Efficient Algorithms for Multivariate Linear Mixed Models in Genome-wide Association Studies

Xiang Zhou [*1] and Matthew Stephens [†1,2]

[1]Department of Human Genetics, University of Chicago
[2]Department of Statistics, University of Chicago

Abstract

Multivariate linear mixed models (mvLMMs) have been widely used in many areas of genetics, and have attracted considerable recent interest in genome-wide association studies (GWASs). However, existing methods for calculating the likelihood ratio test statistics in mvLMMs are time consuming, and, without approximations, cannot be directly applied to analyze even two traits jointly in a typical-size GWAS. Here, we present a novel algorithm for computing parameter estimates and test statistics (Likelihood ratio and Wald) in mvLMMs that i) reduces per-iteration optimization complexity from cubic to linear in the number of samples; and ii) in GWAS analyses, reduces per-marker complexity from cubic to approximately quadratic (or linear if the relatedness matrix is of low rank) in the number of samples. The new method effectively generalizes both the EMMA (Efficient Mixed Model Association) algorithm and the GEMMA (Genome-wide EMMA) algorithm to the multivariate case, making the likelihood ratio tests in GWASs with mvLMM possible, for the first time, for tens of thousands of samples and a moderate number of phenotypes (< 10). With real examples, we show that, as expected, the new method is orders of magnitude faster than competing methods in both variance component estimation in a single mvLMM, and in GWAS applications. The method is implemented in the GEMMA software package, freely available at `http://stephenslab.uchicago.edu/software.html`.

Introduction

Multivariate linear mixed models (mvLMMs) [1] have been applied to many areas of genetics, including, for example, estimating the cross-tissue heritability of gene expression [2], assessing the pleiotropy and genetic correlation between complex phenotypes [3, 4, 5, 6], detecting quantitative trait loci [7], understanding evolutionary patterns [8] and assisting animal breeding programs [9].

[*]xz7@uchicago.edu
[†]mstephens@uchicago.edu

Recently, these models have become increasingly important in genome-wide association studies (GWASs), not only because of their demonstrated effectiveness in accounting for sample relatedness [10, 7, 3] and in controlling for population stratification [3] (as for their univariate counterparts [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]), but also because of a growing appreciation of the potential gains in power from multivariate association analyses [25, 26, 27, 28, 29, 30, 3, 31]. Indeed, [31] emphasizes that, compared with standard univariate analyses, multivariate analyses can increase power not only to detect pleiotropic genetic variants that affect multiple phenotypes simultaneously, but also genetic variants that affect only one of a collection of correlated phenotypes.

Parameter estimation in a mvLMM, however, presents substantial computational challenges, in part because it requires multi-dimensional optimization. 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 [32, 33, 10]. 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 [34]. (For instance, the original paper on the derivative-free method for mvLMM only showed examples for two phenotypes [10].) The derivative-based methods include the expectation maximization (EM) algorithm [35] and its accelerated version using parameter expansion (PX-EM) [36, 37]; and the Newton-Raphson (NR) algorithm [38, 39] and its variant, the average information (AI) algorithm [40]. 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) [41]. This strategy is used in many existing software packages, including the free packages GCTA [42, 4], and WOMBAT [43], and the commercial package ASREML [40].

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 becomes especially problematic in GWASs where the optimizations are performed for every SNP in turn. To address this, [3] recently introduced the multi-trait mixed model (MTMM) method [3], implemented in the MTMM software. This method uses an approximation strategy [44, 15, 16] 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. The accuracy of this approximation is expected to be good for SNPs of small effect, but to decrease with increasing SNP effect as in the univariate case [19, 20].

Here, we present a novel computationally-efficient algorithm for GWAS analyses, based on

a novel PX-NR algorithm for parameter estimation in a standard mvLMM with two covariance components. The algorithm builds on linear algebra techniques previously used for univariate LMMs [18, 19, 20], and, combined with a few additional tricks, extends them to multivariate LMMs. In effect, the algorithm extends both the EMMA algorithm [14] and the FaSTLMM/GEMMA/CM algorithms [18, 19, 20] to the multivariate case, making likelihood ratio tests in GWAS with mvLMM possible, for the first time, for a modest number of phenotypes (e.g. 10) and a reasonably large number of individuals (e.g. 50,000).

Results

Multivariate linear mixed model

We consider the multivariate linear mixed model [1],

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

where n is the number of individuals, d is the number of phenotypes, $\mathbf{Y}$ is a d by n matrix of phenotypes, $\mathbf{W}$ 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, $\mathbf{x}$ 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, $\mathbf{G}$ is a d by n matrix of random effects, $\mathbf{E}$ 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).

Method overview

We have developed efficient algorithms for applying mvLMMs to GWASs. Specifically, for each genetic marker in turn, the algorithms perform the likelihood ratio test (or the Wald test) comparing 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, 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, although they could be easily implemented using similar ideas).

The likelihood ratio test requires the maximum likelihood estimates for the parameters $\mathbf{V}_g, \mathbf{V}_e, \mathbf{A}$ and $\boldsymbol{\beta}$ under both H_0 and H_1. Current algorithms for performing maximum likelihood estimation all require repeated "inversion" (actually, solving a system of linear equations) of an $nd \times nd$ matrix, in every iteration of the EM algorithm, and in every iteration and for every element inside the average information matrix (which is a $d(d+1)$ by $d(d+1)$ matrix) of the AI algorithm, a compu-

tationally expensive procedure which increases cubically with both n and d ($O(n^3d^3)$). Further in GWAS analyses the algorithms must repeat this computational complexity for all p SNPs. So the total complexity is $O(p(t_1n^3d^3 + t_2n^3d^7))$ where t_1 is the maximum number of iterations used for EM, and t_2 for AI.

Here we provide algorithms - both a PX-EM algorithm, and an NR algorithm - that substantially reduce this burden, to a single $O(n^3)$ operation (eigen-decomposition of the relatedness matrix K), followed by a single $O(n^2)$ operation for each SNP, and further $O(nd^2)$ operation at each iteration for PX-EM or $O(nd^6)$ for NR. We also provide similar algorithms to maximize the restricted likelihood. As noted above, there are advantages to combine an EM algorithm with an NR algorithm in this setting. 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.

The methods are described in detail in the Supplementary Information. Briefly, the algorithm combines recently described univariate LMM tricks [18, 19, 20], with the simultaneous diagonalization (known as the canonical transformation in animal breeding literatures [10, 45]) for the PX-EM algorithm, and with a few block-diagonal matrix and low rank matrix properties for the NR algorithm. In effect, the algorithms described here extend both the EMMA algorithm [14], and the FaSTLMM/GEMMA/CM algorithms [18, 19, 20], to the multivariate setting:

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/AI.

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[18, 46]); 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 [40] because our implementation of the NR algorithm has the same time complexity and practical computation time. We implemented the algorithms in the GEMMA software package [19, 22], freely available at `http://stephenslab.uchicago.edu/software.html`.

Efficient variance components estimation

To illustrate the speed gain in fitting a single mvLMM from using our method, we first compared GEMMA with two existing software packages, GCTA [42, 4] and WOMBAT [43]. We used two data sets: a mouse GWAS from the Hybrid Mouse Diversity Panel (HMDP) with four available blood lipid phenotypes (high-density lipoprotein, HDL; total cholesterol, TC; triglycerides, TG;

Table 1: Comparison of computation time of different methods for parameter estimation in a single mvLMM. Results are shown for both HMDP and NFBC1966 data sets. All computing was performed on a single core of an Intel Xeon L5420 2.50GHz CPU. n is the number of individuals, d is the number of traits, c is the number of covariates ($c = 1$ here), t_1 is the number of iterations used in the EM type algorithm and t_2 is the number of iterations used in the NR type algorithm. The implementation of GCTA does not handle more than two traits. The computation time for GEMMA is dominated by the eigen decomposition step and therefore does not vary much with the number of traits. The n^3 step in GEMMA could be replaced with a mn^2 step [14] for the HMDP data set, where m is the number of strains.

