A Commentary on 'Common SNPs Explain a Large Proportion of the Heritability for Human Height' by Yang et al. (2010)

Peter M. Visscher,[1] Jian Yang[1] and Michael E. Goddard[2,3]

[1] *Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Department of Food and Agricultural Systems, University of Melbourne, Australia*

[3] *Biosciences Research Division, Department of Primary Industries, Victoria, Melbourne Australia*

Recently a paper authored by ourselves and a number of co-authors about the proportion of phenotypic variation in height that is explained by common SNPs was published in *Nature Genetics* (Yang et al., 2010). Common SNPs explain a large proportion of the heritability for human height (Yang et al.). During the refereeing process (the paper was rejected by two other journals before publication in *Nature Genetics*) and following the publication of Yang et al. (2010) it became clear to us that the methodology we applied, the interpretation of the results and the consequences of the findings on the genetic architecture of human height and that for other traits such as complex disease are not well understood or appreciated. Here we explain some of these issues in a style that is different from the primary publication, that is, in the form of a number of comments and questions and answers. We also report a number of additional results that show that the estimates of additive genetic variation are not driven by population structure.

Keywords: GWAS, heritability, height, linkage disequilibrium

Rationale for the Study

Genome-wide association studies (GWAS) have found hundreds of SNPs that are significantly associated with complex traits such as height (Gudbjartsson et al., 2008; Lango Allen et al., 2010; Lettre et al., 2008; Weedon et al., 2007) and diseases such as age-related macular degeneration (Jakobsdottir et al., 2009; Maller et al., 2006). However, in most cases, the published SNPs reliably associated with a trait explain only a small proportion of the known genetic variance. For instance, the heritability of human height is about 80% (Fisher, 1918; Visscher et al., 2008) but the published SNPs that are significantly associated with height explain only ~10% of the phenotypic variance (Lango Allen et al., 2010). This has been called the 'missing heritability' problem (Maher, 2008). We proposed two, not mutually exclusive, hypotheses that could explain this missing heritability. It could be that the SNPs used in GWAS explain some or all of the additive genetic variance but most of them have such a small effect that they are not significant and therefore not reported. Alternatively, it could be that some or all of the mutations causing variation in height are not in perfect linkage disequilibrium (LD) with any of the SNPs and therefore part of the genetic variance is undetected by the SNPs. We provided evidence to support both hypotheses.

What was the Purpose of the Study?

The purpose of the study was to estimate the proportion of variation in height that is captured by the SNPs that are used in GWAS. Our study differs from published GWAS in that we estimate the total variance explained by the SNPs without focussing on individual SNPs. Consequently, our estimate is not diminished by the failure of individual SNP effects to reach a significance threshold. If most causal variants for human height have such low frequency in the population that they are not in LD with the (common) SNPs on the commercial SNP arrays then the method we used would not detect much more additional variance than already accounted for by the published genome-wide significant loci. If, however, there are many causal variants that are in LD with the common SNPs but the effect sizes are too small to be detected with genome-wide significance, then our method would pick up their contribution to additive genetic variation.

What Were the Main Results?

We found that the SNPs explain ~45% of the phenotypic variance (Yang et al., 2010). This is substantially

Received 31 August, 2010; accepted 15 September, 2010.

Address for correspondence: Peter Visscher, Queensland Statistical Genetics Laboratory, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia. E-mail: Peter.visscher@qimr.edu.au

more than the ~10% explained by published, significant SNPs but less than the heritability of 80%. Thus the SNPs track approximately half of the known additive genetic variance. The difference between 10% and 45% is due to many SNPs with such small effects that they are not individually significant in GWAS. However, about half the genetic variance is left unaccounted for. We showed that this amount of missing heritability is expected if the mutations causing variation in height are similar to SNPs with minor allele frequency (MAF) < 0.1. The causal variants are expected to have lower MAF than common SNPs because they are more likely to be subject to some form of natural selection that leads to variants negatively associated with reproductive fitness to be at low frequency.

Our study is the first to show that at least half of the heritability for height (typically estimated using twin and family studies) is captured by common SNPs.

