nature genetics

Reevaluation of SNP heritability in complex human traits

Doug Speed[1], Na Cai[2,3], the UCLEB Consortium[4], Michael R Johnson[5], Sergey Nejentsev[6] & David J Balding[1,7]

SNP heritability, the proportion of phenotypic variance explained by SNPs, has been reported for many hundreds of traits. Its estimation requires strong prior assumptions about the distribution of heritability across the genome, but current assumptions have not been thoroughly tested. By analyzing imputed data for a large number of human traits, we empirically derive a model that more accurately describes how heritability varies with minor allele frequency (MAF), linkage disequilibrium (LD) and genotype certainty. Across 19 traits, our improved model leads to estimates of common SNP heritability on average 43% (s.d. 3%) higher than those obtained from the widely used software GCTA and 25% (s.d. 2%) higher than those from the recently proposed extension GCTA-LDMS. Previously, DNase I hypersensitivity sites were reported to explain 79% of SNP heritability; using our improved heritability model, their estimated contribution is only 24%.

The SNP heritability (h^2_{SNP}) of a trait is the fraction of phenotypic variance explained by additive contributions from SNPs[1]. Accurate estimates of h^2_{SNP} are central to resolving the missing heritability debate, indicate the potential utility of SNP-based prediction and help design future genome-wide association studies (GWAS)[2,3]. Whereas techniques for estimating (total) heritability have existed for decades[4,5], the first method[1] for estimating h^2_{SNP} was proposed only in 2010 but has since been applied to many hundreds of traits. Extensions of this method are now being used to partition heritability across chromosomes, according to biological pathways and by SNP function and to calculate the genetic correlation between pairs of traits[6–8].

As the number of SNPs in a GWAS is usually much larger than the number of individuals, estimation of h^2_{SNP} requires steps to avoid overfitting. Most reported estimates of h^2_{SNP} are based on assigning the same Gaussian prior distribution to each SNP effect size, in a way that implies that all SNPs are expected to contribute equal heritability[1,9]. By examining a large collection of real data sets, we derive approximate relationships between the expected heritability of a SNP and MAF, levels of LD with other SNPs and genotype certainty. This provides us with an improved model for heritability estimation and a better understanding of the genetic architecture of complex traits.

RESULTS

When estimating h^2_{SNP}, the 'LDAK model' assumes

$$E[h_j^2] \sim [f_j(1-f_j)]^{1+\alpha} \times w_j \times r_j \quad (1)$$

where $E[h_j^2]$ is the expected heritability contribution of SNP j and f_j is its (observed) MAF. The parameter α determines the assumed relationship between heritability and MAF. In human genetics, it is commonly assumed that heritability does not depend on MAF, which is achieved by setting $\alpha = -1$; however, we consider alternative relationships. The SNP weights $w_1, \ldots, w_m$ are computed on the basis of local levels of LD[9]; w_j tends to be higher for SNPs in regions of low LD, and thus the LDAK model assumes that these SNPs contribute more than those in high-LD regions. Finally, $r_j \in [0,1]$ is an information score measuring genotype certainty; the LDAK model expects that higher-quality SNPs contribute more than lower-quality ones. r_j is defined in the Online Methods, where we also explain how model (1) arises by assuming a genome-wide random regression in which SNP effect sizes are assigned Gaussian distributions.

The 'GCTA model' is obtained from model (1) by setting $w_j = 1$ and $r_j = 1$, and thus assumes that expected heritability does not vary with either LD or genotype certainty. Thus far, most reported estimates of h^2_{SNP} have used the GCTA model with $\alpha = -1$, which corresponds to the assumption that $E[h_j^2]$ is constant, and so the expected contribution of a SNP set depends only on the number of SNPs it contains[1]. To appreciate the major difference between the GCTA and LDAK models, consider a region containing two SNPs: under the GCTA model, the expected heritability of these two SNPs is the same irrespective of the LD between them, whereas under the LDAK model two SNPs in perfect LD are expected to contribute only half the heritability of two SNPs showing no LD. See **Figure 1** for a more detailed example.

An alternative method for estimating h^2_{SNP} is LDSC (LD score regression)[10]. The LDSC model expects that each SNP contributes equal heritability[10,11] and therefore closely resembles the GCTA model with $\alpha = -1$. When applied to the same data set, estimates from LDSC will typically have standard errors 25–100% higher than those from GCTA[11]; this is partly because the LDSC model includes an extra parameter, designed to capture confounding biases, and partly

[1]UCL Genetics Institute, University College London, London, UK. [2]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK. [3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK. [4]A full list of members and affiliations appears in the **Supplementary Note**. [5]Division of Brain Science, Imperial College London, London, UK. [6]Department of Medicine, University of Cambridge, Cambridge, UK. [7]Centre for Systems Genomics, School of BioSciences, and School of Mathematics and Statistics, University of Melbourne, Melbourne, Victoria, Australia. Correspondence should be addressed to D.S. (doug.speed@ucl.ac.uk).

Received 1 September 2016; accepted 18 April 2017; published online 22 May 2017; doi:10.1038/ng.3865

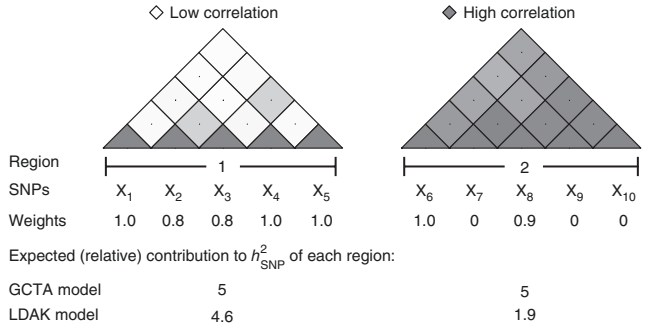

Figure 1 Comparison of the GCTA and LDAK models. Region 1 contains five SNPs in low LD (lighter shadings indicate weaker pairwise correlations). Each SNP contributes unique genetic variation, reflected by SNP weights close to one. Region 2 contains five SNPs in high LD (strong correlations). The total genetic variation tagged by the region is effectively captured by two of the SNPs, and so the others receive zero weight. Under the GCTA model, the regions are expected to contribute heritability proportional to their numbers of SNPs, which are equal here. Under the LDAK model, they are expected to contribute heritability proportional to their sums of SNP weights, which here have the ratio 4.6:1.9. Note that the expected heritability can also depend on the allele frequencies and genotype certainties of the SNPs, but for simplicity these factors are ignored here.

because LDSC estimates are moment based, whereas GCTA (like LDAK) uses restricted maximum likelihood (REML)[12,13]. However, as LDSC requires only summary statistics (*P* values from single-SNP analysis), it can be used on much larger data sets than GCTA and LDAK, which need raw genotype data, and it can be applied to results from large-scale meta-analyses[10].

SNP partitioning

Model (1) can be generalized by dividing SNPs into tranches across which the constant of proportionality is allowed to vary (so $E[h_j^2] = c_k \times [f_j(1-f_j)]^{1+\alpha} \times w_j \times r_j$ for SNPs in tranche k). This is known as SNP partitioning[6]. Two examples are GCTA-MS[14] and GCTA-LDMS[15]; when applied to common SNPs (MAF ≥ 0.01), GCTA-MS divides the genome into five tranches on the basis of MAF, using the boundaries 0.1, 0.2, 0.3 and 0.4, while GCTA-LDMS first divides SNPs into four tranches on the basis of local average LD score[10] and then divides each of these into five on the basis of MAF, resulting in a total of 20 tranches. In general, we prefer to avoid SNP partitioning when estimating h^2_{SNP} because it introduces (often arbitrary) discontinuities in the model assumptions and can cause convergence problems. However, we show below that partitioning based on MAF enables reliable estimation of h^2_{SNP} when rare SNPs (MAF ≤ 0.01) are included. Additionally, SNP partitioning provides a way to visually assess the fit of different heritability models; it allows us to estimate average h_j^2 for different SNP tranches, which can then be compared to the values predicted under different assumptions.