Method	Time Complexity	Computation Time					
		HMDP ($n = 656$)			NFBC1966 ($n = 5255$)		
	# Traits (d)	2	3	4	2	3	4
GEMMA	$O(n^3 + n^2d + n^2c + t_1nc^2d^2 + t_2nc^2d^6)$	< 1.0 s	< 1.0 s	< 1.0 s	6.7 min	6.7 min	6.7 min
WOMBAT	$O(t_1n^3(d + c)^3 + t_2n^3d^7)$	12.5 s	39.2 s	71.0 s	31.0 min	127.6 min	477.3 min
GCTA	$O(t_1n^3(d + c)^3 + t_2n^3d^7)$	11.2 s	–	–	38.2 min	–	–

unesterified cholesterol, UC), and a human GWAS from the Northern Finland Birth Cohort 1966 (NFBC1966) with four selected blood metabolic traits (high-density lipoprotein, HDL; low-density lipoprotein, LDL; triglycerides, TG; C-reactive protein, CRP). We fit the null mvLMM (with $\boldsymbol{\beta} = \mathbf{0}$) for all pairs of traits, all combinations of three traits, and four traits.

As expected, in all cases the three packages give essentially identical results, both for estimated variance components and their standard errors (data not shown). However, GEMMA is about 5-71 times faster (Table 1). In particular, in the NFBC1966 data set, with $d = 4$, GEMMA takes about 7 minutes compared with 8 hours for WOMBAT. The compute time for GEMMA is essentially the same for all d, because in GEMMA the compute time is dominated by the initial $O(n^3)$ eigen-decomposition step; the following optimization iterations are negligible. Thus for larger d the gains of GEMMA over other methods will be even greater (the gain not only increases cubically with the number of phenotypes, but also linearly with the number of iterations, and with larger d more iterations will typically be required).

Efficient GWAS analysis

To illustrate the speed gain in GWAS analysis, we compared GEMMA with the existing software MTMM in the same two example data sets. In contrast to GEMMA, MTMM uses an approximation to compute the likelihood ratio test statistic. Further, based on its computational complexity, GEMMA should speed computation by a factor of approximately d^2 compared with MTMM. Indeed, for both data sets, for two trait analysis, GEMMA is 2-12 times faster than MTMM (Table 2). In particular, in the NFBC1966 data set, GEMMA takes about four hours to finish a two-trait mvLMM analysis that takes MTMM almost two and a half days. In fact, GEMMA can finish the multivariate analysis for four traits within six hours.

Table 2: Computation time for GWAS analysis with mvLMM using different methods for the HMDP and NFBC1966 data sets. All computing was performed on a single core of an Intel Xeon L5420 2.50GHz CPU. The computing time for MTMM only include the multiplication of the genotype matrix with an nd by nd matrix, and does not include the time spent fitting the null model (because MTMM relies on the commercial software ASREML to do so), nor the time spent reading/writing files and inverting the nd by nd matrix. The computing time for GEMMA include all steps. n is the number of individuals, p is the number of SNPs, d is the number of traits, c is the number of covariates ($c = 1$ here), t_1 is the number of iterations used in the EM type algorithm and t_2 is the number of iterations used in the NR type algorithm. The MTMM software implementation handles only two traits. The pn^2 step in GEMMA could be replaced with a pnr step if the relatedness matrix is of rank r.

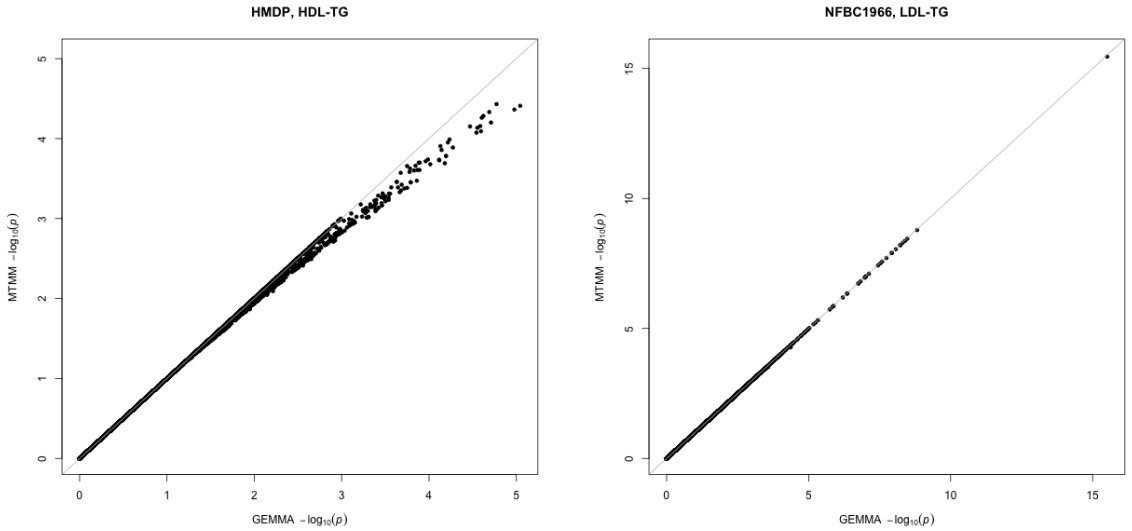
Method	Time Complexity	Computation Time					
		HMDP ($n = 656$, $p = 108,562$)			NFBC1966 ($n = 5255$, $p = 319,111$)		
	# Traits (d)	2	3	4	2	3	4
GEMMA	$O(n^3 + n^2d + n^2c + p(n^2 + t_1nc^2d^2 + t_2nc^2d^6))$	6.2 min	13.7 min	28.5 min	4.4 h	4.8 h	5.8 h
MTMM	$O(t_1n^3(d + c)^3 + t_2n^3d^7 + pn^2d^2)$	16.4 min	–	–	58.0 h	–	–

No other existing software can perform an "exact"[1] GWAS analysis for mvLMM. Further, if existing algorithms were implemented in software, it could easily take them over 14 days to analyze the small HMDP data set, and over 18 years to analyze the large NFBC data set, for only two phenotypes, based on Table 1. Because GEMMA can calculate the exact likelihood ratio test statistics, this allows, for the first time, a genome-wide numerical comparison of p values from the approximate method MTMM with those from an exact method. Results are shown for one pair of traits in Figure 1; results for other pairs of traits are similar (not shown). The difference in p values between the approximate and the exact methods is expected to depend on the genetic marker effect size [16, 47, 19, 20]: when the marker effect size is large, then the difference can be appreciable; when the marker effect size is small, the difference may be negligible. Indeed, in the HMDP data set, the p values from GEMMA can be several fold smaller than that from MTMM (2.8-6.0 fold difference on p values and 12%-16% difference on $\log_{10}(p)$ for six pairs), while in the NFBC1966 data set (where effect sizes are smaller), the p value difference from the two methods are almost negligible (1.2-2.3 fold difference on p values and 1%-2% difference on $\log_{10}(p)$ for six pairs).

Analyzing more than two phenotypes with mvLMM

In contrast to the current implementation of MTMM, GEMMA can analyze more than two phenotypes jointly. We illustrate this feature with the NFBC1966 data set to show how the multivariate analysis can complement univariate analyses. (We omit results for the HMDP data set because neither univariate nor multivariate analyses identified strongly associated SNPs at a significance

[1]Exact is in quotation marks here to acknowledge the inevitable small numerical error in the optimization steps in GEMMA.



(a) GEMMA vs MTMM (HMDP) (b) GEMMA vs MTMM (NFBC1966)

Figure 1: Comparison of -$\log_{10} p$ values obtained from GEMMA with those from MTMM, for (a) paired traits HDL-TG in the HMDP data set and (b) paired traits LDL-TG in the NFBC1966 data set.

level of 0.05 after Bonferroni correction; this does not contradict the results from the original study [48] because here we use a smaller sample size and slightly different phenotype transformations (quantile-transformed instead of log-transformed).)

In total, 53 SNPs from 17 genetic regions pass a significance level of 0.05 after Bonferroni correction in either the four trait multivariate analysis or any of the four univariate analyses (Table 3 and Figure 2). In some cases the multivariate analysis provided the most significant p values, and in others a univariate analysis was more significant. For instance, the multivariate analysis failed to identify four regions that were identified by the univariate analyses. These are the *PCSK9* region (which in the univariate analyses associates most strongly with LDL), *LEPR* region (which in the univariate analyses associates most strongly with CRP), *CR1L* region (which in the univariate analyses associates most strongly with LDL), and *LDLR* region (which in the univariate analyses associates most strongly with LDL). The first three regions were not identified by a previous pairwise analyses of phenotypes in these data either [3]. (In addition, the *PCSK9* region was not reported in the original analysis in these data [49] but was identified in a large-scale meta-analyses of multiple GWAS as associated with LDL [50], while the *CR1L* region was reported in the original analysis in these data [49] but was not reported in the large-scale meta-analysis of lipids [50].) On the other hand, the multivariate analysis also identified two regions, near *PPP1R3B* and *FADS*, that were not identified by univariate analyses. Both these regions were previously identified as associated with lipid traits and/or CRP in two large-scale meta-analyses of multiple GWAS [50, 51], and by

pairwise analyses of phenotypes in these data [3].

