Re-evaluation of SNP heritability in complex human traits: Supplementary Material

Doug Speed,[1,*] Na Cai,[2,3] the UCLEB Consortium,[4] Michael R. Johnson,[5] Sergey Nejentsev,[6] David J Balding.[1,7]
[*]Corresponding author: doug.speed@ucl.ac.uk

[1]*UCL Genetics Institute, University College London, United Kingdom.*
[2]*Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.*
[3]*European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom.*
[4]*A full list of members and affiliations appears in the Supplementary Material.*
[5]*Division of Brain Science, Imperial College London, United Kingdom.*
[6]*Department of Medicine, University of Cambridge, United Kingdom.*
[7]*Centre for Systems Genomics, School of BioSciences and School of Mathematics & Statistics, University of Melbourne, Australia.*

Supplementary Note

Web Resources:

LDAK: `www.ldak.org`
PLINK: `www.cog-genomics.org/plink2`
SHAPEIT: `www.shapeit.fr`
IMPUTE2: `mathgen.stats.ox.ac.uk/impute/impute_v2.html`
Genome Browser: `https://genome.ucsc.edu`
LiftOver: `http://genome.sph.umich.edu/wiki/LiftOver`
Three files referred to in Supplementary Note 1:
`ftp.ncbi.nih.gov/snp/organisms/human_9606/database/organism_data/RsMergeArch.bcp.gz`
`ftp.ncbi.nih.gov/snp/organisms/human_9606/database/organism_data/SNPHistory.bcp.gz`
`www.ldak.org/Imputation-Regions`

Supplementary Note 1: Step-by-step guide for for estimating h^2_{SNP} from raw genotype data. This guide is also available at `www.ldak.org/protocol`; we recommend you use the online version which will be kept up-to-date. This guide is divided into four steps: quality control & imputation, merging cohorts, estimating h^2_{SNP} and advanced analyses. It is designed to be run as bash scripts on a UNIX operating system, but should also work in Terminal on a MAC (LDAK is not compatible with Windows). In addition to awk, it uses LDAK (version 5), PLINK (v. 1.9), SHAPEIT (v. 2.20) and IMPUTE2 (v. 2.3.2); see Web Resources for the software homepages and links to files. This guide also requires a (phased) reference panel; we used 1000 Genomes Phase 3 (October 2014), which we downloaded from the IMPUTE2 website. A backslash (\) at the end of a line indicates a command spans multiple lines. Many of the commands can be run in parallel; access to a computer cluster is highly recommended when performing imputation and when calculating SNP weightings for dense data.

Step 1 - Imputation. Suppose the genotypes for a cohort are stored in PLINK binary format: `data.bed`, `data.bim`, `data.fam`. First, we performed quality control of samples (based on heterozygosity and missingness) and SNPs (MAF, call-rate, HWE). To identify ancestry, we compared samples with HapMap data, stored in `hapmap.bed`, `hapmap.bim`, `hapmap.fam`.

```
#Sample quality control - we suggest using pruned, high-quality, autosomal SNPs
plink --bfile data --indep-pairwise 50 10 .2 --maf 0.01 --geno 0.05 --autosome --out autoQC
plink --bfile data --extract autoQC.prune.in --missing --out stats
plink --bfile data --extract autoQC.prune.in --het --out stats
#Based on stats.het and stats.imissing, identify outlying samples
#Save those to retain in the file keepqc.ind

#Check / determine sex (for human samples, this requires Chromosome 23 data)
plink --bfile data --indep-pairwise 50 10 .2 --maf 0.01 --geno 0.05 --chr 23 --out sexQC
plink --bfile data --extract sexQC.prune.in --check-sex --out sex

#Get (non-ambiguous) SNPs in common with HapMap, compute population axes and projections
#These scripts use awk, installed by default in unix - see www.dougspeed.com/awk for a brief guide
awk '(($5!="A"||$6!="T")&&($5!="T"||$6!="A")&&($5!="C"||$6!="G")&&($5!="G"||$6!="C")){print $4}' \
hapmap.bim > nonamb.snps
awk '(NR==FNR){arr[$1];next}($4 in arr && $1<23){print $2}' nonamb.snps data.bim > hapmap.snps
plink --bfile hapmap --indep-pairwise 50 10 .2 --maf 0.05 --extract hapmap.snps --out hapmap
ldak --calc-kins-direct hapmap --bfile hapmap --extract hapmap.prune.in --ignore-weights YES
    --power -0.25
ldak --pca hapmap --grm hapmap --axes 5
ldak --calc-pca-loads hapmap --bfile hapmap --grm hapmap --pcastem hapmap
ldak --calc-scores hapmap --bfile data --scorefile hapmap.load --allow-flips YES --keep keepqc.ind
#Can compare Columns 5 & 7 of hapmap.profile with Columns 3 & 4 of hapmap.vect
#Save those to retain to keepqcpop.ind

#SNP quality control - will use only samples in keepqcpop.ind
plink --bfile data --keep keepqcpop.ind --freq --out statsb
plink --bfile data --keep keepqcpop.ind --missing --out statsb
plink --bfile data --keep keepqcpop.ind --hardy --out statsb
#Based on statsb.frq, statsb.lmissing and statsb.hwe, identify poor quality SNPs
#Commonly used thresholds are MAF>0.01, CR>0.95 & HWE P>1e-6
#Save those to retain in the file keep.snps

#Remake data retaining only good quality samples and SNPs, and updating sex
plink --bfile data --keep keepqcpop.ind --extract keep.snps --update-sex sex.sexcheck 2 \
--make-bed --out clean
```

Next, it is necessary to ensure SNP annotations are correct and consistent with those in the reference panel. We suggest the following checks. **(i)** Update SNP names based on the file RsMergeArch, and exclude those reported as retired in the file SNPHistory (see LiftOver website for an explanation of these two files). **(ii)** If necessary, convert genomic positions to the correct build (our reference panel was Build 37 / hg 19). For this we obtained locations from Genome Browser. If a SNP name was not present in Genome Browser, we would initially exclude, but would re-introduce if present in the reference panel). We would also exclude SNPs whose alleles did not match those recorded in genome browser (allowing for strand flips). **(iii)** Check positions and alleles consistent with reference panel. In our case, the 1000 Genomes annotations agreed very well with those in Genome Browser, so there were typically at most a handful of SNPs with discordant positions or incompatible alleles. **(iv)** Decide whether to retain ambiguous SNPs (those with alleles A & T or C & G). For a non-ambiguous SNP, if the alleles in the cohort match those in 1000 Genomes, then the strands must also match; for ambiguous SNPs, this is not necessarily the case. Illumina genotyping arrays typically have very few (<5%) ambiguous SNPs, so we decided to exclude these to be certain of consistency. Affymetrix arrays typically have a higher proportion (10-20%), so we preferred to retain. In this case, we aligned SNPs to the forward strand (the alignment of 1000 Genomes) and checked the allele frequencies were highly concordant and that there were no obvious inversions.

We performed the above checks manually in `R`. However, an alternative is to use the software provided on the LiftOver website. Note these checks are required even if subsequently using an imputation server.

Next we phased data using SHAPEIT, then imputed using IMPUTE2. For this we divided the genome into 549 regions (518 autosomal) of approximate size 5Mb (regions provided in the file allreg.txt). Note that we include instructions for imputing Chromosome X, however, for our analyses we considered only autosomal SNPs. Suppose the genotype data, with annotations updated, are stored in `updated.bed`, `updated.bim` and `updated.fam`. We provide approximate times and memory requirements assuming the cohort contains 5000 individuals.

```
#Divide data by chromosome
for i in {1..23}; do
plink --bfile updated --chr $i --make-bed --out split$i;
done

#Phase each chromosome using SHAPEIT
#genetic_map_chr${i}_combined_b37.txt contains genetic mappings for Chromosome i
#The option '--thread' controls the number of cores
#'--effective-size 11418' is recommended for Europeans (for other populations see SHAPEIT website)
for i in {1..23}; do
shapeit -B split$i -M genetic_map_chr${i}_combined_b37.txt --thread 8 \
--effective-size 11418 -O phase$i
done
#Phased data for Chromosome i will be stored in phase$i.haps and phase$i.sample
#Using 8 threads, the longest chromosomes would take 10-20 hours, and require <1Gb per thread

#Impute in regions of approximately 5Mb using IMPUTE2
#allreg.txt contains the chromosome, start and end bp for 549 regions
#Regions 1-518 are autosomes, 520-548 are Chromosome 23, 519 & 549 are pseudo-autosomal
#1000GP_Phase3_chr$i.hap.gz contains phased genotypes for 1000 Genomes individuals
#1000GP_Phase3_chr$i.legend.gz contains SNP annotations for these
#It is possible to exclude SNPs based on MAF using the option `-filt_rules_l' - excluding SNPs
    showing no variation across European 1000 Genomes individuals substantially reduced memory
    requirements
#See the IMPUTE2 website for explanation of other options

#For the 518 autosomal regions
```

```
for j in {1..518}; do
chr=`awk -v j=$j '(NR==j){print $1}' allreg.txt`
start=`awk -v j=$j '(NR==j){print $2}' allreg.txt`
end=`awk -v j=$j '(NR==j){print $3}' allreg.txt`

impute2 -m genetic_map_chr${chr}_combined_b37.txt \
-h 1000GP_Phase3_chr$i.hap.gz -l 1000GP_Phase3_chr$i.legend.gz \
-use_prephased_g -known_haps_g chr${chr}.haps -filt_rules_l 'EUR==0' \
-int $start $end -Ne 11418 -allow_large_reg -o_gz -o chunk$j
done

#For the Chromosome 23 regions, it is necessary to add -chrX (and edit the genetic map)
for j in {520..548}; do
start=`awk -v j=$j '(NR==j){print $2}' allreg.txt`
end=`awk -v j=$j '(NR==j){print $3}' allreg.txt`

impute2 -m genetic_map_chrX_nonPAR_combined_b37.txt
-h 1000GP_Phase3_chrX_NONPAR.hap.gz -l 1000GP_Phase3_chrX_NONPAR.legend.gz \
-use_prephased_g -known_haps_g chr23.haps -filt_rules_l 'EUR==0' \
-int $start $end -Ne 11418 -allow_large_reg -o_gz -o chunk$j -chrX
done

#For the pseudo-autosomal regions, it is necessary to add -chrX and -Xpar
#Note that imputation in these regions tends to be difficult as they tend to contain very few SNPs
j=519
start=`awk -v j=$j '(NR==j){print $2}' allreg.txt`
end=`awk -v j=$j '(NR==j){print $3}' allreg.txt`
impute2 -m genetic_map_chrX_PAR1_combined_b37.txt \
-h 1000GP_Phase3_chrX_PAR1.hap.gz -l 1000GP_Phase3_chrX_PAR1.legend.gz \
-use_prephased_g -known_haps_g chr23.haps -filt_rules_l 'EUR==0' \
-int $start $end -Ne 11418 -allow_large_reg -o_gz -o chunk$j -chrX -Xpar

j=549
start=`awk -v j=$j '(NR==j){print $2}' allreg.txt`
end=`awk -v j=$j '(NR==j){print $3}' allreg.txt`
impute2 -m genetic_map_chrX_PAR2_combined_b37.txt \
-h 1000GP_Phase3_chrX_PAR2.hap.gz -l 1000GP_Phase3_chrX_PAR2.legend.gz \
-use_prephased_g -known_haps_g chr23.haps -filt_rules_l 'EUR==0' \
-int $start $end -Ne 11418 -allow_large_reg -o_gz -o chunk$j -chrX -Xpar

#Imputed data for Region j will be stored in chunk$j.gz and chunk${j}_info
#Regions typically complete in 2-6 hours and require approximately 10Gb memory
```

Step 2 - Quality Control Having imputed each cohort separately, we next merged imputed cohort data to form datasets. We stored data in "Speed Format," a format which accommodates non-integer values and thus allows us to analyze expected allele counts. An alternative is to convert data to hard genotypes and store in PLINK binary format, which will have negligible impact if using only high-quality SNPs. This guide uses strict quality control, matching those we used for the main analyses. Specifically, we retained only bialellic autosomal SNPs which in all cohorts satisfied (expected) MAF>0.01 and $r_j > 0.99$ (if imputed) or `r2_type0` > 0.99 (if directly genotyped), where r_j is our information score (Supplementary Figure 20) and `r2_type0` is a metric computed by IMPUTE2. Suppose we are combining three cohorts, stored in the folders cohortA, cohortB and cohortC.

These instructions are designed to process the output from IMPUTE2, however, they should be easily modifiable for other imputation software (see `www.ldak.org/file-formats` for ways to incorporate different genotype formats). Always keep an eye on the screen output from LDAK to understand how datasets are being processed and for suggested options.

```
#Create badgeno.snps, listing SNPs which in any cohort were directly genotyped and had r2_type0<0.99
#Column 9 of the the info file indicates imputed (0) or genotyped (2); Column 12 provides r2_type0
rm badgeno.A badgeno.B badgeno.C
for j in {1..518}; do
awk < cohortA/chunk${j}_info '($9==2 && $12<.99){print $2}' >> badgeno.A;
awk < cohortB/chunk${j}_info '($9==2 && $12<.99){print $2}' >> badgeno.B;
awk < cohortC/chunk${j}_info '($9==2 && $12<.99){print $2}' >> badgeno.C;
done
cat badgeno.A badgeno.B badgeno.C | sort | uniq > badgeno.snps

#Merge cohorts by region, excluding SNPs in badgeno.snps, and filtering based on MAF and r_j
#'--common-preds YES' restricts t variants present in all (three) cohorts
#'--exclude-odd YES' excludes variants with alleles other than A, C, G, T (i.e., none SNPs)
#'--exclude-dups YES' excludes variants with matching positions (i.e., none diallelic SNPs)
#'--pass-all YES' ensures SNPs must pass QC in all cohorts
#Replace --make-speed with --make-bed to save data in Binary PLINK format

for j in {1..518}; do
echo -e "cohortA/chunk$j cohortA/phase1.sample" > list$j
echo -e "cohortB/chunk$j cohortB/phase1.sample" >> list$j
echo -e "cohortC/chunk$j cohortC/phase1.sample" >> list$j
chr=`awk -v j=$j '(NR==j){print $1}' allreg.txt`
ldak --make-speed merge$j --mgen list$j --oxford-single-chr $chr
--exclude badgeno.snps --exclude-odd YES --exclude-dups YES
--min-maf 0.01 --min-info .99 --pass-all YES --common-preds YES
done
#Mergxed data for Region j will be stored in speed format with prefix merge${j}

#Combine across regions
#Speed format files can be combined using cat; for PLINK data files, use --merge-list in PLINK
cat merge{1..518}.speed > data.speed
cat merge{1..518}.bim > data.bim
cp merge1.fam data.fam
head -n 1 merge1.stats > data.stats
for j in {1..518}; do
tail -q -n +2 merge$j.stats >> data.stats
done
awk < data.stats '(NR>1){print $1, $6}' > data.infos
#The dataset is now saved in data.speed, data.bim and data.stats
```

```
#data.infos contains info scores (although these are not used for our main analysis)

#If hard genotypes are required, allele counts can be rounded to the nearest integer
ldak --make-bed hard --speed data --threshold 0.5
#Genotypes will be saved to hard.bed, hard.bim and hard.fam
```

To reliably estimate h^2_{SNP}, individuals must be unrelated, which we achieve by filtering based on alleleic correlations computed from a pruned subset of SNPs. Next we make the covariate file `covar.covar`, which has 33 columns: two IDs and sex (Columns 1, 2, 5 of `data.fam`), 20 dataset principal axes (Columns 3 to 22 of `prune.vect`) and 10 population axes (Columns 5, 7, ..., 23) of `tg.profile`. To obtain population axes, we use 1000 Genomes data; suppose these are stored in tg.bed, tg.bim and tg.fam. We additionally perform single-SNP analysis; this used as a sanity check, and also to identify any highly-associated SNPs ($P < 10^{-20}$) which we subsequently include (after pruning) as additional fixed-effect covariates.

```
#Thin SNPs, then use these to compute allelic correlations
#We ensure no SNPs within 1Mb have correlation squared >0.2
#Could also exclude SNPs in high-LD regions (e.g., the MHC)
ldak --thin prune --speed data --window-kb 1000 --window-prune 0.2
ldak --calc-kins-direct prune --speed data --extract prune.in --ignore-weights YES --power -0.25
#The list of pruned SNPs are contained in prune.in; the kinship has stem prune

#Filter relatedness (for each cohort, then for all together), then obtain top 20 eigen-vectors
#By default, LDAK filters using the threshold c, where -c is smallest observed kinship
#To instead, filter based on a threshold of (say) 0.05, add --maxrel 0.05
ldak --filter keepA --grm prune --keep cohortA/phase1.sample
ldak --filter keepB --grm prune --keep cohortB/phase1.sample
ldak --filter keepC --grm prune --keep cohortC/phase1.sample
cat keepA.keep keepB.keep keepC.keep > keepABC.keep
ldak --filter prune --grm prune --keep keepABC.keep
ldak --pca prune --grm prune --axes 20 --keep prune.keep

#Compute 10 population axes from 1000 Genomes data
ldak --calc-kins-direct tg --bfile tg --extract prune.in --ignore-weights YES --power -0.25
ldak --pca tg --grm tg --axes 10
ldak --calc-pca-loads tg --bfile tg --grm tg --pcastem tg
ldak --calc-scores tg --speed data --scorefile tg.load --allow-flips YES --keep prune.keep

#Identify highly-associated SNPs from single-SNP analysis (we define as P<1e-20)
ldak --linear linear --speed data --covar covar.covar --keep prune.keep --pheno phen.pheno
awk '(NR>1&&$7<1e-20){print $2}' linear.assoc > top.snps #if using LDAK
#OR
plink --logistic hide-covar --out logistic --bfile hard --pheno phen.pheno \
--covar covar.covar --keep prune.keep
awk '(NR>1&&$9<1e-20){print $2}' logistic.assoc.logistic > top.snps #if using PLINK

#Prune these - we used a correlation-squared threshold of 0.5 and window of 10Mb
ldak --thin top --speed data --extract top.snps --keep prune.keep \
--window-prune 0.5 --window-kb 10000
```

Step 3 - Compute SNP weights, calculate kinships and estimate $h^2_{\mathbf{SNP}}$. At this point, the dataset is stored in `data.speed`, `data.bim` and `data.fam`, while `prune.keep` provides a list of unrelated (and population homogeneous) individuals and `covar.covar` contains covariates. If any highly-associated SNPs were identified, these are listed in `top.in`. Suppose phenotypes are stored in `phen.pheno`. Again, always keep an eye on the screen output from LDAK.

```
#Prepare to compute SNP weightings
ldak --cut-weights sections --speed data --keep prune.keep
#The details of each section will be stored in sections/section.details

#Compute weightings for each section - for this example, suppose there are 200 sections
#The exact number of sections is provided in sections/section.number
#When data come from multiple cohorts (here 3), using Subset Options guards against genotype errors
#For this, we create the files keep1, keep2 & keep3, containing individuals from Cohorts A, B & C
awk '(NR==FNR){arr[$1];next}($1 in arr){print $1, $2}' prune.keep cohortA/phase1.sample > keep1
awk '(NR==FNR){arr[$1];next}($1 in arr){print $1, $2}' prune.keep cohortB/phase1.sample > keep2
awk '(NR==FNR){arr[$1];next}($1 in arr){print $1, $2}' prune.keep cohortC/phase1.sample > keep3
for j in {1..200}; do
ldak --calc-weights sections --speed data --subset-number 3 --subset-prefix keep --section $j
done
#Typically sections complete in under 2 hours (most within 10 minutes)

#Join weightings across sections
ldak --join-weights sections --speed data
#Weights will be stored in sections/weightsALL
```

We will compute kinships for each chromosome separately; this will allow us to incorporate a test for inflation of h^2_{SNP} due to cryptic relatedness. It is normally not necessary here to include highly-associated SNPs as covariates, but if you do, make sure not to double count their contribution. If the trait is quantitative and individuals span multiple cohorts, we can also test for inflation due to genotyping errors.

```
#Get a list of SNPs on each chromosome
awk '{print $4 > "chr"$1""}' data.bim

#Compute kinships for each chromosome
ldak --cut-kins kinships --speed data --by-chr YES
for j in {1..22}; do
ldak --calc-kins kinships --speed data --partition $j --weights sections/weights.all --power -0.25
done

#Join to get genome-wide kinships
ldak --join-kins kinships
#Genome-wide kinships will be saved with prefix kinships/kinship.all

#Construct kinships for Chromosomes 1-3, 4-7, 8-11 & 12-22
rm listA; for j in {1..3}; do echo "kinships/kinships.$j" >> listA; done
rm listB; for j in {4..7}; do echo "kinships/kinships.$j" >> listB; done
rm listC; for j in {8..11}; do echo "kinships/kinships.$j" >> listC; done
rm listD; for j in {12..22}; do echo "kinships/kinships.$j" >> listD; done
for j in {A,B,C,D}; do
ldak --add-grm kins$j --mgrm list$j
```

```
done

#Test each quarter separately and combined
for j in {A,B,C,D}; do
ldak --reml quad$j --grm kins$j --pheno phen.pheno --covar covar.covar --keep prune.keep \
--top-preds top.in --speed data
done
echo -e "kinsA\nkinsB\nkinsC\nkinsD" > listABCD
ldak --reml quadALL --mgrm listABCD --pheno phen.pheno --covar covar.covar --keep prune.keep \
--top-preds top.in --speed data

#The results are stored in quadA.reml, quadB.reml, quadC.reml, quadD.reml and quadALL.reml
#The estimated inflation is (h2A + h2B + h2C + h2D - h2ALL)/3 - useful to view as a percentage
grep Her_K quad{A,B,C,D,ALL}.reml | awk '(NR<=4){sum+=$2}(NR>4){sum2+=$2}END{I=(sum-sum2)/3;print
    "Inflation:", I, "=", I/sum2*100,"%"}'

#Test for inflation due to genotyping errors
ldak --he inflation --grm kinships/kinships.all --pheno phen.pheno --covar covar.covar \
--keep prune.keep --subset-number 3 --subset-prefix keep --top-preds top.in --speed data
#The results are stored in inflation.he and inflation.he.compare
```

If satisfied that there is minimal inflation due to cryptic relatedness (and genotyping error if tested), we are ready to estimate SNP heritability

```
#Estimate SNP heritability
ldak --reml final --grm kinships/kinships.all --pheno phen.pheno --covar covar.covar \
--keep prune.keep --top-preds top.in --speed data
```

Step 4 - Advanced analyses: Here we provide sketch details for some of the other analyses performed in the main text, as well as some additional features of LDAK. If including lower-quality SNPs, then simply add `--infos` when computing kinships. The genotype scaling can be varied using the option `--power`. For example, to use the previous default scaling, use `--power -1`.

After performing REML, the `.share` file provides relative estimates. These are useful when interested in relative contributions (e.g., of different SNP classes). The `.reml` file contains the null and alternative (log) likelihoods and a likelihood ratio test (LRT) statistic (the null model corresponds to only covariates). To test significance, we typically computed the difference in LRT statistics between results from partitioned and non-partitioned model (e.g., when comparing the GCTA and LDAK Models, we performed REML using just the genome-wide kinship matrix, then using two kinships, one computed from low-LD SNPs, the other from high-LD SNPs.

To test the contribution of DNaseI hypersensitivity sites (DHS), we downloaded DHS annotations from `hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz`. From these we created `dhs.txt` which had four columns, providing a unique identifier, chromosome, start and end basepairs for each. Note that if two-way SNP partitioning results in a very uneven divide of SNPs, instead of computing both kinships from scratch, it can be much quicker to first compute kinships for the small tranche, then subtract these from the (previously-computed) genome-wide kinships using the option `--sub-grm`. We did this when testing DHS, as these contain less than 20% of the genome. A similar strategy can be used when considering more than two partitions.

```
#To calculate enrichment of DHS, first compute kinships from DHS SNPs, then its complement (i.e.,
    subtract from genome-wide kinships), then perform two-way REML
ldak --cut-genes dhs --speed data --genefile dhs.txt --ignore-weights YES
ldak --calc-kins-direct dhs --speed data --extract dhs/gene_preds.txt --weights
    sections/weights.all --power -0.25
echo -e "kinships/kinshipALL\ndhs" > listsub
ldak --sub-grm not_dhs --mgrm listsub
echo -e "dhs\nnot_dhs" > listdhs
ldak --reml dhs --mgrm listdhs --pheno phen.pheno --covar covar.covar --keep prune.keep \
--top-preds top.in --speed data

#If including rare variants, then partition based on MAF
#Exact boundaries are not too important, but consider extra boundaries (say at 0.00025) if using
    very rare SNPs
#Likely some of the rare variants will be low quality, so add '--infos'
awk < data.stats '(NR>1 && $3>0.1){print $1}' > maf1
awk < data.stats '(NR>1 && $3>0.01 && $3<=0.1){print $1}' > maf2
awk < data.stats '(NR>1 && $3>0.025 && $3<=0.01){print $1}' > maf3
awk < data.stats '(NR>1 && $3>0.001 && $3<=0.025){print $1}' > maf4
awk < data.stats '(NR>1 && $3<=0.001){print $1}' > maf5
ldak --cut-kins rare --partition-number 5 --partition-prefix maf
for j in {1..5}; do
ldak --calc-kins rare --speed data --partition $j --weights sections/weightsALL \
--power -0.25 --infos data.infos
done
```

Supplementary Note 2: List of analyses. This note summarizes our main analyses; red text reports the conclusions we drew from each. Scripts are provided in Supplementary Note 1. Prior to these analyses, we performed extensive quality control which included testing for inflation of h^2_{SNP} due to cryptic relatedness and genotyping errors (Supplementary Note 5).

(i) - Investigating the relationship between heritability and MAF: The parameter α, used when scaling genotypes, specifies a relationship between heritability and MAF (see Main Text Figure 2A). We compare different α based on model likelihood; higher likelihood indicates more accurate α. Across the 42 traits, and for a variety of SNP filterings, we deduce that $\alpha = -1$, the scaling most commonly used in human genetics and the previous default in SNP-based heritability analysis, provides sub-optimal model fit, whereas $-0.5 < \alpha < 0$ fit the data much better (Main Text Figure 2B, Supplementary Figures 1 & 2 and Supplementary Table 2). We recommend $\alpha = -0.25$ as the new default scaling.

(ii) - The impact of the assumed relationship between heritability and LD: The GCTA Model assumes that heritability contributions are independent of LD; the LDAK Model defines a relationship where SNPs in low-LD regions are expected to contribute more than those in high-LD regions. For each of the 42 traits, we compare estimates of h^2_{SNP} from GCTA with those from LDAK (Main Text Figure 3A, Supplementary Figures 7, 8 & 9 and Supplementary Table 3). We additionally consider GCTA-MS (GCTA partitioned by MAF), GCTA-LDMS (partitioned by MAF and LD) and LDSC, which can be viewed as an implementation of GCTA using summary statistics. Estimates of h^2_{SNP} based on the LDAK Model tend to be substantially higher than those based on the GCTA Model.

(iii) - Comparing the GCTA and LDAK Models through simulation Yang *et al.* reported results from a simulation study in which GCTA outperformed LDAK. However, the superiority of GCTA was to be expected, as when generating phenotypes, they sampled causal SNP effect sizes under the GCTA Model. We repeat their study, showing that if effect sizes are instead sampled under the LDAK Model, then LDAK is superior (Main Text Figure 3B and Supplementary Figures 10 & 11). It is difficult to fairly compare the GCTA and LDAK Models through simulation, so we instead test them using real data.

(iv) - Testing the GCTA and LDAK Models on real data: Across the 42 traits we find that assuming the LDAK Model consistently results in better model fit than assuming the GCTA Model (Supplementary Table 4). This is the case whether we use $\alpha = -1$ or $\alpha - 0.25$, demonstrating that the dependence of heritability on LD is distinct from its dependence on MAF. To visually demonstrate the improvement in model fit, we partition SNP into low-LD and high-LD, showing that the estimated contribution to h^2_{SNP} of each tranche is much closer to that predicted by the LDAK Model (Main Text Figure 4 and Supplementary Figure 12). Across a wide-range of real datasets, the LDAK Model is more realistic than the GCTA Model.

