Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals

Valentin Hivert[1], Julia Sidorenko[1], Florian Rohart[1], Michael E Goddard[2,3], Jian Yang[1,4],

Naomi R Wray[1,5], Loic Yengo[1,6] and Peter M Visscher[1,6,*]

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia

[2]Agriculture Victoria Research, AgriBio, 5 Ring Road, Bundoora, VIC 3083, Australia

[3]Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC 3010, Australia

[4]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[5]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia

[6]These authors contributed equally

*Correspondence: Peter M Visscher (peter.visscher@uq.edu.au)

1

Abstract

Non-additive genetic variance for complex traits is traditionally estimated from data on relatives. It is notoriously difficult to estimate without bias in non-laboratory species, including humans, because of possible confounding with environmental covariance among relatives. In principle, non-additive variance attributable to common DNA variants can be estimated from a random sample of unrelated individuals with genome-wide SNP data. Here, we jointly estimate the proportion of variance explained by additive (h_{SNP}^2), dominance (δ_{SNP}^2) and additive-by-additive (η_{SNP}^2) genetic variance in a single analysis model. We first show by simulations that our model leads to unbiased estimates and provide new theory to predict standard errors estimated using either least squares or maximum likelihood. We then apply the model to 70 complex traits using 254,679 unrelated individuals from the UK Biobank and 1.1M genotyped and imputed SNPs. We found strong evidence for additive variance (average across traits $\bar{h}_{SNP}^2 = 0.207$). In contrast, the average estimate of $\bar{\delta}_{SNP}^2$ across traits was 0.001, implying negligible dominance variance at causal variants tagged by common SNPs. The average epistatic variance $\bar{\eta}_{SNP}^2$ across the traits was 0.058, not significantly different from zero because of the large sampling variance. Our results provide new evidence that genetic variance for complex traits is predominantly additive, and that sample sizes of many millions of unrelated individuals are needed to estimate epistatic variance with sufficient precision.

Introduction

The total genetic variance of a trait can be partitioned into additive, dominance and epistatic variance components[1-5]. The role of non-additive genetic variation (dominance and epistatic) in complex traits in human populations remains elusive, because it is difficult to estimate and also because theory predicts it to be small relative to additive variance. Traditionally, non-additive genetic variance has been estimated from pedigree or twin studies, by contrasting the phenotypic covariance for different kinds of relatives. Non-additive variance leads to closer relatives being more similar than expected from additive genetic variance. However, shared environmental effects may also be expected to be strong among closer relatives, and disentangling these and other sources of familial resemblance remains challenging.

The amount of non-additive genetic variance disproportionally depends on the allele frequencies at causal variants as compared to additive variance, and in general it is lower when such frequency (and locus heterozygosity) is low. Non-additive genetic variance can be estimated in model species, where both the environment and allele frequencies can be controlled[6]. Theory and empirical evidence suggest that genetic variation in complex traits is mainly additive[7,8] and that higher-order epistatic interactions, if present, are still mostly expected to contribute to additive genetic variance[8]. For highly polygenic traits with evidence of directional dominance (also known as inbreeding depression), the expected dominance variance is also very small[9].

Variance components can also be estimated from genome wide SNP genotypes. In humans, Zhu et al.[10] extended the model of Yang et al.[11] to estimate the proportion of phenotypic variance explained by both additive (h^2_{SNP}) and dominance variance (δ^2_{SNP}) from common SNPs in a sample of unrelated individuals using restricted maximum likelihood (REML). Across 79 quantitative traits, in a sample of 6,715 individuals, they reported an average dominance variance of $\bar{\delta}^2_{SNP} = 0.03$, approximately one-fifth of their estimated average narrow sense SNP-based heritability ($\bar{h}^2_{SNP}$), with a large standard error $SE(\bar{\delta}^2_{SNP}) = 0.031$ (computed using the reported standard errors of Zhu et al.[10] and neglecting the covariance between estimates). Therefore, these results are consistent with a small to negligible contribution of dominance variance across these traits.

Epistatic variance attributable to common DNA variants (η^2_{SNP}) can, in principle, also be estimated from a random sample of unrelated individuals with genome-wide SNP data. However, pairwise coefficients for epistatic variance between unrelated individuals are very small, and *a priori* it is expected that non-additive variance contributes less than additive variance. How large sample sizes need to be to reliably detect epistatic variance is currently unknown. Therefore, having a theoretical expectation of the sampling variance of estimators of η^2_{SNP} would allow the quantification of statistical power and thus the required sample size. Moreover, theory predicts that imperfect tagging of causal variants by SNPs leads to a larger reduction in SNP-based estimates of non-additive genetic variance as compared to SNP-based estimates of additive genetic variance. For example, if a causal variant is tagged by a SNP with an linkage disequilibrium (LD) squared correlation r2, then

4

estimates of additive variance are expected to be proportional to r^2, whereas it is r^4 for dominance and additive-by-additive variance[12].

Our study focuses on additive-by-additive epistatic variance as opposed to that generated by higher-order interactions such as dominance-by-additive or dominance-by-dominance components. We derive the theoretical sampling variance of ordinary least squares using Haseman-Elston (HE) regression, and Restricted maximum likelihood (REML) estimators of η^2_{SNP} in a sample of unrelated individuals. We jointly estimate the genome-wide additive, dominance, and additive-by-additive SNP-based genetic variances of 70 human complex traits using a large sample of unrelated individuals ($N = 254{,}679$) from the UK Biobank[13] (UKB).

Results

Overview of the method

Zhu et al.[10] developed a linear mixed model (LMM) where the total genetic variance of a trait attributable to SNPs is partitioned into additive and dominance variance and where the effects of all genotyped or imputed SNPs are fitted together as random effects through the use of additive and dominance genomic relationship matrices (GRM). The different variance components of this multiple-GRM model were then estimated using REML. Here, we extend this model by partitioning the total genetic variance of a trait that is attributable

5

to SNPs into additive, dominance and additive-by-additive variance components. That can

be mathematically written as:

$$\text{var}(\mathbf{y}) = \mathbf{\Theta}_A \sigma_A^2 + \mathbf{\Theta}_D \sigma_D^2 + \mathbf{\Theta}_{AA} \sigma_{AA}^2 + \mathbb{I} \sigma_e^2 \qquad (1)$$

with $\mathbf{y}$ a vector of phenotypic values for N diploid individuals and $\mathbb{I}$ an $N \times N$ identity

matrix. The three components of genetic variance σ_A^2, σ_D^2 and σ_{AA}^2 are respectively the

additive, dominance and additive-by-additive variance explained by SNPs, and σ_e^2 is the

residual variance. $\mathbf{\Theta}_A = G$ and $\mathbf{\Theta}_D$ are the additive and dominance GRMs computed from L

SNPs as previously described in Yang et al.[11] and Zhu et al.[10] (see **Methods** for details), and

$\mathbf{\Theta}_{AA}$ is the additive-by-additive genomic relatedness matrix, defined as:

$$\mathbf{\Theta}_{AA} = \frac{G \odot G}{tr(G \odot G)/N} \qquad (2)$$

with $G \odot G$ the Hadamard product (i.e. coefficient-wise matrix product) of the additive

GRM with itself, and $tr(G \odot G)$ the trace of the matrix. The standardization by the average

of the diagonal elements ensures that the mean of the diagonal elements of $\mathbf{\Theta}_{AA}$ is

approximately one as for $\mathbf{\Theta}_A$ and $\mathbf{\Theta}_D$ (**Supplementary Table 1**), leading to estimates of

genetic variances on the same scale as the residual variance[14,15].

Under Hardy-Weinberg equilibrium, the model defined by **Equation (1)** is theoretically

orthogonal between the additive and dominance components (see **Methods** and Zhu et

al.**[10]**). By orthogonality, we mean that the estimates of one of the genetic variance

components when fitted as the only genetic component would be unbiased even in the

presence of the effects from the other components. Moreover, in a sample of unrelated

6

individuals from an outbred population, the off-diagonal elements of $\mathbf{\Theta}_A$ are expected to be

$\mathbf{\Theta}_{A_{ij}} \sim \mathcal{N}\left(0, Var(G_{ij})\right)$, leading to an expected covariance $\text{Cov}(\mathbf{\Theta}_{A_{ij}}, \mathbf{\Theta}_{AA_{ij}}) = \mathbb{E}[\mathbf{\Theta}_{A_{ij}}^3] = 0$.

The latter property guarantees orthogonality between the additive and additive-by-additive component. Orthogonality between the dominance and additive-by-additive components is more complicated to prove and was investigated empirically.

Equation (1) defines a typical LMM, which variance components (σ_A^2, σ_D^2, σ_{AA}^2 and σ_e^2) can be estimated jointly or separately using REML[16] as well as by HE regression[17-19]. The proportion of phenotypic variance explained by additive (σ_A^2), dominance (σ_D^2) and additive-by-additive (σ_{AA}^2) variance at all SNPs are defined as $h_{SNP}^2 = \sigma_A^2/(\sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_e^2)$, $\delta_{SNP}^2 = \sigma_D^2/(\sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_e^2)$, and $\eta_{SNP}^2 = \sigma_{AA}^2/(\sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_e^2)$. We define the SNP-based broad sense heritability $H_{SNP}^2 = h_{SNP}^2 + \delta_{SNP}^2 + \eta_{SNP}^2$. Details about the statistical model and the computation of the different GRMs are provided in the **Methods** section.