In addition, the multivariate analysis also identified several SNPs in known lipid/CRP-associated regions that were not identified by univariate analyses. These include four SNPs (rs6728178, rs6754295, rs676210 and rs673548) in the *APOB* region, three SNPs (rs16940213, rs473224 and rs261336) in the *LIPC* region, and one SNP (rs2075650) in the *APO cluster* region. All these SNPs, with the only exception of rs16940213 (which is in moderate linkage disequilibrium with rs166358 in the same region; based on the CEU panel in the 1000 Genome Project [52], r^2 between rs16940213 and rs166358 is 0.48), were identified by pairwise analyses of phenotypes in these data [3]. However, the p values from the four trait multivariate analysis can be much lower than any of the p values from the pairwise analyses (Table S1). The most notable examples are two SNPs (rs983309 and rs2126259) in the *PPP1R3B* region and one SNP (rs2075650) in the *APO cluster* region, all of which have a four trait mvLMM p value an order of magnitude smaller than any of the two trait mvLMM p values. (All these three SNPs show moderate univariate associations with three phenotypes.)

Discussion

Here, we present novel algorithms, and a software implementation in the package GEMMA, for multivariate analysis using mvLMMs in genetic association studies. In two example data sets, this method is not only orders of magnitude faster than existing methods in both parameter estimation in a single mvLMM and GWAS analysis, but is the only method currently practical for calculating exact likelihood ratio test statistics, and corresponding p values, genome wide, for joint analysis of more than one phenotype.

One of the main issues concerning multivariate analysis in GWASs, including using mvLMM, is the interpretation of the analysis results. Specifically, rejecting the multivariate null hypothesis, that a SNP is unassociated with *any* phenotype, does not tell us *which* phenotype(s) are associated. And although one can certainly obtain useful information from the univariate tests, the results will not always be clear-cut (e.g. what significance threshold is appropriate to declare a univariate test "significant", in light of a significant multivariate test?). For linear models [31] provides discussion and methods to address these issues; extension of these ideas to mvLMMs would be an interesting avenue for further work.

Although our implementation of the EM algorithm scales only quadratically with the number of phenotypes, d, and so could be applied to reasonably large d, in practice we caution that there could remain both computational and statistical barriers to applying these methods to even quite modest values of d (e.g. $d \approx 10$). Computationally, the number of iterations required to converge for larger d will inevitably increase, and ultimately this issue could be the main barrier to effectively maximizing the likelihood for large d. Statistically, the number of parameters in the mvLMM is also quadratic in the number of phenotypes (the number of parameters in the two variance components

is $d(d+1)$). Therefore, with a moderate sample size, additional assumptions on the structure of the variance components may be necessary to obtain reliable estimates. For example, one could penalize the departure of the estimated variance components to some prior values [53], or could assume the variance components are of low rank [54] or sparse [55]. The effectiveness of such strategies, paired with the mvLMM, in GWASs is another topic for potential further research.

The most computationally expensive part of our method, as in the univariate case, is the eigen-decomposition step. (Although the genotype transformation step can be expensive with a large number of markers, it can be easily parallelized in a computing cluster.) The initial eigen-decomposition step not only requires a large amount of physical memory, but also becomes computationally intractable in practice, for large n (e.g. $> 50,000$ [56]). Low rank approximations to the relatedness matrix [18, 46, 19] can alleviate both computation and memory requirements, and using these kinds of methods mvLMMs could be applied to very large GWASs.

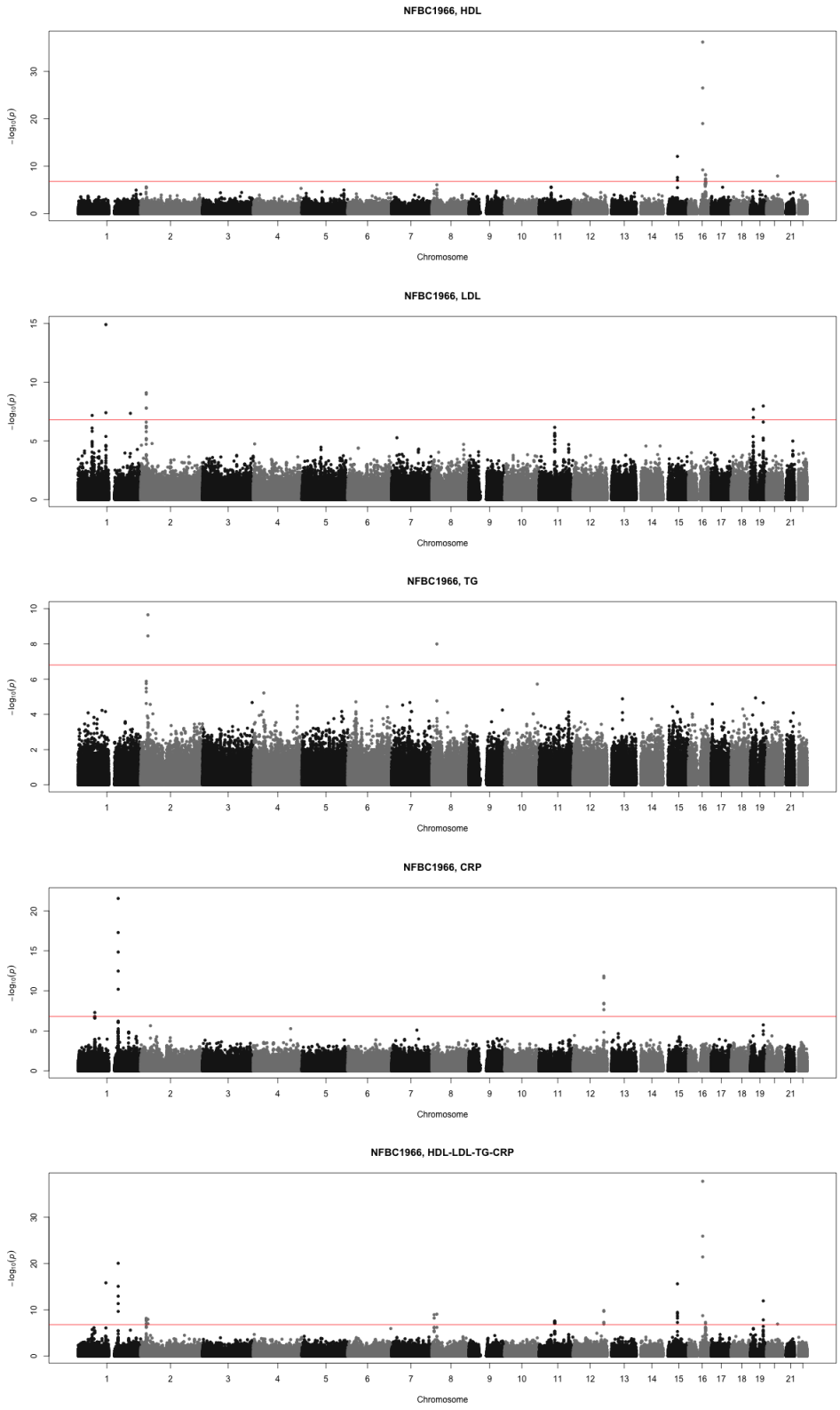


Figure 2: Manhattan plots for the univariate LMM analyses and the joint mvLMM analysis for the four phenotypes in the NFBC1966 data set. The red horizontal line indicates the significance level of 0.05 after Bonferroni correction.

