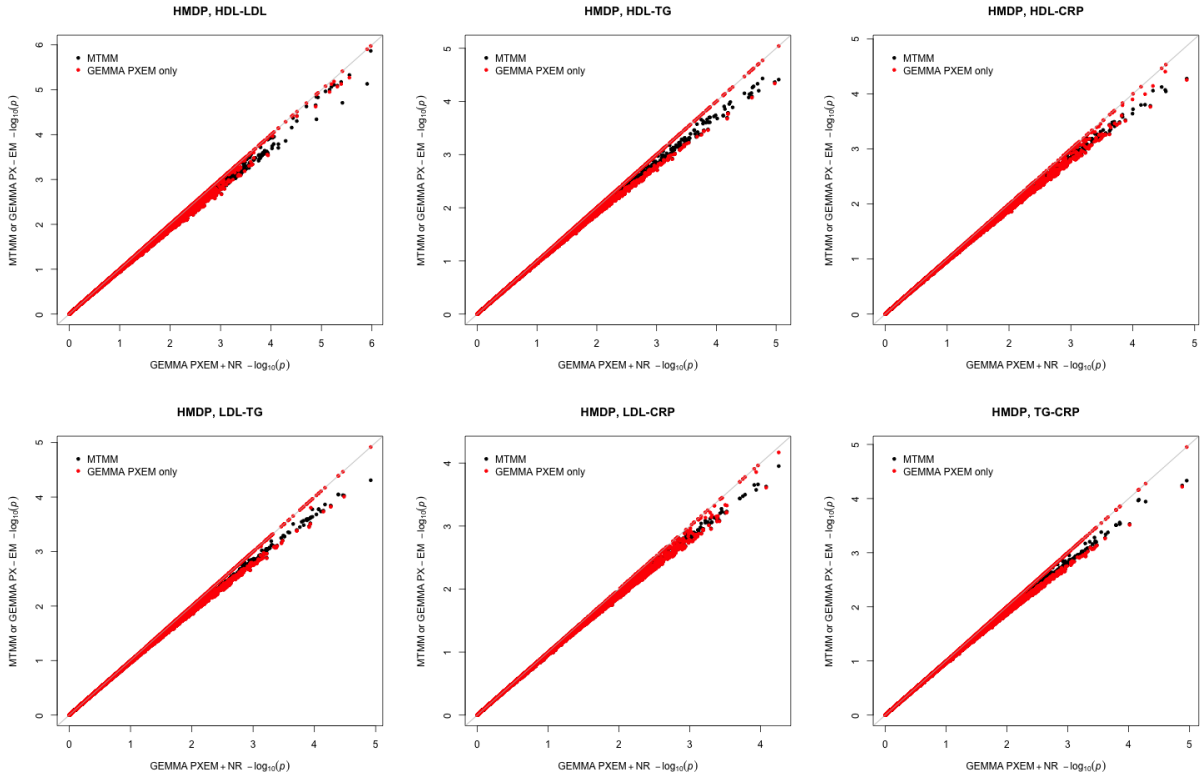
Supplementary Information

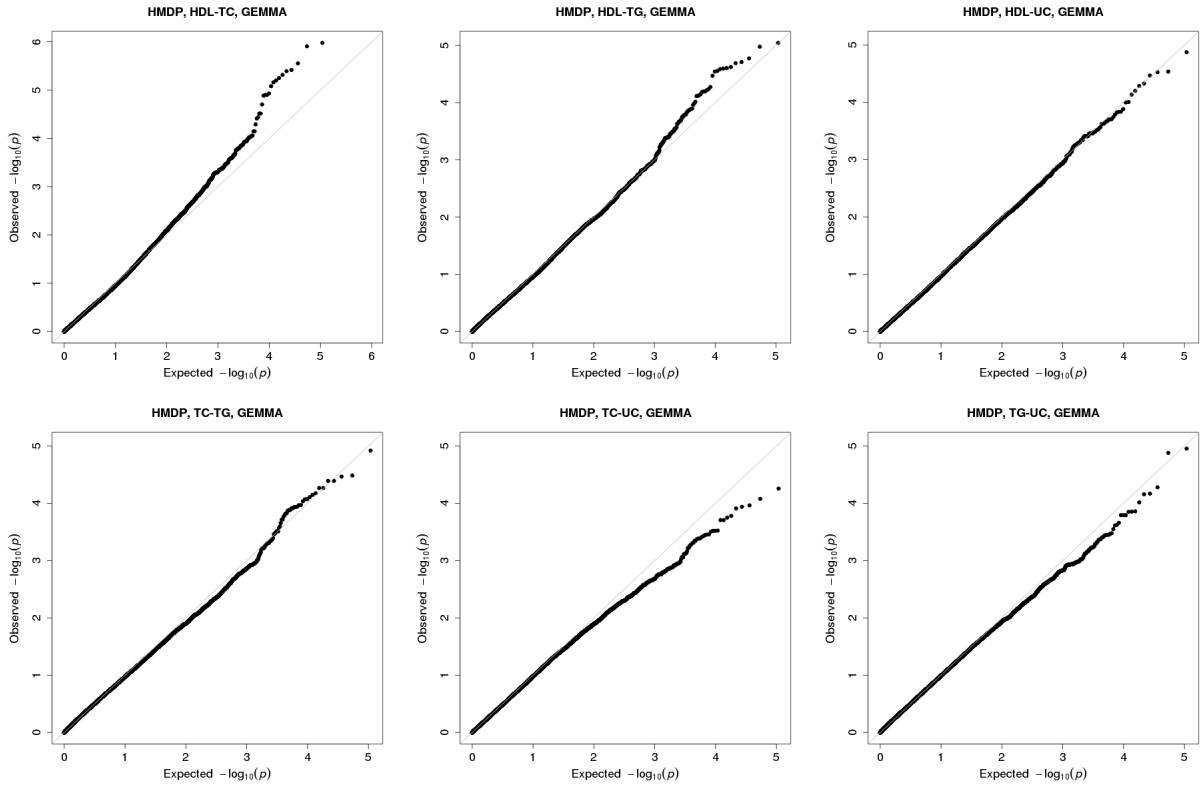
1 Supplementary Figures

Supplementary Figure 1



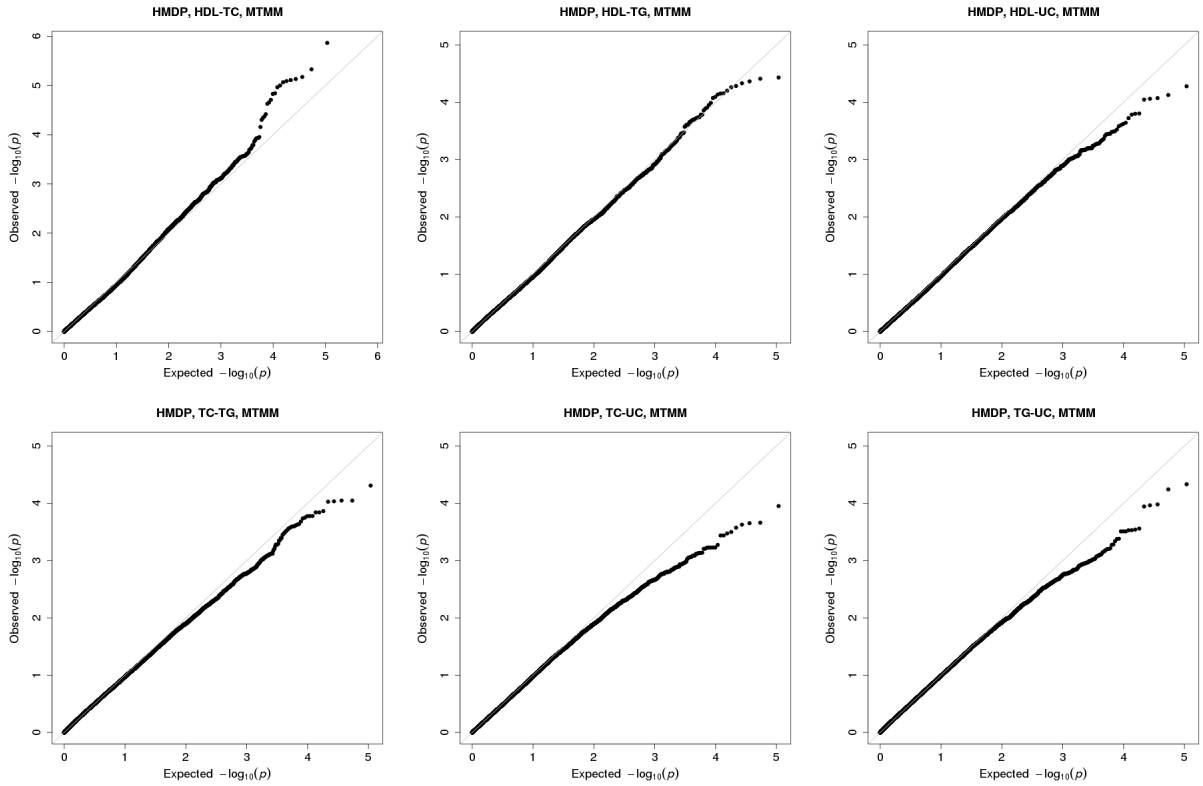
Supplementary Figure 1: Comparison of -$\log_{10} p$ values obtained from GEMMA using both PX-EM and NR, with those from MTMM (black) or those from GEMMA using only PX-EM (red), for all paired traits in the HMDP data set.

Supplementary Figure 2



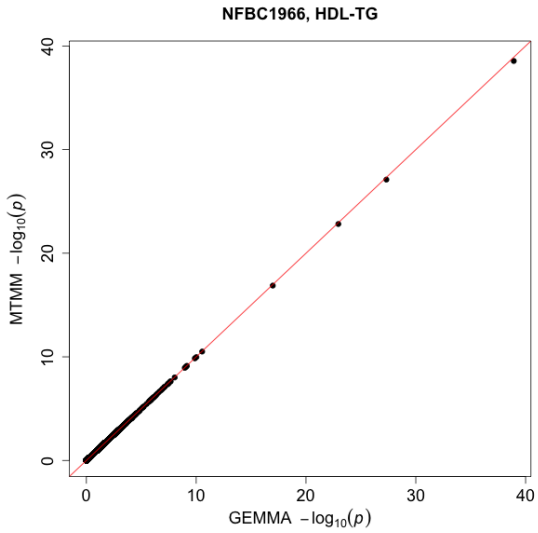
Supplementary Figure 2: Quantile-quantile plots for the p values obtained from GEMMA, for all paired traits in the HMDP data set.

Supplementary Figure 3



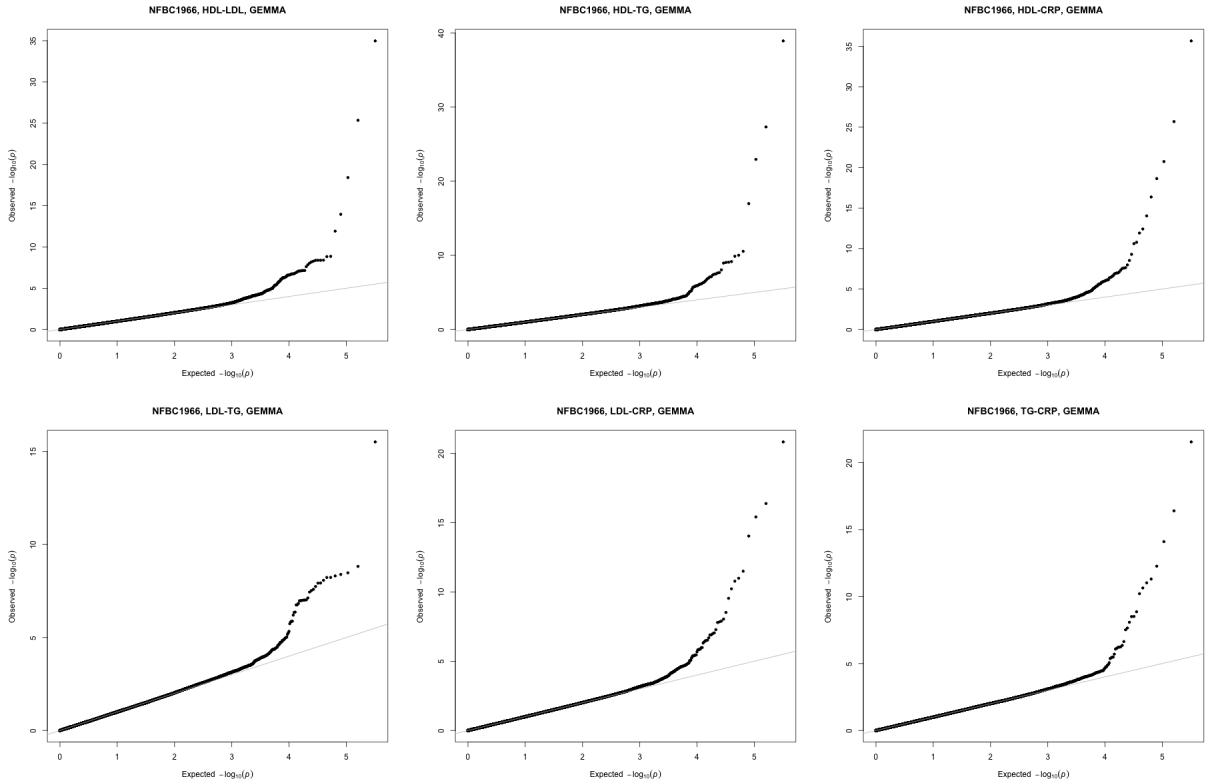
Supplementary Figure 3: Quantile-quantile plots for the p values obtained from MTMM, for all paired traits in the HMDP data set.