What was the Experimental Design?

Our design was the same as that used for GWAS: a sample from a population of individuals not knowingly related to each other were measured for height and genotyped for 300,000 to 600,000 SNPs.

What was the Analysis?

We fit a statistical model for height that includes all the SNPs but the SNP effects (b_i) are treated as random variables from a distribution with variance σ_b^2. The model is $y = \sum w_i b_i + e$ where y is the phenotypic value, b_i is the effect of the i-th SNP, $w_i = (x_i - 2p_i) / \sqrt{[2p_i(1 - p_i)]}$ with p_i the allele frequency and x_i the genotype indicator of the i-th SNP (x_i = 0, 1 or 2), and e is a random environmental effect. The scaling factor w_i is chosen because $\text{var}(w_i b_i) = \text{var}(w_i)\text{var}(b_i) = \text{var}(x_i - 2p_i)\ \text{var}(b_i) / [2p_i(1 - p_i)] = \text{var}(b_i) = \sigma_b^2$, since $\text{var}(x_i - 2p_i) = 2p_i(1 - p_i)$ under Hardy-Weinberg proportions of genotype frequencies. Using matrix notation this can be written y = Wb + e. We implement this analysis by an equivalent model in which y = g + e where g = Wb is a vector of genetic values calculated from the SNP alleles each individual carries, and $\text{var}(g) = WW'\sigma_b^2$. WW' is a matrix of the relationships between all the individuals calculated from the SNPs (Goddard, 2009; Hayes et al., 2009). The variance of the genetic effects (the effects in vector **g**) in this model is the same as the variance explained by all the SNPs together in the original model that fits SNP effects directly. In the twin (or behavior genetics) literature the fitted model y = g + e would be called an AE model (with g a vector of latent additive genetic values), but with the difference that the additive relationships between individuals do not come from pedigree data but are estimated from marker data (and we do not use close relationships). The statistical equivalence of the two models (fitting SNP effects or fitting whole genome additive genetic effects) means that the inference (and, e.g., likelihood) of the two models is identical. For example, one could predict the effects of genome-wide additive values and then transform the estimates to those for individual SNP effects (Strandén and Garrick, 2009; VanRaden , 2008).

How Does This Study Relate to Traditional Designs to Estimate Genetic Variation?

Traditional designs to estimate genetic variation are based upon the resemblance between relatives in pedigrees. From pedigree data we can derive the expected proportion of the genome that is shared between relatives, from probability theory (Lynch & Walsh, 1998). The probabilities that relatives share alleles that are identical-by-descent (IBD) are calculated with respect to a base (reference) population, which is usually defined as the founders in the pedigree. If the model of analysis is correct then the traditional design based upon the pedigree gives an unbiased estimate of the additive genetic variance. The unbiasedness does not depend on the genetic architecture of the trait (e.g., common or rare variants) because the IBD coefficients calculated from the pedigree are the correct probabilities of sharing alleles by descent, whether common or rare. Our design was the same as that used for GWAS: a sample of a population of individuals that are not knowingly related to each other. Therefore it differs from traditional methods of estimating heritability in that we used supposedly unrelated individuals. However, all members of a species are related to some degree because they share common ancestors. We use SNPs to trace small chromosome segments back to a common ancestor in the distant past and hence estimate the relationship between individuals as tracked by the SNPs.

It is also possible to estimate genetic variance from an analysis of within-family segregation in a known pedigree, by correlating actual relationships with phe-

Table 1

Estimates of the Variance Explained by the SNPs on Even Chromosomes from 10 Simulation Replicates

Replicate	h^2	SE
1	0.045	0.055
2	0.025	0.057
3	0.0	0.058
4	0.0	0.057
5	0.0	0.059
6	0.0	0.056
7	0.057	0.056
8	0.0	0.062
9	0.0	0.057
10	0.0	0.054

Note: A total of 1,000 causal variants were simulated on the odd chromosomes, with a total heritability of 0.8. Genetic variance was estimated from a relationship matrix constructed from all SNPs on the even chromosomes. The same genotypes were used as in Yang et al. (2010). If there is population structure then estimated relatedness on the even chromosomes is correlated with relatedness on the odd chromosomes (where the causal variants are simulated) and therefore genetic variance will be associated with the even chromosomes.

