PERSPECTIVE

Concepts, estimation and interpretation of SNP-based heritability

Jian Yang[1,2] , Jian Zeng[1], Michael E Goddard[3,4], Naomi R Wray[1,2] & Peter M Visscher[1,2]

Narrow-sense heritability (h^2) is an important genetic parameter that quantifies the proportion of phenotypic variance in a trait attributable to the additive genetic variation generated by all causal variants. Estimation of h^2 previously relied on closely related individuals, but recent developments allow estimation of the variance explained by all SNPs used in a genome-wide association study (GWAS) in conventionally unrelated individuals, that is, the SNP-based heritability (h^2_{SNP}). In this Perspective, we discuss recently developed methods to estimate h^2_{SNP} for a complex trait (and genetic correlation between traits) using individual-level or summary GWAS data. We discuss issues that could influence the accuracy of $\hat{h}^2_{SNP}$, definitions, assumptions and interpretations of the models, and pitfalls of misusing the methods and misinterpreting the models and results.

Estimation of the variance explained by all SNPs used in a population-based GWAS was initially motivated by the 'missing heritability' problem[1]. The problem was that the estimated variance explained by genome-wide significant (GWS) SNPs discovered in GWAS (denoted $\hat{h}^2_{GWS}$) was only a fraction of the estimated heritability ($\hat{h}^2$) from family or twin studies[2], where $\hat{h}^2_{GWS}$ was estimated in a multi-SNP model to account for linkage disequilibrium (LD) among SNPs and in an independent sample to avoid overestimation due to winner's curse[3]. Taking human height as an example, $\hat{h}^2_{GWS}$ was 5% before 2010 (ref. 4), which is much smaller than a frequently quoted $\hat{h}^2$ of 80% from family or twin studies[5–7]. This raised concerns about the cost-effectiveness of GWAS as an experimental design for the discovery of associated genes[8]. Several explanations of the missing heritability were proposed, including the presence of a large number of common variants of small effect yet to be discovered, rare variants of large effect not tagged by common SNPs on genotyping arrays, and inflation in pedigree-based $\hat{h}^2$ due to shared environmental effects, non-additive genetic variation and/or epigenetic factors[2,9]. The missing heritability question also reignited the debate about the 'common disease, common variant' hypothesis[10], that is, whether the proportion of heritability for

common disease not explained by GWS loci is due to rare variants of large effect not tagged by the current generation of SNP arrays or undetected common variants of small effect[2,11]. It is therefore important to quantify the proportion of variance attributable to all common SNPs (defined here as those with minor allele frequency, MAF ≥ 0.01) used in GWAS. If common SNPs are the major contributor to heritability, then the concern about missing heritability is premature because the extent to which heritability is missing depends on the experimental sample size of GWAS[12].

Estimation of the SNP-based heritability—the GREML approach
SNP-based heritability (or h^2_{SNP}) was initially defined as the proportion of phenotypic variance explained by all SNPs on a genotyping array[13] and is therefore dependent of the number of SNPs on a SNP array. The concept has now been expanded to refer to the variance explained by any set of SNPs, for example, all genetic variants from in-depth whole-genome sequencing (WGS) or imputed from a reference[14]. Yang et al.[13] used a mixed linear model (MLM) approach to estimate h^2_{SNP} in a GWAS data set of unrelated individuals and demonstrated that common SNPs on a genotyping array explain a large proportion (45%) of variance in height. Here 'unrelated individuals' means distantly related individuals rather than individuals with no genetic relatedness, as even random pairs of individuals drawn from a general population would share distant ancestors. Given the small $\hat{h}^2_{GWS}$ (5%) and relatively large $\hat{h}^2_{SNP}$ (45%), it was concluded that, for complex traits like height, there are likely a large number of common variants with effect sizes too small to pass the stringent GWS threshold ($P < 5 \times 10^{-8}$) in GWAS, even with sample sizes that were considered large at that time ($n = 1,000$ to $10,000$ samples before 2010), consistent with a model of polygenic inheritance. It was subsequently predicted that more associated genetic variants could be discovered with larger sample sizes while keeping the same experimental design of GWAS. This prediction has been realized by recent studies with $n > 100,000$ for height, body mass index (BMI), schizophrenia, and many other traits and diseases[15–20]. Under a polygenic model, the amount of heritability unexplained by GWS loci depends on sample size[12]. The aforementioned comparison of 5% versus 80% for height in 2009 (ref. 4) became 16% versus 80% only five years later[15]. Given the nearly linear relationship between the number of GWS loci and the logarithm of sample size ($\log(n)$) observed in published GWAS[12] and the highly polygenic nature of most complex traits[21,22], we predict that the shrinking of the gap between $\hat{h}^2_{GWS}$ and $\hat{h}^2_{SNP}$ will be less than linear with $\log(n)$ because the variance explained by SNPs discovered in studies with larger sample sizes tends to be smaller.

