SumHer better estimates the SNP heritability of complex traits from summary statistics. Supplementary Material

Doug Speed[1,2,3,*] and David J Balding.[3,4]
[*]Corresponding author: doug@aias.au.dk

[1]*Aarhus Institute of Advanced Studies (AIAS), Aarhus University, Denmark.*
[2]*Bioinformatics Research Centre, Aarhus University, Denmark.*
[3]*UCL Genetics Institute, University College London, United Kingdom.*
[4]*Melbourne Integrative Genomics, School of BioSciences and School of Mathematics & Statistics, University of Melbourne, Australia.*

Supplementary Table 1: Details of the 24 functional annotations

Supplementary Table 2: Details of the 25 raw GWAS, and estimates of $h^2_{\mathbf{SNP}}$ and confounding bias

Supplementary Table 3: Details of the 24 summary GWAS and estimates of confounding bias

Supplementary Table 4: Proxies for info scores

Supplementary Table 5: Comparing heritability models for the 25 raw GWAS using SumHer

Supplementary Table 6: Comparing heritability models for the 25 raw GWAS using REML

Supplementary Table 7: Comparing heritability models for the 24 summary GWAS using SumHer

Supplementary Table 8: Average estimates of functional enrichments across the 25 raw GWAS

Supplementary Table 9: Estimates of genetic correlation for the 25 raw GWAS

Supplementary Table 10: Introducing population structure for the 13 WTCCC GWAS

Supplementary Table 11: Introducing relatedness for the 12 eMERGE GWAS

Supplementary Table 12: Number of significant loci for the 24 summary GWAS after correction for confounding bias

Supplementary Table 13: Empirical support for modeling inflation due to confounding as multiplicative

Supplementary Table 14: Average estimates of functional enrichments across the 24 summary GWAS

Supplementary Table 15: Average estimates of functional enrichments across the 24 summary GWAS from the LDSC software following recommendations of Finucane *et al.*

Supplementary Table 16: Average estimates of functional enrichments across the 24 summary GWAS from the LDSC software following recommendations of Gazal *et al.*

Supplementary Table 17: Estimates of genetic correlation for the 24 summary GWAS

Supplementary Table 18: Comparing the predictive performance of Classical and Bayesian PRS

Supplementary Table 19: Average estimates of functional enrichments across the 24 summary GWAS from the hybrid model

Supplementary Table 20: Numbers of large-effect loci for the 25 raw and 24 summary GWAS

Supplementary Note: Instructions for using SumHer. Here we provide step-by-step scripts for using SumHer to estimate confounding bias, SNP heritability, enrichments of heritability and genetic correlation from GWAS results. These analyses require the software LDAK (available at www.ldak.org); we assume that the LDAK executable is saved in current folder, so that LDAK can be run by typing `./ldak5.linux` (or `./ldak5.mac` for the Mac version). They also use PLINK (available at http://cog-genomics.org/plink2) and awk (installed by default in Unix); see http://zzz.bwh.harvard.edu/plink/tutorial.shtml and www.ldak.org/awk for tutorials on each. Note that a backslash (\) at the end of a line indicates the command continues on the next line. This code is also available at www.ldak.org/protocol.

Acquire reference panel. Suppose our reference panel is stored in binary PLINK format in the files `ref.bed`, `ref.bim` and `ref.fam`. Ideally, Column 3 of `ref.bim` contains genetic distances (otherwise, replace `--window-cm X` with `--window-kb 1000X` in the scripts below). The individuals in the reference panel should be ancestrally similar to those used in the GWAS. As we analyzed summary statistics from European-centric GWAS, we used the 404 non-Finnish Europeans from the 1000 Genomes Project,[1] whose data can be obtained as follows.

```
#Download sample IDs and extract Europeans
wget ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/integrated_call_samples_v3.20130502.ALL.panel
awk < integrated_call_samples_v3.20130502.ALL.panel '($3=="EUR" && $2!="FIN"){print $1, $1}' > eur.keep

#Download data for each autosome, and convert using PLINK, extracting European individuals and SNPs with MAF>0.01
for j in {1..22}; do
wget ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/\
ALL.chr$j.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz
./plink --vcf ALL.chr$j.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz \
--make-bed --out chr$j --maf 0.01 --keep eur.keep
done

#Now join these together, excluding multi-allelic SNPs and those with duplicate positions
rm list.txt; for j in {1..22}; do echo chr$j >> list.txt; done
./ldak5.linux --make-bed all --mbfile list.txt --exclude-odd YES --exclude-dups YES

#Download and incorporate genetic distances
wget https://www.dropbox.com/s/slchsd0uyd4hii8/genetic_map_b37.zip?dl=0
unzip genetic_map_b37.zip
for j in {1..22}; do
./plink --bfile all --chr $j --cm-map genetic_map_b37/genetic_map_chr@_combined_b37.txt --make-bed --out map$j
done
cat map{1..22}.bim | awk '{print $2, $3}' > map.all
awk '(NR==FNR){arr[$1]=$2;next}{print $1, $2, arr[$2], $4, $5, $6}' map.all all.bim > ref.bim
cp all.bed ref.bed
cp all.fam ref.fam

#If these scripts were successful, you can remove the files with prefixes chr, all and map
```

Format summary statistics. For the analyses below, we use summary statistics for Alzheimer's Disease from Lambert *et al.*[2] (stored in the file `IGAP_stage_1.txt`, which can be downloaded from http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php), for years of education from Okbay *et al.*[3] (stored in the file `EduYears_Main.txt.gz`, available at http://ssgac.org) and for height from the GIANT Consortium[4] (stored in the file `GIANT_HEIGHT_Wood_et_al_2014_publicrelease_HapMapCeuFreq.txt.gz`, which can be downloaded from https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files). To use with SumHer, the summary statistic files should contain columns labeled `Predictor`, `A1`, `A2`, `Direction`, `Stat` and `n`. The names and alleles should be consistent with those in the reference panel. `Direction` indicates whether the effect of the A1 allele is positive or negative; for quantitative traits, can use the effect size, for binary traits, the log odds. `Stat` is the $\chi^2(1)$ test statistic; for quantitative traits, can use (effect/sd)^2, for binary traits (log odds/sd)^2 (else if only p-values are available, can instead provide these in a column labeled `P`). `n` is the number of individuals used when testing that SNP. The predictor names must be unique, so you should check for (and remove) duplicates.

Having processed the summary statistic files, we also want to identify SNPs within the major histocompatibility complex (Chromosome 6: 25-34 Mb), as well as SNPs which individually explain >1% of phenotypic variation, and SNPs in LD with these. If info scores are provided, we recommend excluding SNPs with score <0.95.

```
#Tidy alzheimers summary statistics - the paper tells us there were 17008 cases and 37154 controls
awk < IGAP_stage_1.txt '(NR>1){snp=$3;a1=$4;a2=$5;dir=$6;stat=($6/$7)^2;n=17008+37154}(NR==1)\
{print "Predictor A1 A2 Direction Stat n"}(NR>1 && (a1=="A"||a1=="C"||a1=="G"||a1=="T") \
&& (a2=="A"||a2=="C"||a2=="G"||a2=="T")){print snp, a1, a2, dir, stat, n}' > alz_all.txt

#Identify summary statistics consistent with the reference panel
awk '(NR==FNR){arr[$2]=$5$6;next}(FNR==1 || ($1 in arr && ($2$3==arr[$1] || $3$2==arr[$1])))' \
ref.bim alz_all.txt > alz.txt

#Check for duplicates using the Unix functions sort and uniq
#If this command results in no output, it means all SNP names are unique
awk < alz.txt '{print $1}' | sort | uniq -d | head

#If there are duplicates, we can remove them using
mv alz.txt alz_temp.txt; awk '!seen[$1]++' alz_temp.txt > alz.txt

#Tidy years education summary statistics - the paper tells us the total sample size was 328917
gunzip -c EduYears_Main.txt.gz | awk '(NR>1){snp=$1;a1=$4;a2=$5;dir=$7;stat=($7/$8)^2;n=328917}\
(NR==1){print "Predictor A1 A2 Direction Stat n"}(NR>1 && $10!="NA" && (a1=="A"||a1=="C"||a1=="G"||a1=="T") \
&& (a2=="A"||a2=="C"||a2=="G"||a2=="T")){print snp, a1, a2, dir, stat, n}' > years_all.txt

#Identify consistent summary statistics and check for duplicates (there are none here)
awk '(NR==FNR){arr[$2]=$5$6;next}(FNR==1 || ($1 in arr && ($2$3==arr[$1] || $3$2==arr[$1])))' \
ref.bim years_all.txt > years.txt
awk < years.txt '{print $1}' | sort | uniq -d | head

#Tidy height summary statistics - now per-SNP sample sizes are provided
gunzip -c GIANT_HEIGHT_Wood_et_al_2014_publicrelease_HapMapCeuFreq.txt.gz | awk '(NR>1){snp=$1;a1=$2;a2=$3;\
dir=$5;p=$7;n=$8}(NR==1){print "Predictor A1 A2 Direction P n"}(NR>1 && (a1=="A"||a1=="C"||a1=="G"||a1=="T") \
&& (a2=="A"||a2=="C"||a2=="G"||a2=="T")){print snp, a1, a2, dir, p, n}' - > height_all.txt

#Identify consistent summary statistics and check for duplicates (there are none here)
awk '(NR==FNR){arr[$2]=$5$6;next}(FNR==1 || ($1 in arr && ($2$3==arr[$1] || $3$2==arr[$1])))' \
ref.bim height_all.txt > height.txt
awk < height.txt '{print $1}' | sort | uniq -d | head

#Get list of MHC SNPs (from reference panel)
awk < ref.bim '($1==6 && $4>25000000 && $4<34000000){print $2}' > mhc.snps

#Identify large-effect SNPs (those explaining more than 1% of phenotypic variance)
#This command uses the fact that the proportion of variance explained by each SNP is stat/(stat+n)
awk < alz.txt '(NR>1 && $5>$6/99){print $1}' > alz.big
awk < years.txt '(NR>1 && $5>$6/99){print $1}' > years.big
awk < height.txt '(NR>1 && $5>$6/99){print $1}' > height.big

#Find SNPs tagging the alzheimers large-effect loci (there were no large-effect loci for years or height)
./ldak5.linux --remove-tags alz --bfile ref --top-preds alz.big --window-cm 1 --min-cor .1

#Create exclusion files, containing mhc and (for alzheimers) SNPs tagging large-effect SNPs
cat mhc.snps alz.out > alz.excl
cat mhc.snps > years.excl
cat mhc.snps > height.excl
```