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press

notypic similarity. We performed such analyses using sibling pairs, estimating genome-wide IBD (actual or realised relationships) from microsatellite markers, and estimated a heritability of ~80% (Visscher et al., 2007; Visscher et al., 2006). As in pedigree analyses, these analyses capture the contribution of all variants, common or rare in the population, because the IBD probabilities are correct for all variants.

What are the Main Innovations in the Paper?

There are a number of new elements in the paper that had not been used before in the human genetics literature. They include (1) The definition of 'relatedness' with respect to a base population that is the current population (Powell et al., 2010), (2) Estimating the genetic variance explained by the SNPs by using a statistically equivalent model with relationships calculated from SNP genotypes, (3) Allowing for imperfect LD between the common SNPs and causal variants of a given MAF spectrum.

How Do You Allow for Imperfect LD Between Common SNPs and Causal Variants?

When we found that only half the genetic variance for height was explained by the genotyped SNPs we reasoned that might be due to imperfect LD between SNPs and causal variants. For the case of a single genotyped SNP and a single ungenotyped causal variant the effect of imperfect LD is easy to quantify, since the variance explained at the SNP is r^2 times the variance explained at the causal locus, with r^2 the standard LD (squared) correlation of alleles at two loci (Hill and Robertson, 1968). The maximum value that r^2 can reach is strongly determined by the allele frequencies at the two loci (Wray, 2005) and the more different the allele frequencies the lower the value of r^2. Therefore, since most genotyped SNPs are common, if causal variants have low minor allele frequency then the amount of variation (heritability) explained at the genotyped SNPs can be substantially lower than the variation explained by the causal variant.

We cannot measure the LD between causal variants and genotyped SNPs directly because we do not know the causal variants. However, we can estimate the LD between SNPs. If the causal variants have similar characteristics to the SNPs in terms of allele frequency spectra and linkage disequilibrium, the LD between causal variants and SNPs should be similar to that between the SNPs themselves. One causal variant can be in LD with multiple SNPs and so the SNPs collectively could trace the causal variant even though no one SNP was in perfect LD with it. Therefore we divided the SNPs randomly into two groups and treated the first group as if they were causal variants and asked how well the second group of SNPs tracked these simulated causal variants. This can be judged by the extent to which the relationship matrices calculated from the SNPs agree with the relationship matrix calculated from the 'causal variants'. The covariance between the estimated relationships for the two sets of SNPs equals the true variance of relatedness whereas the variance of the estimates of relatedness for each set of SNPs equals true variation in relatedness plus estimation error. Therefore, from the regression of pairwise relatedness estimated from one of the set of SNPs onto the estimated pairwise relatedness from the other set of SNPs we can quantify the amount of error and 'regress back' or 'shrink' the estimate of relatedness towards the mean to take account of the prediction error. This is standard practice in making predictions of random effects (Goddard et al., 2009; Lynch & Walsh, 1998; Robinson, 1991).

We found that the relationship between two individuals estimated from the SNPs had to be regressed back towards the mean by about 16% to provide an unbiased estimate of the relationship calculated from the 'causal variants'. This means that the variance explained by causal variants that had similar characteristics, in terms of MAF and LD structure, to common SNPs, would be 54% [~ 45 / (1 – 0.16)] of phenotypic variance not the 45% tracked by the SNPs. It may seem odd that relationships based on 300,000 SNPs are not accurate but this is because the relationships being estimated are very small as the people are 'unrelated' and so small sampling errors caused by a finite but large number of SNPs are still important.

If causal variants have a lower MAF than common SNPs the LD between SNPs and causal variants is likely to be lower than the LD between random SNPs. To investigate the effect of this possibility we used SNPs with low MAF to mimic causal variants. We found that the relationship estimated by random SNPs (with MAF typical of the genotyped SNPs on the array) was a poorer predictor of the relationship at these 'causal variants' than it was of the relationship at other random SNPs. When the relationship matrix at the SNPs is shrunk to provide an unbiased estimate of the relationship at these 'causal variants', we find that the 'causal variants' would explain 80% of the phenotypic variance which is our conventional estimate of the heritability. This does not prove that the causal variants are similar to SNPs with MAF < 0.1 but it shows the data are consistent with this hypothesis.