Data sets

In total, we analyzed data for 42 traits. **Table 1** and **Supplementary Table 1** describe the 19 'GWAS traits' (17 case–control and 2 quantitative traits). For these traits, individuals were genotyped using either genome-wide Illumina or Affymetrix arrays (typically with 500,000 to 1.2 million SNPs). We additionally examined data from eight cohorts of the UCLEB consortium[24], which comprise about 14,000 individuals genotyped using the Metabochip[25] (a relatively sparse array of 200,000 SNPs selected on the basis of previous GWAS) and recorded for a wide range of clinical phenotypes. From these, we considered 23 quantitative phenotypes (average sample size 8,200), which can loosely be divided into anthropomorphic (height, weight, body mass index (BMI) and waist circumference), physiological (lung capacity and blood pressure), cardiac (for example, PR and QT intervals), metabolic (glucose, insulin and lipid levels) and blood chemistry (for example, fibrinogen, IL-6 and hemoglobin levels) traits. In general, our quality control was extremely strict; after imputation, we retained only autosomal SNPs with MAF ≥ 0.01 and information score $r_j \geq 0.99$. We only relaxed quality control when, using the UCLEB data, we explicitly examined the consequences of including lower-quality and rare SNPs.

Further details of our methods and data sets are provided in the Online Methods. In particular, we explain how when estimating h^2_{SNP} we give special consideration to highly associated SNPs, which we define as those with $P < 1 \times 10^{-20}$ from single-SNP analysis, and how for the UCLEB data we confirm that genotyping errors do not correlate with phenotype (which is important for the analyses where we include lower-quality SNPs).

Relationship between heritability and MAF

Varying the value of α in model (1) changes the assumed relationship between heritability and MAF; three example relationships are shown in **Figure 2a**. To determine suitable α values, we analyzed each of the 42 traits using seven values (−1.25, −1, −0.75, −0.5, −0.25, 0 and 0.25), seeing which led to the best model fit (highest likelihood). Full results are provided in **Supplementary Figure 1** and **Supplementary Table 2**. First, to remove any confounding due to LD, we used only a pruned subset of SNPs (with $w_j = 1$); next, we repeated without LD pruning (the results for the GWAS traits are shown in **Fig. 2b**); and, finally, for the UCLEB traits, we repeated including lower-quality and rare SNPs. We found that model fit was typically best for $-0.5 \leq \alpha \leq 0$, whereas the most widely used value, $\alpha = -1$, resulted in suboptimal fit. On the basis that it performs consistently well across different traits and SNP filtering criteria, we recommend that $\alpha = -0.25$ become the default. This value implies that expected heritability declines with increasing MAF; this is seen in **Figure 2a**, which reports, averaged across the 19 GWAS traits, the (weight-adjusted) per-SNP heritability for low- and high-MAF SNPs (see **Supplementary Fig. 2** for further details).

While $\alpha = -0.25$ provided the best fit overall, for individual traits, optimal α may differ, and we therefore investigated the sensitivity of h^2_{SNP} estimates to the value of α (**Supplementary Figs. 3–5**). When analyzing only common SNPs, we found that changes in α had little impact on h^2_{SNP}. For example, across the 23 UCLEB traits, estimates from high-quality, common SNPs using $\alpha = -0.25$ were on average only 5% (s.d. 4%) lower than those using $\alpha = -1$ and 4% (s.d. 4%) higher than those using $\alpha = 0$. However, this was no longer the case when rare SNPs were included in the analysis: for example, when the MAF threshold was reduced to 0.0005, estimates using $\alpha = -0.25$ were on average 18% (s.d. 4%) lower than those using $\alpha = -1$ and 30% (s.d. 6%) higher than those from $\alpha = 0$. Therefore, when including rare SNPs, we guarded against misspecification of α by partitioning on the basis of MAF (with boundaries at 0.001, 0.0025, 0.01 and 0.1); we found that this provided stable estimates of h^2_{SNP} and also allows estimation of the relative contributions of rare and common variants (**Supplementary Fig. 6**).

Relationship between heritability and LD

The LDAK model assumes that heritability varies according to local levels of LD, whereas the GCTA model assumes that heritability is independent of LD. First, we demonstrated that choice of model matters when estimating h^2_{SNP}. For the GWAS traits, **Figure 3a**

Table 1 Properties of data sets and estimates of h^2_{SNP}

						Previous			LDAK	
Collection	Trait (disease prevalence, %)	n	m	$\sum_{j=1}^{m} w_j$	h^2_{GWAS}	h^2_{SNP}	s.d.	Ref.	h^2_{SNP}	s.d.
WTCCC 1	Bipolar disorder (0.5)	1,840 + 2,913	2,729,000	79,000	0.02	0.24	0.04	7	0.35	0.03
	Coronary artery disease (6)	1,907 + 2,918	2,739,000	80,000	0.03	0.25	0.06	7	0.40	0.06
	Crohn's disease (0.5)	1,691 + 2,905	2,724,000	79,000	0.21	0.26	0.01	21	0.32	0.03
	Hypertension (5)	1,918 + 2,916	2,740,000	80,000	<0.01	0.33	0.06	7	0.46	0.06
	Rheumatoid arthritis (0.5)	1,846 + 2,918	2,736,000	80,000	0.19	0.09	0.03	7	0.21	0.03
	Type 1 diabetes (0.5)	1,941 + 2,907	2,732,000	80,000	0.27	0.13	0.03	7	0.31	0.02
	Type 2 diabetes (8)	1,896 + 2,917	2,736,000	80,000	0.08	0.42	0.07	7	0.54	0.07
WTCCC 2	Barrett's esophagus (1.6)	1,861 + 5,138	3,831,000	116,000	<0.01	0.25	0.05	16	0.32	0.04
	Ischemic stroke (2)	3,769 + 5,139	3,797,000	115,000	<0.01	0.25	0.03	17	0.34	0.03
	Parkinson's disease (0.2)	1,687 + 5,136	3,820,000	116,000	0.03	0.27	0.05	18	0.20	0.03
	Psoriasis (0.5)	2,267 + 5,143	3,815,000	116,000	0.21	0.35	0.06	19	0.34	0.02
	Schizophrenia (1)	2,068 + 2,615	3,481,000	111,000	0.07	0.23	0.01	20	0.30	0.04
	Ulcerative colitis (0.2)	2,614 + 5,327	4,062,000	115,000	0.12	0.19	0.01	21	0.28	0.02
WTCCC 2+	Celiac disease (1)	2,492 + 7,376	2,682,000	88,000	0.29	0.33	0.04	22	0.35	0.02
	Multiple sclerosis (0.1)	8,553 + 5,667	3,702,000	113,000	0.17	0.17	0.01	7	0.24	0.01
	Partial epilepsy (0.3)	1,217 + 5,152	3,399,000	108,000	<0.01	0.33	0.05	3	0.27	0.04
RPTB	Pulmonary tuberculosis (4)	5,142 + 5,283	2,987,000	102,000	<0.01	None	None	None	0.26	0.03
Blue Mountains	Intraocular pressure	2,235	4,149,000	125,000	0.02	None	None	None	0.38	0.17
CHOP	Wide-range achievement test	3,747	2,593,000	88,000	<0.01	0.43	0.10	23	0.21	0.09
UCLEB[a]	23 quantitative traits	6,458 to 11,005	353,000	39,000						

n is the sample size (cases + controls), m is the number of SNPs and $\sum_{j=1}^{m} w_j$ is the sum of SNP weights, which can be interpreted as an effective number of independent SNPs. All values are from after quality control. For UCLEB, m and $\sum_{j=1}^{m} w_j$ refer to our main analysis, which considered only high-quality, common SNPs. The final two columns provide our best estimates of h^2_{SNP} from common SNPs, computed using LDAK with $\alpha = -0.25$ (see main text for explanation of α). For comparison, we include previously published estimates of h^2_{SNP} (note that the previous analyses for rheumatoid arthritis, type 1 diabetes and multiple sclerosis excluded major histocompatibility complex (MHC) SNPs, which we estimate contribute 0.07, 0.20 and 0.05, respectively), as well as h^2_{GWAS}, the proportion of phenotypic variance explained by SNPs reported as GWAS significant ($P < 5 \times 10^{-8}$). For disease traits, estimates of h^2_{SNP} and h^2_{GWAS} have been converted to the liability scale assuming the stated prevalence.

[a]Results appear in **Supplementary Table 1**.

reports the relative estimates of h^2_{SNP} from GCTA, GCTA-MS, GCTA-LDMS and LDAK (all using $\alpha = -0.25$); see **Supplementary Figure 7** for an extended version. We found that estimates based on the LDAK model were on average 48% (s.d. 3%) higher than estimates based on the GCTA model. For the UCLEB traits, estimates from LDAK were on average 88% (s.d. 7%) higher than those from GCTA (**Supplementary Fig. 8**). **Figure 3a** also includes results from LDSC, run as described in the original publication[10] (see **Supplementary Table 3** for numerical values). Estimates from LDSC were not significantly different to those from GCTA, which is to be expected considering that GCTA and LDSC assume the same relationship between heritability and LD. In **Supplementary Figure 9**, we consider alternative versions of LDSC (for example, varying how LD scores are computed, forcing the intercept term to be zero and excluding highly associated SNPs). While changing settings can have a large impact, in all cases, the average estimate of h^2_{SNP} from LDSC remained substantially below that from LDAK.