Estimate SNP heritability. To estimate h^2_{SNP} there are two steps: first we compute a (1-part) tagfile which contains $q_j + \sum_{l \in N_j} q_l r^2_{jl}$ for each SNP; then we perform the regression to estimate h^2_{SNP}. To compute an LDAK tagfile, we must first compute LDAK weightings; this can take a few hours, but can be efficiently parallelized. We should use only SNPs in the reference panel for which we have summary statistics, excluding those in the MHC or tagging large-effect loci.

```
#Calculate LDAK weightings (use --extract / --exclude to filter the SNPs)
#Here, we calculate weightings separately for each chromosome then merge
#The on-screen instructions explain how to further parallelize (use --calc-weights instead of --calc-weights-all)
for j in {1..22}; do
./ldak5.linux --cut-weights alz_chr$j --bfile ref --extract alz.txt --exclude alz.excl --chr $j
./ldak5.linux --calc-weights-all alz_chr$j --bfile ref --extract alz.txt --exclude alz.excl --chr $j
done

#Merge weightings across chromosomes
cat alz_chr{1..22}/weights.short > alz.weights

#Calculate the (1-part) LDAK tagfile (no need for --extract / --exclude here as used when computing weightings)
./ldak5.linux --calc-tagging alz_ldak --bfile ref --weights alz.weights --power -.25 --window-cm 1

#Perform the regression (again no need for --extract / --exclude as used when computing weightings)
#Pay attention to the screen output (if it finds errors in the summary statistics, you must add --check-sums NO)
./ldak5.linux --sum-hers alz_ldak --tagfile alz_ldak.tagging --summary alz.txt
#Estimates of SNP heritability will be in alz_ldak.hers

#To convert to the liability scale, add the prevalence (here 0.075) and ascertainment (17008/(17008+37154)=0.314)
./ldak5.linux --sum-hers alz_ldak --tagfile alz_ldak.tagging --summary alz.txt \
--prevalence 0.075 --ascertainment 0.314
#The liability estimates have suffix .liab

#To instead assume the GCTA Model, use --ignore-weights YES and --power -1 when computing the tagfile
#This takes longer (as it uses all SNPs), but we can calculate separately for each chromosome then merge
for j in {1..22}; do
./ldak5.linux --calc-tagging alz_gcta$j --bfile ref --ignore-weights YES --power -1 --window-cm 1 \
--extract alz.txt --exclude alz.excl --chr $j
done

#Join the tagfiles across chromosomes
rm list.txt; for j in {1..22}; do echo "alz_gcta$j.tagging" >> list.txt; done
./ldak5.linux --join-tagging alz_gcta --taglist list.txt

#Then perform the regression (again pay attention to the screen output)
./ldak5.linux --sum-hers alz_gcta --tagfile alz_gcta.tagging --summary alz.txt
#Estimates of SNP heritability are in alz_gcta.hers

#Repeat these steps (calculate weightings and tagging files, then perform the regression) for years and height
```

Estimate confounding bias. To estimate how much test statistics are inflated due to confounding (or to allow for confounding inflation when estimating SNP heritability, heritability enrichments or genetic correlation), you should add either `--genomic-control YES` (to assume inflation is multiplicative) or `--intercept YES` (to assume inflation is additive) when performing the regression.

```
#Allow for multiplicative inflation
./ldak5.linux --sum-hers alz_ldak.gcon --tagfile alz_ldak.tagging --summary alz.txt --genomic-control YES
#The estimated scaling factor C is in alz_ldak.gcon.extra

#Allow for additive inflation
```

```
./ldak5.linux --sum-hers alz_ldak.cept --tagfile alz_ldak.tagging --summary alz.txt --intercept YES
#The estimated intercept 1+A is in alz_ldak.cept.extra (as per-SNP sample sizes are constant, C=1+A)

#Based on the estimate of C (1.03, SD 0.01), we conclude that there is substantial confounding bias
#Therefore we should use the estimate of SNP heritability in alz_ldak.gcon.hers
#This estimate is about a third lower than the one in alz_ldak.hers

#Repeat these steps for years and height - for years we also find confounding, but not for height
./ldak5.linux --sum-hers years_ldak.gcon --tagfile years_ldak.tagging --summary years.txt --genomic-control YES
./ldak5.linux --sum-hers height_ldak.gcon --tagfile height_ldak.tagging --summary height.txt --genomic-control YES
#Confounding bias is substantial for years (0.82, SD 0.01) but not significant for height (0.98, SD 0.02)
```

Estimate functional enrichments. Now we need to create a (multi-part) tagfile, which contains $q_j I_{jk} + \sum_{l \in N_j} q_j I_{lk} r_{jl}^2$ for each SNP and category. To work out which SNPs are in each of the 24 annotations, we use the genefiles in `annotations.zip`, available at www.ldak.org/annotations (these are modified versions of the bedfiles provided at https://data.broadinstitute.org/alkesgroup/LDSCORE).

```
#First work out which (reference panel) SNPs are in each annotation
#Will save these in ann_snps.1, ..., ann_snps.24 (the SNP lists must have consecutive suffixes)
for j in {1..24}; do
./ldak5.linux --cut-genes ann$j --bfile ref --genefile ann$j.genefile --ignore-weights YES
mv ann$j/genes.predictors.used ann_snps.$j
done

#Now compute a tagfile - as the categories overlap, we use --annotation-number and --annotation-prefix
./ldak5.linux --calc-tagging alz_ldak.ann --bfile ref --weights alz.weights --power -.25 --window-cm 1 \
--annotation-number 24 --annotation-prefix ann_snps.

#Perform the regression (our analysis above indicates we should add --genomic-control YES)
./ldak5.linux --sum-hers alz_ldak.ann --tagfile alz_ldak.ann.tagging --summary alz.txt --genomic-control YES
#The enrichment estimates are in (Column 5 of) alz_ldak.ann.enrich

#Repeat the last two steps for years (allow for inflation) and height (not necessary to allow for inflation)

#Given the three sets of enrichment, we can compute inverse-variance-weighted averages using
for i in {1..24}; do
grep "Enrich_A$i " {alz,years,height}_ldak.ann.enrich | awk -v i=$i '{a+=$5/$6^2;b+=1/$6^2}END\
{print "Annotation:", i, "Av. Enr.:", a/b, "SD:", 1/b^.5, "Num. Traits:", NR}'
done

#For the PRS (below), we must save average contributions of each annotation and the base (use the .share files)
rm props.ldak
for i in {1..24}; do
grep "Share_A$i " {alz,years,height}_ldak.ann.share | awk -v i=$i '{a+=$2/$3^2;b+=1/$3^2}END\
{print a/b, 1/b^.5}' >> props.ldak
done
grep Share_Base {alz,years,height}_ldak.ann.share | awk -v i=$i '{a+=$2/$3^2;b+=1/$3^2}END\
{print a/b, 1/b^.5}' >> props.ldak

#To instead assume the GCTA Model, use --ignore-weights YES and --power -1 when computing the tagfile
#Note that to copy Finucane et al, you must use all 52 annotations (not just the first 24)
./ldak5.linux --calc-tagging alz_ldak.ann --bfile ref --ignore-weights YES --power -1 --window-cm 1 \
--extract alz.txt --exclude alz.excl --annotation-number 24 --annotation-prefix ann_snps.

#If categories partition the genome (LDAK allows some overlap), should use --partition-number and --partition-prefix
#In this example, we construct lists part.1 and part.2 containing coding (Annotation 1) and non-coding SNPs
cp ann_snps.1 part.1
```

```
awk '(NR==FNR){arr[$1];next}!($2 in arr){print $2}' part.1 ref.bim > part.2
./ldak5.linux --calc-tagging alz_ldak.coding --bfile ref --weights alz.weights --power -.25 --window-cm 1 \
--partition-number 2 --partition-prefix part.

#Note that we can achieve the same using
./ldak5.linux --calc-tagging alz_ldak.coding --bfile ref --weights alz.weights --power -.25 --window-cm 1 \
--partition-number 1 --partition-prefix part. --background YES
```

Estimate genetic correlation. We now calculate a (1-part) tagfile, using only SNPs common to the two traits. Now the direction of effects matters, so we recommend using only non-ambiguous SNPs (alleles A & C, A & G, C & T or G & T), unless very confident the alleles are consistently aligned (note that to include ambiguous SNPs, you must add `--allow-ambiguous YES` when performing the regression).

```
#Find the overlap of non-ambiguous SNPs for alzheimers and years
awk < alz.txt '(($2=="A"&&$3=="C") || ($2=="A"&&$3=="G") || ($2=="C"&&$3=="A") || ($2=="C"&&$3=="T") || \
($2=="G"&&$3=="A") || ($2=="G"&&$3=="T") || ($2=="T"&&$3=="C") || ($2=="T"&&$3=="G")){print $1}' > alz.non
awk < years.txt '(($2=="A"&&$3=="C") || ($2=="A"&&$3=="G") || ($2=="C"&&$3=="A") || ($2=="C"&&$3=="T") || \
($2=="G"&&$3=="A") || ($2=="G"&&$3=="T") || ($2=="T"&&$3=="C") || ($2=="T"&&$3=="G")){print $1}' > years.non
awk '(NR==FNR){arr[$1];next}($1 in arr){print $1}' alz.non years.non > alz_years.non

#Will exclude SNPs in the exclusion lists for either trait
cat alz.excl years.excl | sort | uniq > alz_years.excl

#Calculate LDAK weightings for the overlap
for j in {1..22}; do
./ldak5.linux --cut-weights alz_years_chr$j --bfile ref --extract alz_years.non --exclude alz_years.excl --chr $j
./ldak5.linux --calc-weights-all alz_years_chr$j --bfile ref --extract alz_years.non --exclude alz_years.excl \
--chr $j
done
cat alz_years_chr{1..22}/weights.short > alz_years.weights

#Calculate LDAK tagging file
./ldak5.linux --calc-tagging alz_years_ldak --bfile ref --weights alz_years.weights --power -.25 --window-cm 1

#Perform the regression (by default, SumHer allows for multiplicative confounding of test statistics)
./ldak5.linux --sum-cors alz_years_ldak --tagfile alz_years_ldak.tagging --summary alz.txt --summary2 years.txt
#The estimate of genetic correlation is in alz_years_ldak.cors

#Could repeat these steps for alzheimers and height, then for years and height
```

Comparing heritability models. The easiest way to compare two heritability models is by likelihood. By default, LDAK excludes SNPs with zero weight when computing tagfiles. However, when comparing heritability models, it is important that each tagfile contain the same SNPs. Therefore, when computing an LDAK tagfile, now add `--reduce NO`.

```
#Comparing the GCTA and LDAK Models for alzheimers by likelihood

#Compute (standard) GCTA tagfile and an unreduced LDAK tagfile
#We made former earlier; for the latter, because unreduced you must now use --extract / --extract to filter the SNPs
for j in {1..22}; do
./ldak5.linux --calc-tagging alz_ldak.full$j --bfile ref --weights alz.weights --power -.25 --window-cm 1 \
--reduce NO --extract alz.txt --exclude alz.excl --chr $j
done
rm list.txt; for j in {1..22}; do echo "alz_ldak.full$j.tagging" >> list.txt; done
./ldak5.linux --join-tagging alz_ldak.full --taglist list.txt
```

```
#Perform the regressions (our earlier analysis indicated we should add --genomic-control YES)
./ldak5.linux --sum-hers alz_gcta.gcon --tagfile alz_gcta.tagging --summary alz.txt --genomic-control YES
./ldak5.linux --sum-hers alz_ldak.full.gcon --tagfile alz_ldak.full.tagging --summary alz.txt --genomic-control YES
#The likelihoods and likelihood ratio test statistics are in alz_gcta.gcon.extra and alz_ldak.full.gcon.extra

#Comparing the GCTA and LDAK Models for alzheimers by creating a merged tagfile
echo "alz_gcta.tagging
alz_ldak.full.tagging" > list.txt
./ldak5.linux --merge-tagging alz_both --taglist list.txt
./ldak5.linux --sum-hers alz_both.gcon --tagfile alz_both.tagging --summary alz.txt --genomic-control YES
#The share of heritability allocated to the LDAK Model (p) is the value Share_P2 in alz_both.gcon.share

#Could repeat these steps for years (allow for inflation) and height (not necessary to allow for inflation)
```