What Would Happen if a Much Denser SNP Chip is Used?

Denser SNPs would provide higher LD with causal variants and so the proportion of variance explained by the SNPs would increase. However, if the characteristics of causal variants differ systematically from those of the genotyped SNPs (e.g., because of lower MAF), then the genotyped SNPs will still not perfectly track the causal variants and so will still explain less than all the genetic variance. This problem can be described in two equivalent ways. As already stated it can be viewed as incomplete LD between SNPs and causal variants due to different MAF distributions. Alternatively, it can be viewed as a difference between

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press

the relationships between people at causal variants and the relationships between individuals at the genotyped SNPs. For instance, if causal variants have low MAF because they are recent mutations, then common genotyped SNPs that trace very ancient relationships do not correctly reflect the relationships at the causal variants. As stated previously, estimates of genetic variance from pedigree relationships do not suffer from this lack of LD because pedigree relationships employ the correct probabilities of IBD at all causal variants, rare or common.

What Would Happen if the Experimental Sample Size is Larger?

Larger sample size would mean that we estimate the variance explained by the SNPs with a lower standard error but it would not systematically affect it because the current estimate is unbiased. A larger sample size in the context of prediction analysis would result in more accurate estimates of individual marker effects and therefore a larger correlation between predictor and outcome (detailed further below and in the Appendix A).

What Would Population Structure do to These Estimates?

Structure in the population, whether due to unknown close relatives in the sample or population substructure, has the effect that the marker similarity for a pair of individuals at one location on the genome is correlated with their similarity at other locations. That is easily seen for very close relatives; for example, full sibs. They share approximately 50% of their genes by descent on all autosomes. So if there are causal variants on one chromosome then they could be 'detected' by SNPs on other chromosomes. This means that LD exists even between genes on different chromosomes. The analysis would still calculate the variance explained by the SNPs but this could include the effect of causal variants that were not even linked to the SNPs being used. Similarly, if we had individuals with ancestry from Holland (tall) and Italy (less tall), then, because Italians are more similar on all chromosomes than they are to the Dutch (and vice versa), causal variants would be correlated with SNPs even if they are not on the same chromosome. We are not interested in these kinds of associations because we want to estimate genetic variance due to causal variants that are in close LD with the SNPs.

What is the Evidence That Population Structure is not Causing the Observed Effects?

We took several steps to avoid population structure inflating the estimate of the variance explained by the SNPs. We excluded one individual from any pair that had an estimated relationship > 0.025 (approximately equivalent to between 3rd and 4th cousins). We fitted the first 20 principal components from the relationship matrix in the statistical model so that any population substructure that they picked up was excluded from the variance explained by the SNPs. Critically, we then estimated the correlation between the relationship matrices estimated from different chromosomes and did not find significant correlation. We tested a set of SNPs that are ancestry-informative in Europe for association with height and did not observe inflation of the test-statistics.

For the purpose of this article, we performed an additional simulation experiment (inspired by comments from Dan Stram) by assuming that the causal variants were all carried on one set of chromosomes (odd numbers) and another set of chromosomes (even numbers) carried SNPs from which we estimated relatedness. If there is structure in the population then this would imply that a pair of individuals that are closely related on odd chromosomes will also be closely related on even chromosomes. We used the observed genotype data of 3,925 individuals and 295K SNPs as the basis of the simulation, and simulated 1,000 causal variants on the odd chromosomes with a total heritability of 80%. Then we performed a restricted maximum likelihood (REML) analysis of the simulated phenotypes on the genetic relationship matrix estimated from the SNPs on the even chromosomes. The estimates and standard errors (SEs) from 10 simulation replicates are shown in Table 1. Since REML estimates of variance are always positive, if the true variance explained is zero, we expect half the replicates to return an estimate of 0.0 and half to return an estimate with mean value 0.8 times the standard error. This is exactly what happened. Therefore we conclude (again) that there is no structure in the data that would inflate the estimate of the variance explained by the SNPs.

