Efficient multivariate linear mixed model algorithms for genome-wide association studies

Xiang Zhou[1,2] & Matthew Stephens[1,2]

Multivariate linear mixed models (mvLMMs) are powerful tools for testing associations between single-nucleotide polymorphisms and multiple correlated phenotypes while controlling for population stratification in genome-wide association studies. We present efficient algorithms in the genome-wide efficient mixed model association (GEMMA) software for fitting mvLMMs and computing likelihood ratio tests. These algorithms offer improved computation speed, power and *P*-value calibration over existing methods, and can deal with more than two phenotypes.

mvLMMs are statistical regression models that relate explanatory variables to multiple correlated outcome variables and have been widely applied in genetics because of their ability to account for relatedness among samples[1]. For example, they have been used to estimate the heritability of gene expression across tissues[2], assess pleiotropy and genetic correlation between complex phenotypes[3–6], detect quantitative trait loci[7], understand evolutionary patterns[8] and assist animal-breeding programs[9]. Recently, mvLMMs have become increasingly important in genome-wide association studies (GWAS), both because of their effectiveness in accounting for sample relatedness[3,7,10] and population stratification[3,11–17], and because of a growing appreciation of the power gains of multivariate association analyses over standard univariate analysis[3,18–22]. Multivariate analyses can increase power not only to detect pleiotropic genetic variants but also genetic variants that affect only one of multiple correlated phenotypes[22].

However, fitting mvLMMs is computationally nontrivial and involves a multidimensional optimization of a potentially non-convex function that may have multiple local optima. Current algorithms for fitting a single mvLMM (implemented in the software packages genome-wide complex trait analysis (GCTA)[4,23], Wombat[24] or ASReml[25]) use two types of optimization algorithm: an initial expectation-maximization–like (EM) algorithm, followed by a Newton-Raphson–like (NR) algorithm. This combines the benefits of the stability of the EM algorithm (every iteration increases the likelihood) with the faster convergence of the NR algorithm[26] (**Supplementary Note**). The computational complexity of these algorithms for n individuals and d phenotypes is $O(t_1n^3d^3 + t_2n^3d^7)$, where t_1, t_2 are the maximum number of iterations of EM and NR algorithms, respectively, and O is the big O notation (**Supplementary Note**). Performing the likelihood ratio test (LRT) in GWAS using these methods would require repeated application for each single nucleotide polymorphism (SNP), which is impractical for GWAS with a large number of SNPs and a moderate number of individuals. Consequently, existing methods cannot be used to perform the LRT for mvLMMs in GWAS settings. The only available method along these lines (multitrait mixed model; MTMM)[3] can only be used to perform an approximate LRT for two phenotypes.

We present a computationally efficient algorithm implemented in the GEMMA software[16,17] for fitting mvLMMs with one covariance component (in addition to the residual error term) and for performing the LRT for association in GWASs (**Supplementary Software** and http://stephenslab.uchicago.edu/software.html). The algorithm builds on linear algebra techniques previously used for univariate LMMs[12,13,17] and, combined with several additional innovations, extends them to multivariate LMMs. Our algorithm substantially reduces the computational burden of computing LRTs by avoiding repeating the expensive $O(n^3)$ operations for every SNP. Specifically, after an initial single $O(n^3)$ operation (eigendecomposition of the relatedness matrix), our algorithm has a per-SNP complexity that is $O(n^2)$, which reduces the overall computational complexity to $O(n^3 + n^2d + s(n^2 + t_1nd^2 + t_2nd^2))$, where s is the number of SNPs. In effect, our algorithm (**Supplementary Note**) provides the multivariate analog of both the univariate algorithm in the program efficient mixed-model association (EMMA)[27] and the improved univariate algorithms in factored spectrally transformed linear mixed models (FaST-LMM)[12], GEMMA[13] and conditional maximization (CM)[12,13,17]. Our work provides a practical approach to computing LRTs for mvLMMs in reasonably large GWAS (e.g., 50,000 individuals) and a modest number of phenotypes (e.g., 2–10 phenotypes).