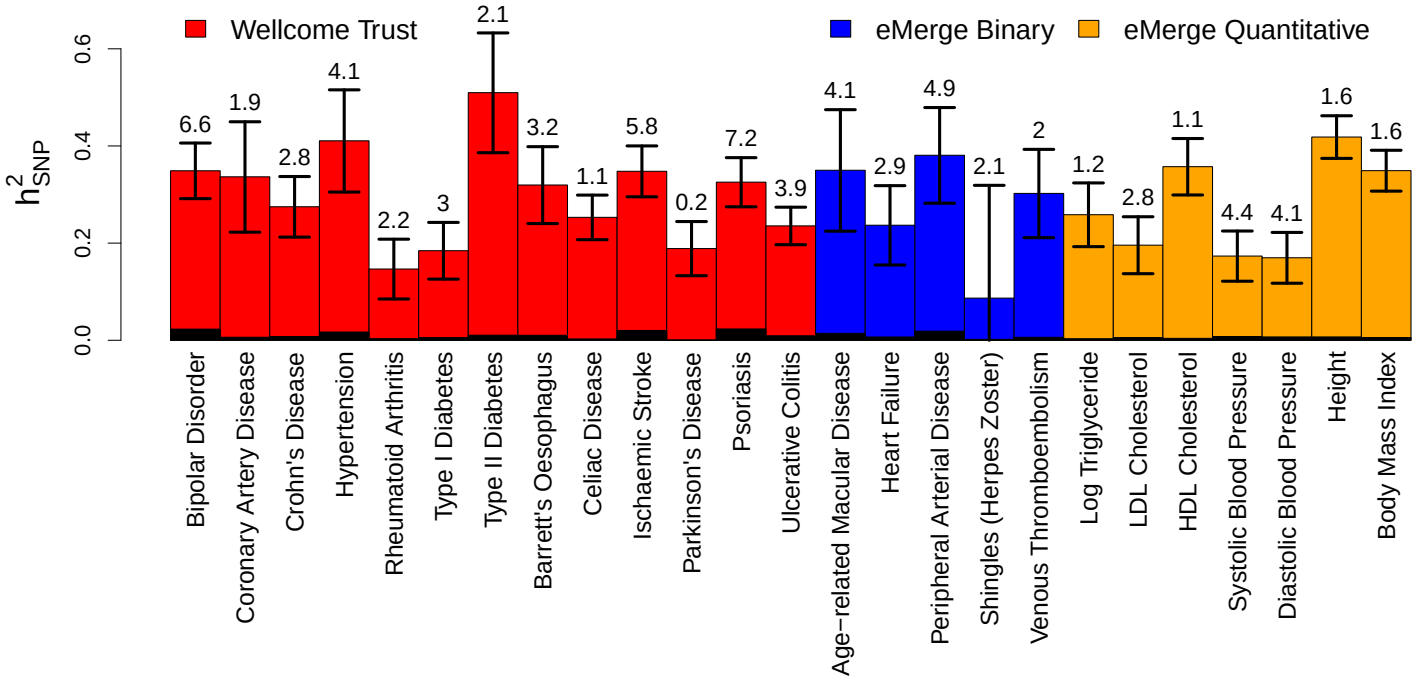
Polygenic risk scores. For each of body mass index, height, HDL & LDL cholesterol and triglyceride, we computed five PRS (one classical and four Bayesian) using the results from the 24 Summary GWAS, then measured how well each of these predicted for the eMERGE data.[5] To avoid strand issue, we reduced to non-ambiguous SNPs. To calculate a Bayesian PRS, we first compute a tagging file (either 1-part or multi-part), then the prior distribution for the heritability tagged by each SNP, and finally the posterior mean effect sizes. We found that the Bayesian PRS benefited from clumping[6,7] (identifying pairs of SNPs in high-LD, then removing the least significant) but not the Classical PRS. Here we demonstrate for height. Suppose the eMERGE data are stored in `eMERGE.bed`, `eMERGE.bim` and `eMERGE.fam`, and the corresponding phenotypes are in `height.pheno`.

```
#Identify the non-ambiguous emerge SNPs
awk < emerge.bim '(($5=="A"&&$6=="C") || ($5=="A"&&$6=="G") || ($5=="C"&&$6=="A") || ($5=="C"&&$6=="T") || \
($5=="G"&&$6=="A") ||($5=="G"&&$6=="T") || ($5=="T"&&$6=="C") || ($5=="T"&&$6=="G") ){print $2}' > emerge.non

#Each PRS scorefile should have the columns Predictor, A1, A2, Centre (can set to NA) and Effect

#For the classical PRS, the effect size of each SNP is its correlation, equal to sign(direction)*(stat/(stat+n)^.5
awk < height.txt '(NR==1){print "Predictor A1 A2 Centre Effect"}(NR>1){r=($5/($5+$6)^.5); \
if($4<0){r=-r};print $1, $2, $3, "NA", r}' > height_class.score

#To calculate the two LDAK Bayesian PRS (normal and enriched), we need unreduced 1-part and 25-part tagfiles
#We made the 1-part version above, so now make the 25-part version
for j in {1..22}; do
./ldak5.linux --calc-tagging height_ldak.ann.full$j --bfile ref --weights height.weights --power -.25 \
--window-cm 1 --reduce NO --extract height.txt --exclude height.excl \
--annotation-number 24 --annotation-prefix ann_snps. --chr $j
done
rm list.txt; for j in {1..22}; do echo "height_ldak.ann.full$j.tagging" >> list.txt; done
./ldak5.linux --join-tagging height_ldak.ann.full --taglist list.txt

#Compute the prior expected heritability tagged by each SNP
#For this we must provide an estimate of SNP heritability (here we use 0.5, but ideally get estimates from SumHer)
#For the enriched LDAK PRS, we must also provide the heritability fractions for the annotations (computed above)
./ldak5.linux --calc-exps height_ldak.full --tagfile height_ldak.full.tagging --her 0.5
./ldak5.linux --calc-exps height_ldak.ann.full --tagfile height_ldak.ann.full.tagging --her 0.5 --props props.ldak

#Next we compute the posterior means
#Will likely have to add --allow-ambiguous YES to ignore warnings (or avoid by adding --extract emerge.non)
./ldak5.linux --calc-posts height_ldak.full --expectations height_ldak.full.exps --summary height.txt \
--allow-ambiguous YES
./ldak5.linux --calc-posts height_ldak.ann.full --expectations height_ldak.ann.full.exps --summary height.txt \
--allow-ambiguous YES
```

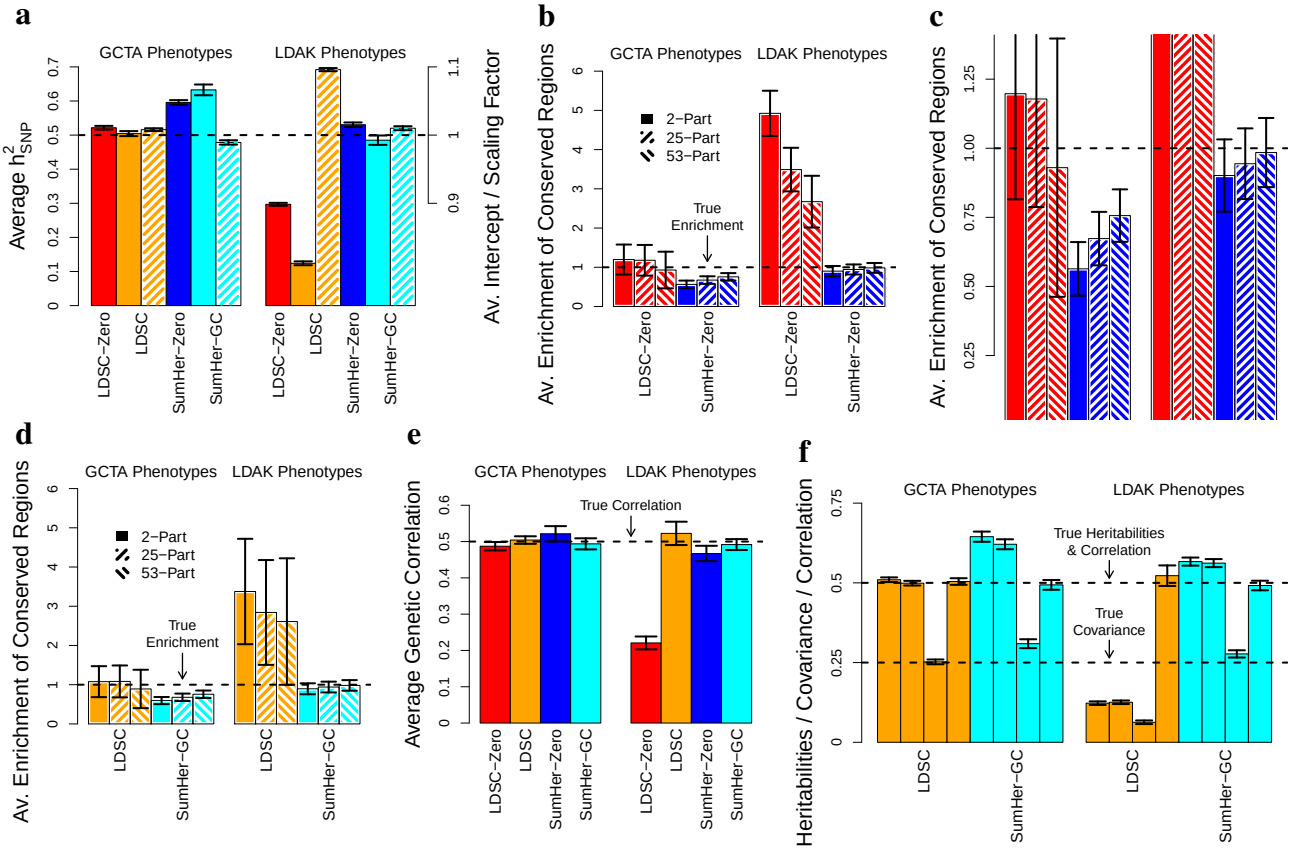
```
#Finally, we clump (we add --extract emerge.non, as we will only use these SNPs when predicting
#Normally when clumping we provide pvalues; here we instead provide "psuedos" (which rank SNPs by Bayes factors)
./ldak5.linux --thin height_ldak.full --bfile ref --extract emerge.non --window-cm 1 --window-prune .5 \
--pvalues height_ldak.full.psuedos
./ldak5.linux --thin height_ldak.ann.full --bfile ref --extract emerge.non --window-cm 1 --window-prune .5 \
--pvalues height_ldak.ann.full.psuedos

#To test each PRS, we project onto the emerge data
#All PRS contain standardized effect sizes, so we use --power -1 (when providing raw effect sizes, use --power 0)
#It is convenient to provide the phenotype (for our analysis, we first regressed on sex and population PCs)
#Read the screen output (as we have excluded ambiguous SNPs, it is ok to allow flips)
./ldak5.linux --calc-scores height_class --bfile emerge --extract emerge.non \
--scorefile height_class.score --power -1 --pheno height.pheno --allow-flips YES
./ldak5.linux --calc-scores height_ldak.full --bfile emerge --extract height_ldak.full.in \
--scorefile height_ldak.full.score --power -1 --pheno height.pheno --allow-flips YES
./ldak5.linux --calc-scores height_ldak.ann.full --bfile emerge --extract height_ldak.ann.full.in \
--scorefile height_ldak.ann.full.score --power -1 --pheno height.pheno --allow-flips YES
#Predicted and observed phenotypes stored in height_class.profile, height_ldak.full.profile
#and height_ldak.ann.full.profile

#We measured performance by calculating rho, the correlation between predicted and observed phenotypes
#We estimated the SD of rho by block-jackknifing 200 times for each PRS; for each jackknife we computed rho_z
#the correlation across 99.5% of individuals, then the estimate of the SD of rho is SD(rho_z)*199^.5
```