Table 3: List of SNPs that exceed significance of 0.05 after Bonferroni correction in the four trait mvLMM analysis or any of the univariate LMM analyses in the NFBC1966 data set. λ_{GC} is the genomic control inflation factor. p values below the threshold are in bold font. SNPs passing the significance level in the multivariate analysis but not in any of the univariate analyses are highlighted in 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 ($\lambda_{GC} = 0.993$)
PCSK9	chromosome 1					
rs207150	55579053	1.87×10^{-6}	5.0×10^{-1}	$\mathbf{6.69 \times 10^{-8}}$	6.58×10^{-1}	9.61×10^{-1}
LEPR	chromosome 1					
rs1892534	65878532	3.33×10^{-6}	2.78×10^{-1}	6.98×10^{-1}	7.63×10^{-1}	$\mathbf{5.09 \times 10^{-8}}$
CELSR2	chromosome 1					
rs611917	109616775	8.48×10^{-7}	2.31×10^{-1}	$\mathbf{3.96 \times 10^{-8}}$	5.91×10^{-1}	7.00×10^{-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}	$\mathbf{6.28 \times 10^{-11}}$
CR1L	chromosome 1					
rs4844614	205941798	2.55×10^{-6}	1.75×10^{-1}	$\mathbf{4.48 \times 10^{-8}}$	6.21×10^{-3}	7.69×10^{-1}
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	$\mathbf{3.66 \times 10^{-10}}$	$\mathbf{8.57 \times 10^{-8}}$	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					
rs6499137	66229305	7.68×10^{-7}	$\mathbf{8.82 \times 10^{-8}}$	7.66×10^{-1}	9.48×10^{-1}	3.94×10^{-1}
rs2271293	66459571	5.32×10^{-7}	$\mathbf{5.98 \times 10^{-8}}$	4.07×10^{-1}	8.39×10^{-1}	8.67×10^{-1}
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}
rs8058517	66937361	9.98×10^{-7}	$\mathbf{4.51 \times 10^{-8}}$	3.63×10^{-1}	3.20×10^{-1}	9.63×10^{-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}
rs2228671	11071912	1.51×10^{-6}	9.64×10^{-1}	$\mathbf{1.00 \times 10^{-7}}$	9.14×10^{-1}	7.63×10^{-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}

Methods

Genotype and Phenotype Data

We analyzed two data sets: the Hybrid Mouse Diversity Panel (HMDP) [48] and the Northern Finland Birth Cohort 1966 (NFBC1966) Study [49].

The HMDP data includes 100 inbred strains with four phenotypes (high-density lipoprotein, HDL; total cholesterol, TC; triglycerides, TG; unesterified cholesterol, UC) and four million high quality fully imputed SNPs (SNPs are downloaded from `http://mouse.cs.ucla.edu/mousehapmap/full.html`). We excluded mice with missing phenotypes for any of these four phenotypes. We excluded non-polymorphic SNPs, and SNPs with a minor allele frequency less than 5%. For SNPs that have identical genotypes, we tried to retain only one of them (by using "–indep-pairwise 100 5 0.999999" option in PLINK [57]). This left us with 98 strains, 656 individuals and 108,562 SNPs. We quantile transformed each phenotype to a standard normal distribution to guard against model mis-specification. We used the product of centered genotype matrix as an estimate of relatedness [58, 59, 22]. Note that the sample size used here is smaller than the original study [48], and the phenotypes are quantile-transformed instead of log transformed for robustness.

The NFBC1966 data contains 5402 individuals with multiple metabolic traits measured and 364,590 SNPs typed. We selected four phenotypes (high-density lipoprotein, HDL; low-density lipoprotein, LDL; triglycerides, TG; C-reactive protein, CRP) among them, following previous studies [3]. We selected individuals and SNPs following previous studies [49, 16] with the software PLINK [57]. Specifically, we excluded individuals with missing phenotypes for any of these four phenotypes or having discrepancies between reported sex and sex determined from the X chromosome. We excluded SNPs with a minor allele frequency less than 1%, having missing values in more than 1% of the individuals, or with a Hardy-Weinberg equilibrium p value below 0.0001. This left us with 5,255 individuals and 319,111 SNPs. For each phenotype, we quantile transformed the phenotypic values to a standard normal distribution, regressed out sex, oral contraceptives and pregnancy status effects [49], and quantile transformed the residuals to a standard normal distribution again. We replaced the missing genotypes for a given SNP with its mean genotype value. We used the product of centered and scaled genotype matrix as an estimate of relatedness [58, 59, 16].

In both data sets, we quantile transformed each single phenotype to a standard normal distribution to guard against model misspecification. Although this strategy does not guarantee that the transformed phenotypes follow a multivariate normal distribution jointly, it often works well in practice when the number of phenotypes is small (see, e.g. [31]). For both data sets, we used a standard mvLMM with an intercept term (without any other covariates), and test each SNP in turn. Because the software MTMM relies on the commercial software ASREML to estimate the variance components in the null model, we modified the MTMM source code so that it can read in the estimated variance components from GEMMA.

Acknowledgment

This research is supported in part by NIH grant HL092206 (PI Y Gilad) and NIH grant HG02585 to MS. We thank A. J. Lusis for making the mouse genotype and phenotype data available. We thank the NFBC1966 Study Investigators for making the NFBC1966 data available. The NFBC1966 Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, UCLA, University of Oulu, and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 Study and does not necessarily reflect the opinions or views of the NFBC1966 Study Investigators, Broad Institute, UCLA, University of Oulu, National Institute for Health and Welfare in Finland and the NHLBI.

Supplementary Information

Multivariate Linear Mixed Model

We consider the multivariate linear mixed model [1],

$$\tilde{\mathbf{Y}} = \mathbf{A}\tilde{\mathbf{W}} + \boldsymbol{\beta}\tilde{\mathbf{x}}^T + \tilde{\mathbf{G}} + \tilde{\mathbf{E}}; \quad \tilde{\mathbf{G}} \sim \mathrm{MN}_{d\times n}(\mathbf{0}, \mathbf{V}_g, \mathbf{K}), \quad \tilde{\mathbf{E}} \sim \mathrm{MN}_{d\times n}(\mathbf{0}, \mathbf{V}_e, \mathbf{I}_{n\times n}), \tag{2}$$

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 [18, 19, 20], 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 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

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{G} + \mathbf{E}; \quad \mathbf{G} \sim \mathrm{MN}(0, \mathbf{V}_g, \mathbf{D}_k), \quad \mathbf{E} \sim \mathrm{MN}(0, \mathbf{V}_e, \mathbf{I}_{n\times n}), \tag{3}$$

which is equivalent to

$$\mathbf{y} = \mathbf{X}^T \otimes \mathbf{I}_{d\times d}\mathbf{b} + \mathbf{g} + \mathbf{e}; \quad \mathbf{g} \sim \mathrm{MVN}(0, \mathbf{D}_k \otimes \mathbf{V}_g), \quad \mathbf{e} \sim \mathrm{MVN}(0, \mathbf{I}_{n\times n} \otimes \mathbf{V}_e), \tag{4}$$

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

$$\mathbf{y}_l = \mathbf{B}\mathbf{x}_l + \mathbf{g}_l + \mathbf{e}_l; \quad \mathbf{g}_l \sim \mathrm{MVN}(0, \delta_l\mathbf{V}_g), \quad \mathbf{e}_l \sim \mathrm{MVN}(0, \mathbf{V}_e), \tag{5}$$

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 $\mathbf{E}$, $\forall l = 1, \cdots, n$.

PX-EM Algorithms

Here, we describe an expectation conditional maximization (ECM) algorithm [60] for finding maximum likelihood estimates (MLE) in mvLMM, an expectation maximization (EM) algorithm [35] for finding restricted maximum likelihood estimates (REMLE) in mvLMM, parameter expansion (PX) versions [36, 37] of the two, and their efficient computations.