A recent article that asserted that GCTA estimates h^2_{SNP} more accurately than LDAK based this claim on a simulation study in which causal SNPs were assigned effect sizes from the same Gaussian distribution, irrespective of LD[6]. This resembles the GCTA model but not the LDAK model, and so it does not seem surprising that GCTA performed better. **Figure 3b** shows that, if effect size variances had instead been scaled by SNP weights and so varied with LD similar to the LDAK model, then the study would have found LDAK to be superior to GCTA. Thus, using simulations to compare different heritability models is problematic because the conclusions will depend on the assumptions used when generating phenotypes. See **Supplementary Figure 10** for a full reanalysis of the reported simulation study and **Supplementary Figure 11** for further simulations.

Rather than using simulations, we compared LDAK and GCTA empirically. **Supplementary Table 4** shows that when $\alpha = -0.25$, assuming the LDAK model led to higher likelihood than assuming the GCTA model for all 19 GWAS traits and for 17 of the 23 UCLEB traits (if we instead used $\alpha = -1$, likelihood was higher under the LDAK model for 31 of the 42 traits). To visually demonstrate the superior fit of the LDAK model, we partitioned SNPs into low- and high-LD tranches (for this, we ranked SNPs according to the average LD score[10] of non-overlapping 100-kb segments, the metric used by GCTA-LDMS[15]). First, we partitioned so that the two tranches contained an equal number of SNPs. The left half of **Figure 4** reports, for each of the GWAS traits, the contribution of the low-LD tranche, estimated using the GCTA model (with $\alpha = -0.25$). Under the GCTA model, the low-LD tranche is expected to contribute 50% of h^2_{SNP}; under the LDAK model, it is expected to contribute 72% of h^2_{SNP}. We saw that the estimated contribution of the low-LD tranche was consistent with the GCTA model (the 95% confidence interval included 50%) for only 5 of the 19 traits, whereas it was consistent with the LDAK model (the confidence interval included 72%) for 18 traits. Next, we partitioned so that the low-LD tranche contained one-quarter of the SNPs; then, the low-LD tranche is predicted to contribute 26% of h^2_{SNP} under the GCTA model but 47% of h^2_{SNP} under the LDAK model. The right half of **Figure 4** shows that its estimated contribution was consistent with the GCTA model for only 7 of the 19 traits, but again was consistent with the LDAK model for 18 traits. Additional results are provided in **Supplementary Figure 12**; these show that, regardless of whether we estimated heritabilities using LDAK (rather

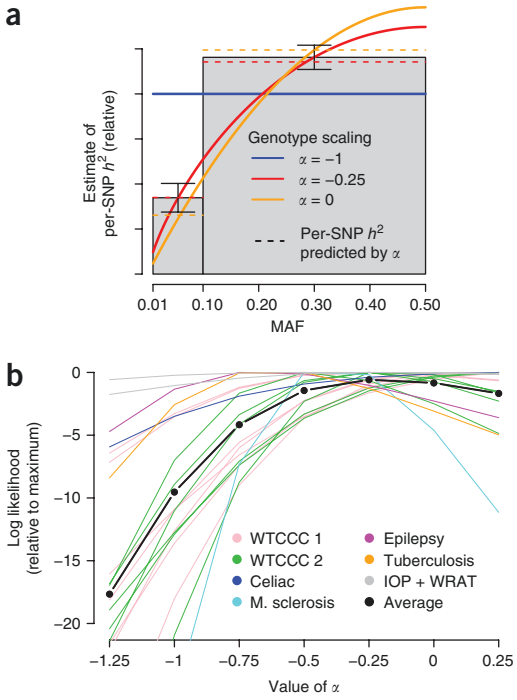

Figure 2 Relationship between heritability and MAF. (**a**) The parameter α specifies the assumed relationship between heritability and MAF: in human genetics, $\alpha = -1$ is typically used (solid blue line), while in animal and plant genetics $\alpha = 0$ is more common (orange); we instead found that $\alpha = -0.25$ (red) provides a better fit to real data. The gray bars report (relative) estimates of the per-SNP heritability for SNPs with MAF < 0.1 and MAF ≥ 0.1, averaged across the 19 GWAS traits (vertical lines provide 95% confidence intervals); the dashed lines indicate the per-SNP heritability predicted by each α value. (**b**) For each tranche, we compare α on the basis of likelihood; higher likelihood indicates better-fitting α. Lines report log likelihoods from LDAK for seven values of α, relative to the highest observed likelihood. Line colors indicate the seven trait categories, while the black line reports averages. M. sclerosis, multiple sclerosis; IOP, intraocular pressure; WRAT, wide-range achievement test.

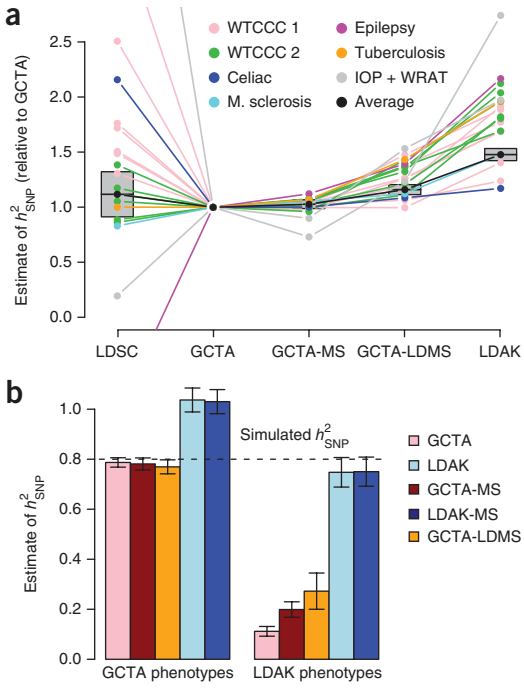

Figure 3 Comparison of methods for estimating h^2_{SNP} for real and simulated data. (**a**) Relative estimates of h^2_{SNP} for the GWAS traits. h^2_{SNP} estimates from LDSC, GCTA-MS (SNPs partitioned by MAF), GCTA-LDMS (SNPs partitioned by LD and MAF) and LDAK are reported relative to those from GCTA. For versions of GCTA and LDAK, we use $\alpha = -0.25$ (see main text for explanation of α). Line colors indicate the seven trait categories; the black line reports the (inverse-variance-weighted) averages, with gray boxes providing 95% confidence intervals for these averages. Numerical values are provided in **Supplementary Table 3**. (**b**) Phenotypes were simulated with 1,000 causal SNPs and $h^2_{SNP} = 0.8$ (black horizontal line) and then analyzed using GCTA, GCTA-MS, GCTA-LDMS, LDAK and LDAK-MS (LDAK with SNPs partitioned by MAF). Bars report average h^2_{SNP} across 200 simulated phenotypes (vertical lines provide 95% confidence intervals). Left, copying the study of Yang *et al.*[15], causal SNP effect sizes are sampled from $N(0,1)$, similar to the GCTA model. Right, causal SNP effect sizes are sampled from $N(0,w_j)$, similar to the LDAK model.

than GCTA), whether we used $\alpha = -1$ (instead of $\alpha = -0.25$) or whether we analyzed the UCLEB traits, it remained the case that the LDAK model better predicted the heritability contribution of each tranche than the GCTA model.

Relationship between heritability and genotype certainty

The LDAK model assumes that SNP heritability contributions vary with genotype certainty (measured by the information score, r_j). Thus far, our analyses have used only very high-quality SNPs ($r_j \geq 0.99$), so this assumption has been redundant. We now also include lower-quality, common SNPs; we focus on the UCLEB traits, as for these we were earlier able to test for correlation between genotyping errors and phenotype (**Supplementary Fig. 13**). **Supplementary Table 5** compares model fit with and without allowance for genotype certainty; it shows that including r_j in the heritability model tends to provide a modest improvement in model fit, resulting in a higher likelihood for 18 of the 31 traits.

Estimates of h^2_{SNP} for the GWAS traits

Table 1 presents our final estimates of h^2_{SNP} for the 19 GWAS traits, obtained using the LDAK model (with $\alpha = -0.25$). For comparison, we include previously reported estimates of h^2_{SNP}, as well as the proportion of phenotypic variance explained by SNPs reported as genome-wide significant (**Supplementary Table 6**). For the disease traits, estimates are on the liability scale, obtained by scaling according to the observed case/control ratio and (assumed) trait prevalence[26,27]. We were unable to find previous estimates of h^2_{SNP} for tuberculosis or intraocular pressure, indicating that, for these two traits, we are the first to establish that common SNPs contribute sizable heritability. Extended results are provided in **Supplementary Table 7**. These show that our final estimates of h^2_{SNP} were on average 43% (s.d. 3%) and 25% (s.d. 2%) higher, respectively, than those obtained using the original versions (with $\alpha = -1$) of GCTA[28] and GCTA-LDMS[15]. Results for the UCLEB traits are provided in **Supplementary Table 1**.