Additional scripts. Our previous paper[8] provided scripts for estimating inflation due to confounding via REML and for performing association analysis, including how to construct covariates (when performing linear regression, we included sex and ten principal components; five derived from the reference panel, five from the 1000 Genomes Project[1]). To simulate phenotypes for the eMERGE data under the GCTA and LDAK Models, we used the second of the following two scripts (`emerge.weights` contains LDAK weightings from the emerge data).

```
#Simulating 100 phenotypes under GCTA and LDAK Models (each with SNP heritability 0.5 and 2000 causal SNPs)
./ldak5.linux --make-phenos gcta --bfile emerge --ignore-weights YES --power -1 --num-causals 2000 \
--num-phenos 100
./ldak5.linux --make-phenos ldak --bfile emerge --weights emerge.weights --power -.25 --num-causals 2000 \
--num-phenos 100

#Simulating 50 pairs of phenotypes under GCTA and LDAK Models so that consecutive pairs have genetic correlation 0.5
./ldak5.linux --make-phenos gcta_bivar --bfile emerge --ignore-weights YES --power -1 --num-causals 2000 \
--num-phenos 50 --bivar .5
./ldak5.linux --make-phenos ldak_bivar --bfile --weights emerge.weights --power -.25 --num-causals 2000 \
--num-phenos 50 --bivar .5
```



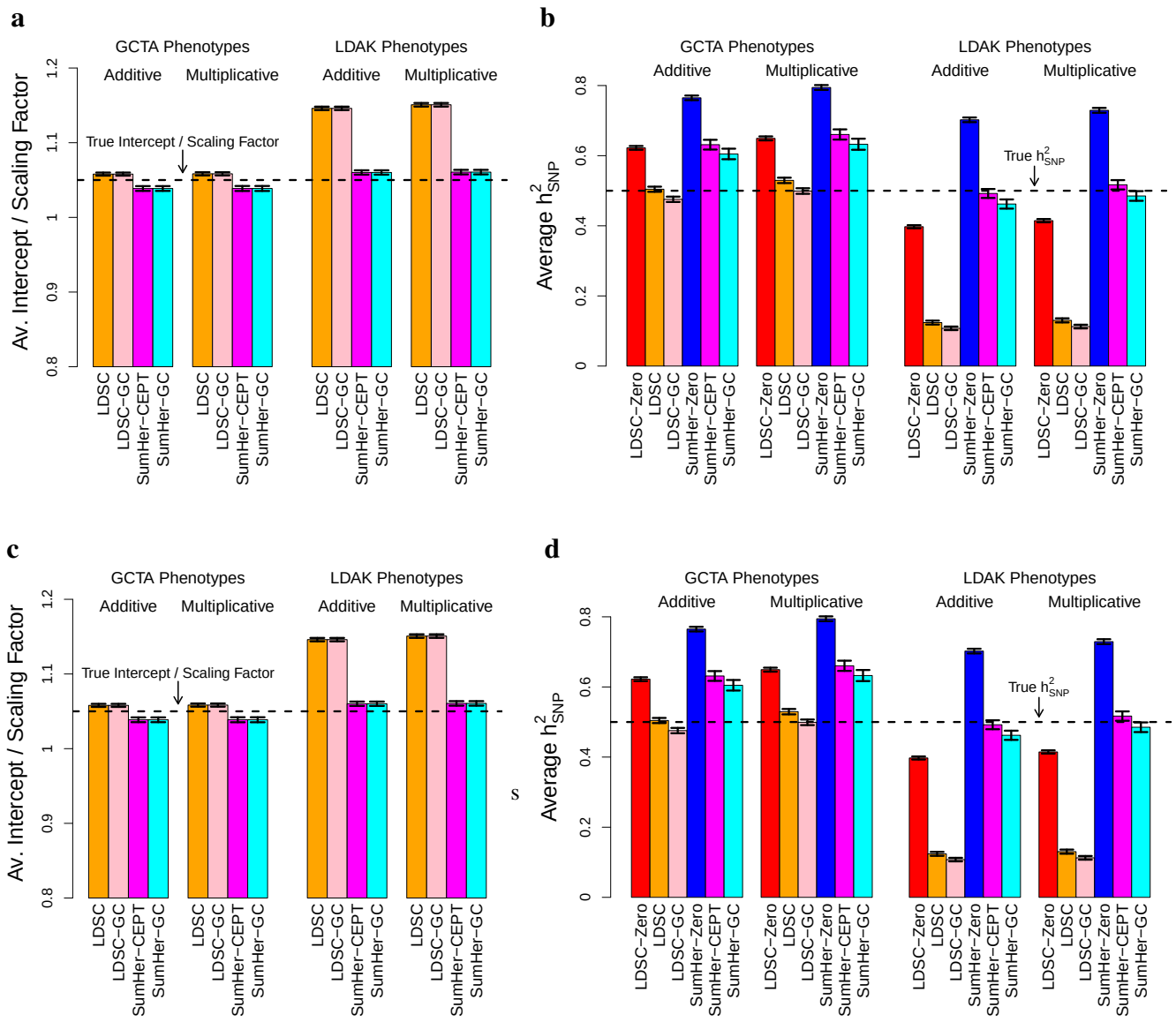
Supplementary Figure 1: REML estimates of confounding for the 25 raw GWAS. For each trait, we compute K_A, K_B, K_C and K_D, kinship matrices from Chromosomes 1-3, Chromosomes 4-7, Chromosomes 8-12 and Chromosomes 13-22, respectively, each time assuming the LDAK Model.[9] Then, using REML, we estimate h^2_A, h^2_B, h^2_C and h^2_D, by regressing the phenotype separately on each of K_A, K_B, K_C and K_D, and h^2_{ALL}, by regressing the phenotype jointly on K_A, K_B, K_C and K_D. Our estimate of inflation due to population structure and cryptic relatedness is then $\frac{1}{3}(h^2_A + h^2_B + h^2_C + h^2_D - h^2_{ALL})/h^2_{ALL}$. Bars report estimates of h^2_{ALL} (for binary traits, estimates are on the liability scale); the vertical line segments mark 95% confidence intervals. The black rectangle at the base of each bar indicates the estimated inflation, while the number above the bar expresses this as a percentage of h^2_{ALL}.



Supplementary Figure 2: Simulations demonstrating the importance of the assumed heritability model. We use genotypes from the eMERGE Network.[5] After the quality control described in Supplementary Figure 19, and excluding SNPs within the major histocompatibility complex (Chr 6: 25-34 Mb) and those not present in our reference panel, there are 25 875 individuals and 4 555 718 autosomal SNPs with MAF$\geq$0.01. We generate 200 phenotypes, each with 2000 causal SNPs (picked uniformly at random), $h^2_{\text{SNP}} = 0.5$ and no confounding bias ($C = 1 + A = 1$), using the model $Y = \sum_j \beta_j X_j + e$, where β_j is the effect size of X_j, the jth causal SNP, and e is Gaussian-distributed noise. For the first 100 phenotypes, we sample effect sizes according to the GCTA Model ($\beta_j \sim \mathbb{N}(0, [f_j(1-f_j)]^{-1})$, where f_j is the MAF of X_j), for the second 100 phenotypes, we sample effect sizes according to the LDAK Model ($\beta_j \sim \mathbb{N}(0, [f_j(1-f_j)]^{-0.25} w_j)$, where w_j is the LDAK weighting for X_j). **(a)** Solid bars (scale on left) report average estimates of h^2_{SNP} from each of LDSC-Zero, LDSC, SumHer-Zero and SumHer-GC, while hatched bars (scale on right) report estimates of confounding bias from LDSC and SumHer-GC. **(b)** Average estimates of enrichment of SNPs in conserved regions (one of the 24 functional categories, see Supplementary Table 1) from LDSC-Zero and SumHer-Zero. For each method, we estimate enrichment using three approaches: a 2-part model (conserved and non-conserved SNPs), a 25-part model (one for each of the 24 functional categories, plus one for all SNPs), and a 53-part model (24 categories, 28 buffer regions and all SNPs, the approach preferred by Finucane *et al.*[10]). **(c)** Zoomed-in version of (b). **(d)** Same as (b), except now we estimate enrichment using LDSC and SumHer-GC

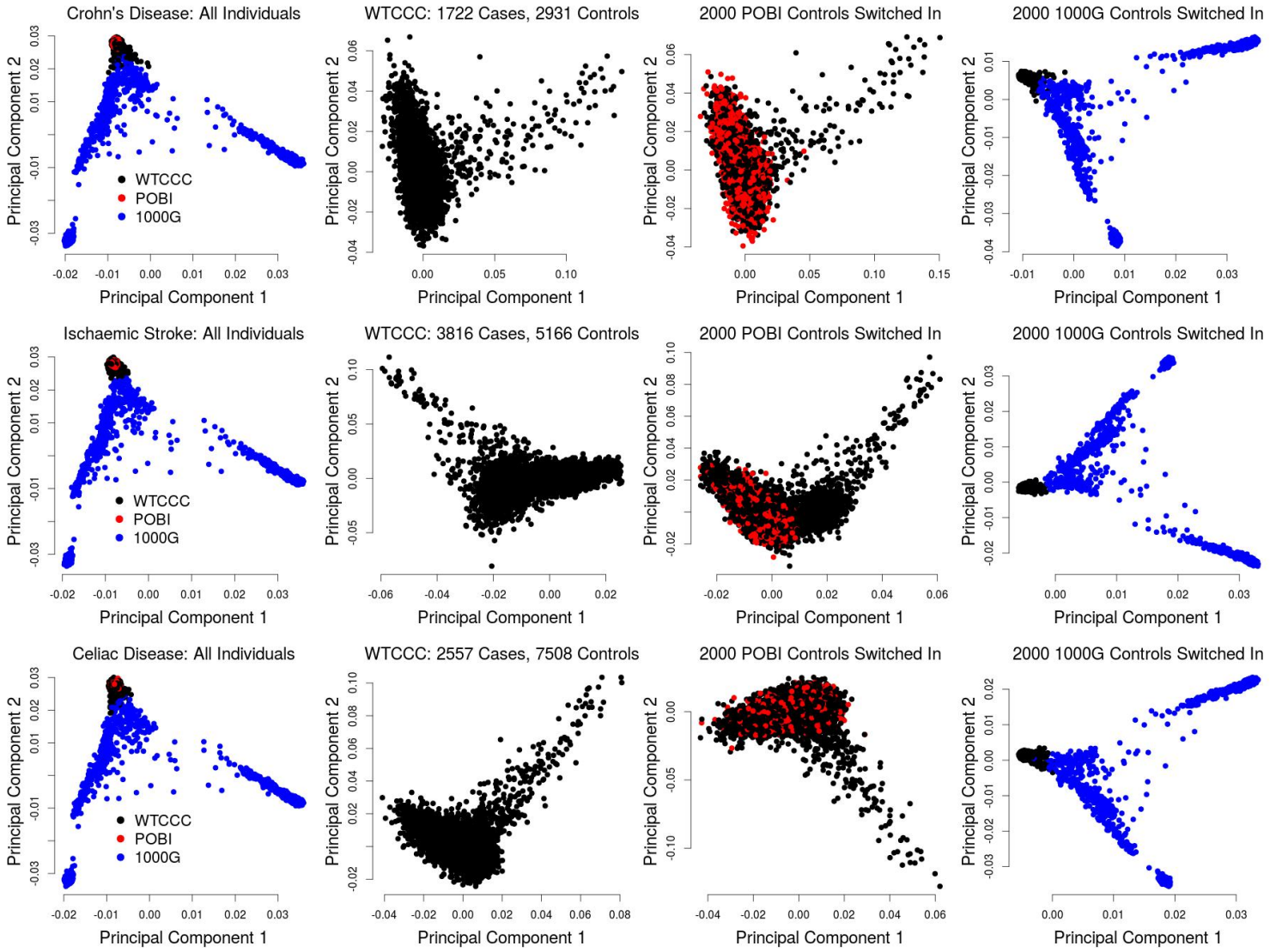
When generating the phenotypes, we ensured that consecutive pairs (Phenotypes 1 & 2, 3 & 4, etc.) have genetic correlation 0.5. **(e)** Average estimates of genetic correlation from LDSC-Zero, LDSC, SumHer-Zero and SumHer-GC. We see that by allowing for confounding bias (i.e., using LDSC or SumHer-GC), it is possible to obtain accurate estimates of genetic correlation regardless of which heritability model is assumed. **(f)** Average estimates from LDSC and SumHer-GC of h^2_A & h^2_B (the SNP heritability of the first and second phenotype in each pair), h^2_{AB} (their genetic covariance) and $h^2_{AB}/\sqrt{h^2_A h^2_B}$ (their genetic correlation). We see that even though assuming a poor heritability model causes over/under-estimation of the SNP heritabilities, it also causes over/under-estimation of the genetic covariance by approximately the same amount, resulting in accurate estimates of genetic correlation.

