




CORE GREML for estimating covariance between random effects in linear mixed models for complex trait analyses

Xuan Zhou^{1,2,3}, Hae Kyung Im⁴  & S. Hong Lee^{1,2,3}  

As a key variance partitioning tool, linear mixed models (LMMs) using genome-based restricted maximum likelihood (GREML) allow both fixed and random effects. Classic LMMs assume independence between random effects, which can be violated, causing bias. Here we introduce a generalized GREML, named CORE GREML, that explicitly estimates the covariance between random effects. Using extensive simulations, we show that CORE GREML outperforms the conventional GREML, providing variance and covariance estimates free from bias due to correlated random effects. Applying CORE GREML to UK Biobank data, we find, for example, that the transcriptome, imputed using genotype data, explains a significant proportion of phenotypic variance for height (0.15, p -value = 1.5e-283), and that these transcriptomic effects correlate with the genomic effects (genome-transcriptome correlation = 0.35, p -value = 1.2e-14). We conclude that the covariance between random effects is a key parameter for estimation, especially when partitioning phenotypic variance by multi-omics layers.

¹ Australian Centre for Precision Health, University of South Australia, Adelaide, SA 5000, Australia. ² UniSA Allied Health and Human Performance, University of South Australia, Adelaide, SA 5000, Australia. ³ South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia. ⁴ Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA. ✉email: hong.lee@unisa.edu.au

Genome-wide association studies (GWASs) have been incredibly successful in identifying genetic variants associated with complex traits. However, the proportion of phenotypic variance explained by genome-wide significant single nucleotide polymorphisms (SNPs) is far lower than the narrow-sense heritability estimate¹. This is largely because GWASs typically examine SNP trait associations one at a time, generating a large number of tests across the genome for which the stringent Bonferroni correction is applied. To overcome this problem, a whole-genome approach that jointly considers all available SNPs has been introduced², allowing estimation of the proportion of phenotypic variance explained by genome-wide SNPs, i.e., SNP-based heritability. Central to this approach is the use of linear mixed models³ (LMMs)—extensions of random-effects models or variance-component models⁴—which treat SNP effects as random.

Using genome-based restricted maximum likelihood (GREML) for parameter estimation, LMMs are a key tool not only for SNP-based heritability of complex traits but also for variance partitioning in general. For example, when heritability is partitioned by functional annotation of SNPs using a LMM with multiple random effects, GREML estimation has provided important insights into the latent genetic architecture of complex traits⁵. As multi-omics data become increasingly available⁶, variance partitioning using LMMs will become indispensable to uncover the relative contributions of multiple ‘omes’ to phenotypic variation. Alongside GREML, linkage disequilibrium score regression (LDSC) provides an alternative way for SNP-based heritability estimation and variance partitioning, using only GWAS summary statistics without the need to access individual genotypes^{7,8}. LDSC or stratified LDSC also treats SNP effects as random⁸, and for the same set of individual-level genotype data, this approach generates similar estimates as GREML^{9–11}.

Following classic LMMs^{3,4}, GREML assumes independence between random effects when estimating variance components. However, it is questionable if this assumption is always valid, especially for genomic analyses of complex traits. For example, gene regulatory networks shared between functional categories may generate non-negligible correlations between effects of these categories on phenotypes¹². In the context of phenotypic variance partitioning by multi-omics layers, effects of genetic variants and their expression levels on phenotypes are likely correlated^{13–15}, as exemplified by overlaps between GWAS loci and expression quantitative trait loci^{16,17} (eQTL). Given these justifiable covariance terms in genomic analyses of complex traits, the naive assumption of independence between random effects held by GREML can lead to a biased partition of phenotypic variance and false inferences on the underlying architecture of complex traits.

Here we introduce an alternative GREML, named CORE GREML (CORE for COvariance between Random Effects), which fits the Cholesky decomposition of kernel matrices in an LMM, to estimate the covariance between a given pair of random effects. Using extensive simulations, we show that CORE GREML outperforms GREML, providing estimates of variance and covariance components that are free from bias due to correlated random effects. We also apply CORE GREML to real data from the UK Biobank, to demonstrate its use for genomic partitioning analyses and for genome-transcriptome partitioning of phenotypic variance. We conclude that the covariance between random effects is a key parameter that needs to be estimated, especially for multi-omics analyses of phenotypic variance.

Results

Methods overview. The proposed method, CORE GREML, is an extension of GREML, in that it uses the Cholesky theorem to

derive the covariance structure between relationship kernel matrices of random effects for LMM-based partitioning of phenotypic variance (see Methods for details). To validate CORE GREML, we simulated 500 replicates of phenotypic data ($n = 10,000$) under settings where the covariance between random effects was zero, positive, and negative (see Supplementary Table 1 for parameter settings). Comparing the model fit of GREML with that of CORE GREML for simulation replicates under the null setting (i.e., no covariance), we estimated the type I error rate of detecting covariance between random effects. Under all settings, we also determined the extent to which CORE GREML recovered the true values of model parameters and compared CORE GREML and GREML estimates to show the impact of neglecting covariance terms. To facilitate interpretation of results from real data analyses, phenotypic data under all settings were simulated using available genomic and transcriptomic data as for real data analyses.

For analysis of real data, we selected ten traits with the highest heritability estimates (see Supplementary Fig. 1 for SNP heritability estimates) from the UK Biobank data that are available to us (reference number 14575). These traits are standing height, sitting height, body mass index (BMI), heel bone mineral density, fluid intelligence, weight, waist circumference, hip circumference, diastolic blood pressure, and years of education. For each trait, we conducted two separate sets of variance partitioning analyses, which are genomic partitioning by functional region and genome-transcriptome partitioning of phenotypic variance. For each analysis, we applied GREML and CORE GREML, and compared the model fit of the two methods to test the significance of the covariance terms between random effects. Where necessary, we performed a five-fold cross-validation to compare the prediction accuracy of CORE GREML against that of GREML. Of note, both GREML and CORE GREML use relationship kernel matrices for variance-components estimation (Methods). The kernel matrices for genomic partitioning analyses were constructed using genotypes of 75,396 SNPs from coding regions, untranslated regions, and promoters (collectively referred to as “regulatory regions” thereafter), 255,665 from the DNase I hypersensitivity sites (DHSs), and 799,935 for all other regions (referred to as “other regions” thereafter). For the genome-transcriptome partitioning of phenotypic variance, the kernel matrix for genetic variance estimation was constructed using genotypes of 1,133,273 genome-wide SNPs and the kernel matrix for the estimation of phenotypic variance explained by the transcriptome was based on imputed expression levels of 227,664 genes from 43 tissues¹⁸ (see Supplementary Table 2), respectively. Importantly, our primary interest was not variance partitioning per se; rather our intention was to demonstrate the use of CORE GREML to detect and estimate covariance terms between random effects in mixed-model-based variance-component analyses and to show the impact of neglecting covariance terms on variance-components estimation.

Method validation by simulation. For phenotypic data simulated under the genomic partitioning model (see Methods) with zero covariance between effects of genomic regions on phenotypes (i.e., the null setting), the CORE GREML vs. GREML comparison yielded significant results (at a significance threshold of 0.05) for 19 replicates out of 500, giving an estimated type I error rate of 0.038 for detecting covariance terms. Similarly, for data simulated using the genome-transcriptome model under the null setting (see Methods for the simulation model), the estimated type I error rate was 0.042. Thus, for both simulation scenarios, type I error rate was not inflated.

In terms of parameter estimation, regardless of the simulation model and the parameter setting, CORE GREML consistently yielded unbiased estimates of all model parameters (Supplementary Figs. 2 and 3). In contrast, GREML only produced unbiased variance estimates of random effects under the null setting where random effects are not correlated ($\text{cov} = 0$ in Supplementary Figs. 2 and 3). As expected, GREML overestimated and underestimated the variance of random effects in the presence of positive and negative correlations between random effects, respectively, and the biased estimation was evident for both genomic partitioning and genome-transcriptome partitioning of phenotypic variance (Supplementary Figs. 2 and 3). Thus, our simulation results validate that CORE GREML properly partitions phenotypic variance whether or not the random effects in a LMM are correlated with each other. These results also indicate that GREML would produce biased variance-components estimates when random effects are correlated.

Assumption on genetic architecture. Incorrect assumptions in the estimation model about the genetic architecture of the trait can also bias variance-components estimation in the context of GREML¹⁹. Therefore, we tested the extent to which CORE GREML estimation is sensitive to a wrong assumption of the genetic architecture in the estimation model. This was achieved by simulating phenotypes under different genetic architectures (see Methods) and comparing CORE GREML estimation from fitting an estimation model that has the correct assumption about the genetic architecture, referred to as the “true model,” with that from fitting a “wrong model” that has an incorrect assumption about the genetic architecture.