Role of DNase I hypersensitivity sites

Gusev *et al.*[7] used SNP partitioning to assess the contributions of SNP classes defined by functional annotations. Across 11 diseases, they concluded that the majority of h^2_{SNP} was explained by DNase I hypersensitivity sites (DHSs), despite these containing fewer than 20% of all SNPs. For **Figure 5**, we performed a similar analysis using the ten traits we had in common with their study (for nine of these, we used the same data). When we copied Gusev *et al.* and assumed the GCTA model with $\alpha = -1$, we estimated that on average DHSs

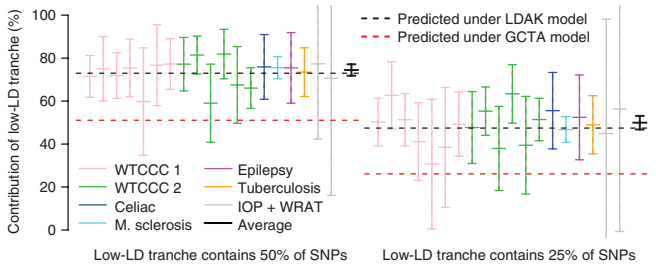

Figure 4 Comparing the GCTA and LDAK models for the GWAS traits. We partition SNPs into those with low or high LD, with the low-LD tranche containing either 50% (left) or 25% (right) of SNPs. For each partition, the horizontal red and black lines indicate the predicted contribution of the low-LD tranche to h^2_{SNP} under the GCTA and LDAK models, respectively. Vertical lines provide point estimates and 95% confidence intervals for the contribution of the low-LD tranche to h^2_{SNP}, estimated assuming the GCTA model. Line colors indicate the seven trait categories, while the black lines provide the (inverse-variance-weighted) averages.

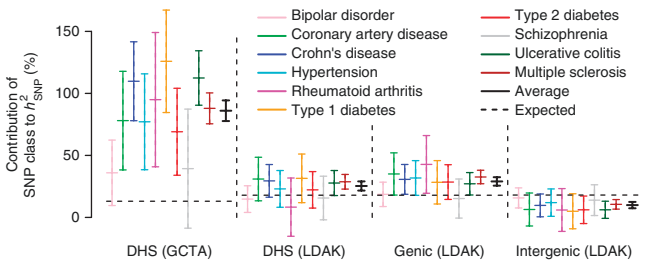

Figure 5 Enrichment of SNP classes. Block 1 reports the contributions to h^2_{SNP} of DHSs, estimated under the GCTA model with $\alpha = -1$ (see the main text for explanation of α). The vertical lines provide point estimates and 95% confidence intervals for each trait and for the (inverse-variance-weighted) average; for three of the traits, the point estimate is above 100%, as was also the case for Gusev *et al.*[7]. Block 2 repeats this analysis but now assuming the LDAK model with $\alpha = -0.25$. Blocks 3 and 4 estimate the contribution of genic SNPs (those inside or within 2 kb of an exon) and intergenic SNPs (further than 125 kb from an exon), again assuming the LDAK model with $\alpha = -0.25$. To assess enrichment, estimated contributions are compared to those expected under the GCTA or LDAK model, as appropriate (horizontal lines).

contributed 86% (s.d. 4%) of h^2_{SNP}, close to the value they reported (79%). When instead we assumed the LDAK model (with $\alpha = -0.25$), the estimated contribution of DHSs was reduced to 25% (s.d. 2%). Under the LDAK model, DHSs were predicted to contribute 18% of h^2_{SNP}, so 25% represents a 1.4-fold enrichment. To add context, we also considered 'genic' SNPs, which we define as SNPs inside or within 2 kb of an exon (using RefSeq annotations[29]), and 'intergenic' SNPs further than 125 kb from an exon; these definitions ensure that these two SNP classes are also predicted to contribute 18% of h^2_{SNP} under the LDAK model. We estimated that genic SNPs contributed 29% (s.d. 2%), while intergenic SNPs contributed 10% (s.d. 2%), representing 1.6-fold and 0.6-fold enrichment, respectively. When we extended this analysis to all 42 traits, DHSs on average contributed 24% (s.d. 2%) of h^2_{SNP}, and, in contrast to Gusev *et al.*, enrichment remained constant when we reduced SNP density (**Supplementary Figs. 14** and **15**, and **Supplementary Table 8**).

Finucane *et al.*[30] performed a similar analysis but considered 52 SNP classes and estimated enrichment using LDSC; across nine traits, they identified five classes with >4-fold enrichment, the highest of which, 'conserved SNPs', had 13-fold enrichment. When we used LDAK to estimate enrichment for our 19 GWAS traits, the results were more modest; the highest enrichment was 2.5-fold, with only 1.3-fold enrichment for conserved SNPs (**Supplementary Fig. 16**).

Relaxing quality control

For the UCLEB data, we considered nine alternative SNP filtering settings. **Supplementary Figure 17** reports estimates of h^2_{SNP} for each trait–filtering combination, while **Figure 6a** provides a summary. First, we varied the information score (r_j) threshold to greater than 0.99, 0.95, 0.9, 0.6, 0.3 and 0 (each time continuing to require MAF ≥ 0.01). Simulations suggested that, by including all 8.8 million common SNPs ($r_j \geq 0$) instead of using just the 353,000 high-quality ones ($r_j \geq 0.99$), we can expect estimates of h^2_{SNP} to increase by 50–60% (**Supplementary Fig. 18**). This is similar to what we observed in practice, as across the 23 traits estimates of h^2_{SNP} (using $\alpha = -0.25$) were on average 45% (s.d. 8%) higher. The simulations further predicted that, even though the Metabochip provides relatively low coverage of the genome (after quality control, it contains only ~60,000 SNPs, predominantly within genes), we can expect estimates of h^2_{SNP} to be approximately 80% as high as those obtained starting from genome-wide genotyping arrays. While we were unable to test this claim directly, it is consistent with our results for height, BMI and QT Interval, the three traits for which reasonably precise estimates of common SNP h^2_{SNP} are available[6] (**Fig. 6b**). For the final three SNP filtering settings, we varied the MAF threshold to be greater than 0.0025, 0.001 and 0.0005 (all with $r_j \geq 0$). Across the 23 traits, we found that rare SNPs contributed substantially to h^2_{SNP}: for example, when we used the 17.3 million SNPs with MAF ≥ 0.0005, estimates of h^2_{SNP} (using $\alpha = -0.25$ and MAF partitioning) were on average 29% (s.d. 12%) higher than those based on the 8.8 million common SNPs (median increase 22%), with rare SNPs contributing on average 33% (s.d. 5%) of h^2_{SNP} (**Fig. 6a**).

DISCUSSION

With estimates of h^2_{SNP} so widely reported, it is easy to forget that calculating the variance explained by large numbers of SNPs is a challenging problem. To avoid overfitting, it is necessary to make strong prior assumptions about SNP effect sizes, but different assumptions can lead to substantially different estimates of h^2_{SNP}. Previous attempts to assess the validity of assumptions have used simulation studies[14,15], but this approach will tend to favor assumptions similar to those used to generate the phenotypes. Instead, we have compared different heritability models empirically, by examining how well they fit real data sets.

We began by investigating the relationship between heritability and MAF. Across 42 traits, we found that the best fit was achieved by setting $\alpha = -0.25$ in model (1), which implies that average heritability varies with $(MAF(1 - MAF))^{0.75}$. As explained in the Online Methods, the value of α corresponds to the scaling of genotypes. Therefore, our result indicates that the performance (detection power and/or prediction accuracy) of many penalized and Bayesian regression methods, for example, Lasso, ridge regression and BayesA[31,32], could be improved simply by changing how genotypes are scaled. Although we recommend $\alpha = -0.25$ as the default value, with sufficient data available, it should be possible to estimate α on a trait-by-trait basis or to investigate more complex relationships between heritability and MAF. In particular, with a better understanding of the relationship between heritability and MAF for low frequencies, it may no longer be necessary to partition by MAF when rare SNPs are included.