An ECM Algorithm for MLE

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

$$\log l(\mathbf{Y}, \mathbf{G} | \mathbf{B}, \mathbf{V}_g, \mathbf{V}_e) = \sum_{l=1}^{n} \{-d \log(2\pi) - \frac{1}{2} \log |\mathbf{V}_e| - \frac{1}{2} \log |\delta_l \mathbf{V}_g| - \frac{1}{2} \mathbf{e}_l^T \mathbf{V}_e^{-1} \mathbf{e}_l - \frac{1}{2} \mathbf{g}_l^T (\delta_l \mathbf{V}_g)^{-1} \mathbf{g}_l\}. \tag{6}$$

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

$$\mathbf{g}_l | \mathbf{y}_l, \mathbf{B}^{(t)}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{g}}_l^{(t)}, \hat{\mathbf{\Sigma}}_l^{(t)}), \tag{7}$$

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

$$\begin{aligned}
&E_{\mathbf{G}|\mathbf{Y}, \mathbf{B}^{(t)}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)}} [\log l(\mathbf{Y}, \mathbf{G} | \mathbf{B}, \mathbf{V}_g, \mathbf{V}_e)] \\
&= \sum_{l=1}^{n} \{-d \log(2\pi) - \frac{1}{2} \log |\delta_l \mathbf{V}_g| - \frac{1}{2} \log |\mathbf{V}_e| - \frac{1}{2} (\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)^T \mathbf{V}_e^{-1} (\mathbf{y}_l - \mathbf{B}\mathbf{x}_l) \\
&- \frac{1}{2} (\hat{\mathbf{g}}_l^{(t)})^T ((\delta_l \mathbf{V}_g)^{-1} + \mathbf{V}_e^{-1})^{-1} \hat{\mathbf{g}}_l^{(t)} - \frac{1}{2} \text{trace}(((\delta_l \mathbf{V}_g)^{-1} + \mathbf{V}_e^{-1})^{-1} \hat{\mathbf{\Sigma}}_l^{(t)}) + \hat{\mathbf{g}}_l^{(t)} \mathbf{V}_e^{-1} (\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)\}.
\end{aligned} \tag{8}$$

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

$$\mathbf{B}^{(t+1)} = (\mathbf{Y} - \hat{\mathbf{G}}^{(t)})\mathbf{X}(\mathbf{X}\mathbf{X}^T)^{-1}, \tag{9}$$

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

$$\mathbf{V}_e^{(t+1)} = \frac{1}{n}\sum_{l=1}^{n} (\hat{\mathbf{e}}_l^{(t)}(\hat{\mathbf{e}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}_l^{(t)}), \tag{11}$$

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 [19].

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 6, 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

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{g} \end{pmatrix} | \mathbf{Y}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}\left(\begin{pmatrix} \hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)} \\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)} \end{pmatrix} \begin{pmatrix} (\mathbf{X} \otimes (\mathbf{V}_e^{(t)})^{-1})\mathbf{y} \\ (\mathbf{I}_{n\times n} \otimes (\mathbf{V}_e^{(t)})^{-1})\mathbf{y} \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)} \\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)} \end{pmatrix} \right), \tag{12}$$

where

$$\begin{pmatrix} \hat{\mathbf{\Sigma}}_{\mathbf{bb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{bg}}^{(t)} \\ \hat{\mathbf{\Sigma}}_{\mathbf{gb}}^{(t)} & \hat{\mathbf{\Sigma}}_{\mathbf{gg}}^{(t)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\mathbf{X}^T \otimes (\mathbf{V}_e^{(t)})^{-1} & \mathbf{X} \otimes (\mathbf{V}_e^{(t)})^{-1} \\ \mathbf{X}^T \otimes (\mathbf{V}_e^{(t)})^{-1} & \mathbf{D}_k^{-1} \otimes (\mathbf{V}_g^{(t)})^{-1} + \mathbf{I}_{n\times n} \otimes (\mathbf{V}_e^{(t)})^{-1} \end{pmatrix}^{-1}. \tag{13}$$

Therefore,

$$\mathbf{g}_l | \mathbf{Y}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{g}}_l^{(t)}, \hat{\mathbf{\Sigma}}_{l,\mathbf{gg}}^{(t)}), \tag{14}$$

$$\mathbf{e}_l = \mathbf{y}_l - \mathbf{B}\mathbf{x}_l - \mathbf{g}_l | \mathbf{Y}, \mathbf{V}_g^{(t)}, \mathbf{V}_e^{(t)} \sim \text{MVN}(\hat{\mathbf{e}}_l^{(t)}, \hat{\mathbf{\Sigma}}_{l,\mathbf{ee}}^{(t)}), \tag{15}$$

where

$$\hat{\mathbf{b}}^{(t)} = \hat{\mathbf{\Sigma}}^{(t)}_{\mathbf{bb}} \sum_{l=1}^{n} \mathbf{x}_k \otimes (\mathbf{V}_l^{(t)})^{-1} \mathbf{y}_l, \tag{16}$$

$$\hat{\mathbf{g}}_l^{(t)} = \sum_{l=1}^{n} \delta_l \mathbf{V}_g^{(t)} (\mathbf{V}_l^{(t)})^{-1} (\mathbf{y}_l - \hat{\mathbf{B}}^{(t)} \mathbf{x}_l), \tag{17}$$

$$\hat{\mathbf{e}}_l^{(t)} = \mathbf{y}_l - \hat{\mathbf{B}}^{(t)} \mathbf{x}_l - \hat{\mathbf{g}}_l^{(t)}, \tag{18}$$

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

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

$$\hat{\mathbf{\Sigma}}^{(t)}_{l,\mathbf{ee}} = \delta_l \mathbf{V}_g^{(t)} (\mathbf{V}_l^{(t)})^{-1} + (\mathbf{x}_l^T \otimes (\mathbf{V}_l^{(t)})^{-1}) (\sum_{l=1}^{n} \mathbf{x}_l \mathbf{x}_l^T \otimes \delta_l \mathbf{V}_g^{(t)} (\mathbf{V}_l^{(t)})^{-1})^{-1} (\mathbf{x}_l \otimes (\mathbf{V}_l^{(t)})^{-1}), \tag{21}$$

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

$$\begin{aligned}
&E_{\mathbf{B},\mathbf{G}|\mathbf{Y},\mathbf{V}_g^{(t)},\mathbf{V}_e^{(t)}}[\log l(\mathbf{Y}, \mathbf{G}, \mathbf{B} | \mathbf{V}_g, \mathbf{V}_e)] \\
&= \sum_{l=1}^{n} \{ -d \log(2\pi) - \frac{1}{2} \log |\mathbf{V}_e| - \frac{1}{2} \log |\delta_l \mathbf{V}_g| - \frac{1}{2} (\hat{\mathbf{e}}_l^{(t)})^T \mathbf{V}_e^{-1} \hat{\mathbf{e}}_l^{(t)} - \frac{1}{2} \text{trace}(\mathbf{V}_e^{-1} \hat{\mathbf{\Sigma}}^{(t)}_{l,\mathbf{ee}}) \\
&- \frac{1}{2} (\hat{\mathbf{g}}_l^{(t)})^T (\delta_l \mathbf{V}_g)^{-1} \hat{\mathbf{g}}_l^{(t)} - \frac{1}{2} \text{trace}((\delta_l \mathbf{V}_g)^{-1} \hat{\mathbf{\Sigma}}^{(t)}_{l,\mathbf{gg}}) \}.
\end{aligned} \tag{22}$$

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

$$\mathbf{V}_g^{(t+1)} = \frac{1}{n} \sum_{l=1}^{n} \delta_l^{-1} (\hat{\mathbf{g}}_l^{(t)} (\hat{\mathbf{g}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}^{(t)}_{l,\mathbf{gg}}), \tag{23}$$

$$\mathbf{V}_e^{(t+1)} = \frac{1}{n} \sum_{l=1}^{n} (\hat{\mathbf{e}}_l^{(t)} (\hat{\mathbf{e}}_l^{(t)})^T + \hat{\mathbf{\Sigma}}^{(t)}_{l,\mathbf{ee}}). \tag{24}$$

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

$$\mathbf{y}_l = \mathbf{B}\mathbf{x}_l + \mathbf{V}_a \mathbf{g}_l^* + \mathbf{e}_l \quad \mathbf{g}_l^* \sim \text{MVN}(0, \delta_l \mathbf{V}_g^*) \quad \mathbf{e}_l \sim \text{MVN}(0, \mathbf{V}_e), \tag{25}$$

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

$$\mathbf{V}_a^{(t+1)} = (\frac{1}{n}\sum_{l=1}^{n} E((\mathbf{y}_l - \mathbf{B}\mathbf{x}_l)(\mathbf{g}_l^*)^T))(\frac{1}{n}\sum_{l=1}^{n} E(\mathbf{g}_l^*(\mathbf{g}_l^*)^T))^{-1}. \tag{26}$$

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).

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. [10, 45] 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 $O(nc^2d^2)$.

Newton-Raphson's Algorithms

Here, we describe Newton-Raphson's algorithms for MLE and REMLE estimation in mvLMM. Although an average-information algorithm [40] 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.