It is worth noting that both the GCTA and LDAK Models assume every SNP tags some heritability, whereas it is likely that for many traits, a substantial fraction of the genome tags no heritability. However, these simulations, for which we generated phenotypes with a relatively small number (2 000) of causals SNPs, demonstrate that it is possible to obtain accurate estimates despite this assumption. For a more detailed look at this and other fundamental assumptions, see the paper in which we proposed LDAK.[11]

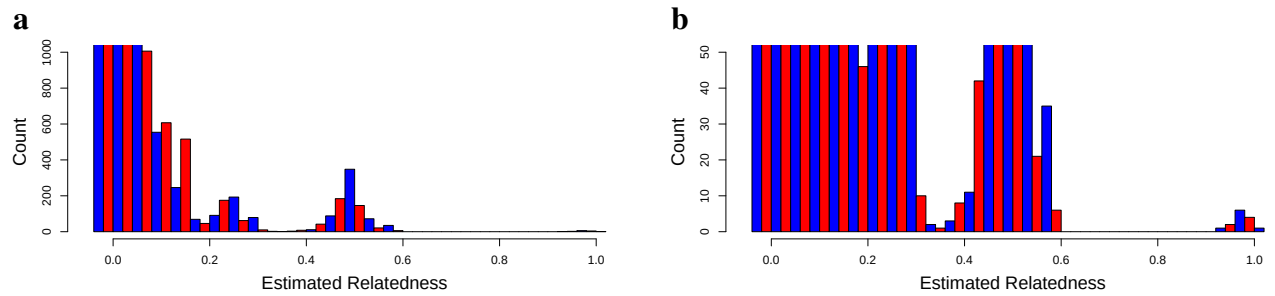


Supplementary Figure 3: Simulations investigating the additive and multiplicative models for inflation due to confounding. We now construct summary statistics where there is confounding bias. Using the results of single-SNP analysis for each of the 200 phenotypes generated for Supplementary Figure 2, we introduce either additive or multiplicative inflation: for the former, we add 0.05 to each test statistic; for the latter we multiply each test statistic by 1.05 (Supplementary Tables 10 & 11 indicate that, for a GWAS of this size, there would have to be substantial confounding to inflate test statistics by 5%). **(a)** Average estimates of the intercept, $1 + A$, from LDSC and SumHer-CEPT, and of the scaling factor, C, from LDSC-GC and SumHer-GC (LDSC-GC assumes the GCTA Model and multiplicative inflation, while SumHer-CEPT assumes the LDAK Model and additive inflation). **(b)** Average estimates of h^2_{SNP} from LDSC-Zero, LDSC, LDSC-GC, SumHer-Zero, SumHer-CEPT and SumHer-GC. **(c)** and **(d)** We find no observable difference in results if instead of inflating test statistics by a constant, we either add on $0.1a$ or multiply by $1 + 0.1a$, where $a \sim \text{Uniform}(0, 1)$.

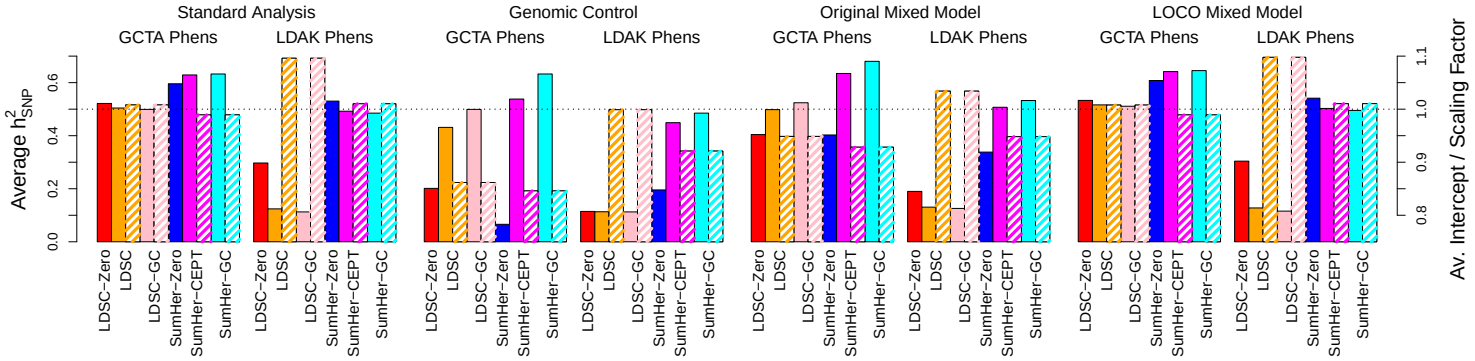
These simulations demonstrate that when per-SNP sample sizes are constant, the choice of modeling inflation as additive or multiplicative has no impact on the estimate of confounding. The choice will impact estimates of h^2_{SNP}; specifically, the estimate when allowing for additive inflation will be $C = 1 + A$ times the estimate when allowing for multiplicative inflation. Our analyses of real data indicate that for a carefully-performed GWAS, C will be close to one; for example, only five of the 24 summary GWAS had estimated inflation $>10\%$ (Supplementary Table 3), and for each of these, the bias was negative (C<1), indicating that it was predominantly due to genomic control. Therefore, the impact on estimates of h^2_{SNP} of changing the confounding model will in general be slight compared to the impact of changing the heritability model.



Supplementary Figure 4: Introducing population structure for the 13 WTCCC GWAS. For each GWAS, we replace 2 000 randomly picked controls with either 2 000 unrelated individuals from People of the British Isles[12] (POBI) or 2 000 non-European individuals from the 1000 Genome Project[1] (1000G). These principal component plots demonstrate the impact of switching controls for Crohn's Disease (top row), Ischaemic Stroke (middle row) and Celiac Disease (bottom row), the smallest and two largest GWAS, respectively. Points are colored according to collection (black: WTCCC, red: POBI, blue: 1000G). Although both the WTCCC controls and POBI individuals were recruited from the UK, switching in the latter generates modest structure, because the POBI individuals came from predominantly isolated, rural regions (observe that the red points in Column 3 are not uniformly spread across the black points). Switching in 1000G individuals generates extreme structure (Column 4), because these are Admixed Americans, Africans and Asians, whereas the WTCCC individuals are Caucasian. In practice, we would expect population structure to be modest, rather than extreme, as when performing a GWAS it is standard practice to identify and exclude ancestral outliers.[13]

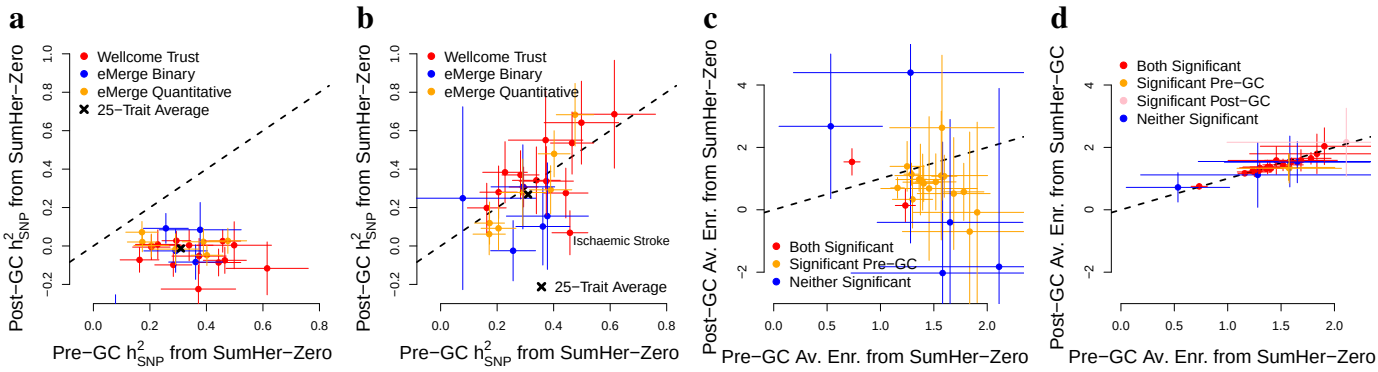


Supplementary Figure 5: Introducing relatedness for the 12 eMERGE GWAS. The eMERGE data contain 28 803 Caucasian individuals. For our main analyses, we restricted to 25 875 unrelated individuals (obtained by filtering so that no pair of individuals remained with allelic correlation[14] >0.05). In order to generate familial relatedness, we repeat the analysis twice, first restricting to 27 575 individuals (obtained by filtering so that no pair of individuals remains with allelic correlation >0.175), then using all 28 803 individuals. These two histograms, **(b)** is a zoomed-in version of **(a)**, summarize the allelic correlation for the ${}^{28803}C_2$ pairs of individuals; they show that using all 28 803 individuals results in the inclusion of approximately 14 identical twins, 950 full-sibs and 700 half-sibs.