We also examined the relationship between heritability and LD. Thus far, most estimates of h^2_{SNP} have been based on the GCTA model;

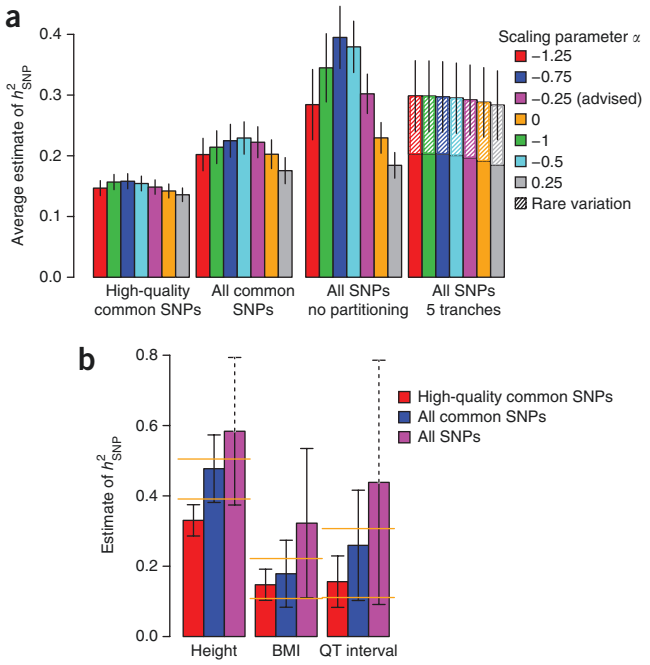

Figure 6 Varying quality control for the UCLEB traits. We consider three SNP filtering settings: 353,000 high-quality common SNPs (information score ≥ 0.99, MAF ≥ 0.01), 8.8 million common SNPs (MAF ≥ 0.01) and all 17.3 million SNPs (MAF ≥ 0.0005). (**a**) Blocks indicate SNP filtering; bars report (inverse-variance-weighted) average estimates of h^2_{SNP} using LDAK (vertical lines provide 95% confidence intervals). Bar color indicates the value of α used. For blocks 1–3, h^2_{SNP} is estimated using the non-partitioned model. For block 4, SNPs are partitioned by MAF; we find this is necessary when rare SNPs are included and it also allows estimation of the contribution of SNPs with MAF < 0.01 (hatched areas). (**b**) Bars report our final estimates of h^2_{SNP} for height, BMI and QT interval—the three traits for which common SNP heritability has previously been estimated with reasonable precision[6] (orange lines mark the 95% confidence intervals from these previous studies). Bar colors now indicate SNP filtering; all estimates are based on $\alpha = -0.25$, using either a non-partitioned model (red and blue bars) or with SNPs partitioned by MAF (purple bars).

this model can be motivated by a belief that each SNP is expected to have the same effect on the phenotype, from which it follows that the expected heritability of a region should depend on the number of SNPs it contains. By contrast, the LDAK model views highly correlated SNPs as tagging the same underlying variant and therefore believes that the expected heritability of a region should vary according to the total amount of distinct genetic variation it contains. Across our traits, we found that the relationship between heritability and LD specified by the LDAK model consistently provided a better description of reality.

This finding has important consequences for complex trait genetics. First, it implies that, for many traits, common SNPs explain considerably more phenotypic variance than previously reported, which represents a major advance in the search for missing heritability[2]. It also affects a large number of closely related methods. For example, LDSC[10], like GCTA, assumes that heritability contributions are independent of LD, and it therefore also tends to underestimate h^2_{SNP}. Similarly, we have shown that estimates of the relative importance of SNP classes via SNP partitioning can be misleading when the GCTA model is assumed[7,30]. Further afield, most software for mixed-model association analyses (for example, FAST-LMM, GEMMA, MLM-LOCO and BOLT) use an extension of the GCTA model[33–36], which is also the case for most bivariate analyses, including those performed by LDSC[8,37,38]. It remains to be seen how much these methods would be affected if they employed more realistic heritability models.

Attempts have been made to improve the accuracy of heritability models via SNP partitioning[14,15,39]. We find that partitioning by MAF can be advantageous, as it guards against misspecification of the relationship between heritability and MAF when rare variants are included. **Figure 3a** and **Supplementary Figure 7** indicate that the realism of the GCTA model can be improved by partitioning based on LD; for example, across the GWAS traits, estimates from GCTA-LDMS are on average 16% (s.d. 2%) higher than those from GCTA and only 23% (s.d. 2%) lower than those from LDAK. The improvement arises because model misspecification is reduced by allowing SNPs in lower-LD tranches to have higher average heritability. However, **Supplementary Table 9** illustrates why we consider such an approach suboptimal; in particular, SNP partitioning can be computationally expensive and, even with LD partitioning, model fit tends to be worse than that from LDAK.

While we have investigated the role of MAF, LD and genotype certainty, there remain other factors on which heritability could depend, in particular the available functional annotations of genomes[40]. For example, our comparison of genic and intergenic SNPs indicates that the effect size prior distribution could be improved by taking into account proximity to coding regions. By way of demonstration, **Supplementary Table 10** shows that model fit is improved by assuming $E[h_j^2] = c_k \times [f_j(1-f_j)]^{1+\alpha} \times w_j \times r_j \times \exp(-(D_j+50)/500)$ where D_j is the distance (in kb) between SNP j and the nearest exon (under this model, genic SNPs are expected to have about twice the heritability of intergenic SNPs). In general, we believe that modifications of this type will have a relatively small impact; we note that, across the 19 GWAS traits, this modification increases model log likelihood by on average only 1.5, much less than the average increase obtained by using $\alpha = -0.25$ instead of $\alpha = -1$ (8.9) or by choosing the LD model specified by LDAK instead of GCTA (17.7), and does not significantly change estimates of h^2_{SNP}. However, with sufficient data, it may be possible to obtain more substantial improvement by tailoring model assumptions to individual traits.

When estimating h^2_{SNP}, care should be taken to avoid possible sources of confounding. Previously, we advocated a test for inflation of h^2_{SNP} due to population structure and familial relatedness[3]. The conclusions of a recent paper claiming that h^2_{SNP} estimates are unreliable[41] would have changed substantially had this test been applied (**Supplementary Fig. 19**). We also recommend testing for inflation due to genotyping errors, particularly before including lower-quality and/or rare SNPs. For the 23 UCLEB traits, we showed that including poorly imputed SNPs resulted in significantly higher estimates of h^2_{SNP} and made it possible to capture the majority of genome-wide heritability, despite the very sparse genotyping provided by the Metabochip. We found that including rare SNPs also led to significantly higher h^2_{SNP}. Although sample size prevented us from obtaining precise estimates of h^2_{SNP} for individual traits, our analyses indicate that, for larger data sets, including rare SNPs will be both practical and fruitful in the search for the remaining missing heritability[2].

URLs. LDAK, http://www.ldak.org/; PLINK, http://www.cog-genomics.org/plink2; SHAPEIT, http://www.shapeit.fr/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; DHS annotations, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz; RefSeq annotations, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

Access to Wellcome Trust Case Control Consortium data was authorized as work related to the project "Genome-wide association study of susceptibility and clinical phenotypes in epilepsy," while access to Children's Hospital of Philadelphia (CHOP) data was granted under Project 49228-1, "Assumptions underlying estimates of SNP heritability." We thank A. Molloy, J. Mills and L. Brody for permission to use genotype data from the Trinity College Dublin Student Study and S. Langley for help accessing the CHOP data. This work is funded by the UK Medical Research Council under grant MR/L012561/1 (awarded to D.S.) and the British Heart Foundation under grant RG/10/12/28456 (the UCLEB Consortium) and is supported by researchers at the National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre. N.C. is an ESPOD Fellow from the European Molecular Biology Laboratory, European Bioinformatics Institute, and Wellcome Trust Sanger Institute. M.R.J. receives funding from the Imperial College NIHR Biomedical Research Centre (BRC) Scheme. S.N. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science and is also supported by the NIHR Cambridge Biomedical Research Centre. Analyses were performed with the use of the UCL Computer Science Cluster and the help of the CS Technical Support Group, as well as the use of the UCL Legion High-Performance Computing Facility (Legion@UCL) and associated support services.

