K\times(K-L)}$. We then have $\boldsymbol{X} = \boldsymbol{B}_1 diag(s_1, .., s_R)\boldsymbol{C}_1^T$. Together, we have

$$\boldsymbol{P} = \boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{B}_0\boldsymbol{B}_0^T - \boldsymbol{B}_1\boldsymbol{B}_1^T = \boldsymbol{B}_2\boldsymbol{B}_2^T$$

where $\boldsymbol{B}_2$ denotes an orthonormal basis of $\mathcal{C}(\boldsymbol{X})^\perp$ and $\mathcal{C}(\boldsymbol{X})$ is the column space of $\boldsymbol{X}$.

We then define:

$$\begin{aligned} \boldsymbol{\omega} &:= \boldsymbol{B}_2^T\boldsymbol{y} \\ &\sim \mathcal{N}(\boldsymbol{B}_2^T\boldsymbol{X\alpha}, \sigma^2_{quad}\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2 + \sigma^2_\epsilon\boldsymbol{B}_2^T\boldsymbol{I}_N\boldsymbol{B}_2) \\ &\sim \mathcal{N}(\boldsymbol{0}_{N-L}, \sigma^2_{quad}\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2 + \sigma^2_\epsilon\boldsymbol{I}_{N-L}) \end{aligned}$$

Let the eigendecomposition of $\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2 = \boldsymbol{U}_b diag(s_1, \ldots, s_{N-L})\boldsymbol{U}_b^T$ where $\boldsymbol{U}_b$ is a $(N-L)\times(N-L)$ orthonormal matrix of eigenvectors of $\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2$. We then define

$$\begin{aligned} \tilde{\boldsymbol{y}} &:= \boldsymbol{U}_b^T\boldsymbol{\omega} = \boldsymbol{U}_b^T\boldsymbol{B}_2^T\boldsymbol{y} \\ &\sim \mathcal{N}(\boldsymbol{0}_{N-L}, \sigma^2_{quad}diag(s_1, \ldots, s_{N-L}) + \sigma^2_\epsilon\boldsymbol{I}_{N-L}) \end{aligned} \tag{13}$$

From Equation 13, the restricted log-likelihood function can be written as:

$$\ell_R(\sigma^2_{quad}, \sigma^2_\epsilon) = -\frac{N-L}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N-L}\log\left(s_i\sigma^2_{quad} + \sigma^2_\epsilon\right) - \frac{1}{2}\sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{s_i\sigma^2_{quad} + \sigma^2_\epsilon} \tag{14}$$

which, in turn, yields the profile restricted log-likelihood (as a function of $\delta = \frac{\sigma^2_\epsilon}{\sigma^2_{quad}}$):

$$\ell_{PR}(\delta) = -\frac{1}{2}[(N-L)\log\frac{1}{N-L}\left(\sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{s_i+\delta}\right) + \sum_{i=1}^{N-L}\log(s_i+\delta)] + Const \tag{15}$$

The values of σ^2_{quad} and σ^2_ϵ that maximize the restricted log-likelihood can be efficiently computed using the same technique as described in Supplementary Note S1.2 once the eigendecomposition of $\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2$ has been computed.

S1.3.1 Efficient computation

Using the fact that we have $\boldsymbol{P} = \boldsymbol{B}_2\boldsymbol{B}_2^T$, we write

$$\begin{aligned} \boldsymbol{PKP} &= \boldsymbol{B}_2\boldsymbol{B}_2^T\boldsymbol{K}\boldsymbol{B}_2\boldsymbol{B}_2^T \\ &= \boldsymbol{B}_2\boldsymbol{U}_b diag(s_1, \ldots, s_{N-L})\boldsymbol{U}_b^T\boldsymbol{B}_2^T \\ &= \boldsymbol{B} diag(s_1, \ldots, s_{N-L})\boldsymbol{B}^T \end{aligned}$$

Where $\boldsymbol{B} = \boldsymbol{B}_2\boldsymbol{U}_b$ is a $N \times (N-L)$ matrix with orthonormal columns

$$\begin{aligned} \boldsymbol{B}^T\boldsymbol{B} &= \boldsymbol{U}_b^T\boldsymbol{B}_2^T\boldsymbol{B}_2\boldsymbol{U}_b \\ &= \boldsymbol{U}_b\boldsymbol{I}_{N-L}\boldsymbol{U}_b^T = \boldsymbol{U}_b\boldsymbol{U}_b^T = \boldsymbol{I}_{N-L} \end{aligned}$$