(v) - Testing the relationship between heritability and genotype certainty: Using the UCLEB traits, we demonstrate that when lower-quality SNPs are included, model specification is improved by taking into account genotype certainty (Supplementary Table 5). If lower-quality SNPs are included, the effect-size prior assumptions should be adjusted to incorporate genotype certainty.

(vi) - Final estimates of h^2_{SNP}: We analyze each trait using LDAK with $\alpha = -0.25$ (Main Text Table 1 and Supplementary Table 1). For many traits, we find that common SNPs tag substantially more heritability than previously appreciated.

(vii) - Enrichment of DNaseI hypersensitivity sites (DHS): By partitioning each dataset into SNPs inside and outside of DHS, we estimate that across the 42 traits, DHS on average contribute 24% (SD 2) of h^2_{SNP}. This is higher than they are expected to contribute under the LDAK Model (18%), but considerably lower than the estimate of 79% from Gusev *et al.*[2] (Main Text Figure 5, Supplementary Figure 14 and Supplementary Table 8). DHS are enriched for heritability, but to a much lesser extent than previously reported.

(viii) - Relaxing quality control for the UCLEB data: We first reduce the information score threshold, then the MAF threshold. We find that across the 23 traits, including lower-quality (common) SNPs leads to significantly higher estimates of h^2_{SNP}, and to a large extent makes up for the relatively low coverage of the Metabochip (Main Text Figure 6 and Supplementary Figures 5 & 18). Likewise, we show that including rare SNPs results in a significant increase in h^2_{SNP}, and we estimate that common and rare variants contribute to h^2_{SNP} in an approximate 2:1 ratio (Main Text Figure 6 and Supplementary Figure 6). Our improved heritability model allows us to appreciate the causal variation tagged only by lower-quality and rare SNPs.

Supplementary Note 3: Prevalence of pulmonary tuberculosis. When analyzing diseases, it is preferable to report estimates of h^2_{SNP} on the liability scale, as these are invariant to ascertainment and so can be compared across studies.[3,4] However, to do this requires an estimate of (lifetime) prevalence. Where possible, we copied the prevalences used by previous heritability analyses. For pulmonary tuberculosis, we are unable to find any reported estimates of h^2_{SNP}, so instead we use statistics provided by the World Health Organization.[5]. Our estimate of its prevalence in Russia is 4%, which we obtain by multiplying the fraction of the population reported (in 2014) to have developed either clinically diagnosed or bacterially confirmed tuberculosis (78 190/143 M = 0.00055) by an average life span (70 years).[5] Based on this value, our estimate of h^2_{SNP} is 0.26 (SD 0.03). If the true prevalence was in fact half as high, our estimate would instead be 0.21 (SD 0.02); if twice as high, 0.32 (SD 0.03). For comparison, using data reported from a small twin study,[6] we estimate the total heritability to be 0.5 (SD 0.2), while the two loci identified so far through GWAS (near *ASAP1*[7] and within the *HLA*[8]) together explain less than 0.01.

Supplementary Note 4: Additional cohort details. Supplementary Table 12 describes the 40 cohorts from which the GWAS and UCLEB datasets were constructed. Details of Cohorts 1 to 19, which all come from WTCCC 1 or 2, are provided in the original publication of the Wellcome Trust Case Control Consortium,[9] on the consortium website (`www.wtccc.org.uk`) and on the European Genome-Phenome Archive (`www.ebi.ac.uk/ega`), from where we downloaded the data. Below we describe aspects of the remaining 21 cohorts.

Cohorts 20 & 21 (celiac). These represent the two UK case/control collections described in Dubois *et al.*,[10] and were provided by David van Heel (Queen Mary University of London). Note that a raw Cohort 21 includes WTCCC 2 control samples. Celiac individuals were diagnosed according to standard clinical, serological and histopathological criteria, including small intestinal biopsy (see van Heel *et al.*[11] for more details).

Cohorts 22 & 23 (multiple sclerosis). These are samples recruited by the International Multiple Sclerosis Genetics Consortium (`www.imsgenetics.org`). Samples were recruited from 15 countries, although to reduce genetic heterogeneity, we exclude the 652 Finnish samples. To make imputation feasible we divided into Sweden, UK or US (Cohort 22) and Australia, Belgium, Denmark, France, Germany, Italy, New Zealand, Northern Ireland, Norway, Poland or Spain (Cohort 23). Except for approximately 700 Swedish controls, all individuals are cases, classified according to clinical and para-clinical criteria which establish that focal areas consistent with inflammatory demyelination have occurred in more than one part of the brain and spinal cord and on more than one occasion, and for which there is no better explanation than the diagnosis of multiple sclerosis.[12–14]

Cohorts 24 & 25 (partial epilepsy). These are from the Imperial-Liverpool-Melbourne collaboration used by the International League Against Epilepsy (ILAE) meta-analysis consortium[15]. Patients were classified by clinicians as partial, generalized or unclassified, according to ILAE guidelines[16,17] (unclassified essentially means there was insufficient evidence to classify as either partial or generalized). We excluded generalized and unclassified patients.

Cohorts 26 & 27 (pulmonary tuberculosis). The dataset (post-quality-control) used by Curtis *et al.*[7] Cases were recruited from two Russian cities, St. Petersburg and Samara, were all HIV negative and had active pulmonary tuberculosis confirmed by culture of *M. tuberculosis* from sputum; controls were healthy adults from Russia. To make imputation feasible, we divided samples (at random) into two equal-size cohorts.

Cohort 28 (intraocular pressure). This study comprises individuals living within the Blue Mountains region (west of Sydney, Australia) invited at random to attend an eye examination; those who consented had intraocular pressure measured using a Goldmann applanation tonometer.[18] As in the original analysis, we included age (in addition to sex) as a covariate.

Cohorts 29, 30 & 31 (wide-range achievement test). These were recruited by the Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia (CHOP) in collaboration with the Brain Behavior Laboratory at the University of Pennsylvania. Individuals were aged 8-21 who had volunteered to participate in genomic studies of complex pediatric disorders. In total, the phenotype file reports approximately 900 clinical variables; these include questions on eating disorders, anxieties, head injuries, vision problems, obsessive compulsive issues and phobias, although most were sparsely recorded. Based on the number of measurements available and clarity of phenotype definition, we picked Column 905, "Wide-Range Assessment Test (WRAT) Total Standard Score". Of the 4 429 individuals

with measurements, we excluded 202 outliers (those with scores below 58 or above 143), leaving 4 227. As with intraocular pressure, we included age as a covariate in all analyses.

Cohort 32 (Irish controls). The schizophrenia cases in WTCCC 2 (Cohort 14) were recruited at Trinity College Dublin, and are poorly matched by the two WTCCC 2 UK population control cohorts. Therefore, we instead use Cohort 32 as controls. This cohort was also recruited at Trinity College Dublin, comprises healthy young adult volunteers of Irish ancestry aged between 18 and 28 years at time of collection,[19,20] and was previously used as Irish Controls by the International League Against Epilepsy Consortium.[15]

Cohorts 33 to 40 (UCLEB). These are eight studies participating in the UCL-LSHTM-Edinburgh-Bristol Consortium, described in Shah *et al.*[21] From the many hundreds of phenotypes reported in one or more cohorts, we reduce to 23 as follows. First, for each cohort, we removed phenotypes recorded for less than 80% of individuals. Then, having matched up phenotypes across cohorts, we retain only the 56 with values recorded for at least half of individuals (i.e., for >7 000 individuals). Finally, we exclude binary traits (e.g., smoking and alcohol) and derivatives of other traits (e.g., logarithms of lipid levels). For each of the 28 remaining traits, we divided individuals by cohort and regressed on age. We then quantile-normalized: suppose n' individuals within Cohort 32 have values recorded for height; the tallest individual (after regressing out age), is assigned the new phenotype $\phi^{-1}(1/(n'+1))$, the next tallest is assigned $\phi^{-1}(2/(n'+1))$, and so on, where ϕ^{-1} denotes the inverse cumulative density function of the standard Gaussian distribution. We performed this transformation for each cohort in turn, and separately for males and females in the cohorts that include both sexes (36 to 40). Finally, using these normalized values, we tested for inflation due to genotyping errors (Supplementary Figure 13), leading us to exclude five phenotypes, so that 23 remained.

Supplementary Note 5: Quality control and imputation. This note summarizes the steps we took to construct our 20 datasets, starting from the 40 cohorts listed in Supplementary Table 12. See Supplementary Note 1 for detailed scripts.

(i) - Pre-imputation sample quality control: We excluded samples with extreme missingness or heterogeneity (computed from a pruned set of high-quality, autosomal SNPs), and population outliers (inferred by comparison with HapMap[22] samples). See Supplementary Figure 23 for examples. We then used (high-quality, pruned) SNPs on Chromosome X to infer sex.

(ii) - Pre-imputation SNP quality control: We excluded SNPs with MAF<0.01, call-rate <0.95 or Hardy-Weinberg $P < 10^{-6}$. Next we ensured SNP names (rs numbers) were up-to-date and that genomic positions were Build 37 (hg 19), then compared annotations with those in 1000 Genomes (Phase 3). In general, we erred on the side of caution. Ideally, SNPs would match by name, genomic position and alleles (allowing for strand flips). For a SNP which matched by name and alleles (but not position), we retained and updated the position to 1000 Genomes; for a SNP which matched by position and alleles (but not name), we retained only if the position agreed with Genome Browser. Notably, this strategy means we excluded SNPs with names and positions not present in 1000 Genomes, on the basis that we could not guarantee that these were aligned, nor would the imputation process be able to assess their accuracy (a feature we relied on for deciding which directly genotyped SNPs are reliably called). For most cohorts, this step resulted in relatively few SNPs being excluded (<1%). However, for Metabochip it reduced the number of SNPs used for imputation from about 110 K to 60 K. For Illumina cohorts, we additionally excluded ambiguous SNPs (alleles A & T or C & G); for Affymetrix cohorts, for which often over 10% of SNPs are ambiguous, we kept these but verified that their allele frequencies were highly concordant with those in 1000 Genomes (which indicated strand alignment).

(iii) - Imputation: We first phased the data using SHAPEIT.[23] We specified the effective population size as 11 418, the recommended value for European individuals, but otherwise left all parameters at their default values. The phased data were then passed to IMPUTE2. For this we provided window breakpoints approximately 5 Mb apart (where possible we placed breakpoints in SNP deserts, such as the centromere, and we extended the window length in regions of low SNP density). We again specified the effective population size as 11 418, but otherwise used default settings.

(iv) - Merging imputed cohorts and post-imputation SNP quality control: Supplementary Table 11 indicates which cohorts were combined to construct each dataset. For the GWAS datasets, we retained only autosomal SNPs which in all cohorts had (expected) MAF>0.01 and r_j >0.99. By definition, directly genotyped SNPs have $r_j = 1$, and so for these we instead required `r2_type0`>0.99, where `r2_type0` is an information metric reported by IMPUTE2[24]; we found that this dramatically reduced the number of directly genotyped SNPs, on average from 840 K to 280 K. While compared to previous heritability analyses, our SNP quality control thresholds

might seem unnecessarily strict, we demonstrate in Supplementary Note 6 how easily estimates of h^2_{SNP} can be inflated due to genotyping errors. When merging the UCLEB cohorts, we excluded non-autosomal SNPs, then computed MAF and r_j across all samples (rather than per-cohort); we required MAF>0.01 and r_j >0.99, except for the analyses in which these thresholds are explicitly relaxed.

(v) - Post-imputation sample quality control: For each dataset, we pruned SNPs so that no pair had squared correlation > 0.2, then computed unweighted allelic correlations. We used these for filtering out relatedness; first, we considered each cohort separately, then all (remaining) samples together. For each filtering, we removed individuals so that no pair remained with allelic correlation $> c$, where $-c$ is the smallest correlation observed across the individuals being considered. The rationale is that the observed negative correlations are from pairs of distantly-related individuals, and so positive correlations of greater magnitude indicate closer relatedness.[25] The top 20 eigen-vectors computed from the remaining samples are included as covariates in subsequent regressions. Eigen-vectors represent the most prominent axes of variation in the data. While these should implicitly capture population structure,[26] to be explicit, we additionally include as covariates projections onto ten population axes computed via principal component analysis of the 1000 Genomes data.

Supplementary Note 6: Control-control GWAS. We construct a dataset using the Illumina and Affymetrix versions of the WTCCC 2 controls (Cohorts 16-19 in Supplementary Table 12). We restrict ourselves to 4 572 unrelated individuals for whom we have both Illumina and Affymetrix genotypes (we confirm that matching IDs correspond to the same individual by checking the corresponding samples have genetic correlation >0.9). We consider only genotyped SNPs, excluding those which in any of the four cohorts have (expected) MAF <0.01, Hardy-Weinberg $P < 10^{-6}$ or call-rate <0.95 (these thresholds are typical of those used in GWAS[27]). We also exclude ambiguous SNPs (alleles A & T or C & G) to ensure no strand mismatches when merging. At this point, 259 867 SNPs remain. We construct a phenotype file where 1958 Birth Cohorts samples (Cohorts 16 & 17) are recorded as cases, while National Blood Samples (Cohorts 18 & 19) are controls.

We first estimate h^2_{SNP} using only Illumina genotypes (Cohorts 16 & 18). As we are comparing two sets of UK population controls, h^2_{SNP} should be close to zero. Instead our estimate of h^2_{SNP} is 0.22 (SD 0.08), significantly greater than zero (P=0.004). If we use Affymetrix genotypes (Cohorts 17 & 19), our estimate is 0.37 (SD 0.09; $P < 10^{-5}$). These estimates are unchanged if we repeat including 20 principal components as covariates. Using the test described in Supplementary Figure 19, we estimate that population inflation contributes minimally to these estimates of h^2_{SNP} (0.1% for Illumina, 0.9% for Affymetrix). These results suggest that the inflated estimates of h^2_{SNP} are mainly due to genotyping errors. To test this hypothesis, we compute a kinship matrix $\boldsymbol{K'}$ using both genotyping chips: if $\boldsymbol{X}^I$ and $\boldsymbol{X}^A$ contain Illumina and Affymetrix genotypes, respectively, then the kinship for Individuals i and k is

$$K'_{i,k} = \frac{1}{2}\sum_{j=1}^{m} w_j \frac{(X^A_{i,j} - 2p_j)(X^I_{k,j} - 2p_j)}{W} + \frac{1}{2}\sum_{j=1}^{m} w_j \frac{(X^I_{i,j} - 2p_j)(X^A_{k,j} - 2p_j)}{W},$$

where $2p_j$ is the allele fraction for SNP j among all 9 144 individuals) genotypes, and W is the sum of SNP weights. In practice, $\boldsymbol{K'}$ can readily be constructed from $\boldsymbol{K}$, the matrix of allelic correlation calculated across all 9 144 individuals: if individuals are ordered by cohort, then

$$\boldsymbol{K} = \begin{bmatrix} K_{16,16} & K_{16,17} & K_{16,18} & K_{16,19} \\ K_{17,16} & K_{17,17} & K_{17,18} & K_{17,19} \\ K_{18,16} & K_{18,17} & K_{18,18} & K_{18,19} \\ K_{19,16} & K_{19,17} & K_{19,18} & K_{19,19} \end{bmatrix} \quad \text{and} \quad \boldsymbol{K'} = \frac{1}{2}\begin{bmatrix} K_{18,16} & K_{18,17} \\ K_{19,16} & K_{19,17} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} K_{16,18} & K_{16,19} \\ K_{17,18} & K_{17,19} \end{bmatrix},$$

where the submatrix $K_{a,b}$ contains allelic correlations between individuals in Cohort a and those in Cohort b. Genotyping errors are problematic when they correlate with phenotype, and this is a particular risk for case-control studies where cases and controls have been genotyped separately (as is the case for most of our GWAS datasets). For example, suppose in Cohort 16 (Illumina), there is a tendency for individuals with allele count 1 for a particular SNP to be wrongly called as having count 2. This will cause pairs of individuals within this cohort (phenotypically concordant) to have (slightly) higher genetic similarity than pairs where one individual is in Cohort 16 and the other in Cohort 18 (phenotypically discordant). Thus for the Illumina GWAS (Cohort 16 vs Cohort 18), this SNP will show (artificial) association with the phenotype and will contribute towards inflation of h^2_{SNP}. Using $\boldsymbol{K'}$ to estimate pairwise genetic similarities should reduce this inflation, because we do not expect the Illumina and Affymetrix arrays to have the same patterns of miss-calls (i.e., it is

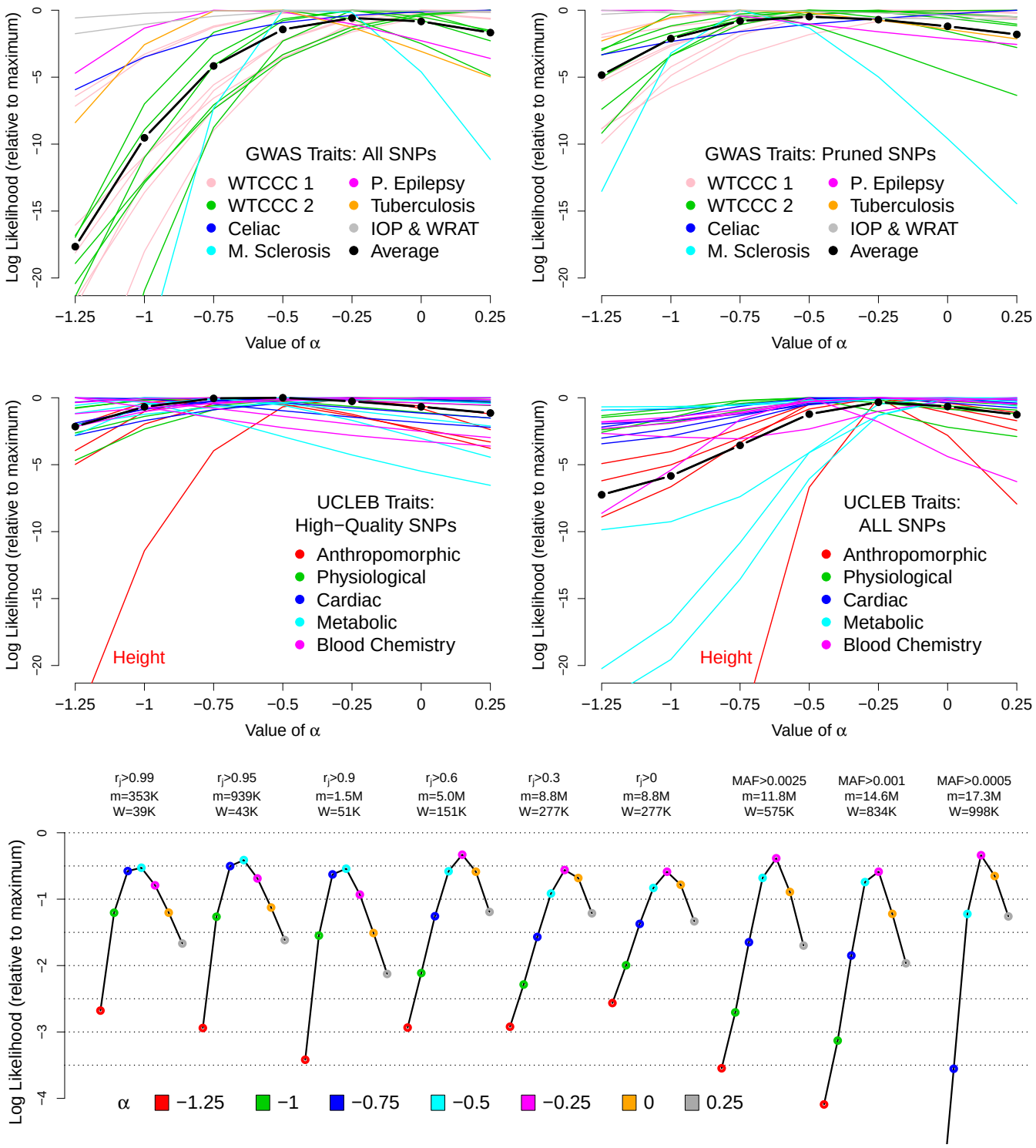
unlikely to also be the case in Cohort 17 that this SNP has allele counts 1 wrongly called as 2). When we instead use $\boldsymbol{K}'$, our estimate of h^2_{SNP} is 0.14 (SD 0.08), only borderline significant (P=0.04), and lower than the previous estimates (0.22 and 0.37) indicating that the inflation previously observed was largely due to genotyping errors. While there remains suggestive evidence for inflation, it should be realized that estimates based on $\boldsymbol{K}'$ are not immune to confounding. In particular, when calculating kinship matrices, missing genotypes are set to the mean, so that each SNP actually takes four values: 0, 1, 2 or $2p_j$. Therefore, if a group of individuals have similar patterns of missingness, perhaps because there are SNPs which are difficult to genotype regardless of genotyping platform, these individuals will appear more genetically similar, and potentially lead to inflation of h^2_{SNP}.

These analyses illustrate the importance of very strict quality control, and motivate our decision to use only the highest quality SNPs (those with MAF>0.01 and information score r_j >0.99). To verify that these thresholds suffice, we construct two additional GWAS datasets (Dataset 20 combines Cohorts 16 & 18, while Dataset 21 combines Cohorts 17 & 19), again making a fake phenotype file where individuals in Cohorts 16 and 17 are cases. This time we include imputed SNPs, and follow the quality control steps in Supplementary Note 5. We confirm that the subsequent estimates of h^2_{SNP}, 0.11 (SD 0.07) and 0.08 (SD 0.08), are not significantly greater than zero.

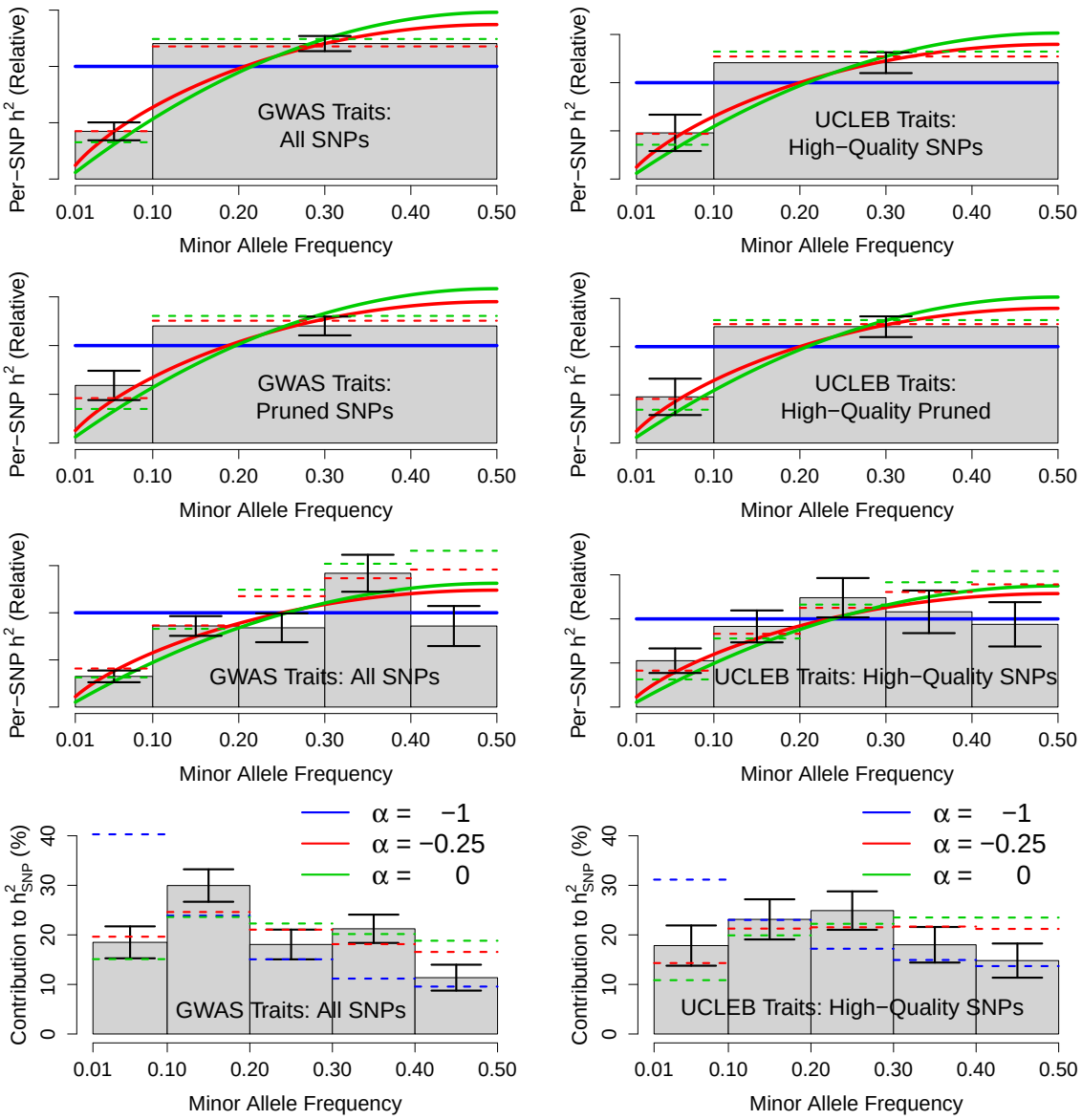
Supplementary Note 7: Simulation datasets. Simulation Dataset I, used in Supplementary Figures 18 & 22, combines the Illumina WTCCC 2 Controls (Cohorts 16 & 18 in Supplementary Table 12). We process this dataset as described in Supplementary Note 5, after which it contains 5 134 individuals, 4 710 536 SNPs, and has sum of SNP weights 136 407. For Simulation Dataset II, used in Supplementary Figure 10, we also combine Cohorts 16 & 18, but now copy as closely as possible the quality control used by Yang *et al.*[1]. First, to convert the imputed data to hard genotypes, we set allele counts based on the most likely state probability. Next, we retain only diallelic SNPs with minor allele count > 3 (equivalent to MAF>0.0002) and Hardy-Weinberg $P > 10^{-6}$. Finally, we filter individuals so that no pair remains with allelic correlation >0.05. After these steps, the dataset contains 4 869 individuals and 16 172 209 SNPs (with sum of weights 3 085 873). By comparison, the dataset used by Yang *et al.* contained 3 642 individuals and 17.6 M SNPs.

Supplementary Note 8: UCLEB Consortium Members. Tina Shah,[1] Jorgen Engmann,[1] Amand Floriaan Schmidt,[1] Chris Finan,[1] Caroline Dale,[2] Pimphen Charoen,[2] Jon White,[3] Stela McLachlan,[4] Andrew Wong,[5] Barbara Jefferis,[6] Tom R. Gaunt,[7] Fotios Drenos,[8] Jackie Cooper,[8] Antoinette Amuzu,[2] Reecha Sofat,[9] Ken Ong,[5,10] Rebecca Hardy,[5] Diana Kuh,[5] Debbie A. Lawlor,[7] George Davey Smith,[7] Nicholas Wareham,[10] Goya Wannamethee,[6] Philippa J. Talmud,[8] Steve E. Humphries,[8] Christine Power,[11] Elina Hypponen,[11] Claudia Langenberg,[2,12] Mika Kivimaki,[13] Meena Kumari,[13,14] Yoav Ben-Shlomo,[15] Richard Morris,[15] Peter Whincup,[16] Frank Dudbridge,[17] Jacqueline Price,[4] Juan P. Casas,[2] Aroon D. Hingorani.[1]

1. Institute of Cardiovascular Science, University College London, London, UK. **2.** Farr Institute of Health Informatics, University College London, London, UK. **3.** University College London Genetics Institute, Department of Genetics, Environment and Evolution, London, UK. **4.** Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK. **5.** MRC Unit for Lifelong Health and Ageing, London, UK. **6.** Department of Primary Care & Population Health, UCL Institute of Epidemiology & Health Care, University College London, London, UK. **7.** MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. **8.** Centre for Cardiovascular Genetics, Dept. of Medicine, British Heart Foundation Laboratories, Rayne Building, Royal Free and University College Medical School, London, UK. **9.** Centre for Clinical Pharmacology, University College London, London, UK. **10.** MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. **11.** MRC Centre of Epidemiology for Child Health, Department of Population Health Sciences, UCL Institute of Child Health, University College London, London, UK. **12.** MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. **13.** Department of Epidemiology & Public Health, UCL Institute of Epidemiology & Health Care, University College London, London, UK. **14.** Institute for Social and Economic Research, University of Essex, Colchester, UK. **15.** School of Social and Community Medicine, University of Bristol, Bristol, UK. **16.** Population Health Research Institute, St George's, University of London, London, UK. **17.** Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.