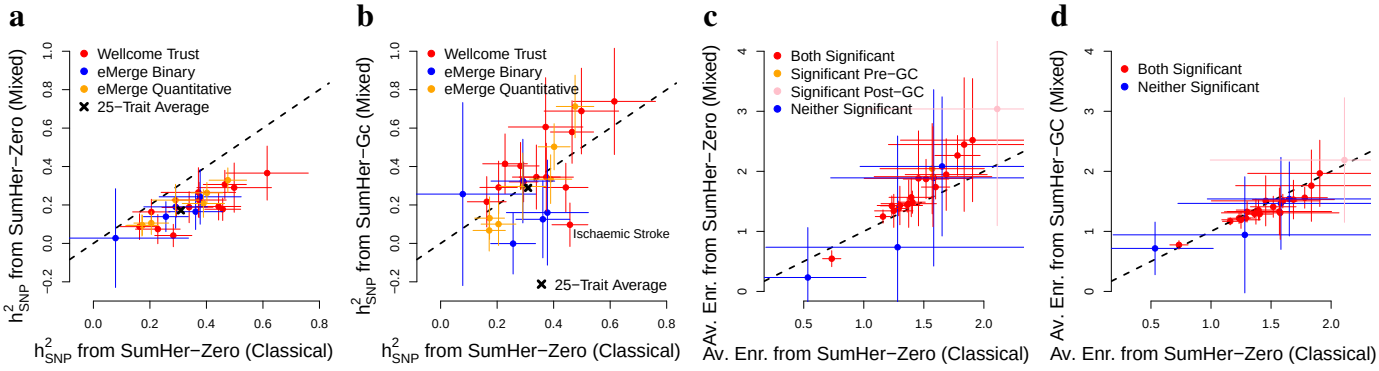


Supplementary Figure 6: Simulations investigating the impact of genomic control and mixed-model association analysis. For each of the 200 simulated phenotypes described in Supplementary Figure 2 ($h^2_{\text{SNP}} = 0.5$ and no confounding bias), we perform four types of association analysis: **Standard Analysis** tests each SNP using classical (least-squares) linear regression; **Genomic control** also performs classical linear regression, but then divides test statistics by the genomic inflation factor; **Original Mixed Model** tests each SNP using mixed-model association analysis[15] (to generate the kinship matrix, we thinned SNPs then computed allelic correlations[14]); **LOCO Mixed Model** also uses mixed-model association analysis, but when testing each SNP, excludes from the kinship matrix all SNPs on the same chromosome (for this we use Bolt-LMM[16,17]). Solid bars (scale on left) report estimates of h^2_{SNP} from LDSC-Zero, LDSC, LDSC-GC, SumHer-Zero SumHer-CEPT and SumHer-GC, while hatched bars (scale on right) report estimate of confounding bias from LDSC, LDSC-GC, SumHer-CEPT and SumHer-GC (LDSC-GC assumes the GCTA Model and that inflation of test statistics due to confounding is multiplicative, while SumHer-CEPT assumes the LDAK Model and additive inflation). Note that because per-SNP effect sizes are constant, estimates of confounding bias from LDSC and SumHer-CEPT match those from LDSC-GC and SumHer-GC, respectively (Supplementary Figure 3).

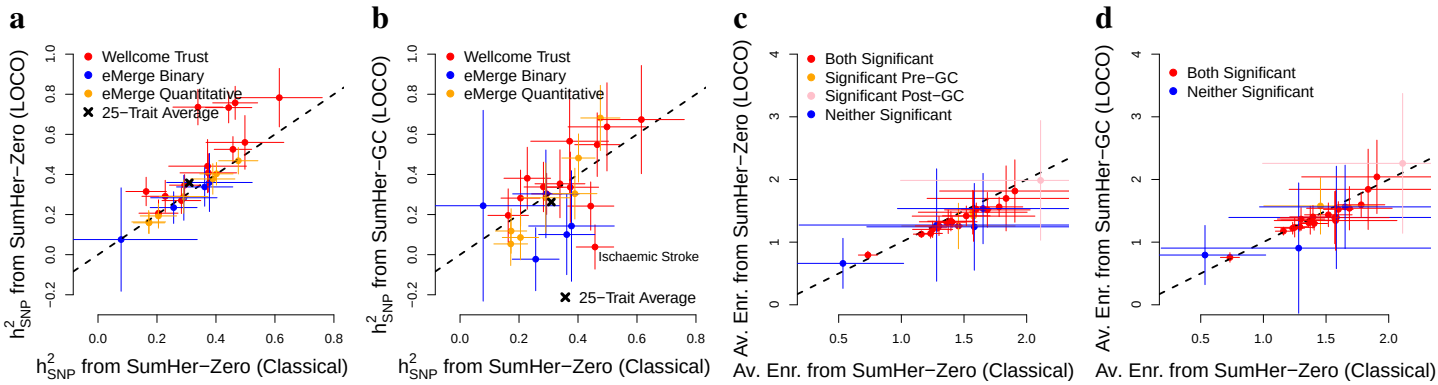
If we focus on the LDAK Phenotypes, as these best reflect real data, we see that when test statistics have been subjected to genomic control or obtained from ordinary mixed-model analysis, SumHer-Zero will substantially under-estimate h^2_{SNP}, but that reliable estimates can be obtained by using instead SumHer-GC. For mixed-model analysis, the under-estimation occurs because when testing each SNP for association, some of its variance explained is attributed to the kinship matrix (proximal contamination[18]). However, this can be avoided by using leave-one-chromosome-out (LOCO) analysis.[16] We note that SumHer-CEPT performs as well as SumHer-GC when applied to test statistics from ordinary mixed-model analysis, suggesting the deflation of test statistics is a mixture of additive and multiplicative. These simulations explain why we view genomic control as a source of confounding, whose impact is often larger than that due to, say, population structure or relatedness. They also emphasize that when genomic control has been performed, the accuracy of SumHer-GC estimates depends on the appropriateness of the LDAK Model. In Supplementary Figure 7, we see that most of the estimates from SumHer-GC post genomic control were consistent with those from SumHer-Zero pre genomic control, reflecting that in general the LDAK Model performs well; however, the inconsistent results for Ischaemic Stroke remind that the LDAK Model will be sub-optimal for some traits.



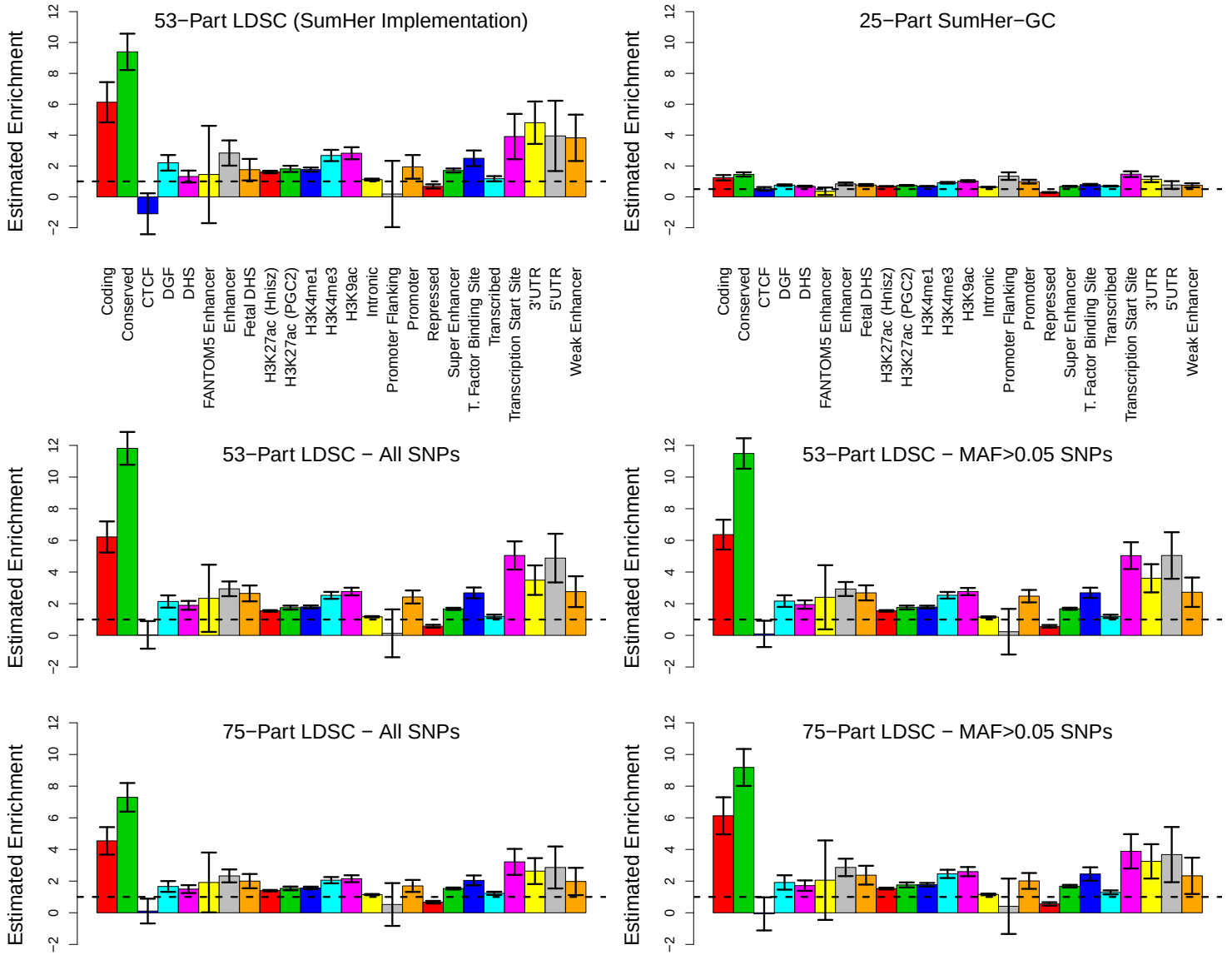
Supplementary Figure 7: Performing genomic control for the 25 raw GWAS. Each plot compares estimates from SumHer using raw test statistics (x-axis) with those using test statistics subjected to genomic control (y-axis). **(a)** Estimates of h^2_{SNP} using SumHer-Zero both pre and post genomic control. **(b)** Estimates of h^2_{SNP} using SumHer-Zero pre genomic control and SumHer-GC post genomic control. **(c)** Average estimates of enrichments for the 24 functional categories using SumHer-Zero both pre and post genomic control. **(d)** Average estimates of enrichments for the 24 functional categories using SumHer-Zero pre genomic control and SumHer-GC post genomic control. In the first two plots, estimates for binary traits are on the observed scale, while colors distinguish between the 13 WTCCC, the 5 binary eMERGE and 7 quantitative eMERGE traits (black denotes the 25-trait average). In the last two plots, both SumHer-Zero and SumHer-GC estimates of enrichments were obtained using a 25-part model, while colors indicate significant enrichments ($P < 0.05$) from one or both methods. In all plots, horizontal and vertical line segments mark 95% confidence intervals. Genomic control divides test statistics by the genomic inflation factor (GIF).[19] As the GIF tends to over-estimate confounding bias,[20] performing genomic control will generally result in deflated test statistics. We find that if not accounted for (i.e., using SumHer-Zero, which assumes test statistics are not confounded), genomic control will result in under-estimation of h^2_{SNP} and inaccurate estimates of heritability enrichments, but that reliable estimates can be obtained by allowing for multiplicative inflation of test statistics (i.e., using SumHer-GC).