Unbiased estimates when all GRMs are fitted simultaneously

We use simulation to quantify empirically the bias and precision of the HE and REML estimators of genetic variances, as well as to assess consistency with theory. We first validate our model using simulations from 254,679 unrelated participants of the UKB with 1.1M autosomal genotyped and imputed SNPs from the HapMap 3 (HM3) panel[20] with minor allele frequency (MAF) $\geq$ 1%. In our simulations, we sampled putative causal variants from a pool of 100,000 pre-defined HM3 SNPs (**Methods**). We used the same pool

of causal variants across all simulation replicates from which we randomly sampled 1000 causal variants for each replicate. The phenotypes were simulated using **Equation (5)** (in **Methods**) where the additive, dominance, and additive-by-additive effects were generated from a standard normal distribution and adjusted to the expected variance of the additive, dominance and additive-by-additive genome-wide effects (simulated $h_{SNP}^2, \delta_{SNP}^2$ and η_{SNP}^2). The residuals were then generated from a normal distribution with mean 0 and variance $1 - (h_{SNP}^2 + \delta_{SNP}^2 + \eta_{SNP}^2)$. The genotype data were used to calculate the three GRMs Θ_A, Θ_D and Θ_{AA} as defined above (summary statistics for each GRM are provided in **Supplementary Table 1 and 2**).

We estimated $h_{SNP}^2, \delta_{SNP}^2$ and η_{SNP}^2 either jointly (hereafter referred to as ADAA model) or by including their corresponding GRM one at a time in the model. Variance components were then estimated using REML and HE regression, which are both implemented in the GCTA[16] software. The full dataset of 254,679 individuals was analyzed using HE regression with pairwise phenotypic cross-products. However, because of its heavy computational burden, REML estimates were derived from 8 sub-datasets, each of $\simeq$ 32K individuals. We meta-analyzed estimates from these 8 sub-datasets using inverse-variance weighting (IVW) (**Methods**), recognizing that the standard errors of these meta-analysed estimates are expected to be 2.8 fold higher than if analysis of a single combined dataset had been possible for additive and dominance variance, and 1.3 fold higher for additive-by-additive variance. Unless stated otherwise, simulations results are shown for analyses including the causal variants.

From REML analysis, the model shows good orthogonal properties (fitting all the three GRMs together or only one at a time does not change the estimates) for $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$. However, we found a significant deviation from regressing $\hat{\eta}^2_{SNP}$ estimates from the ADAA model and that from a univariate model fitting Θ_{AA} only (intercept = 0.65, see **Figure 1-A**). The same conclusion applies to HE analysis where $\hat{\eta}^2_{SNP}$ from the univariate model show an even stronger deviation (intercept = 1.20) from the joint estimates (**Figure 1-B**), as compared to REML. A lack of orthogonality is expected when using linked markers as orthogonal estimates of epistatic variance are not possible[21]. However, we observed a lack of orthogonality in simulations using simulated unlinked markers and unrelated individuals (**Supplementary Figure 1**). Although this observation was initially surprising, we derived theoretically why this occurs, and reflects induced collinearity between the GRM because allele frequencies are estimated with error from finite sample sizes (**Supplementary Note 1** and **Supplementary Figure 9**). The main and practical conclusion from these simulations is that Θ_A and Θ_D, the additive and dominance GRMs, should always be fitted when estimating additive-by-additive variance using Θ_{AA}.

After assessing the orthogonality properties of the model, we evaluated its performance with respect to loss of information due to tagging and its sampling variance. Consistent with Yang et al.[22], we define unbiasedness as when the average estimates across simulation replicates is not statistically different from the true expected value (t-test *P*-value>0.05). **Figure 2** shows that REML analysis performed with all observed HM3 SNPs including

causal variants yields nearly unbiased estimates of the three variance components ($\hat{h}^2_{SNP} = 0.30, \hat{\delta}^2_{SNP} = 0.30, \hat{\eta}^2_{SNP} = 0.29$). A slight downward bias is expected when dealing with missing genotypes that are imputed during the simulation process. Using simulation of unrelated individuals and unlinked markers, we can show that $\hat{h}^2_{SNP} \simeq (1 - r)h^2_{SNP}$, with r the missing genotype rate (**Supplementary Figure 4**). When the pool of causal variants was excluded from the analysis, $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$ both showed a downward bias with respective mean estimates of 0.28 and 0.26 whereas $\hat{\eta}^2_{SNP} = 0.30$ did not appear deflated. However, we lacked power to detect a small bias because of the large sampling variance of $\hat{\eta}^2_{SNP}$. To verify that the biases for $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$ are consistent with loss of tagging, we estimated the adjusted multiple correlation R^2 computed for the first 10K causal SNPs with nearby SNPs in 1Mb windows (**Methods**). We found values of 0.96 and 0.86, for additive and dominance effects, respectively, leading to expected values $\hat{h}^2_{SNP} = 0.29$ and $\hat{\delta}^2_{SNP} = 0.26$, consistent with the estimated variance components. These results show that despite incomplete LD we are still able to detect additive and dominance variation due to common causal variants in the UKB data. Results from HE analysis imply the same conclusion (**Figure 2**). Finally, we found the standard deviation (SD) of $\hat{\eta}^2_{SNP}$ across simulation replicates to be one order of magnitude larger than that of $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$. More specifically, $SD(\hat{h}^2_{SNP}) = 0.008$, $SD(\hat{\delta}^2_{SNP}) = 0.014$ and $SD(\hat{\eta}^2_{SNP}) = 0.16$, in agreement with theoretical expectations presented below.

Standard error of the estimate of additive-by-additive variance

We next sought to derive the theoretical sampling variance of estimators of η^2_{SNP}. Assuming a sample of N unrelated individuals from an outbred population, the diagonal and off-diagonal elements of the G matrix are respectively $G_{ii} \sim \mathcal{N}\left(1, Var(G_{ii})\right)$ and $G_{ij} \sim \mathcal{N}\left(0, Var(G_{ij})\right)$. Using results from Visscher et al.[23] and Visscher and Goddard[24], we derived different sampling variance for HE regression using phenotypic cross-products and REML (**Supplementary Note 2**),

$$Var\left(\hat{\eta}^2_{SNP\,HE}\right) \simeq \frac{1}{N^2 Var(G_{ij})^2} \qquad (3)$$

and

$$Var\left(\hat{\eta}^2_{SNP\,REML}\right) \simeq \frac{2}{N(4Var(G_{ii}) + 2Var(G_{ii})^2 + 3(N-1)Var(G_{ij})^2)} \qquad (4)$$

This result implies that statistical power to detect η^2_{SNP} is substantially larger when using REML as compared to HE. Note that the sampling variance of $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$ under REML or HE regression are approximately the same. For an infinite sample size, $\frac{Var(\hat{\eta}^2_{SNP\,HE})}{Var(\hat{\eta}^2_{SNP\,REML})} \simeq 1.5$ (**Supplementary Note 2**), while this ratio is in fact much larger for finite sample sizes (**Supplementary Figure 2**).

From simulations of 35K unrelated individuals with simulated unlinked markers, we found approximations from **Equations (3) and (4)** to be accurate. For REML analyses across 100 replicates, we observed $SD(\hat{h}^2_{SNP}) = 0.012$ (0.013 under HE), $SD(\hat{\delta}^2_{SNP}) = 0.013$ (0.014

11

under HE) and $SD(\hat{\eta}^2_{SNP}) = 0.611$ (2.817 under HE), as compare to theoretical standard

errors (SE) $\mathbb{E}(SE(\hat{h}^2_{SNP})) = \mathbb{E}(SE(\hat{\delta}^2_{SNP})) = 0.013$ (under REML and HE), and

$\mathbb{E}(SE(\hat{\eta}^2_{SNP})) = 0.599$ under REML (2.858 under HE). We then used simulations based on

actual genotypes from UKB participants to assess the accuracy of our theoretical

expectation on real data. For each replicate, we quantified the empirical SE of estimates of

variance components as the SD across 8 sub-datasets of 32K individuals. The observed

standard deviations were then averaged over replicates and compared to the theoretical

expectations. Overall, we found both our theoretical expectations to be accurate (**Table 1**).

We noticed a very small absolute bias downward for the theoretical SE in the order of 10^{-3}

for $\mathbb{E}\left(SE(\hat{h}^2_{SNP})\right)$ and $\mathbb{E}(SE(\hat{\delta}^2_{SNP}))$, and upward in the order of 10^{-2} for $\mathbb{E}(SE(\hat{\eta}^2_{SNP}))$, both

under REML and HE.

We also re-assessed the validity of the SE reported by GCTA and found it to be accurate for

REML (**Table 1)**. When performing HE analysis, GCTA reports two estimates of SE based on

OLS (SE_{OLS}) or jackknife (SE_{Jack}). Comparing the observed SD to the reported SE from

GCTA, we found SE_{OLS} to be slightly biased downward for $SE_{OLS}(\hat{h}^2_{SNP})$ and $SE_{OLS}(\hat{\delta}^2_{SNP})$

and upward for $SE_{OLS}(\hat{\eta}^2_{SNP})$ as observed with the theoretical SE. However, the reported

SE_{Jack} by GCTA is always more conservative. Therefore, we chose to use the latter and the

corresponding *P*-values for the real phenotype analysis.