AUTHOR CONTRIBUTIONS

D.S. and N.C. performed the analyses. D.S. and D.J.B. wrote the manuscript with assistance from N.C., M.R.J., S.N. and members of the UCLEB Consortium.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
2. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
3. Speed, D. *et al.* Describing the genetic architecture of epilepsy through heritability analysis. *Brain* **137**, 2680–2689 (2014).
4. Henderson, C., Kempthorne, O., Searle, S. & von Krosigk, C. The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218 (1959).
5. Falconer, D. & Mackay, T. *Introduction to Quantitative Genetics* 4th edn (Longman, 1996).
6. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
7. 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).
8. 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).
9. 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).
10. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
11. Bulik-Sullivan, B. Relationship between LD score and Haseman–Elston regression. Preprint at *bioRxiv* http://dx.doi.org/10.1101/018283 (2015).
12. Corbeil, R. & Searle, S. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* **18**, 31–38 (1976).
13. 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).
14. Lee, S.H. *et al.* Estimation of SNP heritability from dense genotype data. *Am. J. Hum. Genet.* **93**, 1151–1155 (2013).
15. 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).
16. Ek, W.E. *et al.* Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. *J. Natl. Cancer Inst.* **105**, 1711–1718 (2013).
17. Bevan, S. *et al.* Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. *Stroke* **43**, 3161–3167 (2012).
18. Keller, M.F. *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).
19. Yin, X. *et al.* Common variants explain a large fraction of the variability in the liability to psoriasis in a Han Chinese population. *BMC Genomics* **15**, 87 (2014).
20. Lee, S.H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
21. Chen, G.B. *et al.* Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and Immunochip data. *Hum. Mol. Genet.* **23**, 4710–4720 (2014).
22. Stahl, E.A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).
23. Robinson, E.B. *et al.* The genetic architecture of pediatric cognitive abilities in the Philadelphia Neurodevelopmental Cohort. *Mol. Psychiatry* **20**, 454–458 (2015).
24. Shah, T. *et al.* Population genomics of cardiometabolic traits: design of the University College London–London School of Hygiene and Tropical Medicine–Edinburgh–Bristol (UCLEB) Consortium. *PLoS One* **8**, e71345 (2013).
25. Voight, B.F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
26. Dempster, E.R. & Lerner, I.M. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).
27. 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).
28. 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)..
29. Pruit, K., Brown, G., Tatusova, T. & Maglott, D. in *The NCBI Handbook* (eds. McEntyre, J. & Ostell, J.) Chapter. 18 (National Center for Biotechnology Information, 2002).
30. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
31. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, 2001).
32. Habier, D., Fernando, R.L., Kizilkaya, K. & Garrick, D.J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186 (2011).
33. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
34. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
35. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).
36. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
37. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
38. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
39. Gazal, S. *et al.* Linkage disequilibrium dependent architecture of human complex traits reveals action of negative selection. Preprint at *bioRxiv* http://dx.doi.org/10.1101/082024 (2017).
40. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
41. 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).

ONLINE METHODS

The **Supplementary Note** summarizes the different analyses we performed and the conclusions we drew from each. In general, we assume there are n individuals, recorded for p covariates and genotyped (either directly or via imputation) for m SNPs: the length–n vector $\mathbf{Y}$ contains phenotypic values the $n \times p$ matrix $\mathbf{Z}$ contains covariates, and the $n \times m$ matrix $\mathbf{S}$ contains (expected) allele counts.

Information score r_j. Let the vector $\mathbf{S}_j = (S_{1,j}, \ldots, S_{n,j})^{\mathrm{T}} \in [0,2]^n$ denote the allele counts for SNP j ($\mathbf{S}_j$ is column j of $\mathbf{S}$). Our information score r_j estimates the squared correlation between $\mathbf{S}_j$ and $\mathbf{G}_j = (G_{1,j}, \ldots, G_{n,j})^{\mathrm{T}} \in \{0,1,2\}^n$, the true genotypes for SNP j. When using imputed data, $\mathbf{G}_j$ is typically not known; instead, for each individual, we have a triplet of state probabilities $p_{i,j,0}$, $p_{i,j,1}$, $p_{i,j,2}$, where $p_{i,j,g} = P(G_{1,j} = g)$ and $p_{i,j,0} + p_{i,j,1} + p_{i,j,2} = 1$. Therefore, we define r_j by taking expectations over the 3^n possible realizations of $\mathbf{G}_j$

$$r_j = \frac{E\left[\sum\left(S_{i,j} - \bar{S}_j\right)\left(G_{i,j} - \bar{G}_j\right)\right]^2}{\left(\sum\left(S_{i,j} - \bar{S}_j\right)^2\right) E\left[\sum\left(G_{i,j} - \bar{G}_j\right)^2\right]}$$

where

$$\bar{S}_j = \frac{1}{n}\sum S_{i,j}$$

and

$$\bar{G}_j = \frac{1}{n}\sum G_{i,j}$$

$\mathbf{S}_j$ is known, so computing $\sum\left(S_{i,j} - \bar{S}_j\right)^2$ is straightforward. The two expectations can also be calculated explicitly

$$E\left[\sum\left(S_{i,j} - \bar{S}_j\right)\left(G_{i,j} - \bar{G}_j\right)\right] = \sum\left(S_j - \bar{S}_j\right) E\left[G_{i,j} - \mu\right]$$
$$= \sum\left(S_j - \bar{S}_j\right)\left(p_{i,j,1} + 2p_{i,j,2} - \mu\right)$$

$$E\left[\sum\left(G_j - \bar{G}_j\right)^2\right] = \sum E\left[\left(G_j - \mu\right)^2\right]$$
$$= \sum\left[p_{i,j,0}\left(-\mu\right)^2 + p_{i,j,1}\left(1-\mu\right)^2 + p_{i,j,2}\left(2-\mu\right)^2\right]$$

where

$$\mu = E\left[\bar{G}_j\right] = \frac{1}{n}\sum\left(p_{i,j,1} + 2p_{i,j,2}\right)$$

For our analyses, we use expected allele counts (dosages), so $S_{i,j} = p_{i,j,1} + 2p_{i,j,2}$. In this case,

$$E\left[\sum\left(S_{i,j} - \bar{S}_j\right)\left(G_{i,j} - \bar{G}_j\right)\right] = \sum\left(S_{i,j} - \bar{S}_j\right)^2$$

and so the score reduces to

$$r_j = \sum\left(S_{i,j} - \bar{S}_j\right)^2 / \sum\left(G_{i,j} - \bar{G}_j\right)^2$$

For a directly genotyped SNP, each triplet of state probabilities will be (1,0,0), (0,1,0) or (0,0,1), which will result in $S_{i,j} = G_{i,j}$ for all i and $r_j = 1$; so for these SNPs, in place of r_j, we use the metric r2_type2 reported by IMPUTE2 (ref. 42). Additional details on our information score are provided in **Supplementary Figure 20**.

Estimating h^2_{SNP}. We first construct the $n \times m$ genotype matrix $\mathbf{X}$, by centering and scaling the allele counts for each SNP according to $X_{ij} = (S_{ij} - 2f_j) \times [2f_j(1 - f_j)]^{\alpha/2}$, where $f_j = \Sigma_i S_{ij}/2n$. If w_j and r_j denote the LD weight[9] and information score for SNP j, then the LDAK model for estimating SNP heritability $h^2_{\mathrm{SNP}} = \sigma^2_{\mathrm{g}}/(\sigma^2_{\mathrm{g}} + \sigma^2_{\mathrm{e}})$ is

$$Y_i = \sum \theta_k Z_{i,k} + \sum \beta_j X_{i,j} + e_i$$
$$\text{with } \beta_j \sim N\left(0, r_j w_j \sigma^2_{\mathrm{g}}/W\right),\ e_i \sim N\left(0, \sigma^2_{\mathrm{e}}\right) \text{ and } W = \sum r_j w_j \left[2f_j\left(1-f_j\right)\right]^{1+\alpha} \quad (2)$$

θ_k denotes the fixed-effect coefficient for the kth covariate and β_j and e_i are random effects indicating the effect size of SNP j and the noise component for individual i, while σ^2_{g} and σ^2_{e} are interpreted as genetic and environmental variances, respectively. Note that the introduction of r_j is an addition to the model we proposed in 2012 (ref. 9). Model (2) is equivalent to assuming[43,44],

$$Y \sim N(Z\theta, K\sigma^2_g + I\sigma^2_e), \quad (3)$$

with

$$K = \frac{X\Omega X^T}{W}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix and Ω denotes a diagonal matrix with diagonal entries $(r_1 w_1, \ldots, r_m w_m)$. The kinship matrix $\mathbf{K}$, also referred to as a genetic relationship matrix (GRM)[1] or genomic similarity matrix (GSM)[45], consists of average allelic correlations across the SNPs (adjusted for LD and genotype certainty). Model (3) is typically solved using REML[12], which returns estimates of $\theta_1, \ldots, \theta_p, \sigma^2_{\mathrm{g}}$ and σ^2_{e} (ref. 12).