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:

$$l(\mathbf{V}_g, \mathbf{V}_e) = -\frac{nd}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{H}| - \frac{1}{2}\mathbf{y}^T\mathbf{P}\mathbf{y}, \tag{27}$$

$$\begin{aligned} l_r(\mathbf{V}_g, \mathbf{V}_e) = &-\frac{(n-c-1)d}{2}\log(2\pi) + \frac{d}{2}\log|\mathbf{X}\mathbf{X}^T| - \frac{1}{2}\log|\mathbf{H}| \\ &- \frac{1}{2}\log|(\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d})| - \frac{1}{2}\mathbf{y}^T\mathbf{P}\mathbf{y}, \end{aligned} \tag{28}$$

where

$$\mathbf{H} = \mathbf{D}_k\otimes\mathbf{V}_g + \mathbf{I}_{n\times n}\otimes\mathbf{V}_e = \mathrm{diag}(\mathbf{V}_l), \tag{29}$$

$$\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d})((\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{X}^T\otimes\mathbf{I}_{d\times d}))^{-1}(\mathbf{X}\otimes\mathbf{I}_{d\times d})\mathbf{H}^{-1}. \tag{30}$$

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

$$\frac{\partial\mathrm{vec}(\mathbf{D}_k\otimes\mathbf{V}_g)}{\partial v_{g,ij}} = \mathrm{vec}(\mathbf{D}_k\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\frac{1}{1+1_{i=j}}, \tag{31}$$

$$\frac{\partial\mathrm{vec}(\mathbf{I}_{n\times n}\otimes\mathbf{V}_e)}{\partial v_{e,ij}} = \mathrm{vec}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\frac{1}{1+1_{i=j}}, \tag{32}$$

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 [19], we obtain the first order partial derivatives for the log-likelihood and the log-restricted likelihood functions

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

$$\frac{\partial l}{\partial v_{e,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{34}$$

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

$$\frac{\partial l_r}{\partial v_{e,ij}} = \frac{1}{1+1_{i=j}}\{-\frac{1}{2}\mathrm{trace}(\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))) + \frac{1}{2}\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}\mathbf{y}\}, \tag{36}$$

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

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

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

$$\frac{\partial l^2}{\partial v_{e,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{H}^{-1}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{H}^{-1}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$-\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}, \quad (39)$$

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

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

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

$$\frac{\partial l_r^2}{\partial v_{e,ij}\partial v_{e,i'j'}} = \frac{1}{(1+1_{i=j})(1+1_{i'=j'})}\{\frac{1}{2}\text{trace}(\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'})))$$
$$-\mathbf{y}^T\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{ij}+\mathbf{I}_{ji}))\mathbf{P}(\mathbf{I}_{n\times n}\otimes(\mathbf{I}_{i'j'}+\mathbf{I}_{j'i'}))\mathbf{P}\mathbf{y}\}. \quad (42)$$

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 low rank 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

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

$$\mathbf{q}_{ij}^{g} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}\mathbf{y}, \tag{44}$$

$$q_{ij}^{g} = \mathbf{y}^T\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}\mathbf{y}, \tag{45}$$

$$\mathbf{Q}_{ij,i'j'}^{gg} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}(\mathbf{X}^T \otimes \mathbf{I}_{d\times d}), \tag{46}$$

$$\mathbf{q}_{ij,i'j'}^{gg} = (\mathbf{X} \otimes \mathbf{I}_{d\times d})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}\mathbf{y}, \tag{47}$$

$$q_{ij,i'j'}^{gg} = \mathbf{y}^T\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{H}^{-1}\mathbf{y}. \tag{48}$$

For the trace terms, we have

$$\mathrm{trace}(\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) = \mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}), \tag{49}$$

$$\begin{aligned}\mathrm{trace}(\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) &= \mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) \\ &\quad - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij,i'j'}^{gg}) - \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{i'j',ij}^{gg}) + \mathrm{trace}(\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}).\end{aligned} \tag{50}$$

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

$$\mathbf{y}^T\mathbf{P}\mathbf{y} = q - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}, \tag{51}$$

$$\mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}\mathbf{y} = q_{ij}^{g} - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}_{ij}^{g} - (\mathbf{q}_{ij}^{g})^T\mathbf{Q}^{-1}\mathbf{q} + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{q}, \tag{52}$$

$$\begin{aligned}\mathbf{y}^T\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{P}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})\mathbf{P}\mathbf{y} &= q_{ij,i'j'}^{gg} - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{q}_{ij,i'j'}^{gg} - (\mathbf{q}_{j'i',ji}^{gg})^T\mathbf{Q}^{-1}\mathbf{q} - (\mathbf{q}_{ji}^{g})^T\mathbf{Q}^{-1}\mathbf{q}_{i'j'}^{g} \\ &\quad + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{q}_{i'j'}^{g} + (\mathbf{q}_{ji}^{g})^T\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}\mathbf{Q}^{-1}\mathbf{q} + \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij,i'j'}^{gg}\mathbf{q} \\ &\quad - \mathbf{q}^T\mathbf{Q}^{-1}\mathbf{Q}_{ij}^{g}\mathbf{Q}^{-1}\mathbf{Q}_{i'j'}^{g}\mathbf{Q}^{-1}\mathbf{q}.\end{aligned} \tag{53}$$

Therefore, it suffices to efficiently evaluate

$$|\mathbf{H}| = \sum_{l=1}^{n} |\mathbf{V}_l|, \tag{54}$$

$$\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})) = \sum_{l=1}^{n} \delta_l(\mathbf{I}_j^T\mathbf{V}_l^{-1}\mathbf{I}_i), \tag{55}$$

$$\mathrm{trace}(\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{ij})\mathbf{H}^{-1}(\mathbf{D}_k \otimes \mathbf{I}_{i'j'})) = \sum_{l=1}^{n} \delta_l^2(\mathbf{I}_{j'}^T\mathbf{V}_l^{-1}\mathbf{I}_i)(\mathbf{I}_j^T\mathbf{V}_l^{-1}\mathbf{I}_{i'}), \tag{56}$$

and

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

$$\mathbf{q} = \sum_{l=1}^{n} \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{y}_l), \tag{58}$$

$$q = \sum_{l=1}^{n} \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{y}_l, \tag{59}$$

and

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

$$\mathbf{q}_{ij}^{g} = \sum_{l=1}^{n} \delta_l \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{y}_l), \tag{61}$$

$$q_{ij}^{g} = \sum_{l=1}^{n} \delta_l \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{y}_l, \tag{62}$$

$$\mathbf{Q}_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 (\mathbf{x}_l \mathbf{x}_l^T) \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1}), \tag{63}$$

$$\mathbf{q}_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 \mathbf{x}_l \otimes (\mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1} \mathbf{y}_l), \tag{64}$$

$$q_{ij,i'j'}^{gg} = \sum_{l=1}^{n} \delta_l^2 \mathbf{y}_l^T \mathbf{V}_l^{-1} \mathbf{I}_{ij} \mathbf{V}_l^{-1} \mathbf{I}_{i'j'} \mathbf{V}_l^{-1} \mathbf{y}_l. \tag{65}$$

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 $O(nc^2d^6)$.

Test Statistics and p Values

The effect size estimate $\hat{\boldsymbol{\beta}}$ and its variance $V(\hat{\boldsymbol{\beta}})$ can be obtained in the following way. With the variance component estimates $\hat{\mathbf{V}}_g$ and $\hat{\mathbf{V}}_e$ ($\hat{\mathbf{V}}_l = \delta_l \hat{\mathbf{V}}_g + \hat{\mathbf{V}}_e$), we have

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

and its precision

$$V(\hat{\boldsymbol{\beta}})^{-1} = \sum_{l=1}^{n} x_l^2 \hat{\mathbf{V}}_l^{-1} - (\sum_{l=1}^{n} (x_l \mathbf{w}_l) \otimes \hat{\mathbf{V}}_l^{-1})(\sum_{l=1}^{n} (\mathbf{w}_l \mathbf{w}_l^T) \otimes \hat{\mathbf{V}}_l^{-1})^{-1}(\sum_{l=1}^{n} (x_l \mathbf{w}_l) \otimes \hat{\mathbf{V}}_l^{-1}). \quad (67)$$

where 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 $\mathbf{W}$.

Under the null hypothesis, both the Wald test statistic $\hat{\boldsymbol{\beta}}^T V(\hat{\boldsymbol{\beta}})^{-1} \hat{\boldsymbol{\beta}}$ and the likelihood ratio test statistics follow a $\chi^2(d)$ distribution asymptotically (for small samples, the variance of $\hat{\boldsymbol{\beta}}$ may require some approximations; see e.g. [61, 62]). These asymptotic distributions are used to obtain p values.

Supplementary Results

Table S1 lists SNPs that exceed significance of 0.05 after Bonferroni correction in either the four trait mvLMM analysis or any of the two trait mvLMM analyses, in the NFBC1966 data set.