The approach of Yang et al.[13] was subsequently termed genomic relatedness matrix (GRM) restricted maximum likelihood (GREML)[23] and implemented in the GCTA software tool[24] (**Box 1**). GREML shares fea-

[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [2]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia. [3]Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia. [4]Biosciences Research Division, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia. Correspondence should be addressed to J.Y. (jian.yang@uq.edu.au) or P.M.V. (peter.visscher@uq.edu.au).
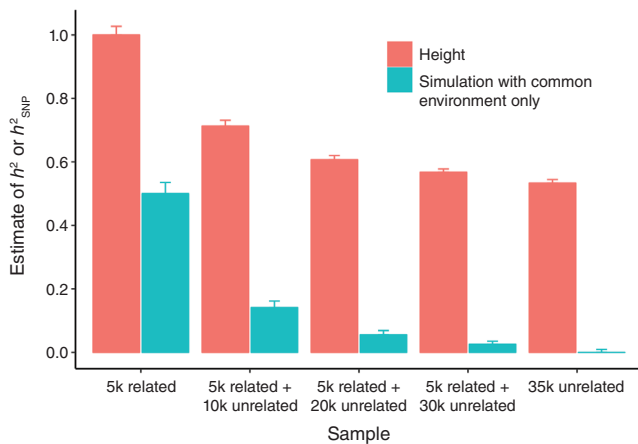
Figure 1 Interpretation of estimated genetic variance depends on ascertainment of the sample. Shown in red are the pedigree-based heritability estimate ($\hat{h}^2_{ped}$) for height from 2,824 pairs of full siblings in the UK Biobank data[61] ("5k related"; sibling correlation = 0.520), $\hat{h}^2_{SNP}$ from a GREML analysis of 35,000 unrelated UK Biobank individuals using all the imputed SNPs in common with HapMap 3 ("35k unrelated") and the estimates in between from GREML analyses in a mixed sample of unrelated individuals and close relatives (part 2 of the **Supplementary Note**). The difference between $\hat{h}^2_{ped}$ and $\hat{h}^2_{SNP}$ demonstrates the genetic variation (due to rare variants in particular) not tagged by common HapMap 3 SNPs and/or confounding in $\hat{h}^2_{ped}$ from common environmental effects and non-additive genetic variation. Shown in blue are the results from the same analyses for a simulated phenotype based on a common environment model without genetic effect (part 2 of the **Supplementary Note**). Each bar is a single estimate, and each error bar indicates the SE of the estimate.

Figure 2 Relationship between SNP-based heritability on the liability scale ($h^2_{SNP(l)}$) and SNP-based heritability estimated from case–control samples. (**a**–**d**) The plots show that the same estimate of $h^2_{SNP(l)}$ of 0.1 (**a**), 0.2 (**b**), 0.4 (**c**) or 0.6 (**d**) on the liability scale can correspond to a wide range of SNP-based heritability estimates from case–control samples on the observed 0–1 scale (part 5 of the **Supplementary Note**), depending on the proportion of cases in the sample (P) and the assumed lifetime risk of disease (K) used to transform the estimates to the liability scale. For each plotted line, the minimum value assumes a population sample with $P = K$. In real application, we advise investigating the sensitivity of estimates of $h^2_{SNP(l)}$ to choice of K, but we find that the impact is small when $K < 0.05$. As shown in **c** and **d**, for a rare disease with high $h^2_{SNP(l)}$, $h^2_{SNP(O)}$ is expected to be larger than 1 because of the nonlinear relationship between genetic variance and phenotypic variance on the observed 0–1 scale.

tures with a pedigree-based analysis (part 1 of the **Supplementary Note**) but is usually applied to a sample of unrelated individuals (note that this is also the usual experimental design for GWAS), and hence $\hat{h}^2_{SNP}$ is unlikely to be confounded by common environmental effects (**Fig. 1**). For pairs of distantly related individuals, the amount of the genome shared is small and highly variable, and it is unlikely that pairs who share slightly more of the genome than average will also have greater sharing of environment in a relatively homogeneous population. The use of unrelated individuals also means that $\hat{h}^2_{SNP}$ is unlikely to be contaminated with contributions from non-additive genetic effects, as the correlation between the additive and non-additive genetic relationships is tiny, whereas such contamination could be a problem in $\hat{h}^2$ estimated from families depending on the study design. In addition, GREML can be applied to family data, but the resulting estimates should be interpreted with caution (part 3 of the **Supplementary Note**).

The GREML estimate directly quantifies the proportion of phenotypic variance explained by all SNPs used in GWAS and therefore provides the upper limit of $\hat{h}^2_{GWS}$ given the same experimental design. The information to estimate h^2_{SNP} comes from very small coefficients of genetic relationship for pairs of individuals, but small standard error (SE) for $\hat{h}^2_{SNP}$ (part 4 of the **Supplementary Note**) can be achieved because of the large number of pairwise relationships (for example, 50 million for a study using 10,000 individuals), although these pairs are not independent. Subsequent work has extended the method to estimate h^2_{SNP} in disease data[25] (part 5 of the **Supplementary Note**) and genetic correlation (r_g) between traits[26,27] (part 6 of the **Supplementary Note**). There are several caveats to estimating h^2_{SNP} using data from case–control studies (part 5 of the **Supplementary Note**) and interpreting the estimates on different scales (**Fig. 2**).