Finally, using our theoretical expectations, we computed the expected power to detect a significant additive-by-additive effect from our UKB data. The sampling variance of $\hat{\eta}^2_{SNP}$ is expected to be very large under HE for small sample sizes, when it is much smaller under REML (**Supplementary Figure 2**), leading to a substantially larger power (**Figure 3**). However, the increase of sample size reduces the gap between the two approaches and we expect $\frac{Var(\hat{\eta}^2_{SNP\,HE})}{Var(\hat{\eta}^2_{SNP\,REML})} \simeq 2$ for a sample size of 1M unrelated individuals (**Supplementary Figure 2**). We show that even if we were to be able to analyze 254K individuals jointly with REML, our power (at $\alpha = 5\%$) to detect an additive-by-additive effect of 0.2 (which is unlikely in real data) would only be $\simeq 0.45$ under REML and $\simeq 0.17$ under HE. Under the REML IVW meta-analysis, the expected standard error becomes $SE_{IVW} = \frac{SE_{32K}}{\sqrt{8}}$, where SE_{32K} is the expected SE for a sample of 32K individuals. Hence, our expected power under REML slightly decreases because of the meta-analysis strategy (**Supplementary Figure 3**) and we expect a power of $\simeq 0.29$ ($\alpha = 0.05$) for $\eta^2_{SNP} = 0.2$. Comparing the expected SE under the IVW-REML and the HE analysis of 254K individuals, we expect slightly larger $SE\left(\hat{h}^2_{SNP\,IVW-REML}\right) = 0.004$ and $SE\left(\hat{\delta}^2_{SNP\,IVW-REML}\right) = 0.005$ as compared to $SE\left(\hat{h}^2_{SNP\,HE}\right) = 0.001$ and $SE\left(\hat{\delta}^2_{SNP\,HE}\right) = 0.002$. However, the meta-analysis under REML is expected to produce a smaller $SE\left(\hat{\eta}^2_{SNP\,IVW-REML}\right) = 0.14$ as compared to $SE\left(\hat{\eta}^2_{SNP\,HE}\right) = 0.19$.

In summary, our simulations and analytical results demonstrate that reliable inference of non-additive variance components is achievable using REML and HE regression in large samples, and therefore that these methods can be applied to real data.

13

Estimation of non-additive genetic variation in human complex traits

We estimated h^2_{SNP}, δ^2_{SNP} and η^2_{SNP} for 70 quantitative traits of the UKB (**Methods**), with an average number of phenotyped individuals $\simeq$ 200,550 (range: 36,990 to 251,805). For REML analysis, IVW meta-analysis was applied when $N > 60K$. Otherwise, the entire set of individuals was analysed (**Methods**). Estimates of h^2_{SNP}, δ^2_{SNP} and η^2_{SNP} using REML and HE are reported in **Supplementary Tables 4 and 5** and the distributions of REML and HE estimates across traits are shown in **Figure 4**. The mean estimate of h^2_{SNP} across traits was $\bar{h}^2_{SNP\,REML} = 0.207 \pm 0.001$ (CI95% = [0.205;0.208]) for REML and slightly lower for HE with $\bar{h}^2_{SNP\,HE} = 0.194 \pm 0.001$ (CI95% = [0.192;0.196]). The average estimate of δ^2_{SNP} across traits was $\bar{\delta}^2_{SNP} = 0.001 \pm 0.001$ (CI95% = [-0.001;0.003]) for both REML and HE. This result is in agreement with theory suggesting than σ^2_D is much smaller than σ^2_A. Finally, we found a mean estimate of η^2_{SNP} equal to $\bar{\eta}^2_{SNP\,REML} = 0.058 \pm 0.032$ (CI95% = [-0.005;0.122]) for REML and $\bar{\eta}^2_{SNP\,HE} = -0.054 \pm 0.112$ (CI95% = [-0.278;0.171]) for HE. Estimates are consistent between REML and HE analysis for $\hat{h}^2_{SNP}$ and $\hat{\delta}^2_{SNP}$ (squared correlations r²=0.97 and 0.43 respectively), but the two methods show poor agreement for $\hat{\eta}^2_{SNP}$ (squared correlation r²=0.04, see **Supplementary Figure 5**), consistent with their high standard errors.

From the 70 traits analysed, we estimated an effective number of 43 traits (**Methods**) and test for 3 variance components, however, because we ascertained traits with significant

additive variance (**Methods**), we only included two variance components for multiple testing correction. Therefore, our P-value threshold for declaring statistical significance at an experiment-wise error rate $\alpha = 0.05$ is $0.05/(43 \times 2) = $ 5.8e-4. After multiple-testing correction, urate concentration (N = 238,773) was the only trait showing a small but significant dominance variance ($\hat{\delta}^2_{SNP} = 0.012 \pm 0.003$, P = 9.6e-5) under HE analysis. However, significance was not obtained using REML ($\hat{\delta}^2_{SNP} = 0.003 \pm 0.005$, P = 5.1e-1).

We did not find significant evidence of epistatic variance. However, we unexpectedly found two traits with an apparent negative $\hat{\eta}^2_{SNP}$ but significant before multiple-testing correction under REML, Bone mineral density (BMD) ($\hat{\eta}^2_{SNP} = -0.58 \pm 0.176$, P = 1.0e-3) and Corneal resistance factor ($\hat{\eta}^2_{SNP} = -0.80 \pm 0.233$, $P = 6.3e-4$). This unusual result implies that close relatives would be disproportionally less phenotypically similar than distant relatives. We therefore quantified phenotypic covariance in family data in the UKB as a function of relatedness but found no evidence for "phenotypic repulsion"[25] (**Supplementary Figure 6**). All these estimates of non-additive genetic variation are implausibly large, so the most parsimonious explanation is the large sampling variance.

Discussion

The role of non-additive genetic variance in human complex traits has been a topic of much discussion and debate[7,26-30]. In this study, we jointly estimate the proportions of phenotypic variance of 70 traits that is explained by additive, dominance and additive-by-additive

15

genetic variation tagged by common SNPs, in a large sample of 254,679 unrelated individuals from the UKB. Using common variants, we found no evidence of significant non-additive variance ($\bar{\delta}^2_{SNP}$=0.001 and $\bar{\eta}^2_{SNP} = 0.055$) across traits while we confirm the evidence for additive variance ($\bar{h}^2_{SNP} = 0.208$). We also derived theoretical standard errors for REML and HE regression estimators of additive-by-additive variance and validated our theory through extensive simulations. Our theoretical and empirical results suggest that REML should be preferred over HE regression for the same sample size, as the former estimator leads to a substantially larger statistical power.

A lack of dominance variance is expected from theory[7] and for traits that are polygenic and subjected to inbreeding depression.[9] However, most of our 70 traits do not show evidence of inbreeding depression. Nevertheless, the argument of high degree of polygenicity could be enough to expect a lack of dominance variance too[31,32]. Indeed and as quoted by Crow[31] for continuous quantitative traits, "in general, the smaller the effects, the more nearly additive they are" (see **Supplementary Note 3**). In a large sample, our average dominance variance estimated across traits $\bar{\delta}^2_{SNP} = 0.001$, is much lower than the 0.03 previously reported by Zhu et al.[10] across 79 traits, although the two estimates are not significantly different from each other. Therefore our findings are consistent with that from Zhu et al.[10] as they confirm a negligible contribution of dominance variance for causal variants that are tagged by common SNPs. By extrapolation, our results lead to the conclusion that dominance variance is likely to contribute very little to the broad sense heritability of human complex traits. Urate concentration was the only trait showing a significant

16

dominance variance in our HE analysis ($\hat{\delta}^2_{SNP} = 0.012 \pm 0.003$, P = 9.6e-5), but the estimate was lower and non-significant in our REML analysis ($\hat{\delta}^2_{SNP} = 0.003 \pm 0.005$, P = 5.1e-1). Similarly, we observed that h^2_{SNP} estimates for urate concentration were higher using HE compared to REML ($\hat{h}^2_{SNP} = 0.383 \pm 0.005$ under HE and 0.304 ± 0.004 under REML). If REML and HE estimates are expected to converge under a classical polygenic model, the genetic architecture of a trait can bias the estimates of the two methods in different ways[33-35]. Previous studies[36,37] have suggested that urate concentration is a trait on the low spectrum of polygenicity and controlled by three main genes and two large effect QTLs on chromosome 4. Moreover, urate concentration also displays sex-specific effects and heritability[38]. Analysis performed on each sex separately (with a 3 fold IVW meta-analysis for REML) confirmed a larger estimated heritability in females ($\hat{h}^2_{SNP} = 0.549 \pm 0.009$ under HE and 0.354 ± 0.005 under REML) than in males ($\hat{h}^2_{SNP} = 0.254 \pm 0.007$ under HE and 0.237 ± 0.006 under REML), and a significant dominance variance for females only under HE analysis ($\hat{\delta}^2_{SNP} = 0.017 \pm 0.005$, P = 1.4e-3 for females, and $\hat{\delta}^2_{SNP} = 0.005 \pm 0.006$, P = 3.9e-1 for males). Importantly, the significant dominance variance estimate detected in females with HE vanishes when excluding SNPs on chromosome 4 from our analysis ($\hat{\delta}^2_{SNP} = 0.0004 \pm 0.006$, P = 0.93 without chromosome 4). Similarly, chromosome 4 also entirely accounts for the observed differences in $\hat{h}^2_{SNP}$ estimates between HE and REML analysis in females ($\hat{h}^2_{SNP} = 0.175 \pm 0.005$ under HE and 0.185 ± 0.005 under REML), as well as the discrepancies between males and females (for males, $\hat{h}^2_{SNP} = 0.151 \pm 0.005$ under HE and 0.165 ± 0.006 under REML). Altogether, low polygenicity of urate concentration combined with substantial sex differences in the genetic architecture on chromosome 4 constitute strong departure from assumptions