We found that misspecification of genetic architecture in the estimation model in general biased CORE GREML estimation of variance components but not the covariance term for genome-transcriptome partitioning of phenotypic variance (see Supplementary Fig. 4). Nonetheless, under any given genetic architecture, misspecification can be feasibly diagnosed by comparing the likelihood of estimation models that assume a wide range of possible genetic architectures. As shown in Supplementary Fig. 5, differences in the likelihood of estimation models are highly indicative of deviations from the true underlying genetic architecture. In fact, a grid search approach has been practiced, choosing the model closest to the true underlying genetic architecture in the GREML context¹⁹.

In light of the above results, to reduce the chance of misspecification of genetic architecture for real data analyses, we fitted two estimation models, the Genome-wide Complex Trait Analysis (GCTA) model² and the Linkage Disequilibrium Adjusted Kinships (LDAK) model (with parameter α , which controls the extent to which minor allele frequency (MAF) affects the variance of SNP-specific effects on phenotypes, set at the recommended default -0.25 ¹⁹). We found that for all traits, the GCTA model had a better fit than the LDAK model, irrespective of estimation method (i.e., GREML or CORE GREML; see Supplementary Tables 3 and 4), indicating that the GCTA model is closer to the true genetic architecture than the LDAK model for our selected traits. Nonetheless, heritability estimates by the two models do not differ substantially (Supplementary Table 3) and significant covariance terms detected by the GCTA model remain significant when using the LDAK model (Supplementary Table 4; although the GCTA model seems more conservative than the LDAK model for detecting covariance terms). This is consistent with the previous observation that heritability estimates based on high-quality common SNPs are robust to variations in the assumed genetic architecture (more specifically, parameter α -values of the LDAK model; see

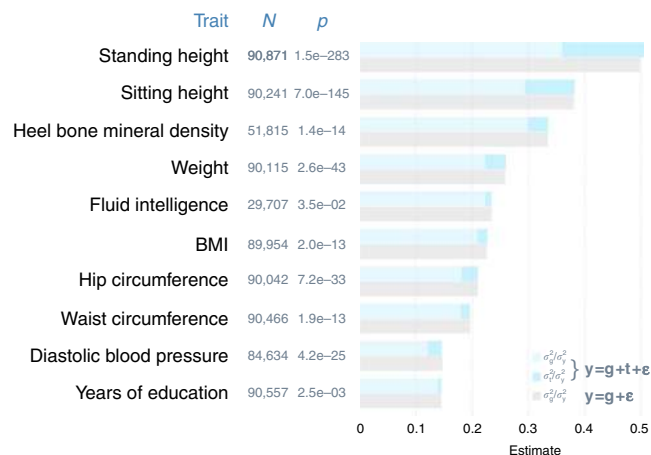


Fig. 1 Imputed transcriptome contributes to phenotypic variance. Shown are estimated variance components as a proportion of total phenotypic variance from a linear mixed model that includes the random effects of the imputed transcriptome and another model that does not, denoted as $\mathbf{y} = \mathbf{g} + \mathbf{t} + \boldsymbol{\varepsilon}$ and $\mathbf{y} = \mathbf{g} + \boldsymbol{\varepsilon}$, respectively. N = sample size; p = p -values from likelihood ratio tests ($df = 1$) that compare the two models to detect significant effects of the imputed transcriptome. \mathbf{g} = the random effects of the genome; \mathbf{t} = the random effects of the imputed transcriptome; $\boldsymbol{\varepsilon}$ = residuals; σ_t^2 = phenotypic variance explained by the imputed transcriptome; σ_g^2 = phenotypic variance explained by the genome; σ_y^2 = total phenotypic variance. The imputed transcriptome consists of expression levels of 227,664 genes from 43 non-sex-specific tissues. Source data are provided as a Source Data file.

Speed et al.¹⁹). Given the above, unless specified otherwise, results presented in the main text below are GCTA-based; LDAK-based results are included in Supplementary Tables 3 and 4.

Real data analyses. Intuitively, the covariance between any pair of random effects would not exist if the variance of any of the random effects is negligible. Therefore, prior to covariance estimation, we tested whether the variance components of interest differ from zero for the ten selected traits. For genomic partitioning, we estimated genetic variance by functional region using GREML and found that for all traits all variance components were different from zero by Wald’s tests (Supplementary Table 5). For genome-transcriptome partitioning of phenotypic variance, given that all selected traits are high in heritability (i.e., large genetic variances), we tested whether the imputed transcriptome could explain a significant proportion of phenotypic variance. To do so, we fitted two models using GREML, a “G model” that breaks phenotypic effects into the random effects of the genome and residuals, i.e., $\mathbf{y} = \mathbf{g} + \boldsymbol{\varepsilon}$, and a “G-T model” that decomposes phenotypic effects into the random effects of the genome and the imputed transcriptome and residuals, i.e., $\mathbf{y} = \mathbf{g} + \mathbf{t} + \boldsymbol{\varepsilon}$. We declared the presence of significant effects of the imputed transcriptome when the G-T model had a better fit than the G model using the likelihood ratio test with one degree of freedom. We found significant effects of the imputed transcriptome for all traits, except fluid intelligence (Fig. 1). Interestingly, although the G-T model had a much better fit than the G model for the nine traits, the two models explained a similar amount of phenotypic variance (Fig. 1), which was verified by additional simulations (see Supplementary Note 1 and Supplementary Fig. 6). This suggests that the partition of phenotypic variance represented by the G-T model is closer to the true underlying model than that represented by the G model.

Importantly, the effects of the imputed transcriptome on phenotypes are orthogonal to the effects of the SNPs that were

used to impute the transcriptome. To show this, we performed two analyses. First, we constructed a GRM with the 1,316,391 SNPs that were used for the transcriptome imputation and denoted their effects as \mathbf{g}_1 in a model that expresses the phenotypes of a trait, \mathbf{y} , as $\mathbf{y} = \mathbf{g}_1 + \mathbf{t} + \boldsymbol{\varepsilon}$, where \mathbf{t} is the random effects of the imputed transcriptome and $\boldsymbol{\varepsilon}$ is residuals. If the imputed transcriptome has any effects on phenotypes that are orthogonal to \mathbf{g}_1 alone, this model would have a better fit than a reduced model with \mathbf{g}_1 only, i.e., $\mathbf{y} = \mathbf{g}_1 + \boldsymbol{\varepsilon}$. Based on the likelihood ratio test with one degree of freedom, $\mathbf{y} = \mathbf{g}_1 + \mathbf{t} + \boldsymbol{\varepsilon}$ had a better fit than $\mathbf{y} = \mathbf{g}_1 + \boldsymbol{\varepsilon}$ for all traits, except years of education and fluid intelligence (see Supplementary Table 6), indicating that \mathbf{t} is orthogonal to \mathbf{g}_1 for most traits.

Second, we compared two models, $\mathbf{y} = \mathbf{g}_0 + \mathbf{g}_1 + \boldsymbol{\varepsilon}$ and $\mathbf{y} = \mathbf{g}_0 + \mathbf{t} + \boldsymbol{\varepsilon}$, where \mathbf{g}_0 denotes the random effects of 1,131,002 SNPs in our genome-transcriptome partitioning analyses of a complex trait (i.e., the G-T model mentioned above). It is noted that the set of SNPs for estimating \mathbf{g}_0 were ones that remained after applying our quality control criteria, whereas the 1,316,391 SNPs for the transcriptome imputation (hence for estimating \mathbf{g}_1) were based on the transcriptome imputation protocol¹⁸. The two sets of SNPs overlap by 716,636. If the two-GRM model ($\mathbf{y} = \mathbf{g}_0 + \mathbf{g}_1 + \boldsymbol{\varepsilon}$) is no different from the genome-imputed transcriptome model ($\mathbf{y} = \mathbf{g}_0 + \mathbf{t} + \boldsymbol{\varepsilon}$), their likelihoods would be very similar. Contrary to this, we found that $\mathbf{y} = \mathbf{g}_0 + \mathbf{t} + \boldsymbol{\varepsilon}$ had a much better fit than $\mathbf{y} = \mathbf{g}_0 + \mathbf{g}_1 + \boldsymbol{\varepsilon}$ for all traits, except heel bone mineral density, years of education, and fluid intelligence (see Supplementary Table 6), confirming that \mathbf{g}_1 and \mathbf{t} are distinct for most traits. In addition, the results also indicate that $\mathbf{y} = \mathbf{g}_0 + \mathbf{t} + \boldsymbol{\varepsilon}$ is closer to the true underlying model than $\mathbf{y} = \mathbf{g}_0 + \mathbf{g}_1 + \boldsymbol{\varepsilon}$ for most traits.

