Web Resources

The URLs for data presented herein are as follows:

GCTA, http://www.complextraitgenomics.com/software/gcta/
LDAK, http://www.ldak.org

References

1. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021.
2. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
3. Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J., and Goddard, M.E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet. *6*, e1001139.
4. Stahl, E.A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B.F., Kraft, P., Chen, R., Kallberg, H.J., Kurreeman, F.A.S., et al.; Diabetes Genetics Replication and Meta-analysis Consortium; Myocardial Infarction Genetics Consortium (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat. Genet. *44*, 483–489.
5. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. *9*, e1003264.
6. Zaitlen, N., and Kraft, P. (2012). Heritability in the genome-wide association era. Hum. Genet. *131*, 1655–1664.
7. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R.; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS) (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat. Genet. *44*, 247–250.
8. 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., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
9. Lee, S.H., and van der Werf, J.H.J. (2006). An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. Genet. Sel. Evol. *38*, 25–43.
10. Leutenegger, A.-L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., and Thompson, E.A. (2003). Estimation of the inbreeding coefficient through use of genomic data. Am. J. Hum. Genet. *73*, 516–523.
11. Amin, N., van Duijn, C.M., and Aulchenko, Y.S. (2007). A genomic background based method for association analysis in related individuals. PLoS ONE *2*, e1274.
12. VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. *91*, 4414–4423.
13. Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al.; Multicenter Genetic Studies of Schizophrenia Consortium; Psychosis Endophenotypes International Consortium; Wellcome Trust Case Control Consortium 2 (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat. Genet. *45*, 1150–1159.
14. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., and Wray, N.R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism -derived genomic relationships and restricted maximum likelihood. Bioinformatics *28*, 2540–2542.
15. Lee, S.H., Harold, D., Nyholt, D.R., Goddard, M.E., Zondervan, K.T., Williams, J., Montgomery, G.W., Wray, N.R., and Visscher, P.M.; ANZGene Consortium; International Endogene Consortium; Genetic and Environmental Risk for Alzheimer's disease Consortium (2013). Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. Hum. Mol. Genet. *22*, 832–841.

Response to Lee et al.: SNP-Based Heritability Analysis with Dense Data

To the Editor: In Speed et al.,[1] we identified two potential issues when performing SNP-based heritability estimation: (1) estimates of h^2 can be biased when the tagging of causal variants differs from that of the SNPs used for calculating the genomic-relationship matrix (GRM), and (2) the accuracy of h^2 estimates depends on how closely the assumed relationship between a causal variant's minor allele frequency (MAF) and effect size matches the true relationship (this relationship can be modeled with a scale parameter s, where the standard assumption is $s = -1$). To resolve the

Table 1. Performance of Standard and MAF-Stratified Versions of GCTA and LDAK

Architecture	GCTA	LDAK	MAF-GCTA	MAF-LDAK
Causal SNPs Picked at Random				
Architecture A ($s = 1$)	0.50 (0.05)	0.51 (0.08)	0.50 (0.05)	0.51 (0.08)
Architecture B ($s = 0$)	0.52 (0.04)	0.50 (0.07)	0.50 (0.04)	0.50 (0.07)
70% Causal SNPs with MAF < 0.1				
Architecture C ($s = 1$)	0.47 (0.04)	0.51 (0.06)	0.50 (0.04)	0.51 (0.06)
Architecture D ($s = 0$)	0.52 (0.05)	0.53 (0.07)	0.52 (0.05)	0.53 (0.07)
All Causal SNPs with MAF < 0.1				
Architecture E ($s = 1$)	0.45 (0.05)	0.51 (0.08)	0.52 (0.04)	0.53 (0.07)
Architecture F ($s = 0$)	0.45 (0.04)	0.52 (0.06)	0.51 (0.03)	0.53 (0.05)
Well-Tagged Causal SNPs				
Architecture G ($s = 1$)	0.89 (0.03)	0.56 (0.06)	0.89 (0.04)	0.56 (0.06)
Architecture H ($s = 0$)	0.88 (0.03)	0.54 (0.08)	0.84 (0.03)	0.54 (0.07)
Poorly Tagged Causal SNPs				
Architecture I ($s = 1$)	0.13 (0.05)	0.49 (0.06)	0.14 (0.05)	0.49 (0.06)
Architecture J ($s = 0$)	0.13 (0.05)	0.50 (0.07)	0.13 (0.05)	0.51 (0.07)

Architectures A–F have been defined by Lee et al.; we additionally considered architectures G and H, where causal variants are well-tagged SNPs, and architectures I and J, where causal variants are poorly tagged SNPs. The true (simulated) h^2 is 0.5. Each value reports the mean estimate of h^2 over 50 replicates (the empirical SD is provided in parentheses). Note that Lee et al. also reported the Akaike information criterion (AIC), but we omit this score because it can be highly misleading; for example, in our simulations, the highest AIC was achieved with GCTA for architecture G, where the estimates of h^2 were on average 80% higher than the true value.

first issue, we proposed computing an adjusted GRM, where uneven tagging is accounted for by weighted SNPs. This approach is implemented in our software LDAK.[1]

Lee et al. proposed continuing to use an unadjusted GRM computed via GCTA and to instead take a MAF-stratification approach to estimating h^2. We summarize the major claims of their paper as follows:

(1) Their MAF-stratification approach accounts for uncertainty about s, (2) the weightings from LDAK are suboptimal for dense SNP data, and (3) their MAF-stratification approach gives less biased estimates of h^2. We agree with the first, show that the second can be avoided by appropriate parameter selection in LDAK, and disagree with the third. In our opinion, adjusting for uneven tagging becomes even more important in the analysis of dense SNP data, and the MAF-stratification method of Lee et al. can be improved by incorporating the weightings used in LDAK.