The heritability of SNP j can be estimated by $h^2_j = \beta^2_j\,\mathrm{var}(\mathrm{X}_j)/\mathrm{var}(\mathrm{Y})$, which under model (2) and assuming Hardy–Weinberg equilibrium[46,47] has expectation

$$E\left[h^2_j\right] = \frac{E\left[\beta^2_j\right] \times Var\left(X_j\right)}{Var(Y)} = \frac{r_j w_j \sigma^2_g / W \times \left[2f_j\left(1-f_j\right)\right]^{1+\alpha}}{Var(Y)} \quad (4)$$

If P_1 and P_2 index two sets of SNPs of size $|P_1|$ and $|P_2|$, then under the LDAK model they are expected to contribute heritability in the ratio $W_1{:}W_2$, where

$$W_l = \sum r_j w_j \left[2f_j\left(1-f_j\right)\right]^{1+\alpha}$$

The GCTA model corresponds to setting $w_j = r_j = 1$, in which case

$$W_l = \sum \left[2f_j\left(1-f_j\right)\right]^{1+\alpha}$$

Most applications of GCTA have further assumed $\alpha = -1$, so that $W_1 = |P_1|$, which corresponds to the assumption that SNP sets are expected to contribute heritability proportional to the number of SNPs they contain.

Model (2) assumes that all effect sizes can be described by a single prior distribution. This assumption is relaxed by SNP partitioning. Suppose that the SNPs are divided into tranches $P_1 \ldots, P_L$ of sizes $|P_1|, ., |P_L|$; typically, these will partition the genome so that each SNP appears in exactly one tranche and $\Sigma_l |P_l| = m$, but this is not required. This corresponds to generalizing model (2), so that SNPs in tranche l have effect size prior distribution $\beta_j \sim N\left(0, r_j w_j \sigma^2_l / W_l\right)$. Letting $\Sigma = \sigma^2_1, \ldots, \sigma^2_L$ then $h^2_{\mathrm{SNP}} = \Sigma / \left(\Sigma + \sigma^2_{\mathrm{e}}\right)$ and σ^2_1 / Σ represents the contribution to h^2_{SNP} of SNPs in tranche l. This model can equivalently be expressed as $\mathbf{Y} \sim N\left(\mathbf{Z\theta}, \mathbf{K}\sigma^2_1 + \ldots + \mathbf{K}_L\sigma^2_L + \mathbf{I}\sigma^2_e\right)$, where $\mathbf{K}_1$ represents allele correlations across the SNPs in tranche 1.

For analyses under the LDAK model, we used LDAK v.5; for analyses under the GCTA model, we used GCTA v.1.26. For about one-third of GCTA-LDMS analyses, the GCTA REML solver failed with the error "information matrix is not invertible," in which case we reran the analysis using LDAK (while the GCTA and LDAK solvers are both based on average information REML[28,48], subtle differences mean that, when using a large number of tranches, one might complete while the other fails). For the few occasions when both solvers failed, we instead used GCTA-LD (SNPs divided only by LD, rather than by LD and MAF), which we found gave very similar results to GCTA-LDMS for traits where both completed (**Supplementary Fig. 7**). For diseases, we converted estimates of h^2_{SNP} to the liability scale on the basis of the observed case/control ratio and assumed prevalence[26,27]. In general, we copied the prevalences used by previous studies; however, for tuberculosis, where no previous estimate of h^2_{SNP} was available, we derived an estimate of prevalence from World Health Organization data[49] (**Supplementary Note**).

LDSC. Originally designed as a way to quantify confounding in a GWAS, LDSC[10] also provides a method for estimating h^2_{SNP}, which requires only summary statistics from single-SNP analysis (rather than raw genotype and phenotype data).

LDSC is based on the principle that, in a single-SNP analysis, the $\chi^2(1)$ test statistic for SNP j has expected value $E[X^2(1)] = 1 + nh_j^2 + n\Sigma_{k \neq j} r_{j,k}^2 + na_j$, where $r_{j,k}^2$ denotes the squared correlation between SNPs j and k, while a_j represents bias due to confounding factors (for example, population structure and familial relatedness)[10]. Under a polygenic model where every SNP is expected to contribute equally ($E[h_j^2] = h_{SNP}^2 / m$) and the (widely used) assumption that bias is constant across SNPs ($a_j = a$), we have $E[X^2(1)] = 1 + nl_j h_{SNP}^2 / m + na$, where $l_j = 1 + \Sigma_{k \neq j} r_{j,k}^2$ is referred to as the LD score of SNP j (as it is not feasible to compute pairwise correlations across all SNPs, in practice these are approximated using a sliding window of, say, 1 cM). Therefore, LDSC estimates h_{SNP}^2 and a by regressing test statistics on LD scores. In the absence of confounding ($a = 0$), LDSC can be viewed as estimating h_{SNP}^2 under the GCTA model with $\alpha = -1$ (as this satisfies the assumption that every SNP is expected to contribute equal heritability). As the authors of LDSC point out[10], it is straightforward to accommodate alternative relationships between $E[h_j^2]$ and MAF ($\alpha \neq -1$) by changing how genotypes are scaled when computing LD scores, and genotype certainty could potentially be accommodated. However, the similarity with the GCTA model appears intrinsic to LDSC; while the assumption that heritability is independent of LD can be relaxed via SNP partitioning[39], we cannot envisage how the method could be modified to accommodate the LDAK SNP weights. For LDSC analyses, we used LDSC v.1.0.0 both for calculating LD scores and estimating h_{SNP}^2.

Accommodating loci with very large effects. Equation (2) assumes that all SNP effect sizes can be modeled by a single Gaussian distribution. Estimates are generally robust to violations of this assumption[9], but problems can occur when individual SNPs have very large effect sizes because a single Gaussian distribution cannot accommodate both these SNPs and the very many with small effect sizes. This is a common concern when analyzing autoimmune traits, for which the MHC can contribute substantial heritability. In response to this problem, some authors exclude MHC SNPs from analyses[7,28,50,51]. Another approach is to model effect sizes as a mixture of Gaussians[52,53], but this is not computationally feasible for millions of SNPs and many thousands of individuals. Therefore, our proposed strategy is to first identify SNPs with $P < 1 \times 10^{-20}$ from single-SNP analysis, to prune these using a correlation-squared threshold of 0.5 and then to include those that remain as fixed-effect covariates. Thus, in place of equation (3), we assume $\mathbf{Y} \sim N\left(\mathbf{Z\theta} + \mathbf{T\Phi}, \mathbf{K}\sigma_g^2 + \mathbf{I}\sigma_e^2\right)$, where columns of the matrix $\mathbf{T}$ contain allele counts of the highly associated SNPs (that is, $\mathbf{T}$ is a submatrix of $\mathbf{S}$) and the vector Φ represents their effect sizes. In contrast to standard (non-SNP) covariates, the variance explained by $\mathbf{T}$ counts toward SNP heritability: $h_{SNP}^2 = (\sigma_g^2 + \sigma_T^2)/(\sigma_g^2 + \sigma_T^2 + \sigma_e^2)$, where $\sigma_T^2 = (\mathbf{T\Phi})^T (\mathbf{T\Phi})$. **Supplementary Figures 21** and **22** provide further details. In particular, we appreciate that our definition of highly associated is somewhat arbitrary, so we confirm that estimates of h_{SNP}^2 are almost unchanged if instead we use $P < 5 \times 10^{-8}$.

Data sets. When searching for GWAS data sets, we preferred those with sample size at least 4,000 to ensure reasonable precision of h_{SNP}^2 (ref. 54). In total, our data sets were constructed from 40 independent cohorts, all of which have been previously described (see **Supplementary Tables 11** and **12** for references and details of how cohorts were merged to form data sets). For the UCLEB data, there were in total 28 quantitative traits with measurements recorded for 7,000 individuals. For each of these, we quantile normalized, then applied a test for inflation due to genotyping errors (**Supplementary Fig. 13**). Specifically, our test, inspired by Bhatia *et al.*[55] and valid for quantitative phenotypes where individuals are recruited from multiple cohorts, first estimates h_{SNP}^2 using only pairs of individuals in different cohorts and then using only pairs of individuals in the same cohort; a significant difference between the two estimates indicates possible inflation due to genotyping errors. We excluded five traits that showed evidence of inflation ($P < 0.05/28$), leaving us with 23: height, weight, BMI, waist circumference, forced vital capacity, 1-s forced vital capacity, systolic blood pressure (adjusted), diastolic blood pressure (adjusted), PR interval, QT interval, corrected QT interval, QRS voltage product, Sokolow Lyon, glucose, insulin, total cholesterol (adjusted), LDL cholesterol (adjusted), triglycerides (adjusted), viscosity, fibrinogen, IL-6, C-reactive protein and hemoglobin. Approximately 40% of individuals were receiving medication to reduce blood pressure and 25% to reduce lipid levels, so, where indicated, phenotypes had been adjusted for this: for individuals on medication, their raw measurements had been increased either by adding on (blood pressure) or scaling by (lipid levels) a constant[56,57]. We note that some pairs of traits are highly correlated. However, as the overall correlation is not that extreme (we estimate the effective number of independent traits to be about 15) and most of our UCLEB analyses serve to support conclusions drawn from the GWAS traits, we decided to retain all 23 traits (rather than, say, consider only a subset). See the **Supplementary Note** for further details on phenotyping.