To illustrate the benefits of our mvLMM algorithm, we used two data sets: a mouse GWAS from the Hybrid Mouse Diversity Panel (HMDP) with four blood lipid phenotypes and a human GWAS from the Northern Finland Birth Cohort 1966 (NFBC1966) with four blood metabolic traits (Online Methods). The HMDP data are a small GWAS with strong relatedness among many individuals; the NFBC1966 data are a larger GWAS with weak relatedness among most individuals.

Even for fitting a single mvLMM, our algorithm in GEMMA is substantially faster than GCTA and Wombat (**Table 1**). For example, for the NFBC1966 data with four phenotypes, GEMMA takes about 7 min compared with 8 h for Wombat. Moreover, unlike other methods, computing time for GEMMA is essentially the

[1]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [2]Department of Statistics, University of Chicago, Chicago, Illinois, USA. Correspondence should be addressed to X.Z. (xz7@uchicago.edu) or M.S. (mstephens@uchicago.edu).

RECEIVED 6 MAY 2013; ACCEPTED 13 JANUARY 2014; PUBLISHED ONLINE 16 FEBRUARY 2014; DOI:10.1038/NMETH.2848

Table 1 | Computation durations for parameter estimation in a single mvLMM and for association analysis in GWAS

		Computation time					
		HMDP data (n = 656, s = 108,562)			NFBC1966 data (n = 5,255, s = 319,111)		
Method	**Time complexity**	**2 traits**	**3 traits**	**4 traits**	**2 traits**	**3 traits**	**4 traits**
Fitting a single mvLMM							
GEMMA	$O(n^3 + n^2d + n^2c + t_1nc^2d^2 + t_2nc^2d^6)$	<1 s	<1 s	<1 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	–	–
Genome-wide applications							
GEMMA	$O(n^3 + n^2d + n^2c + s(n^2 + t_1nc^2d^2 + t_2nc^2d^6))$	6.2 min	13.7 min	28.5 min	4.4 h	4.8 h	5.8 h
MTMM	$O(t_1n^3(d + c)^3 + t_2n^3d^7 + sn^2d^2)$	16.4 min	–	–	58.0 h	–	–

Computation was performed on a single core of an Intel Xeon L5420 2.50 GHz processor. n, number of individuals; s, number of SNPs; d, number of traits; c, number of covariates (c = 1 here); t_1, number of iterations used in the EM type algorithm and t_2, number of iterations used in the NR-type algorithm. –, not applicable.

same for any number of phenotypes because the computing time is dominated by the initial eigendecomposition. Thus, the gains for more phenotypes would be even greater.

The more important gains of our algorithm are in GWAS applications, where existing algorithms are not practical for computing the LRT for even two phenotypes. An extrapolation from the data in **Table 1** suggests that computation with existing algorithms might take over 14 d for HMDP data and over 18 years for NFBC1966 data. The MTMM software[3] can analyze two phenotypes, but it uses an approximate LRT[11,15] to reduce per-SNP computation time to $O(n^2)$. The approximate LRT avoids the expensive repeated optimization of the variance components for each SNP but is guaranteed to underestimate the LRT (**Supplementary Note**), and in the univariate setting this has been shown to produce miscalibrated P values and/or loss of power[13,17].

To illustrate the benefits of LRT over approximate LRT in the multivariate setting, we performed null and alternative simulations using the HMDP data (Online Methods). Consistent with the univariate findings, MTMM P values were systematically larger than expected under the null, with the most significant P values rising by almost an order of magnitude (**Fig. 1a**). In contrast, P values from the GEMMA LRT were well-calibrated (**Fig. 1a**). Although in principle the mvLMM likelihood surface could harbor multiple local optima, causing our P values to be miscalibrated, in practice this did not happen in any situation we examined. However, we found that obtaining well-calibrated P values requires both the EM and NR algorithms: use of only the EM algorithm can lead to poor convergence of the LRT, which results in underestimation of P values similar to the case with MTMM (**Supplementary Fig. 1**). The systematic inflation of MTMM P values under the null presumably accounts for MTMM's loss of power relative to GEMMA in simulations under the alternative (**Fig. 1c**).