Supplementary Figure 4



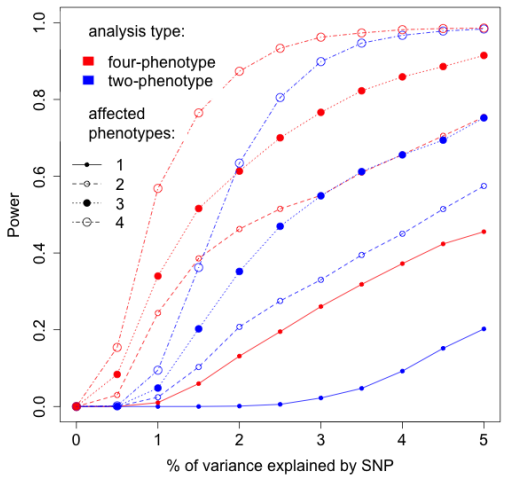
Supplementary Figure 4: GEMMA p values are similar compared with MTMM p values for the NFBC data.

Supplementary Figure 5

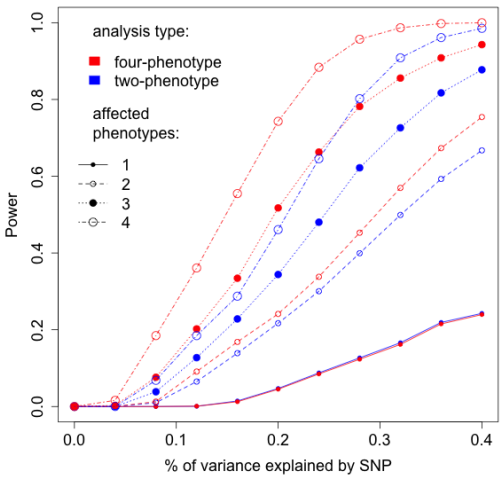


Supplementary Figure 5: Quantile-quantile plots for the p values obtained from GEMMA, for all paired traits in the NFBC1966 data set.

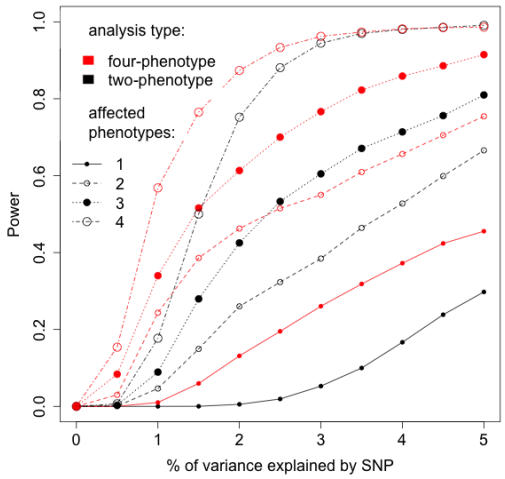
Supplementary Figure 6



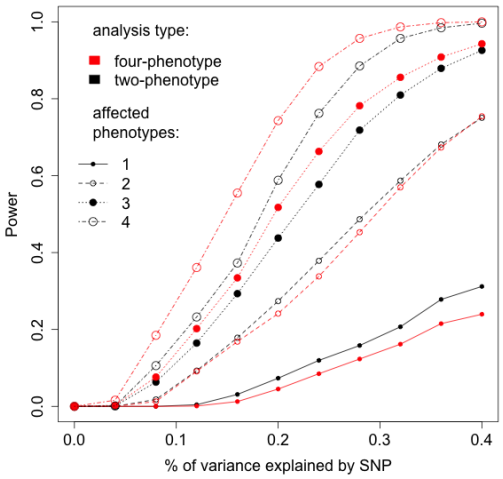
(a) HMDP-based simulations



(b) NFBC1966-based simulations



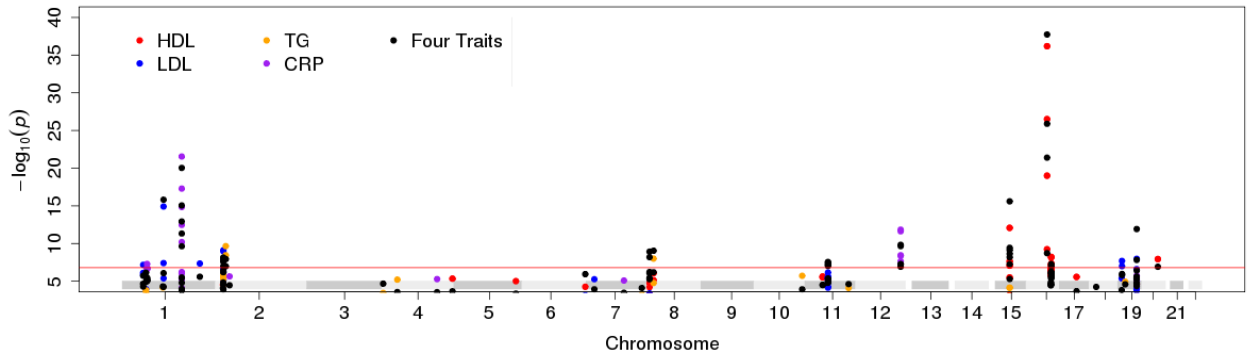
(c) HMDP-based simulations



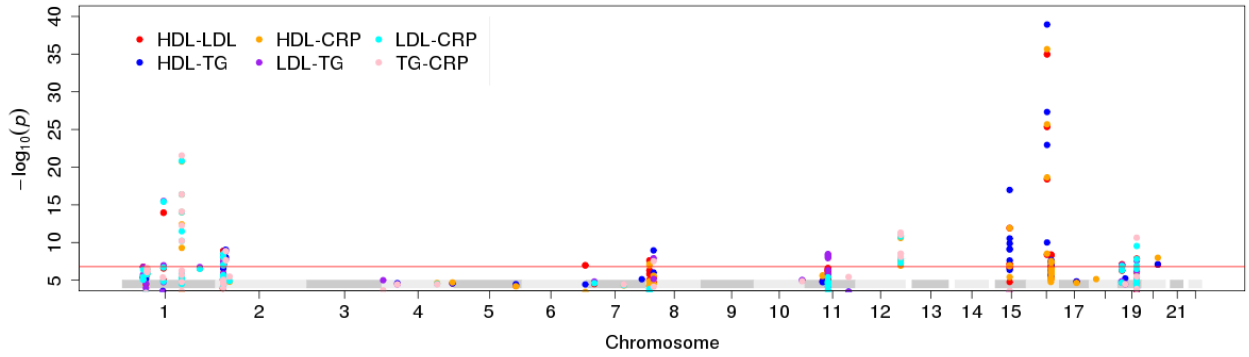
(d) NFBC1966-based simulations

Supplementary Figure 6: Power comparison between the four-phenotype analysis and the two-phenotype analysis with GEMMA, using simulations based on the HMDP data (a, c) or the NFBC1966 data (b, d). x-axis shows the proportion of phenotypic variance in the affected traits explained (PVE) by the SNP. Symbol size and line type indicate the number of phenotypes affected by the causal SNP. For four-phenotype analysis (red), the genome-wide significance threshold after Bonferroni correction for the number of SNPs is used (4.6×10^{-7} for HMDP-based simulations and 1.6×10^{-7} for NFBC1966-based simulations). For two-phenotype analysis, either the same genome-wide significance threshold is used (black) (c, d), or a significance threshold further corrected for the six tests performed by the two-trait analysis (7.6×10^{-8} for HMDP-based simulations and 2.6×10^{-8} for NFBC1966-based simulations) is used (blue) (a, b).

Supplementary Figure 7



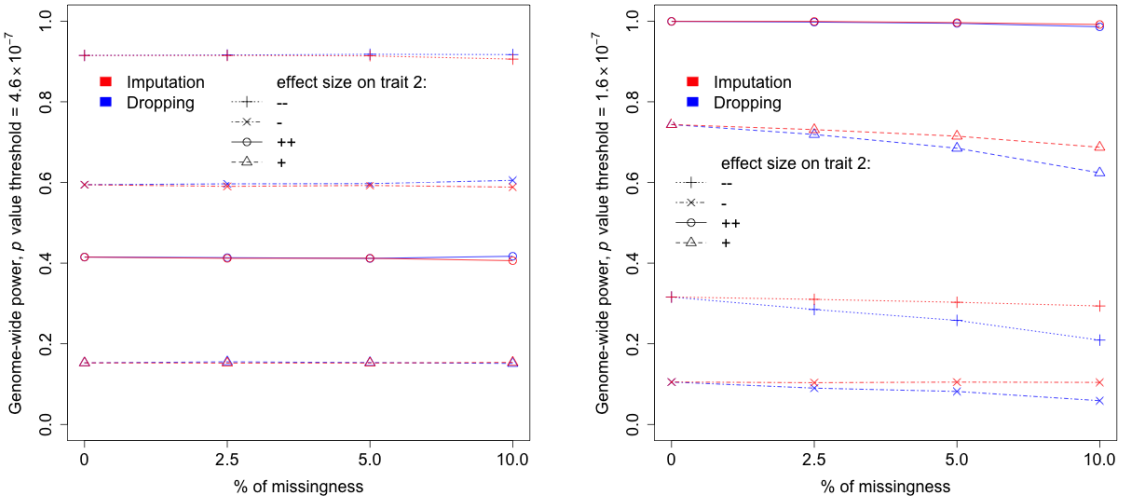
(a) univariate and four-phenotype analyses



(b) two-phenotype analyses

Supplementary Figure 7: GEMMA -$\log_{10} p$ values for SNPs from the univariate LMM analyses (a), the two-phenotype mvLMM analyses (b) and the joint four-phenotype mvLMM analyses (a) in the NFBC1966 data are plotted against their physical chromosomal positions. Chromosomes are shown in alternate colors. The red horizontal line indicates the significance level of 0.05 after Bonferroni correction for the number of SNPs (1.6×10^{-7}). Only SNPs with a p value below 1×10^{-5} in any of the eleven tests are displayed.

Supplementary Figure 8



(a) Imputation vs Dropping (HMDP)

(b) Imputation vs Dropping (NFBC1966)

Supplementary Figure 8: Power comparison of the two approaches to deal with missing phenotypes in GEMMA, using simulations based on the HMDP data (a) or the NFBC1966 data (b). The first approach use only individuals with fully observed phenotypes (dropping, blue). The second approach imputes phenotypes before association tests (imputation, red). x-axis shows the percentage of individuals having one missing phenotype. the point symbol and line type indicate the SNP effect direction (compared with its effect on the first phenotype) and size (quantified by PVE) on the second phenotype (+: opposite direction, 0.8PVE; ×: opposite direction, 0.2PVE; o: same direction, 0.8PVE; Δ: same direction, 0.2PVE).