Quality control. We processed each of the 40 cohorts in identical fashion; see the **Supplementary Note** for full details. In summary, after excluding apparent population outliers, samples with extreme missingness or heterozygosity and SNPs with MAF < 0.01, call rate < 0.95 or $P < 1 \times 10^{-6}$ from a test for Hardy–Weinberg equilibrium, we phased using SHAPEIT[58] and then imputed using IMPUTE2 (ref. 42) and the 1000 Genomes Project Phase 3 (2014) reference panel[59]. When merging cohorts to construct the GWAS data sets, we retained only autosomal SNPs that in all cohorts had MAF ≥ 0.01 and $r_j \geq 0.99$ (using IMPUTE2 r2_type2 in place of r_j for directly genotyped SNPs). For the eight UCLEB cohorts, we applied these filters only after merging. We only relaxed quality control for the analyses of the UCLEB data where we explicitly examined the consequences of including lower-quality and rare SNPs. When possible, the matrix $\mathbf{S}$ contained expected allele counts (dosages); that is, $S_{i,j} = p_{i,j,1} + 2p_{i,j,2}$, where $p_{i,j,1}$ and $p_{i,j,2}$ denote the probabilities of allele counts 1 and 2, respectively. If hard genotypes were required, for example, when using LDSC to compute LD scores[10], we rounded $S_{i,j}$ to the nearest integer. As this was only necessary when considering high-quality SNPs ($r_j \geq 0.99$), we expect this rounding to have negligible impact on results. For each trait, **Table 1** reports m, the total number of SNPs after imputation, and $\Sigma_{j=1}^{m} w_j$, the sum of SNP weights; the aim of these weights is to remove duplication of signal due to LD, and their sum can loosely be interpreted as an effective number of independent SNPs. For the GWAS data sets, $\Sigma_j w_j$ ranged from 79,000 to 125,000. By contrast, when restricted to only high-quality SNPs, the UCLEB data had $\Sigma_j w_j = 39{,}000$, reflecting that the Metabochip directly captures a much smaller amount of genetic variation than standard genome-wide SNP arrays.

When analyzing quantitative traits, genotyping errors will tend only to be a concern when there are systematic differences between phenotypes across cohorts, and this is something we are able to explicitly test (**Supplementary Fig. 13**). However, for disease traits, when cases and controls have been genotyped separately (as was the design of most of our GWAS data sets), any errors will almost certainly correlate with phenotype and therefore cause inflation of h_{SNP}^2[9,27]. To test the effectiveness of our quality control for the GWAS traits, we constructed a pseudo case–control study using two control cohorts; we confirmed that the resulting estimate of h_{SNP}^2 was not significantly greater than zero, suggesting that the quality control steps we used for the GWAS data sets were sufficiently strict (**Supplementary Note**).

Accurate estimation of h_{SNP}^2 requires samples of unrelated individuals with similar ancestry. Prior to imputation, we removed ancestry outliers identified through principal-component analyses (**Supplementary Fig. 23**). After imputation, we computed (unweighted) allelic correlations using a pruned set of SNPs and then filtered individuals so that no pair remained with correlation greater than c, where $-c$ is the smallest observed pairwise correlation (c ranged from 0.029 to 0.038, depending on the data set). For our data sets, this filtering excluded relatively few individuals (on average 3.8%, with maximum 11.6%). For all analyses, we included a minimum of 30 covariates: the top 20 eigenvectors from the allelic correlation matrix just described and projections onto the top 10 principal components computed from 1000 Genomes Project samples[59]. For the 19 GWAS traits, we also included sex as a covariate, while for intraocular pressure and wide-range achievement test scores we additionally included age. **Supplementary Figure 24** reports the proportion of phenotypic variance explained by each covariate. To check our filtering and covariate choices, we estimated the inflation of h_{SNP}^2 due to population structure and residual relatednesss[3] (**Supplementary Fig. 19**). For the GWAS traits, we estimated that on average h_{SNP}^2 estimates were inflated by at most 3.1%, with the highest observed for ischemic stroke (7.1%). For the 23 UCLEB traits, the average inflation was 0.3% (highest 2.3%).

doi:10.1038/ng.3865

Single-SNP analysis. Supplementary Figure 25 provides Manhattan plots from logistic (case–control traits) and linear regression (quantitative traits), performed using PLINK v.1.9. These analyses provide the summary statistics required by LDSC. For the GWAS traits, we identified highly associated SNPs ($P < 1 \times 10^{-20}$) within the MHC for six of the GWAS traits (rheumatoid arthritis, type 1 diabetes, psoriasis, ulcerative colitis, celiac disease and multiple sclerosis), while rs2476601, a SNP within *PTPN22*, was highly associated with both rheumatoid arthritis and type 1 diabetes[60,61]. For the UCLEB traits, we found highly associated SNPs within *SCN10A* (PR interval), *APOE* (total cholesterol, LDL cholesterol and C-reactive protein) and *ZPR1* (triglyceride levels). For heritability analysis, these SNPs were pruned and then included as additional fixed-effect covariates as described above.

Computational requirements. The most time-consuming aspect of analysis was genotype imputation; for a typically sized cohort (~3,000 individuals), this took approximately 1 CPU-year (a few days on a 100-node cluster). Next is computation of SNP weights, which for the GWAS traits (~4 million SNPs) took approximately 1 CPU-day (again, this can be nearly perfectly parallelized). Finally, solving the mixed model via REML would take between a few minutes for the smaller traits (~5,000 individuals) and a few hours for the largest (~14,000 individuals). Memory-wise, the most onerous task is solving the mixed model, for which memory demands scale with n^2; however, even for the largest data set, this was less than 5 GB (when using multiple kinship matrices, LDAK allows for these to be read on the fly, so that the memory demands are no higher than when using only one).

Code availability. Step-by-step instructions for estimating h^2_{SNP} starting from raw genotype data, as well as for performing our other analyses, are provided in the **Supplementary Note**.

Data availability. In total, we analyze data from 40 cohorts; 25 of these were downloaded (after completing a data access request) from the European Genome-phenome Archive or dbGaP, while the remaining 15 (which include the 8 UCLEB cohorts) were obtained directly from the relevant custodians. Full details of the cohorts (with accession codes where applicable) are provided in the **Supplementary Note**.

42. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* **1**, 457–470 (2011).
43. 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).
44. Habier, D., Fernando, R.L. & Dekkers, J.C. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**, 2389–2397 (2007).
45. Speed, D. & Balding, D.J. Relatedness in the post-genomic era: is it still useful? *Nat. Rev. Genet.* **16**, 33–44 (2015).
46. Hardy, G.H. Mendelian proportions in a mixed population. *Science* **28**, 49–50 (1908).
47. Weinberg, W. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins fur Vaterländische Naturkd. Württemb.* **64**, 368–382 (1908).
48. Lee, S.H. & van der Werf, J.H. An efficient variance component approach implementing an average information REML suitable for combined LD and linkage mapping with a general complex pedigree. *Genet. Sel. Evol.* **38**, 25–43 (2006).
49. World Health Organization. *Global Tuberculosis Report* (World Health Organization, 2014).
50. Gusev, A. *et al.* Quantifying missing heritability at known GWAS loci. *PLoS Genet.* **9**, e1003993 (2013).
51. Speed, D. & Balding, D.J. MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res.* **24**, 1550–1557 (2014).
52. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
53. Moser, G. *et al.* Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genet.* **11**, e1004969 (2015).
54. 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).
55. Bhatia, G. *et al.* Haplotypes of common SNPs can explain missing heritability of complex diseases. Preprint at *bioRxiv* http://dx.doi.org/10.1101/022418 (2016).
56. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. & Burton, P.R. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.* **24**, 2911–2935 (2005).
57. Asselbergs, F.W. *et al.* Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am. J. Hum. Genet.* **91**, 823–838 (2012).
58. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
59. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
60. Todd, J.A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
61. Plenge, R.M. *et al. TRAF1—C5* as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).