To support the third claim, Lee et al. presented simulations in which the causal variants tend to be poorly tagged. However, the design of the simulations means that potential biases in their MAF-stratification approach are not evident. For example, for architecture E, causal variants are restricted to SNPs with a MAF < 0.1, so estimates of h^2 tend to be biased downward if they are based on a GRM computed from all SNPs (demonstrated by the performance of standard GCTA in their simulations). MAF < 0.1 defines one of the (arbitrary) tranches in their analysis; had there been a mismatch between the MAF tranche from which causal variants were randomly selected and the MAF tranche used for analysis, their approach would have experienced biases similar to those suffered by GCTA.

Here, we demonstrate the continued importance of adjusting for uneven tagging through simulation (50 replicates in each case). We considered a data set of 6,387 individuals, who, after imputation against the 1000 Genomes reference panel, were genotyped for 4,238,038 SNPs. In addition to considering architectures A–F defined by Lee et al., we also considered architectures G and H, in which the 10,000 causal SNPs are well tagged, and architectures I and J, in which they are poorly tagged. (Tagging is measured by T, effectively the multiplicity of a signal.[1] Here, T ranges from 1 to 2,032 and has a median of 92; a SNP is defined as well tagged if T > 179 and poorly tagged if T < 45.) We compared four analysis methods: GCTA (standard GRM), LDAK (weighted GRM), and MAF versions of both of these; in these MAF versions, GRMs were computed for each of five MAF tranches. MAF-LDAK was implemented with the "region" option in LDAK. Each GRM was computed with $s = -1$, and we used the default settings for LDAK.

Our results for architectures A–F (Table 1) agree with those of Lee et al. in that MAF-GCTA outperformed GCTA. Any overestimation of h^2 by LDAK appeared to be slight, less than what was observed by Lee et al.; we return to this point below. When causal variants were well or poorly tagged (architectures G–J), the estimates of h^2 from MAF-GCTA (similar to those of GCTA) tended to be biased upward or downward, respectively, whereas those from MAF-LDAK were much closer to the truth, indicating that adjusting for uneven tagging and stratifying by MAF are both advantageous. We repeated these simulations with 500, 1,000, 2,000, and 5,000 causal variants and obtained similar results each time (data not shown).

The above-reported upward bias of LDAK is smaller than that found by Lee et al. (the second claim above), which might be a result of fewer SNPs (approximately four million versus eight million) because we imposed stricter quality control (QC). We believe that strict QC is crucial in the estimation of h^2 for binary traits,[1,2] but to allow direct comparison, we relaxed the QC thresholds to match those of Lee et al. Afterwards, our data set had 7,190,149 SNPs. Because of the limited time to prepare this response, we only used chromosomes 1 and 2 (1,153,686 SNPs) and simulated 1,000 causal variants. Focusing on the

Table 2. Performance of GCTA and LDAK for Very Dense Data

Architecture	GCTA	LDAK	LDAK2
Causal SNPs Picked at Random			
Architecture A ($s = 1$)	0.49 (0.03)	0.52 (0.04)	0.50 (0.04)
Architecture B ($s = 0$)	0.55 (0.02)	0.54 (0.03)	0.52 (0.03)
70% Causal SNPs with MAF < 0.1			
Architecture C ($s = 1$)	0.49 (0.02)	0.52 (0.04)	0.50 (0.04)
Architecture D ($s = 0$)	0.51 (0.02)	0.52 (0.03)	0.50 (0.03)
All Causal SNPs with MAF < 0.1			
Architecture E ($s = 1$)	0.42 (0.03)	0.51 (0.04)	0.49 (0.04)
Architecture F ($s = 0$)	0.46 (0.03)	0.51 (0.04)	0.49 (0.04)

LDAK employs the default parameter settings when computing weightings; LDAK2 turns off the linkage-disequilibrium-decay function. The true (simulated) h^2 is 0.5. Each value reports the mean estimate of h^2 over 50 replicates (the empirical SD is provided in parentheses).

architectures with $s = -1$ in Table 2, we observed overestimation of h^2 by LDAK, although to a lesser extent than did Lee et al. (2%, 2%, and 1% compared with 6%, 5%, and 3%). When weightings are calculated, LDAK by default models linkage-disequilibrium decay with distance in order to give more weight to local correlations than to long-range correlations that might be due to relatedness. For unrelated individuals, this is unnecessary, and so a large value (such as 100 Mb) can be used for the decay parameter. With this change, no bias is apparent (LDAK2).

In summary, although we agree with Lee et al. that MAF stratification is effective in reducing biases caused by misspecification of the scale parameter s, we feel that it remains important to adjust for uneven tagging. In addition to achieving improved accuracy by incorporating the LDAK weightings, this approach effects SNP pruning (for imputed data, approximately 90% of SNPs will receive weight zero and so can be discarded), thus reducing the subsequent task of computing GRMs by a factor of about ten.

Doug Speed,[1,*] Gibran Hemani,[2] Michael R. Johnson,[3] and David J. Balding[1]
[1]UCL Genetics Institute, University College London, London WC1E 6BT, UK; [2]Diamantina Institute, University of Queensland, Brisbane, Queensland 4102, Australia; [3]Division of Brain Sciences, Imperial College London, London W6 8RF, UK
*Correspondence: doug.speed@ucl.ac.uk (D.S.)

References

1. Speed, D., Hemani, G., Johnson, M.R., and Balding, D.J. (2012). Improved heritability estimation from genome-wide SNPs. Am. J. Hum. Genet. *91*, 1011–1021.
2. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. *88*, 294–305.