2 Supplementary Tables

Supplementary Table 1: List of SNPs that exceed the significance of 0.05 after Bonferroni correction for both the number of SNPs and the number of tests from either the four-phenotype mvLMM analysis or the univariate LMM analysis in the NFBC1966 data set. λ_{GC} is the genomic control inflation factor. p values in the four-phenotype analysis below the threshold (1.6×10^{-7}), and p values in the univariate analysis below the threshold (3.9×10^{-8}) are in bold font. SNPs that are significant only in the four-phenotype analysis but not in the univariate analysis are highlighted in red, while SNPs that are significant only in the univariate analysis are highlighted in blue. SNPs that are more significant in the four-phenotype analysis than in the univariate analysis (after correcting for the number of tests performed) are highlighted in magenta (in addition to red).

SNP	Position	mvLMM (p value) Four Traits ($\lambda_{GC} = 0.979$)	LMM (p value) HDL ($\lambda_{GC} = 0.998$)	LDL ($\lambda_{GC} = 1.003$)	TG ($\lambda_{GC} = 1.000$)	CRP ($\lambda_{GC} = 0.993$)
CELSR2	chromosome 1					
rs646776	109620053	$\mathbf{1.54 \times 10^{-16}}$	9.17×10^{-12}	$\mathbf{1.23 \times 10^{-15}}$	9.30×10^{-1}	6.32×10^{-2}
CRP	chromosome 1					
rs1811472	157908973	$\mathbf{1.19 \times 10^{-13}}$	7.05×10^{-2}	8.81×10^{-1}	4.90×10^{-1}	$\mathbf{1.42 \times 10^{-15}}$
rs12093699	157914612	$\mathbf{4.71 \times 10^{-12}}$	8.75×10^{-1}	5.77×10^{-1}	8.39×10^{-1}	$\mathbf{3.39 \times 10^{-13}}$
rs2592887	157919563	$\mathbf{8.79 \times 10^{-16}}$	4.99×10^{-2}	9.34×10^{-1}	3.17×10^{-1}	$\mathbf{5.13 \times 10^{-18}}$
rs2794520	157945440	$\mathbf{9.29 \times 10^{-21}}$	2.98×10^{-1}	7.34×10^{-1}	9.91×10^{-1}	$\mathbf{2.84 \times 10^{-22}}$
rs11265260	157966663	$\mathbf{2.27 \times 10^{-10}}$	5.85×10^{-2}	1.55×10^{-1}	6.90×10^{-1}	$\mathbf{6.28 \times 10^{-11}}$
APOB	chromosome 2					
rs6728178	21047434	$\mathbf{7.37 \times 10^{-9}}$	3.46×10^{-6}	7.57×10^{-7}	3.29×10^{-6}	3.62×10^{-1}
rs6754295	21059688	$\mathbf{7.99 \times 10^{-9}}$	3.83×10^{-6}	6.27×10^{-7}	5.25×10^{-6}	3.48×10^{-1}
rs676210	21085029	$\mathbf{2.09 \times 10^{-8}}$	2.74×10^{-6}	8.01×10^{-6}	1.79×10^{-6}	4.65×10^{-1}
rs693	21085700	$\mathbf{9.50 \times 10^{-8}}$	4.19×10^{-2}	$\mathbf{1.05 \times 10^{-9}}$	4.00×10^{-3}	3.48×10^{-1}
rs673548	21091049	$\mathbf{1.36 \times 10^{-8}}$	2.19×10^{-6}	6.71×10^{-6}	1.32×10^{-6}	4.79×10^{-1}
rs1429974	21154275	6.33×10^{-7}	8.02×10^{-2}	$\mathbf{1.65 \times 10^{-8}}$	1.84×10^{-1}	9.48×10^{-1}
rs754524	21165046	$\mathbf{2.44 \times 10^{-8}}$	9.44×10^{-2}	$\mathbf{7.82 \times 10^{-10}}$	1.57×10^{-1}	6.10×10^{-1}
rs754523	21165196	5.68×10^{-7}	8.55×10^{-2}	$\mathbf{1.55 \times 10^{-8}}$	1.93×10^{-1}	9.66×10^{-1}
GCKR	chromosome 2					
rs1260326	27584444	$\mathbf{1.16 \times 10^{-8}}$	1.58×10^{-1}	1.14×10^{-1}	$\mathbf{2.21 \times 10^{-10}}$	5.25×10^{-2}
rs780094	27594741	$\mathbf{1.02 \times 10^{-7}}$	3.38×10^{-1}	2.63×10^{-1}	$\mathbf{3.50 \times 10^{-9}}$	1.44×10^{-1}
PPP1R3B	chromosome 8					
rs983309	9215142	$\mathbf{6.40 \times 10^{-9}}$	5.99×10^{-5}	2.57×10^{-3}	7.57×10^{-1}	1.55×10^{-3}
rs2126259	9222556	$\mathbf{1.19 \times 10^{-9}}$	1.67×10^{-5}	5.79×10^{-4}	4.24×10^{-1}	5.93×10^{-3}
LPL	chromosome 8					
rs10096633	19875201	$\mathbf{8.96 \times 10^{-10}}$	8.23×10^{-7}	9.36×10^{-1}	$\mathbf{1.01 \times 10^{-8}}$	5.22×10^{-1}
FADS	chromosome 11					
rs174537	61309256	$\mathbf{5.55 \times 10^{-8}}$	3.41×10^{-2}	3.76×10^{-6}	2.45×10^{-2}	8.44×10^{-1}
rs102275	61314379	$\mathbf{3.25 \times 10^{-8}}$	2.22×10^{-2}	3.06×10^{-6}	2.27×10^{-2}	9.05×10^{-1}
rs174546	61326406	$\mathbf{2.80 \times 10^{-8}}$	3.99×10^{-2}	2.25×10^{-6}	2.03×10^{-2}	9.04×10^{-1}
rs174556	61337211	$\mathbf{6.88 \times 10^{-8}}$	1.67×10^{-1}	7.02×10^{-7}	5.44×10^{-2}	9.89×10^{-1}
rs1535	61354548	$\mathbf{9.40 \times 10^{-8}}$	4.76×10^{-2}	4.11×10^{-6}	2.77×10^{-2}	8.71×10^{-1}
HNF1A	chromosome 12					
rs2650000	119873345	$\mathbf{1.48 \times 10^{-10}}$	2.16×10^{-1}	9.36×10^{-1}	6.48×10^{-1}	$\mathbf{1.44 \times 10^{-12}}$
rs7953249	119888107	$\mathbf{2.21 \times 10^{-10}}$	1.43×10^{-1}	9.16×10^{-1}	6.10×10^{-1}	$\mathbf{2.38 \times 10^{-12}}$
rs1169300	119915608	$\mathbf{5.29 \times 10^{-8}}$	6.76×10^{-1}	4.05×10^{-1}	5.27×10^{-1}	$\mathbf{4.28 \times 10^{-9}}$
rs2464196	119919810	$\mathbf{5.82 \times 10^{-8}}$	5.95×10^{-1}	5.44×10^{-1}	5.74×10^{-1}	$\mathbf{3.58 \times 10^{-9}}$
rs735396	119923227	$\mathbf{1.19 \times 10^{-7}}$	7.17×10^{-1}	3.22×10^{-1}	3.18×10^{-1}	$\mathbf{2.24 \times 10^{-8}}$
LIPC	chromosome 15					
rs166358	56468097	$\mathbf{3.66 \times 10^{-10}}$	8.57×10^{-8}	3.80×10^{-2}	2.29×10^{-1}	5.43×10^{-1}
rs1532085	56470658	$\mathbf{2.52 \times 10^{-16}}$	$\mathbf{8.33 \times 10^{-13}}$	3.46×10^{-1}	6.05×10^{-2}	6.21×10^{-1}
rs415799	56478046	$\mathbf{7.47 \times 10^{-10}}$	$\mathbf{2.32 \times 10^{-8}}$	5.72×10^{-1}	1.13×10^{-1}	9.06×10^{-1}
rs16940213	56482629	$\mathbf{5.51 \times 10^{-8}}$	3.19×10^{-6}	2.67×10^{-1}	1.31×10^{-1}	4.29×10^{-1}
rs473224	56524633	$\mathbf{6.54 \times 10^{-9}}$	1.57×10^{-3}	5.30×10^{-1}	7.64×10^{-5}	1.73×10^{-1}
rs261336	56529710	$\mathbf{2.29 \times 10^{-9}}$	6.51×10^{-4}	2.30×10^{-1}	7.08×10^{-5}	3.00×10^{-1}
CETP	chromosome 16					
rs9989419	55542640	$\mathbf{1.88 \times 10^{-9}}$	$\mathbf{5.79 \times 10^{-10}}$	7.63×10^{-1}	9.62×10^{-1}	6.12×10^{-1}
rs3764261	55550825	$\mathbf{1.85 \times 10^{-38}}$	$\mathbf{6.56 \times 10^{-37}}$	7.09×10^{-2}	2.52×10^{-1}	2.06×10^{-1}
rs1532624	55562980	$\mathbf{1.30 \times 10^{-26}}$	$\mathbf{3.15 \times 10^{-27}}$	1.64×10^{-1}	1.20×10^{-1}	1.05×10^{-1}
rs7499892	55564091	$\mathbf{4.01 \times 10^{-22}}$	$\mathbf{9.93 \times 10^{-20}}$	8.36×10^{-1}	4.51×10^{-1}	8.09×10^{-1}
LCAT	chromosome 16					
rs255049	66570972	$\mathbf{5.34 \times 10^{-8}}$	$\mathbf{6.86 \times 10^{-9}}$	1.44×10^{-1}	1.60×10^{-1}	7.50×10^{-1}
rs255052	66582496	$\mathbf{9.72 \times 10^{-8}}$	$\mathbf{5.98 \times 10^{-9}}$	2.01×10^{-1}	2.36×10^{-1}	7.02×10^{-1}
LDLR	chromosome 19					
rs11668477	11056030	1.09×10^{-6}	5.95×10^{-2}	$\mathbf{2.04 \times 10^{-8}}$	1.66×10^{-2}	7.18×10^{-1}
APO cluster	chromosome 19					
rs157580	50087106	$\mathbf{1.51 \times 10^{-8}}$	9.82×10^{-3}	$\mathbf{1.06 \times 10^{-8}}$	1.18×10^{-3}	2.42×10^{-1}
rs2075650	50087459	$\mathbf{1.21 \times 10^{-12}}$	4.95×10^{-2}	8.54×10^{-6}	2.20×10^{-5}	1.01×10^{-5}
HNF4A	chromosome 20					
rs1800961	42475778	$\mathbf{1.21 \times 10^{-7}}$	$\mathbf{1.15 \times 10^{-8}}$	3.78×10^{-1}	4.59×10^{-3}	4.99×10^{-1}

Supplementary Table 2: List of SNPs that exceed the significance of 0.05 after Bonferroni correction for both the number of SNPs and the number of tests from either the four-phenotype mvLMM analysis or the two-phenotype mvLMM analysis in the NFBC1966 data set. λ_{GC} is the genomic control inflation factor. p values in the four-phenotype analysis below the threshold (1.6×10^{-7}), and p values in the two-phenotype analysis below the threshold (2.6×10^{-8}) are in bold font. SNPs that are significant only in the four-phenotype analysis but not in the two-phenotype analysis are highlighted in red. No SNP is significant only in the two-phenotype analysis. SNPs that are more significant in the four-phenotype analysis than in the two-phenotype analysis (after correcting for the number of tests performed) are highlighted in magenta (in addition to red).