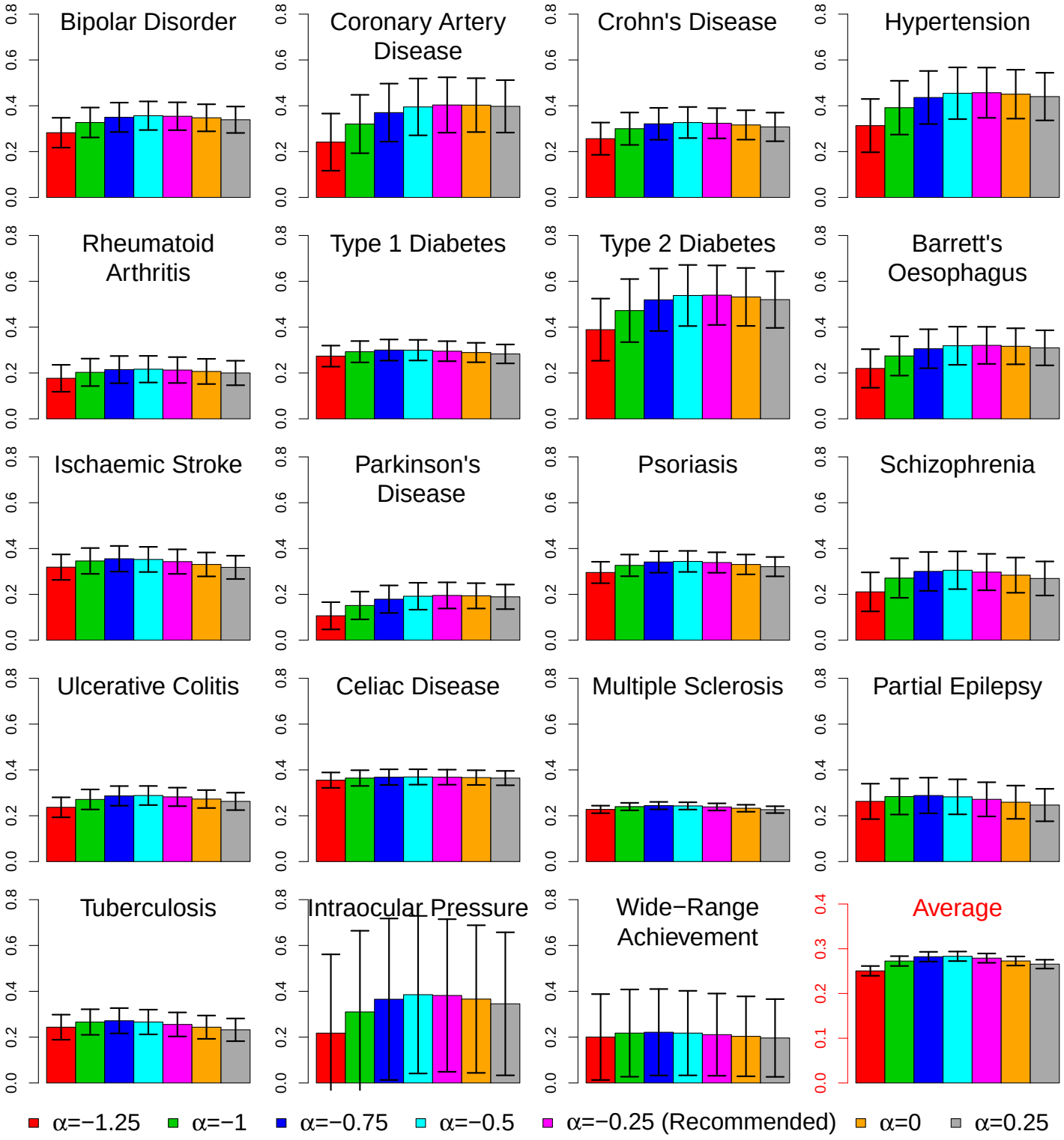
Supplementary Figure 1: Inferring the relationship between heritability and MAF for GWAS and UCLEB traits. The choice of α when scaling genotypes determines the assumed relationship between heritability and MAF. Lines show how log likelihood varies with α; values are reported relative to the highest observed for the corresponding trait. The top four plots report log likelihoods for the 19 GWAS traits (first using all SNPs, then only a pruned subset), then for the 23 UCLEB traits, (first using high quality common SNPs, then all SNPs); line colors indicate trait categories, while the black lines report averages. The bottom plot reports averages across the UCLEB traits for nine SNP filterings, determined by MAF and our information score r_j.



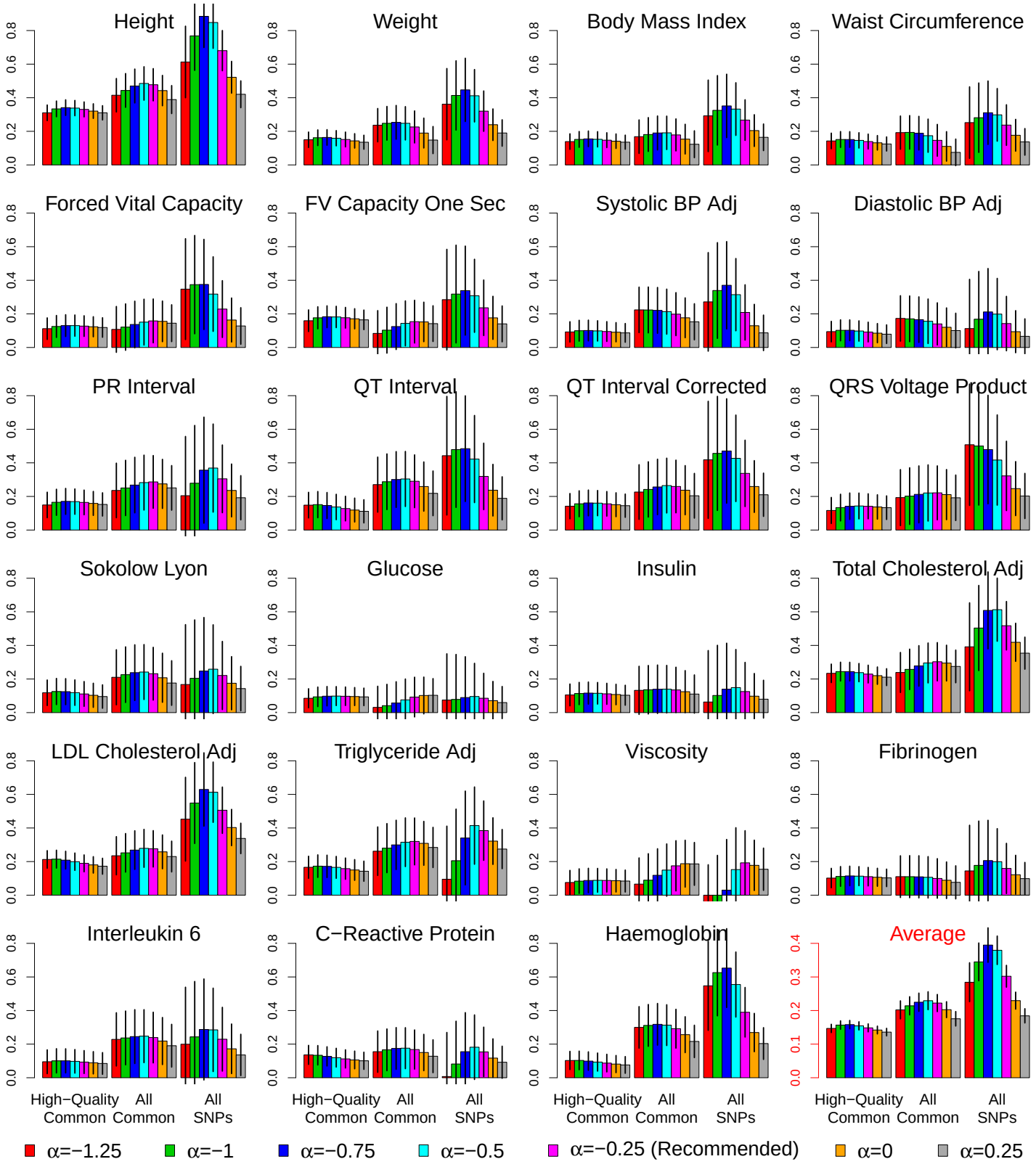
Supplementary Figure 2: Partitioning by MAF for GWAS and UCLEB traits. SNP partitioning allows us to estimate heritabilities for different MAF tranches; we then divide these by the sums of SNP weights for each tranche to obtain estimates of (weight-adjusted) per-SNP heritabilities. Grey bars report per-SNP heritabilities averaged across either the GWAS (left) or UCLEB traits (right); vertical lines provide 95% confidence intervals. The assumed relationship between heritability and MAF is determined by α: the solid lines show the relationship for α =-1, -0.25 and 0, while the dashed lines mark the expected per-SNP heritabilities for these three values.

The top pair of plots divides SNPs into MAF$\leq$0.1 or MAF$>$0.1. There is a clear drop in heritability: for example, across the GWAS traits, a SNP with MAF$\leq$0.1 on average contributes 35% as much heritability as a SNP with MAF$>$0.1. For the second pair of plots, we repeat this analysis using only a pruned subset of SNPs, showing that the decline in per-SNP heritability persists even when SNPs are in (approximate) linkage equilibrium. For the third pair, we use all SNPs but now five MAF tranches. We note that across both GWAS and UCLEB traits, the estimated per-SNP heritability for SNPs with MAF$>$0.4 is significantly below that predicted by $\alpha = -0.25$, suggesting that there is scope to improve the assumed relationship between heritability and MAF (although our sensitivity analyses in Supplementary Figures 3 & 4 indicate that doing so is unlikely to have much impact on estimates of h^2_{SNP}). The bottom pair of plots report the estimated contribution to h^2_{SNP} for each of five MAF tranches, averaged across traits. Although MAF$\leq$0.1 SNPs have lower-than-average per-SNP heritability, this tranche has highest sum of SNP weights, and as a result its total heritability contribution is comparable with those of the other tranches.

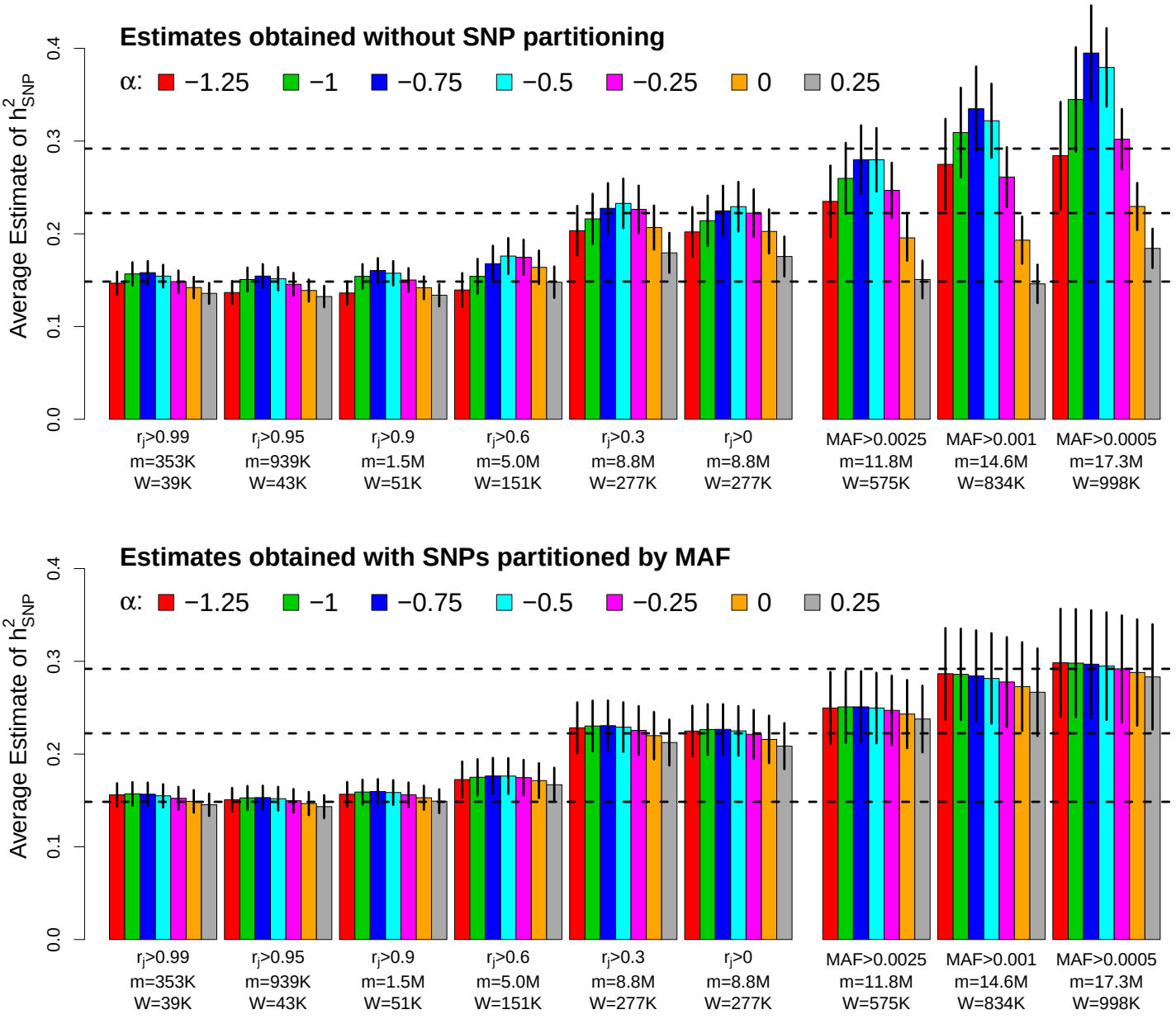
These estimates are obtained using the LDAK Model with $\alpha = -0.25$. Interestingly, if we instead use $\alpha = -1$, the estimates for MAF$\leq$0.1 SNPs tend to decrease, despite this value assigning more weight to low-MAF SNPs. For example, across the GWAS traits, when assuming $\alpha = -1$, MAF$\leq$0.1 SNPs are estimated to contribute 24% as much heritability as MAF$>$0.1 SNPs (rather than 40%), indicating that misspecification of α results in SNP contributions being assigned to the wrong MAF tranches.



Supplementary Figure 3: Estimates of h^2_{SNP} using different values of α for GWAS traits. Each plot corresponds to a different trait, while the final plot considers the (inverse variance weighted) average across all 19. Within each plot, the bars report estimates of h^2_{SNP} obtained using LDAK with seven values of α; vertical lines provide 95% confidence intervals. For binary traits, values have been converted to the liability scale. These estimates are based on common SNPs (MAF>0.01), and so in general, the impact of varying α is limited, particularly for $-0.75 \leq \alpha \leq 0$ (Bars 3-6 of each plot).

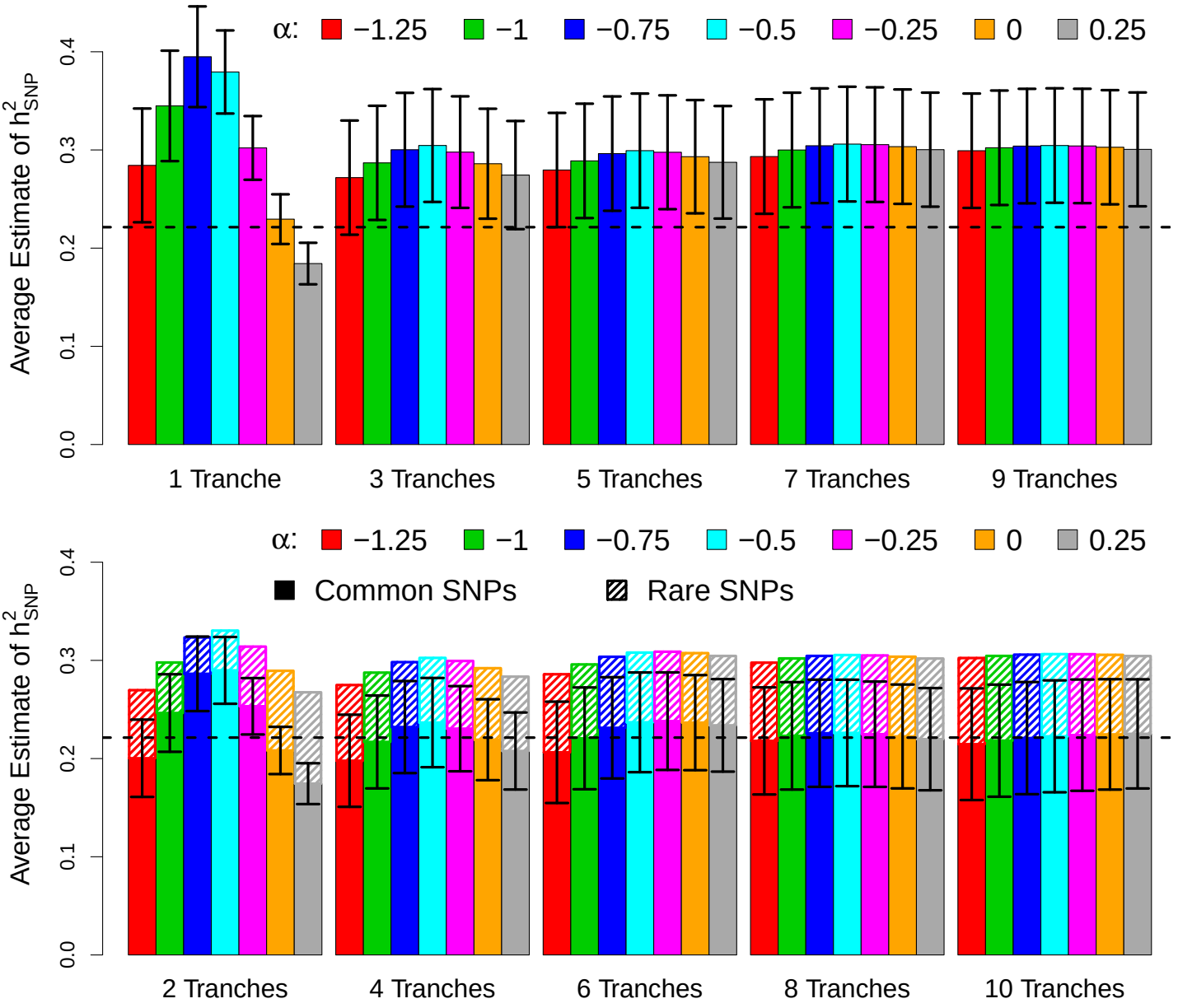


Supplementary Figure 4: Estimates of h^2_{SNP} using different values of α for UCLEB traits. Each plot corresponds to a different trait, while the final plot considers the (inverse variance weighted) average across all 23 traits. Within each plot, the three blocks indicate the SNP filtering (high-quality common SNPs, all common SNPs or all SNPs). Within each block, the bars report estimates of h^2_{SNP} obtained using LDAK with seven values of α; vertical lines provide 95% confidence intervals. We see that when using only common SNPs (Blocks 1 & 2), varying α has a limited impact on estimates of h^2_{SNP}, particularly for $-0.75 \leq \alpha \leq 0$ (Bars 3-6 and 10-13 of each plot). However, when rare SNPs are included, the impact can be large, so to obtain stable estimates, we partition rare SNPs based on MAF (Supplementary Figure 6).

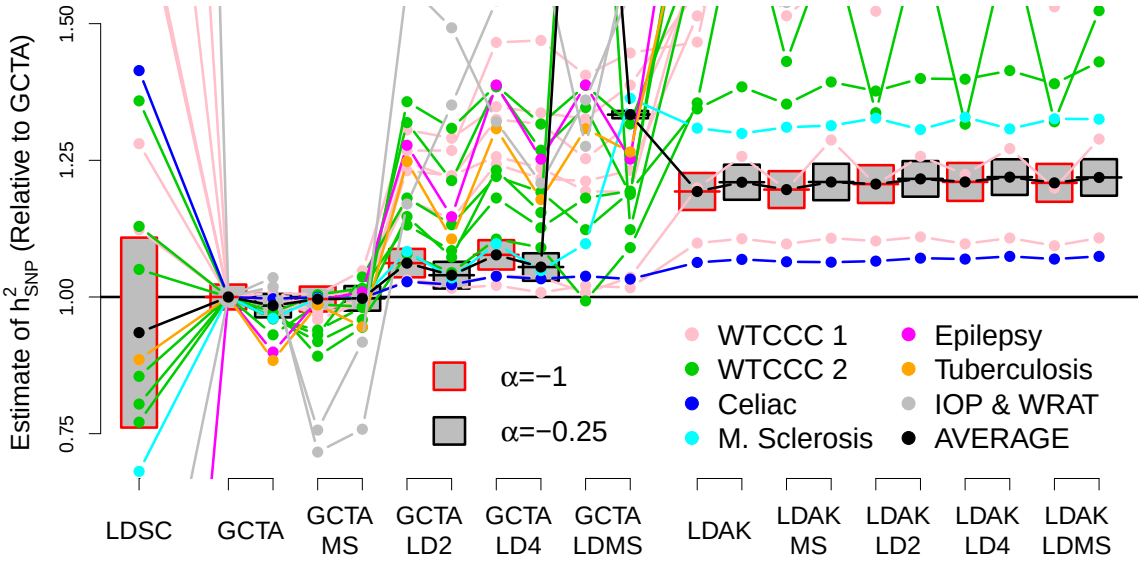


Supplementary Figure 5: Average Estimates of h^2_{SNP} using different values of α for UCLEB traits. Each block corresponds to a different SNP filtering, defined by MAF and r_j threshold: m = number of SNPs, $W = \sum_j r_j w_j$. Within each block, the bars report the (inverse variance weighted) average estimate of h^2_{SNP} across the 23 traits, obtained using LDAK with seven values of α. **Top:** Estimates are obtained without partitioning. We see that when restricted to common SNPs (Blocks 1-6), estimates of h^2_{SNP} are largely insensitive to the value of α, particularly for $-0.75 \leq \alpha \leq 0$, the range in which best-fitting α is most likely to lie (Supplementary Figure 1). However, when rare SNPs are included (Blocks 7, 8 & 9), varying α results in significantly different estimates of h^2_{SNP} **Bottom:** Now SNPs are partitioned by MAF: for Blocks 1-6, there are two tranches (MAF $\leq$0.1 and MAF $>$0.1); for Blocks 7, 8 & 9, there are three, four and five tranches, respectively, constructed by successively adding extra boundaries, at 0.01, 0.0025 and 0.001. We find that partitioning SNPs by MAF produces estimates robust to choice of α.

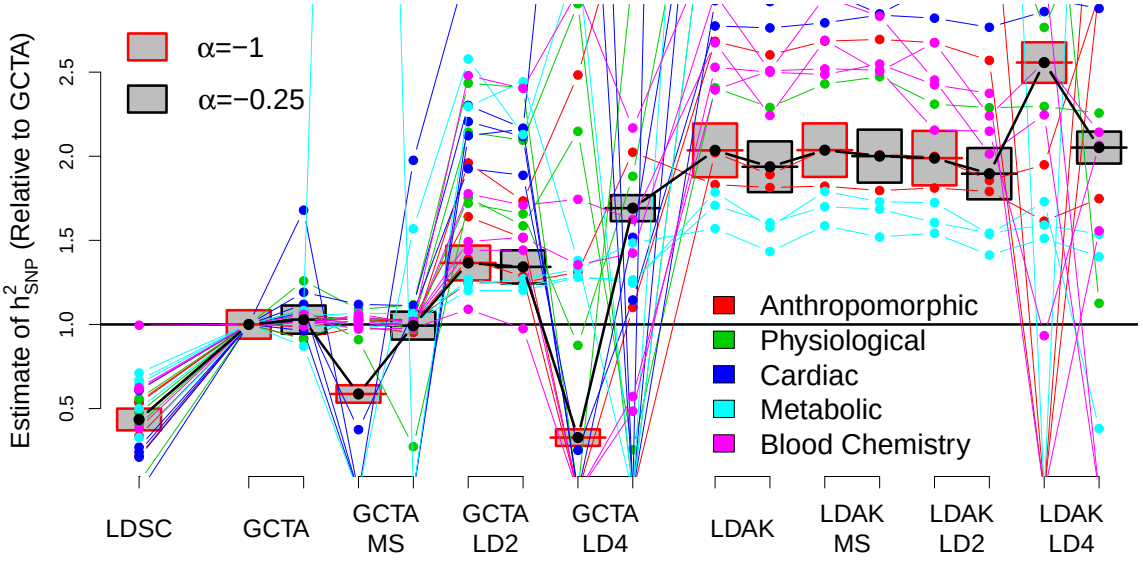
The three horizontal lines (at heights 0.15, 0.22 and 0.29) mark the average estimate of h^2_{SNP} when using only high-quality common SNPs (Block 1), all common SNPs (Block 6) and all SNPs (Block 9). We note that the heights of the bars under the partitioned model are close to the heights of the purple bars when using the non-partitioned model, indirectly lending support to $\alpha = -0.25$ being the most appropriate value. These results may suggest it is always worthwhile to estimate h^2_{SNP} with SNPs partitioned by MAF. However, partitioning typically results in less precise estimates (higher SDs), and there are computational advantages to using the non-partitioned model, particularly if multiple phenotypes are recorded for the same individuals.