17

underlying the consistency of HE and REML estimators. There is also prior evidence of female sex hormones effects on urate concentration[38], which we did not accounted for in our analysis, and further investigations of the sex-specific genetic architecture of the trait are needed. Finally, we detected significant dominance deviation ($P < $ 5e-8) at two SNPs on chromosome 1 (top hit for rs12124078, a SNP associated with kidney function[39]) and 4 (larger signal with top hit for rs9998811 located in *SLC2A9* gene region[36]), (**Supplementary Figure 7**) although the total dominance variance explained by these two SNPs remains very small ($\hat{\delta}^2_{SNP} = 0.002$).

The lack of evidence for the additive-by-additive variance in our analyses is mostly due to the very large sampling variance of $\hat{\eta}^2_{SNP}$. Therefore, potentially real effects on individual traits cannot be ruled out. Nevertheless, we can provide an upper limit for the role of epistatic variance associated with SNPs across the 70 traits of approximately 0.12 (upper bound of the 95% CI from REML analysis, see **Figure 4-A**), based on the SE of the mean $\bar{\eta}^2_{SNP}$ of 0.032. This upper limit (across traits) remains smaller than the well-estimated mean h^2_{SNP} of 0.207. Our power calculations based on REML suggest that $\simeq$2 million unrelated individuals would be necessary to ensure >76% statistical power to detect $\eta^2_{SNP} = 0.05$ at $\alpha = 0.05$. The same calculations based on HE regression would only yield a statistical power of 52%. Analysis of such a large sample raise new computational challenges to be addressed in future research.

18

Besides these limitations, theory and prior evidence also suggest that epistatic variance is likely to be small[7,8,21], with first order additive-by-additive epistasis expected to be the largest contributor to epistatic variance. In this study, we therefore focused on the additive-by-additive variance and did not estimate higher order interactions or epistatic components involving dominance. Because epistatic GRMs result from the Hadamard products of lower order GRMs, they will quickly tend to identity with increasing order of interactions. It will lead to very large standard errors and estimates of genetic variance indistinguishable from the residual (non-genetic) variance. The same argument holds for any epistatic interaction involving dominance as it will result in even larger standard errors, the variance of the off-diagonal elements of Θ_D in unrelated European individuals being approximately 10^{-5}, half of the variance of the off-diagonal elements of Θ_A. Finally, Mäki-Tanila and Hill[8] showed for polygenic traits that when gene-gene interactions (aa) are of same magnitude as single-locus effects (a), epistatic variance is expected to decrease and eventually disappear with the increase number of causal variants. In this context, epistatic variance lead to mainly additive variance, while dominance is expected to remain the main contributor to non-additive genetic variance. Therefore, and in light of our results showing a lack of dominance variance, that would suggest that epistatic variance is likely to be extremely small in human complex traits.

Our findings can be directly applied to human diseases using a liability threshold model. Yet, it is worth emphasizing that the observed 0-1 scale is expected to show non-additive variance even when variance in liability is fully additive[40]. An analysis of non-additive

variance on liability would require the use of generalized non-linear mixed models and would be less powerful than the analysis of quantitative traits. For these reasons we have not attempted to estimate dominance or epistatic variance for liability to common disease in our study. Results for additive SNP-based heritability are by-and-large the same for quantitative traits and common disease, so it seems reasonable to assume that there is likely to be little non-additive variance for liability of disease.

We showed by simulation that we were able to capture a large proportion of the genetic variation from common SNPs in our data even when causal variants were not included. However, because the expected loss of non-additive variance is disproportionally larger than additive variance with the decay of linkage disequilibrium between causal and tagging variants, the contribution of rare variants poorly tagged in our study is expected to be missed. There is evidence that rare variants contribute to narrow sense heritability[41,42]. However, using simulations, Zhu et al.[10] showed that the observed difference between the additive and dominance variance across traits was unlikely to be explained by a disproportionally missing contribution of rare variants to dominance variation. Moreover, because the amount of non-additive variance also disproportionally depends on allele frequencies as compared to additive variance, with the largest amount expected for intermediate frequencies[8], contribution of rare variants to the non-additive genetic variance is expected to be minute.

Importantly, the absence of evidence for epistatic variance does not imply the absence of functional epistasis[43,44]. In addition, any significant signal of epistatic variance should be investigated in depth due to potential inferential problems. Indeed, phantom epistasis, that is a non-additive signal generated from incomplete linkage between variants,[45,46] can induce a bias in the estimates of epistatic variance. A recent study suggested that the effect of phantom epistasis is likely to increase for low density markers which is not necessarily relevant for human data. However, the genetic architecture of a trait, such as local polygenicity (i.e. the non-random spatial distribution of small effect loci) and large effects QTL, could also favor phantom epistasis.

To conclude, the analysis of 70 human complex traits from a large sample of unrelated individuals provides new evidence that genetic variance for complex traits is predominantly additive and suggests negligible dominance variance due to causal variants that are associated with common SNPs. Because of a large standard error, we cannot draw firm conclusions regarding additive-by-additive variance for individual traits, but we can conclude that its upper value is about half of the additive genetic variance captured by common SNPs. We showed that REML lead to substantially larger power as compared to HE at a given sample size, and that sample sizes of many millions of unrelated individuals will be necessary to estimate epistatic variance with sufficient precision.

Methods

Genetic model

We assume a model with additive, dominance and epistatic interaction effects (ADAA model). For epistatic variance, we only focus on additive-by-additive interactions. Consider one diploid individual genotyped at L loci, each with major and minor alleles A and B respectively. Let $a_i = (\mu_{AA})/2$ and $d = \mu_{AB} - (\mu_{BB} + \mu_{AA})/2$, with μ_{BB}, μ_{AB} and μ_{AA} being the phenotypic means in the three genotypic classes at one locus. Let p_i be the allele frequency of A at locus i, a_i the additive effect, d_i the dominance effect, aa_{ij} the additive-by-additive interaction effect between locus i and j. Under the Hardy-Weinberg equilibrium assumption, we can define the average effect of allele substitution, i.e., $\beta_i = a_i + (1 - 2p_i)d_i + \sum_{j \neq i}^{L} 2 p_j aa_{ij}$, which contains the additive effect, a term due to dominance interaction between two alleles, and a term due to additive-by-additive interactions between pairs of loci[8]. Then, the additive variance at locus i ($\sigma^2_{A(i)}$) is $2p_i(1 - p_i)[a_i + (1 - 2p_i)d_i + \sum_{j \neq i}^{L} 2 p_j aa_{ij}]^2$, the variance of the average effect of allele substitution, dominance variance ($\sigma^2_{d(i)}$) is $[2p_i(1 - p_i)d_i^2]$, and additive-by-additive variance ($\sigma^2_{aa(i)}$) is $\sum_{j \neq i}^{L} 2 p_i(1 - p_i)2p_j(1 - p_j)(aa_{ij})^2$. Hence, the genotypic variance ($\sigma^2_{g(i)}$) is $\sigma^2_{a(i)} + \sigma^2_{d(i)} + \sigma^2_{aa(i)}$.

Let us define the additive genotype coding $x_{A(i)}$ as 0,1 or 2 for genotypes BB, AB and AA respectively , and dominance coding $x_{D(i)}$ as 0, $2p_i$ or $(4p_i - 2)$. This parametrization of $x_{d(i)}$ following Zhu et al.[10] ensures the orthogonality with $x_{A(i)}$, compare to the classical 0,

22

1, 0 code use for the different genotypes, which lead to $Cov(x_{A(i)}, x_{D(i)}) = 2p_i(1 - 2p_i)$.

Now including all L biallelic loci, we can fit the additive, dominance and additive-by-additive effect of all SNPs as random effects in a mixed linear model:

$$y = \mu + \sum_{i=1}^{L}\left[w_{A(i)}u_{A(i)} + w_{D(i)}u_{D(i)}\right] + \sum_{i=1}^{L-1}\sum_{j>i}^{L} w_{A(i)}\, w_{A(j)}u_{AA(ij)} + e \tag{5}$$

with $w_{A(i)} = (x_{A(i)} - 2p_i)/\sqrt{2p_i(1 - p_i)}$ and $w_{D(i)} = (x_{D(i)} - 2p_i^2)/(2p_i(1 - p_i))$ are respectively the standardized form of x_A and x_D. u_A, u_D and u_{AA} are the additive, dominance and additive-by-additive random effects of the standardized genotypes, μ is the mean term and $e \sim \mathcal{N}(0, \sigma_e^2)$ is the residual.