SNP	Position	mvLMM (p value)						
		Four Traits ($\lambda_{GC} = 0.979$)	HDL-LDL ($\lambda_{GC} = 0.989$)	HDL-TG ($\lambda_{GC} = 0.991$)	HDL-CRP ($\lambda_{GC} = 0.992$)	LDL-TG ($\lambda_{GC} = 0.993$)	LDL-CRP ($\lambda_{GC} = 0.998$)	TG-CRP ($\lambda_{GC} = 0.994$)
CELSR2	chromosome 1							
rs646776	109620053	$\mathbf{1.54 \times 10^{-16}}$	$\mathbf{1.1 \times 10^{-14}}$	1.98×10^{-1}	1.92×10^{-2}	$\mathbf{3.06 \times 10^{-16}}$	$\mathbf{3.84 \times 10^{-16}}$	1.61×10^{-1}
CRP	chromosome 1							
rs1811472	157908973	$\mathbf{1.19 \times 10^{-13}}$	1.82×10^{-1}	2.04×10^{-1}	$\mathbf{9.17 \times 10^{-15}}$	7.18×10^{-1}	$\mathbf{9.23 \times 10^{-15}}$	$\mathbf{7.87 \times 10^{-15}}$
rs12093699	157914612	$\mathbf{4.71 \times 10^{-12}}$	8.28×10^{-1}	9.70×10^{-1}	$\mathbf{3.82 \times 10^{-13}}$	7.99×10^{-1}	$\mathbf{3.17 \times 10^{-12}}$	$\mathbf{5.40 \times 10^{-13}}$
rs2592887	157919563	$\mathbf{8.79 \times 10^{-16}}$	1.45×10^{-1}	1.49×10^{-1}	$\mathbf{4.24 \times 10^{-17}}$	5.80×10^{-1}	$\mathbf{4.12 \times 10^{-17}}$	$\mathbf{3.97 \times 10^{-17}}$
rs2794520	157945440	$\mathbf{9.29 \times 10^{-21}}$	5.20×10^{-1}	5.34×10^{-1}	$\mathbf{1.77 \times 10^{-21}}$	9.31×10^{-1}	$\mathbf{1.47 \times 10^{-21}}$	$\mathbf{2.82 \times 10^{-22}}$
rs11265260	157966663	$\mathbf{2.27 \times 10^{-10}}$	3.98×10^{-2}	7.62×10^{-2}	$\mathbf{5.11 \times 10^{-10}}$	3.68×10^{-1}	$\mathbf{5.94 \times 10^{-11}}$	$\mathbf{6.07 \times 10^{-11}}$
APOB	chromosome 2							
rs6728178	21047434	$\mathbf{7.37 \times 10^{-9}}$	$\mathbf{1.48 \times 10^{-9}}$	2.03×10^{-7}	2.63×10^{-5}	3.02×10^{-8}	4.92×10^{-6}	2.23×10^{-5}
rs6754295	21059688	$\mathbf{7.99 \times 10^{-9}}$	$\mathbf{1.37 \times 10^{-9}}$	2.98×10^{-7}	2.78×10^{-5}	3.61×10^{-8}	4.01×10^{-6}	3.40×10^{-5}
rs676210	21085029	$\mathbf{2.09 \times 10^{-8}}$	$\mathbf{9.04 \times 10^{-9}}$	1.11×10^{-7}	2.03×10^{-5}	1.07×10^{-7}	4.91×10^{-5}	1.18×10^{-5}
rs693	21085700	$\mathbf{9.50 \times 10^{-8}}$	$\mathbf{3.95 \times 10^{-9}}$	9.83×10^{-3}	1.26×10^{-1}	$\mathbf{5.94 \times 10^{-9}}$	$\mathbf{9.09 \times 10^{-9}}$	1.81×10^{-2}
rs673548	21091049	$\mathbf{1.36 \times 10^{-8}}$	$\mathbf{6.45 \times 10^{-9}}$	7.78×10^{-8}	1.64×10^{-5}	7.52×10^{-8}	4.18×10^{-5}	8.73×10^{-6}
rs754524	21165046	$\mathbf{2.44 \times 10^{-8}}$	$\mathbf{4.15 \times 10^{-9}}$	1.79×10^{-1}	1.63×10^{-1}	$\mathbf{4.90 \times 10^{-9}}$	$\mathbf{3.01 \times 10^{-9}}$	2.46×10^{-1}
GCKR	chromosome 2							
rs1260326	27584444	$\mathbf{1.16 \times 10^{-8}}$	1.38×10^{-1}	$\mathbf{9.00 \times 10^{-10}}$	8.54×10^{-2}	$\mathbf{1.51 \times 10^{-9}}$	5.77×10^{-2}	$\mathbf{1.36 \times 10^{-9}}$
rs780094	27594741	$\mathbf{1.02 \times 10^{-7}}$	3.83×10^{-1}	$\mathbf{9.08 \times 10^{-9}}$	2.61×10^{-1}	$\mathbf{1.82 \times 10^{-8}}$	2.11×10^{-1}	$\mathbf{2.23 \times 10^{-8}}$
PPP1R3B	chromosome 8							
rs983309	9215142	$\mathbf{6.40 \times 10^{-9}}$	5.04×10^{-7}	1.24×10^{-4}	7.54×10^{-8}	3.98×10^{-3}	1.71×10^{-4}	3.30×10^{-3}
rs2126259	9222556	$\mathbf{1.19 \times 10^{-9}}$	$\mathbf{2.41 \times 10^{-8}}$	6.04×10^{-5}	1.02×10^{-7}	3.11×10^{-4}	1.46×10^{-4}	6.95×10^{-3}
LPL	chromosome 8							
rs10096633	19875201	$\mathbf{8.96 \times 10^{-10}}$	4.38×10^{-6}	$\mathbf{1.10 \times 10^{-9}}$	2.53×10^{-6}	$\mathbf{1.20 \times 10^{-8}}$	8.12×10^{-1}	3.01×10^{-8}
FADS	chromosome 11							
rs174537	61309256	$\mathbf{5.55 \times 10^{-8}}$	5.10×10^{-7}	3.43×10^{-2}	8.73×10^{-2}	$\mathbf{8.47 \times 10^{-9}}$	2.32×10^{-5}	5.96×10^{-2}
rs102275	61314379	$\mathbf{3.25 \times 10^{-8}}$	2.54×10^{-7}	2.51×10^{-2}	6.23×10^{-2}	$\mathbf{5.90 \times 10^{-9}}$	1.85×10^{-5}	5.91×10^{-2}
rs174546	61326406	$\mathbf{2.80 \times 10^{-8}}$	3.6×10^{-7}	3.32×10^{-2}	1.06×10^{-1}	$\mathbf{3.35 \times 10^{-9}}$	1.38×10^{-5}	5.34×10^{-2}
rs174556	61337211	$\mathbf{6.88 \times 10^{-8}}$	5.58×10^{-7}	1.32×10^{-1}	3.84×10^{-1}	$\mathbf{4.08 \times 10^{-9}}$	4.04×10^{-6}	1.45×10^{-1}
rs1535	61354548	$\mathbf{9.40 \times 10^{-8}}$	8.18×10^{-7}	4.48×10^{-2}	1.21×10^{-1}	$\mathbf{1.18 \times 10^{-8}}$	2.50×10^{-5}	6.91×10^{-2}
HNF1A	chromosome 12							
rs2650000	119873345	$\mathbf{1.48 \times 10^{-10}}$	4.54×10^{-1}	4.53×10^{-1}	$\mathbf{1.67 \times 10^{-11}}$	8.76×10^{-1}	$\mathbf{1.02 \times 10^{-11}}$	$\mathbf{4.90 \times 10^{-12}}$
rs7953249	119888107	$\mathbf{2.21 \times 10^{-10}}$	3.31×10^{-1}	3.35×10^{-1}	$\mathbf{2.45 \times 10^{-11}}$	8.41×10^{-1}	$\mathbf{1.67 \times 10^{-11}}$	$\mathbf{9.40 \times 10^{-12}}$
rs1169300	119915608	$\mathbf{5.29 \times 10^{-8}}$	6.16×10^{-1}	6.20×10^{-1}	$\mathbf{2.55 \times 10^{-8}}$	6.81×10^{-1}	$\mathbf{1.26 \times 10^{-8}}$	$\mathbf{3.13 \times 10^{-9}}$
rs2464196	119919810	$\mathbf{5.82 \times 10^{-8}}$	6.90×10^{-1}	5.99×10^{-1}	$\mathbf{2.31 \times 10^{-8}}$	7.92×10^{-1}	$\mathbf{1.44 \times 10^{-8}}$	$\mathbf{3.05 \times 10^{-9}}$
rs735396	119923227	$\mathbf{1.19 \times 10^{-7}}$	5.45×10^{-1}	4.23×10^{-1}	1.17×10^{-7}	4.96×10^{-1}	5.24×10^{-8}	$\mathbf{8.3 \times 10^{-9}}$
LIPC	chromosome 15							
rs166358	56468097	$\mathbf{3.66 \times 10^{-10}}$	2.34×10^{-7}	$\mathbf{8.28 \times 10^{-10}}$	1.39×10^{-7}	1.58×10^{-2}	8.50×10^{-2}	4.45×10^{-1}
rs1532085	56470658	$\mathbf{2.52 \times 10^{-16}}$	$\mathbf{1.21 \times 10^{-12}}$	$\mathbf{1.06 \times 10^{-17}}$	$\mathbf{1.21 \times 10^{-12}}$	1.67×10^{-1}	5.87×10^{-1}	1.63×10^{-1}
rs415799	56478046	$\mathbf{7.47 \times 10^{-10}}$	7.02×10^{-8}	$\mathbf{2.87 \times 10^{-11}}$	1.03×10^{-7}	2.90×10^{-1}	8.39×10^{-1}	2.39×10^{-1}
rs16940213	56482629	$\mathbf{5.51 \times 10^{-8}}$	1.71×10^{-5}	$\mathbf{2.37 \times 10^{-8}}$	3.84×10^{-6}	7.46×10^{-2}	3.62×10^{-1}	2.85×10^{-1}
rs473224	56524633	$\mathbf{6.54 \times 10^{-9}}$	3.80×10^{-3}	$\mathbf{7.00 \times 10^{-10}}$	9.57×10^{-4}	3.09×10^{-4}	3.47×10^{-1}	2.80×10^{-4}
rs261336	56529710	$\mathbf{2.29 \times 10^{-9}}$	7.29×10^{-4}	$\mathbf{1.33 \times 10^{-10}}$	6.20×10^{-4}	3.68×10^{-4}	3.19×10^{-1}	3.35×10^{-4}
CETP	chromosome 16							
rs9989419	55542640	$\mathbf{1.88 \times 10^{-9}}$	$\mathbf{4.05 \times 10^{-9}}$	$\mathbf{9.67 \times 10^{-11}}$	$\mathbf{2.94 \times 10^{-9}}$	9.21×10^{-1}	8.41×10^{-1}	8.43×10^{-1}
rs3764261	55550825	$\mathbf{1.85 \times 10^{-38}}$	$\mathbf{1.06 \times 10^{-35}}$	$\mathbf{1.19 \times 10^{-39}}$	$\mathbf{2.19 \times 10^{-36}}$	1.64×10^{-1}	1.07×10^{-1}	3.02×10^{-1}
rs1532624	55562980	$\mathbf{1.30 \times 10^{-26}}$	$\mathbf{4.42 \times 10^{-26}}$	$\mathbf{4.79 \times 10^{-28}}$	$\mathbf{2.04 \times 10^{-26}}$	1.89×10^{-1}	1.25×10^{-1}	1.22×10^{-1}
rs7499892	55564091	$\mathbf{4.01 \times 10^{-22}}$	$\mathbf{3.88 \times 10^{-19}}$	$\mathbf{1.13 \times 10^{-23}}$	$\mathbf{2.25 \times 10^{-19}}$	7.52×10^{-1}	9.44×10^{-1}	6.62×10^{-1}
LCAT	chromosome 16							
rs255049	66570972	$\mathbf{5.34 \times 10^{-8}}$	$\mathbf{4.07 \times 10^{-9}}$	3.25×10^{-8}	2.85×10^{-8}	4.75×10^{-2}	3.29×10^{-1}	3.32×10^{-1}
rs255052	66582496	$\mathbf{9.72 \times 10^{-8}}$	$\mathbf{5.24 \times 10^{-9}}$	$\mathbf{2.10 \times 10^{-8}}$	4.58×10^{-8}	1.06×10^{-1}	3.81×10^{-1}	5.12×10^{-1}
APO cluster	chromosome 19							
rs157580	50087106	$\mathbf{1.51 \times 10^{-8}}$	$\mathbf{1.42 \times 10^{-8}}$	1.95×10^{-3}	7.83×10^{-3}	$\mathbf{2.57 \times 10^{-8}}$	$\mathbf{1.60 \times 10^{-8}}$	7.35×10^{-4}
rs2075650	50087459	$\mathbf{1.21 \times 10^{-12}}$	1.93×10^{-5}	1.17×10^{-4}	7.21×10^{-7}	6.24×10^{-7}	$\mathbf{2.88 \times 10^{-10}}$	$\mathbf{2.27 \times 10^{-11}}$
HNF4A	chromosome 20							
rs1800961	42475778	$\mathbf{1.21 \times 10^{-7}}$	8.41×10^{-8}	7.61×10^{-8}	$\mathbf{1.05 \times 10^{-8}}$	1.81×10^{-2}	5.05×10^{-1}	6.03×10^{-3}