Multiple terms and notations that have been used to describe the parameter estimated by GREML. We recommend using the term
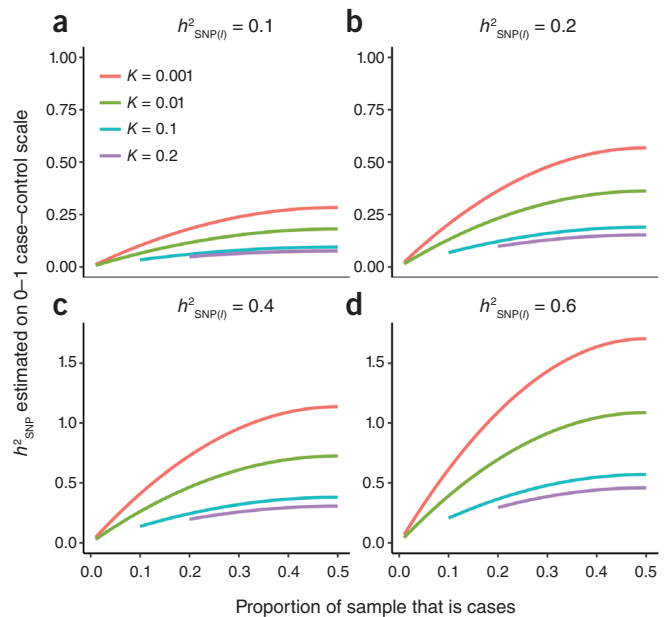
'SNP-based heritability' and the notation h^2_{SNP}. Unlike h^2, which is a population-level parameter irrespective of experimental design, h^2_{SNP} is a parameter given a set of SNPs. We likewise believe that it is also necessary to use a specific notation, $\hat{h}^2_{ped}$, to represent h^2 estimated from pedigrees (including twins) because of the potential biases in pedigree-based $\hat{h}^2$ due to confounding factors such as common environmental effects. We have shown above that h^2_{SNP} is by definition smaller than h^2 because not all causal variants, in particular those with low frequency, can be perfectly tagged by SNPs used in GWAS (**Fig. 3a** and part 1 of the **Supplementary Note**). Here, by 'causal variant', we mean a genetic mutation that causes a different cascade of events in biological pathways and consequent phenotypic change rather than an associated variant identified from GWAS. In the particular case where h^2_{SNP} is defined as the variance explained by all such causal variants, $h^2_{SNP} = h^2$. In reality, however, causal variants are unknown. An unbiased estimate of h^2 might be achieved by estimating h^2_{SNP} from in-depth WGS data assuming that all causal variants have been sequenced and that there is no difference in LD between causal variants and other sequence variants[14] (see below for more discussion).

Both GWAS and estimation of h^2_{SNP} by GREML use LD

GWAS relies, by design, on genotyped common SNPs tagging unknown causal variants in the same chromosomal region. Estimating how much trait variation is tagged when fitting all SNPs simultaneously also makes use of the LD between SNPs and unobserved causal variants. A sparse
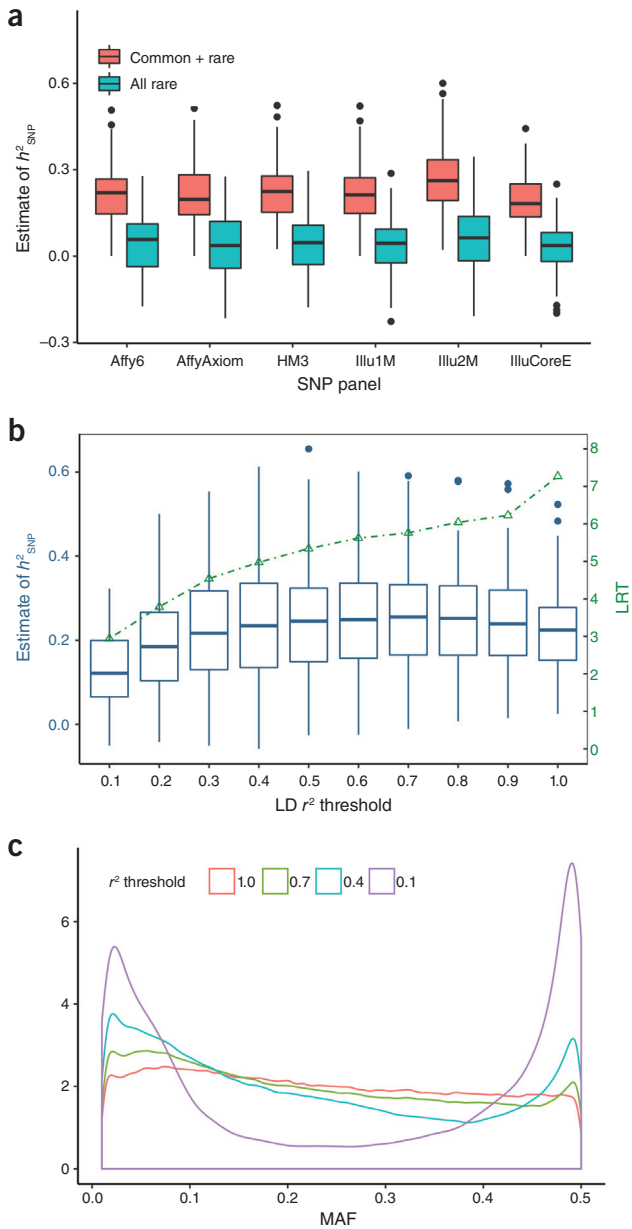
Figure 3 Estimation of genetic variance depends on ascertainment of SNPs and genetic architecture. (**a**) Estimates of $\hat{h}^2_{SNP}$ using SNPs on six different SNP panels for a simulated trait under two scenarios: (i) causal variants are random, with both common and rare variants (red), and (ii) causal variants are rare (blue) (see part 7 of the **Supplementary Note** for details of the simulation). The six SNP panels are the Affymetrix 6.0 array (Affy6), Affymetrix Axiom array (AffyAxiom), HapMap 3 Project (HM3), Illumina OmniExpress array (Illu1M), Illumina Omni2.5 array (Illu2M) and Illumina CoreExome array (IlluCoreE). (**b**) Effect of LD pruning on $\hat{h}^2_{SNP}$ and the likelihood-ratio test (LRT) statistic. LD pruning was performed on the basis of HapMap 3 SNPs in PLINK (–indep-pairwise 50 5 r^2) with the LD r^2 threshold shown on the x axis. The last column with an r^2 threshold of 1 represents the result without LD pruning (with all HapMap 3 SNPs). GREML analyses were performed using common SNPs on the HapMap 3 panel. (**c**) Distribution of MAFs of HapMap 3 variants after LD pruning with different r^2 thresholds (no pruning for the r^2 threshold of 1.0). In the box plots shown in a and b, the band inside the box is the median; the bottom and top of the box are the first and third quartiles, respectively (Q1 and Q3); the lower and upper whiskers are Q1 − 1.5 IQR and Q3 + 1.5 IQR, respectively, where IQR = Q3 − Q1; and the dots are the data not included between the whiskers.