Table S1: List of SNPs that exceed significance of 0.05 after Bonferroni correction in either the four trait mvLMM analysis or any of the two trait mvLMM analyses in the NFBC1966 data set. λ_{GC} is the genomic control inflation factor. p values below the threshold are in bold font. SNPs having smaller p value in the four trait analysis are highlighted in 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							
rs611917	109616775	8.48×10^{-7}	2.58×10^{-7}	4.78×10^{-1}	4.19×10^{-1}	$\mathbf{1.04 \times 10^{-7}}$	2.03×10^{-7}	7.91×10^{-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							
rs3923037	21011755	3.75×10^{-7}	$\mathbf{1.47 \times 10^{-7}}$	3.70×10^{-5}	7.22×10^{-3}	$\mathbf{9.71 \times 10^{-8}}$	3.96×10^{-6}	1.19×10^{-4}
rs6728178	21047434	$\mathbf{7.37 \times 10^{-9}}$	$\mathbf{1.48 \times 10^{-9}}$	2.03×10^{-7}	2.63×10^{-5}	$\mathbf{3.02 \times 10^{-8}}$	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}	$\mathbf{3.61 \times 10^{-8}}$	4.01×10^{-6}	3.40×10^{-5}
rs676210	21085029	$\mathbf{2.09 \times 10^{-8}}$	$\mathbf{9.04 \times 10^{-9}}$	$\mathbf{1.11 \times 10^{-7}}$	2.03×10^{-5}	$\mathbf{1.07 \times 10^{-7}}$	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	$\mathbf{1.36 \times 10^{-8}}$	$\mathbf{6.45 \times 10^{-9}}$	$\mathbf{7.78 \times 10^{-8}}$	1.64×10^{-5}	$\mathbf{7.52 \times 10^{-8}}$	4.18×10^{-5}	8.73×10^{-6}
rs1429974	21154275	6.33×10^{-7}	$\mathbf{7.27 \times 10^{-8}}$	1.70×10^{-1}	1.98×10^{-1}	$\mathbf{1.03 \times 10^{-7}}$	$\mathbf{1.01 \times 10^{-7}}$	3.93×10^{-1}
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}
rs754523	21165196	5.68×10^{-7}	$\mathbf{7.08 \times 10^{-8}}$	1.82×10^{-1}	2.01×10^{-1}	$\mathbf{9.53 \times 10^{-8}}$	$\mathbf{8.98 \times 10^{-8}}$	3.94×10^{-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}	$\mathbf{2.23 \times 10^{-8}}$
intergenic region between FRMD1 and DACT2			chromosome 6					
rs2171981	168269089	1.15×10^{-6}	$\mathbf{1.06 \times 10^{-7}}$	3.62×10^{-5}	3.58×10^{-4}	2.16×10^{-3}	1.01×10^{-3}	4.43×10^{-1}
PPP1R3B	chromosome 8							
rs983309	9215142	$\mathbf{6.40 \times 10^{-9}}$	5.04×10^{-7}	1.24×10^{-4}	$\mathbf{7.54 \times 10^{-8}}$	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}	$\mathbf{1.02 \times 10^{-7}}$	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}	$\mathbf{3.01 \times 10^{-8}}$
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}	$\mathbf{1.17 \times 10^{-7}}$	4.96×10^{-1}	$\mathbf{5.24 \times 10^{-8}}$	$\mathbf{8.3 \times 10^{-9}}$
LIPC	chromosome 15							
rs166358	56468097	$\mathbf{3.66 \times 10^{-10}}$	2.34×10^{-7}	$\mathbf{8.28 \times 10^{-10}}$	$\mathbf{1.39 \times 10^{-7}}$	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	$\mathbf{7.47 \times 10^{-10}}$	$\mathbf{7.02 \times 10^{-8}}$	$\mathbf{2.87 \times 10^{-11}}$	$\mathbf{1.03 \times 10^{-7}}$	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							
rs6499137	66229305	7.68×10^{-7}	5.69×10^{-7}	$\mathbf{3.48 \times 10^{-8}}$	8.54×10^{-7}	9.39×10^{-1}	6.84×10^{-1}	6.52×10^{-1}
rs2271293	66459571	5.32×10^{-7}	$\mathbf{1.24 \times 10^{-7}}$	$\mathbf{4.32 \times 10^{-8}}$	3.72×10^{-7}	6.29×10^{-1}	6.80×10^{-1}	9.75×10^{-1}
rs255049	66570972	$\mathbf{5.34 \times 10^{-8}}$	$\mathbf{4.07 \times 10^{-9}}$	$\mathbf{3.25 \times 10^{-8}}$	$\mathbf{2.85 \times 10^{-8}}$	4.75×10^{-2}	3.29×10^{-1}	3.32×10^{-1}
rs255052	66582496	$\mathbf{9.72 \times 10^{-8}}$	$\mathbf{5.24 \times 10^{-9}}$	$\mathbf{2.10 \times 10^{-8}}$	$\mathbf{4.58 \times 10^{-8}}$	1.06×10^{-1}	3.81×10^{-1}	5.12×10^{-1}
rs8058517	66937361	9.98×10^{-7}	$\mathbf{8.18 \times 10^{-8}}$	$\mathbf{1.40 \times 10^{-7}}$	2.65×10^{-7}	2.62×10^{-1}	6.46×10^{-1}	6.30×10^{-1}
LDLR	chromosome 19							
rs11668477	11056030	1.09×10^{-6}	$\mathbf{7.87 \times 10^{-8}}$	3.47×10^{-2}	1.52×10^{-1}	$\mathbf{1.48 \times 10^{-7}}$	$\mathbf{1.27 \times 10^{-7}}$	4.80×10^{-2}
rs2228671	11071912	1.53×10^{-6}	5.20×10^{-7}	9.95×10^{-1}	9.50×10^{-1}	$\mathbf{9.81 \times 10^{-8}}$	4.83×10^{-7}	9.53×10^{-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}	$\mathbf{2.88 \times 10^{-10}}$	$\mathbf{2.27 \times 10^{-11}}$
rs405509	50100676	3.86×10^{-7}	4.56×10^{-3}	4.60×10^{-1}	5.00×10^{-5}	4.80×10^{-4}	$\mathbf{1.24 \times 10^{-7}}$	1.40×10^{-4}
HNF4A	chromosome 20							
rs1800961	42475778	$\mathbf{1.21 \times 10^{-7}}$	$\mathbf{8.41 \times 10^{-8}}$	$\mathbf{7.61 \times 10^{-8}}$	$\mathbf{1.05 \times 10^{-8}}$	1.81×10^{-2}	5.05×10^{-1}	6.03×10^{-3}

References

[1] Charles R. Henderson. *Applications of linear models in animal breeding.* University of Guelph, Guelph, 1984.

[2] Alkes L. Price, Agnar Helgason, Gudmar Thorleifsson, Steven A. McCarroll, Augustine Kong, and Kari Stefansson. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLOS Genetics*, 7:e1001317, 2011.

[3] Arthur Korte, Bjarni J Vilhjlmsson, Vincent Segura, Alexander Platt, Quan Long, and Magnus Nordborg. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics*, 44:1066–1071, 2012.

[4] Sang Hong Lee, Jian Yang, Michael E. Goddard, Peter M. Visscher, and Naomi R. Wray. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, 28:2540–2542, 2012.

[5] Shashaank Vattikuti, Juen Guo, and Carson C. Chow. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLOS Genetics*, 8:e1002637, 2012.

[6] M. Trzaskowski, Jian Yang, Peter M. Visscher, and R Plomin. DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12. *Molecular Psychiatry*, 0:advance online, 2013.

[7] Christophe I. Amos. Robust variance-components approach for assessing genetic linkage in pedigrees. *American Journal of Human Genetics*, 54:535–543, 1994.

[8] Loeske E B Kruuk. Estimating genetic parameters in natural populations using the animal model. *Philosophical Transactions of the Royal Society B*, 359:873–890, 2004.

[9] Karin Meyer, David J. Johnston, and Hans-Ulrich Graser. Estimates of the complete genetic covariance matrix for traits in multi-trait genetic evaluation of Australian Hereford cattle. *Australian Journal of Agricultural Research*, 55:195–210, 2004.

[10] Karin Meyer. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genetics Selection Evolution*, 23:67–83, 1991.

[11] Mark Abney, Carole Ober, and Mary Sara McPeek. Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: Fasting serum-insulin level in the hutterites. *The American Journal of Human Genetics*, 70:920–934, 2002.