3 Supplementary Note

3.1 Multivariate Linear Mixed Model

We consider the multivariate linear mixed model[1],

$$\tilde{\mathbf{Y}} = \mathbf{A}\tilde{\mathbf{W}} + \boldsymbol{\beta}\tilde{\mathbf{x}}^T + \tilde{\mathbf{G}} + \tilde{\mathbf{E}}; \quad \tilde{\mathbf{G}} \sim \mathrm{MN}_{d\times n}(\mathbf{0}, \mathbf{V}_g, \mathbf{K}), \quad \tilde{\mathbf{E}} \sim \mathrm{MN}_{d\times n}(\mathbf{0}, \mathbf{V}_e, \mathbf{I}_{n\times n}), \tag{1}$$

where n is the number of individuals, d is the number of phenotypes, $\tilde{\mathbf{Y}}$ is a d by n matrix of phenotypes, $\tilde{\mathbf{W}}$ is a c by n matrix of covariates including a row of 1s as intercept and $\mathbf{A}$ is a d by c matrix of corresponding coefficients, $\tilde{\mathbf{x}}$ is a n-vector of genotype for a particular marker and $\boldsymbol{\beta}$ is a d-vector of its effect sizes for the d phenotypes, $\tilde{\mathbf{G}}$ is a d by n matrix of random effects, $\tilde{\mathbf{E}}$ is a d by n matrix of residual errors, $\mathbf{K}$ is a known n by n relatedness matrix, $\mathbf{I}_{n\times n}$ is a n by n identity matrix, $\mathbf{V}_g$ is a d by d symmetric matrix of genetic variance component, $\mathbf{V}_e$ is a d by d symmetric matrix of environmental variance component and $\mathrm{MN}_{d\times n}(\mathbf{0}, \mathbf{V_1}, \mathbf{V_2})$ denotes the $d \times n$ matrix normal distribution with mean 0, row covariance matrix $\mathbf{V}_1$ (d by d), and column covariance matrix $\mathbf{V}_2$ (n by n).

We group all covariates together into a $(c+1)$ by n matrix $\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{W}} \\ \tilde{\mathbf{x}}^T \end{pmatrix}$, and group all coefficients together into a d by $(c+1)$ matrix $\mathbf{B} = (\mathbf{A}, \boldsymbol{\beta})$.

Following[2,3,4], we perform an eigen-decomposition of the relatedness matrix $\mathbf{K} = \mathbf{U}_k\mathbf{D}_k\mathbf{U}_k^T$, where $\mathbf{U}_k$ is a n by n orthogonal matrix of eigen vectors and $\mathbf{D}_k$ is a diagonal n by n matrix filled with the corresponding eigen values, or $\mathrm{diag}(\delta_1, \cdots, \delta_n)$. We then obtain transformed phenotype matrix $\mathbf{Y} = \tilde{\mathbf{Y}}\mathbf{U}_k$ and transformed covariate matrix $\mathbf{X} = \tilde{\mathbf{X}}\mathbf{U}_k$. We further denote $\mathbf{G} = \tilde{\mathbf{G}}\mathbf{U}_k$ as the transformed random effect matrix, and $\mathbf{E} = \tilde{\mathbf{E}}\mathbf{U}_k$ as the transformed residual error matrix. Now, the transformed phenotypes given the transformed covariates follow

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{G} + \mathbf{E}; \quad \mathbf{G} \sim \mathrm{MN}(0, \mathbf{V}_g, \mathbf{D}_k), \quad \mathbf{E} \sim \mathrm{MN}(0, \mathbf{V}_e, \mathbf{I}_{n\times n}), \tag{2}$$

which is equivalent to

$$\mathbf{y} = \mathbf{X}^T \otimes \mathbf{I}_{d\times d}\mathbf{b} + \mathbf{g} + \mathbf{e}; \quad \mathbf{g} \sim \mathrm{MVN}(0, \mathbf{D}_k \otimes \mathbf{V}_g), \quad \mathbf{e} \sim \mathrm{MVN}(0, \mathbf{I}_{n\times n} \otimes \mathbf{V}_e), \tag{3}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\mathbf{b} = \mathrm{vec}(\mathbf{B})$, $\mathbf{g} = \mathrm{vec}(\mathbf{G})$, $\mathbf{e} = \mathrm{vec}(\mathbf{E})$, vec denotes vectorization (i.e. stacking columns), MVN denotes multivariate normal distribution and $\otimes$ denotes Kronecker product.

Therefore, for each individual l, the transformed phenotypes given the transformed covariates follow independent (but not identical) multivariate normal distributions

$$\mathbf{y}_l = \mathbf{B}\mathbf{x}_l + \mathbf{g}_l + \mathbf{e}_l; \quad \mathbf{g}_l \sim \mathrm{MVN}(0, \delta_l\mathbf{V}_g), \quad \mathbf{e}_l \sim \mathrm{MVN}(0, \mathbf{V}_e), \tag{4}$$

with variance $\mathbf{V}_l = \delta_l \mathbf{V}_g + \mathbf{V}_e$, where $\mathbf{y}_l$ is the lth column vector of $\mathbf{Y}$, $\mathbf{x}_l$ is lth column vector of $\mathbf{X}$, $\mathbf{g}_l$ is lth column vector of $\mathbf{G}$, $\mathbf{e}_l$ is lth column vector of $\mathbf{E}$, $\forall l = 1, \cdots, n$.

3.2 Optimization Method Overview

We are interested in obtaining parameter estimates from this model, which are used further to obtain statistics and p values to test the null hypothesis that the marker effect sizes for all phenotypes are zero, $H_0 : \boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{0}$ is a d-vector of zeros, against the general alternative $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$. This test, in the bivariate case, corresponds to the "full test" in MTMM. We do not consider either the "interaction test" or the "common test" in MTMM here.

Parameter estimation in a mvLMM presents substantial computational challenges, in part because it requires multi-dimensional optimization for a potentially non-convex function. Procedures for multi-dimensional optimization can be classified into two categories based on whether or not they use derivatives. Derivative-free methods evaluate the (restricted) likelihood function for every combination of parameters along a searching path[5,6,7]. They are easy to implement, but are often computationally inefficient: their time complexity grows exponentially with the number of parameters, making them impractical for a reasonably large number of phenotypes[8]. (For instance, the original paper on the derivative-free method for mvLMM only showed examples for two phenotypes[7].) The derivative-based methods include the expectation maximization (EM) algorithm[9] and its accelerated version using parameter expansion (PX-EM)[10,11]; and the Newton-Raphson (NR) algorithm[12,13] and its variant, the average information (AI) algorithm[14]. Because of the stability of EM-type algorithms (each iteration is guaranteed to increase the likelihood), and the faster convergence rate of NR-type algorithms, the two are often combined to gain the benefits of both (e.g. PX-AI algorithm)[15]. This strategy is used in many existing software packages, including the free packages GCTA[16,17], and WOMBAT[18], and the commercial package ASREML[14].

Unfortunately, even with the PX-AI algorithm, the per-iteration computation time for fitting a mvLMM still increases cubically, or worse, both with the number of individuals (n) and with the number of phenotypes (d) (the computational complexity is $O(n^3d^3)$ for EM and $O(n^3d^7)$ for AI). This is because existing method require repeated "inversion" (actually, solving a system of linear equations) of an $nd \times nd$ matrix, in every iteration of the EM-like algorithm, and for evaluating every element inside the average information matrix (which is a $d(d+1)$ by $d(d+1)$ matrix) during each iteration of the NR-like algorithm, a computationally expensive procedure which increases cubically with both n and d ($O(n^3d^3)$). This becomes especially problematic in GWASs where the optimizations are performed for every SNP in turn. To address this issue,[19] recently introduced the multi-trait mixed model (MTMM) method[19], implemented in the MTMM software, to use an approximation strategy[20,21,22] to reduce computation time from cubic to quadratic in n. Specifically, the approximation avoids repeatedly re-optimizing the variance components under the alternative model for each SNP, by re-using part of the pre-estimated variance components under the null

model (fit using the software ASREML) to approximate the likelihood ratio statistic. However, the approximated likelihood ratio statistic $2l_1(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_0) - 2l_0(\hat{\boldsymbol{\theta}}_0', \hat{\boldsymbol{\theta}}_0)$ is guaranteed to be less than the true likelihood ratio statistic $2l_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_1) - 2l_0(\hat{\boldsymbol{\theta}}_0', \hat{\boldsymbol{\theta}}_0)$, since $l_1(\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_0) < l_1(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_1)$, where $\boldsymbol{\theta}$ are part of the variance component parameters to be fixed from H_0, $\boldsymbol{\theta}'$ are part of the variance component parameters to be estimated from H_1, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}'$ and $\hat{\boldsymbol{\beta}}$ are MLE estimates, while $\tilde{\boldsymbol{\theta}}_1'$ and $\tilde{\boldsymbol{\beta}}$ are conditional MLE estimates given $\hat{\boldsymbol{\theta}}_0$.

Here, we present novel algorithms that substantially reduce the computation burden. Our algorithms combine recently described univariate LMM tricks[2,3,4], with the simultaneous diagonalization (known as the canonical transformation in animal breeding literatures[7,23]) for the PX-EM algorithm, and with a few block-diagonal matrix and sparse matrix properties for the NR algorithm. In effect, our algorithms provide the multivariate analogue of the univariate algorithms EMMA[24], and FaSTLMM/GEMMA/CM[2,3,4]. Specifically, with one $O(n^3)$ operation upfront,

1. The EMMA algorithm reduced the computational cost per iteration for a *single* univariate LMM ($d = 1$) from $O(n^3)$ to $O(n)$; in the multivariate case our algorithms reduce $O(n^3d^3)$ to $O(nd^2)$ for EM and reduce $O(n^3d^7)$ to $O(nd^6)$ for NR .

2. FaSTLMM/GEMMA/CM reduced the computation cost per SNP for univariate LMMs from $O(n^3)$ to $O(n^2)$ (or $O(n)$ if $\mathbf{K}$ has low rank[2,25]); in the multivariate case our algorithms reduce $O(n^3d^3)$ per SNP to $O(n^2)$ (or $O(n)$ if $\mathbf{K}$ has low rank).

Our algorithms also obviate the need for the widely used AI algorithm[14] because our implementation of the NR algorithm has the same time complexity and practical computation time.

For numerical optimization in the null model, we initialize the two variance components to be both diagonal matrices, with diagonal elements estimated from the corresponding univariate LMMs. We then perform the PX-EM algorithm, as described in details below, for 10,000 iterations or until the log likelihood increase between two consecutive iterations is below 10^{-4}. Afterwards, we perform the NR algorithm, as described in details below, using variance component estimates from the previous PX-EM algorithm, for another 100 iterations or until the log likelihood increase between two consecutive iterations is below 10^{-4}. For GWAS applications, for each SNP tested, we use the variance components estimated from the null model as initial values. Because for moderate d the PX-EM algorithm is considerably faster than the NR algorithm, we perform the NR algorithm only for markers where the p value after the PX-EM algorithm is below 1.0×10^{-3}. With the above thresholds, it often takes hundreds to thousands PX-EM iterations followed by a dozen NR iterations to optimize the null model, and often takes a few dozen PX-EM iterations followed by a couple NR iterations to optimize the alternative model for each SNP. Notice that all the precision thresholds and maximal iterations listed above can be adjusted in GEMMA.