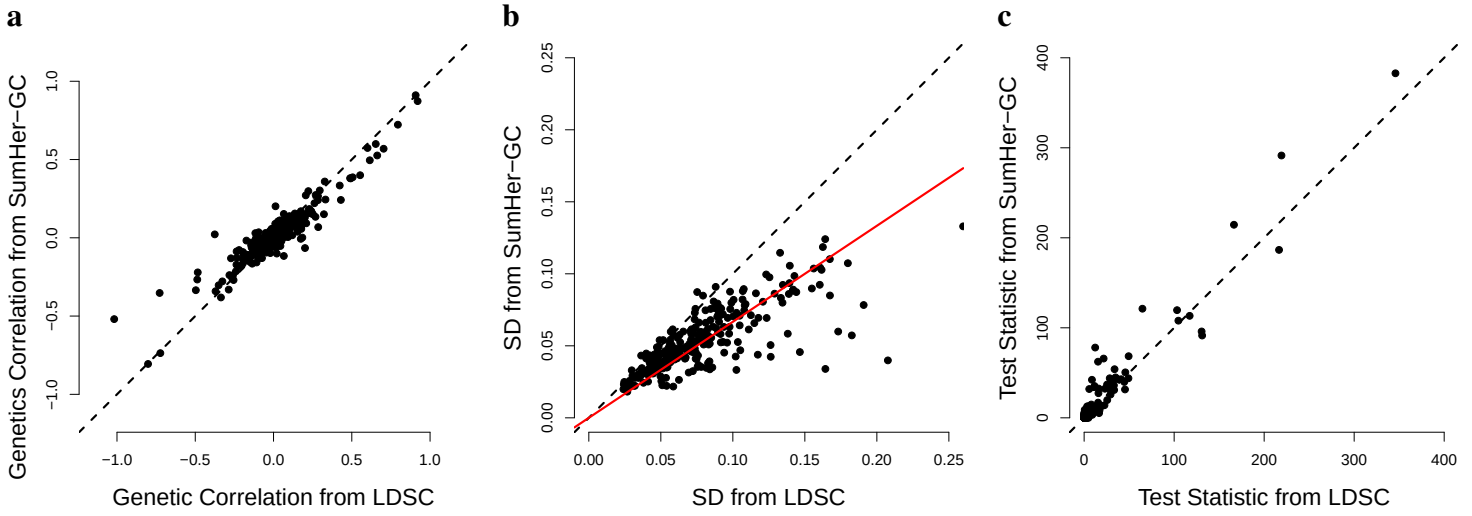
Supplementary Figure 8: Mixed-model association analysis for the 25 raw GWAS. Details are same as Supplementary Figure 7, except now plots compare estimates from SumHer using test statistics from classical linear regression (x-axis) with those using test statistics from mixed-model association analysis (y-axis). To generate the kinship matrix required for the latter, we thinned SNPs then computed allelic correlations.[14] Supplementary Figure 7 showed how genomic control resulted in estimates of h^2_{SNP} centered on zero and less accurate estimates of enrichments. Here we see that mixed-model association analysis has a similar, albeit less severe, impact, but that accuracy can be improved by using SumHer-GC instead of SumHer-Zero. For these analyses, we used ordinary mixed-model association analysis, where no allowance is made for proximal contamination;[18] see Supplementary Figure 9 for results using instead leave-one-chromosome-out mixed-model association analysis.[16]



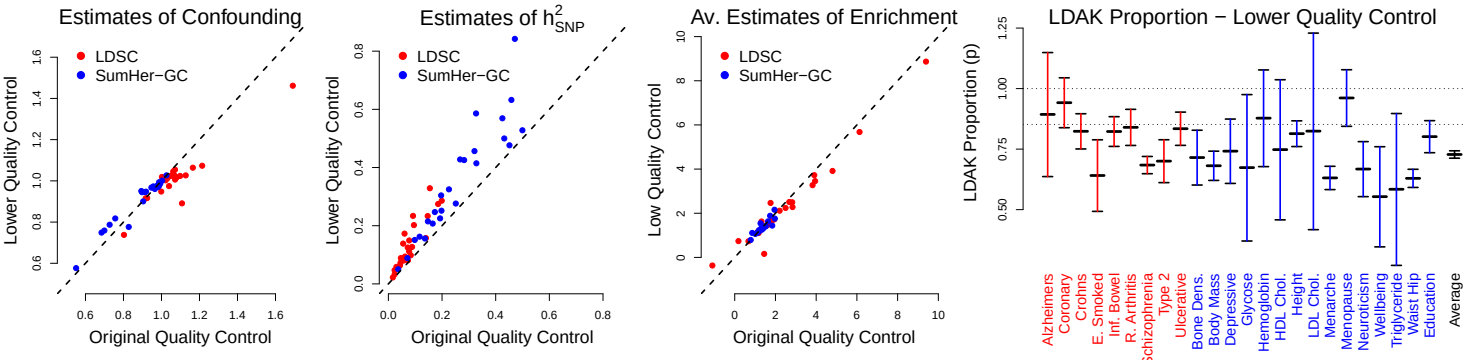
Supplementary Figure 9: Leave-one-chromosome-out mixed-model association analysis for the 25 raw GWAS. Details are same as Supplementary Figure 7, except now plots compare estimates from SumHer using test statistics from classical linear regression (x-axis) with those using test statistics from Bolt-LMM[17] (y-axis). When testing each SNP, Bolt-LMMM avoids proximal contamination by excluding from the kinship matrix all SNPs on the same chromosome.[16] Whereas ordinary mixed-model association analysis results in deflated estimates of h^2_{SNP} (Supplementary Figure 8), this is no longer the case with leave-one-chromosome-out mixed-model association analysis (LOCO), and as such there appears limited advantage to using SumHer-GC instead of SumHer-Zero. However, we appreciate that the use of summary statistics from mixed-model (rather than classical) regression requires closer investigation. In particular, it is recognized that with sufficient sample size, mixed-model association analysis can increase detection power compared to classical association analysis[16] (this is because the inclusion of a random effect "soaks up" the polygenic contribution of SNPs, resulting in a lower residual variance). This increase in power might be the reason why in Panel (a) the average estimate of h^2_{SNP} from SumHer-Zero is slightly above the diagonal (to be expected if test statistics for associated SNPs from mixed-model association analysis are slightly higher than those from classical regression), and we believe that for much larger sample sizes (e.g., upwards of 200 000 individuals), the inflation could be substantial.



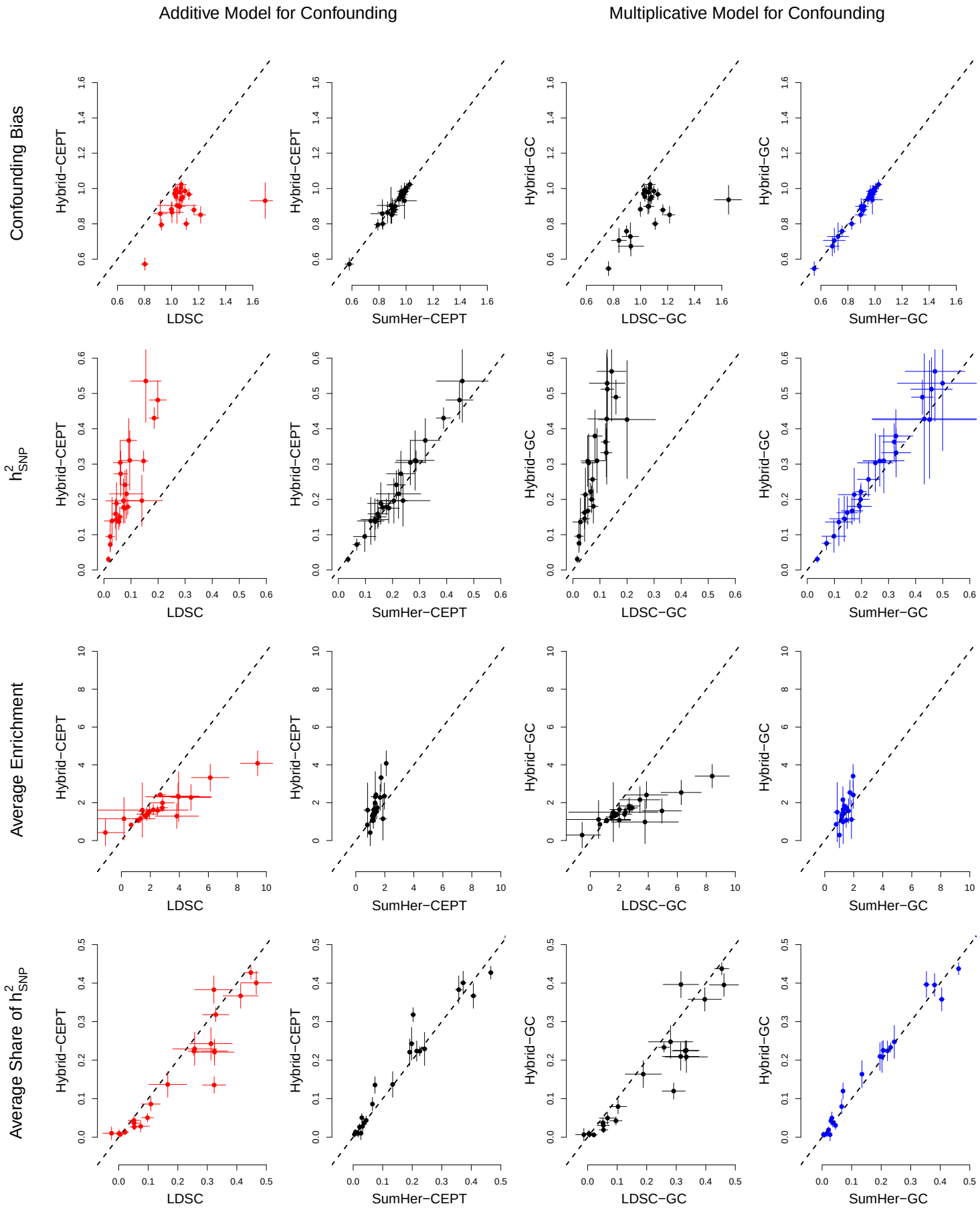
Supplementary Figure 10: Average estimates of functional enrichments across the 24 summary GWAS from the LDSC software. Bars report average estimates of enrichments for each of the 24 functional annotations; vertical line segments mark 95% confidence intervals. Numerical values are provided in Supplementary Tables 14, 15 & 16. The top row presents results from the SumHer implementation of LDSC, using a 53-part model, and from SumHer-GC, using a 25-part model. These are the same results presented in Figure 3a in the Main Text. When running SumHer, we ensure that the SNPs in the reference panel match those used for the regression (we achieve this by restricting to SNPs which are both present in the 1000 Genome Project[1] and for which we have summary statistics). By contrast, the authors of LDSC[21] recommend that the reference panel contains as many SNPs as possible, but that the regression should use only HapMap 3 SNPs[22] with MAF $\geq$0.05. To examine the impact this has, we now estimate enrichments using the LDSC software,[21] following the recommendations on https://github.com/bulik/ldsc (in particular, we use the reference panel they supply, which comprises 489 European individuals from the 1000 Genomes Project,[1] and also use the precomputed LD scores). For the middle row, we use the approach of Finucane *et al.*,[10] (who used the 53-part model); for the bottom row we use the approach of Gazal *et al.*,[23] who instead used a 75-part model (to the 53 categories, they add 3 more functional annotations, 3 extra buffers, 10 MAF tranches and 6 continuous LD-related annotations). By adding the option `--not-M-5-50`, LDSC will perform the regression using HapMap 3 SNPs with MAF $\geq$0.01, rather than just those with MAF $\geq$0.05. Regardless of settings, we see that estimates from the LDSC software are very similar to those from the SumHer implementation of LDSC, and very different from SumHer-GC estimates.