To validate the transcriptomic effects on phenotypes revealed by the G-T model, we performed a five-fold cross-validation, in which the phenotypic prediction accuracy of the G-T model was compared against that of the G model. For each trait, we randomly split the sample into a training set (~80%) and a validation set (~20%), and iterated this process five times in a manner such that validation sets did not overlap across iterations. To derive the prediction accuracy for the two models, we computed the Pearson's correlation coefficient between the observed and predicted phenotypes of each trait in each iteration and averaged correlation estimates across five iterations. Figure 2a shows that the gain in the phenotypic prediction accuracy by the G-T model relative to that by the G model grew as the estimated transcriptomic contribution to phenotypic variance increased ($p = 1.86e - 06$), suggesting that the transcriptomic effects on phenotypes of the selected traits are genuine. Taken together, our results thus far indicate that the variance components of interest differ from zero for all traits with fluid intelligence being the exception for the genome-transcriptome partitioning of phenotypic variance. Importantly, these results established the basis for our subsequent covariance estimation.

By comparing model fit by GREML and CORE GREML, we detected significant covariance between the random effects of the regulatory regions and DHS for height and sitting height (Fig. 3). Of note, the genomic partitioning model included three covariance terms, but two of these terms were not significant for height and sitting height, based on the Wald's test with one degree of freedom. We therefore reduced the model for these two traits by dropping nonsignificant covariance terms, noting that the fit of the reduced model did not differ from the full model ($p = 0.85$ and 0.37 for height and sitting height, respectively). We subsequently used estimates from the reduced model for these two traits. In genome-transcriptome analyses, we found significant covariance between the random effects of the genome and those of the imputed transcriptome for height, sitting height, heel bone mineral density, and diastolic blood pressure (Fig. 4).

We standardized all estimated covariance terms using respective variance estimates to derive correlation estimates (Figs. 3 and 4 far right), noting that all significant estimates were positive and small to moderate in size, ranging from 0.14 (standing height in Fig. 3) to 0.58 (heel bone mineral density in Fig. 4).

In a subsequent sensitivity analysis, all significant covariance terms emerging from the genomic partitioning and genome-transcriptome analyses remained after applying a rank-based inverse normal transformation to phenotypic observations (see *p*-values in Supplementary Figs. 7 and 8). Thus, the estimated covariance terms were robust against the violation of the normality assumption held by GREML and CORE GREML. In another sensitivity analysis, we checked if the estimated genome-transcriptome covariance found for height, sitting height, heel bone mineral density, and diastolic blood pressure is robust to SNP selection, by comparing our CORE GREML results, where the genetic effects are based on ~1.1 million HapMap phase III SNPs, with results where the genetic effects are based on ~1.3 million HapMap phase II SNPs (Supplementary Table 7). The significant genome-transcriptome covariance remains for height, sitting height, and diastolic blood pressure.

To show that covariance terms estimated by CORE GREML are genuine biological parameters, we validated the covariance between the random effects of the genome and the imputed transcriptome using the same five-fold cross-validation procedure as before (i.e., for the validation of transcriptomic effects on phenotypes). In this instance, the phenotypic prediction accuracy of CORE GREML was compared against that of GREML. We chose genome-transcriptome analyses for validation since significant covariance emerged from four traits in contrast to two traits in genomic partitioning analyses. Figure 2b shows that the gain in the phenotypic prediction accuracy of CORE GREML relative to that of GREML grew as the magnitude of covariance estimates increased ($p = 1.61e - 05$).

To show the impact of neglecting significant covariance terms, we compared variance-component estimates from GREML with those from CORE GREML. In both genomic partitioning and genome-transcriptome analyses, variance estimates by GREML for correlated random effects were larger than those by CORE GREML (Figs. 5 and 6a for significant results; see Supplementary Figs. 7 and 8 for full results), noting that the differences in estimates between the two methods were proportional to the magnitude of covariance estimates. This is in line with our simulation results under positive covariance settings ($\text{cov} > 0$ in Supplementary Figs. 2 and 3). As expected, neglecting covariance did not affect variance estimates for uncorrelated random effects. For example, variance estimates of the random effects of other genomic regions by CORE GREML for standing and sitting height agreed with those by GREML (Fig. 6a). Similarly, for traits without any significant covariance term, there were minimal differences between GREML and CORE GREML estimates (Supplementary Figs. 7 and 8), which aligned with simulation results under settings of zero covariance ($\text{cov} = 0$ in Supplementary Figs. 2 and 3). Based on these observations, the larger variance estimates by GREML for correlated random effects compared to CORE GREML estimates are most likely due to bias from neglecting the correlations between these random effects.

We also considered the impact of neglecting covariance between random effects on functions of variance-component estimates, including the following: (1) proportions of phenotypic variance explained by the genome (i.e., narrow-sense heritability) and by the imputed transcriptome (Supplementary Fig. 9); (2) heritability partitioned by functional genomic region (Supplementary Fig. 10); and (3) proportions of genetic variance by functional genomic region (Fig. 6b for significant results; Supplementary Fig. 11 for full results). These functions are useful

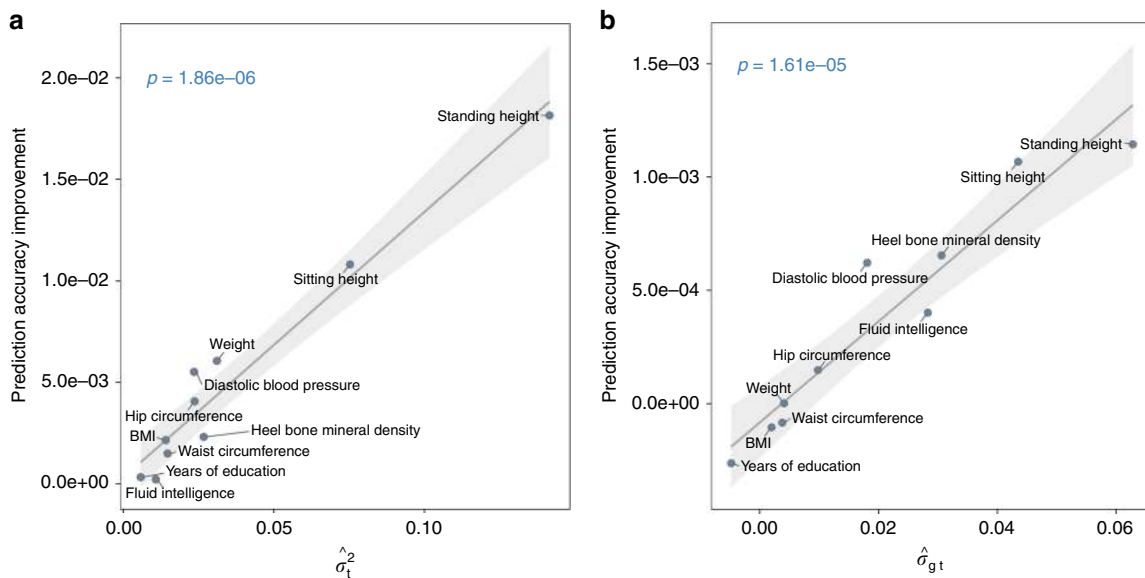


Fig. 2 Summary of five-fold cross-validation results. Phenotypic prediction accuracy improves when accounting for the effects of the imputed transcriptome and the covariance between the effects of the genome and the imputed transcriptome. **a** The more the imputed transcriptome contributes to phenotypic variance, the greater the gain in phenotypic prediction accuracy. $\hat{\sigma}_t^2$ denotes the phenotypic variance explained by the imputed transcriptome. Prediction accuracy was computed using the Pearson’s correlation coefficient between the observed and the predicted for models with and without the imputed transcriptome component, denoted as $\mathbf{y} = \mathbf{g} + \mathbf{t} + \boldsymbol{\epsilon}$ and $\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon}$, respectively. \mathbf{g} = the random effects of the genome; \mathbf{t} = the random effects of the imputed transcriptome; $\boldsymbol{\epsilon}$ = residuals. Prediction accuracy improvement (i.e., y-axis) was derived by subtracting the prediction accuracy of the model $\mathbf{y} = \mathbf{g} + \boldsymbol{\epsilon}$ from that of the model $\mathbf{y} = \mathbf{g} + \mathbf{t} + \boldsymbol{\epsilon}$. The least squares line is based on a linear model that regressed prediction accuracy improvement on phenotypic variance explained by the imputed transcriptome. The shaded area represents 95% confidence band based on s.e.m. A two-sided *t*-test (*df* = 8) was used to determine if the slope of the regression line differs from zero (see *p*-value on the left corner). **b** The larger the estimated covariance between the random effects of the genome and those of the imputed transcriptome (i.e., σ_{gt}), the greater the gain in phenotypic prediction accuracy. Prediction accuracy was computed using the Pearson’s correlation coefficient between the observed and the predicted for two models of the form $\mathbf{y} = \mathbf{g} + \mathbf{t} + \boldsymbol{\epsilon}$, but one assuming $\sigma_{gt} = 0$, i.e., GREML, and the other setting σ_{gt} as a free parameter for estimation, i.e., CORE GREML. Prediction accuracy improvement was derived by subtracting the prediction accuracy of GREML from that of CORE GREML. The least squares line is based on a linear model that regressed prediction accuracy improvement on σ_{gt} estimates. The shaded area represents 95% confidence band based on s.e.m. A two-sided *t*-test (*df* = 8) was used to determine if the slope of the regression line differs from zero (see *p*-value on the left corner). Source data are provided as a Source Data file.