This model can be expressed as an individual based model for a sample of N diploid individuals and the presence of fixed covariates, written in matrix form as:

$$\mathbf{y} = \mathbf{Cb} + \mathbf{W}_A\mathbf{u}_A + \mathbf{W}_D\mathbf{u}_D + \mathbf{W}_{AA}\mathbf{u}_{AA} + \mathbf{e}$$

where $\mathbf{y}$ is a $N \times 1$ vector of individuals phenotypes, $\mathbf{W}_A$ and $\mathbf{W}_D$ are $N \times L$ matrices with one row per individual containing the corresponding $w_{A(i)}$ and $w_{D(i)}$ vectors. $\mathbf{W}_{AA}$ is a $N \times \frac{L(L-1)}{2}$ matrix with one row per individual containing all the pairwise $w_{A(i)}w_{A(j)}$. $\mathbf{u}_A$ and $\mathbf{u}_D$ are $L \times 1$ vectors of locus specific additive and dominance effects, $\mathbf{u}_{AA}$ is a $\frac{L(L-1)}{2} \times 1$ vector of additive-by-additive effects, $\mathbf{e}$ is an $N \times 1$ vector of residuals. $\mathbf{C}$ is an $N \times c$ matrix of c covariates, and $\mathbf{b}$ is a $c \times 1$ vector of corresponding covariate effects.

23

If we define $\mathbf{g}_A = \mathbf{W_A}\mathbf{u_A}$, $\mathbf{g}_D = \mathbf{W}_D\mathbf{u}_D$ and $\mathbf{g}_{AA} = \mathbf{W}_{AA}\mathbf{u}_{AA}$, then we have:

$$\mathbf{y} = \mathbf{Cb} + \mathbf{g}_A + \mathbf{g}_D + \mathbf{g}_{AA} + \mathbf{e}$$

The (co)variance matrix of phenotypes becomes:

$$\begin{aligned} \text{var}(\mathbf{y}) &= \text{var}(\mathbf{g}_A) + \text{var}(\mathbf{g}_D) + \text{var}(\mathbf{g}_{AA}) + \text{var}(\mathbf{e}) \\ &= \mathbf{\Theta}_A\sigma_A^2 + \mathbf{\Theta}_D\sigma_D^2 + \mathbf{\Theta}_{AA}\sigma_{AA}^2 + \mathbb{I}\sigma_e^2 \end{aligned}$$

where $\mathbf{\Theta}_A = G = \frac{\mathbf{W_A}\mathbf{W_A'}}{L}$ and $\mathbf{\Theta}_D = \frac{\mathbf{W_D}\mathbf{W_D'}}{L}$ are the additive and dominance genomic relatedness matrices (GRM) as described in Yang et al.[16] and Zhu et al.[10], and $\mathbf{\Theta}_{AA}$ is the additive-by-additive genomic relatedness matrix define as:

$$\mathbf{\Theta}_{AA} = \frac{G \odot G}{tr(G \odot G)/N}$$

Genotype data

We analyze a large dataset of 347,849 unrelated (genomic relatedness < 0.05) individuals of European descent (188,088 females and 159,761 males), from the UK Biobank (UKB). Informed consent was obtained from all the subjects and those who expressed the which to be withdrawn have been removed from analysis. We used the release 3 of the UKB where individuals were genotyped on the Affymetrix UK Biobank Axiom array before imputation using the HRC and UK10K reference panel and IMPUTE2[47]. SNPs were filtered for quality control by removing those with missing genotyping rate > 0.05, Hardy-Weinberg equilibrium test $P < 10^{-6}$ and minor allele frequency (maf) < 0.01. After filtering, we extracted autosomal HapMap phase 3 (HM3) markers, resulting in 1,130,561 SNPs.

24

The Additive and Dominance GRM were computed using the standard algorithm of GCTA software[16] version 1.93.0b as previously described, and the additive-by-additive GRM was computed using R version 3.6.2[48] from the additive GRM as described in **Equation (2)**. In order to remove any cryptic relatedness, we trimmed the GRMs by removing one individual for each pair with an additive genomic relatedness > 0.025. This resulted in a final dataset of 254,679 unrelated individuals (138,196 females and 116,483 males).

Phenotypes selection and quality control

We chose 70 continuous trait from the UKB (listed in the **Supplementary Table 3**) with a total number of phenotyped individuals in the all UKB ≥ 60,000, a significant SNP heritability $h^2_{SNP} \geq 0.05$ (at $\alpha = 0.05$ level, based on Neale's lab h^2_{SNP} estimates https://nealelab.github.io/UKBB_ldsc/h2_browser.html) and a square pairwise phenotypic correlation $r^2 < 0.8$. First assessment data only were used, and for phenotypes with left and right measure, only one was chosen randomly. Outlier individuals were removed following Tukey's method[49] separately for males and females in every phenotype. This step resulted in a mean and minimum number of phenotyped individuals across traits of 200,550 and 36,690 respectively (among a total of 254,679 individuals). Each sex-specific dataset has been further corrected for age, as well as month of assessment for blood biochemistry traits only, and standardized to z-score. The first 20 eigenvectors of the principal components (PCs), estimated from the 254,679 individuals genotype data using

25

flashPCA2 software[50], were included as fixed covariates in the REML analyses and the phenotypes were pre-corrected for HE analysis.

Effective number of phenotypes

For multiple testing correction purpose, we computed the effective number of phenotypes (P_{Eff}) from a general purpose estimator[51] using the Shannon entropy:

$$P_{Eff} = \prod_{j=1}^{P} \left(\frac{\lambda_j}{\sum_{i=1}^{P} \lambda_i} \right)^{-\frac{\lambda_j}{\sum_{i=1}^{P} \lambda_i}}$$

where P is the total number of phenotypes and λ_j the jth eigenvalue of the phenotypes' correlation matrix. Using the correlation matrix of our 70 traits, we found $P_{Eff} = 43$ and used it for Bonferroni corrections. Because we tested for three variance components but chose only traits with significant SNP-based heritability, we also used two components for the Bonferroni corrections, resulting in a threshold P-value of 0.05/(43x2) = 5.8e-4 at $\alpha = 5\%$.

Estimation of genetic variance components

Genetic variance components were estimated using HE regression with phenotypic cross-products and REML. As stated in the main text, HE regression analyses were performed on the full dataset as well as REML when the number of phenotyped individuals was smaller or equal to 60K. Otherwise, we used an inverse-variance weighted (IVW) meta-analysis of 8

26

sub-datasets of $\simeq$32K individuals. For a single trait, the genetic variance estimate from the IVW meta-analysis of i sub-datasets and its associated standard error are:

$$\hat{\theta}_{IVW} = \frac{\Sigma_i\left(\frac{\hat{\theta}_i}{\sigma^2_{\theta_i}}\right)}{\Sigma_i\left(\frac{1}{\sigma^2_{\theta_i}}\right)} \text{ and } SE_{IVW} = \sqrt{\frac{1}{\Sigma_i\left(\frac{1}{\sigma^2_{\theta_i}}\right)}}$$

where $\hat{\theta}_i$ is the genetic variance estimate from the sub-dataset i, and $\sigma^2_{\theta_i}$ the associated sampling variance. Indications on whether we use a single or meta-analysis for each phenotype can be find in the **Supplementary Table 3**. To obtain unbiased estimates of the variance components, we did not constrain REML estimates to be non-negative. For REML, P-values of the estimates were computed as the probability of the likelihood ratio test $LRT \simeq (\frac{Estimate}{SE})^2$ in a Chi-square distribution with one degree of freedom. For HE analysis, we used the reported jackknife SE and P-values by GCTA.

We estimated the mean of each variance component estimate $\hat{\theta}$, $\mu = \frac{1}{P}\sum_{i=1}^{P}\hat{\theta}_i$ across our P traits, as well as its standard error $SE(\mu) = \sqrt{\frac{1}{P^2}\sum_{i=1}^{P}\left[\sigma^2_{\theta_i} + \sum_{i \neq j} C\ ov(\theta_i; \theta_j)\right]}$ where $Cov(\theta_i; \theta_j)$ accounts for the non-independence of the traits. We can show from simulations (**Supplementary Figure 8**) that a good approximation of the covariance between genetic estimates is $Cov(\theta_i; \theta_j) \simeq r_p^2 \times \sigma_{\theta_i} \times \sigma_{\theta_j}$ with r_p^2 the squared phenotypic correlation between traits i and j, and σ_{θ_i} the standard error of the estimate which can be approximate from our analysis (using the reported SE by GCTA). Dominance variance explained by

genome-wide significant SNPs was calculated as $\sum\left(p_i(1-p_i)\hat{d}_i\right)^2$ with p_i the empirical allele frequency at SNP i and $\hat{d}_i$ the estimated dominance effect from the genome-wide association analysis. We used R version 3.6.2[48] to perform the analysis of GCTA outputs.