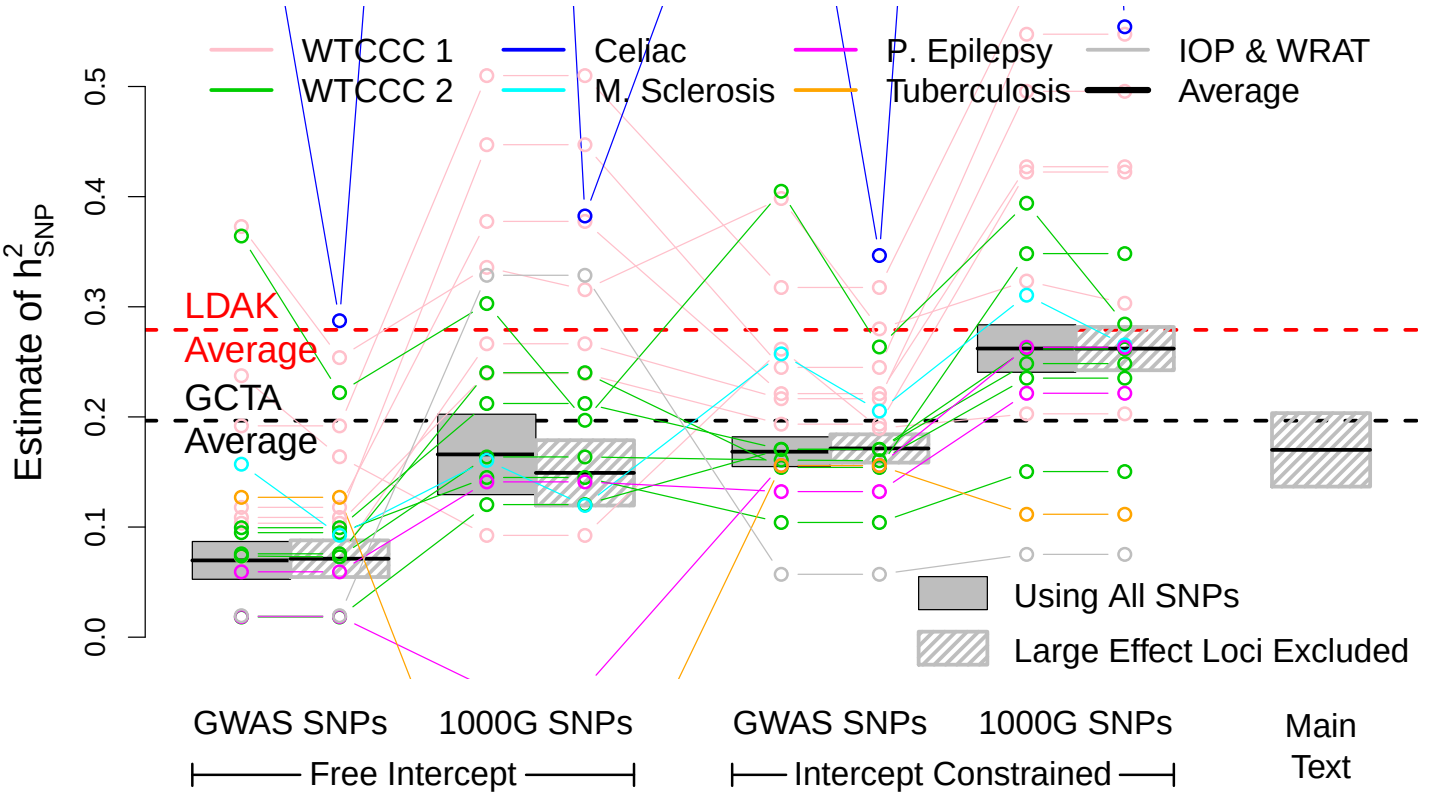
Supplementary Figure 6: Including rare SNPs when estimating h^2_{SNP} for UCLEB traits. For the UCLEB data, we sort the 17.3 M SNPs by MAF, then partition these evenly into L tranches, where L ranges from 1 to 10. For each L, we analyze using LDAK for seven values of α. Bars report the (inverse variance weighted) averages across the 23 traits; bar colors indicate the value of α used. **Top:** We examine how estimates of h^2_{SNP} depend on L (vertical lines provide 95% confidence intervals). When $L = 1$ (the non-partitioned model), varying α has a large impact on estimates of h^2_{SNP}. However, the impact reduces as L is increased. **Bottom:** For even L, the central MAF boundary is at 0.0111 (the median MAF across all SNPs). For $L = 2$, we can estimate rare heritability as the heritability assigned to Tranche 1, for $L = 4$, it is the heritability assigned to Tranches 1 or 2, for $L = 6$, it's the heritability assigned to Tranches 1, 2 or 3, and so on (note that for this example only, we are defining rare as MAF$\leq$0.011, whereas in general we use MAF$\leq$0.01). We now examine how the ability to divide h^2_{SNP} into common and rare depends on L (vertical lines provide 95% confidence intervals for the common SNP contribution). Again, we see that as L increases, sensitivity to choice of α decreases. For reference, the horizontal line in each plot reports the average estimate of h^2_{SNP} when using only common SNPs. For the main analysis, we used five tranches (two common, three rare), which is closest to $L = 6$ here.



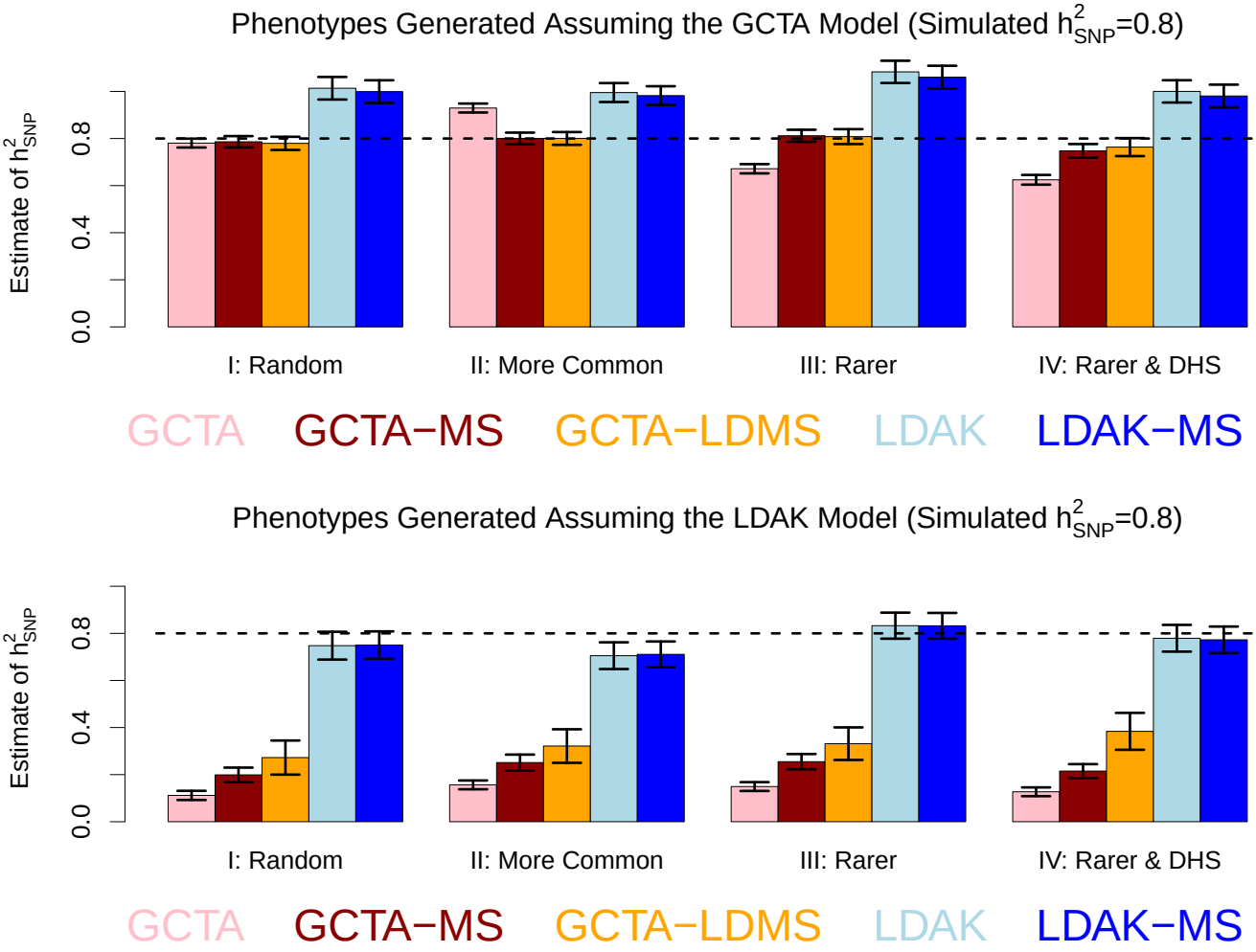
Supplementary Figure 7: Relative estimates of h^2_{SNP} for GWAS traits. We report estimates of h^2_{SNP} from versions of LDAK and GCTA, using either $\alpha = -1$ (red-edged boxes) or $\alpha = -0.25$ (black-edged). All estimates are relative to those from GCTA with $\alpha = -1$. Lines report individual trait estimates (colored by category), and the (inverse variance weighted) averages; gray boxes provide 95% confidence intervals for these averages. As well as the non-partitioned versions of LDAK and GCTA, we consider partitioning by MAF (MS: 5 tranches), by LD (using either 2 or 4 tranches) and by both LD & MAF (LDMS: 20 tranches). We also report results from LDSC, which assumes the GCTA Model with $\alpha = -1$. Variance components are estimated using Average Information REML[28–30]. The REML algorithm often fails to converge with large numbers of tranches (say, >10), particularly when the corresponding kinship matrices are highly-correlated. For example, when running GCTA-LDMS, the REML solver in GCTA v.1.26 failed (with the error "information matrix is not invertible") for 7 of the 19 traits when $\alpha = -1$ and for 4 traits when $\alpha = -0.25$. When this happened, we then tried to solve using the LDAK v.5 REML solver, but if this also failed, we substituted in results from GCTA-LD4 (as our analyses indicate that GCTA-LD4 and GCTA-LDMS tend to give similar results).



Supplementary Figure 8: Relative estimates of h^2_{SNP} for UCLEB traits. Details as for Supplementary Figure 7, except that we omit GCTA-LDMS and LDAK-LDMS because it was not feasible to partition by both MAF and LD (e.g., GCTA-LDMS completed for less than a quarter of the traits, possibly due to the sparse genotyping used for the UCLEB data).

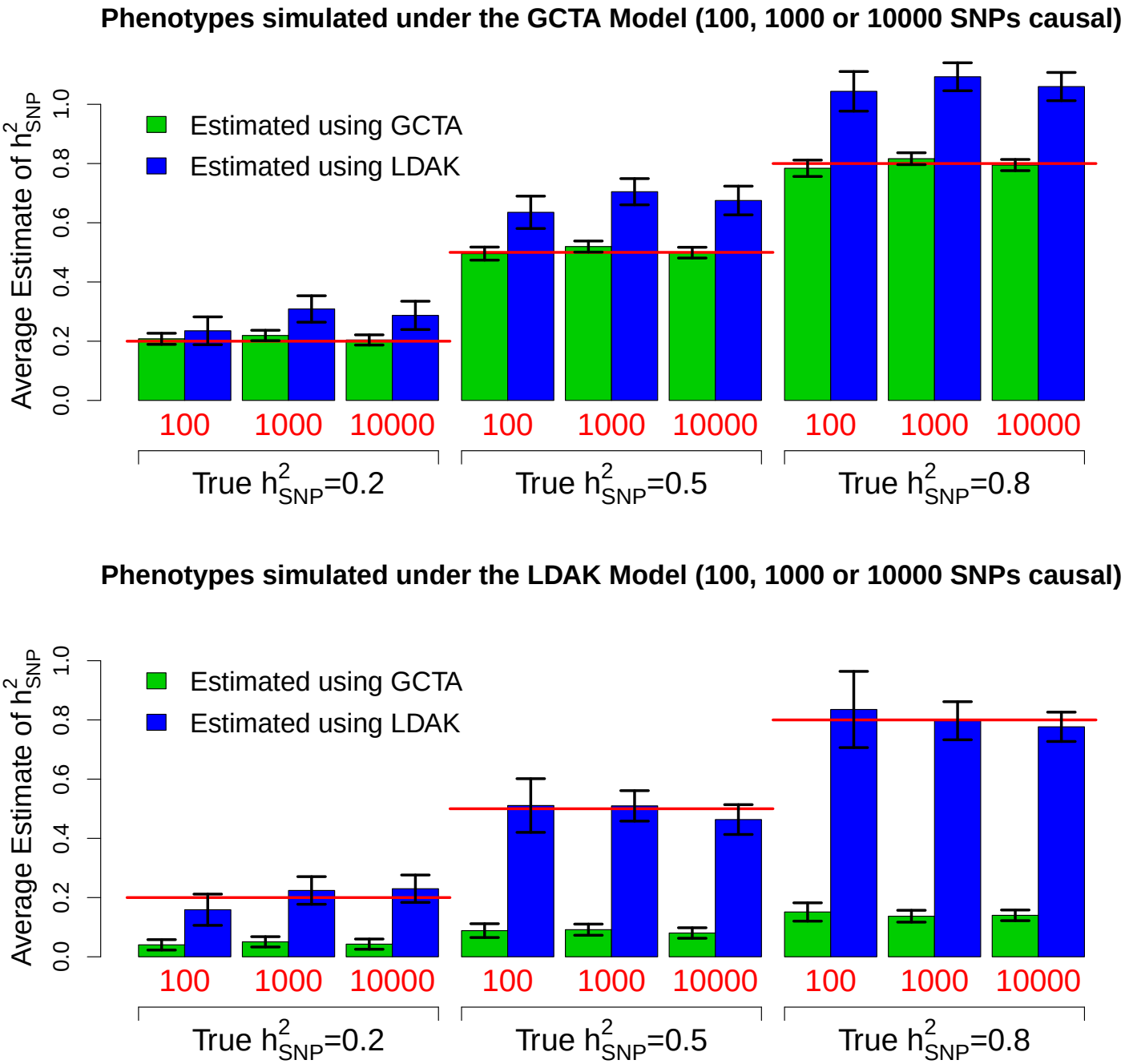


Supplementary Figure 9: Estimates of h^2_{SNP} from versions of LDSC for GWAS traits. Colored lines report individual estimates of h^2_{SNP} from LDSC using different settings, while gray boxes report (inverse variance weighted) averages. To run LD Score Regression, it is necessary to provide "reference" LD Scores. These can be computed either using the GWAS SNPs or from an independent (typically more dense) dataset. For our 23 GWAS traits, using LD Scores computed from 1000 Genomes data (provided on the LDSC website, `https://github.com/bulik/ldsc`) tends to result in higher estimates of h^2_{SNP} (Boxes 3, 4, 7 & 8) than when using only the GWAS SNPs (Boxes 1, 2, 5 & 6). A possible explanation is that in LDSC h^2_{SNP} represents the total phenotypic variance explained by all (common) SNPs in the reference dataset, so will tend to increase when a dataset with higher coverage is used. We find that performing the analysis with highly-associated SNPs ($P < 10^{-20}$ from marginal testing) excluded has limited impact (Boxes 2, 4, 6 & 8). The intercept term represents inflation due to confounders; when confident confounders are not an issue, LDSC suggests constraining the intercept to one. We find this leads to considerably larger, and more precise,[31] estimates of h^2_{SNP} (Boxes 5, 6, 7 & 8). In addition to providing reference LD Scores, it is also necessary to specify "weighting" LD Scores, which are used to reduce heteroscedasticity and account for local correlations in the regression. For these analyses, we use the same LD Scores for reference and weighting, the current recommendations on the LDSC website). However, in the main text, we used the method outlined in the original paper,[32] which advised using 1000 Genomes data for reference LD Scores and the GWAS SNPs for weighting LD Scores (Box 9). The red and black horizontal lines mark the average estimates of h^2_{SNP} from LDAK and GCTA. Note that in the main text, our conclusion that estimates of h^2_{SNP} from LDSC are not significantly different to those from GCTA was based on the examining the ratio (LDSC divided by GCTA); based on the absolute values, GCTA is higher. However, regardless of method of comparison, it remains that estimates from LDSC are typically no higher than those from GCTA, and on average are always lower than those from LDAK.

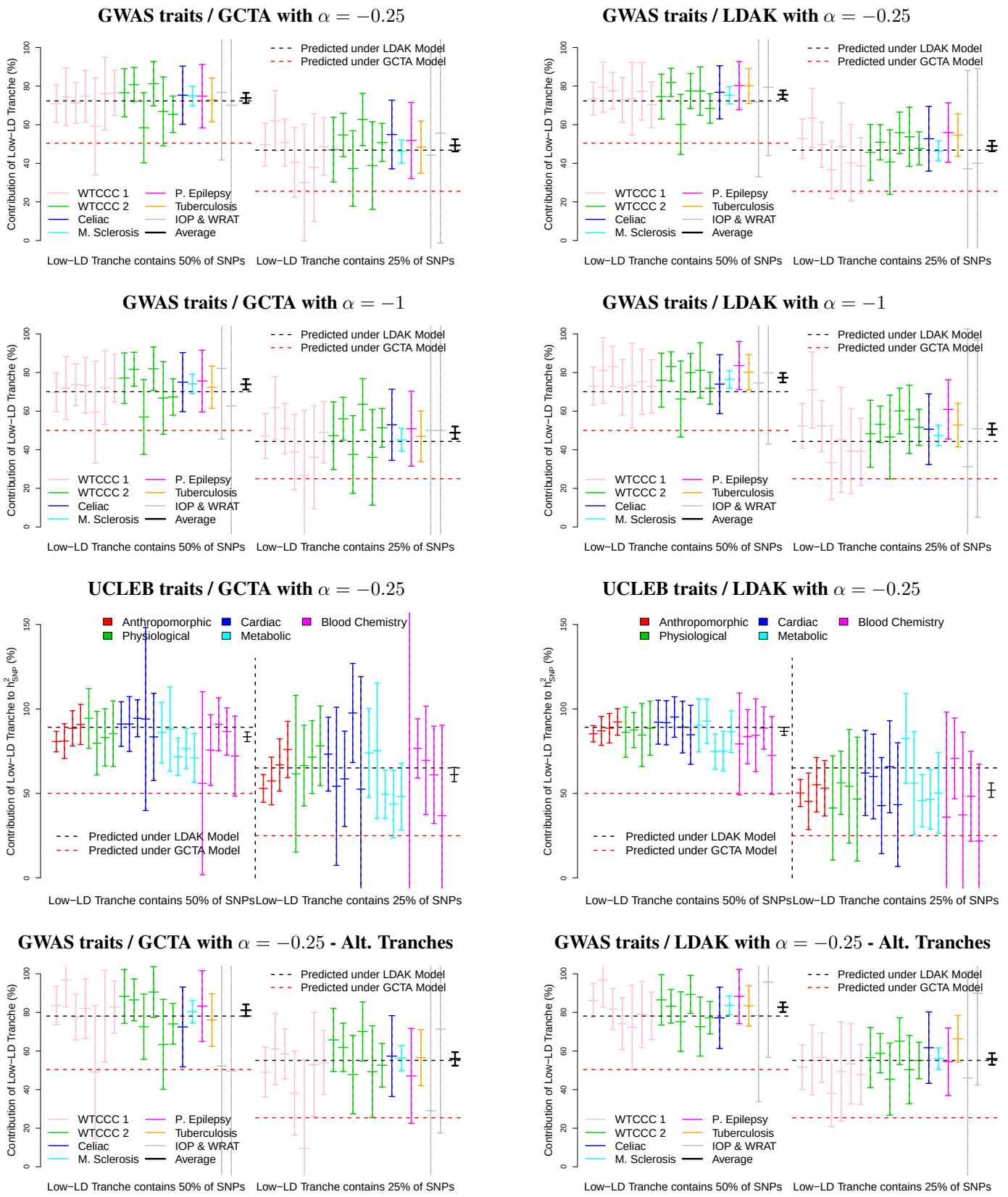


Supplementary Figure 10: Repeating the simulation study of Yang *et al.*[1] Yang *et al.*[1] compared versions of GCTA and LDAK via simulation. We investigate the dependence of their conclusion that GCTA outperforms LDAK on their choice of simulation model. For this we use Simulation Dataset II ($n = 5\,134$ individuals; $m = 16\,172\,209$ SNPs (MAF $>$0.0002); sum of SNP weights $\sum_j w_j = 3\,085\,873$; see Supplementary Note 7). Like Yang *et al.*, we generate phenotypic values using the model $Y = \sum_{j \in J} \beta_j X_j + e$, where the set J indicates which SNPs are causal, X_j denotes the genotypes for SNP j (centered and scaled to have variance one), β_j denotes its effect size and $e \sim \mathbb{N}(0, \sigma_e^2)$ represents Gaussian-distributed noise. For all simulations, σ_e^2 is chosen to ensure that $h^2_{\text{SNP}} = 0.8$. Again like Yang *et al.*, we consider four ways of selecting causal variants (i.e., deciding J). **Scenario I (random):** 1 000 causal variants randomly sampled from all SNPs. **Scenario II (more common):** 1 000 random and 500 additional common (MAF$>$0.01) causal variants. **Scenario III (rarer):** 1 000 random and 500 additional rare (MAF$<$0.01) causal variants. **Scenario IV (rarer and DHS):** 1 000 random and 500 additional rare causal variants sampled from the DNaseI hypersensitive sites (based on annotations provided at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz).

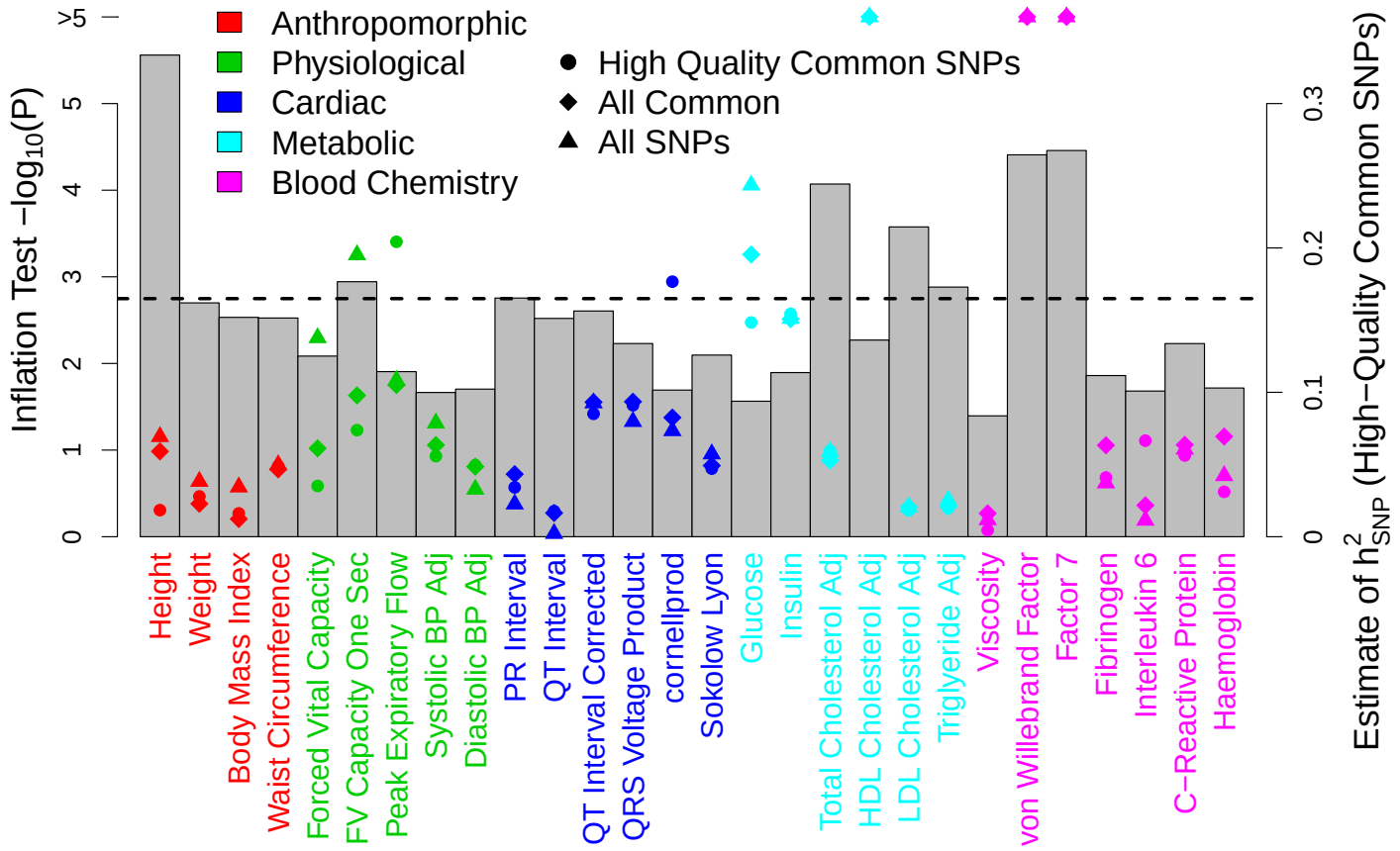
Top: We follow Yang *et al.* and sample β_j from $\mathbb{N}(0, 1)$, which is the assumed distribution of effect sizes under the GCTA model. The bars report mean estimates (across 200 repetitions) of h^2_{SNP} estimated using GCTA, GCTA-MS[33] (SNPs partitioned by MAF) and GCTA-LDMS[1] (SNPs partitioned by MAF and LD), as well as from LDAK and LDAK-MS; the vertical lines mark 95% confidence intervals for the means. Like Yang *et al.*, we observe that overall, GCTA provides more accurate estimates of h^2_{SNP} than LDAK. **Bottom:** Next we sample each effect size β_j from $\mathbb{N}(0, w_j)$, mimicking the LDAK Model. As expected, LDAK (greatly) outperforms GCTA.



Supplementary Figure 11: Additional simulations comparing the GCTA and LDAK Models We use genotypes from Simulation Dataset II ($n = 5\,134$ individuals; $m = 16\,172\,209$ SNPs (MAF >0.0002); sum of SNP weights $\sum_j w_j = 3\,085\,873$; see Supplementary Note 7). We generate phenotypes with either 100, 1 000 or 10 000 randomly-chosen causal variants and h^2_{SNP} equal to 0.2, 0.5 or 0.8. For the top plot, we sample effect sizes under the GCTA Model: $\beta_j \mathbb{N}(0, 1)$. For the bottom plot we sample effect sizes under the LDAK Model: $\beta_j \mathbb{N}(0, w_j)$. Green boxes report estimates of h^2_{SNP} using GCTA, blue boxes report estimates using LDAK, while red lines mark true h^2_{SNP}. We see that regardless of the number of SNPs contributing heritability, or the total heritability they contribute, it remains that when phenotypes are generated under the GCTA Model, estimates using GCTA are most accurate, while when phenotypes are generated under the LDAK Model, LDAK performs best.



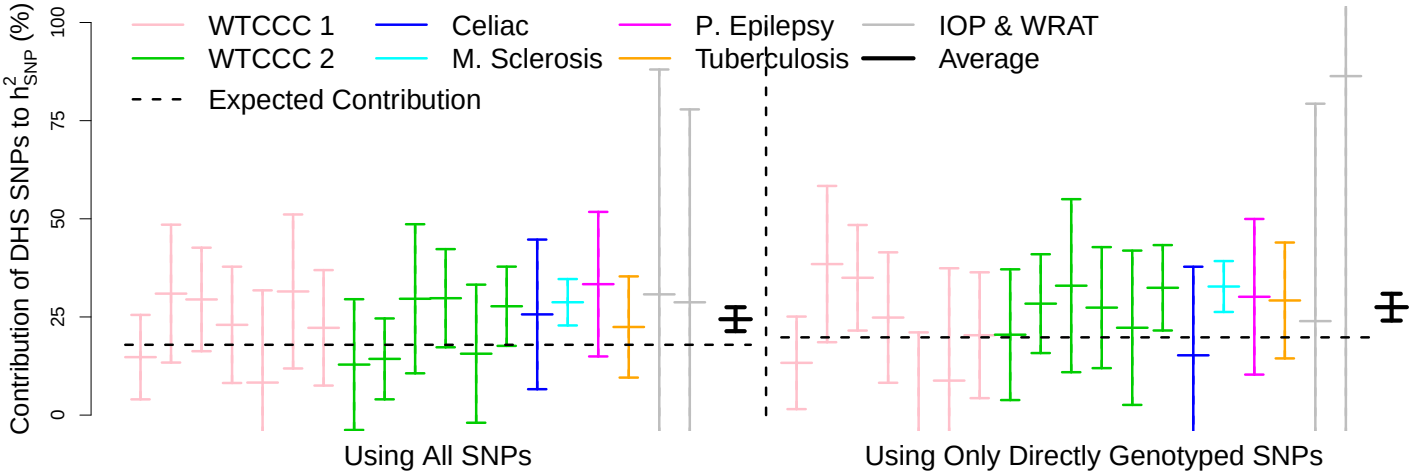
Supplementary Figure 12: Partitioning by LD for GWAS and UCLEB traits. SNPs are divided into low- and high-LD, so that either the low-LD tranche contains 50% or 25% of SNPs. The vertical lines report the point estimates and 95% confidence intervals for the estimated contribution of the low-LD tranche; lines are colored according to trait category, while the black line provides the (inverse variance weighted) average. We are interested in how close estimates are to the predicted contributions under the GCTA and LDAK Models (red and black horizontal lines). The plot titles indicate which traits the results correspond to, and which model was assumed when estimating heritabilities. In general we divided SNPs into low- and high-LD based on average LD Score of non-overlapping 100 kb segments;[1] but for the final two plots, we instead rank segments based on average LDAK SNP weight.