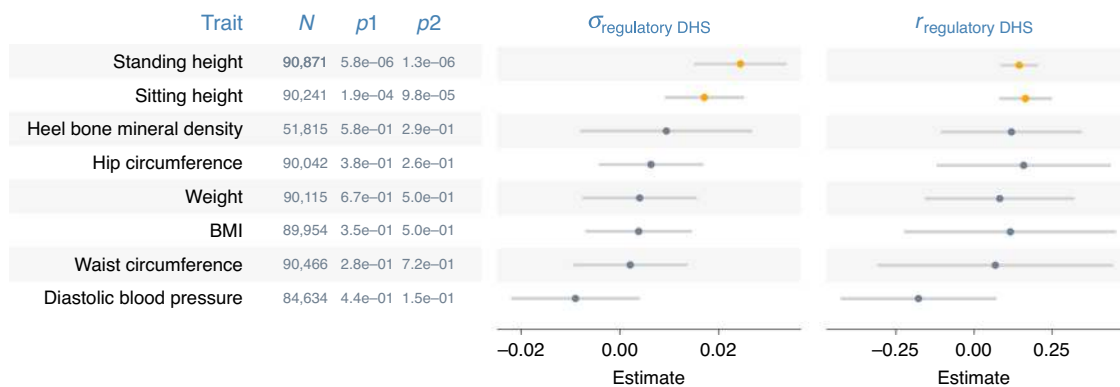


Fig. 3 Covariance estimates from genomic partitioning analyses. Shown are estimated covariance ($\sigma_{\text{regulatory DHS}}$) and correlations ($r_{\text{regulatory DHS}}$) between genetic effects attributable to the regulatory regions and those attributable to the DNase I hypersensitivity sites (DHS) on phenotypes. Error bars are 95% confidence intervals (based on s.e.m.). *N* = sample size; *p*₁ = *p*-values from likelihood ratio tests that compare GREML with CORE GREML to detect $\sigma_{\text{regulatory DHS}}$; *p*₂ = *p*-values based on the Wald’s test statistic under the null hypothesis that $r_{\text{regulatory DHS}} = 0$ (i.e., a two-sided test). Highlighted in orange are significant $\sigma_{\text{regulatory DHS}}$ and $r_{\text{regulatory DHS}}$ after a Bonferroni adjustment for multiple comparisons. Fluid intelligence and years of education are excluded because either the random effects of the regulatory regions or those of the DHS on their phenotypes were not significant. Source data are provided as a Source Data file.

for inferring the omic architecture of complex traits. For example, the relative phenotypic contributions of the genome and the imputed transcriptome can be inferred from the proportions of phenotypic variance explained by the two omes (note: these

estimates are essentially identical to variance-component estimates, because the phenotypes of traits have been standardized prior to analyses). The functional significance of SNPs from a given genomic region can be tested by assessing if the proportion

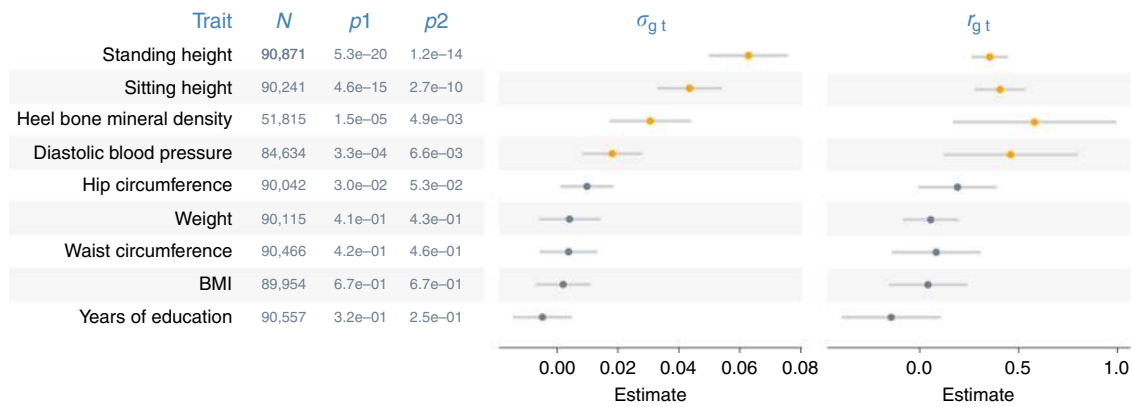


Fig. 4 Covariance estimates from genome-transcriptome partitioning analyses. Shown are estimated covariances (σ_{gt}) and correlations (r_{gt}) between the random effects of the genome and those of the imputed transcriptome. Error bars are 95% confidence intervals (based on s.e.m.). *N* = sample size; *p*₁ = *p*-values from likelihood ratio tests that compare GREML with CORE GREML to detect σ_{gt} ; *p*₂ = *p*-values based on the Wald’s test statistic under the null hypothesis that $r_{gt} = 0$ (i.e., a two-sided test). Highlighted in orange are significant σ_{gt} and r_{gt} after a Bonferroni adjustment for multiple comparisons. Fluid intelligence is excluded because the effects of the imputed transcriptome on this trait was not significant after Bonferroni correction. Source data are provided as a Source Data file.

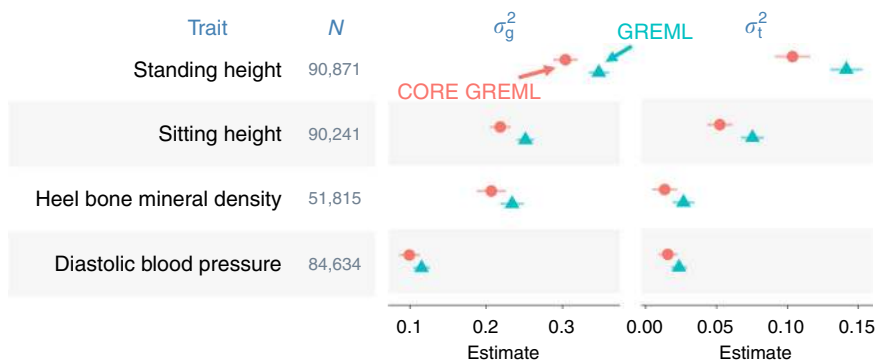


Fig. 5 Variance estimates from genome-transcriptome partitioning analyses. *N* = sample size; σ_g^2 and σ_t^2 denote the phenotypic variances explained by the genome and by the imputed transcriptome, respectively. Model parameters were estimated using the traditional method, i.e., GREML, and the proposed method, i.e., CORE GREML. Error bars are 95% confidence intervals (based on s.e.m.). Residual variance estimates are omitted for simplicity. Source data are provided as a Source Data file.

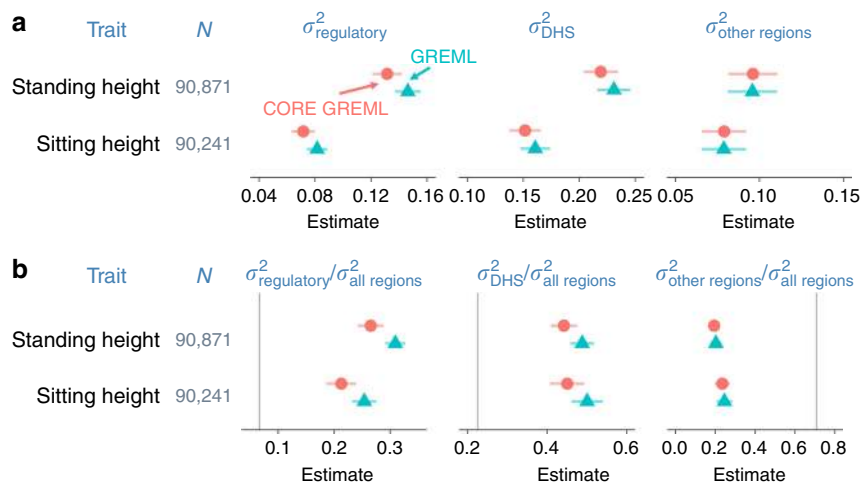


Fig. 6 Variance estimates from genomic partitioning analyses. Three functional regions of the genome are regulatory regions, DNase I hypersensitivity sites (DHS), and all other regions. *N* = sample size; $\sigma_{regulatory}^2$, σ_{DHS}^2 , and $\sigma_{other\ regions}^2$ denote phenotypic variances explained by the three functional regions. Model parameters were estimated using the traditional method, i.e., GREML, and the proposed method, i.e., CORE GREML. **a** Variance-component estimates. **b** Estimated proportions of total genetic variance attributable to three functional regions of the genome. Error bars are 95% confidence intervals (based on s.e.m.). Vertical lines in **b** are percentages of SNPs from the three functional regions; conceptually, they are expected proportions of total genetic variance explained by the three functional regions of the genome assuming all genome-wide SNPs have an equal contribution to phenotypic variation. Residual variance estimates are omitted for simplicity. Source data are provided as a Source Data file.