Box 1 Statistical model used in the GREML approach to estimate h^2_{SNP}

The statistical model used by GREML can be described in its simplest form as

$$\mathbf{y} = \mathbf{Wu} + \mathbf{e}$$

where **y** is an n x 1 vector of standardized phenotypes with n equal to the sample size, $\mathbf{W} = \{w_{ij}\}$ is an n x m standardized SNP genotype matrix where m is the number of SNPs, $\mathbf{u} = \{u_i\}$ is an m x 1 vector of the additive effects of all variants when fitted jointly in the model, $\mathbf{u} \sim N(0, \mathbf{I}\sigma^2_u)$ with **I** being an identity matrix, and **e** is a vector of residuals, $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$. An equivalent model is

$$\mathbf{y} = \mathbf{g} + \mathbf{e}$$

where $\mathbf{g} \sim N(0, \mathbf{A}\sigma^2_{g(SNP)})$ with $\sigma^2_{g(SNP)}$ being the additive genetic variance captured by SNPs, $\sigma^2_{g(SNP)} = m\sigma^2_u$, $\mathbf{A} = \mathbf{WW}'/m$ and $h^2_{SNP} = \sigma^2_{g(SNP)} / [\sigma^2_{g(SNP)} + \sigma^2_e]$. The parameters to be estimated are $\sigma^2_{g(SNP)}$ (or σ^2_u) and σ^2_e. The matrix **A** describes the variance–covariance structure of the random effects **g** and is assumed to be known in the estimation process. In practice, **A** is called the SNP-derived genetic (or genomic) relationship matrix (GRM) and is estimated from the SNP data. The estimate of $\sigma^2_{g(SNP)}$ from GREML can be described as the estimated variance explained by all the SNPs ($m\hat{\sigma}^2_u$) or equivalently as the estimated genetic variance by contrasting the phenotypic similarity between unrelated individuals to their SNP-derived genetic similarity[13,58,59].

SNP array that does not cover common variation in the genome well is less likely to lead to the discovery of trait-associated variants (even with a large sample size), and fitting those SNPs together in a GREML analysis will result in a smaller proportion of phenotypic variance explained than with a denser SNP array (**Fig. 3a**). Because the maximum possible LD correlation between two genetic variants declines as their difference in MAF increases[28], genetic variation at rare variants (MAF < 0.01) is unlikely to be well tagged by common SNPs on genotyping arrays (**Fig. 3a**). If causal variants are located in genomic regions with a different LD property from the rest of the genome, this can lead to bias in $\hat{h}^2_{SNP}$ (refs. 14,29,30; see below for more discussion).