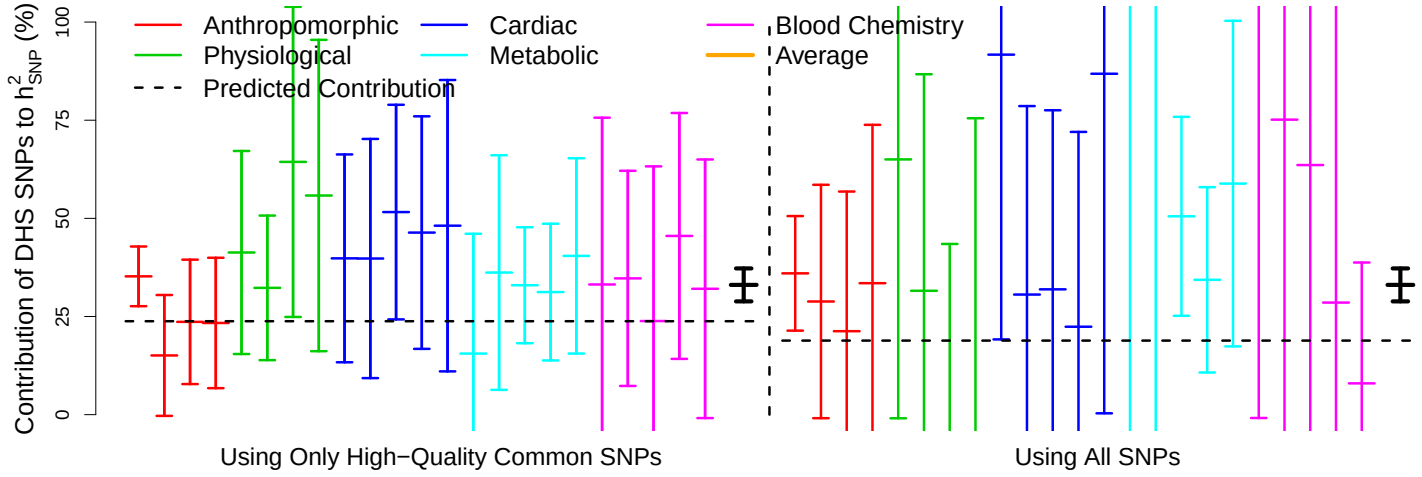
Supplementary Figure 13: Testing for inflation due to genotyping errors for UCLEB traits. Haseman-Elston regression represents an alternative way to estimate h^2_{SNP}. If we assume the standard mixed regression model $Y \sim \mathbb{N}(0, \boldsymbol{K}\sigma_g^2 + \boldsymbol{I}\sigma_e^2)$, where $\boldsymbol{K}$ and $\boldsymbol{I}$ denote the kinship matrix and an identity matrix, respectively, while σ_g^2 and σ_e^2 are the genetic and environmental variances, then $V_{i,k} = (Y_i - Y_k)^2$ has expected value $2(\sigma_g^2 + \sigma^2) - 2\sigma_g^2 K_{i,k}$. Haseman-Elston regression fits the linear model $V_{i,k} = \alpha + \beta K_{i,k}$, then $-\beta/\alpha$ represents an estimate of $h^2_{\text{SNP}} = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. Suppose the vectors $\boldsymbol{V}_S$ and $\boldsymbol{L}_S$ contain $(Y_i - Y_k)^2$ and $K_{i,k}$ for pairs of individuals in the same cohort, while $\boldsymbol{V}_D$ and $\boldsymbol{L}_D$ refer to pairs of individuals in different cohorts. The essence of our test is to fit the separate models $\boldsymbol{V}_S = \alpha_S + \boldsymbol{L}_S\beta_S$ and $\boldsymbol{V}_D = \alpha_D + \boldsymbol{L}_D\beta_D$, then observe how similar $-\beta_S/\alpha_S$ and $-\beta_D/\alpha_D$, the resulting estimates of h^2_{SNP}. Genotyping errors will cause pairs of individuals within the same cohort to appear more similar compared to pairs in different cohorts. If these errors are correlated with phenotype (i.e., if phenotypic similarity is higher within cohorts than across), then we would expect to observe $-\beta_S/\alpha_S \gg -\beta_D/\alpha_D$

We can formally test for inflation by performing a likelihood ratio test (LRT). A general strategy is to measure the improvement in fit of the alternative model, where α_S, α_D, β_S and β_D are free to vary, compared to the null model, where $\alpha_S = \alpha_D$ and $\beta_S = \beta_D$; a p-value can be obtained by comparing the LRT statistic with a $\chi^2(2)$ distribution. However, for the UCLEB data, phenotypes were quantile normalized prior to analysis (Supplementary Note 4), which ensures $\alpha_S = \alpha_D$; therefore, for our alternative model, we allow β_S and β_D to vary but set $\alpha_S = \alpha_D$, and we instead compare the LRT statistic with a $\chi^2(1)$ distribution. The figure reports $-\log_{10}(P)$ from this test for the original 28 UCLEB traits, computed using high-quality common SNPs (MAF>0.01; $r_j > 0.99$), all common SNPs (MAF>0.01) or all SNPs (MAF>0.0005). For subsequent analyses, we use only the 23 traits which when computed from high-quality common SNPs have P >0.05/28 (traits where the circle is below the horizontal line). This test is performed in LDAK by adding the options `--subset-prefix` and `--subset-number` when performing Haseman-Elston regression. For example, if data comprise two cohorts, with IDs in `list1` and `list2`, then one should add `--subset-prefix list --subset-number 2` when using the command `--he`.

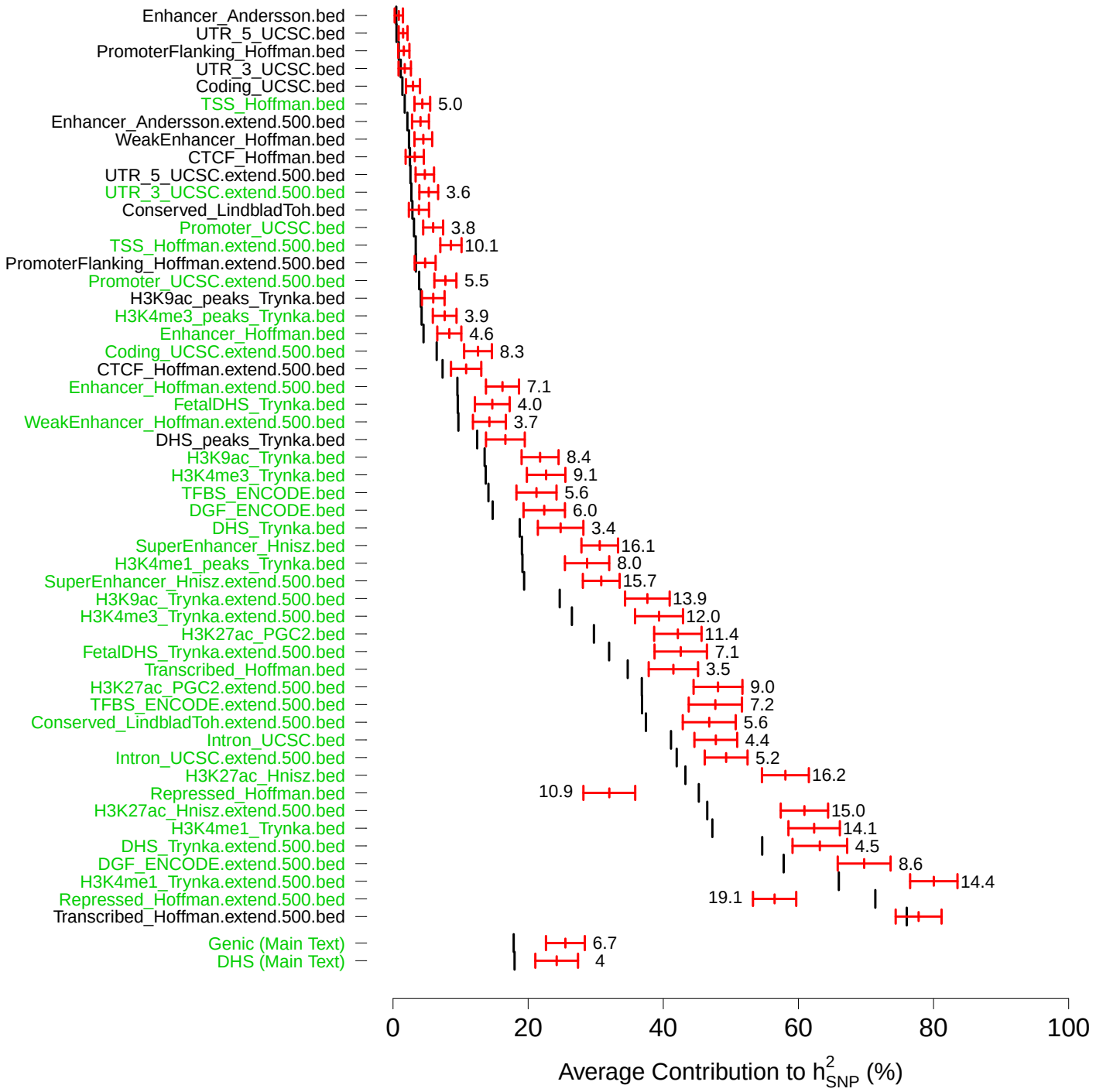
Note that when covariates are provided, we first regress their contribution from the phenotype, then test using the residuals. We anticipate that a more advanced test could be developed which accommodates covariates directly,[34] although for our application we expect this to have minimal impact. Additionally, our test is currently valid only for quantitative phenotypes, but we are working on a version suitable for case-control studies, where genotyping errors are typically more of a concern.



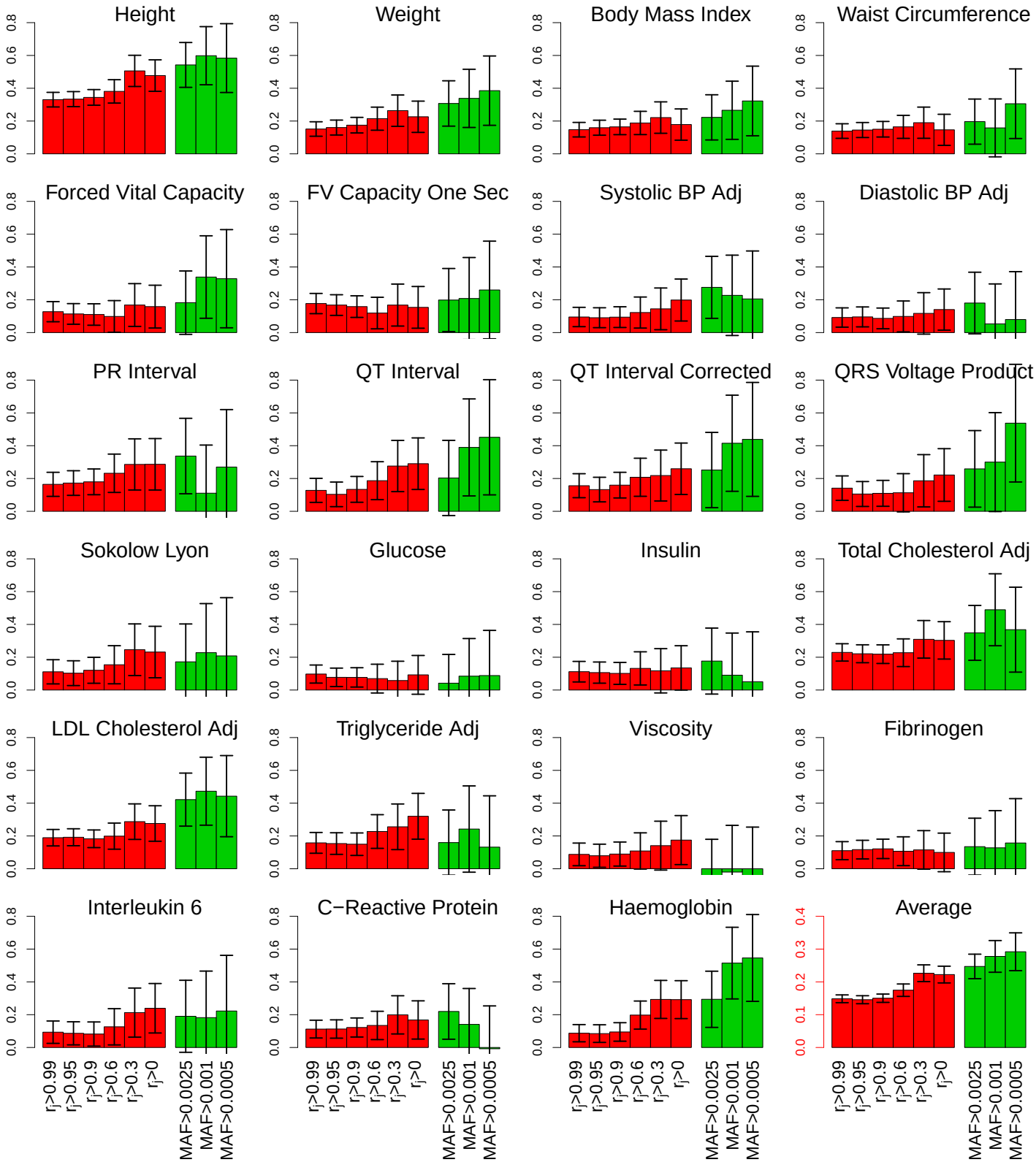
Supplementary Figure 14: Estimating enrichment of DHS for GWAS traits Annotations of DNaseI hypersensitive sites are from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz. Using the LDAK Model with $\alpha = -0.25$, vertical lines report estimates of the contribution to h^2_{SNP} of DHS for each of the 19 GWAS traits, then the (inverse variance weighted) average, first using all SNPs (on average 5.1 M), then only directly genotyped (235 K). Although the expected contribution (horizontal lines) varies as SNP density is reduced, we find that the enrichment (estimated divided by expected) remains constant (at 1.4-fold), contrasting the results of Gusev *et al.*[2] who found that enrichment dropped from 5.1-fold to 1.6-fold. See Supplementary Table 8 for numerical values.



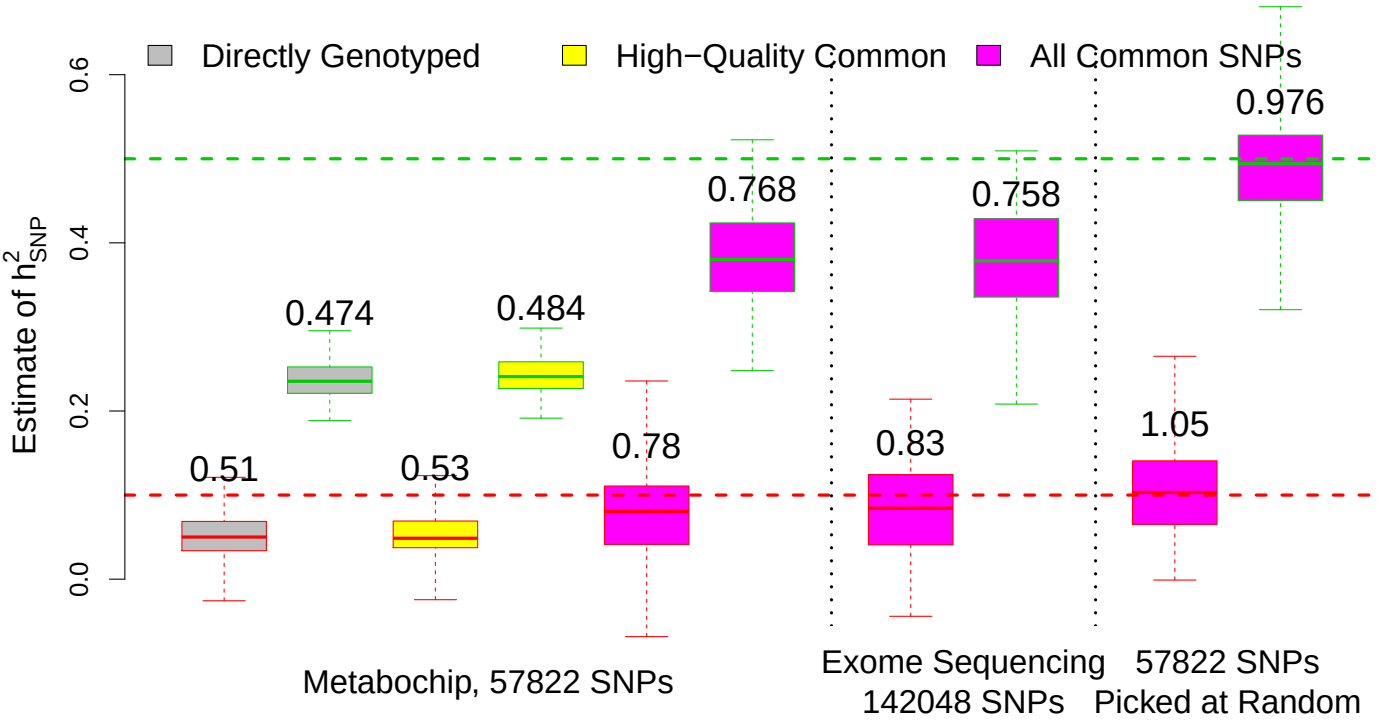
Supplementary Figure 15: Estimating enrichment of DHS for UCLEB traits. Annotations of DNaseI hypersensitive sites are from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz. Using the LDAK Model with $\alpha = -0.25$, we estimate the contribution to h^2_{SNP} of DHS for each of the 23 UCLEB traits, then the (inverse variance weighted) average, first using only the 353 K high-quality common SNPs, then all 17.3 M SNPs. We estimate that the average contributions are 33% (SD 2) and 36% (SD 4), which compared to the expected contributions (24% and 19%; horizontal lines) represent enrichment of 1.4-fold (SD 0.1) and 1.9-fold (SD 0.2).



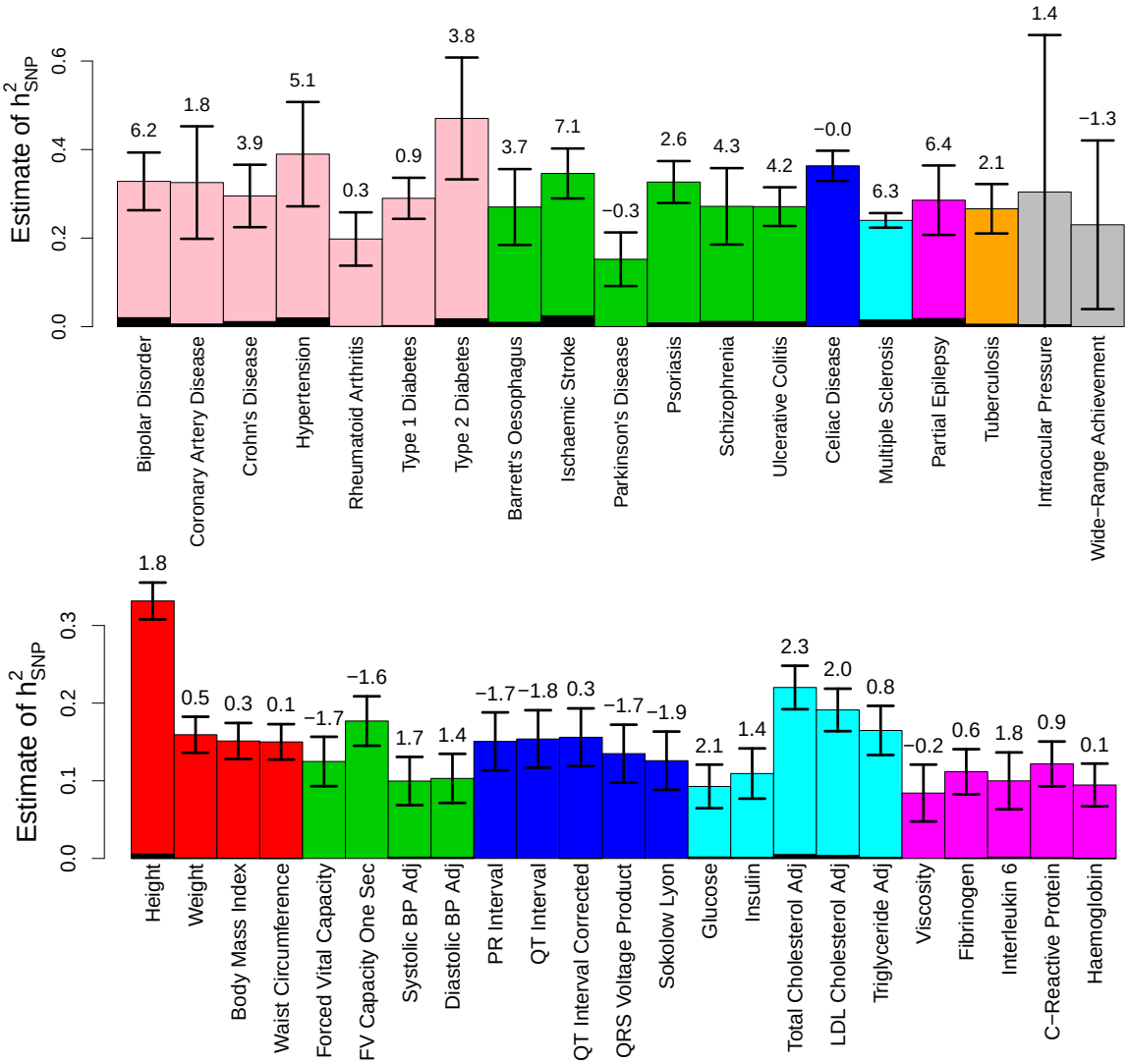
Supplementary Figure 16: Estimating enrichment for 52 SNP classes for GWAS traits. We consider the 52 SNP classes examined by Finucane *et al.*,[35] downloadable at https://data.broadinstitute.org/alkesgroup/LDSCORE (annotations are contained in the file `baseline_bedfiles.tgz`). For comparison, we also include "Genic" SNPs (those inside or within 2 kb of exons) and DHS SNPs (as defined by Gusev *et al.*[2]). Assuming the LDAK Model with $\alpha = -0.25$, red lines provide point estimates and 95% confidence intervals for the fraction of h^2_{SNP} contributed by each SNP class, averaged across the 19 GWAS traits, while black lines mark the expected fractions. Although many SNP classes show significant enrichment (green labels indicate classes with $P < 0.001$, while numbers report $-\log(P)$ for these classes), we note that overall, our estimated enrichments are much more modest than those computed by Finucane *et al.* using LDSC; for example, our highest enrichment is 2.5-fold as opposed to 13-fold. It may appear concerning that significant positive enrichment is observed for 35 out of 52 classes, however, it should be noted that SNP classes are highly-correlated; in particular, there is very significant overlap between all 35 of these classes and genic SNPs ($P < 10^{-100}$ from a hypergeometric test).



Supplementary Figure 17: Estimates of h^2_{SNP} for different SNP filterings for UCLEB traits. Each plot corresponds to a different trait, while the final plot considers the (inverse variance weighted) average across all 23 traits. Within each plot, the bars report estimates of h^2_{SNP} for each of nine SNP filterings, determined by MAF and r_j threshold (vertical lines provide 95% confidence intervals). All estimates are obtained using LDAK with $\alpha = -0.25$. For filterings which consider only common SNPs (red bars), these are based on the non-partitioned model; for filterings including also rare SNPs (green bars), h^2_{SNP} is estimated using a partitioned model, with SNPs divided by MAF. We see that in general reducing the r_j threshold initially has little impact on estimates of h^2_{SNP}, however, admitting low-quality SNPs ($0 \leq r_j \leq 0.6$) results in a significant increase, as does including rare SNPs.

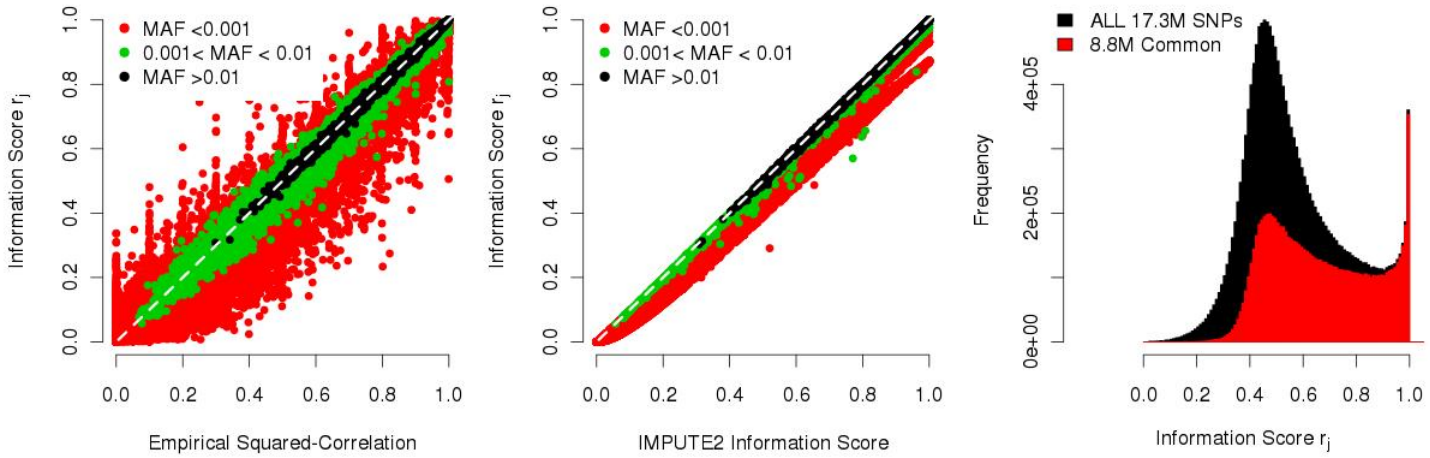


Supplementary Figure 18: Estimating h^2_{SNP} from sparse genotyping, proof of principal. For this analysis, we use Simulation Dataset I (see Supplementary Note 7). In order to increase precision of h^2_{SNP} estimates, we restrict to Chromosomes 1 & 2 (711 164 SNPs). We generate phenotypes each with 200 causal SNPs and effect sizes sampled under the LDAK Model, with $h^2_{\text{SNP}} = 0.1$ or 0.5 (we refer to this as true h^2_{SNP}). For three different starting subsets of SNPs, we investigate how much of true h^2_{SNP} can be recovered via imputation. First we reduce to the 57 822 SNPs present on the Metabochip (of which 9 218 are on Chromosomes 1 or 2). Boxes report estimates of h^2_{SNP} for 100 simulated phenotypes with true h^2_{SNP} 0.1 (red borders) or 0.5 (green borders); numbers above boxes report the average fraction of true h^2_{SNP} recovered. When we use only the 9 218 starting SNPs, we recover approximately half of h^2_{SNP}. If we impute, but retain only the 49 668 high-quality ($r_j > .99$) common SNPs, the fraction increases by only 2-3%. However, if we impute and retain all 1 611 157 common SNPs (allowing for genotype certainty), the fraction of h^2_{SNP} recovered is close to four-fifths (an increase of 49% or 58% compared to using only high-quality SNPs). For comparison, when we instead start with the 142 048 SNPs inside or within 100 basepairs of exons (of which 23 951 are on Chromosomes 1 or 2), the fraction of true h^2_{SNP} recovered after imputation is about four-fifths, indicating that the coverage of the Metabochip is approximately equivalent to that of exome sequencing. We also perform the analysis starting with 9 218 SNPs picked at random, showing that were the Metabochip SNPs evenly spread (rather than predominantly gene-centric), then with imputation we could expect to recover almost all of h^2_{SNP}.