of genetic variance attributable to the region is substantially higher than the proportion of SNPs from the region⁵. Notably, GREML estimates of these functions were larger than CORE GREML estimates whenever there was a significant covariance term (Supplementary Figs. 9–11), but the two methods agreed with each other otherwise. For example, the relative phenotypic contributions by the genome and the imputed transcriptome inferred from GREML were larger than those inferred from CORE GREML for heel bone mineral density, diastolic blood pressure, sitting height and standing height (Fig. 5). Similarly, the functional significance of SNPs from regulatory regions inferred by GREML was larger than that by CORE GREML for sitting height and standing height (Fig. 6b). Taken together, our results indicate that GREML can lead to incorrect inferences on the underlying architecture of complex traits unless correlations between random effects are properly modeled.

It is important to note that kernel matrices used for variance-components estimation in an LMM can be similar, for instance, due to linkage disequilibrium (LD) in a genomic partitioning analysis. In fact, the correlations between off-diagonal entries of kernel matrices used in our analyses are moderate to high (0.35–0.98; see Supplementary Table 8). This similarity might give rise to the covariance between random effects. However, this possibility is unlikely for at least two reasons. First, if covariance is driven by the similarity between kernel matrices, then we would expect that in the null setting of our simulations, type I error rate is inflated, given the high similarities between kernel matrices in our analyses. Contrary to this, we found that type I error rate is controlled. Second, the kernel matrix constructed using genotypes from DHS is more similar to the kernel matrix constructed using genotypes from other regions than to the one for regulatory regions; but significant covariance was only detected between effects of regulatory regions and those of DHS on standing height and sitting height. Therefore, covariance between random effects is unlikely driven by the similarity between kernel matrices for variance-components estimation.

Discussion

When applying the classic LMMs for standard heritability estimation, where phenotypic variance is only partitioned into genetic and residual variances, the model assumption of negligible covariance between random effects (i.e., genetic and residual effects) may be met in many cases. However, when phenotypic variance is further partitioned, e.g., by functional genomic region or omic layer, using a model with multiple random effects, the covariance terms between these random effects can be substantial, as we demonstrated using the genomic partitioning analyses and the genome-transcriptome partitioning analyses for complex traits. Unless these non-negligible covariance terms are properly accounted for, variance-components estimation would be biased, resulting in misleading inferences on the latent omic architecture of complex traits, as shown by simulation results. Therefore, we recommend that covariance terms between random effects need to be carefully checked and properly modeled for genomic analyses of complex traits.

CORE GREML can serve as a useful tool for detecting and estimating covariance terms between random effects, as demonstrated using analyses of both simulated and real data. Prior to the proposal of CORE GREML, there have been several attempts to relax the assumption of independence between random effects^{20,21}, but they are specific to experimental studies and are not readily applicable to genome-wide analyses for human complex traits. To our knowledge, CORE GREML is the first of its kind in variance partitioning analyses, which correctly models the covariance between random effects.

We demonstrated the use of CORE GREML in genomic partitioning analyses and genome-transcriptome partitioning of phenotypic variance, and found that significant covariance terms mostly emerged from the latter (for four traits out of ten). Although genome-transcriptome associations have been reported by numerous studies^{12–15}, they were based on a limited number of SNPs and genes. In contrast, the association estimates by CORE GREML were based on aggregated effects of genome-wide SNPs and those of all available gene expression levels jointly on phenotypes, thereby providing an overall picture of the proportion of phenotypic variance shared by the whole genome and the transcriptome.

Of note, our study used imputed, as opposed to measured, gene expression. As demonstrated by our cross-validation results, the imputed transcriptome already improves phenotypic prediction accuracy, hence it allows phenotype forecasting, e.g., for newborns, solely on the basis of genotype information. It is noted though, the gene expression imputation models used in our study on average explains 13.7% of gene expression variation¹⁸. As a result, the imputed transcriptome would only have captured part of transcriptomic effects on phenotypes; hence, the phenotypic variance explained by the transcriptome would have been underestimated. However, our intention of using the imputed transcriptome is not to accurately estimate variance explained by the transcriptome; but as a proxy of the transcriptome to demonstrate the use of CORE GREML for genome-transcriptome partitioning analyses. When actual gene expression levels become available for future analyses, we expect an additional gain in explained phenotypic variance.

Although the model with the random effects of the imputed transcriptome fit the phenotypic data much better than the reduced model without the imputed transcriptomic effects, the two models explained the same amount of total phenotypic variance (Fig. 1). Notably, this result aligns with a recent notion of total genetic effects on complex traits, which is partitioned into genetic effects mediated by gene expression and ones not²². Although the former is essentially the effects of imputed gene expression on phenotypes, the latter is the effects of common SNPs (see Eq. (3) in Yao et al.²²). Despite the conceptual similarity, our study is different from Yao et al.²² in two key aspects. First, implied from the model in Yao et al.²², gene expression levels were based on SNPs at *cis*-eQTLs only; however, in our study gene expression levels were computed using genome-wide SNPs. This may explain that the estimated phenotypic variance due to imputed gene expression in their study tends to be smaller than that in ours for BMI, standing height, heel bone mineral density and years of education. Second, unlike CORE GREML used in our study, the model used by Yao et al.²² does not account for covariance between effects of the genome and the imputed transcriptome on phenotypes.

Importantly, the proposed CORE GREML can be used to analyze and dissect the shared effects among omic layers, beyond the genome and the transcriptome, including proteome, metabolome and exposome, when multi-omics data become available. We anticipate, based on our genome-transcriptome analyses of complex traits, that covariance between the random effects of omic layers is a key parameter, such that CORE GREML will be an important tool for multi-omics partitioning analyses. Other potential applications of CORE GREML include phenotypic variance partitioning by chromosome²³ or MAF bin^{24,25}, where correlations between the random effects in the model are intuitive. Even for the simplest partitioning of phenotypic variance that separates genetic variance apart from residual variance, CORE GREML can be useful, if genetic effects and residuals are correlated due to confounding, associations or interactions between genetic and environmental effects²⁶. One may also be

interested to partition genome-wide SNPs into subsets by LD structure and MAF bin and specify a separate random-effect term for each subset in CORE GREML. However, this would inevitably result in many random-effect terms in the model. Given that CORE GREML is computationally intensive, estimation would eventually become infeasible for models with a large number of random effects (e.g., the baseline model in Finucane et al.⁸ with >100 random-effects terms). Further studies are required to develop computationally efficient algorithms for CORE GREML, e.g., using summary statistics. In addition, we only validated CORE GREML for quantitative traits in this study. Validation of our method for binary traits is required in future studies.

Finally, we showed, using simulations, that misspecification of LD and MAF dependent genetic architecture can cause substantial bias in variance-components estimation by CORE GREML, although covariance estimation seems robust. However, the likelihood of the true estimation model in general is much greater than a wrong model, suggesting that likelihood-based comparisons of models that assume different genetic architectures is a useful way to reduce the chance of mis-specifying genetic architecture. For demonstration, we fitted the GCTA model and the LDAK model with the recommended default setting¹⁹ and found the GCTA model in general had a better fit than the LDAK model for our traits. In the absence of the knowledge of the true genetic architecture, it is recommended to vary parameter settings of the LDAK model more systematically (as in Speed et al.¹⁹) and choose the best fitting model via likelihood comparison before applying CORE GREML.

In this study, we introduce a generalized GREML, referred to as CORE GREML, which relaxes the assumption of independence between random effects held by classic mixed-effects models for variance-component analyses. Using both simulations and real data, we showed that in the presence of non-negligible covariance terms, CORE GREML improved genomic partitioning and multi-omics partitioning analyses by the conventional GREML. We conclude that the covariance between random effects for analysis of complex traits is a key parameter for estimation and, hence, recommend that covariance terms should be carefully checked and properly modeled.