What about G × G and G × E Interactions?

The narrow sense heritability of height is estimated to be around 0.8. This is the additive genetic variance as a proportion of the phenotypic variance — it does not include non-additive genetic variance (G × G or dominance) or genotype × environment interactions (G × E). These may form part of the 20% of phenotypic variance that is not additive genetic.

What are Our Results Consistent With?

Our results are consistent with a highly polygenic model because we detect variation across the entire genome. In the published GWAS for height, the largest proportion of variance explained by any one SNP is about 0.003, that is, 0.3% (Lango Allen et al., 2010). This means there must be many genes affecting height and consequently most must explain a small proportion of the variance. We estimate that half of the additive genetic variation in height is tagged by common SNPs. This does not mean that causal variants are necessarily common or necessarily SNPs, only that they are in sufficient LD with the genotyped SNPs to be detected. Since we do not explain all genetic variation that we believe exists in the population, our results are consistent with a model in which the

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press

remainder of variation is caused by variants not in strong LD with the SNPs. One reason why such variants are not in strong LD with the SNPs is because they have lower MAF. Another reason could be that they are in regions of the genome not well covered by SNPs; for example, structural/repetitive variation. Our results could be caused by causal variants that have MAF < 0.1.

What are our Results not Consistent With?

It is possible that some causal variants are due to a very rare allele that has a large effect on height. For instance, many mutations causing dwarfism are known. Mutations at the gene FBN1 that causes Marfan syndrome increase height by approximately 10cm, but their frequencies are low (~1/5000 or less). Under the (unrealistic) assumption of no selection, these mutations explain about 0.2 to 0.3% of the variance for height (using $2p(1-p)[a + (1-2p)d]^2$, with a and d the additive and dominance effects and p the frequency of the mutation). It is unlikely that any mutation with such a large effect explains a large proportion of the variance because they are rare. Our results show that half the genetic variance is tracked by common SNPs and this variance is split among hundreds or even thousands of SNPs. The remaining half of the genetic variance could well be split among a large number of causal variants with MAF < 0.1 but there is no reason to believe that this missing variance is explained by only a few variants and therefore most must explain a small amount of the variance. This has implications for the design of resequencing studies. Genome sequence data will include causal variants and so increase the power to detect rare variants. However, the power will still be limited by the proportion of variance explained by each variant and therefore large sample size will still be needed.

Why are the Same Proportions of Variance Not Found When Doing Prediction Analyses?

From the prediction analyses in Lango Allen et al. (2010), the regression analysis explains only ~10%-12% of variation in height. This is not at all inconsistent with estimating that 45% of phenotypic variation is explained by common SNPs. The prediction equation depends on the effects of individual SNPs being estimated accurately. Since these effects are small, even small standard errors are important and reduce the accuracy of a prediction equation based on SNPs (Goddard et al., 2009).

The prediction analyses in Lango Allen et al. (and in other published papers) are from a regression of the phenotype (y) on a predictor ($\hat{y}$). The predictor is typically a linear combination of estimated SNP effects ($\hat{b}_i$ for the i-th SNP) and genotype indicator variables (x_i for the i-th SNP, with, for example, $x_i = 0$, 1 or 2 for genotypes AA, AB and BB). That is, $\hat{y} = \sum x_i \hat{b}_i$. Importantly, the SNP effects are usually estimated using least squares. SNPs that are in the predictor can be variants that are known to be associated with the trait or a subset of all GWAS SNPs, for example ranked on statistical significance (Purcell et al., 2009; Wray et al., 2007).