We also compared performance of GEMMA and MTMM on real phenotypes for both data sets. As MTMM is implemented only for two phenotypes, we analyzed all pairs of traits. For these data, GEMMA ran 2–12 times faster than MTMM (**Table 1**). For NFBC1966 data, GEMMA took ~4 h for a two-phenotype analysis that took MTMM almost 2.5 d. Consistent with the simulations and with theory, the MTMM P values for HMDP analyses were consistently less significant (up to sixfold) than P values from GEMMA (**Fig. 1b**) and in many cases were substantially less significant than would be expected even under the null (**Supplementary Figs. 2** and **3**). For NFBC1966, the two methods produced similar P values (**Supplementary Figs. 4** and **5**), consistent with univariate assessments that show that the approximate

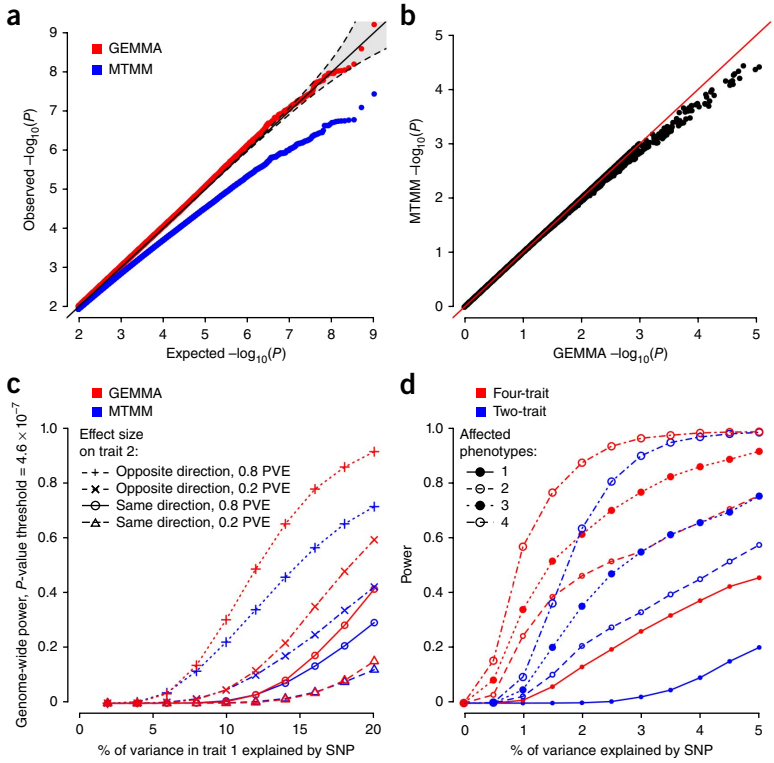

Figure 1 | Statistical benefits of the mvLMM algorithm implemented in GEMMA. (**a**) Quantile-quantile plot showing the improved calibration of GEMMA P values compared with those from MTMM for simulated null data. Gray shaded area between dashed lines indicates 0.025 and 0.975 point-wise quantiles of the ordered P values under the null distribution. Solid diagonal line shows the line $y = x$. (**b**) Comparison of MTMM P values against GEMMA P values in the HMDP data. (**c**) Gain in power for GEMMA compared with MTMM in four different simulation scenarios based on HMDP data. x axis shows the percentage of phenotypic variance in the first phenotype explained (PVE) by the SNP, and symbols indicate the SNP effect direction (compared with its effect on the first phenotype) and size (quantified by PVE) on the second phenotype. (**d**) Simulation results illustrating the potential gain in power from four-phenotype versus two-phenotype analyses.

LRT works well in large samples in which individuals are not closely related and the marker effect size is small.

Our method also makes mvLMM analysis possible for GWAS with more than two phenotypes. On simulations based on HMDP and NFBC1966 data, we compared the power of performing the multivariate LRT on all four phenotypes to conducting all six two-phenotype analyses with a Bonferroni correction for multiple testing (**Fig. 1d** and **Supplementary Fig. 6**). The four-phenotype analysis was consistently as powerful or more so than the two-phenotype analyses, even when only one or two of the four phenotypes were truly associated with the genotype. Although this may seem counterintuitive when exactly two phenotypes are associated with a genotype, it is actually expected because unassociated phenotypes in the multivariate analysis can increase power if they are correlated with the associated phenotypes[22].