Thus, $\boldsymbol{B}$ is the matrix of eigenvectors of $\boldsymbol{PKP}$ while $s_1, \ldots, s_{N-L}$ are its eigenvalues. We can identify the columns of $\boldsymbol{B}$ with those of $\boldsymbol{A}$ in Equation 8, we have

$$\boldsymbol{B} = \boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2] \tag{16}$$

The eigenvalues of $\boldsymbol{B}_2^{\mathrm{T}}\boldsymbol{K}\boldsymbol{B}_2$ are equivalent to those of $\boldsymbol{P}\boldsymbol{K}\boldsymbol{P}$. The number of non-zero eigenvalues of $\boldsymbol{B}_2^{\mathrm{T}}\boldsymbol{K}\boldsymbol{B}_2$ is $S \leq min(R, N-L)$ so that $s_i = 0$ for $i > S$ and $s_i = \lambda_i$ for $i \leq S$. We also have the relation: $\tilde{\boldsymbol{y}} = \boldsymbol{B}^T\boldsymbol{y}$.

We can now rewrite the restricted log-likelihood from Equation 14 in this setting as:

$$\begin{aligned}
\ell_R(\sigma^2_{quad}, \sigma^2_\epsilon) &= -\frac{1}{2}\left[\sum_{i=1}^{N-L}\log\left(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon\right) + \sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{\rho_i\sigma^2_{quad}+\sigma^2_\epsilon}\right] + Const \\
&= -\frac{1}{2}\left[\sum_{i=1}^{S}\log\left(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon\right) + (N-L-S)\log\left(\sigma^2_{quad}+\sigma^2_\epsilon\right)\right. \\
&\quad \left. + \sum_{i=1}^{S}\frac{\tilde{y}_i^2}{\rho_i\sigma^2_{quad}+\sigma^2_\epsilon} + \sum_{i=S+1}^{N-L}\frac{\tilde{y}_i^2}{\sigma^2_{quad}+\sigma^2_\epsilon}\right] + Const \\
&= -\frac{1}{2}\left[\sum_{i=1}^{S}\log\left(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon\right) + (N-L-S)\log\left(\sigma^2_{quad}+\sigma^2_\epsilon\right)\right. \\
&\quad \left. - \sum_{i=1}^{S}\frac{\rho_i\sigma^2_{quad}\tilde{y}_i^2}{(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon)(\sigma^2_{quad}+\sigma^2_\epsilon)} + \sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{\sigma^2_{quad}+\sigma^2_\epsilon}\right] + Const \\
&= -\frac{1}{2}\left[\sum_{i=1}^{S}\log\left(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon\right) + (N-L-S)\log\left(\sigma^2_{quad}+\sigma^2_\epsilon\right)\right. \\
&\quad \left. - \sum_{i=1}^{S}\frac{\rho_i\sigma^2_{quad}\tilde{y}_i^2}{(\rho_i\sigma^2_{quad}+\sigma^2_\epsilon)(\sigma^2_{quad}+\sigma^2_\epsilon)} + \frac{\|\tilde{\boldsymbol{y}}\|_2^2}{\sigma^2_{quad}+\sigma^2_\epsilon}\right] + Const
\end{aligned} \tag{17}$$

Analogously, we can now rewrite the profile restricted log-likelihood from Equation 15 in this setting as:

$$\begin{aligned}
\ell_{PR}(\delta) &= -\frac{1}{2}\left[(N-L)\log\left(\frac{1}{N-L}\left(\sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{\rho_i+\delta}\right)\right) + \sum_{i=1}^{N-L}\log(\rho_i+\delta)\right] + Const \\
&= -\frac{1}{2}\left[(N-L)\log\left(\frac{1}{N-L}\left(\sum_{i=1}^{S}\frac{\tilde{y}_i^2}{\rho_i+\delta} + \sum_{i=S+1}^{N-L}\frac{\tilde{y}_i^2}{\delta}\right)\right) + \sum_{i=1}^{S}\log(\rho_i+\delta) + (N-L-S)\log(\delta)\right] + Const \\
&= -\frac{1}{2}\left[(N-L)\log\left(\frac{1}{N-L}\left(\sum_{i=1}^{S}\frac{\tilde{y}_i^2}{\rho_i+\delta} + \sum_{i=1}^{N-L}\frac{\tilde{y}_i^2}{\delta} - \sum_{i=1}^{S}\frac{\tilde{y}_i^2}{\delta}\right)\right) + \sum_{i=1}^{S}\log(\rho_i+\delta) + (N-L-S)\log(\delta)\right] + Const \\
&= -\frac{1}{2}\left[(N-L)\log\left(\frac{1}{N-L}\left(-\sum_{i=1}^{S}\frac{\rho_i\tilde{y}_i^2}{(\rho_i+\delta)\delta} + \frac{1}{\delta}\sum_{i=1}^{N-L}\tilde{y}_i^2\right)\right) + \sum_{i=1}^{S}\log(\rho_i+\delta) + (N-L-S)\log(\delta)\right] + Const \\
&= -\frac{1}{2}\left[(N-L)\log\left(\frac{1}{N-L}\left(-\sum_{i=1}^{S}\frac{\rho_i\tilde{y}_i^2}{(\rho_i+\delta)\delta} + \frac{1}{\delta}\|\tilde{\boldsymbol{y}}\|_2^2\right)\right) + \sum_{i=1}^{S}\log(\rho_i+\delta) + (N-L-S)\log(\delta)\right] + Const
\end{aligned} \tag{18}$$

To evaluate the functions in Equations 17 and 18, we need to compute $\tilde{y}_i$ for $i \in \{1, \ldots, S\}$ and $\|\tilde{\boldsymbol{y}}\|_2^2 = \sum_{i=1}^{N-L}\tilde{y}_i^2$.

From Equation 16, we can write $\tilde{\boldsymbol{y}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{y} \\ \boldsymbol{A}_2^{\mathrm{T}}\boldsymbol{y}\end{bmatrix}$. We see that $\tilde{y}_i$ is given by the i-th entry of $\boldsymbol{A}_1^{\mathrm{T}}\boldsymbol{y}$ for $i \leq S$. Here $\boldsymbol{A}_1$ denotes the left singular vectors of $\boldsymbol{P}\boldsymbol{\Phi}$ and can be computed in $\mathcal{O}(ND^2)$ time (as outlined in Section S1.1.2).

Let $\boldsymbol{B}_4 = [\boldsymbol{B}_1, \boldsymbol{B}]$. Since $\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{B}_2\boldsymbol{U} = \boldsymbol{0}$, we have:

$$\begin{aligned}
\boldsymbol{B}_4^{\mathrm{T}}\boldsymbol{B}_4 &= [\boldsymbol{B}_1\boldsymbol{B}]^{\mathrm{T}}\begin{bmatrix}\boldsymbol{B}_1 \\ \boldsymbol{B}\end{bmatrix} \\
&= \begin{bmatrix}\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{B}_1 & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}_1 \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}_1 & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\end{bmatrix} \\
&= \begin{bmatrix}\boldsymbol{I}_L & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{N-L}\end{bmatrix} = \boldsymbol{I}_N
\end{aligned}$$

Thus, $\boldsymbol{B}_4$ is an orthonormal matrix and we have $\boldsymbol{B}_4\boldsymbol{B}_4^{\mathrm{T}} = \boldsymbol{I}_N$. Further, $\boldsymbol{B}_4^{\mathrm{T}}\boldsymbol{y} = [\boldsymbol{B}_1, \boldsymbol{B}]^{\mathrm{T}}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{B}_1^{\mathrm{T}}\boldsymbol{y} \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{y}\end{bmatrix} = \begin{bmatrix}\boldsymbol{y}_1 \\ \tilde{\boldsymbol{y}}\end{bmatrix}$. As a

result:

$$\begin{aligned}\|\boldsymbol{y}\|_2^2 = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} &= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{I}_N\boldsymbol{y} \\ &= \boldsymbol{y}^{\mathrm{T}}\boldsymbol{B}_4\boldsymbol{B}_4^{\mathrm{T}}\boldsymbol{y} \\ &= \left(\boldsymbol{B}_4^{\mathrm{T}}\boldsymbol{y}\right)^{\mathrm{T}}\boldsymbol{B}_4^{\mathrm{T}}\boldsymbol{y} \\ &= \boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{y}_1 + \tilde{\boldsymbol{y}}^{\mathrm{T}}\tilde{\boldsymbol{y}} = \|\boldsymbol{y}_1\|_2^2 + \|\tilde{\boldsymbol{y}}\|_2^2\end{aligned}$$