Interpretation and misinterpretation of the GREML model

There are several circumstances where the principle of GREML is misinterpreted and the method is misapplied, and this could potentially lead to misleading or confusing inference. GREML is based on a random-effect model (**Box 1**). If the number of SNPs (m) is smaller than the sample size (n), this model is similar to a linear regression analysis (fixed-effect model) in terms of estimating h^2_{SNP} (note that the adjusted R^2 from multiple regression is an unbiased estimate of variance explained in a fixed-effect model). Such a hypothetical experiment would not rely on selecting SNPs to be individually GWS, nor would it rely on assumptions about the genetic architecture. In either a linear regression or random-effect model, the effect sizes of SNPs are fitted jointly (therefore accounting for LD among SNPs), meaning that the effect of any SNP is interpreted as the effect size of this SNP conditioning on the joint effects of all other SNPs. In GWAS, m is normally larger than n, in which case there is no unique solution to the fixed-effect model, a well-known overfitting problem in statistics. In a random-effect model, there is an additional assumption that the joint SNP effects $\mathbf{u} = \{u_1, u_2, ..., u_m\}$ follow a normal distribution with mean 0 and variance σ^2_u (see **Box 1** for notations) so that the
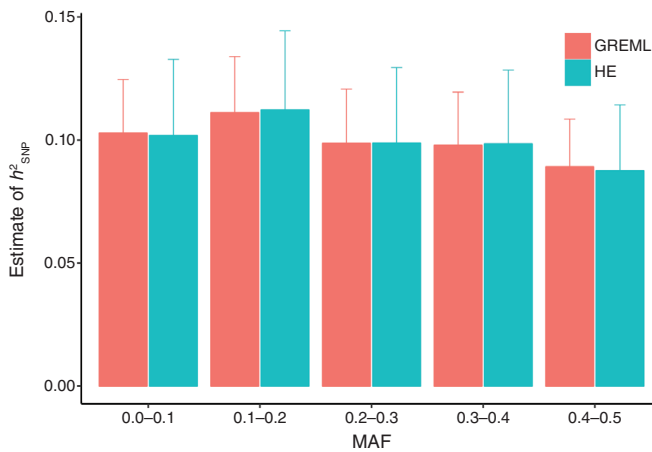
Figure 4 Multiple-component GREML or HE regression for sets of SNPs stratified by MAF. Results are shown as $\hat{h}^2_{SNP}$ with SE (error bar) in each MAF group averaged over 200 simulation replicates using ~11,500 unrelated individuals (SNP-based relatedness < 0.05) and ~550,000 genotyped SNPs after standard quality controls. In each simulation replicate, 1,000 SNPs were selected at random as causal variants with their effects sampled from a standard normal distribution with mean 0 and variance 1. The true heritability was 0.5 (roughly 0.1 per MAF bin). The SE of the estimate from HE regression was calculated using the jackknife approach where one individual was left out at a time.

model parameters are estimable even when m is larger than n, where σ^2_u is interpreted as per-SNP genetic variance when all SNPs are fitted jointly, hence accounting for LD[31]. Therefore, σ^2_u is not consistent across models having different numbers of SNPs. There is a misunderstanding that GREML does not account for LD because it does not have a covariance matrix for **u** (ref. 32). This is incorrect. In fact, the LD correlations among SNPs have been modeled by fitting the SNP genotype matrix **W**, similar to that in linear regression analysis[31]. Because σ^2_u is the variance of a SNP effect conditioning on the joint effects of all other SNPs and w_{ij} is the standardized SNP genotype, the additive genetic variance captured by all SNPs is $\sigma^2_{g(SNP)} = m\sigma^2_u$ (**Box 1**).

In part 8 of the **Supplementary Note**, we list five scenarios where GREML (or the GCTA tool) is misused, resulting in potentially misleading results. In addition, there is often a question about whether the SNPs included in GREML analysis need to be pruned for LD. As discussed above, GREML accounts for LD so that LD pruning is not necessary (but see the later discussion on bias due to the nonrandom distribution of causal variants with respect to LD). LD pruning using a high r^2 threshold might increase the estimate, but the likelihood of the model is not improved as compared to that without LD pruning (**Fig. 3b**). Caution is needed in interpreting the GREML estimate from pruned SNPs because of the change in the MAF spectrum of SNPs resulting from LD pruning (**Fig. 3c**). Changing the set of SNPs means that the underlying parameter being estimated (h^2_{SNP} for a set of LD-pruned SNPs) is different from the original parameter (h^2_{SNP} for all SNPs).

Bias due to the nonrandom distribution of causal variants with respect to LD

We have mentioned above that $\hat{h}^2_{SNP}$ from WGS data could be a biased estimate of h^2 if the LD property of causal variants is different from that of the other variants[14,29,30,33]. The unbiasedness of GREML in estimating h^2 using WGS data depends on the ratio of $\overline{r^2_{MQ}}$ (mean LD r^2 between causal and non-causal variants) to $\overline{r^2_{MM}}$ (mean LD r^2 between non-causal variants)[14]. Note that, because r^2 is a function of MAF, a

difference in MAF spectrum between causal and non-causal variants will lead to a difference in LD (MAF-mediated LD bias), resulting in a bias in $\hat{h}^2_{SNP}$. One solution is to stratify SNPs by MAF (MAF-stratified GREML, GREML-MS)[14,34,35], which reduces bias in the estimate due to MAF-mediated LD bias. However, a more general approach is to not rely on a specific model of the interplay between allele frequency, effect size and LD, but instead stratify SNPs by MAF and LD jointly and estimate genetic variance with MAF–LD subsets. This approach, termed GREML-LDMS, appears to provide unbiased estimates of h^2 as well as the contributions of common and rare variants to h^2 in simulations based on WGS data, regardless of the underlying genetic architecture and distribution of causal variants with respect to MAF and LD[14,36]. We recommend the use of GREML-LDMS to estimate h^2_{SNP} in imputed data (part 9 of the **Supplementary Note**). The applications of GREML-LDMS to WGS data sets with rich phenotypes in the future will be able to provide nearly unbiased estimates of h^2 in unrelated individuals and quantify the variance explained by all rare variants for a range of complex traits. However, large sample sizes are required to estimate h^2_{SNP} with useful precision because var($\hat{h}^2_{SNP}$) depends on sample size and variant density[37] (part 4 of the **Supplementary Note**); for example, a sample size of ~33,000 is needed to obtain an SE of 0.02 for WGS data.