In our experience, both the PX-EM and NR algorithms are required for optimization, consistent with previous observations (e.g.[14,15,18,16,17]). On one hand, although the PX-EM algorithm is

stable (each iteration is guaranteed to increase the likelihood), it is often too slow to converge. For example, in many data sets we have worked on, even thousands of PX-EM iterations may not be sufficient to maximize the null model (e.g. the difference between the resulting log likelihood and the true maximal value can be as large as one). And in many GWAS applications, using the PX-EM algorithm alone will fail to optimize the alternative model in a large fraction of SNPs (an average of 83% of the p values from the pairwise analyses using PX-EM alone are closer to the MTMM p values than to the p values obtained by the PX-EM plus NR algorithms in the HMDP data set; Supplementary Figure 1). On the other hand, the NR algorithm is fast to converge given a good initial value (only takes a few iterations), but can easily fail to do so given a bad starting point. Therefore, we follow previous approaches[14,15,18,16,17] and combine the two algorithms together, with the PX-EM algorithm providing a good starting value for the following NR algorithm. In addition, for moderate d the PX-EM algorithm is considerably faster than the NR algorithm, and so for GWAS applications, we perform the NR algorithm only for markers where the p value after the PX-EM algorithm is $< 1.0 \times 10^{-3}$ (or a user adjusted threshold). This strategy makes GWAS analysis a few times faster than using NR algorithm for every marker, without noticeable loss of accuracy.

3.3 PX-EM Algorithms

Here, we describe an expectation conditional maximization (ECM) algorithm[26] for finding maximum likelihood estimates (MLE) in mvLMM, an expectation maximization (EM) algorithm[9] for finding restricted maximum likelihood estimates (REMLE) in mvLMM, parameter expansion (PX) versions[10,11] of the two, and their efficient computations.

3.3.1 An ECM Algorithm for MLE

We view $\mathbf{G}$ as missing values, and we have the joint likelihood function as

$$\log l(\mathbf{Y}, \mathbf{G}|\mathbf{B}, \mathbf{V}_g, \mathbf{V}_e) = \sum_{l=1}^{n} \{-d\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_e| - \frac{1}{2}\log|\delta_l \mathbf{V}_g| - \frac{1}{2}\mathbf{e}_l^T \mathbf{V}_e^{-1}\mathbf{e}_l - \frac{1}{2}\mathbf{g}_l^T(\delta_l \mathbf{V}_g)^{-1}\mathbf{g}_l\}. \tag{5}$$

The conditional distribution of $\mathbf{G}$ given $\mathbf{Y}$ and the current values of $\mathbf{B}^{(t)}$, $\mathbf{V}_g^{(t)}$, $\mathbf{V}_e^{(t)}$ follows

$$\mathbf{g}_l|\mathbf{y}_l, \mathbf{B}^{(t)}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{g}}_l^{(t)}, \hat{\mathbf{\Sigma}}_l^{(t)}), \tag{6}$$

where $\mathbf{V}_l^{(t)} = \delta_l \mathbf{V}_g^{(t)} + \mathbf{V}_e^{(t)}$, $\hat{\mathbf{g}}_l^{(t)} = \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1}(\mathbf{y}_l - \mathbf{B}^{(t)}\mathbf{x}_l)$ and $\hat{\mathbf{\Sigma}}_l^{(t)} = \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1}\mathbf{V}_e^{(t)}$.

The expected value of the log likelihood function, with respect to the conditional distribution

of $\mathbf{G}$ given $\mathbf{Y}$ and the current values of $\mathbf{B}^{(t)}$, $\mathbf{V}_g^{(t)}$, $\mathbf{V}_e^{(t)}$, is

$$
\begin{aligned}
&E_{\mathbf{G}|\mathbf{Y},\mathbf{B}^{(t)},\mathbf{V}_g^{(t)},\mathbf{V}_e^{(t)}}[\log l(\mathbf{Y},\mathbf{G}|\mathbf{B},\mathbf{V}_g,\mathbf{V}_e)] \\
&= \sum_{l=1}^{n}\{-d\log(2\pi) - \frac{1}{2}\log|\delta_l\mathbf{V}_g| - \frac{1}{2}\log|\mathbf{V}_e| - \frac{1}{2}(\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)^T\mathbf{V}_e^{-1}(\mathbf{y}_l - \mathbf{B}\mathbf{x}_l) \\
&- \frac{1}{2}(\hat{\mathbf{g}}_l^{(t)})^T((\delta_l\mathbf{V}_g)^{-1} + \mathbf{V}_e^{-1})^{-1}\hat{\mathbf{g}}_l^{(t)} - \frac{1}{2}\text{trace}(((\delta_l\mathbf{V}_g)^{-1} + \mathbf{V}_e^{-1})^{-1}\hat{\mathbf{\Sigma}}_l^{(t)}) + \hat{\mathbf{g}}_l^{(t)}\mathbf{V}_e^{-1}(\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)\}.
\end{aligned} \tag{7}
$$

We optimize the above expectation using two conditional maximization steps, in which $\mathbf{B}^{(t+1)}$ is updated conditional on $\mathbf{V}_g^{(t)}$ and $\mathbf{V}_e^{(t)}$, and $\mathbf{V}_g^{(t+1)}$, $\mathbf{V}_e^{(t+1)}$ are updated conditional on $\mathbf{B}^{(t+1)}$, $\mathbf{V}_g^{(t)}$ and $\mathbf{V}_e^{(t)}$, or

$$\mathbf{B}^{(t+1)} = (\mathbf{Y} - \hat{\mathbf{G}}^{(t)})\mathbf{X}(\mathbf{X}\mathbf{X}^T)^{-1}, \tag{8}$$

$$\mathbf{V}_g^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n}\delta_l^{-1}(\hat{\mathbf{g}}_l^{(t)}(\hat{\mathbf{g}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}_l^{(t)}), \tag{9}$$

$$\mathbf{V}_e^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n}(\hat{\mathbf{e}}_l^{(t)}(\hat{\mathbf{e}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}_l^{(t)}), \tag{10}$$

where $\hat{\mathbf{G}}^{(t)}$ is a d by n matrix with lth column $\hat{\mathbf{g}}_l^{(t)}$, $\hat{\mathbf{e}}_l^{(t)} = \mathbf{y}_l - \mathbf{B}^{(t+1)}\mathbf{x}_l - \hat{\mathbf{g}}_l^{(t)}$. We note that the derivation of the last two equations requires obtaining the partial derivatives with respect to $\text{vec}(\mathbf{V}_g)$ and $\text{vec}(\mathbf{V}_e)$ based on a few matrix calculus properties listed in[3].

3.3.2 An EM Algorithm for REMLE

We view both $\mathbf{B}$ and $\mathbf{G}$ as missing values. The joint likelihood function remains the same as in equation 5, and the joint conditional distribution of $\mathbf{B}$, $\mathbf{G}$ given $\mathbf{Y}$ and the current values of $\mathbf{V}_g^{(t)}$, $\mathbf{V}_e^{(t)}$ is

$$\begin{pmatrix}\mathbf{b}\\ \mathbf{g}\end{pmatrix}|\mathbf{Y},\mathbf{V}_g^{(t)},\mathbf{V}_e^{(t)} \sim \text{MVN}(\begin{pmatrix}\hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)}\\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)}\end{pmatrix}\begin{pmatrix}(\mathbf{X}\otimes(\mathbf{V}_e^{(t)})^{-1})\mathbf{y}\\ (\mathbf{I}_{n\times n}\otimes(\mathbf{V}_e^{(t)})^{-1})\mathbf{y}\end{pmatrix}, \begin{pmatrix}\hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)}\\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)}\end{pmatrix}), \tag{11}$$

where

$$\begin{pmatrix}\hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)}\\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\mathbf{X}^T\otimes(\mathbf{V}_e^{(t)})^{-1} & \mathbf{X}\otimes(\mathbf{V}_e^{(t)})^{-1}\\ \mathbf{X}^T\otimes(\mathbf{V}_e^{(t)})^{-1} & \mathbf{D}_k^{-1}\otimes(\mathbf{V}_g^{(t)})^{-1} + \mathbf{I}_{n\times n}\otimes(\mathbf{V}_e^{(t)})^{-1}\end{pmatrix}^{-1}. \tag{12}$$

Therefore,

$$\mathbf{g}_l|\mathbf{Y}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{g}}_l^{(t)}, \hat{\mathbf{\Sigma}}_{l,\mathbf{gg}}^{(t)}), \tag{13}$$

$$\mathbf{e}_l = \mathbf{y}_l - \mathbf{B}\mathbf{x}_l - \mathbf{g}_l|\mathbf{Y}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{e}}_l^{(t)}, \hat{\mathbf{\Sigma}}_{l,\mathbf{ee}}^{(t)}), \tag{14}$$

where

$$\hat{\mathbf{b}}^{(t)} = \hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} \sum_{l=1}^{n} \mathbf{x}_k \otimes (\mathbf{V}_l^{(t)})^{-1}\mathbf{y}_l, \tag{15}$$

$$\hat{\mathbf{g}}_l^{(t)} = \sum_{l=1}^{n} \delta_l \mathbf{V}_g^{(t)} (\mathbf{V}_l^{(t)})^{-1}(\mathbf{y}_l - \hat{\mathbf{B}}^{(t)}\mathbf{x}_l), \tag{16}$$

$$\hat{\mathbf{e}}_l^{(t)} = \mathbf{y}_l - \hat{\mathbf{B}}^{(t)}\mathbf{x}_l - \hat{\mathbf{g}}_l^{(t)}, \tag{17}$$

$$\hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} = \sum_{l=1}^{n} \mathbf{x}_l\mathbf{x}_l^T \otimes (\mathbf{V}_l^{(t)})^{-1}, \tag{18}$$

$$\hat{\mathbf{\Sigma}}_{l,\mathbf{gg}}^{(t)} = \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1} + (\mathbf{x}_l^T \otimes \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1})(\sum_{l=1}^{n} \mathbf{x}_l\mathbf{x}_l^T \otimes \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1})^{-1}(\mathbf{x}_l \otimes \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1}), \tag{19}$$

$$\hat{\mathbf{\Sigma}}_{l,\mathbf{ee}}^{(t)} = \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1} + (\mathbf{x}_l^T \otimes (\mathbf{V}_l^{(t)})^{-1})(\sum_{l=1}^{n} \mathbf{x}_l\mathbf{x}_l^T \otimes \delta_l \mathbf{V}_g^{(t)}(\mathbf{V}_l^{(t)})^{-1})^{-1}(\mathbf{x}_l \otimes (\mathbf{V}_l^{(t)})^{-1}), \tag{20}$$

and $\hat{\mathbf{B}}^{(t)}$ is the matrix satisfies $\text{vec}(\hat{\mathbf{B}}^{(t)}) = \hat{\mathbf{b}}^{(t)}$.