Let us take the extreme but illustrative example that we know m causal variants and that these variant together explain h^2 of the phenotypic variation in the population. That is, $\Sigma[2p_i(1-p_i)b_i^2] / \text{var}(y) = \text{var}(g) / \text{var}(y) = h^2$, with p_i the frequency of the causal variant at the i-th locus and var(g) the additive genetic variance explained by these m causal variants. However, although we know which loci contribute additive genetic variation, we do not know the true effects (b_i), we only have their estimates ($\hat{b}_i$). Now we can contrast the estimate of the proportion of variance explained by these loci with the prediction accuracy when we estimate the effects in a discovery sample and use those estimated effects in an independent test sample. Details are given in Appendix A. They show that the proportion of variance explained by all SNPs in the population is a different parameter to the squared correlation between phenotype and a predictor constructed from the causal variants when the effects are estimated with error. The measurement error on the causal variants (or SNPs) decreases the correlation between predictor and phenotype whereas the estimate of the total variance explained by all SNPs is not influenced by these errors.

Why Did We Leave Out Close Relatives?

The reason for leaving out closer relatives (e.g., 3rd cousins or closer) was to avoid the possibility that the resemblance between close relatives could be due to non-genetic effects (shared environment) so that we would be picking up environmental rather than genetic effects. In fact, leaving these few pairs in or out made very little difference to the results. If we had included many close relatives such as twin pairs (MZ and DZ pairs), fullsibs and parents and offspring, then the estimate of heritability would be dominated by the phenotypic resemblance of these relatives because their estimated relationships (1 for MZs, and approximately 1/2 for first-degree relatives) are so much larger than the estimates of relatedness between 'unrelated' pairs (on average zero with a SD of approximately 0.004). The estimate of heritability from an analysis with many close relatives would be similar to the estimate using only those relatives and fitting an AE model. Such an analysis would not tell us something new and would not be informative with respect to variation due to causal variants that are in LD with common SNPs.

Concluding Remarks

In Yang et al. (2010) we focussed on the estimation of additive genetic variation explained by all SNPs together, using the standard model of quantitative (biometrical) genetics. We relied on the use of a statistical equivalent model of fitting all SNPs in the model of analysis (with their effects random, i.e. from a distribution of effect sizes) and a model that fits random effects

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press

of individuals and uses all SNPs to estimate relationships between people. The method we used differs from standard GWAS in that there is no selection of SNPs based upon test statistics for association between height and SNPs. Consequently, we do not suffer from the 'Winner's Curse' and we have shown elsewhere (Goddard et al., 2009) that treating SNPs effects as random (rather than fixed, as is done in standard GWAS analysis) is logical and leads to unbiased estimated of their effects, in the sense that $E(b|\hat{b}) = \hat{b}$. Statistical analysis of genome-wide effects of individuals or of individual markers effects when fitting all markers in the model is routinely done in plant and animal breeding (Goddard & Hayes, 2009; Meuwissen et al., 2001).

Why have we encountered so much apparent misunderstanding of the methods and results in the human genetics community? The core of our method is heavily steeped in the tradition of prediction of random effects and the estimation of variance due to random (latent) effects. While estimation and partitioning of variance has a long history in human genetics, in particular in twin research, the prediction of random effects is alien to many human geneticists and, surprisingly, also to statisticians (Robinson, 1991). Another reason could be the simultaneous use of population genetics and quantitative genetics concepts and theory in our paper, since these are usually applied in different applications, e.g., gene mapping or estimation of heritability. All concepts and methods that we used are extensively described in the textbooks by Falconer and Mackay (1996; chapters 1, 3, 4, 7–10) and Lynch and Walsh (1998; chapters 4, 7, 26, 27).

Acknowledgments

We thank Dan Stram, Bruce Weir, Alkes Price and Naomi Wray for their insightful comments and helpful discussions on the Yang et al. (2010) paper. We thank the referees for helpful comments on the manuscript. We acknowledge funding from the Australian National Health and Medical Research Council (NHMRC grants 389892, 613672) and the Australian Research Council (ARC grants DP0770096 and DP1093900).

References

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. London: Longman.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399–433.

Goddard, M. E. (2009). Genomic selection, prediction of accuracy and maximisation of long term. *Genetica, 136,* 245–257.

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics, 10,* 381–391.