Speed et al.[29] proposed a method called LDAK to correct for the LD bias. The basic idea is to weight each SNP by a factor inversely proportional to its LD with SNPs nearby. This weighting strategy can introduce MAF bias because it gives more weight to SNPs with lower MAF (Supplementary Fig. 2 of Yang et al.[14]), as LD is a function of MAF[28]. The LDAK model implicitly assumed that the variance explained by a rare variant (for example, 0.001 < MAF < 0.01) is more than ten times larger than that explained by a common variant (for example, 0.1 < MAF < 0.5) (based on the LDAK weights calculated from a sequenced reference set[14]). This is an unrealistic model because it predicts that the power to detect rare variants would be orders of magnitude higher than that to detect common variants, a prediction not consistent with empirical results in the cases of human height[15,38], schizophrenia[17,39] and type 2 diabetes[40]. The LDAK-induced MAF bias can be substantial, especially when there is a large number of rare variants (as in a WGS data set), leading to an inflated estimate of h^2_{SNP} (ref. 14).

The LDAK model has recently been changed substantially[41]. Two new parameters have been added: one is a weighting according to MAF and the other is a weighting according to imputation accuracy. Although it is not the justification for these two new parameters, both give more weight to common variants than the original LDAK model[41]. The revised LDAK model is now more similar to GREML-LDMS[14], but not identical, as Speed et al.[41] estimate a higher SNP-based heritability from their empirical analyses on a range of traits. In simulation studies to compare the methods, the results depend on the model used to simulate the data. Unfortunately, we cannot be sure which is the closer-fitting model for any given trait. GREML-LDMS makes fewer assumptions about the relationship between causal variants, LD and MAF and thereby appears to be more robust than the revised LDAK method[36], although at the expense of estimating more parameters. On balance, we conclude that this topic merits further investigation[36], as the relationship between local LD, locus heterozygosity and additive genetic variance for complex traits has not yet been resolved, and indeed may differ across the genome and between traits.

Assumptions about the relationship between effect sizes and allele frequencies

Under an evolutionarily neutral model, the proportion of variance in a polygenic trait explained by all variants in a MAF bin is linearly proportional to the width of the MAF bin[14] (the variance explained by a

rare variant, on average, is tiny, but there are a large number of them). Therefore, a significant deviation of the observed variance explained in a MAF bin from the expected value is evidence that the trait has been under natural selection[14,42]. In GCTA-GREML, we standardize the SNP genotypes and assume that the effect size per standardized genotype (u_i) follows a normal distribution. This implicitly assumes a larger per-allele effect (b_i) for a SNP with lower MAF, consistent with a model of purifying selection where variants with larger effect sizes tend to be under higher selection and therefore are more likely to be at lower frequencies (for example, MAF < 0.1). There is an option in GCTA to run GREML assuming that effect size is independent of MAF (neutral model). However, the difference between the two models is trivial in GREML-MS analysis[14]. Moreover, GREML-MS allows the data to reveal the relationship between variance explained and MAF. One of the important extensions of GREML in the future is to estimate directly from the data a parameter to quantify the relationship between b_i and allele frequency while fitting a mixture distribution to the joint effects of SNPs[43] (part 10 of the **Supplementary Note**).

Comparison with HE regression

As described in **Box 1**, the GREML analysis is based on an MLM that is equivalent to fitting the additive genetic values of all individuals, that is, $\mathbf{y} = \mathbf{g} + \mathbf{e}$ with $\mathrm{var}(\mathbf{y}) = \mathbf{A}\sigma^2_{g(\mathrm{SNP})} + \mathbf{I}\sigma^2_e$. The variance components in this model are usually estimated using the REML approach. However, the REML algorithm is computationally intensive (part 11 of the **Supplementary Note**). Alternatively, $\sigma^2_{g(\mathrm{SNP})}$ can be estimated from Haseman–Elston (HE) regression[37,44], that is, $y_i y_j = b_0 + b_1 A_{ij} + e_{ij}$, where $b_1 = \sigma^2_{g(\mathrm{SNP})}$. The performance of GREML has been compared using extensive simulations in Golan *et al.*[45] in ascertained case–control studies where GREML estimates can be biased, especially when m/n is small and disease prevalence is low. We also performed simulation to compare the two methods with an emphasis on the SE under a polygenic model (part 12 of the **Supplementary Note**). HE regression is computationally much more efficient but slightly less powerful than REML, as the SE of $\hat{h}^2_{\mathrm{SNP}}$ from HE regression is larger than that from REML (**Supplementary Table 1** and part 12 of the **Supplementary Note**). The small difference in SE between the methods might not be important when the sample size becomes very large. For example, given $\hat{h}^2_{\mathrm{SNP}} > 0.1$, whether the SE is 0.01 (REML) or 0.015 (HE regression) does not make any difference in statistical inference of whether $h^2_{\mathrm{SNP}} = 0$. HE regression can also be used to estimate multiple genetic components, for example, multiple sets of SNPs stratified by MAF or chromosome (**Fig. 4**), or to estimate genetic correlations between traits (**Supplementary Table 2**). These analyses have been implemented in the latest version of GCTA (see URLs). In addition, phenotype correlation–genotype correlation (PCGC) regression is an implementation of HE regression designed for disease data to attenuate the biases in ascertained case–control studies[22,45] (see URLs).