Methods

Ethics statement. We used data from the UK Biobank (<https://www.ukbiobank.ac.uk>) for our analyses.

The UK Biobank’s scientific protocol has been reviewed and approved by the North West Multi-center Research Ethics Committee, National Information Governance Board for Health & Social Care, and Community Health Index Advisory Group. UK Biobank has obtained informed consent from all participants. Our access to the UK Biobank data was under the reference number 14575. The research ethics approval of the current study was obtained from the University of South Australia Human Research Ethics Committee.

Generalizing GREML. A LMM can be written as

$$y = Xb + Zg + \epsilon \tag{1}$$

where y is a vector of trait phenotypes, b is a vector of fixed effects, g is a vector of additive genetic effects, and ϵ is a vector of residual effects. X and Z are incidence matrices. The random effects, g and ϵ , are assumed to be normally distributed with mean zeros and variances $A\sigma_g^2$ and $I\sigma_\epsilon^2$, respectively, where A and I are the genetic relationship kernel matrix^{2,27,28} and an identity matrix, respectively. The variance-covariance matrix of all observations, $\text{var}(y)$, can be written as

$$\text{var}(y) = A\sigma_g^2 + I\sigma_\epsilon^2 \tag{2}$$

This is the standard definition of variance-covariance matrix used in LMM, which assumes no correlation between g and ϵ , i.e., $\text{cor}(g, \epsilon) = 0$. When relaxing this classic assumption, Eq. (2) can be expressed as

$$\text{var}(y) = A\sigma_g^2 + I\sigma_\epsilon^2 + \left[\sqrt{A} \cdot \sqrt{I} + \left(\sqrt{A} \cdot \sqrt{I} \right)' \right] \sigma_{g,\epsilon} \tag{3}$$

where \sqrt{A} is the Cholesky decomposition of the genetic relationship kernel matrix

with $A = \sqrt{A} \cdot \sqrt{A}$, $\sqrt{I} = I$, and $\sigma_{g,\epsilon}$ is the covariance between g and ϵ , i.e., $\sigma_{g,\epsilon} = \text{cor}(g, \epsilon) \cdot \sqrt{\sigma_g^2 \cdot \sigma_\epsilon^2}$.

When considering multiple random effects in the LMM (e.g., genomic partitioning approach), the model can be written as

$$y = Xb + \sum_{i=1}^k Zg_i + \epsilon \tag{4}$$

where g_i is the random genetic effects of the i th pre-defined functional category, e.g., regulatory regions.

Such LMMs with multiple random effects typically assume no correlation between g_i and g_j . However, this assumption can be violated if the effects of two categories on phenotypes are associated, e.g., through the same gene pathway. We relax this assumption and write the variance-covariance matrix of all observations, $\text{var}(y)$, as

$$V = \sum_{i=1}^k A_i \sigma_{g_i}^2 + \sum_{i=1}^k \sum_{j=1}^{i-1} \left[\sqrt{A_i} \cdot \sqrt{A_j} + \left(\sqrt{A_i} \cdot \sqrt{A_j} \right)' \right] \cdot \sigma_{g_i g_j} + I\sigma_\epsilon^2 \tag{5}$$

where A_i is the genetic relationship kernel matrix constructed using SNPs from the i th functional category, $\sqrt{A_i}$ is the Cholesky decomposition of A_i , and $\sigma_{g_i g_j}$ is the genetic covariance between g_i and g_j . It is noted that the correlation term between genetic effects (g_i) and residuals (ϵ) is not included and hence assumed to be zero in Eq. (4), which is usually valid, although it is possible to parameterize this term.

The log likelihood of the proposed model, which can be generally applied to Eqs. (1) and (4), is

$$\log L = -\frac{1}{2} [\ln|V| + \ln|X'V^{-1}X| + y'Py] \tag{6}$$

where \ln is the natural log and $||$ the determinant of the associated matrices. The projection matrix is defined as $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$. By maximizing the log likelihood, the direct average information algorithm^{29,30} can be used to obtain CORE GREML estimates of parameters including the covariance terms between random effects.

This CORE GREML approach can be easily extended to phenotypic variance partitioning using multi-omics data, e.g. genome-transcriptome analyses (see Genome-Transcriptome Partitioning Model section below).

Heritability. For Eq. (1), the standard definition of heritability is

$$h^2 = \frac{\sigma_g^2}{\sigma_y^2} \tag{7}$$

where the phenotypic variance is $\sigma_y^2 = \sigma_g^2 + \sigma_\epsilon^2$ in the absence of $\text{cor}(g, \epsilon)$. When there is non-negligible $\text{cor}(g, \epsilon)$, the phenotypic variance should be written as $\sigma_y^2 = \sigma_g^2 + \sigma_\epsilon^2 + 2 \cdot \text{cov}(g, \epsilon)$.

For Eq. (4), a general expression of heritability for the i th genetic component is

$$h^2 = \frac{\sigma_{g_i}^2}{\sigma_y^2} \tag{8}$$

where $\sigma_y^2 = \sum_{i=1}^k \sigma_{g_i}^2 + \sum_{i=1}^k \sum_{j=1}^{i-1} 2 \cdot \sigma_{g_i g_j} + \sigma_\epsilon^2$.

Using the Delta method³¹, the sampling variance of heritability for this example can be obtained as

$$\text{var}(h^2) = \left(\frac{\sigma_{g_i}^2}{\sigma_y^2} \right)^2 \left(\frac{\text{var}(\sigma_{g_i}^2)}{\sigma_{g_i}^4} - \frac{2 \cdot \text{cov}(\sigma_{g_i}^2, \sigma_y^2)}{\sigma_{g_i}^2 \cdot \sigma_y^2} + \frac{\text{var}(\sigma_y^2)}{\sigma_y^4} \right) \tag{9}$$

where $\text{var}(\sigma_{g_i}^2)$, $\text{var}(\sigma_y^2)$, and $\text{cov}(\sigma_{g_i}^2, \sigma_y^2)$ can be obtained from the average information matrix^{29,32} of CORE GREML.

Correlation between two random effects. The correlation between two random (genetic) effects ($r_{g_i g_j}$) can be defined as the genetic covariance scaled by the square root of the product of the genetic variances of the two random effects, i.e.,

$$r_{g_i g_j} = \frac{\sigma_{g_i g_j}}{\sqrt{\sigma_{g_i}^2 \cdot \sigma_{g_j}^2}} \tag{10}$$

Using the Delta method³¹, the sampling variance of genetic correlation can be obtained as

$$\text{var}(r_{g_i g_j}) = r_{g_i g_j}^2 \left(\frac{\text{var}(\sigma_{g_i}^2)}{4 \cdot \sigma_{g_i}^4} + \frac{\text{var}(\sigma_{g_j}^2)}{4 \cdot \sigma_{g_j}^4} + \frac{\text{var}(\sigma_{g_i g_j})}{\sigma_{g_i g_j}^2} + \frac{2 \cdot \text{cov}(\sigma_{g_i}^2, \sigma_{g_j}^2)}{4 \cdot \sigma_{g_i}^2 \cdot \sigma_{g_j}^2} - \frac{2 \cdot \text{cov}(\sigma_{g_i}^2, \sigma_{g_i g_j})}{2 \cdot \sigma_{g_i}^2 \cdot \sigma_{g_i g_j}} - \frac{2 \cdot \text{cov}(\sigma_{g_j}^2, \sigma_{g_i g_j})}{2 \cdot \sigma_{g_j}^2 \cdot \sigma_{g_i g_j}} \right) \tag{11}$$

where the variance and covariance terms used are from the information matrix of CORE GREML.

Computational requirements. CORE GREML analyses were performed using open-access software, MTG2 (version 2.14 or later versions; <https://sites.google.com/site/honglee0707/mtg2>). The computational requirements for fitting a model with two random effects using CORE GREML are outlined in Supplementary Table 9.

Genotype data. The UK Biobank (project approval number 14575) contains health-related data from ~500,000 participants aged between 40 and 69 years, who were recruited throughout the UK between 2006 and 2010³³. Prior to data analysis, we applied stringent quality control to exclude unreliable genotypic data. We filtered SNPs with an INFO score (used to indicate the quality of genotype imputation) < 0.6, an MAF < 0.01, a Hardy–Weinberg equilibrium *p*-value < 1e−4, or a call rate < 0.95. We then selected HapMap phase III SNPs, which are known to yield reliable and robust estimates of SNP-based heritability^{25,34,35}, for downstream analyses. We filtered individuals who had a genotype-missing rate > 0.05, were non-white British ancestry, or had the first or second ancestry principal components outside 6 SDs of the population mean. We also applied relatedness cut-off quality control to exclude one of any pair of individuals with a genomic relationship > 0.025. From the remaining individuals, we selected those who were included in both the first and second release of UK Biobank genotype data. We calculated the discordance rate of imputed genotypes between the two versions and excluded individuals with a discordance rate > 0.05. Eventually, genotypes of 1,131,002 SNPs from 91,472 individuals remained for data analysis.