Goddard, M. E., Wray, N. R., Verbyla, K., & Visscher, P. M. (2009). Estimating effects and making predictions from genome-wide marker data. *Statistical Science, 24,* 517–529.

Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadottir, A., Ingason, A., Steinthorsdottir, V., Olafsdottir, E. J., Olafsdottir, G. H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K. K., Witjes, J. A., Swinkels, D. W., den Heijer, M., Franke, B., Verbeek, A. L., Becker, D. M., Yanek, L. R., Becker, L. C., Tryggvadottir, L., Rafnar, T., Gulcher, J., Kiemeney, L. A., Kong, A., Thorsteinsdottir, U., & Stefansson, K. (2008). Many sequence variants affecting diversity of adult human height. *Nature Genetics, 40,* 609–615.

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

Hill, W. G., & Robertson, A. (1968). The effects of inbreeding at loci with heterozygote advantage. *Genetics, 60,* 615–628.

Jakobsdottir, J., Gorin, M. B., Conley, Y. P., Ferrell, R. E., & Weeks, D. E. (2009). Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genetics, 5,* e1000337.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Ferreira, T., Wood, A. R., Weyant, R. J., Segre, A. V., & Raychaudhuri, S. (in press). Hundreds of variants influence human height and cluster within genomic loci and biological pathways. *Nature.*

Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F. R., Berndt, S. I., Sanna, S., Eyheramendy, S., Voight, B. F., Butler, J. L., Guiducci, C., Illig, T., Hackett, R., Heid, I. M., Jacobs, K. B., Lyssenko, V., Uda, M., Boehnke, M., Chanock, S. J., Groop, L. C., Hu, F. B., Isomaa, B., Kraft, P., Peltonen, L., Salomaa, V., Schlessinger, D., Hunter, D. J., Hayes, R. B., Abecasis, G. R., Wichmann, H. E., Mohlke, K. L., & Hirschhorn, J. N. (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. *Nature Genetics, 40*, 584–591.

Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Maher, B. (2008) Personal genomes, The case of the missing heritability. *Nature, 456,* 18–21.

Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M. J., Seddon, J. M. (2006) Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genetics, 38*, 1055–1059.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157,* 1819–1829.

Powell, J. E., Visscher, P. M., & Goddard, M. E. (in press). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460,* 748–752.

Robinson, G. K. (1991). That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science, 6,* 15–32.

Strandén, I., & Garrick, D. J. (2009). Technical note, Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *Journal of Dairy Science, 92,* 2971–2975.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science, 91,* 4414–4423.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era — Concepts and misconceptions. *Nature Reviews Genetics, 9,* 255–266.

Visscher, P. M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., Hill, W. G., Hottenga, J. J., Willemsen, G., Boomsma, D. I., Liu, Y. Z., Deng, H. W., Montgomery, G. W., & Martin, N. G. (2007). Genome partitioning of genetic variation for height from 11,214 sibling pairs. *American Journal of Human Genetics, 81,* 1104–1110.

Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Cornes, B. K., Montgomery G. W., Martin N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics, 2,* e41.

Weedon, M. N., Lettre, G., Freathy, R. M., Lindgren, C. M., Voight, B. F., Perry, J. R., Elliott, K. S., Hackett, R., Guiducci, C., Shields, B., Zeggini, E., Lango, H., Lyssenko, V., Timpson, N. J., Burtt, N. P., Rayner, N. W., Saxena, R., Ardlie, K., Tobias, J. H., Ness, A. R., Ring, S. M., Palmer, C. N., Morris, A. D., Peltonen, L., Salomaa, V., Davey, Smith, G., Groop, L. C., Hattersley, A. T., McCarthy, M. I., Hirschhorn J. N., & Frayling T. M. (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics, 39,* 1245–1250.

Wray, N. R. (2005). Allele frequencies and the r2 measure of linkage disequilibrium, impact on design and interpretation of association studies. *Twin Research and Human Genetics, 8,* 87–94.

Wray, N. R., Goddard, M. E., & Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Research, 17,* 1520–1528.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden P. A., Heath, A. C., Martin, N. G., Montgomery, G.W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics, 42,* 565–569.