We also applied four-phenotype, two-phenotype and univariate analyses to NFBC1966 data. From all of these, 45 SNPs from 14 genetic regions passed a significance level of 0.05 after Bonferroni correction (both for the number of SNPs and, in univariate and two-phenotype analyses, for the number of tests). As expected, some SNPs showed stronger signals in the four-phenotype analysis, whereas others showed stronger signals in the other analyses. When we compared four-phenotype and univariate analyses (**Supplementary Table 1** and **Supplementary Fig. 7**), 16 SNPs were significant in the four-phenotype analysis and not the univariate analysis, whereas 3 SNPs were significant only in the univariate analysis. When we compared four-phenotype and two-phenotype analyses (**Supplementary Table 2** and **Supplementary Fig. 7**), 1 SNP was significant in the four-phenotype analysis and not the two-phenotype analysis, whereas no SNPs were significant only in the two-phenotype analysis.

These results support the idea that multivariate tests can be more powerful than multiple univariate or pairwise tests. It is also clear, however, that in a GWAS setting no single test will be the most powerful in detecting the many different types of genetic effects that could occur. It is possible to manufacture simulations so that any given test is most powerful[22]. Thus, different multivariate and univariate tests should be viewed as complementary rather than competing.

Our algorithm is not without limitations. Perhaps the most fundamental is that, like its univariate counterparts, our algorithm only applies to mvLMMs with one variance component (in addition to the residual error term). However, with additional assumptions it may be extended to more variance components[28]. Our method also requires complete phenotype data, but this can be addressed by imputing missing phenotypes before performing association tests (**Supplementary Note** and **Supplementary Fig. 8**). Finally, although our implementation of the EM algorithm could theoretically be applied to a reasonably large number of phenotypes because it scales only quadratically in this dimension, in practice computational and statistical barriers remain for even modest numbers of phenotypes (e.g., ~10 phenotypes). Computationally, the number of iterations required to converge for more phenotypes will inevitably increase, and ultimately this could be the main barrier that limits the number of phenotypes. Statistically, the number of parameters in the mvLMM is also quadratic in the number of phenotypes ($d(d+1)$ parameters in the two variance components). Therefore, with moderate sample size, it may be desirable to assume structure for the variance components or incorporate additional prior information[29].

The most computationally expensive part of our method, as in the univariate case, is the initial eigendecomposition. This requires a large amount of physical memory and also becomes computationally intractable for large n values (e.g., >50,000). Low rank approximations to the relatedness matrix[12,17,30] can alleviate both computation and memory requirements, and could allow mvLMMs to be applied to very large GWAS.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was supported in part by US National Institutes of Health (NIH) grant HL092206 (principal investigator, Y. Gilad) and NIH grant HG02585 to M.S. We thank A.J. Lusis for making the mouse genotype and phenotype data available, and 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, University of California Los Angeles, 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 their views or those of their host institutions.