Simulation studies

We performed simulation studies to validate our model and theoretical sampling variance of the estimates. We performed simulations based on the real genotypes observed on the 1,130,561 autosomal HM3 SNPs of the 254,679 unrelated European individuals from UKB. Following Zhu et al.[10], we randomly sampled 100,000 SNPs (across autosomes) as a pool of causal variants and used the remaining SNPs as the observed SNPs. To measure at which extent we were able to capture the additive and dominance variation of causal variants that are associated with common SNPs, we computed the multiple regression R^2 for the first 10K causal SNPs by regressing the $x_{A(i)}$ ($x_{D(i)}$ under dominance) of the target SNPs with the $x_{A(i)}$ ($x_{D(i)}$ under dominance) of the neighboring SNPs in a 1Mb window.

We generated the phenotypes using a custom C++ program following **Equation (5)** where the additive, dominance, and additive-by-additive effects were generated from a standard normal distribution and adjusted to the expected variance of the additive, dominance and additive-by-additive genome-wide effects (h^2_{SNP}, δ^2_{SNP} and η^2_{SNP}). The residuals were then generated from a normal distribution with mean 0 and variance $1 - (h^2_{SNP} + \delta^2_{SNP} + \eta^2_{SNP})$. The simulated phenotypes were standardized to z-scores. We chose $h^2_{SNP} = \delta^2_{SNP} = \eta^2_{SNP} =$

28

0.3 and simulated 100 replicates, each with 1000 randomly sampled causals SNPs from the pool of 100,000 putative causal variants. Missing genotypes were imputed to the mean genotype during the simulation process. The different variance components were estimated using **Equation (1)** under REML and HE regression, including or not causal variants.

To compare results from simulations based on observed genotypes of UKB participants with those under Hardy-Weinberg and linkage equilibrium, we used R v3.6.2[48] to generate a sample of 35K unrelated individuals genotyped at 100K unlinked markers with uniform allele frequency distributions in the range [0.01;0.99]. Phenotypes were then simulated as described above with the same setting ($h^2_{SNP} = \delta^2_{SNP} = \eta^2_{SNP} = 0.3$). Causal markers were randomly sampled among the 100K unlinked SNPs and all variants were included in the analysis.

Acknowledgments

Author contributions

P.M.V conceived and designed the study. Theory was derived by P.M.V, L.Y and V.H. The UKB phenotypic data were extracted by J.S who also assisted V.H on quality control. V.H and L.Y performed simulations and statistical analysis under the assistance and guidance of P.M.V, N.R.W and J.Y. The manuscript has been written by V.H, L.Y and P.M.V with the participation of all authors. All authors reviewed and approved the final manuscript.

Competing Interests

The authors declare no competing interests.

Web Resources

The URLs for data and software used in this paper are as follow:

UK Biobank: https://www.ukbiobank.ac.uk/

FlashPCA2: https://github.com/gabraham/flashpca

GCTA: https://cnsgenomics.com/software/gcta/#Overview

References

1. Cockerham, C.C. An extension of the concept of partitionning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* (1954).
2. Kempthorne, O. The correlation between relatives in a random mating population. *Proc. R. Soc. Lond. B*, 103-113 (1954).
3. Falconer, D.S. & Mackay, T.F.C. *Introduction to quantitative genetics*, (Burnt Mill, England : Longman, 1996).
4. Fisher, R.A. The Correlation between Relatives on the Supposition of Mendelian Inherit- ance. *TRANS. ROY SOC. EDIN.* **LII**, 399-433 (1918).
5. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*, (Sinauer Associates., 1998).
6. Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L.V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-237 (2013).
7. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* **4**(2008).
8. Mäki-Tanila, A. & Hill, W.G. Influence of gene interaction on complex trait variation with multilocus models. *Genetics* **198**, 355-367 (2014).
9. Robertson, A. & Hill, W.G. Population and quantitative genetics of many linked loci in finite populations. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **219**, 253-264 (1983).
10. Zhu, Z. *et al.* Dominance genetic variation contributes little to the missing heritability for human complex traits. *American Journal of Human Genetics* **96**, 377-385 (2015).
11. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-569 (2010).
12. Weir, B.S. Linkage disequilibrium and association mapping. *Annu Rev Genomics Hum Genet* **9**, 129-42 (2008).
13. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
14. Vitezica, Z.G., Legarra, A., Toro, M.A. & Varona, L. Orthogonal Estimates of Variances for Additive. *Genetics* **206**, 1297-1307 (2017).
15. Xu, S. Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* **195**, 1209-1222 (2013).
16. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76-82 (2011).
17. Elston, R.C., Buxbaum, S., Jacobs, K.B. & Olson, J.M. Haseman and Elston Revisited. *Genetic Epidemiology* **19**, 1-17 (2000).
18. Haseman, J.K. & Elston, R.C. The Investigation of Linkage Between a Quantitative Trait and a Marker Locus. *Behavior Genetics* **2**, 3-19 (1972).
19. Yang, J., Zeng, J., Goddard, M.E., Wray, N.R. & Visscher, P.M. Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics* **49**, 1304-1310 (2017).
20. International HapMap, C. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
21. Hill, W.G. & Mäki-Tanila, A. Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *Journal of Animal Breeding and Genetics* **132**, 176-186 (2015).

22.   Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* **47**, 1114-20 (2015).

23.   Visscher, P.M. *et al.* Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genetics* **10**(2014).

24.   Visscher, P.M. & Goddard, M.E. A general unified framework to assess the sampling variance of heritability estimates using pedigree or marker-based relationships. *Genetics* **199**, 223-232 (2015).

25.   Steinsaltz, D., Dahl, A. & Wachter, K.W. On Negative Heritability and Negative Estimates of Heritability. *Genetics* **215**, 343-357 (2020).

26.   Carlborg, Ö. & Haley, C.S. Epistasis: Too often neglected in complex trait studies? *Nature Reviews Genetics* **5**, 618-625 (2004).

27.   Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).

28.   Wei, W.H., Hemani, G. & Haley, C.S. Detecting epistasis in human complex traits. *Nat Rev Genet* **15**, 722-33 (2014).

29.   Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1193-1198 (2012).

30.   Zaitlen, N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet* **9**, e1003520 (2013).

31.   Crow, J.F. On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 1241-1244 (2010).

32.   Keightley, P.D. Models of quantitative variation of flux in metabolic pathways. *Genetics* **121**, 869-876 (1989).

33.   Chen, G.B. Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Front Genet* **5**, 107 (2014).

34.   Chen, G.B. On the reconciliation of missing heritability for genome-wide association studies. *Eur J Hum Genet* **24**, 1810-1816 (2016).

35.   Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* **91**, 1011-21 (2012).

36.   Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).

37.   Tin, A. *et al.* Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat Genet* **51**, 1459-1474 (2019).

38.   Halperin Kuhns, V.L. & Woodward, O.M. Sex Differences in Urate Handling. *Int J Mol Sci* **21**(2020).

39.   Pattaro, C. *et al.* Genome-Wide Association and Functional Follow-Up Reveals New Loci for Kidney Function. *PLOS Genetics* **8**, e1002584 (2012).

40.   Dempster, E.R. & Lerner, I.M. Heritability of threshol characters. *Genetics* **35**, 212-236 (1950).

41.   Hernandez, R.D. *et al.* Ultrarare variants drive substantial cis heritability of human gene expression. *Nat Genet* **51**, 1349-1355 (2019).

42.   Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv* (2019).

43. Barton, N.H. & Keightley, P.D. Understanding quantitative genetic variation. *Nat Rev Genet* **3**, 11-21 (2002).

44. Hansen, T.F. Why epistasis is important for selection and adaptation. *Evolution* **67**, 3501-11 (2013).

45. de Los Campos, G., Sorensen, D.A. & Toro, M.A. Imperfect Linkage Disequilibrium Generates Phantom Epistasis (& Perils of Big Data). *G3 (Bethesda)* **9**, 1429-1436 (2019).

46. Schrauf, M.F. *et al.* Phantom Epistasis in Genomic Selection: On the Predictive Ability of Epistatic Models. *G3 (Bethesda)* **10**, 3137-3145 (2020).

47. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

48. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2019).

49. Tukey, J.W. *Exploratory Data Analysis*, (Addison-Wesley, 1977).

50. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics (Oxford, England)* **33**, 2776-2778 (2017).