We can then compute $\|\tilde{\boldsymbol{y}}\|_2^2 = \|\boldsymbol{y}\|_2^2 - \|\boldsymbol{y}_1\|_2^2$. Computing $\boldsymbol{y}_1$ requires the top L left singular vectors of $\boldsymbol{X}$ that can, in turn, be computed in $\mathcal{O}(NL^2)$ time. To compute $\tilde{y}_i, i \in \{1, \ldots, S\}$, we need to compute eigenvectors (columns of $\boldsymbol{B}$) corresponding to the non-zero eigenvalues of $\boldsymbol{PKP}$ which is equivalent to the corresponding left singular vectors of $\boldsymbol{P\Phi}$. Finally, $\rho_i, i \in \{1, \ldots, S\}$ are the non-zero eigenvalues of $\boldsymbol{PKP}$ which are also obtained from the corresponding singular values of $\boldsymbol{P\Phi}$. Both these quantities can be obtained in $\mathcal{O}(NKD + NK^2 + K^3 + ND^2)$ time. Thus, the profile restricted log-likelihood as represented in Equation 18 can be optimized with a $\mathcal{O}(NKD + NK^2 + K^3 + ND^2)$ one-time computation followed by $\mathcal{O}(D)$ time to evaluate this function subsequently.

S1.4 Estimating the effects of interactions

Given estimates of the variance components $(\sigma^2_{quad}, \sigma^2_\epsilon)$, we can compute the posterior probability distribution of the effects of each interaction in a set. Assuming

$$\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \frac{\sigma^2_{quad}}{D}\boldsymbol{I}_D)$$
$$\boldsymbol{y}|\boldsymbol{\Phi}, \boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\Phi\gamma}, \sigma^2_\epsilon \boldsymbol{I}_N)$$

the posterior probability is:

$$\begin{aligned}
P(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\Phi}) &\propto P(\boldsymbol{y}|\boldsymbol{\Phi}, \boldsymbol{\gamma}) \times P(\boldsymbol{\gamma}) \\
&= \frac{1}{(2\pi\sigma^2_\epsilon)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2_\epsilon}(\boldsymbol{y}-\boldsymbol{\Phi\gamma})^T(\boldsymbol{y}-\boldsymbol{\Phi\gamma})\right) \times \left(\frac{D}{2\pi\sigma^2_{quad}}\right)^{\frac{D}{2}} \exp\left(-\frac{D}{2\sigma^2_{quad}}\sum_{i=1}^{D}\gamma_i^2\right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2_\epsilon}(\boldsymbol{y}-\boldsymbol{\Phi\gamma})^T(\boldsymbol{y}-\boldsymbol{\Phi\gamma}) - \frac{D}{2\sigma^2_{quad}}\sum_{i=1}^{D}\gamma_i^2\right) \\
&= \exp\left(-\frac{1}{2\sigma^2_\epsilon}\left(\boldsymbol{\gamma}^T(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \frac{D\sigma^2_\epsilon}{\sigma^2_{quad}}\boldsymbol{I}_D)\boldsymbol{\gamma} + 2\boldsymbol{y}^T\boldsymbol{\Phi\gamma}\right)\right) \\
&= \exp\left(-\frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\mu}_{posterior})^T\boldsymbol{\Sigma}^{-1}_{posterior}(\boldsymbol{\gamma}-\boldsymbol{\mu}_{posterior})\right)
\end{aligned}$$

where $\boldsymbol{\mu}_{posterior} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda\boldsymbol{I}_D)^{-1}\boldsymbol{\Phi}^T\boldsymbol{y}$, $\boldsymbol{\Sigma}_{posterior} = \sigma^2_\epsilon(\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda\boldsymbol{I}_D)^{-1}$, and $\lambda = \frac{D\sigma^2_\epsilon}{\sigma^2_{quad}}$.

Therefore, given σ^2_{quad} and σ^2_ϵ, we have

$$\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\Phi} \sim \mathcal{N}(\boldsymbol{\mu}_{posterior}, \boldsymbol{\Sigma}_{posterior})$$

For a given interaction t, we estimate the posterior mean μ_t as the corresponding entry in $\boldsymbol{\mu}_{posterior}$ while the posterior standard deviation σ_t is estimated as the square root of the corresponding diagonal entry of $\boldsymbol{\Sigma}_{posterior}$.