Non-additive genetic variation

The GREML approach has been extended to estimate dominance genetic variance tagged by SNPs in unrelated individuals on the basis of a classical quantitative genetics model[46]. Similar to the additive GREML method, the dominance GREML model fits the additive and dominance effects of all SNPs as two sets of random effects in an MLM. This is an orthogonal model because the additive and dominance genotype variables and, thereby, the additive and dominance GRMs are independent. On average, across 79 quantitative traits, additive genetic variation explained ~15% of the phenotypic variance and dominance genetic variation explained ~3% of the variance[46]. The ratio of additive to dominance variance is consistent with what is expected from theory[47]. The

method can be further extended to estimate genetic variance attributable to epistasis[48] on the basis of the classical quantitative genetics model[49], $y = g_A + g_D + g_{AA} + g_{AD} + g_{DD} + e$, where g_A and g_D are the additive and dominance genetic values of an individual and g_{AA}, g_{AD} and g_{DD} are additive-by-additive, additive-by-dominance and dominance-by-dominance epistatic genetic values, respectively. However, the sample size will need to be very large to get a precise estimate of epistatic variance because the variance in the epistatic genetic relationship between unrelated individuals is very small. For instance, the genetic relationship for g_{AA} is A_{ij}^2, which has a variance of $2[\mathrm{var}(A_{ij})]^2$ (ref. 49). For HapMap 3 SNPs, $\mathrm{var}(A_{ij}) \approx 2.0 \times 10^{-5}$ so that the variance in genetic relationship for g_{AA} is ~1.0×10^{-9}, meaning that over 1 million unrelated individuals will be needed to estimate the variance explained by g_{AA} with SE <0.05 (>4 million unrelated individuals to get SE <0.01). The variance in the dominance genetic relationship is smaller than for the additive genetic relationship. Therefore, it will be even more difficult to estimate variance for g_{AD} or g_{DD}.

Estimating h^2_{SNP} and r_g from GWAS summary data

We have discussed above the MLM-based approaches to estimate h^2_{SNP} using individual-level GWAS data. There are other methods that are able to estimate h^2_{SNP} from GWAS summary data (estimated SNP effects and their SE for all SNPs analyzed in a study)[50]. For example, the AVENGEME method uses maximum likelihood to estimate the genetic variance of a trait, the proportion of genetic variants affecting the trait and the genetic covariance (and therefore genetic correlation) between traits from the test statistic for association between phenotype and polygenic risk score (PRS)[51,52]. We can also estimate h^2_{SNP} directly from summary data using the deviation of the observed χ^2 test statistic for a SNP from its expected value under the null hypothesis of no association[53] (part 13 of the **Supplementary Note**). This is the basic principle of the recently developed LD score regression approach (LDSC)[54]. This approach requires only summary data from GWAS because LD scores can be estimated from a reference sample (for example, the 1000 Genomes Project). LDSC has been extended to estimate r_g between traits using summary data[55], which allows the traits to be measured on different samples regardless of whether there is an overlap between samples, and to partition $\hat{h}^2_{\mathrm{SNP}}$ by functional annotation[56]. This method provides great flexibility for researchers to estimate r_g between any two GWAS data sets. Both GREML and LDSC aim to estimate the variance explained by all SNPs used in GWAS. However, there are distinct differences between the two methods. LDSC is orders of magnitude faster than GREML, and the computing time for LDSC does not scale up with sample size. LDSC only requires summary data, which allows the reanalysis of summary data available from published meta-analyses. There are also limitations for LDSC. LDSC is not applicable in estimating the variance explained by rare variants (for example, MAF < 0.01) using either imputed or WGS data[36] nor the variance explained by SNPs in small genomic regions (although the latter has been overcome by the HESS method developed recently[53]), and it is more sensitive to the genetic architecture of a trait (**Supplementary Table 3**). A previous study showed that $\hat{h}^2_{\mathrm{SNP}}$ estimates from LDSC are consistently smaller than those from GREML in the same data set[57], which is likely owing to errors in LD scores estimated from the reference (by default, LDSC uses LD scores from HapMap 3 SNPs in the 1000 Genomes Project). We therefore advise using LD scores from the data used to generate the GWAS summary statistics. Although this may not be possible for published summary statistics, it should be possible for large cohorts such as the UK Biobank. It is noteworthy that LDSC will suffer bias in a similar way as GREML if causal variants are not randomly distributed with respect to LD. The estimate of r_g from bivariate LDSC is consistent with that

Box 2 Future applications of SNP-based heritability to large data sets

The methods for estimating h^2_{SNP} can be extended and applied to large data sets in the future. These future directions include:

1. Applications of GREML-LDMS or similar approaches (that account for bias in $\hat{h}^2_{SNP}$ due to LD bias) to in-depth WGS data to obtain nearly unbiased estimates of h^2 for a range of complex traits and quantify the variance attributable to all rare variants;

2. Methods that provide an unbiased estimate of h^2 from identity-by-descent information inferred from SNP array data[60];

3. Methods to estimate $\hat{h}^2_{ped}$ from pedigree data accounting for common environmental effects and assortative mating;

4. Fast Bayesian MLM approaches based on flexible models that are applicable to WGS data to estimate the distribution of effect sizes of all variants;

5. Methods to estimate h^2_{SNP} free of assumptions about the relationship between per-allele effect and allele frequency[43].

from bivariate GREML, but the jackknife SE of $\hat{r}_g$ from LDSC is larger than that expected from the approximation theory[37,55,57].