[12] Jianming Yu, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, Brandon S. Gaut, Dahlia M. Nielsen, James B. Holland, Stephen Kresovich, and Edward S. Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38:203–208, 2006.

[13] Wei-Min Chen and Gonalo R. Abecasis. Family-based association tests for genomewide association scans. *American Journal of Human Genetics*, 81:913–926, 2007.

[14] Hyun Min Kang, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, David Heckerman, Mark J. Daly, and Eleazar Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 178:1709–1723, 2008.

[15] Zhiwu Zhang, Elhan Ersoz, Chao-Qiang Lai, Rory J. Todhunter, Hemant K. Tiwari, Michael A. Gore, Peter J. Bradbury, Jianming Yu, Donna K. Arnett, Jose M. Ordovas, and Edward S. Buckler. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42:355–360, 2010.

[16] Hyun Min Kang, Jae Hoon Sul, Susan K. Service, Noah A. Zaitlen, Sit-Yee Kong, Nelson B. Freimer, Chiara Sabatti, and Eleazar Eskin. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42:348–354, 2010.

[17] Alkes L. Price, Noah A. Zaitlen, David Reich, and Nick Patterson. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11:459–463, 2010.

[18] Christoph Lippert, Jennifer Listgarten, Ying Liu, Carl M. Kadie, Robert I. Davidson, and David Heckerman. FaST linear mixed models for genome-wide association studies. *Nature Methods*, 8:833–835, 2011.

[19] Xiang Zhou and Matthew Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44:821–824, 2012.

[20] Peter Donnelly Matti Pirinen and Chris C. A. Spencer. 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.

[21] Gulnara R Svishcheva, Tatiana I Axenovich, Nadezhda M Belonogova, Cornelia M van Duijn, and Yurii S Aulchenko. Rapid variance componentsbased method for whole-genome association analysis. *Nature Genetics*, 44:1166–1170, 2012.

[22] Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modelling with Bayesian sparse linear mixed models. *PLoS Genetics*, 9:e1003264, 2013.

[23] Jae Hoon Sul and Eleazar Eskin. Mixed models can correct for population structure for genomic regions under selection. *Nature Review Genetics*, 14:300, 2013.

[24] Alkes L. Price, Noah A. Zaitlen, David Reich, and Nick Patterson. Response to Sul and Eskin. *Nature Review Genetics*, 14:300, 2013.

[25] Samprit Banerjee, Brian S. Yandell, and Nengjun Yi. Bayesian quantitative trait loci mapping for multiple traits. *Genetics*, 179:2275–2289, 2008.

[26] Seyoung Kim and Eric P. Xing. Statistical estimation of correlated genome associations to a quantitative trait network. *PLOS Genetics*, 5:e1000587, 2009.

[27] Seyoung Kim, Kyung-Ah Sohn, and Eric P. Xing. A multivariate regression approach to association analysis of a quantitative trait network. *Bioinformatics*, 25:204–212, 2009.

[28] Manuel A. R. Ferreira and Shaun M. Purcell. A multivariate test of association. *Bioinformatics*, 25:132–133, 2009.

[29] Lei Zhang, Yu-Fang Pei, Jian Li, Christopher J. Papasian, and Hong-Wen Deng. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. *PLOS One*, 4:e6502, 2009.

[30] Paul F. OReilly, Clive J. Hoggart, Yotsawat Pomyen, Federico C. F. Calboli, Paul Elliott, Marjo-Riitta Jarvelin, and Lachlan J. M. Coin. Multiphen: Joint model of multiple phenotypes can increase discovery in GWAS. *PLOS One*, 7:e34861, 2012.

[31] Matthew Stephens. A unified framework for association analysis with multiple related phenotypes. *Submitted*, 0, 2013.

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

[33] Karin Meyer. 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.

[34] Karin Meyer and S. P. Smith. Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genetics Selection Evolution*, 28:23–49, 1996.

[35] Arthur P. Dempster, Nan M. Laird, and Donald B. Rubin. Maximum likelihood from incomplete data via the EM algirhtm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

[36] Chuanhai Liu, Donald B. Rubin, and Ying Nian Wu. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, 85:755–770, 1998.

[37] Jean-Louis Foulley and David A. van Dyk. The PX-EM algorithm for fast stable fitting of Henderson's mixed model. *Genetics Selection Evolution*, 32:143–163, 2000.

[38] H. D. Patterson and R. Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554, 1971.

[39] R. Thompson. The estimation of variance, covariance components with an application when records are subject to culling. *Biometrics*, 29:527–550, 1973.

[40] Arthur R. Gilmour, Robin Thompson, and Brian R. Cullis. Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 51:1440–1450, 1995.

[41] Karin Meyer. *PX×AI: Algorithmics for better convergence in restricted maximum likelihood estimation.* 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brasil, 2006.

[42] Jian Yang, Sang Hong Lee, Michael E. Goddard, and Peter M. Visscher. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88:76–82, 2011.

[43] Karin Meyer. 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.

[44] Yurii S. Aulchenko, Dirk-Jan de Koning, and Chris Haley. 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.

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

[46] Jennifer Listgarten, Christoph Lippert, Carl M Kadie, Robert I Davidson, Eleazar Eskin, and David Heckerman. Improved linear mixed models for genome-wide association studies. *Nature Methods*, 9:525–526, 2012.

[47] The International Multiple Sclerosis Genetics Consortium and The Wellcome Trust Case Control Consortium. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, 10:214–219, 2011.

[48] Brian J. Bennett, Charles R. Farber, Luz Orozco, Hyun Min Kang, Anatole Ghazalpour, Nathan Siemers, Michael Neubauer, Isaac Neuhaus, Roumyana Yordanova, Bo Guan, Amy Truong, Wen-Pin Yang, Aiqing He, Paul Kayne, Peter Gargalovic, Todd Kirchgessner, Calvin Pan, Lawrence W. Castellani, Emrah Kostem, Nicholas Furlotte, Thomas A. Drake, Eleazar Eskin, and Aldons J. Lusis. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Research*, 20:281–290, 2010.

[49] Chiara Sabatti, Susan K Service, Anna-Liisa Hartikainen, Anneli Pouta, Samuli Ripatti, Jae Brodsky, Chris G Jones, Noah A Zaitlen, Teppo Varilo, Marika Kaakinen, Ulla Sovio, Aimo Ruokonen, Jaana Laitinen, Eveliina Jakkula, Lachlan Coin, Clive Hoggart, Andrew Collins, Hannu Turunen, Stacey Gabriel, Paul Elliot, Mark I McCarthy, Mark J Daly, Marjo-Riitta Jrvelin, Nelson B Freimer, and Leena Peltonen. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genetics*, 41:35–46, 2008.

[50] Tanya M. Teslovich, Kiran Musunuru, Albert V. Smith, Andrew C. Edmondson, Ioannis M. Stylianou, Masahiro Koseki, and etc. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, 466:707–713, 2010.

[51] Abbas Dehghan, Josee Dupuis, Maja Barbalic, and ect. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*, 123:731–738, 2011.

[52] The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491:56–65, 2012.

[53] Karin Meyer and Mark Kirkpatrick. Better estimates of genetic covariance matrices by bending using penalized maximum likelihood. *Genetics*, 185:1097–1110, 2010.

[54] Mark Kirkpatrick and Karin Meyer. Direct estimation of genetic principal components. *Genetics*, 168:2295–2306, 2004.

[55] Daniel E. Runcie and Sayan Mukherjee. Bayesian sparse factor analysis of genetic covariance matrices. *Genetics*, 0:in press, 2013.

[56] Xiuwen Zheng, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28:3326–3328, 2012.

[57] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, and Pak C. Sham. PLINK: a toolset for whole-genome association and population-based linkage analysis. *The American Journal of Human Genetics*, 81:559–575, 2007.

[58] Ben J. Hayes, Peter M. Visscher, and Michael E. Goddard. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, 91:47–60, 2009.

[59] William Astle and David J. Balding. Population structure and cryptic relatedness in genetic association studies. *Statistical Science*, 24:451–471, 2009.

[60] Xiao-Li Meng and Donald B. Rubin. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80:267–278, 1993.

[61] Raghu N. Kackar and David A. Harville. Approximations for standard errors of estimators of fixed and random effect in mixed linear models. *Journal of the American Statistical Association*, 79:853–862, 1984.

[62] Michael G. Kenward and James H. Roger. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53:983–997, 1997.