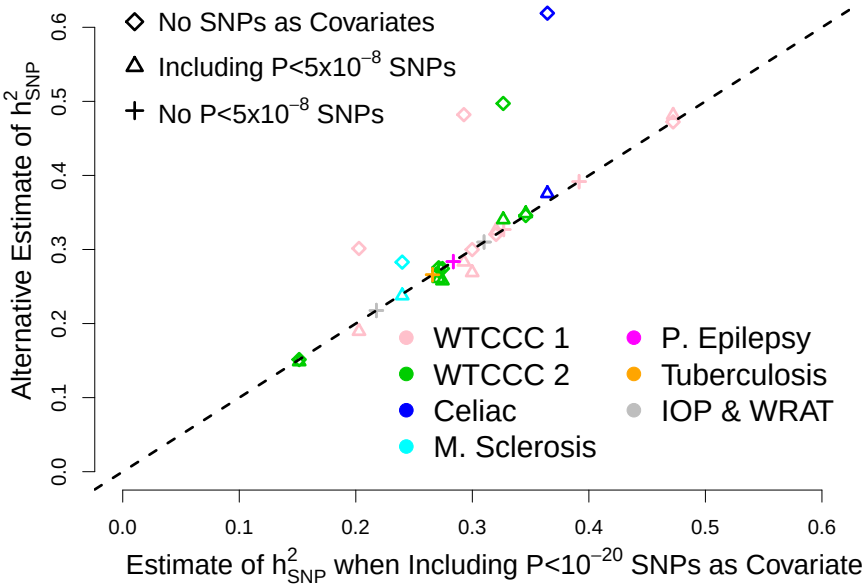


Supplementary Figure 19: Testing for inflation due to cryptic relatedness for GWAS and UCLEB traits. Accurate estimation of h^2_{SNP} relies on individuals being "unrelated" (only distantly related) and population homogeneous. Otherwise, estimates will reflect not only the variance explained by the GWAS SNPs (and variants in local LD with these), but also that of other genetic factors tagged through long-range LD and of environmental contributions which correlate with familial relatedness. The fundamental premise of our test[36] is that genetic similarities due to population structure and residual relatedness should be approximately uniform across the genome. For example, we would expect the allelic correlation for a pair of full-sibs to be close to 0.5 whether we consider all SNPs or just those on Chromosome 1. Therefore, inflation of h^2_{SNP} due to these phenomena should also be approximately uniform; i.e., an estimate of h^2_{SNP} computed from all SNPs should be as inflated as an estimate from only Chromosome 1 SNPs. For each dataset, we partition SNPs into (approximate) quarters (Chromosomes 1-3, 4-7, 8-11 & 12-23). First we analyze all SNPs together (using the partitioned model), from which we estimate $A = h^2_1 + h^2_2 + h^2_3 + h^2_4 + h^2_C$, where h^2_j is the variance explained by SNPs in Quarter j and h^2_C is inflation due to cryptic relatedness. Next we analyze each quarter separately, which provides estimates of $B_j = h^2_j + h^2_C$. Therefore, an estimate of h^2_C is $(B_1 + B_2 + B_3 + B_4 - A)/3$. In this figure, colored bars report estimates of h^2_{SNP} for the 19 GWAS and 23 UCLEB trait: within each, the black bar indicates the estimated inflation; numbers above bars express the inflation as a percent of h^2_{SNP}.

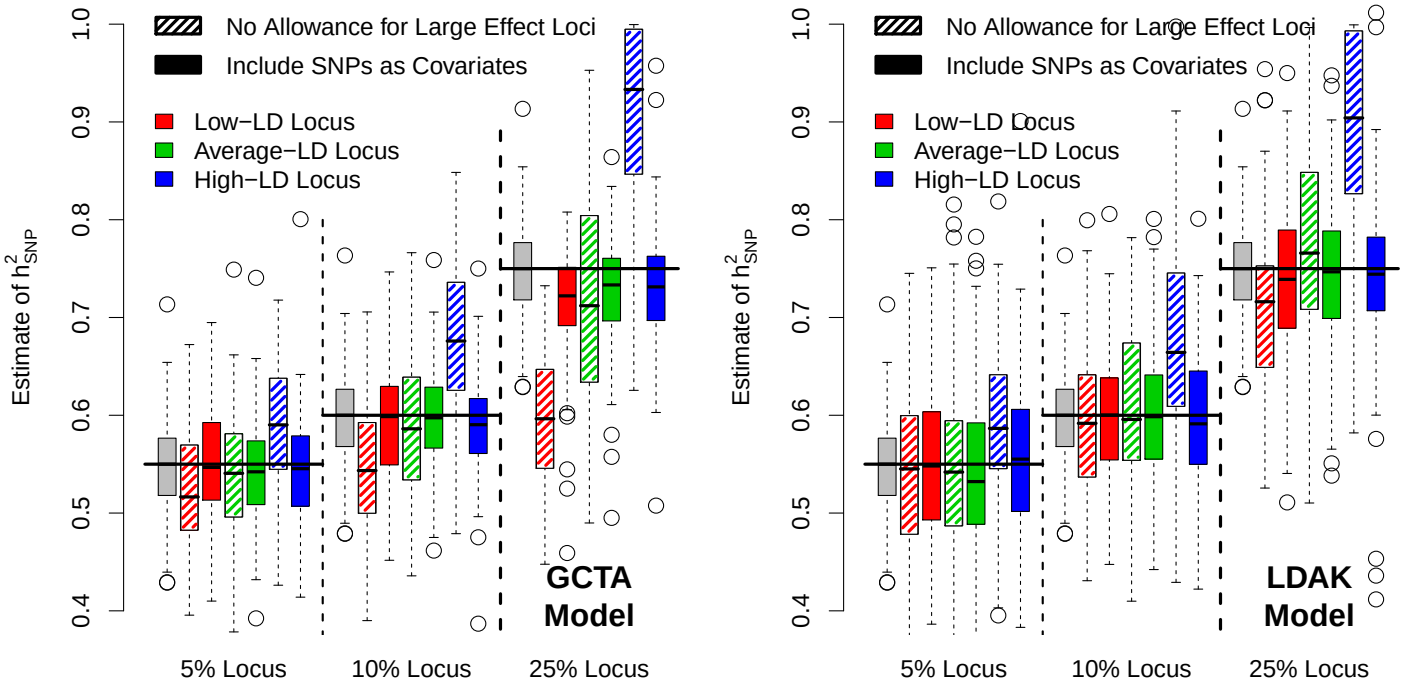
Recently, Kumar *et al.*[37] estimated $h^2_{\text{SNP}} = 0.26$ (SD 0.05) for systolic blood pressure, based on a dataset containing 49 214 SNPs in approximate linkage equilibrium. Next they repeatedly analyzed a random 10% of the genome (5 000 SNPs), noticing that the resulting estimate was almost always higher than 0.26/10. They presented this as evidence that the estimation process is flawed. We would argue that it instead shows their dataset is wholly unsuitable for estimation of h^2_{SNP}. We would expect their first analysis to provide an estimate of $A = h^2_{\text{SNP}} + h^2_C$, and their second to estimate $B = 0.1\, h^2_{\text{SNP}} + h^2_C$, and so an estimate of h^2_C is $(10\, B - A)/9$. Although numerical values are not reported, Figure 4 of their paper suggests the average estimate of B was at least 0.1, which indicates that h^2_C is at least 0.08, and their estimate of 0.26 is inflated by at least a third.



Supplementary Figure 20: Information score r_j. Our information score r_j is designed to approximate the squared correlation between S_j, the genotypes used for SNP j in the analysis, and its true genotypes. To test the accuracy of r_j, we consider 50 000 imputed SNPs from the middle of Chromosome 1: 71-76 Mb. We use the Blue Mountains Eye Study data, as this is our smallest dataset (n=2 635), so inaccuracies are likely to be most apparent. For each SNP, we sample a set of genotypes from the IMPUTE2 genotype probabilities, then compute the squared-correlation between these and the dosages. We repeat this ten times to obtain average squared-correlation. Plot 1 shows very high concordance between r_j and empirical squared-correlation, although it does decline with MAF (point colors). We note that r_j is very similar in design to the information score reported by Beagle,[38] and we find it correlates very highly with that from IMPUTE2[24] (Plot 2). This suggests that when adjusting effect-size prior variance for genotype certainty, other information scores can be used in place of r_j. Plot 3 reports the distribution of r_j for all 17.3 M SNPs (black bars) used in the UCLEB data, and for the 7.8 M of those with MAF >0.01 (red bars).

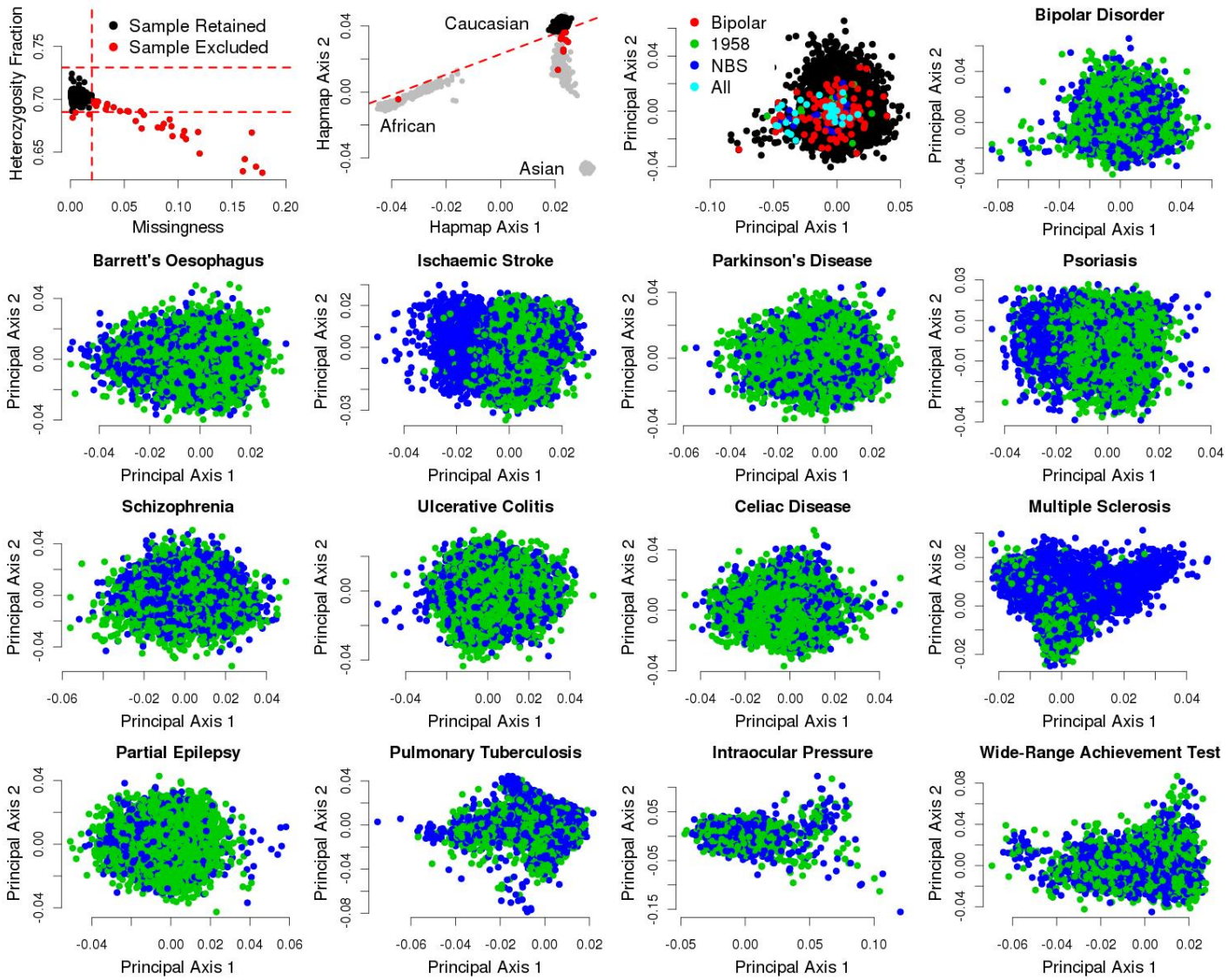


Supplementary Figure 21: Including highly-associated SNPs as covariates for GWAS traits. For reasons explained in Supplementary Figure 22, for all GCTA and LDAK analyses, we include as covariates SNPs with $P < 10^{-20}$ from marginal analysis (but include their contribution to phenotypic variance within σ^2_g when computing $h^2_{\text{SNP}} = \sigma^2_g/(\sigma^2_g+\sigma^2_e)$). Our definition of highly-associated is somewhat arbitrary; however, we confirm estimates of h^2_{SNP} are largely unchanged if instead we use $P < 5 \times 8^{-8}$, the conventional GWAS significance threshold (i.e., the triangles are close to the diagonal). By contrast, large inaccuracies can result if h^2_{SNP} is estimated without allowing for large effect loci (the diamonds are noticeably above the diagonal for rheumatoid arthritis, type 1 diabetes, psoriasis, celiac disease and multiple sclerosis, five autoimmune traits with a substantial contribution from the major histocompatibility complex). Points marked by a cross have no SNPs with $P < 5 \times 10^{-8}$ so are not affected for either threshold.

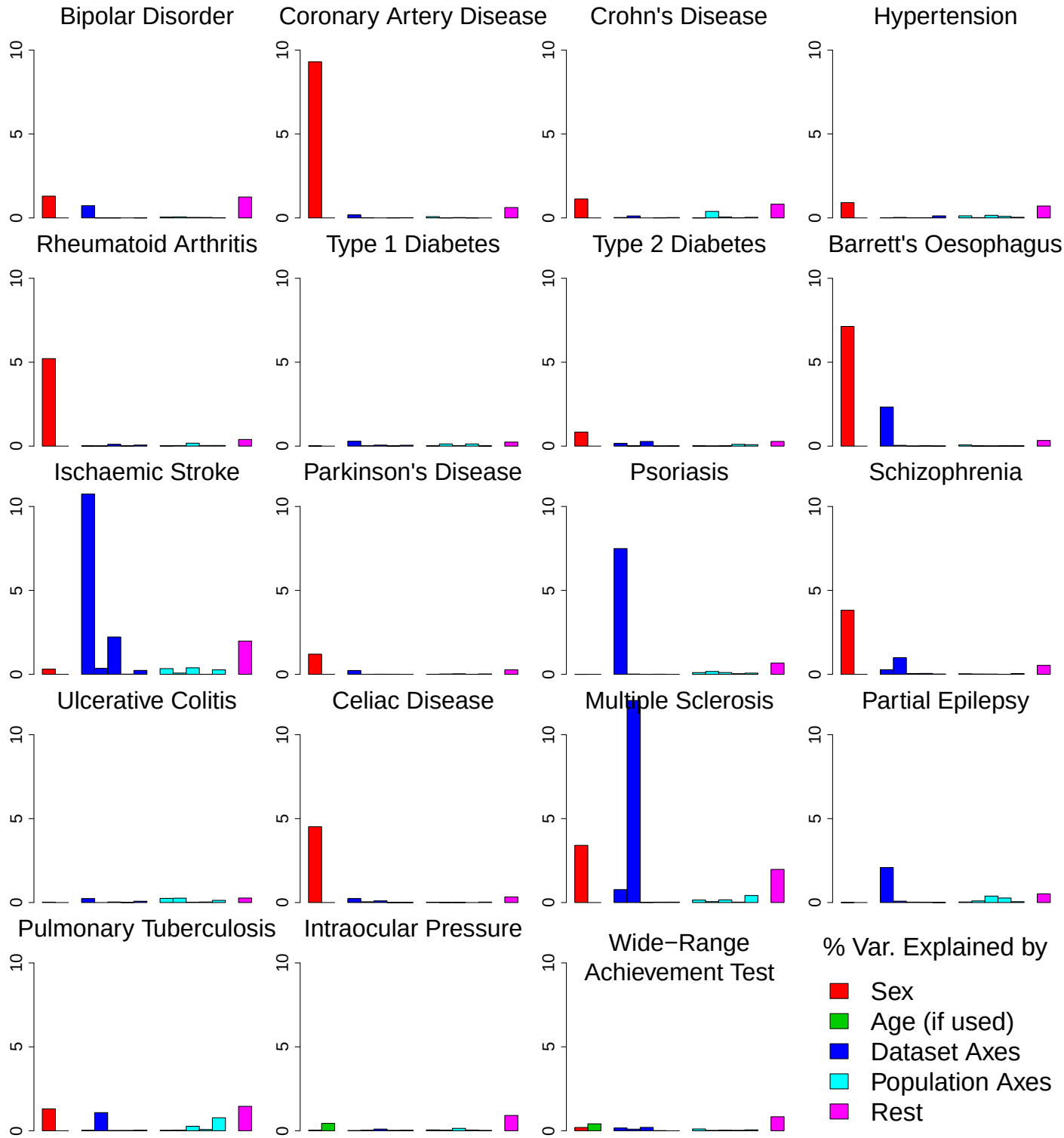


Supplementary Figure 22: Accommodating very large effect loci. Before estimating h^2_{SNP}, we recommend performing a single-SNP association analysis to identify highly-associated SNPs ($P < 10^{-20}$), pruning these, then including those which remain as fixed-effect covariates in subsequent regressions. The pruning is mainly for logistical reasons, to reduce computational demands and avoid colinearity, so we suggest using a reasonably high correlation squared threshold (say 0.5 or 0.8). Main Text Equation (1) becomes $\boldsymbol{y} \sim \mathbb{N}(\boldsymbol{Z\theta} + \boldsymbol{T\phi}, \boldsymbol{K}\sigma^2_g + \boldsymbol{I}\sigma^2_e)$, where columns of the matrix $\boldsymbol{T}$ contain the genotypes of the (pruned) highly-associated SNPs (i.e., $\boldsymbol{T}$ is a submatrix of the full genotype matrix $\boldsymbol{S}$). The final estimate of h^2_{SNP} will be $(\sigma^2_g + \sigma^2_T)/(\sigma^2_g + \sigma^2_T + \sigma^2_e)$, where $\sigma^2_T = (\boldsymbol{T\phi})^T(\boldsymbol{T\phi})$, the variance explained by the highly-associated SNPs. For example, suppose the total variance of a phenotype is 10, of which fixed-effect covariates explain 2 and all SNPs explain 4. When using conventional covariates (e.g., age and sex), these are treated as nuisance variables, so the variance they explain is ignored and the estimate of h^2_{SNP} would be $4/(10-2) = 0.5$. By contrast, when the covariates are highly-associated SNPs, their contributions are included, and so the estimate of h^2_{SNP} is $(4+2)/10 = 0.6$. In LDAK, highly-associated SNPs are included as covariates by adding the option `--top-snps` when performing REML using the command `--reml`.

The first plot demonstrates the problem caused by individual loci with very large effect. Using Simulation Dataset I ($n = 5\,134$ individuals; $m = 4\,710\,536$ common SNPs; sum of SNP weights $\sum_j w_j = 136\,407$; see Supplementary Note 7), we generate phenotypes with $h^2_{\text{SNP}} = 0.55, 0.6$ or 0.75. For each phenotype, 1000 causal SNPs explain 50% of phenotypic variance, with the remainder of h^2_{SNP} (0.05, 0.1 or 0.25) explained by a single locus picked from either a region of low LD (average LD-score <68.0; red boxes), median LD (96.8-133.7; green boxes) or high LD (>200.2; blue boxes); these boundaries represent the 20$^{\text{th}}$, 40$^{\text{th}}$, 60$^{\text{th}}$ and 80$^{\text{th}}$ percentiles of average LD-Score (computed using non-overlapping 100 kb segments[1]). First, we sample the effect sizes of causal SNPs in line with the the GCTA Model, $\beta_j \sim \mathbb{N}(0, 1)$. Gray boxes represent the "gold standard," where the large-effect locus and its effect size is known, so its contribution to the phenotype can be excluded. Hatched boxes report estimates of h^2_{SNP} when no allowance is made for highly-associated SNPs; we see that estimates of h^2_{SNP} are deflated when the large-effect SNP is in a low-LD region (red boxes), but inflated when in a high-LD region (blue boxes), and in all cases, precision is lower than the gold standard. Solid boxes report estimates of h^2_{SNP} when highly-associated SNPs are included as fixed effects using the protocol described above (SNPs with $P < 10^{-20}$ from single-SNP regression are pruned then included as covariates). We see that estimates now appear to be unbiased and precision is comparable with the gold standard. The second plot shows that conclusions are similar if instead effect sizes are sampled under the LDAK Model, $\beta_j \sim \mathbb{N}(0, w_j)$.

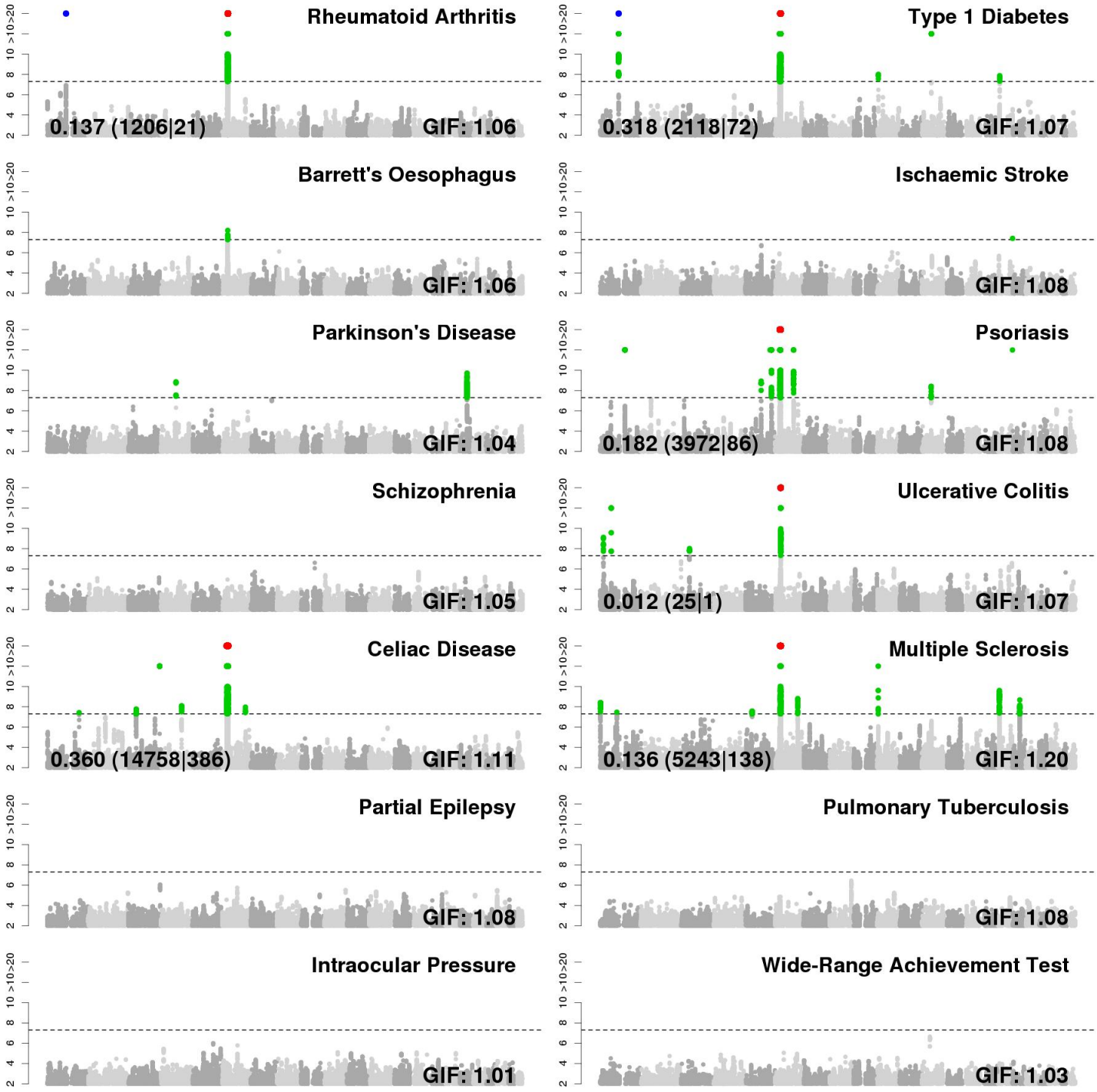


Supplementary Figure 23: Sample quality control for GWAS traits. Supplementary Note 1 describes our quality control steps; here we provide some examples. **Plot 1:** For each cohort in turn (the plot demonstrates for Cohort 1; see Supplementary Table 12), we exclude outliers based on sample missingness and heterozygosity rate. These two metrics tend to be correlated, so we find it useful to consider them jointly when deciding suitable exclusion thresholds (red lines). **Plot 2:** We then project each cohort onto population axes computed from HapMap data, excluding samples (red points) which stray from the Caucasian cluster. **Plot 3:** Having imputed cohorts, then combined these to form datasets, we filter out relatedness based on (unweighted) allelic correlations computed from a pruned set of SNPs. Specifically, we exclude individuals until no pair remains with genetic correlation higher than c, where $-c$ is the smallest correlation observed (achieved using the command `--filter` in LDAK). We first perform this filtering for each cohort in turn, then for all samples at once. The plot corresponds to Dataset 1 (see Supplementary Table 11, showing which samples were excluded when considering the bipolar (red points), 1958 Birth Cohort (green) and National Blood Service (dark blue) samples separately, then when considering the remaining samples together (light blue). **Plot 4-16:** Having filtered relatedness, we performed principal component analysis. We plot the leading two axes for a selection of traits (to save space, we exclude six of the well-studied WTCCC 1 cases, coronary artery disease, Crohn's Disease, hypertension, rheumatoid arthritis, type 1 diabetes and type 2 diabetes, for which the plots closely resemble that for bipolar disorder). Blue points indicate cases, green controls (for the two quantitative GWAS traits, we treat individuals with phenotype above the median as cases). For most datasets, cases and controls appear to be well-mixed for the leading axes. The least satisfactory dataset is multiple sclerosis (Plot 7), where (even after excluding Finnish samples), the cases, (recruited from 14 countries), were more heterogeneous than the controls (mainly from UK, with some from Sweden). However, our test for inflation of h^2_{SNP} due to cryptic relatedness (Supplementary Figure 19) indicates this heterogeneity is not a serious problem. For all heritability analyses, we included the top 20 principal components as covariates (in addition to 10 population axes derived from 1000 Genomes data, and for some traits, sex and age; see Supplementary Figure 24).



Supplementary Figure 24: Phenotypic variance explained by covariates for GWAS traits. For all analyses we include a minimum of 30 covariates: 20 "Dataset Axes" (eigen-vectors of the allelic correlation matrix computed from pruned SNPs) and 10 "Population Axes" (obtained by projecting our dataset onto principal components from 1000 Genomes.[39]. For the GWAS traits, we additionally included sex, and for intraocular pressure and wide-range achievement test also age. The plots report, for each trait, the proportion of phenotypic variance explained by each covariate. We individually report for sex, age, and the first 5 dataset and population axes, then for the remaining 15+5 dataset and population axes combined.

The apparent sex effects may be caused by ascertainment (different sampling of males and females between cases and controls). The population axes appear largely redundant. For the traits where the dataset axes explain substantial phenotypic variance, notably schizophrenia, ischaemic stroke, psoriasis, and multiple sclerosis, this appears localized to the top two axes — in particular, the final 20 axes ("Rest") at most explain 2% of variance — indicating that possible confounders are well taken care of in our analyses (consistent with the results of Supplementary Figure 19).



Supplementary Figure 25: Single-SNP association testing for GWAS traits. Plots report, for a selection of traits, $-log_{10}(P)$ from single-SNP association analysis, including as covariates sex, 20 principal component axes from the dataset, and projections of the data onto 10 principal axes from 1000 Genomes;[39] for intraocular pressure and wide-range achievement test, we also included age as a covariate. Green points mark SNPs with $P < 5 \times 10^{-8}$, red those with $P < 10^{-20}$; the latter are the SNPs which, after pruning, are included as covariates when estimating h^2_{SNP}. We report the genomic inflation factor (GIF), and for traits with highly-associated SNPs, the proportion of phenotypic variance these SNPs explain, and their number pre- and post-pruning. To save space, we exclude bipolar disorder, coronary artery disease, Crohn's disease, hypertension and type 2 diabetes, five of the well-studied WTCCC 1 traits,[9] none of which had highly-associated SNPs.