Summary

We have provided a perspective of the methods for estimating SNP-based heritability in unrelated individuals using GWAS data. We emphasized that the GREML approach accounts for LD when estimating h^2_{SNP} and actually uses LD to tag causal variants if they are not observed. We discussed the concepts and assumptions of the methods and scenarios under which the estimates could be biased, the methods could be misused and the results could be misinterpreted. We further discussed the extensions and applications of the methods in large data sets in the future (**Box 2**). These future directions could expand understanding of the genetic architecture for human complex traits and inform the design of future experiments to fully dissect genetic variation and genetic correlations.

1. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
2. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
3. Xiao, R. & Boehnke, M. Quantifying and correcting for the winner's curse in genetic association studies. *Genet. Epidemiol.* **33**, 453–462 (2009).
4. Visscher, P.M. Sizing up human height variation. *Nat. Genet.* **40**, 489–490 (2008).
5. Fisher, R.A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
6. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).
7. Macgregor, S., Cornes, B.K., Martin, N.G. & Visscher, P.M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.* **120**, 571–580 (2006).
8. Goldstein, D.B. Common genetic variation and human traits. *N. Engl. J. Med.* **360**, 1696–1698 (2009).
9. Eichler, E.E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–450 (2010).
10. Schork, N.J., Murray, S.S., Frazer, K.A. & Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **19**, 212–219 (2009).
11. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
12. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
13. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
14. 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).
15. Wood, A.R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
16. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
17. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
18. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
19. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
20. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48**, 1162–1170 (2016).
21. Yang, J. *et al.* Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. *PLoS Genet.* **9**, e1003355 (2013).
22. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
23. Benjamin, D.J. *et al.* The genetic architecture of economic and political preferences. *Proc. Natl. Acad. Sci. USA* **109**, 8026–8031 (2012).
24. 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).
25. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
26. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. & Wray, N.R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism–derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).
27. Lee, S.H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
28. Wray, N.R. Allele frequencies and the r^2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res. Hum. Genet.* **8**, 87–94 (2005).
29. Speed, D., Hemani, G., Johnson, M.R. & Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).
30. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet.* **9**, e1003993 (2013).
31. Yang, J., Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proc. Natl. Acad. Sci. USA* **113**, E4579–E4580 (2016).
32. Krishna Kumar, S., Feldman, M.W., Rehkopf, D.H. & Tuljapurkar, S. Limitations of GCTA as a solution to the missing heritability problem. *Proc. Natl. Acad. Sci. USA* **113**, E61–E70 (2016).
33. 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).
34. Lee, S.H. *et al.* Estimation of SNP heritability from dense genotype data. *Am. J. Hum. Genet.* **93**, 1151–1155 (2013).
35. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
36. Evans, L. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *bioRxiv* https://dx.doi.org/10.1101/115527 (2017).
37. Visscher, P.M. *et al.* Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* **10**, e1004269 (2014).
38. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).

39. Purcell, S.M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).

40. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

41. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S. & Balding, D.J. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).

42. Gazal, S. *et al.* Linkage disequilibrium dependent architecture of human complex traits reveals action of negative selection. *bioRxiv* https://dx.doi.org/10.1101/082024 (2016).

43. Zeng, J. *et al.* Widespread signatures of negative selection in the genetic architecture of human complex traits. *bioRxiv* https://dx.doi.org/10.1101/145755 (2017).

44. Haseman, J.K. & Elston, R.C. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19 (1972).

45. Golan, D., Lander, E.S. & Rosset, S. Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci. USA* **111**, E5272–E5281 (2014).

46. Zhu, Z. *et al.* Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* **96**, 377–385 (2015).

47. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).

48. Rönnegård, L., Pong-Wong, R. & Carlborg, O. Defining the assumptions underlying modeling of epistatic QTL using variance component methods. *J. Hered.* **99**, 421–425 (2008).

49. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).

50. Pasaniuc, B. & Price, A.L. Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.* **18**, 117–127 (2017).

51. Palla, L. & Dudbridge, F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* **97**, 250–259 (2015).

52. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).

53. Shi, H., Kichaev, G. & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *Am. J. Hum. Genet.* **99**, 139–153 (2016).

54. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

55. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

56. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

57. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Hum. Mol. Genet.* **24**, 7445–7449 (2015).

58. Lynch, M. & Ritland, K. Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753–1766 (1999).

59. Hayes, B.J., Visscher, P.M. & Goddard, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)* **91**, 47–60 (2009).

60. Browning, B.L. & Browning, S.R. A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* **88**, 173–182 (2011).

61. 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**, e1001779 (2015).

EDITORIAL SUMMARY

Jian Yang and colleagues explore the uses and abuses of heritability estimates derived from pedigrees and from GWAS SNPs and make recommendations for best practice in future applications of SNP-based heritability.