AUTHOR CONTRIBUTIONS

X.Z. and M.S. conceived the idea and designed the study. X.Z. developed the algorithms, implemented the software and performed the analyses. X.Z. and M.S. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Henderson, C.R. *Applications of Linear Models in Animal Breeding* (University of Guelph, 1984).
2. Price, A.L. *et al. PLoS Genet.* **7**, e1001317 (2011).
3. Korte, A. *et al. Nat. Genet.* **44**, 1066–1071 (2012).
4. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. *Bioinformatics* **28**, 2540–2542 (2012).
5. Trzaskowski, M., Yang, J., Visscher, P.M. & Plomin, R. *Mol. Psychiatry* doi:10.1038/mp.2012.191 (29 January 2013).
6. Vattikuti, S., Guo, J. & Chow, C.C. *PLoS Genet.* **8**, e1002637 (2012).
7. Amos, C.I. *Am. J. Hum. Genet.* **54**, 535–543 (1994).
8. Kruuk, L.E. *Phil. Trans. R. Soc. Lond. B* **359**, 873–890 (2004).
9. Meyer, K., Johnston, D.J. & Graser, H.U. *Aust. J. Agric. Res.* **55**, 195–210 (2004).
10. Meyer, K. *Genet. Sel. Evol.* **23**, 67–83 (1991).
11. Kang, H.M. *et al. Nat. Genet.* **42**, 348–354 (2010).
12. Lippert, C. *et al. Nat. Methods* **8**, 833–835 (2011).
13. Pirinen, M., Donnelly, P. & Spencer, C.C.A. *Ann. Appl. Stat.* **7**, 369–390 (2013).
14. Yu, J.M. *et al. Nat. Genet.* **38**, 203–208 (2006).
15. Zhang, Z.W. *et al. Nat. Genet.* **42**, 355–360 (2010).
16. Zhou, X., Carbonetto, P. & Stephens, M. *PLoS Genet.* **9**, e1003264 (2013).
17. Zhou, X. & Stephens, M. *Nat. Genet.* **44**, 821–824 (2012).
18. Banerjee, S., Yandell, B.S. & Yi, N.J. *Genetics* **179**, 2275–2289 (2008).
19. Ferreira, M.A.R. & Purcell, S.M. *Bioinformatics* **25**, 132–133 (2009).
20. Kim, S. & Xing, E.P. *PLoS Genet.* **5**, e1000587 (2009).
21. O'reilly, P.F. *et al. PLoS One* **7**, e34861 (2012).
22. Stephens, M. *PLoS One* **8**, e65245 (2013).
23. Yang, J.A., Lee, S.H., Goddard, M.E. & Visscher, P.M. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
24. Meyer, K. *J. Zhejiang Univ. Sci. B* **8**, 815–821 (2007).
25. Gilmour, A.R., Thompson, R. & Cullis, B.R. *Biometrics* **51**, 1440–1450 (1995).
26. Meyer, K. PX × AI: Algorithmics for better convergence in restricted maximum likelihood estimation. in *8th World Congress on Genetics Applied to Livestock Production* (Belo Horizonte, Brasil, 2006).
27. Kang, H.M. *et al. Genetics* **178**, 1709–1723 (2008).
28. Kostem, E. & Eskin, E. *Am. J. Hum. Genet.* **92**, 558–564 (2013).
29. Runcie, D.E. & Mukherjee, S. *Genetics* **194**, 753–767 (2013).
30. Listgarten, J. *et al. Nat. Methods* **9**, 525–526 (2012).

ONLINE METHODS

Genotype and phenotype data. We analyzed two data sets: HMDP[31] and NFBC1966 study[32].

The HMDP data include data for 100 inbred strains with four phenotypes (high-density lipoprotein, HDL; total cholesterol, TC; triglycerides, TG; and unesterified cholesterol, UC) and four million high-quality fully imputed SNPs (SNPs were downloaded from http://mouse.cs.ucla.edu/mousehapmap/full.html). We excluded mice with missing phenotypes for any of these four phenotypes. We excluded nonpolymorphic 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[33]). 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 misspecification. We used the product of centered genotype matrix as an estimate of relatedness[16,17,34,35]. Note that the sample size used here is smaller than in the original study[31], and the phenotypes were quantile-transformed instead of log-transformed for robustness.

The NFBC1966 data contain information for 5,402 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[11,32] with the software PLINK[33]. 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 effects of sex, oral contraceptives and pregnancy status[32], 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[11,17,34,35].

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 (e.g., ref. 22). For both data sets, we used a standard mvLMM with an intercept term (without any other covariates), and tested 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.

Simulations. To check whether GEMMA and MTMM produce calibrated P values, we randomly selected 100,000 real genotypes in the HMDP data. We simulated 10,000 phenotypes under the null, based on the real relatedness matrix and the estimated genetic and environmental covariance matrices (for HDL and TG). We calculated P values for each SNP-phenotype pair in turn (one billion pairs). We did not perform comparisons based on the NFBC1966 data, partly because GEMMA and MTMM produced identical P values there and partly because the sample size in NFBC1966 data makes it computationally impractical to perform billions of association tests to check for the type I error at the genome-wide significance level.