Phenotype data. To preclude negligible heritability as a possibility for negative findings (i.e., no covariance between random effects), we deliberately chose ten UK Biobank traits available to us with the largest heritability estimates by an independent open source (https://nealelab.github.io/UKBB_ldsc/), which included standing height, sitting height, BMI, heel bone mineral density, fluid intelligence, weight, waist circumference, hip circumference, diastolic blood pressure, and years of education³⁶. Heritability estimates for all selected traits were at least 20 times greater than their SEs, to ensure that they were significantly different from zero. We further verified SNP-based heritability of these traits using GREML and estimates are shown in Supplementary Fig. 1.

Prior to model fitting, phenotypic data were prepared using R (v3.4.3) in three sequential steps as follows: (1) adjustment for age, sex, birth year, social economic status (by Townsend Deprivation Index), population structure (by the first ten principal components of the genomic relationship matrix estimated using PLINK v1.9), assessment center, and genotype batch using linear regression; (2) standardization; and (3) removal of data points outside ±3 SDs from the mean. The distributions of phenotypes of the ten traits are shown in Supplementary Fig. 12. We noted mild to strong deviations from normality for traits such as BMI and years of education. This motivated a subsequent sensitivity analysis to test the robustness of our findings against the violation of the normality assumption held by GREML and CORE GREML. Specifically, we applied a rank-based inverse normal transformation to phenotypes of all traits and repeated our analyses on the transformed phenotypes.

Functional annotation of the genome. The genome was annotated using three pre-defined functional categories (<http://gusevlab.org/software/>) as follows: (1) regulatory regions that consist of coding regions, untranslated regions and promoters; (2) DHSs; and (3) all other regions. We assigned each SNP into 1 of the 3 categories, resulting in 75,396 SNPs in the regulatory regions, 255,665 in the DHS, and 799,935 in all other regions. Prior to the assignment, genotype data were quality-controlled (see above for details).

Gene expression imputation. Using PrediXcan¹⁸ (<https://github.com/hakyimlab/PrediXcan>), we imputed expression levels of 2028 to 9630 genes for 43 non-sex-specific tissues (Supplementary Table 2) by projecting estimated SNP effects of expression onto genotypes of 1,316,391 SNPs for 91,472 individuals from the UK Biobank. Selected SNPs had an INFO score > 0.6, an MAF > 0.01, a *p*-value for the Hardy–Weinberg test > 0.0001, and missingness < 0.05. SNP effect estimates were sourced from GTEx v7 models (2018-01-08 release; <http://predictdb.org>), which were trained using 2,496,846 SNPs of European individuals from the Genotype-Tissue Expression project³⁷.

Variance partitioning models. The phenotypic variance of each selected trait was partitioned using two separate random-effects models (see below for model description). The “genome-transcriptome” model partitions phenotypic variance into variation from the genome, the transcriptome and unknown sources (i.e., residual variance), whereas the “genomic partitioning” model assumes phenotypic variation comes from the genome and residuals, and further partitions genetic variance by functional category of the genome. Three functional categories were under consideration, namely, regulatory regions (encompassing coding regions, untranslated regions, and promoters), DHSs, and all other regions.

Each partitioning model was fitted using the conventional method, i.e., GREML, and the proposed alternative, i.e., CORE GREML. Essentially, GREML sets all covariance terms between random effects to zero, whereas CORE GREML treats these terms as free parameters for estimation. To detect significant covariance terms, we performed likelihood ratio tests to determine whether the model fit by CORE GREML was better than that by GREML.

Assuming that the phenotypes are pre-adjusted for fixed effects, the genome-transcriptome partitioning model can be expressed as

$$y = \mu + g + t + \epsilon \tag{12}$$

where y is a $n \times 1$ vector of phenotype data, μ is the grand mean, g , t , and ϵ are the main genetic, transcriptomic, and residual effects, following $g \sim N(0, A_{n \times n} \sigma_g^2)$, $t \sim N(0, T_{n \times n} \sigma_t^2)$, and $\epsilon \sim N(0, I_{n \times n} \sigma_\epsilon^2)$, respectively. The terms, σ_g^2 , σ_t^2 , and σ_ϵ^2 denote phenotypic variances attributable to the genome, the transcriptome, and residuals, respectively. $A_{n \times n}$ and $T_{n \times n}$ are relationship kernel matrices and $I_{n \times n}$ is an identity matrix. $A_{n \times n}$ is derived by $W_{n \times m} W'_{n \times m} / m$ and $T_{n \times n}$ by $Q_{n \times p} Q'_{n \times p} / p$, where $W_{n \times m}$ contains standardized genotype information of m ($=1,131,002$) SNPs for n ($=91,472$) individuals, and $Q_{n \times p}$ contains standardized imputed expression of p ($=227,664$) genes collapsed across 43 tissues for the n individuals. Essentially, entries of $A_{n \times n}$ and $T_{n \times n}$ describe pairwise similarities between individuals based on their genotypes and imputed gene expression, respectively.

The variance–covariance matrix of phenotypic observations is

$$\text{var}(y) = A_{n \times n} \sigma_g^2 + T_{n \times n} \sigma_t^2 + \left[\sqrt{A_{n \times n}} \cdot \sqrt{T_{n \times n}} + \left(\sqrt{A_{n \times n}} \cdot \sqrt{T_{n \times n}} \right)' \right] \cdot \sigma_{gt} + I_{n \times n} \sigma_\epsilon^2 \tag{13}$$

where σ_{gt} is the covariance between the effects of the genome and the transcriptome on phenotypes. Here we assume no correlation between residuals and genomic or transcriptomic effects.

The genomic partitioning model can be expressed as

$$y = \mu + g_{\text{regulatory}} + g_{\text{DHS}} + g_{\text{other}} + \epsilon \tag{14}$$

where y is a $n \times 1$ vector of phenotype data that are decomposed into the grand mean μ , the genetic effects due to regulatory, $g_{\text{regulatory}} \sim N(0, A_{\text{regulatory}} \sigma_{\text{regulatory}}^2)$, DHS, $g_{\text{DHS}} \sim N(0, A_{\text{DHS}} \sigma_{\text{DHS}}^2)$ and other genomic regions $g_{\text{other}} \sim N(0, A_{\text{other}} \sigma_{\text{other}}^2)$, and residuals, $\epsilon \sim N(0, I_{n \times n} \sigma_\epsilon^2)$. The terms $\sigma_{\text{regulatory}}^2$, σ_{DHS}^2 , σ_{other}^2 , and σ_ϵ^2 denote phenotypic variances attributable to the three functional regions and residuals, respectively. The kernel matrices $A_{\text{regulatory}}$, A_{DHS} , and A_{other} were constructed using 75,396 SNPs from regulatory regions, 255,665 from DHS, and 799,935 from all other genomic regions. I is a $n \times n$ identity matrix.

The variance–covariance matrix of phenotypic observations is

$$\begin{aligned} \text{var}(y) = & A_{\text{regulatory}} \sigma_{\text{regulatory}}^2 + A_{\text{DHS}} \sigma_{\text{DHS}}^2 + A_{\text{other}} \sigma_{\text{other}}^2 \\ & + \left[\sqrt{A_{\text{regulatory}}} \cdot \sqrt{A_{\text{DHS}}} + \left(\sqrt{A_{\text{regulatory}}} \cdot \sqrt{A_{\text{DHS}}} \right)' \right] \sigma_{\text{regulatory DHS}} \\ & + \left[\sqrt{A_{\text{regulatory}}} \cdot \sqrt{A_{\text{other}}} + \left(\sqrt{A_{\text{regulatory}}} \cdot \sqrt{A_{\text{other}}} \right)' \right] \sigma_{\text{regulatory other}} \\ & + \left[\sqrt{A_{\text{DHS}}} \cdot \sqrt{A_{\text{other}}} + \left(\sqrt{A_{\text{DHS}}} \cdot \sqrt{A_{\text{other}}} \right)' \right] \sigma_{\text{DHS other}} + I \sigma_\epsilon^2 \end{aligned} \tag{15}$$

where σ_{ij} is the covariance between genetic effects of functional regions i and j , for i and $j \in \{\text{regulatory regions, DHS, other regions}\}$ and $i \neq j$. We assume no correlation between residuals and any of the genetic effects.