51. Del Giudice, M. Effective Dimensionality: A Tutorial. *Multivariate Behavioral Research* **0**, 1-16 (2020).
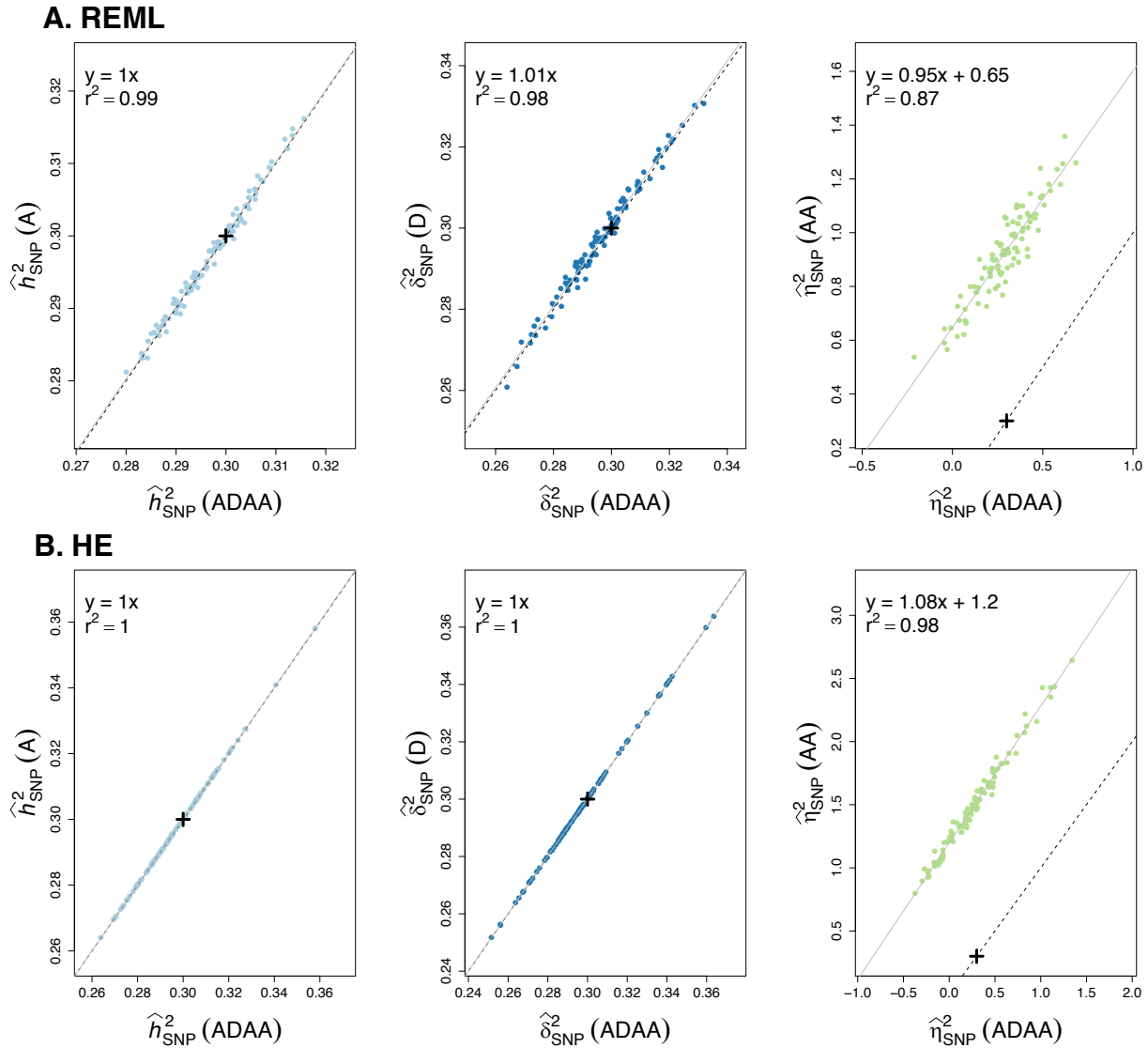
Figure 1: REML and HE estimates from simulations based on observed genotypes of UKB participants (simulated $h^2_{SNP} = \delta^2_{SNP} = \eta^2_{SNP} = 0.3$). We compared (**A**) IVW-REML (meta-analysis of 8 sub-datasets of 32K individuals) and (**B**) HE estimates (analysis of 254,679 individuals) for 100 replicates of simulations including the pool of causal variants when we jointly estimate the different variance components (ADAA) or only one at a time for h^2_{SNP}, δ^2_{SNP} and η^2_{SNP}. Simulated values are depicted by a black cross, the black dashed

line depicts the Y=X line while the solid grey line depicts the linear regression between the

corresponding ADAA and single component model. The squared correlation r^2 is close to 1

for all the variance components and η^2 is the only one showing a large intercept.
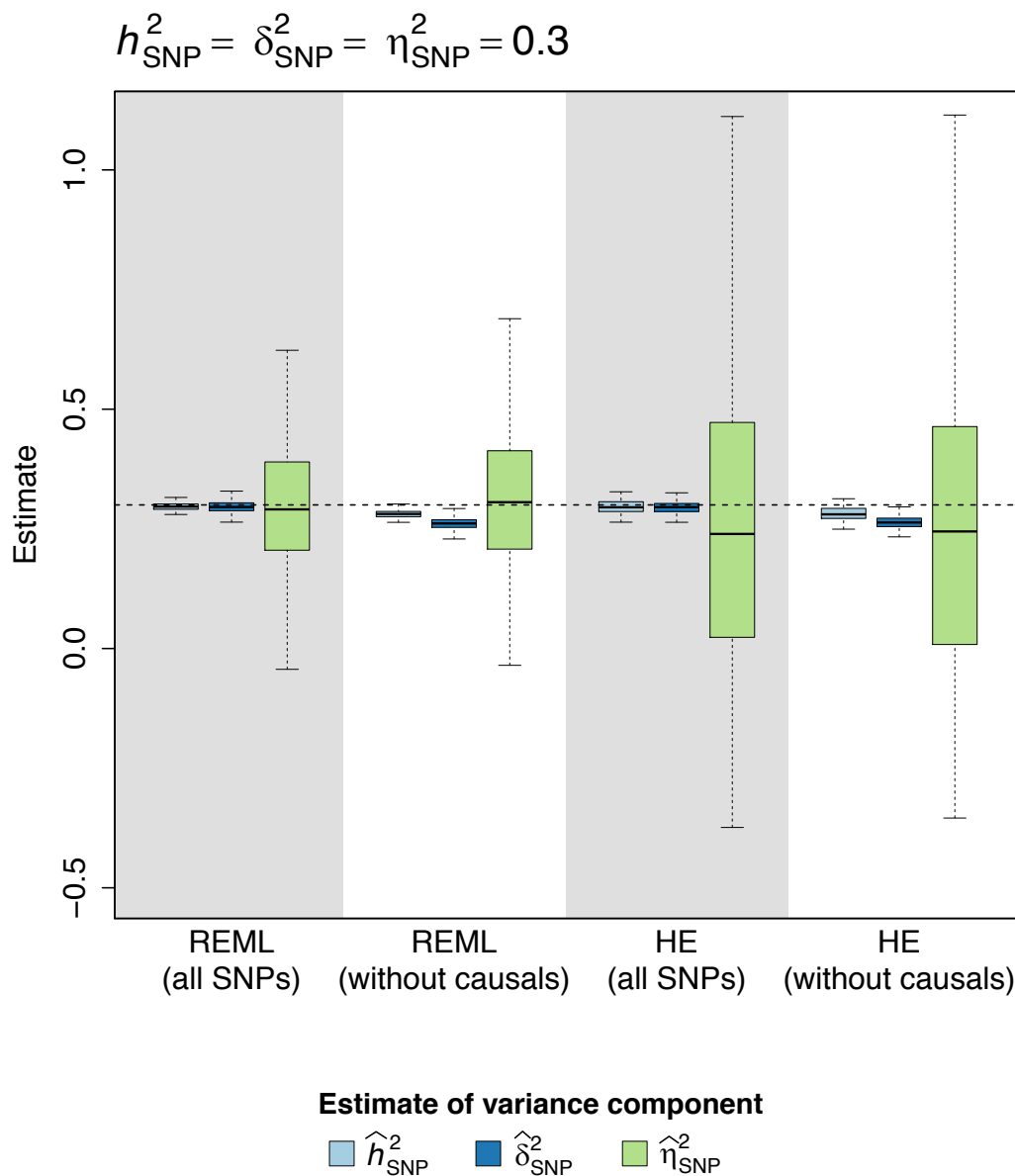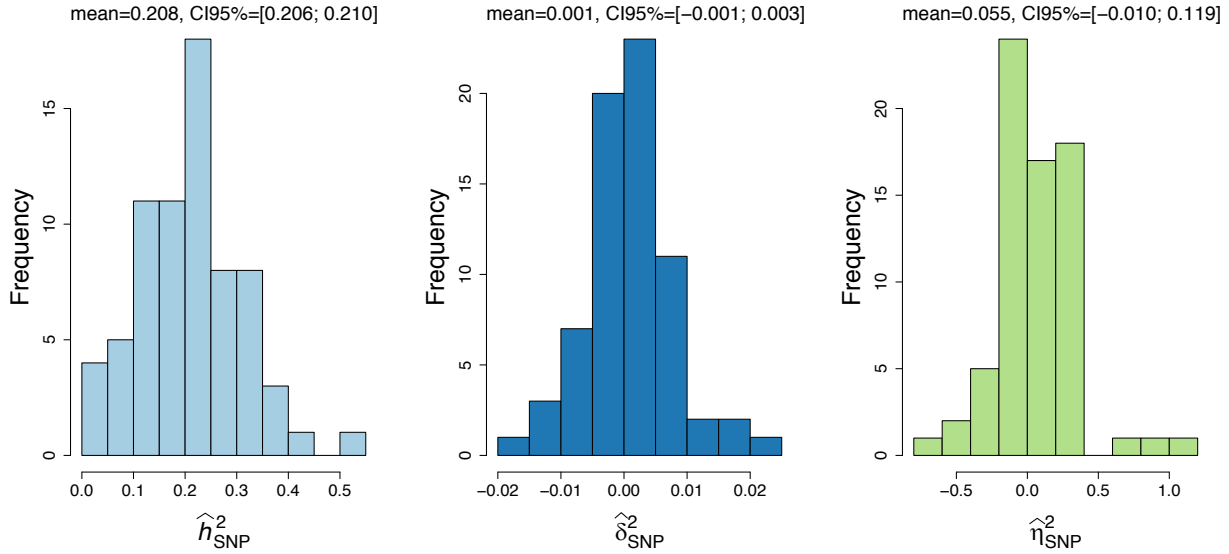
$$h^2_{\text{SNP}} = \delta^2_{\text{SNP}} = \eta^2_{\text{SNP}} = 0.3$$