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press

APPENDIX A

Contrasting the Estimate of Genetic Variance With Accuracy of Prediction

From the theory in Yang et al. (2010) and the description in the main text of the current paper we have an estimate of the relationship between individuals j and k defined at the causal variants are,

$$A_{jk} = \frac{1}{m}\sum \frac{\text{cov}(x_{ij}, x_{ik})}{\text{var}(x_i)}$$

We also have the covariance between their phenotypes:

$$\text{cov}(y_j, y_k) = \text{cov}(\Sigma x_{ij} b_i + e_j,\ x_{ik} b_i + e_k) = \Sigma[b_i^2 \text{cov}(x_{ij}, x_{ik})]$$
$$= \Sigma[b_i^2 \text{var}(x_i)\, \text{cov}(x_{ij}, x_{ik}) / \text{var}(x_i)]$$
$$= \text{var}(g) A_{jk}$$

Hence, the phenotypic resemblance between any pair of individuals is proportional to the relationship at the causal variants. This is exactly analogous to using pedigree relationships (instead of the A_{jk} from identity at the causal variants) and contrasting phenotypic resemblance with identity at the causal variants will lead to an unbiased estimate of genetic variance contributed by all these variants.

Now we contrast the estimate of heritability from relatives to the proportion of variance that is explained by the predictor in the least squares regression analysis (R^2). Least squares estimates have the properties of $\text{E}(\hat{b}\,|\,\text{b}) = b$ and $\text{var}(\hat{b}\,|\,\text{b})$ is the sampling variance that depends on the allele frequency and sample size: $\text{var}(\hat{b}\,|\,\text{b}) \sim \text{var}(y) / [2p(1-p)N]$ with N being the sample size of discovery set. Hence the expected value of the covariance between predictor and phenotype is

$$\text{E}\{\Sigma[b_i \hat{b}_i\, 2p_i(1-p_i)]\} = \Sigma[2p_i(1-p_i)\, b_i^2] = \text{var}(g).$$
$$\text{R}^2(y, \hat{y}) = \text{cov}(y, \hat{y})^2 / [\text{var}(y)\, \text{var}(\hat{y})]$$
$$= \text{var}(g)^2 / [\text{var}(y)\, \text{var}(\hat{y})] = h^2\, \text{var}(g)/\text{var}(\hat{y})$$

This equation is only equal to h^2 if the SNP effects are known without error because then $\text{var}(\hat{y}) = \text{var}(g)$. Using least squares estimates of b results in the variance of the predictor being much larger than it should be if the SNP effects are estimated with error:

$$\text{var}(\hat{y}) \approx \Sigma[2p_i(1-p_i)\hat{b}_i^2]$$

The expected value of $\hat{b}^2$ is $\text{E}(\hat{b})^2 + \text{var}(\hat{b}) = b^2 + \text{SE}^2(\hat{b})$, so
$\text{E}[\text{var}(\hat{y})] = \Sigma[2p_i(1-p_i)b_i^2] + \Sigma[2p_i(1-p_i)\text{SE}^2(\hat{b}_i)] \approx \text{var}(g) + m\text{var}(y) / N$
and finally,

$$\text{E}[R^2(y, \hat{y})] \approx h^2\, \text{var}(g) / [\text{var}(g) + m\, \text{var}(y)/\text{N}] = h^2 / [1 + m / (h^2 \text{N})]$$

This is a very long-winded way of saying that the accuracy of prediction from estimated SNP effects can be very different from the proportion of variance explained in the population by those effects.

As an example, if we assume that for height $m = 1000$ and $h^2 = 0.5$. Then even if we knew all causal variants and had a sample size of 10,000 to estimate their effect sizes, the expected regression R^2 in a different sample is ~0.42. In reality, we do not know the causal variants. The estimated effect sizes for associated loci for height are about 1-4 mm and the SE from a sample size of 100,000 and MAF of 0.3 and a phenotypic SD of 70 mm is ~0.3mm. Hence there is considerable error in the estimation of the effects.

https://doi.org/10.1375/twin.13.6.517 Published online by Cambridge University Press