Simulation. To validate CORE GREML, we simulated 500 replicates of phenotypic data using the two variance partitioning models shown above under each of three parameter settings: zero (i.e., null setting), positive and negative covariance between random effects in the variance partitioning model. Simulations were based on quality-controlled genotype data and imputed transcriptome data from a random sample of 10,000 UK Biobank individuals. The genotype data contained a total of 1,131,002 SNPs (see Genotype data above) and the imputed transcriptome contained imputed expressions of 227,664 genes collapsed across 43 non-sex-specific tissues (see Gene expression imputation above).

For genome-transcriptome analysis, phenotypes were simulated using Eq. (12) according to the following variance-covariance structure of random effects:

$$\text{var} \begin{pmatrix} g \\ t \\ \epsilon \end{pmatrix} = \begin{pmatrix} \sigma_g^2 & \sigma_{gt} & 0 \\ \sigma_{gt} & \sigma_t^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{pmatrix} = \begin{pmatrix} 0.4 & \sigma_{gt} & 0 \\ \sigma_{gt} & 0.4 & 0 \\ 0 & 0 & 0.2 \end{pmatrix} \tag{16}$$

where the value of σ_{gt} varied across parameter settings (Supplementary Table 1).

For genomic partitioning analysis, phenotypes were simulated using Eq. (14) according to the following variance–covariance structure of random effects:

$$\text{var} \begin{pmatrix} \mathbf{g}_{\text{regulatory}} \\ \mathbf{g}_{\text{DHS}} \\ \mathbf{g}_{\text{other}} \\ \boldsymbol{\varepsilon} \end{pmatrix} = \begin{pmatrix} \sigma_{\text{regulatory}}^2 & \sigma_{\text{regulatory DHS}} & \sigma_{\text{regulatory other}} & 0 \\ \sigma_{\text{regulatory DHS}} & \sigma_{\text{DHS}}^2 & \sigma_{\text{DHS other}} & 0 \\ \sigma_{\text{regulatory other}} & \sigma_{\text{DHS other}} & \sigma_{\text{other}}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon}^2 \end{pmatrix} \quad (17)$$

$$= \begin{pmatrix} 0.3 & \sigma_{\text{regulatory DHS}} & \sigma_{\text{regulatory other}} & 0 \\ \sigma_{\text{regulatory DHS}} & 0.4 & \sigma_{\text{DHS other}} & 0 \\ \sigma_{\text{regulatory other}} & \sigma_{\text{DHS other}} & 0.1 & 0 \\ 0 & 0 & 0 & 0.2 \end{pmatrix}$$

where the values of covariance terms varied across parameter settings (Supplementary Table 1).

For each replicate, we fitted the two variance partitioning models using both GREML and CORE GREML as for analysis of real data. Under the null setting, we assessed if CORE GREML can detect covariance term(s) at a controlled rate of type I errors, by comparing the model fit of CORE GREML with that of GREML using likelihood ratio tests. Under all settings, we assessed if CORE GREML can produce unbiased estimates of variance and covariance components. To show the impact of neglecting genuine covariance terms, we also provided parameter estimates by GREML.

To test the sensitivity of CORE GREML estimation to a wrong assumption about the genetic architecture in the estimation, we also simulated phenotypes under genetic architecture. Following past studies^{19,38}, we parameterized genetic architecture by MAF and LD, assuming the variance of SNP-specific effects on phenotypes, $\text{var}(\beta_i)$, for any given SNP i , is proportional to its LD score, w , and MAF, f , expressed as $\text{var}(\beta_i) \propto w_i^\gamma [f_i(1-f_i)]^{1+\alpha}$, where α and γ control the extents to which w and f affect $\text{var}(\beta_i)$. By altering values of α (either -1 or -0.25) and γ (either 0 or 1), we simulated phenotypic data under three different genetic architectures ($\alpha = -1, \gamma = 1$; $\alpha = -0.25, \gamma = 1$; $\alpha = -0.25, \gamma = 0$), each including three scenarios of the covariance between the random effects of the genome and those of the transcriptome, namely, null ($\sigma_{\text{gt}} = 0$), positive ($\sigma_{\text{gt}} = 0.2$), and negative ($\sigma_{\text{gt}} = -0.2$), as for the CORE GREML validation simulations (Supplementary Table 1; note the genetic architecture for the CORE GREML validation simulations is under the setting $\alpha = -1, \gamma = 0$). Each scenario had 500 replicates of simulated phenotypic data (each with $n = 10,000$). For a given genetic architecture, we fitted two estimation models, both using CORE GREML, but one assuming values of α and γ the same as those of the simulation model (i.e., “true model”) and the other always assuming $\alpha = -1, \gamma = 0$ (i.e., “wrong model”).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The genotype and phenotype data of the UK Biobank can be accessed through procedures described on its webpage (<https://www.ukbiobank.ac.uk/using-the-resource>). Simulated data used in this paper can be obtained from the authors upon request. Source data are provided with this paper.

Code availability

The source code for MTG2 and example code along with related files for fitting CORE GREML using MTG2 can be accessed without any restrictions from <https://sites.google.com/site/honglee0707/mtg2> or from <https://doi.org/10.5061/dryad.bk3j9kd8c>.

Received: 19 December 2019; Accepted: 3 August 2020;

Published online: 21 August 2020

References

- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
- Henderson, C. R., Kempthorne, O., Searle, S. R. & Von Krosigk, C. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218 (1959).
- Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
- Gusev, A. et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
- Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18**, <https://doi.org/10.1186/s13059-017-1215-1> (2017).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- Johannesson, M., Magnusson, P., Ikram, M. & Visscher, P. Equivalence of LD-score regression and individual-level-data methods. *Behav. Genet.* **47**, 642–719 (2017).
- Ni, G. et al. Estimation of genetic correlation via linkage disequilibrium score regression and genomic restricted maximum likelihood. *Am. J. Hum. Genet.* **102**, 1185–1194 (2018).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- Mercer, T. R. et al. DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* **45**, 852 (2013).
- Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).
- Shu, W., Chen, H., Bo, X. & Wang, S. Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains. *Nucleic Acids Res.* **39**, 7428–7443 (2011).
- Wang, Y.-M. et al. Correlation between DNase I hypersensitive site distribution and gene expression in HeLa S3 cells. *PLoS ONE* **7**, <https://doi.org/10.1371/journal.pone.0042414> (2012).
- Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, <https://doi.org/10.1371/journal.pgen.1004383> (2014).
- Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
- Speed, D. et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
- Dumont, C., Chenel, M. & Mentré, F. Influence of covariance between random effects in design for nonlinear mixed-effect models with an illustration in pediatric pharmacokinetics. *J. Biopharm. Stat.* **24**, 471–492 (2014).
- Frossard, J. & Renaud, O. The correlation structure of mixed effects models with crossed random effects in controlled experiments. Preprint at <https://arxiv.org/abs/1903.10766> (2019).
- Yao, D. W., O’connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat. Genet.* **52**, 626–633 (2020).
- Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
- Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).
- Lee, S. H. et al. Estimation of SNP heritability from dense genotype data. *Am. J. Hum. Genet.* **93**, 1151–1155 (2013).
- VanderWeele, T. J., Ko, Y.-A. & Mukherjee, B. Environmental confounding in gene-environment interaction studies. *Am. J. Epidemiol.* **178**, 144–152 (2013).
- Amin, N., Van Duijn, C. M. & Aulchenko, Y. S. A genomic background based method for association analysis in related individuals. *PLoS ONE* **2**, <https://doi.org/10.1371/journal.pone.0001274> (2007).
- VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
- Lee, S. H. & Van Der Werf, J. H. An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol.* **38**, 25–43 (2006).
- Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, 1998).
- Gilmour, A. R., Thompson, R. & Cullis, B. R. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 1440–1450 (1995).
- Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, <https://doi.org/10.1371/journal.pmed.1001779> (2015).
- Lee, S. H. et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).

35. Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
36. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
37. Lonsdale, J. et al. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
38. Hou, K. et al. Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nat. Genet.* **51**, 1244–1251 (2019).

Acknowledgements

This research is supported by the Australian National Health and Medical Research Council (1087889) and the Australian Research Council (DP190100766, FT160100229). The UK Biobank is funded by the UK Department of Health, the Medical Research Council, the Scottish Executive, and the Wellcome Trust medical research charity. Funding bodies had no role in the study design, the collection, analysis, and interpretation of the data, and the writing of the manuscript. We thank the staff and participants of the UK Biobank for their important contributions. Work was performed using computational resources provided by the Australian Government through Gadi under the National Computational Merit Allocation Scheme (NCMAS).

Author contributions

S.H.L. conceived the idea and directed the study. X.Z. performed analyses. H.K.I. provided critical feedback and key elements in interpreting the results. S.H.L. and X.Z. drafted the manuscript. All authors contributed to the editing and approval of the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-18085-5>.

Correspondence and requests for materials should be addressed to S.H.L.

Peer review information *Nature Communications* thanks Doug Speed and Martin Zhang for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020