Figure 2: **Distributions of REML and HE estimates from simulations based on observed genotypes of 254,679 UKB participants** (simulated $h^2_{SNP} = \delta^2_{SNP} = \eta^2_{SNP} = 0.3$). IVW-REML (analysis of 8 sub-datasets of 32K individuals) and HE estimates (analysis of 254,679 individuals) for 100 replicates of simulations are shown for the three different

variance components including all variants (all SNPs) or excluding the causal variants

(without causal). The black dashed line indicates the simulated value of 0.3.
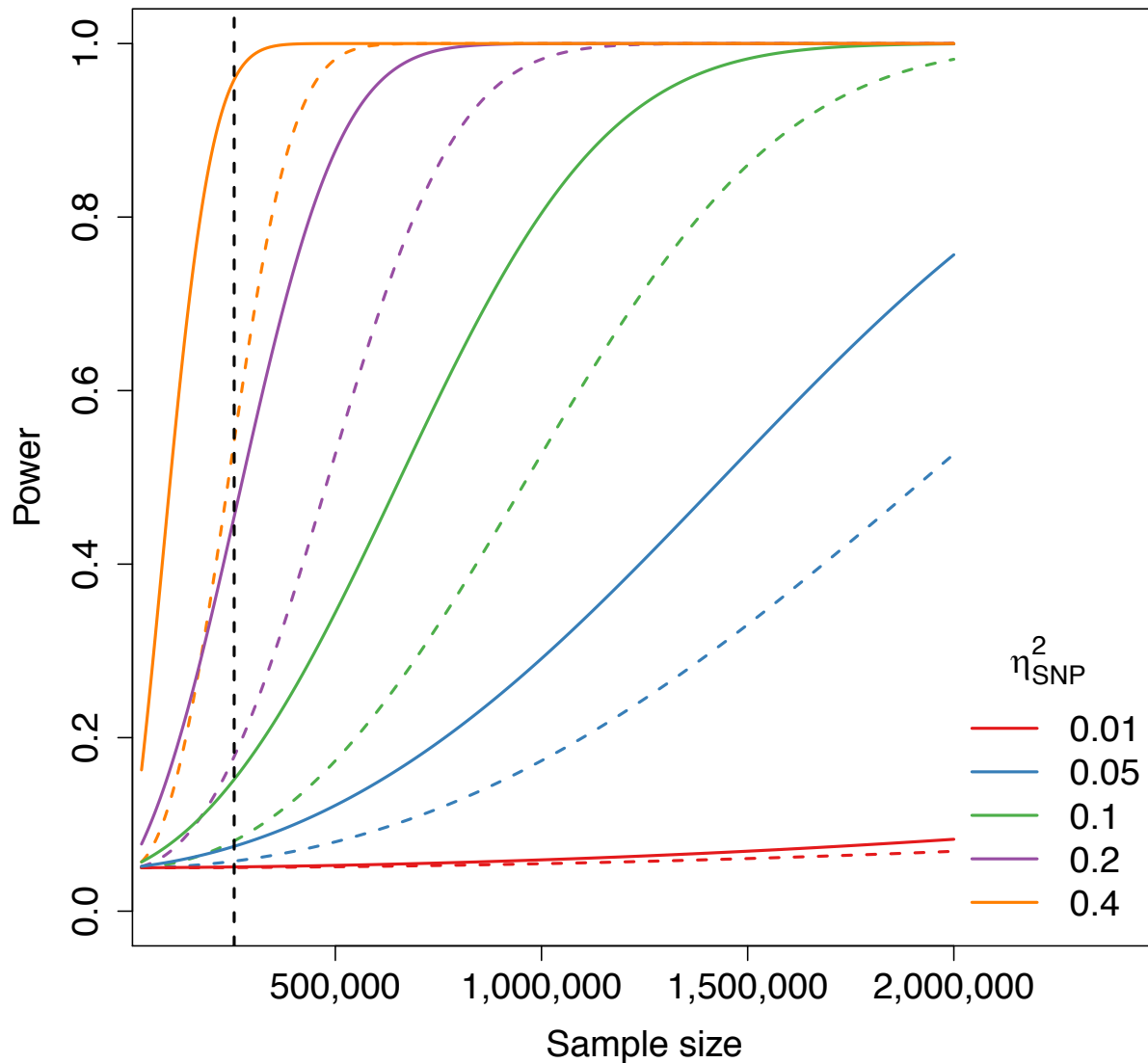
Figure 3: Statistical power to detect additive-by-additive variance as a function of sample size for REML (solid lines) and HE regression (dashed lines) on additive-by-additive GRMs computed on UKB data. Results are shown for an additive-by-additive heritability equal to 0.01, 0.05, 0. 1, 0.2 and 0.4. The current sample size of 254,679 unrelated individuals is depicted by the vertical black dashed line.

A. REML

B. HE
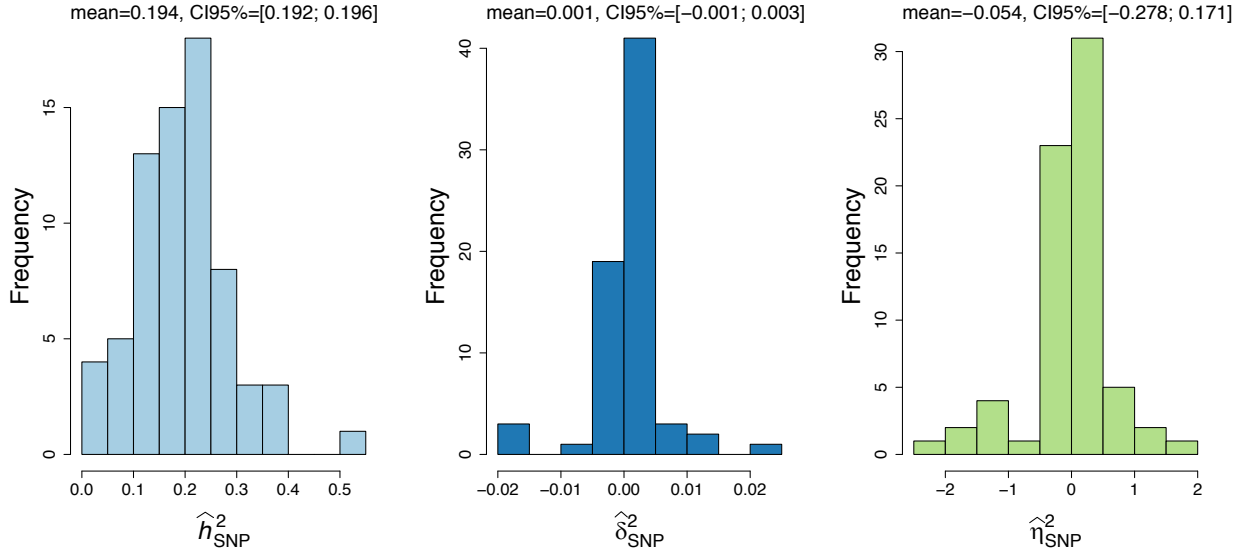
Figure 4: Distributions of the (A) REML and (B) HE estimates of SNP-based

h_{SNP}^2, δ_{SNP}^2 **and** η_{SNP}^2 **for 70 continuous traits in the UK Biobank.** For each distribution of

variance components estimates, we indicate the mean estimate as well as the 95%

confidence interval (CI95%).

Table 1: Observed standard deviations (SD) from the analysis of simulations based on observed genotypes of 254,679 UKB participants (simulated $h^2_{SNP} = \delta^2_{SNP} = \eta^2_{SNP} = 0.3$), as well as theoretical and reported standard errors (SE) by GCTA.

Analysis	Variance component	Observed SD	Theoretical SE	reported SE
$REML_{32K}$	$\widehat{h^2}$	0.011	0.010	0.012
	$\widehat{\delta^2}$	0.015	0.014	0.016
	$\widehat{\eta^2}$	0.37	0.399	0.393
HE$_{32K}$	$\widehat{h^2}$	1.4×10^{-2}	9.9×10^{-3}	9.9×10^{-3} (1.8×10^{-2})
	$\widehat{\delta^2}$	1.8×10^{-2}	1.4×10^{-2}	1.5×10^{-2} (2.2×10^{-2})
	$\widehat{\eta^2}$	1.45	1.57	1.55 (2.19)

We performed 100 replicates of simulations based on observed genotypes of 254,679 UKB participants. In order to assess the accuracy of the theoretical and reported SE, 8 sub-datasets of 32K individuals were analyzed for each replicate both with REML ($REML_{32K}$) and HE (HE$_{32K}$). The observed SD have been computed within a replicate and averaged across them for each variance component. Reported SE by GCTA were averaged within and across replicates and the format for HE analysis is $SE_{OLS}(SE_{jackknife})$.