The expected value of the log likelihood function, with respect to the conditional distribution of $\mathbf{B}$, $\mathbf{G}$ given $\mathbf{Y}$ and the current values of $\mathbf{V}_g^{(t)}$, $\mathbf{V}_e^{(t)}$, is

$$\begin{aligned}
&E_{\mathbf{B},\mathbf{G}|\mathbf{Y},\mathbf{V}_g^{(t)},\mathbf{V}_e^{(t)}}[\log l(\mathbf{Y}, \mathbf{G}, \mathbf{B}|\mathbf{V}_g, \mathbf{V}_e)] \\
&= \sum_{l=1}^{n}\{-d\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_e| - \frac{1}{2}\log|\delta_l\mathbf{V}_g| - \frac{1}{2}(\hat{\mathbf{e}}_l^{(t)})^T\mathbf{V}_e^{-1}\hat{\mathbf{e}}_l^{(t)} - \frac{1}{2}\text{trace}(\mathbf{V}_e^{-1}\hat{\mathbf{\Sigma}}_{l,\mathbf{ee}}^{(t)}) \\
&- \frac{1}{2}(\hat{\mathbf{g}}_l^{(t)})^T(\delta_l\mathbf{V}_g)^{-1}\hat{\mathbf{g}}_l^{(t)} - \frac{1}{2}\text{trace}((\delta_l\mathbf{V}_g)^{-1}\hat{\mathbf{\Sigma}}_{l,\mathbf{gg}}^{(t)})\}.
\end{aligned} \tag{21}$$

We update $\mathbf{V}_g^{(t+1)}$ and $\mathbf{V}_e^{(t+1)}$ to maximize the above expectation

$$\mathbf{V}_g^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n} \delta_l^{-1}(\hat{\mathbf{g}}_l^{(t)}(\hat{\mathbf{g}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}_{l,\mathbf{gg}}^{(t)}), \tag{22}$$

$$\mathbf{V}_e^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n} (\hat{\mathbf{e}}_l^{(t)}(\hat{\mathbf{e}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}_{l,\mathbf{ee}}^{(t)}). \tag{23}$$

3.3.3 PX versions of ECM and EM

We introduce a latent parameter $\mathbf{V}_a$ as a d by d matrix to ensure $\mathbf{V}_g = \mathbf{V}_a \mathbf{V}_g^* \mathbf{V}_a$. The mvLMM with the new parameterization becomes

$$\mathbf{y}_l = \mathbf{B}\mathbf{x}_l + \mathbf{V}_a \mathbf{g}_l^* + \mathbf{e}_l \quad \mathbf{g}_l^* \sim \text{MVN}(0, \delta_l \mathbf{V}_g^*) \quad \mathbf{e}_l \sim \text{MVN}(0, \mathbf{V}_e), \tag{24}$$

The expectation of the log likelihood function in the ECM or the EM algorithm is taken at $\mathbf{V}_a^{(t)} = \mathbf{I}_{d\times d}$. The updates for $\mathbf{V}_g^*$, $\mathbf{V}_e$ (and $\mathbf{B}$ for ECM) remain the same, and the update for $\mathbf{V}_a$ is

$$\mathbf{V}_a^{(t+1)} = (\frac{1}{n}\sum_{l=1}^{n} E((\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)(\mathbf{g}_l^*)^T))(\frac{1}{n}\sum_{l=1}^{n} E(\mathbf{g}_l^*(\mathbf{g}_l^*)^T))^{-1}. \tag{25}$$

where the expectations are taken with respect to the conditional distribution of $\mathbf{G}^*$ (and $\mathbf{B}$ for EM) given $\mathbf{Y}$ and the current values of $\mathbf{V}_g^{(t)}$, $\mathbf{V}_e^{(t)}$ (and $\mathbf{B}^{(t+1)}$ for ECM).

3.3.4 Efficient computation

The most computationally expensive part of the ECM/EM algorithm is the evaluation of each $\mathbf{V}_l^{-1}$ and further the calculation of quantities that involve these inverses. A naive brute force approach will make the computation cubic in the number of traits, which can be avoided, by performing a transformation that further converts correlated traits into uncorrelated ones (in addition to the transformation that we have already performed to convert correlated individuals into uncorrelated ones). The idea behind this is commonly referred to as the canonical transformation in animal breeding literatures (e.g.[7,23] and references there in), or as the simultaneous diagonalization in linear algebra.

More specifically, for each value of $\mathbf{V}_g$ and $\mathbf{V}_e$, we perform an eigen decomposition of the matrix $\mathbf{V}_e^{-\frac{1}{2}} \mathbf{V}_g \mathbf{V}_e^{-\frac{1}{2}} = \mathbf{U}_\lambda \mathbf{D}_\lambda \mathbf{U}_\lambda^T$, and we transform each phenotype vector $\mathbf{y}_l$ and each covariate vector $\mathbf{x}_l$ by multiplying $\mathbf{U}_\lambda^T \mathbf{V}_e^{-\frac{1}{2}}$. After that, for each individual, the transformed phenotypes given the transformed genotypes will follow independent univariate normal distributions (rather than multivariate normal distributions). Subsequently, each $\mathbf{V}_l^{-1} = \mathbf{V}_e^{-\frac{1}{2}} \mathbf{U}_\lambda (\mathbf{D}_\lambda + \mathbf{I}_{d\times d})^{-1} \mathbf{U}_\lambda^T \mathbf{V}_e^{-\frac{1}{2}}$ and quantities involving $\mathbf{V}_l^{-1}$ can be calculated efficiently.

Therefore, the computation complexity for each iteration in the (PX) ECM/EM algorithm is $O(nc^2d^2)$.

3.4 Newton-Raphson's Algorithms

Here, we describe Newton-Raphson's algorithms for MLE and REMLE estimation in mvLMM. Although an average-information algorithm[14] has often been used in place of a standard Newton-Raphson's algorithm, we found it unnecessary when we use the efficient algorithms described below.

3.4.1 Target functions and partial derivatives

Both the log-likelihood function and the log-restricted likelihood function can be expressed as functions for $\mathbf{V}_g$ and $\mathbf{V}_e$ only:

$$l(\mathbf{V}_g, \mathbf{V}_e) = -\frac{nd}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}| - \frac{1}{2}\mathbf{y}^T\mathbf{P}\mathbf{y}, \tag{26}$$

$$\begin{aligned} l_r(\mathbf{V}_g, \mathbf{V}_e) = &-\frac{(n-c-1)d}{2}\log(2\pi) + \frac{d}{2}\log|\mathbf{X}\mathbf{X}^T| - \frac{1}{2}\log|\mathbf{H}| \\ &- \frac{1}{2}\log|(\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d})| - \frac{1}{2}\mathbf{y}^T\mathbf{P}\mathbf{y}, \end{aligned} \tag{27}$$

where

$$\mathbf{H} = \mathbf{D}_k\otimes\mathbf{V}_g + \mathbf{I}_{n\times n}\otimes\mathbf{V}_e = \mathrm{diag}(\mathbf{V}_l), \tag{28}$$

$$\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d})((\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d}))^{-1}(\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}. \tag{29}$$

With a slight abuse of notation, we denote $v_{g,ij}$ as the ijth element of $\mathbf{V}_g$, $v_{e,ij}$ as the ijth element of $\mathbf{V}_e$, $\mathbf{I}_i$ as a d-vector with ith element 1 and other elements 0, and $\mathbf{I}_{ij} = \mathbf{I}_i\mathbf{I}_j^T$ as a d by d matrix with ijth element 1 and other elements 0. We have

$$\frac{\partial\mathrm{vec}(\mathbf{D}_k\otimes\mathbf{V}_g)}{\partial v_{g,ij}} = \mathrm{vec}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\frac{1}{1+1_{i=j}}, \tag{30}$$

$$\frac{\partial\mathrm{vec}(\mathbf{I}_{n\times n}\otimes\mathbf{V}_e)}{\partial v_{e,ij}} = \mathrm{vec}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\frac{1}{1+1_{i=j}}, \tag{31}$$

where $1_{i=j}$ is an indicator function that takes value 1 when i equals j and 0 otherwise.

With a few matrix calculus properties listed in[3], we obtain the first order partial derivatives for the log-likelihood and the log-restricted likelihood functions

$$\frac{\partial l}{\partial v_{g,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{32}$$

$$\frac{\partial l}{\partial v_{e,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{33}$$

$$\frac{\partial l_r}{\partial v_{g,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{P}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{34}$$

$$\frac{\partial l_r}{\partial v_{e,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{35}$$

and the second order partial derivatives for the log-likelihood function

$$\frac{\partial l^2}{\partial v_{g,ij}\partial v_{g,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{H}^{-1}(\mathbf{D}_k \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (36)$$

$$\frac{\partial l^2}{\partial v_{g,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{H}^{-1}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (37)$$

$$\frac{\partial l^2}{\partial v_{e,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{H}^{-1}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{H}^{-1}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (38)$$

and second order partial derivatives for the log-restricted likelihood function

$$\frac{\partial l_r^2}{\partial v_{g,ij}\partial v_{g,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (39)$$

$$\frac{\partial l_r^2}{\partial v_{g,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (40)$$

$$\frac{\partial l_r^2}{\partial v_{e,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$- \mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n} \otimes (\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}. \quad (41)$$

3.4.2 Efficient computation

We describe in this section the efficient calculations of the target functions, the first-order partial derivatives with respect to $v_{g,ij}$, and the second-order partial derivatives with respect to $v_{g,ij}$ and $v_{g,i'j'}$. The first-order and second-order partial derivatives with respect to other parameters can be calculated in a similar fashion. The calculations described here use basic properties of block diagonal matrices and sparse matrices.

We denote $\mathbf{Q} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{X}^T \otimes \mathbf{I}_{d\times d})$, $\mathbf{q} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}\mathbf{y}$, $q = \mathbf{y}^T\mathbf{H}^{-1}\mathbf{y}$, and with a

slight abuse of notation, denote

$$\mathbf{Q}_{ij}^{g} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{X}^T \otimes \mathbf{I}_{d\times d}), \tag{42}$$

$$\mathbf{q}_{ij}^{g} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}\mathbf{y}, \tag{43}$$

$$q_{ij}^{g} = \mathbf{y}^T\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}\mathbf{y}, \tag{44}$$

$$\mathbf{Q}_{ij,i'j'}^{gg} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}(\mathbf{X}^T \otimes \mathbf{I}_{d\times d}), \tag{45}$$

$$\mathbf{q}_{ij,i'j'}^{gg} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}\mathbf{y}, \tag{46}$$

$$q_{ij,i'j'}^{gg} = \mathbf{y}^T\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}\mathbf{y}. \tag{47}$$

For the trace terms, we have

$$\mathrm{trace}(\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) = \mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}), \tag{48}$$

$$\begin{aligned}\mathrm{trace}(\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) &= \mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) \\ &\quad - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij,i'j'}^{gg}) - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{i'j',ij}^{gg}) + \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}).\end{aligned} \tag{49}$$

For the vector-matrix-vector product terms, we have

$$\mathbf{y}^T\mathbf{P}\mathbf{y} = q - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}, \tag{50}$$

$$\mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}\mathbf{y} = q_{ij}^{g} - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}_{ij}^{g} - (\mathbf{q}_{ij}^{g})^T\mathbf{Q}^{-1}\mathbf{q} + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{q}, \tag{51}$$

$$\begin{aligned}\mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{P}\mathbf{y} &= q_{ij,i'j'}^{gg} - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}_{ij,i'j'}^{gg} - (\mathbf{q}_{j'i',ji}^{gg})^T\mathbf{Q}^{-1}\mathbf{q} - (\mathbf{q}_{ji}^{g})^T\mathbf{Q}^{-1}\mathbf{q}_{i'j'}^{g} \\ &\quad + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{q}_{i'j'}^{g} + (\mathbf{q}_{ji}^{g})^T\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}\mathbf{Q}^{-1}\mathbf{q} + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij,i'j'}^{gg}\mathbf{q} \\ &\quad - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}\mathbf{Q}^{-1}\mathbf{q}.\end{aligned} \tag{52}$$

Therefore, it suffices to efficiently evaluate

$$|\mathbf{H}| = \sum_{l=1}^{n} |\mathbf{V}_l|, \tag{53}$$

$$\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) = \sum_{l=1}^{n} \delta_l(\mathbf{I}_j^T\mathbf{V}_l^{-1}\mathbf{I}_i), \tag{54}$$

$$\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) = \sum_{l=1}^{n} \delta_l^2(\mathbf{I}_{j'}^T\mathbf{V}_l^{-1}\mathbf{I}_i)(\mathbf{I}_j^T\mathbf{V}_l^{-1}\mathbf{I}_{i'}), \tag{55}$$

and

$$\mathbf{Q} = \sum_{l=1}^{n} (\mathbf{x}_l \mathbf{x}_l^T) \otimes \mathbf{V}_l^{-1}, \tag{56}$$

$$\mathbf{q} = \sum_{l=1}^{n} \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{y}_l), \tag{57}$$

$$q = \sum_{l=1}^{n} \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{y}_l, \tag{58}$$

and

$$\mathbf{Q}_{ij}^{g} = \sum_{l=1}^{n} \delta_l (\mathbf{x}_l \mathbf{x}_l^T) \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1}), \tag{59}$$

$$\mathbf{q}_{ij}^{g} = \sum_{l=1}^{n} \delta_l \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{y}_l), \tag{60}$$

$$q_{ij}^{g} = \sum_{l=1}^{n} \delta_l \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{y}_l, \tag{61}$$

$$\mathbf{Q}_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 (\mathbf{x}_l \mathbf{x}_l^T) \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1}), \tag{62}$$

$$\mathbf{q}_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1} \mathbf{y}_l), \tag{63}$$

$$q_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1} \mathbf{y}_l. \tag{64}$$

Notice that one key trick for calculating all the above quantities is observing that $\mathbf{I}_{ij} = \mathbf{I}_i \mathbf{I}_j^T$. In this way, many of the above quantities only involve scalar multiplications or rank one matrix updates.