Trait	n	Estimates of h^2_{SNP} (SD) for Different SNP filterings		
		High-Quality SNPs	**All Common SNPs**	**All SNPs**
		m=353 090, W=38 818	m=8 819 943, W=277 469	m=17 260 213, W=998 445
Height	10 965	0.33 (0.02)	0.48 (0.05)	0.58 (0.11)
Weight	11 005	0.15 (0.02)	0.23 (0.05)	0.39 (0.11)
Body Mass Index	10 933	0.15 (0.02)	0.18 (0.05)	0.32 (0.11)
Waist Circumference	10 956	0.14 (0.02)	0.15 (0.05)	0.31 (0.11)
Forced Vital Capacity	7 871	0.13 (0.03)	0.16 (0.07)	0.33 (0.15)
FV Capacity One Sec	7 871	0.18 (0.03)	0.15 (0.06)	0.26 (0.15)
Systolic BP Adj	7 906	0.10 (0.03)	0.20 (0.07)	0.21 (0.15)
Diastolic BP Adj	7 906	0.09 (0.03)	0.14 (0.06)	0.08 (0.15)
PR Interval	6 460	0.16 (0.04)	0.29 (0.08)	0.27 (0.18)
QT Interval	6 647	0.13 (0.04)	0.29 (0.08)	0.45 (0.18)
QT Interval Corrected	6 647	0.16 (0.04)	0.26 (0.08)	0.44 (0.18)
QRS Voltage Product	6 458	0.14 (0.04)	0.22 (0.08)	0.54 (0.18)
Sokolow Lyon	6 460	0.11 (0.04)	0.23 (0.08)	0.21 (0.18)
Glucose	8 421	0.10 (0.03)	0.09 (0.06)	0.09 (0.14)
Insulin	7 457	0.11 (0.03)	0.13 (0.07)	0.05 (0.16)
Total Cholesterol Adj	8 678	0.23 (0.03)	0.30 (0.06)	0.37 (0.13)
LDL Cholesterol Adj	9 177	0.19 (0.03)	0.28 (0.06)	0.44 (0.13)
Triglyceride Adj	7 341	0.16 (0.03)	0.32 (0.07)	0.13 (0.16)
Viscosity	6 647	0.09 (0.04)	0.17 (0.08)	-0.08 (0.17)
Fibrinogen	8 543	0.11 (0.03)	0.10 (0.06)	0.16 (0.14)
Interleukin 6	6 753	0.09 (0.03)	0.24 (0.08)	0.22 (0.17)
C-Reactive Protein	8 581	0.11 (0.03)	0.17 (0.06)	-0.01 (0.13)
Haemoglobin	8 852	0.09 (0.03)	0.29 (0.06)	0.55 (0.14)
Average	**8 197**	**0.15 (0.01)**	**0.22 (0.01)**	**0.29 (0.03)**
Relative	**- - -**	**1 (- - -)**	**1.45 (0.08)**	**1.91 (0.18)**
Relative	**- - -**	**0.64 (0.02)**	**1 (- - -)**	**1.29 (0.12)**

Supplementary Table 1: Estimates of h^2_{SNP} for UCLEB traits. n = sample size, m = number of SNPs, $W = \sum_j r_j w_j$, where r_j and w_j denote the information scores and SNP weights, respectively. For high-quality common SNPs (r_j >0.99 and MAF>0.01) and all common SNPs (MAF>0.01), we estimate h^2_{SNP} using the (non-partitioned) LDAK Model with $\alpha = -0.25$; for all SNPs (MAF>0.0005) we use the LDAK Model with $\alpha = -0.25$ and SNPs divided by MAF (with boundaries at 0.001, 0.0025, 0.01 and 0.1).

Trait	$\alpha = -1.25$	$\alpha = -1.0$	$\alpha = -0.75$	$\alpha = -0.5$	$\alpha = -0.25$	$\alpha = 0$	$\alpha = 0.25$
Bipolar Disorder	-3188	-3176	-3166	-3161	-3158	**-3158**	-3158
Coronary Artery Disease	-3114	-3109	-3105	-3102	-3100	-3099	**-3098**
Crohn's Disease	-3071	-3061	-3055	-3051	-3049	**-3049**	-3049
Hypertension	-3318	-3310	-3304	-3300	-3297	-3296	**-3296**
Rheumatoid Arthritis	-2864	-2861	-2859	-2858	**-2858**	-2858	-2859
Type 1 Diabetes	-2311	-2308	-2306	-2304	**-2304**	-2304	-2305
Type 2 Diabetes	-3305	-3298	-3293	-3290	-3288	**-3287**	-3287
Barrett's Oesophagus	-3840	-3832	-3826	-3823	-3820	-3819	**-3819**
Ischaemic Stroke	-5537	-5527	-5521	**-5520**	-5520	-5522	-5525
Parkinson's Disease	-3866	-3860	-3855	-3851	-3849	-3848	**-3847**
Psoriasis	-4025	-4014	-4007	-4004	**-4003**	-4004	-4006
Schizophrenia	-3197	-3189	-3183	-3180	**-3180**	-3180	-3181
Ulcerative Colitis	-5113	-5095	-5083	-5077	**-5075**	-5075	-5076
Celiac Disease	-4519	-4517	-4515	-4514	-4513	-4513	**-4513**
Multiple Sclerosis	-7938	-7906	-7887	**-7880**	-7880	-7884	-7891
Partial Epilepsy	-2940	-2937	**-2935**	-2935	-2936	-2938	-2939
Tuberculosis	-7323	-7318	-7315	**-7315**	-7316	-7318	-7320
Intraocular Pressure	-5381	-5380	-5380	-5379	**-5379**	-5379	-5379
Wide-range Achievement Test	-14787	-14787	-14786	-14786	**-14786**	-14786	-14786
Average	-4718	-4710	-4704	-4702	**-4701**	-4701	-4702

Supplementary Table 2: Inferring the relationship between heritability and MAF for GWAS traits. This is a numerical version of the first plot in Figure 1. Values report log likelihoods from the LDAK Model for seven α. Higher likelihoods indicate better-fitting α; the highest value for each trait is marked in **bold**.

Trait	LDSC	GCTA	GCTA-MS	GCTA-LDMS	LDAK
Bipolar Disorder	2.50 (0.48)	1	1.04 (0.11)	1.27 (0.13)	1.69 (0.15)
Coronary Artery Disease	1.31 (0.96)	1	1.03 (0.20)	1.47 (0.24)	1.88 (0.29)
Crohn's Disease	1.72 (0.73)	1	0.97 (0.12)	1.30 (0.14)	1.69 (0.18)
Hypertension	3.61 (0.71)	1	1.00 (0.17)	1.20 (0.19)	1.91 (0.23)
Rheumatoid Arthritis	1.42 (0.63)	1	1.00 (0.12)	1.01 (0.14)	1.36 (0.18)
Type 1 Diabetes	1.55 (0.45)	1	1.02 (0.06)	0.98 (0.07)	1.21 (0.09)
Type 2 Diabetes	1.76 (0.79)	1	1.04 (0.15)	1.38 (0.18)	1.77 (0.22)
Barrett's Oesophagus	1.38 (0.55)	1	0.96 (0.18)	1.34 (0.20)	1.96 (0.25)
Ischaemic Stroke	0.86 (0.36)	1	1.06 (0.12)	1.43 (0.14)	2.03 (0.16)
Parkinson's Disease	1.17 (0.58)	1	1.00 (0.19)	1.13 (0.21)	1.82 (0.27)
Psoriasis	1.02 (0.36)	1	1.06 (0.08)	1.33 (0.09)	1.65 (0.11)
Schizophrenia	0.89 (0.82)	1	1.08 (0.21)	1.24 (0.23)	2.12 (0.29)
Ulcerative Colitis	1.11 (0.33)	1	1.02 (0.10)	1.20 (0.10)	1.79 (0.13)
Celiac Disease	2.02 (1.22)	1	1.00 (0.03)	1.07 (0.04)	1.15 (0.05)
Multiple Sclerosis	0.81 (0.16)	1	1.04 (0.04)	1.10 (0.04)	1.43 (0.05)
Partial Epilepsy	-0.86 (0.67)	1	1.12 (0.21)	1.39 (0.24)	2.16 (0.30)
Pulmonary Tuberculosis	1.00 (0.50)	1	1.07 (0.15)	1.33 (0.16)	1.96 (0.21)
Intraocular Pressure	0.20 (1.56)	1	0.90 (0.49)	1.53 (0.70)	1.97 (0.88)
Wide-range Achievement Test	6.40 (4.01)	1	0.73 (0.71)	1.58 (0.89)	2.73 (1.19)
Average	1.10 (0.10)	1	1.02 (0.02)	1.14 (0.02)	1.45 (0.03)

Supplementary Table 3: Relative estimates of h^2_{SNP} for GWAS traits. This is a numerical version of Figure 3a in the main text. Values report estimates of h^2_{SNP} (with SDs in brackets) from LDSC, GCTA-MS (partitioned by MAF), GCTA-LDMS (partitioned by MAF and LD) and LDAK, relative to those from GCTA; the final row provides (inverse variance weighted) averages.

	$P < 10^{-20}$ SNPs included as fixed effects				No allowance for highly-associated SNPs			
	$\alpha = -0.25$		$\alpha = -1.0$		$\alpha = -0.25$		$\alpha = -1.0$	
Trait	GCTA	LDAK	GCTA	LDAK	GCTA	LDAK	GCTA	LDAK
Bipolar Disorder	-3176	**-3158**	-3179	**-3176**	-3176	**-3158**	-3179	**-3176**
Coronary Artery Disease	-3107	**-3100**	**-3108**	-3109	-3107	**-3100**	**-3108**	-3109
Crohn's Disease	-3055	**-3049**	-3062	**-3061**	-3055	**-3049**	-3062	**-3061**
Hypertension	-3311	**-3297**	-3313	**-3310**	-3311	**-3297**	-3313	**-3310**
Rheumatoid Arthritis	-2859	**-2858**	**-2860**	-2861	**-3064**	-3114	**-3072**	-3122
Type 1 Diabetes	-2313	**-2304**	-2311	**-2308**	**-3147**	-3230	**-3145**	-3230
Type 2 Diabetes	-3296	**-3288**	-3298	**-3298**	-3296	**-3288**	-3298	**-3298**
Barrett's Oesophagus	-3832	**-3820**	-3835	**-3832**	-3832	**-3820**	-3835	**-3832**
Ischaemic Stroke	-5565	**-5520**	-5565	**-5527**	-5565	**-5520**	-5565	**-5527**
Parkinson's Disease	-3852	**-3849**	**-3854**	-3860	-3852	**-3849**	**-3854**	-3860
Psoriasis	-4035	**-4003**	-4034	**-4014**	**-4081**	-4163	**-4086**	-4164
Schizophrenia	-3192	**-3180**	-3195	**-3189**	-3192	**-3180**	-3195	**-3189**
Ulcerative Colitis	-5107	**-5075**	-5113	**-5095**	-5144	**-5114**	-5151	**-5138**
Celiac Disease	-4519	**-4513**	-4519	**-4517**	**-4252**	-4638	**-4268**	-4664
Multiple Sclerosis	-7977	**-7880**	-7968	**-7906**	**-7894**	-7978	**-7903**	-8029
Partial Epilepsy	-2951	**-2936**	-2950	**-2937**	-2951	**-2936**	-2950	**-2937**
Tuberculosis	-7335	**-7316**	-7331	**-7318**	-7335	**-7316**	-7331	**-7318**
Intraocular Pressure	-5380	**-5379**	**-5380**	-5380	-5380	**-5379**	**-5380**	-5380
Wide-range Achievement Test	-14788	**-14786**	-14788	**-14787**	-14788	**-14786**	-14788	**-14787**
Average	-4718	**-4701**	-4719	**-4710**	**-4759**	-4785	**-4762**	-4796
Height	-15298	**-15297**	-15316	**-15309**	-15298	**-15297**	-15316	**-15309**
Weight	**-15474**	-15474	-15475	**-15474**	**-15474**	-15474	-15475	**-15474**
Body Mass Index	-15398	**-15387**	-15399	**-15389**	-15398	**-15387**	-15399	**-15389**
Waist Circumference	-15446	**-15431**	-15445	**-15430**	-15446	**-15431**	-15445	**-15430**
Forced Vital Capacity	-11074	**-11067**	-11074	**-11069**	-11074	**-11067**	-11074	**-11069**
FV Capacity One Sec	-11078	**-11068**	-11079	**-11071**	-11078	**-11068**	-11079	**-11071**
Systolic BP Adj	-11114	**-11113**	**-11113**	-11113	-11114	**-11113**	**-11113**	-11113
Diastolic BP Adj	-11110	**-11109**	-11111	**-11108**	-11110	**-11109**	-11111	**-11108**
PR Interval	-9031	**-9028**	-9032	**-9029**	-9076	**-9072**	-9077	**-9073**
QT Interval	-9334	**-9332**	-9336	**-9331**	-9334	**-9332**	-9336	**-9331**
QT Interval Corrected	-9329	**-9326**	-9331	**-9327**	-9329	**-9326**	-9331	**-9327**
QRS Voltage Product	-9063	**-9057**	-9063	**-9058**	-9063	**-9057**	-9063	**-9058**
Sokolow Lyon	-9071	**-9068**	-9071	**-9067**	-9071	**-9068**	-9071	**-9067**
Glucose	-11839	**-11837**	-11839	**-11838**	-11839	**-11837**	-11839	**-11838**
Insulin	-10500	**-10494**	-10500	**-10495**	-10500	**-10494**	-10500	**-10495**
Total Cholesterol Adj	**-12004**	-12016	**-12007**	-12015	**-12132**	-12140	**-12132**	-12135
LDL Cholesterol Adj	**-12620**	-12626	**-12621**	-12623	**-12853**	-12853	-12852	**-12846**
Triglyceride Adj	**-10211**	-10224	**-10214**	-10223	**-10257**	-10272	**-10259**	-10270
Viscosity	**-9343**	-9344	**-9344**	-9344	**-9343**	-9344	**-9344**	-9344
Fibrinogen	-12028	**-12027**	-12029	**-12028**	-12028	**-12027**	-12029	**-12028**
Interleukin 6	-9484	**-9483**	-9484	**-9483**	-9484	**-9483**	-9484	**-9483**
C-Reactive Protein	-12026	**-12025**	-12026	**-12023**	-12073	**-12072**	-12073	**-12069**
Haemoglobin	**-12411**	-12415	**-12412**	-12413	**-12411**	-12415	**-12412**	-12413
Average	-11491	**-11489**	-11492	**-11490**	-11512	**-11510**	-11514	**-11510**

Supplementary Table 4: Comparing the GCTA and LDAK Models for GWAS and UCLEB traits. Values report log likelihoods and are presented in pairs; the highest of each pair is marked in **bold**. When $\alpha = -0.25$, our recommended value, the LDAK Model fits better than the GCTA Model for all 19 GWAS traits and for 17 of the 23 UCLEB traits (first pair of columns). The LDAK remains superior when $\alpha = -1$, albeit the average improvement in log likelihood is reduced (second pair). This reduction is because the GCTA Model implicitly gives more weight to high-MAF SNPs than the LDAK Model (high-MAF SNPs tend to be in regions of higher-LD), which partially makes up for using α too high.

Estimates of h^2_{SNP} are based on an infinitesimal model, which is violated when there are loci of very large effect. For this reason, we advocate including highly-associated SNPs as fixed effects (Supplementary Figures 21 & 22). For the final four columns, these SNPs are no longer given special consideration (traits with highly-associated SNPs are marked in red). For 5 of the 6 affected GWAS traits, model fit is now higher under the GCTA Model. This is because for these traits, almost all the large effect loci are located in the major histocompatibility complex (MHC), a region of very high LD, which is implicitly given higher weighting under the GCTA Model.

Trait	Including r_j	Excluding r_j	Trait	Including r_j	Excluding r_j
Height	**-15365**	-15373	Sokolow Lyon	**-9068**	-9069
Weight	**-15488**	-15489	Glucose	**-11842**	-11842
Body Mass Index	**-15404**	-15404	Insulin	-10499	**-10499**
Waist Circumference	**-15446**	-15447	Total Cholesterol Adj	**-12037**	-12040
Forced Vital Capacity	-11073	**-11073**	LDL Cholesterol Adj	**-12716**	-12718
FV Capacity One Sec	**-11082**	-11082	Triglyceride Adj	**-10225**	-10226
Systolic BP Adj	**-11114**	-11115	Viscosity	**-9344**	-9345
Diastolic BP Adj	**-11111**	-11112	Fibrinogen	**-12034**	-12034
PR Interval	**-9030**	-9032	Interleukin 6	**-9482**	-9483
QT Interval	-9331	**-9331**	C-Reactive Protein	**-12019**	-12020
QT Interval Corrected	-9330	**-9329**	Haemoglobin	**-12407**	-12410
QRS Voltage Product	-9060	**-9059**	**Average**	**-11500**	-11501

Supplementary Table 5: Examining the relationship between heritability and genotype certainty for UCLEB traits. The LDAK Model allows for genotype certainty by scaling effect-size prior variance by r_j, our information score. Here we consider all common SNPs. Values report log likelihoods for each trait, with and without allowance for genotype certainty. Including r_j in the heritability model provides a modest overall improvement in model fit, and results in higher likelihood for 18 of the 23 traits.

Trait Keyword	#SNPs	h^2_{GWAS}	Trait Keyword	#SNPs	h^2_{GWAS}	Trait Keyword	#SNPs	h^2_{GWAS}
Bipolar disorder	24	0.02	Coronary artery disease	13	0.03	Crohn's disease	151	0.21
Hypertension	8	<0.01	Rheumatoid arthritis	122	0.19	Type 1 diabetes	65	0.27
Type 2 diabetes	83	0.08	Barrett's esophagus	2	<0.01	Ischemic stroke	6	<0.01
Parkinson's disease	57	0.03	Psoriasis	56	0.21	Schizophrenia	146	0.07
Ulcerative colitis	91	0.12	Celiac disease	36	0.29	Multiple sclerosis	82	0.17
Epilepsy	5	<0.01	Pulmonary Tuberculosis	7	<0.01	Intraocular pressure	16	0.02

Supplementary Table 6: Proportion of phenotypic variance explained by genome-wide significant SNPs for the GWAS traits. For each trait, we searched the most recent version (August 2016) of the GWAS Catalog[40] (www.ebi.ac.uk/gwas/docs/downloads) for SNPs reported as genome-wide significant ($P < 5 \times 10^{-8}$). We restrict to SNPs present in 1000 Genomes[39] and where the corresponding GWAS was described as European, Canadian or Ashkenazi Jews, or if the SNP was discovered through a global meta-analysis consortium, where individuals were primarily of Caucasian ancestry. The table lists for each trait the (case-sensitive) keyword we used, how many SNPs this returned, and h^2_{GWAS}, the proportion of phenotypic variance these SNPs explained in our data (calculated using ordinary least squares regression). For disease traits, h^2_{GWAS} has been transformed to the liability scale.[3,4] Note that by using the keyword "Epilepsy", we did not restrict to SNPs specifically associated with partial epilepsy.[15] We were unable to find any $P < 5 \times 10^{-8}$ SNPs for wide-range achievement test, even when we consider related keywords such as "Cognitive."

Trait (Prevalence)	h^2_{GWAS}	Estimates of h^2_{SNP} (SD) from Different Methods						
		Previous	LDSC	GCTA	GCTA-MS	GCTA-LDMS	LDAK$_{\alpha=-1}$	LDAK$_{\alpha=-0.25}$
Bipolar Disorder (0.5)	0.02	0.24 (0.04)[2]	0.53 (0.10)	0.22 (0.02)	0.21 (0.02)	0.26 (0.03)	0.33 (0.03)	0.35 (0.03)
Coronary Artery Disease (6)	0.03	0.25 (0.06)[2]	0.28 (0.21)	0.22 (0.04)	0.22 (0.04)	0.31 (0.05)	0.32 (0.07)	0.40 (0.06)
Crohn's Disease (0.5)	0.21	0.26 (0.01)[41]	0.33 (0.14)	0.19 (0.02)	0.18 (0.02)	0.24 (0.03)	0.30 (0.04)	0.32 (0.03)
Hypertension (5)	<0.01	0.33 (0.06)[2]	0.86 (0.17)	0.24 (0.04)	0.23 (0.04)	0.29 (0.05)	0.39 (0.06)	0.46 (0.06)
Rheumatoid Arthritis (0.5)	0.19	0.09 (0.03)[2]	0.22 (0.10)	0.15 (0.02)	0.14 (0.02)	0.15 (0.02)	0.19 (0.03)	0.21 (0.03)
Type 1 Diabetes (0.5)	0.27	0.13 (0.03)[2]	0.38 (0.11)	0.26 (0.01)	0.26 (0.01)	0.27 (0.02)	0.31 (0.02)	0.31 (0.02)
Type 2 Diabetes (8)	0.08	0.42 (0.07)[2]	0.54 (0.24)	0.30 (0.05)	0.30 (0.05)	0.40 (0.05)	0.47 (0.07)	0.54 (0.07)
Barrett's Oesophagus (1.6)	<0.01	0.25 (0.05)[42]	0.23 (0.09)	0.17 (0.03)	0.15 (0.03)	0.22 (0.04)	0.27 (0.04)	0.32 (0.04)
Ischaemic Stroke (2)	<0.01	0.25 (0.03)[43]	0.15 (0.06)	0.18 (0.02)	0.18 (0.02)	0.25 (0.02)	0.35 (0.03)	0.34 (0.03)
Parkinson's Disease (0.2)	0.03	0.27 (0.05)[44]	0.13 (0.06)	0.11 (0.02)	0.10 (0.02)	0.11 (0.02)	0.15 (0.03)	0.20 (0.03)
Psoriasis (0.5)	0.21	0.35 (0.06)[45]	0.21 (0.07)	0.21 (0.02)	0.21 (0.02)	0.27 (0.02)	0.32 (0.02)	0.34 (0.02)
Schizophrenia (1)	0.07	0.23 (0.01)[46]	0.12 (0.11)	0.15 (0.03)	0.14 (0.03)	0.16 (0.04)	0.27 (0.04)	0.30 (0.04)
Ulcerative Colitis (0.2)	0.12	0.19 (0.01)[41]	0.17 (0.05)	0.16 (0.02)	0.15 (0.02)	0.20 (0.02)	0.27 (0.02)	0.28 (0.02)
Celiac Disease (1)	0.29	0.33 (0.04)[47]	0.65 (0.39)	0.30 (0.01)	0.30 (0.01)	0.33 (0.01)	0.35 (0.02)	0.35 (0.02)
Multiple Sclerosis (0.1)	0.17	0.17 (0.01)[2]	0.13 (0.03)	0.17 (0.01)	0.17 (0.01)	0.19 (0.01)	0.24 (0.01)	0.24 (0.01)
Partial Epilepsy (0.3)	<0.01	0.33 (0.05)[36]	-0.11 (0.08)	0.14 (0.03)	0.14 (0.03)	0.19 (0.03)	0.28 (0.04)	0.27 (0.04)
Pulmonary Tuberculosis (4)	<0.01	None Found	0.13 (0.07)	0.15 (0.02)	0.15 (0.02)	0.19 (0.02)	0.27 (0.03)	0.26 (0.03)
Intraocular Pressure	0.02	None Found	0.04 (0.30)	0.19 (0.11)	0.14 (0.10)	0.26 (0.14)	0.31 (0.18)	0.38 (0.17)
Wide-Range Achievement Test	<0.01	0.43 (0.10)[48]	0.49 (0.31)	0.07 (0.06)	0.05 (0.06)	0.10 (0.07)	0.22 (0.10)	0.21 (0.09)
Average	---	--- ---	**0.17 (0.02)**	**0.19 (0.00)**	**0.19 (0.00)**	**0.22 (0.00)**	**0.27 (0.01)**	**0.28 (0.01)**
Relative	---	--- ---	**1.05 (0.10)**	**1 (---)**	**0.99 (0.02)**	**1.15 (0.02)**	**1.39 (0.03)**	**1.43 (0.03)**
Relative	---	--- ---	**0.86 (0.08)**	**0.87 (0.02)**	**0.86 (0.02)**	**1 (---)**	**1.22 (0.02)**	**1.25 (0.02)**

Supplementary Table 7: Estimates of h^2_{SNP} for GWAS traits. We report estimates from LDSC[32] (LD Score Regression), GCTA, GCTA-MS[33] (GCTA stratified by MAF) and GCTA-LDMS[1] (stratified by MAF and LD); all of these methods use $\alpha = -1$ when scaling genotypes. The final two columns provide estimates from LDAK, first using $\alpha = -1$, then using $\alpha = -0.25$ (our recommended value). For disease traits, estimates have been converted to the liability scale based on the assumed prevalence.[3,4] For comparison, we include previous estimates of h^2_{SNP}; where possible, these are based on Caucasian samples, but for psoriasis, the only available estimate used Han Chinese individuals. We also report h^2_{GWAS}, the proportion of phenotypic variance explained by SNPs reported as genome-wide significant (see Supplementary Table 6)

Trait	Contribution to h^2_{SNP} (SD) ‖ Predicted Contribution ‖ Enrichment (SD)																	
	DHS			Exonic			Genic			Inter-Genic			DHS - Genotyped			DHS - GCTA		
Bipolar Disorder	15 (5)	17.5	0.8 (0.3)	4.1 (2)	1.9	2.1 (1.1)	19 (5)	17.6	1.1 (0.3)	16 (4)	18.7	0.8 (0.2)	13 (6)	18.9	0.7 (0.3)	36 (13)	12.7	2.8 (1.1)
Coronary Artery Disease	31 (9)	17.5	1.8 (0.5)	7 (4)	1.9	3.7 (1.9)	35 (9)	17.5	2.0 (0.5)	6 (7)	18.7	0.3 (0.4)	38 (10)	18.8	2.0 (0.5)	78 (20)	12.7	6.2 (1.6)
Crohn's Disease	29 (7)	17.5	1.7 (0.4)	2.7 (2)	1.9	1.5 (1.3)	31 (6)	17.6	1.7 (0.3)	10 (5)	18.7	0.5 (0.3)	35 (7)	18.8	1.9 (0.4)	110 (16)	12.7	8.7 (1.3)
Hypertension	23 (8)	17.5	1.3 (0.4)	3.6 (3)	1.9	1.9 (1.5)	32 (7)	17.6	1.8 (0.4)	12 (6)	18.7	0.6 (0.3)	25 (8)	18.9	1.3 (0.4)	77 (20)	12.7	6.1 (1.6)
Rheumatoid Arthritis	8 (12)	17.5	0.5 (0.7)	7.8 (5)	1.9	4.1 (2.5)	43 (12)	17.6	2.4 (0.7)	6 (9)	18.7	0.3 (0.5)	-8 (15)	18.9	-0.4 (0.8)	95 (28)	12.7	7.5 (2.2)
Type 1 Diabetes	31 (10)	17.5	1.8 (0.6)	10.3 (4)	1.9	5.5 (2.1)	28 (9)	17.5	1.6 (0.5)	5 (7)	18.7	0.3 (0.4)	9 (15)	18.9	0.5 (0.8)	126 (21)	12.7	9.9 (1.7)
Type 2 Diabetes	22 (8)	17.6	1.3 (0.4)	4.2 (3)	1.9	2.2 (1.5)	28 (7)	17.6	1.6 (0.4)	6 (6)	18.7	0.3 (0.3)	20 (8)	18.8	1.1 (0.4)	69 (18)	12.7	5.5 (1.4)
Barrett's Oesophagus	13 (8)	18.3	0.7 (0.5)	3.3 (3)	1.9	1.8 (1.7)	27 (8)	18.1	1.5 (0.4)	17 (6)	17.4	1.0 (0.3)	20 (8)	20.7	1.0 (0.4)	63 (20)	13.4	4.7 (1.5)
Ischaemic Stroke	14 (5)	18.3	0.8 (0.3)	-0.3 (2)	1.9	-0.1 (1.0)	14 (5)	18.1	0.7 (0.3)	15 (4)	17.4	0.9 (0.2)	28 (6)	20.5	1.4 (0.3)	59 (13)	13.3	4.5 (1.0)
Parkinson's Disease	30 (10)	18.4	1.6 (0.5)	8.5 (4)	1.9	4.6 (2.0)	24 (9)	18.1	1.3 (0.5)	10 (6)	17.4	0.6 (0.4)	33 (11)	20.6	1.6 (0.5)	74 (20)	13.4	5.5 (1.5)
Psoriasis	30 (6)	18.2	1.6 (0.3)	5.4 (2)	1.8	3.0 (1.3)	29 (6)	18.1	1.6 (0.3)	7 (4)	17.4	0.4 (0.2)	27 (8)	20.7	1.3 (0.4)	105 (15)	13.4	7.9 (1.1)
Schizophrenia	16 (9)	17.4	0.9 (0.5)	3.5 (3)	1.7	2.1 (1.9)	15 (8)	17.8	0.9 (0.5)	14 (6)	18.0	0.8 (0.4)	22 (10)	20.8	1.1 (0.5)	39 (24)	12.8	3.1 (1.9)
Ulcerative Colitis	28 (5)	18.4	1.5 (0.3)	4.4 (2)	2.0	2.2 (1.0)	27 (5)	17.6	1.5 (0.3)	6 (4)	17.8	0.3 (0.2)	32 (6)	20.1	1.6 (0.3)	112 (11)	13.6	8.3 (0.8)
Celiac Disease	26 (10)	18.0	1.4 (0.5)	8.2 (4)	2.1	3.9 (1.9)	38 (9)	18.7	2.0 (0.5)	-1 (7)	17.9	-0.0 (0.4)	15 (12)	19.5	0.8 (0.6)	74 (21)	13.0	5.7 (1.6)
Multiple Sclerosis	29 (3)	18.2	1.6 (0.2)	4.5 (1)	1.8	2.5 (0.6)	33 (3)	18.0	1.8 (0.2)	11 (2)	17.4	0.6 (0.1)	33 (3)	20.5	1.6 (0.2)	88 (6)	13.3	6.6 (0.5)
Partial Epilepsy	33 (9)	18.0	1.8 (0.5)	-3.7 (3)	1.8	-2.1 (1.8)	9 (8)	18.1	0.5 (0.5)	16 (6)	17.5	0.9 (0.4)	30 (10)	20.2	1.5 (0.5)	64 (22)	13.1	4.9 (1.7)
Pulmonary Tuberculosis	22 (7)	18.1	1.2 (0.4)	5.2 (2)	1.9	2.8 (1.3)	26 (6)	17.7	1.5 (0.3)	7 (5)	18.1	0.4 (0.3)	29 (8)	20.1	1.5 (0.4)	75 (15)	13.5	5.6 (1.1)
Intraocular Pressure	31 (29)	18.8	1.6 (1.6)	0.1 (10)	2.1	0.0 (4.9)	-21 (29)	18.6	-1.1 (1.5)	17 (19)	17.4	1.0 (1.1)	24 (28)	20.8	1.2 (1.4)	6 (70)	13.9	0.4 (5.0)
Wide-range Achievement Test	29 (25)	17.5	1.6 (1.4)	8.9 (10)	1.8	5.0 (5.5)	4 (24)	18.1	0.2 (1.4)	-6 (19)	18.0	-0.3 (1.1)	86 (89)	19.7	4.4 (4.5)	129 (87)	12.6	10.2 (6.9)
Average	**24 (2)**	**18.0**	**1.4 (0.1)**	**4.0 (1)**	**1.9**	**2.1 (0.3)**	**27 (1)**	**17.9**	**1.5 (0.1)**	**10 (1)**	**17.8**	**0.6 (0.1)**	**27 (2)**	**20.0**	**1.4 (0.1)**	**82 (4)**	**13.1**	**6.3 (0.3)**