To compare power between GEMMA and MTMM, we used real genotypes from the HMDP and NFBC1966 data, and we simulated phenotypes by adding genotype effects back to the original phenotypes[15,17]. Specifically, we first identified SNPs unassociated with the four phenotypes based on one-phenotype, two-phenotype and four-phenotype analyses (LRT $P > 0.1$ in any of the 11 tests). We ordered the SNPs (76,780 in HMDP data and 208,145 in NFBC1966 data) that satisfied these criteria by their genomic location, and selected from these SNPs 10,000 evenly spaced SNPs to act as causal SNPs. For each causal SNP, we specified its effect size for the first trait (HDL) to explain a particular percentage of the phenotypic variance (proportion of variance explained, or PVE). Afterward, we specified its effect for the second trait (TG) so that the proportion of variance in the second trait explained by the SNP equaled either 20% or 80% of the PVE in the first trait. We considered effect sizes for the two traits to be either in the same direction or in the opposite direction, and we added the simulated effects back to the original phenotypes to form the new simulated phenotypes. For each prespecified PVE (ranged from 2% to 20% in HMDP data and 0.04% to 0.4% in NFBC1966 data), we simulated 10,000 sets of phenotypes, one for each causal SNP, and calculated the P value for each SNP-phenotype pair. We calculated statistical power as the proportion of P values exceeding the genome-wide significance level at the conventional 0.05 level after Bonferroni correction ($P = 4.6 \times 10^{-7}$ for HMDP data and $P = 1.6 \times 10^{-7}$ for NFBC1966 data). Note that we simulated phenotypes based on HDL and TG in both data sets, and the two phenotypes were positively correlated in HMDP data but negatively correlated in NFBC1966 data.

Our algorithms rely on fully observed phenotypes. To make the method more widely applicable, we developed a phenotype imputation scheme to impute missing phenotypes where necessary (**Supplementary Note**). To show the power gain of our imputation scheme versus simply dropping data for individuals with partially missing phenotypes, we performed a simulation study. Specifically, we used the same set of simulated phenotypes described above, but randomly made 2.5%, 5% or 10% of the individuals to have one phenotype missing. We calculated P values for each SNP-phenotype pair from the two approaches using GEMMA, and calculated statistical power at the conventional 0.05 level after Bonferroni correction.

Finally, we performed a power comparison between the four-phenotype analysis and the two-phenotype analysis using GEMMA, using simulations based on the two data sets. Specifically, we used the same set of 10,000 SNPs described above to act as causal SNPs, and we simulated phenotypes by adding genotype effects to the observed phenotypes, as above. For each causal SNP, we made it to affect either one, two, three or four phenotypes. When the causal SNP affected two or four phenotypes, its effects on a randomly selected half of the traits were in the opposite direction of its effects on the other half. When the causal SNP affected three phenotypes, its effects on two randomly selected traits were in the opposite direction of its effect

doi:10.1038/nmeth.2848

on the third trait. The SNP effect size for each affected phenotype was simulated independently to account for a pre-specified PVE of that phenotype (ranged from 0.5% to 5% in HMDP data and 0.04% to 0.4% in NFBC1966 data), which was further scaled with a random factor draw from a uniform distribution U(0.8, 1). The simulated effects were added back to the original phenotypes to form the new simulated phenotypes. For the four-phenotype analysis, we calculated the *P* value for each SNP-phenotype pair and we calculated statistical power at the conventional 0.05 level after Bonferroni correction ($P = 4.6 \times 10^{-7}$ for HMDP and $P = 1.6 \times 10^{-7}$ for NFBC1966). For the two-phenotype analysis, we obtained the minimal *P* value from the six pair-wise analyses for each SNP, and calculated statistical power as the proportion of these *P* values exceeding either the same significance level ($P = 4.6 \times 10^{-7}$ for HMDP data and $P = 1.6 \times 10^{-7}$ for NFBC1966 data), or a significance level that was further adjusted to account for the six tests performed ($P = 7.6 \times 10^{-8}$ for HMDP and $P = 2.6 \times 10^{-8}$ for NFBC1966).

31. Bennett, B.J. *et al. Genome Res.* **20**, 281–290 (2010).
32. Sabatti, C. *et al. Nat. Genet.* **41**, 35–46 (2009).
33. Purcell, S. *et al. Am. J. Hum. Genet.* **81**, 559–575 (2007).
34. Astle, W. & Balding, D.J. *Stat. Sci.* **24**, 451–471 (2009).
35. Hayes, B.J., Visscher, P.M. & Goddard, M.E. *Genet. Res.* **91**, 143–143 (2009).