The most time consuming part is the calculation of $\mathbf{Q}_{ij,i'j'}^{gg}$, each requiring $O(nc^2d^2)$ computation time. The computation complexity for each iteration in the Newton-Raphson's algorithm is therefore $O(nc^2d^6)$.

3.5 Test Statistics and p Values

3.5.1 Test Statistics

We consider three common tests for mvLMMs: the likelihood ratio test, the Wald test and the score test.

The likelihood ratio test calculates the maximum log likelihoods for both the alternative ($\hat{l}_1(\hat{\mathbf{V}}_{g,1}, \hat{\mathbf{V}}_{e,1})$) and the null ($\hat{l}_0(\hat{\mathbf{V}}_{g,0}, \hat{\mathbf{V}}_{e,0})$) models. It computes a test statistics based on the difference between the two: $z_{LR} = 2(\hat{l}_1(\hat{\mathbf{V}}_{g,1}, \hat{\mathbf{V}}_{e,1}) - \hat{l}_1(\hat{\mathbf{V}}_{g,0}, \hat{\mathbf{V}}_{e,0}))$.

The Wald estimates the effect size

$$\hat{\boldsymbol{\beta}} = \sum_{l=1}^{n} x_l \hat{\mathbf{V}}_l^{-1} \mathbf{y}_l - (\sum_{l=1}^{n} (x_l \mathbf{w}_l) \otimes \hat{\mathbf{V}}_l^{-1})(\sum_{l=1}^{n} (\mathbf{w}_l \mathbf{w}_l^T) \otimes \hat{\mathbf{V}}_l^{-1})^{-1}(\sum_{l=1}^{n} \mathbf{w}_l \otimes (\hat{\mathbf{V}}_l^{-1} \mathbf{y}_l)), \tag{65}$$

and its precision $V(\hat{\boldsymbol{\beta}})^{-1}$

$$V(\hat{\boldsymbol{\beta}})^{-1} = \sum_{l=1}^{n} x_l^2 \hat{\mathbf{V}}_l^{-1} - (\sum_{l=1}^{n} (x_l \mathbf{w}_l) \otimes \hat{\mathbf{V}}_l^{-1})(\sum_{l=1}^{n} (\mathbf{w}_l \mathbf{w}_l^T) \otimes \hat{\mathbf{V}}_l^{-1})^{-1}(\sum_{l=1}^{n} (x_l \mathbf{w}_l) \otimes \hat{\mathbf{V}}_l^{-1}), \tag{66}$$

with the variance component estimates $\hat{\mathbf{V}}_g$ and $\hat{\mathbf{V}}_e$ ($\hat{\mathbf{V}}_l = \delta_l \hat{\mathbf{V}}_{g,1} + \hat{\mathbf{V}}_{e,1}$) obtained under the alternative. Above, x_l is the lth element of the transformed genotype vector $\mathbf{x}$ and $\mathbf{w}_l$ is the lth column vector of the transformed covariance matrix $\mathbf{W}$. Afterwards, it computes $z_{Wald} = \sqrt{\hat{\boldsymbol{\beta}}^T V(\hat{\boldsymbol{\beta}})^{-1} \hat{\boldsymbol{\beta}}}$.

The score test computes the score, a $d(d+1)$ vector $\hat{\mathbf{s}}$, with each element equals to the corresponding first order partial derivate described in the previous sections, and the observed information matrix, a $d(d+1)$ by $d(d+1)$ matrix $\hat{\mathbf{I}}$, whose formula is also described above. These values are evaluated with parameter estimates obtained under the null, and are used to compute $z_{Score} = \sqrt{\hat{\mathbf{s}}^T \hat{\mathbf{I}}^{-1} \hat{\mathbf{s}}}$.

3.5.2 p Value Calibration

Under the null hypothesis, all three test statistics follow a $\chi^2(d)$ distribution asymptotically. However, when the sample size is small *or* the relatedness structure is strong, then the test statistics will not follow the asymptotic distribution exactly, and p values calculated from the asymptotic distribution will not be be calibrated (see, e.g.[27,28]). To correct for this, we follow[28] and use Edgeworth-corrected critical values for the three tests. Specifically, the corrected z scores from the three tests for a given marker is

$$z_{LR}^c = z_{LR}/(1 + \frac{\hat{a}}{2d}), \tag{67}$$

$$z_{Wald}^c = \frac{-(2d + \hat{a} + \hat{b})(d+2) + \sqrt{(2d + \hat{a} + \hat{b})^2 (d+2)^2 + 8d(d+2)\hat{c} z_{Wald}}}{2\hat{c}}, \tag{68}$$

$$z_{Score}^c = \frac{(2d + \hat{a} - \hat{b})(d+2) - \sqrt{(2d + \hat{a} - \hat{b})^2 (d+2)^2 - 8d(d+2)\hat{c} z_{Score}}}{2\hat{c}}, \tag{69}$$

where

$$\hat{a} = 2\hat{b} - \hat{c}, \tag{70}$$

$$\hat{b} = 2\sum_{i<j}\sum_{i'<j'}\hat{\Lambda}_{ij,i'j'}\text{trace}((\mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{ij,i'j'}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T - \mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{ij}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{i'j'}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)(\mathbf{R}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)^{-1}), \tag{71}$$

$$\hat{c} = \sum_{i<j}\sum_{i'<j'}\hat{\Lambda}_{ij,i'j'}(\text{trace}((\mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{ij}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)(\mathbf{R}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)^{-1}(\mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{i'j'}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)(\mathbf{R}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)^{-1}),$$
$$+\frac{1}{2}\text{trace}((\mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{ij}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)(\mathbf{R}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)^{-1})\text{trace}((\mathbf{R}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{Q}}_{i'j'}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)(\mathbf{R}\hat{\mathbf{Q}}^{-1}\mathbf{R}^T)^{-1})) \tag{72}$$

are evaluated with the variance component estimates from the alternative model. $\mathbf{\Lambda}$ is the inverse of the Hessian matrix and Λ_{ij} is its ijth element, and $\mathbf{R}$ is a d by cd matrix with right most d by d matrix a diagonal matrix and all other elements 0.

Notice that the corrections are marker-specific, and require estimates and partial derivatives from the alternative model. In our experience, without the corrections, the score test is often too conservative, the Wald test is often too anti-conservative, while the likelihood ratio test behaves between the two and has the correct control for type I error.

3.6 Phenotype Imputation

The tricks used in our mvLMM algorithms require complete or imputed phenotypes. Since in a typical study many individuals may have partially missing phenotypes, removing all such individuals could substantially reduce power. To address this we developed a phenotype imputation scheme for mvLMMs, which can be applied to impute missing phenotypes before applying our LRT methods.

The imputation method first estimates parameters of the mvLMM under the null model using individuals with fully observed phenotypes, and then, conditional on these estimates and the observed phenotype data, imputes missing phenotypes using their conditional means.

Specifically, we first estimate $\hat{\mathbf{b}}$, $\hat{\mathbf{V}}_g$ and $\hat{\mathbf{V}}_e$ in the null model using individuals with completely observed phenotypes. Afterwards, we impute missing phenotype values using the conditional mean given observed phenotypes and estimated parameters. Denote n_o as the number of observed values, n_m as the number of missing values ($n_o + n_m = nd$), $\mathbf{y}_o$ as a n_o vector of observed values and $\mathbf{X}_o$ as a n_o by dc matrix of corresponding covariates, $\mathbf{y}_m$ as a n_m vector of missing values and $\mathbf{X}_m$ as a n_m by dc matrix of corresponding covariates. Under the null mvLMM, $\mathbf{y} = [\mathbf{y}_o^T, \mathbf{y}_m^T]^T$ follows a multivariate normal distribution with covariance matrix $\hat{\mathbf{H}} = \mathbf{K} \otimes \hat{\mathbf{V}}_g + \mathbf{I} \otimes \hat{\mathbf{V}}_e$. Let $\hat{\mathbf{H}}_{oo}$ be the n_o by n_o sub-matrix of $\mathbf{H}$ that corresponds to the n_o observed values, and $\hat{\mathbf{H}}_{mo}$ be the n_m by n_o sub-matrix of $\hat{\mathbf{H}}$ that corresponds to the n_m missing values and n_o observed values. We use the

conditional mean of $\mathbf{y}_m$ given $\mathbf{y}_o$ and estimated parameters as an estimate for the missing values

$$\hat{\mathbf{y}}_m = \mathbf{X}_m^T \hat{\mathbf{b}} + \hat{\mathbf{H}}_{mo} \hat{\mathbf{H}}_{oo}^{-1} (\mathbf{y}_o - \mathbf{X}_o^T \hat{\mathbf{b}}). \tag{73}$$

Supplementary Figure 8 shows the results of simulations, based on both HMDP and NFBC1966 data, comparing the power of this imputation-based approach with the alternative approach of dropping individuals with partially missing phenotypes. For the HMDP simulations, because of the high relatedness, both methods achieve almost identical power as if all phenotypes are fully observed. For the NFBC1966 data, phenotype imputation achieves consistently greater power than dropping individuals, and in many simulation scenarios achieves power similar to that achieved if all phenotypes are observed (0% missingness in Supplementary Figure 8).

References

1. Henderson, C. R. *Applications of linear models in animal breeding.* University of Guelph, Guelph, (1984).

2. Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. FaST linear mixed models for genome-wide association studies. *Nature Methods* **8**, 833–835 (2011).

3. Zhou, X. and Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–824 (2012).

4. Pirinen, M., Donnelly, P., and Spencer, C. C. A. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Annals of Applied Statistics* **7**, 369–390 (2013).

5. Graser, H. U., Smith, S. P., and Tier, B. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science* **64**, 1362–1370 (1987).

6. Meyer, K. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genetics Selection Evolution* **21**, 317–340 (1989).

7. Meyer, K. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics Selection Evolution* **23**, 67–83 (1991).

8. Meyer, K. and Smith, S. P. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genetics Selection Evolution* **28**, 23–49 (1996).

9. Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood from incomplete data via the EM algirhtm. *Journal of the Royal Statistical Society: Series B* **39**, 1–38 (1977).

10. Liu, C., Rubin, D. B., and Wu, Y. N. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85**, 755–770 (1998).

11. Foulley, J.-L. and van Dyk, D. A. The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genetics Selection Evolution* **32**, 143–163 (2000).

12. Patterson, H. D. and Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).

13. Thompson, R. The estimation of variance, covariance components with an application when records are subject to culling. *Biometrics* **29**, 527–550 (1973).

14. Gilmour, A. R., Thompson, R., and Cullis, B. R. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440–1450 (1995).

15. Meyer, K. *PX×AI: Algorithmics for better convergence in restricted maximum likelihood estimation*. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brasil, (2006).

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

17. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., and Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).

18. Meyer, K. WOMBAT – A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science B* **8**, 815–821 (2007).

19. Korte, A., Vilhjlmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* **44**, 1066–1071 (2012).

20. Aulchenko, Y. S., de Koning, D.-J., and Haley, C. Genomewide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177**, 577–585 (2007).

21. Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* **42**, 355–360 (2010).

22. Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-Y., Freimer, N. B., Sabatti, C., and Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* **42**, 348–354 (2010).

23. Ducrocq, V. and Chapuis, H. Generalizing the use of the canonical transformation for the solution of multivariate mixed model equations. *Genetics Selection Evolution* **29**, 205–224 (1997).

24. Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).

25. Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., and Heckerman, D. Improved linear mixed models for genome-wide association studies. *Nature Methods* **9**, 525–526 (2012).

26. Meng, X.-L. and Rubin, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278 (1993).

27. Evans, G. B. A. and Savin, N. E. Conflict among the criteria revisited; the W, LR and LM tests. *Econometrica* **3**, 737–748 (1982).

28. Rothenberg, T. J. Hypothesis testing in linear models when the error covariance matrix is nonscalar. *Econometrica* **4**, 827–842 (1984).