Supplementary Table 8: Enrichment of DHS and other SNP classes for GWAS traits. Triplets report the estimated contribution to h^2_{SNP}, the predicted contribution, and the enrichment (estimated contribution divided by predicted) for DHS, exonic, genic and inter-genic SNPs. Except for the final block of results (see below), estimates and predictions are calculated assuming the LDAK Model with $\alpha = -0.25$. For DHS (DNaseI hypersensitive sites), we use annotations provided at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeRegDnaseClustered/wgEncodeRegDnaseClusteredV3.bed.gz. Exonic SNPs are those within exons according to RefSeq gene annotations.[49] Genic SNPs are defined (here) as those inside or within 2 kb of exons, while inter-genic are those further than 125 kb from exons (we chose these thresholds to ensure that DHS, genic and inter-genic SNPs are predicted to contribute equally to h^2_{SNP}). In general, results are based on imputed data (on average 5.1 M SNPs per trait); for DHS, we also perform the analysis using only directly genotyped SNPs (on average 235 K SNPs), noting that enrichment remains 1.4-fold, despite the reduction in SNP density.

Gusev *et al.*[2] performed a similar analysis, but instead assumed the GCTA Model with $\alpha = -1$; across 11 traits (10 of which are included in our 19), they estimated that DHS contribute on average 79% (SD 8) of h^2_{SNP}. For the final block, we copy their approach, obtaining a comparable estimate of 82% (SD 4); as DHS on average contain 13% of SNPs, this corresponds to 6.3-fold enrichment (note that Gusev *et al.*[2] reported enrichment as 5.1-fold, as they instead compared 79% to an estimate of the effective size of DHS). There are three main reasons why the GCTA and LDAK Models result in such different conclusions. Firstly, their contrasting estimates of h^2_{SNP}: 16.2% (SD 0.4) vs 25.0% (SD 0.6). Secondly, their contrasting estimates of the (absolute) heritability of DHS: 13.6% (SD 0.6) vs 6.3% (SD 0.4). Thirdly, their contrasting estimates of the predicted contributions of DHS: 18.0% vs 13.1% (a difference mainly due to DHS tending to have lower-than-average LD).

	GCTA (2)			GCTA-LD4 (5)			GCTA-LDMS (21)			LDAK (2)			LDAK-LD4 (5)			LDAK-LDMS (21)		
Trait	**logL**	**LRT**	**△ AIC**	**logL**	**LRT**	**△ AIC**	**logL**	**LRT**	**△ AIC**	**logL**	**LRT**	**△ AIC**	**logL**	**LRT**	**△ AIC**	**logL**	**LRT**	**△ AIC**
Bipolar Disorder	-3176	101.4	99.4	-3167	120.3	112.3	-3161	131.5	91.5	-3158	136.8	**134.8**	-3157	138.8	130.8	-3151	131.5	111.9
Coronary Artery Disease	-3107	29.5	27.5	-3098	48.3	40.3	-3092	59.7	19.7	-3100	43.9	41.9	-3097	50.2	**42.2**	-3089	59.7	26.4
Crohn's Disease	-3055	83.6	81.6	-3046	102.5	**94.5**	-3035	125.1	85.1	-3049	96.0	94.0	-3049	96.8	88.8	-3041	125.1	71.8
Hypertension	-3311	43.0	41.0	-3305	55.9	47.9	-3296	73.6	33.6	-3297	70.6	**68.6**	-3294	76.2	68.2	-3284	73.6	56.6
Rheumatoid Arthritis	-2859	20.7	18.7	-2858	24.0	16.0	-2850	40.0	0.0	-2858	23.5	**21.5**	-2857	24.4	16.4	-2850	40.0	-1.0
Type 1 Diabetes	-2313	19.0	17.0	-2310	24.9	16.9	-2304	35.4	-4.6	-2304	35.6	**33.6**	-2303	38.2	30.2	-2299	35.4	6.8
Type 2 Diabetes	-3296	53.6	51.6	-3288	68.5	60.5	-3280	85.6	45.6	-3288	69.2	**67.2**	-3285	74.2	66.2	-3281	85.6	43.2
Barrett's Oesophagus	-3832	44.6	42.6	-3826	56.5	48.5	-3819	71.4	31.4	-3820	67.5	**65.5**	-3820	69.4	61.4	-3813	71.4	42.9
Ischaemic Stroke	-5565	84.6	82.6	-5547	121.1	113.1	-5547	121.1	81.1	-5520	174.4	172.4	-5515	185.4	**177.4**	-5515	121.1	145.4
Parkinson's Disease	-3852	41.5	39.5	-3851	42.8	34.8	-3843	59.2	19.2	-3849	47.7	**45.7**	-3848	50.5	42.5	-3835	59.2	35.9
Psoriasis	-4035	45.6	43.6	-4021	73.2	65.2	-4012	90.8	50.8	-4003	108.8	106.8	-4000	115.5	**107.5**	-3989	90.8	96.4
Schizophrenia	-3192	34.0	32.0	-3190	38.8	30.8	-3185	49.0	9.0	-3180	59.4	**57.4**	-3179	61.7	53.7	-3170	49.0	38.3
Ulcerative Colitis	-5107	124.2	122.2	-5095	149.0	141.0	-5095	149.0	109.0	-5075	189.8	**187.8**	-5073	192.3	184.3	-5073	149.0	152.3
Celiac Disease	-4519	29.7	27.7	-4513	41.4	33.4	-4506	54.7	14.7	-4513	40.5	**38.5**	-4513	42.4	34.4	-4506	54.7	15.2
Multiple Sclerosis	-7977	430.0	428.0	-7941	502.1	494.1	-7941	502.1	462.1	-7880	624.0	**622.0**	-7877	629.7	621.7	-7877	502.1	589.7
Partial Epilepsy	-2951	28.3	26.3	-2945	40.6	32.6	-2945	40.6	0.6	-2936	57.5	**55.5**	-2934	61.3	53.3	-2934	40.6	21.3
Pulmonary Tuberculosis	-7335	60.4	58.4	-7327	75.9	67.9	-7310	110.9	70.9	-7316	97.6	**95.6**	-7315	101.1	93.1	-7300	110.9	89.6
Intraocular Pressure	-5380	3.9	1.9	-5377	8.4	0.4	-5367	29.2	-10.8	-5379	5.1	**3.1**	-5379	5.3	-2.7	-5370	29.2	-16.1
Wide-Range Achievement	-14788	2.5	0.5	-14787	3.8	-4.2	-14784	10.2	-29.8	-14786	5.4	**3.4**	-14784	9.7	1.7	-14776	10.2	-14.6
Mean	-4718	67.4	65.4	-4710	84.1	76.1	-4704	96.8	56.8	-4701	102.8	**100.8**	-4699	106.5	98.5	-4692	96.8	79.6
Median	-3832	43.0	41.0	-3826	55.9	47.9	-3819	71.4	31.4	-3820	67.5	**65.5**	-3820	69.4	61.4	-3813	71.4	42.9

Supplementary Table 9: Is it beneficial to partition by LD? As well as the non-partitioned versions of LDAK and GCTA, we consider partitioning by LD (4 tranches) and by both LD & MAF (LDMS: 20 tranches); for all methods we use $\alpha = -0.25$. Triplets of values report the log likelihood, the likelihood ratio test statistic (twice the improvement in log likelihood compared to the null model) and the change in Akaike information criterion (explained below), for each of the 19 GWAS traits.

Yang *et al.* claimed that GCTA-LDMS is superior to LDAK, however we have shown the evidence they provided to support this claim is invalid, as it relied on a simulation study which by design favored the GCTA Model (Supplementary Figures 10 & 11). Comparing models of different complexity is not straightforward. If two heritability models are of equal complexity (i.e., use the same number of SNP tranches), then we can compare simply by model likelihood. Therefore, we can conclude from the above values that LDAK tends to outperform GCTA, LDAK-LD4 tends to outperform GCTA-LD4, and LDAK-LDMS tends to outperform GCTA-LDMS. To compare models of different complexity, it is necessary to introduce penalization, to allow for the fact that model fit is expected to increase with complexity. The AIC is defined as twice the number of parameters minus twice the log likelihood; we report change in AIC, obtained by subtracting the AIC of the null model. We see that for all except one trait, the improvement in model fit from partitioning by LD or by LD & MAF, is not sufficient to justify the increase in complexity, and that based on AIC, LDAK performs approximately twice as well as GCTA-LDMS.

Trait	No weighting		Exon weighting	
	logL	h^2_{GWAS}	logL	h^2_{GWAS}
Bipolar Disorder	**-3158**	0.71 (0.06)	-3159	0.70 (0.06)
Coronary Artery Disease	-3100	0.43 (0.07)	**-3099**	0.43 (0.06)
Crohn's Disease	-3049	0.64 (0.07)	**-3045**	0.65 (0.06)
Hypertension	-3297	0.52 (0.06)	**-3297**	0.52 (0.06)
Rheumatoid Arthritis	-2858	0.41 (0.06)	**-2858**	0.41 (0.06)
Type 1 Diabetes	-2304	0.64 (0.04)	**-2304**	0.63 (0.04)
Type 2 Diabetes	-3288	0.53 (0.06)	**-3285**	0.54 (0.06)
Barrett's Oesophagus	**-3820**	0.40 (0.05)	-3821	0.39 (0.05)
Ischaemic Stroke	**-5520**	0.51 (0.04)	-5522	0.50 (0.04)
Parkinson's Disease	-3849	0.37 (0.05)	**-3848**	0.37 (0.05)
Psoriasis	-4003	0.60 (0.04)	**-4001**	0.60 (0.04)
Schizophrenia	**-3180**	0.53 (0.07)	-3181	0.52 (0.07)
Ulcerative Colitis	-5075	0.60 (0.04)	**-5068**	0.61 (0.04)
Celiac Disease	-4513	0.48 (0.02)	**-4510**	0.49 (0.02)
Multiple Sclerosis	-7880	0.65 (0.02)	**-7871**	0.65 (0.02)
Partial Epilepsy	**-2936**	0.39 (0.05)	-2937	0.39 (0.05)
Tuberculosis	-7316	0.32 (0.03)	**-7315**	0.32 (0.03)
Intraocular Pressure	**-5379**	0.38 (0.17)	-5379	0.37 (0.17)
Wide-range Achievement Test	-14786	0.21 (0.09)	**-14786**	0.22 (0.09)
Average	-4701	0.52 (0.01)	**-4699**	0.52 (0.01)

Supplementary Table 10: Incorporating distance from exons for GWAS traits. The LDAK Model assumes $\mathbb{E}[h_j^2] \propto [f_j(1-f_j)]^{1+\alpha} \times w_j \times r_j$. Here, we consider also $\mathbb{E}[h_j^2] \propto [f_j(1-f_j)]^{1+\alpha} \times w_j \times r_j \times \exp(\frac{-(D_j+50)}{500})$, where D_j is the distance (in kb) between SNP j and the nearest exon. This model is designed so that genic SNPs are expected to have approximately twice the heritability of inter-genic SNPs. This table reports, for each of the GWAS traits, the log likelihood and estimate of h^2_{SNP} (with SD in brackets) under the standard and exon-weighted LDAK Models. We see that weighting SNPs based on distance from exons typically improves model fit, to be expected considering that genic SNPs tend to have higher heritability than inter-genic SNPs (Supplementary Table 8). However, overall, the improvement in likelihood is slight, and there is negligible effect on estimates of h^2_{SNP}.

Trait	Cohorts	m	W	Trait	Cohorts	m	W
Bipolar Disorder	1,8,9	2 729 104	79 491	Barrett's Oesophagus	10,16,18	3 830 533	116 165
Coronary Artery Disease	2,8,9	2 738 568	79 671	Ischaemic Stroke	11,16,18	3 797 362	115 334
Crohn's Disease	3,8,9	2 723 516	79 331	Parkinson's Disease	12,16,18	3 819 831	116 270
Hypertension	4,8,9	2 739 910	79 673	Psoriasis	13,16,18	3 814 818	115 935
Rheumatoid Arthritis	5,8,9	2 735 531	79 567	Schizophrenia	14,32	3 481 050	110 868
Type 1 Diabetes	6,8,9	2 731 788	79 609	Ulcerative Colitis	15,17,19	4 061 726	114 724
Type 2 Diabetes	7,8,9	2 735 552	79 643	Pulmonary Tuberculosis	26,27	2 987 383	102 256
Celiac Disease	20,21	2 681 987	88 382	Intraocular Pressure	28	4 149 456	124 792
Multiple Sclerosis	16,18,22,23	3 702 284	113 154	Wide-range Achievement Test	29,30,31	2 593 016	87 823
Partial Epilepsy	16,18,24,25	3 399 056	108 134	UCLEB Traits	33-40	353 090	38 892
Illumina Control-Control	16,18	4 710 536	136 407	Affymetrix Control-Control	17,19	4 179 761	119 137

Supplementary Table 11: Construction of Datasets. m = number of SNPs, $W = \sum_j w_j$ = sum of SNP weights. Cohort numbers are explained in Supplementary Table 12. For UCLEB, m and W refer to our main analysis, which considers only high-quality common SNPs (MAF>0.01, r_j >0.99). We use the two control-control datasets (bottom row) to examine the adequacy of our quality control (see Supplementary Note 6).

	Cohort (Assession Code)	Genotyping Platform	Quality Control 1		QC 2
			n	m	m
1	WTCCC 1: Bipolar Disease (EGAD00000000003)	Affymetrix 500 K	1 963	399 943	3 015 567
2	WTCCC 1: Coronary Artery Disease (EGAD00000000004)	Affymetrix 500 K	1 956	401 194	3 034 574
3	WTCCC 1: Crohn's Disease (EGAD00000000005)	Affymetrix 500 K	1 916	404 644	3 005 015
4	WTCCC 1: Hypertension (EGAD00000000006)	Affymetrix 500 K	1 969	402 326	3 041 849
5	WTCCC 1: Rheumatoid ArthritisEGAD00000000007)	Affymetrix 500 K	1 932	402 817	3 033 111
6	WTCCC 1: Type 1 Diabetes (EGAD00000000008)	Affymetrix 500 K	1 980	402 621	3 020 006
7	WTCCC 1: Type 2 Diabetes (EGAD00000000009)	Affymetrix 500 K	1 957	401 226	3 020 679
8	WTCCC 1: 1958 Birth Cohort* (EGAD00000000001)	Affymetrix 500 K	1 493	406 566	3 048 680
9	WTCCC 1: National Blood Service* (EGAD00000000002)	Affymetrix 500 K	1 480	401 357	3 028 805
10	WTCCC 2: Barrett's Oesophagus (EGAD00010000506)	Illumina 670 K	1 901	529 702	4 191 898
11	WTCCC 2: Ischaemic Stroke (EGAD00010000264)	Illumina 670 K	3 972	530 785	4 124 693
12	WTCCC 2: Parkinson's Disease (EGAD00000000057)	Illumina 610 K	1 705	531 134	4 180 706
13	WTCCC 2: Psoriasis (EGAD00010000124)	Illumina 670 K	2 519	530 376	4 162 924
14	WTCCC 2: Schizophrenia† (EGAD00010000262)	Affymetrix 6.0	3 016	741 463	4 376 041
15	WTCCC 2: Ulcerative Colitis (EGAD00000000025)	Affymetrix 6.0	2 775	901 267	4 300 826
16	WTCCC 2: 1958 Birth Cohort‡	Illumina 1.2 M	2 699	741 574	4 897 381
17	(EGAD00000000022 & EGAD00000000021)	Affymetrix 6.0	2 706	901 662	4 374 316
18	WTCCC 2: National Blood Service‡	Illumina 1.2 M	2 501	738 972	4 895 769
19	(EGAD00000000024 & EGAD00000000023)	Affymetrix 6.0	2 674	741 483	4 344 699
20	Celiac Collection 1 from Dubois *et al.*[10]	Illumina Hap300v1 & Hap550v3	3 293	289 333	2 783 302
21	Celiac Collection 2§ from Dubois *et al.*[10]	Illumina 670 K & 1.2 M	6 772	515 664	4 132 558
22	International Multiple Sclerosis Genetics Consortium (IMSGC)¶[50]	Illumina 610 K	4 473	529 485	4 182 239
23	(EGAD00000000120)		4 948	529 485	4 089 368
24	Imperial & Liverpool Samples from ILAE Consortium[15]	Illumina 660 K	1 349	532 173	4 185 985
25	Melbourne Samples from ILAE Consortium[15]	Illumina 660 K	348	525 777	3 954 127
26	Russian Pulmonary Tuberculosis (RPTB)¶ from Curtis *et al.*[7]	Affymetrix 6.0	5 569	687 789	3 077 219
27			5 569	687 789	3 085 649
28	Blue Mountains Eye Study (EGAD00010000584)	Illumina 670 K	2 635	529 786	4 149 456
29	Children's Hospital of Philadelphia (CHOP) (phs000607.v1.p1)	Illumina 610 K	2 127	458 182	3 228 597
30		Illumina OmniExpressv2	1 337	644 856	4 278 434
31		Illumina Hap550v3	1 029	483 412	3 478 209
32	Trinity College Dublin from ILAE Consortium[15]	Illumina OmniQuad	2 232	732 310	4 802 327
33	British Regional Heart Study[51]	Metabochip	2 255	63 318	303 687
34	British Women's Heart and Health Study[52]	Metabochip	1 953	64 400	310 846
35	Caerphilly Prospective Study[53]	Metabochip	1 329	64 775	317 876
36	Edinburgh Artery Study[54]	Metabochip	0 749	63 385	305 564
37	English Longitudinal Study of Ageing[55]	Metabochip	1 868	64 703	312 027
38	Edinburgh Type 2 Diabetes Study[56]	Metabochip	0 990	65 181	316 859
39	MRC National Survey of Health[57]	Metabochip	2 421	64 630	310 176
40	Whitehall II Study[58]	Metabochip	3 046	65 119	309 585

Supplementary Table 12: List of Cohorts. n = sample size, m = number of SNPs. Quality Control 1 (performed just prior to imputation) filters samples based on missingness and heterozygosity, and excludes SNPs with MAF<0.01, call-rate <0.95 or Hardy-Weinberg $P < 10^{-6}$. Quality Control 2 (performed just after imputation) excludes SNPs based on MAF and information score. Wellcome Trust Case Control Consortium (WTCCC), IMSGC and Blue Mountains Eye Study data were downloaded from the European Genome-Phenome Archive (www.ebi.ac.uk/ega); CHOP data were downloaded from dbGaP (www.ncbi.nlm.nih.gov/gap); Cohort 33-40 are available upon application from the UCLEB Consortium.[21] The remaining datasets were obtained directly from the authors cited. *WTCCC 1 controls are subsets of the WTCCC 2 controls. † Unlike the other WTCCC 2 cohorts, the Schizophrenia cases are of Irish origin (recruited at Trinity College Dublin). ‡WTCCC 2 Controls were genotyped twice, on Illumina and Affymetrix SNP arrays. §Celiac Collection 2 uses WTCCC 2 Controls. ¶For both IMSGC and RPTB data, to make imputation feasible, we divided the samples into two cohorts.

References:

1. 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).

2. 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).

3. Dempster, E. & Lerner, I. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).

4. Lee, S., Wray, N., Goddard, M. & Visscher, P. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).

5. World Health Organization. Global tuberculosis report (2014).

6. Comstock, G. Tuberculosis in twins: a re-analysis of the Prophit survey. *Am. Rev. Respir. Dis.* **4**, 621–624 (1978).

7. Curtis, J. *et al.* Susceptibility to tuberculosis is associated with variants in the *asap1* gene encoding a regulator of dendritic cell migration. *Nat. Genet.* **47**, 523–527 (2015).

8. Sveinbjornsson, G. *et al.* HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. *Nat. Genet.* **48**, 318–322 (2016).

9. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

10. Dubois, P. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).

11. van Heel, D. *et al.* A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*. *Nat. Genet.* **39**, 827–829 (2007).

12. Poser, C. *et al.* New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann. Neurol.* **13**, 227–231 (1983).

13. McDonald, W. *et al.* Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.* **50**, 121–127 (2001).

14. Polman, C. *et al.* Diagnostic criteria for multiple sclerosis: 2005 revisions to the mcdonald criteria". *Ann. Neurol.* **58**, 840–846 (2005).

15. International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **13**, 893–903 (2014).

16. Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia* **30**, 389–399 (1989).

17. Berg, A., Berkovic, S. & Brodie, M. Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology. *Epilepsia* **51**, 676–85 (2010).

18. The Blue Mountains Eye Study (BMES) and The Wellcome Trust Case Control Consortium 2 (WTCCC2). Genome-wide association study of intraocular pressure identifies the *GLCCI1/ICA1* region as a glaucoma susceptibility locus. *Hum. Mol. Genet.* **22**, 4653–4660 (2013).

19. Mills, J. *et al.* Do high blood folate concentrations exacerbate metabolic abnormalities in people with low vitamin B-12 status? *Am. J. Clin. Nutr.* **94**, 495–500 (2011).

20. Molloy, A. *et al.* A common polymorphism in HIBCH influences methylmalonic acid concentrations in blood independently of cobalamin. *Am. J. Hum. Genet.* **5**, 869–882 (2016).

21. 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).

22. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).

23. Delaneau, O., Zagury, J. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).

24. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).

25. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

26. Price, A. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).

27. Anderson, C. *et al.* Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).

28. Corbeil, R. & Searle, S. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics* **18**, 31–38 (1976).

29. Gilmour, A., Thompson, R. & Cullis, B. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440–1450 (1995).

30. Lee, S. & van der Werf, J. 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).

31. Bulik-Sullivan, B. Relationship between LD Score and Haseman-Elston Regression (2015). Preprint available on BioRχiv.

32. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2014).

33. Lee, S. *et al.* Estimation of SNP-heritability from dense genotype data. *Am. J. Hum. Genet.* **93**, 1151–1155 (2013).

34. Golan, D., Lander, E. & Rosset, S. Measuring missing heritability: Inferring the contribution of common variants. *PNAS* **111**, E5272E5281 (2014).

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

36. Speed, D. *et al.* Describing the genetic architecture of epilepsy through heritability analysis. *Brain* **137**, 26802689 (2014).

37. Kumar, S., Feldman, M., Rehkopf, D. & Tuljapurkar, S. Limitations of GCTA as a solution to the missing heritability problem. *PNAS* **113**, E61E70 (2015).

38. Browning, B. & Browning, S. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).

39. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

40. Welter, D. *et al.* The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Am. J. Clin. Nutr.* **42**, D1001–1006 (2013).

41. Chen, G. *et al.* Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum. Mol. Genet.* **23**, 4710–4720 (2014).

42. Ek, W. *et al.* Germline genetic contributions to risk for esophageal adenocarcinoma, Barretts Esophagus, and gastroesophageal reflux. *J. Nal. Cancer Inst.* **105**, 1711–1718 (2013).

43. 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).

44. Keller, M. *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in parkinson's disease. *Hum. Mol. Genet.* **21**, 4996–5009 (2012).

45. 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** (2014).

46. Lee, S. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).

47. Stahl, E. *et al.* Bayesian inference of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* **44**, 483–489 (2012).

48. Robinson, E. *et al.* The genetic architecture of pediatric cognitive abilities in the Philadelphia Neurodevelopmental Cohort. *Mol. Psychiatry* **20**, 454–458 (2015).

49. Bethesda (MD): National Library of Medicine (US), N. C. f. B. I. The ncbi handbook [internet] (2002).

50. The International Multiple Sclerosis Genetics Consortium *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).

51. Shaper, A. *et al.* British Regional Heart Study: cardiovascular risk factors in middle-aged men in 24 towns. *Brit. Med. J.* **283**, 179–186 (1981).

52. Lawlor, D., Bedford, C., Taylor, M. & Ebrahim, S. Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *57* **57**, 134–140 (2003).

53. Caerphilly, T. & Studies", S. C. H. D. Plasma triglyceride and high density lipoprotein cholesterol as predictors of ischaemic heart disease in british men. *Brit. Heart J.* **68**, 60–66 (1992).

54. Fowkes, F. *et al.* Edinburgh artery study: prevalence of asymptomatic and symptomatic peripheral arterial disease in the general population. *Int. J. Epidemiol.* **20**, 384–392 (1991).

55. Marmot, M., Banks, J., Bludell, R., Lessof, C. & Nazroo, J. *Health, wealth and lifestyles of the older population in England: ELSA 2002* (Institute for Fiscal Studies, 2002).

56. Price, J. *et al.* The Edinburgh Type 2 Diabetes Study: study protocol. *BMC Endocr. Disord.* **8** (2008).

57. Wadsworth, M., Kuh, D., Richards, M. & Hardy, R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int. J. Epidemiol.* **35**, 49–54 (2006).

58. Marmot, M. *et al.* Health inequalities among British civil servants: the Whitehall II study. *Lancet* **337**, 1387–1393 (1991).