Supplementary Figure 11: Estimates of genetic correlation for the 24 summary GWAS. We analyze each of the 276 pairs of traits using LDSC (x-axis) and SumHer-GC (y-axis). **(a)** Estimates of genetic correlation. **(b)** Standard deviations of these estimates; the red line marks $y = 2x/3$. **(c)** $\chi^2(1)$ test statistics (the estimate divided by its s.d, all squared); two extreme test statistics are not shown, those for Crohn's Disease & Inflammatory Bowel Disease (x=1434,y=1418) and Ulcerative Colitis & Inflammatory Bowel Disease (x=1153,y=2517). We see that while LDSC and SumHer-GC estimates of genetic correlation are highly concordant (correlation 0.94), the latter are on average a third more precise, resulting in higher test statistics.



Supplementary Figure 12: Impact of introducing poorly-genotyped SNPs for the 24 summary GWAS. When analyzing Crohn's Disease, inflammatory bowel disease, schizophrenia and ulcerative colitis, we excluded SNPs with info score <0.95; for the remaining 20 traits, for which info scores were not available, we restricted to SNPs present in the eMERGE data (Supplementary Table 4). Now we omit this filtering. The first three plots show that reducing quality control does not change the main conclusions; it remains that SumHer-GC gives substantially lower estimates of confounding than LDSC, higher estimates of h^2_{SNP} and more modest estimates of enrichments. However, we note that reducing quality control reduces the estimate of p, the LDAK Proportion in the hybrid model, from 0.85 (marked by the lower horizontal line) to 0.73 (SD 0.01). This is to be expected; there will be a correlation between the heritability tagged by a SNP and its genotyping certainty (a SNP genotyped with error will tag less causal variation than were it perfectly typed),[8] and similarly, there will be a correlation between genotyping certainty and local levels of LD (low-LD regions tend to contain more low-MAF SNPs, which are often hard to genotype reliably, while imputation is easier in high-LD regions). Therefore, including lower-certainty SNPs in a GWAS will generate correlation between the heritability tagged by each SNP and levels of LD. This will reduce the fit of the LDAK Model (whose core assumption is that the amount of heritability tagged by each SNP is independent of levels of LD[11]), and improve the fit of the GCTA Model (under which the amount of heritability tagged by each SNP is assumed to be proportional to levels of LD[21]), and as such, traits will appear more "GCTA-like".

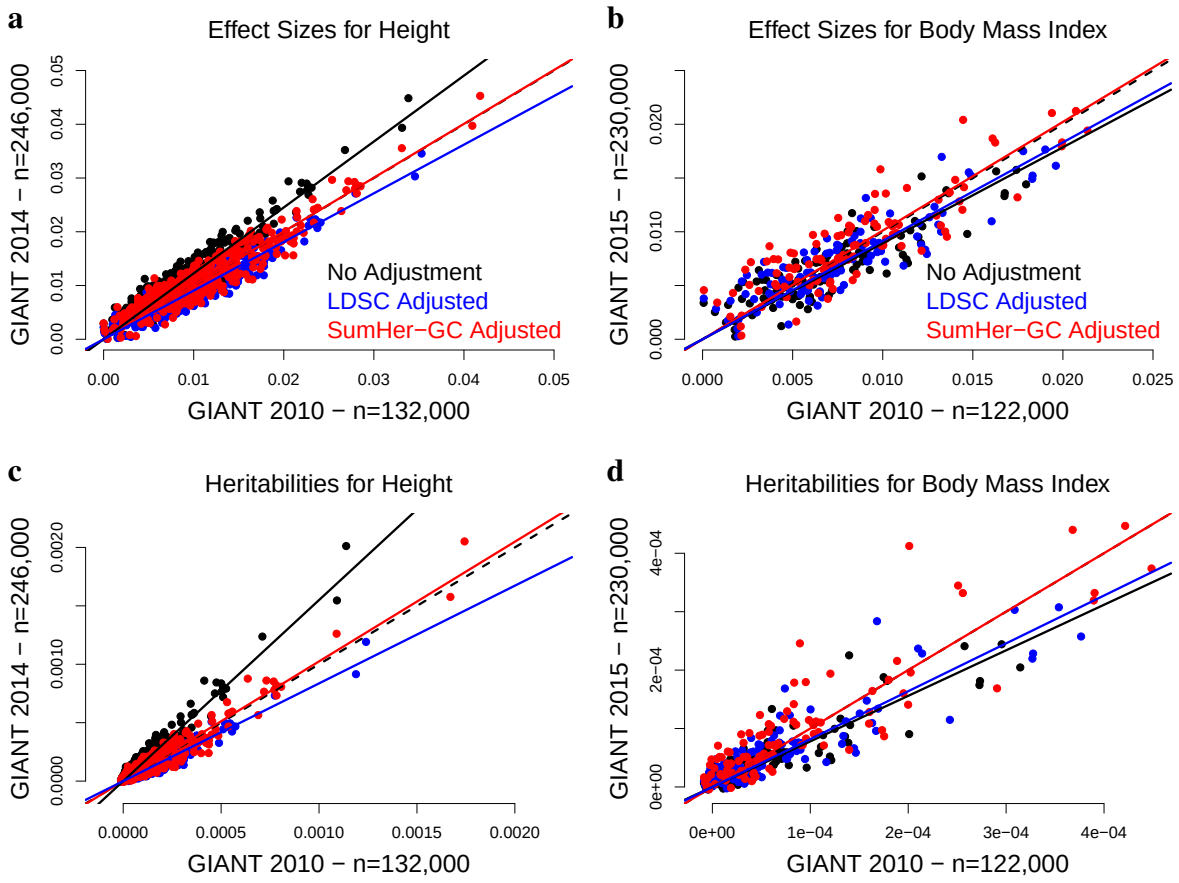


Supplementary Figure 13: Is it beneficial to use a hybrid model? Caption on next page.

Supplementary Figure 13. Is it beneficial to use a hybrid model? Continued from previous page. In order to compare the fit of the GCTA and LDAK Models, we ran SumHer using a hybrid heritability model where the fractions $1-p$ and p indicate the proportions of GCTA and LDAK, respectively. This heritability model can also be used to obtain estimates of confounding bias, h^2_{SNP}, heritability enrichments and genetic correlation. Here we compare estimates for the 24 summary GWAS obtained using the hybrid model to those obtained using only the GCTA Model or only the LDAK Model. Full details are provided below, however, our main conclusions are that estimates from the hybrid model are more similar to those based on the LDAK Model (blue points) than those based on the LDAK Model (red points), and that estimates based on the LDAK Model are more precise than those based on the hybrid model (for the blue points, the horizontal segments tend to be shorter than the vertical segments).

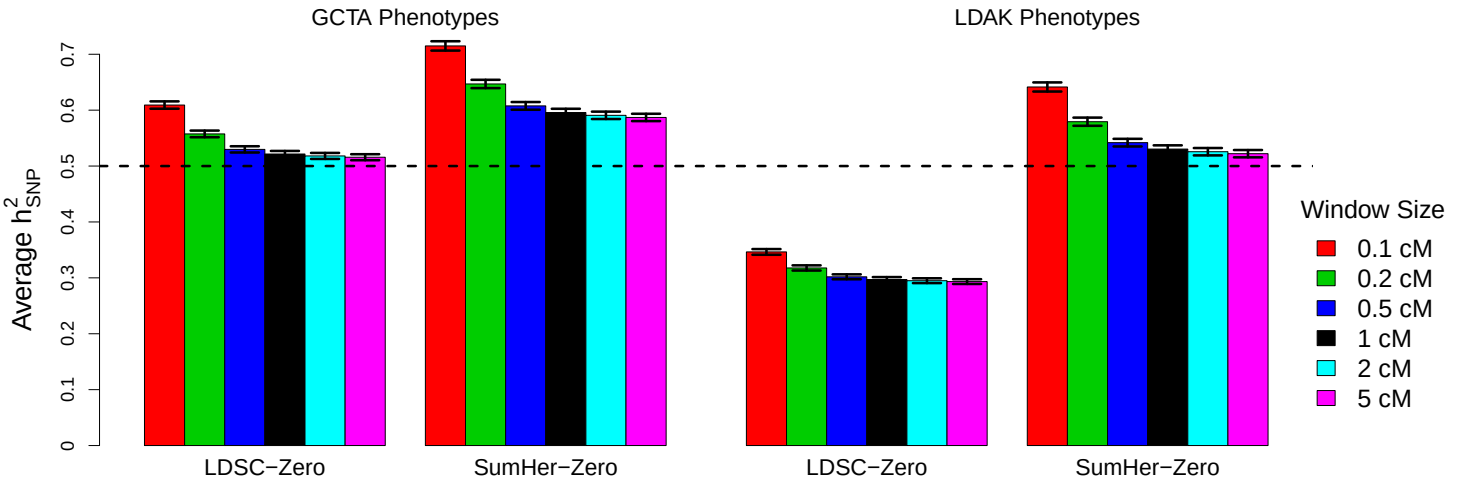
For the left two columns (LDSC vs SumHer-CEPT vs Hybrid-CEPT), we assume that the inflation of test statistics due to confounding is additive; for the right two columns (LDSC-GC vs SumHer-GC vs Hybrid-GC), we assume that the inflation due to confounding is multiplicative. The top two rows show that estimates of confounding bias and h^2_{SNP} from the hybrid model match very closely those using only the LDAK Model, but not those using only the GCTA Model. The third row compares estimates of enrichments, averaged across the 24 traits. For these, LDSC and LDSC-CEPT use a 53-part model, SumHer-CEPT and SumHer-GC use a 25-part model, meaning that Hybrid-CEPT and Hybrid-GC use a 78-part model. Again, estimates from the hybrid model are closer to those from the LDAK Model than those from the GCTA Model (for numerical values see Supplementary Table 19). To calculate the expected share of each category under the hybrid model (the denominators when computing enrichments), it is necessary to choose a value for p; we set $p = 0.5$, so that the expected shares are the averages of those for the GCTA and LDAK Models separately. Note that p is only fixed for the purpose of calculating the expected shares, and not when calculating the estimated shares. Moreover, the expected shares under the GCTA and LDAK Models are so similar (Supplementary Table 1), that the enrichment estimates would generally not change by more than 10% were we to have instead set $p = 0$ or $p = 1$. In the fourth row we compare estimates of the share of h^2_{SNP} contributed by each category (the numerators when computing enrichments), averaged across the 24 traits. Once more, estimates from the hybrid model are closer to those from the LDAK Model than those from the GCTA Model (again, numerical values are in Supplementary Table 19). Note that for computational reasons, we have not considered estimates of genetic correlation (to do so, it would be necessary to compute three sets of tagging files for each of the 276 pairs of traits). However, the fact that, estimates of genetic correlations from LDSC and SumHer-GC are so similar (Supplementary Figure 11), indicates that estimates from all three heritability models will be highly concordant.

While we showed that it is better to use the LDAK Model than the GCTA Model, in order to protect against model misspecification, it might appear better still to use the hybrid model. We advise against this for four reasons. Firstly, using the hybrid model is much more computationally demanding that using the LDAK Model (it typically requires 20 times as much memory, and 20 times as long to run), and when using multi-part models, we often encountered convergence issues. Secondly, the hybrid model generally produces less precise estimates; this is particularly true when estimating enrichments (see Supplementary Table 19). Thirdly, we showed that introducing poorly-genotyped SNPs makes traits appear more GCTA-like (Supplementary Figure 12), indicating that using the hybrid model will make results more sensitive to genotyping errors. Finally, the above results show that even if we switched to the hybrid model, it would not affect our main conclusions, that LDSC tends to over-estimate confounding bias, under-estimate h^2_{SNP} and produce misleading estimates of functional enrichments.

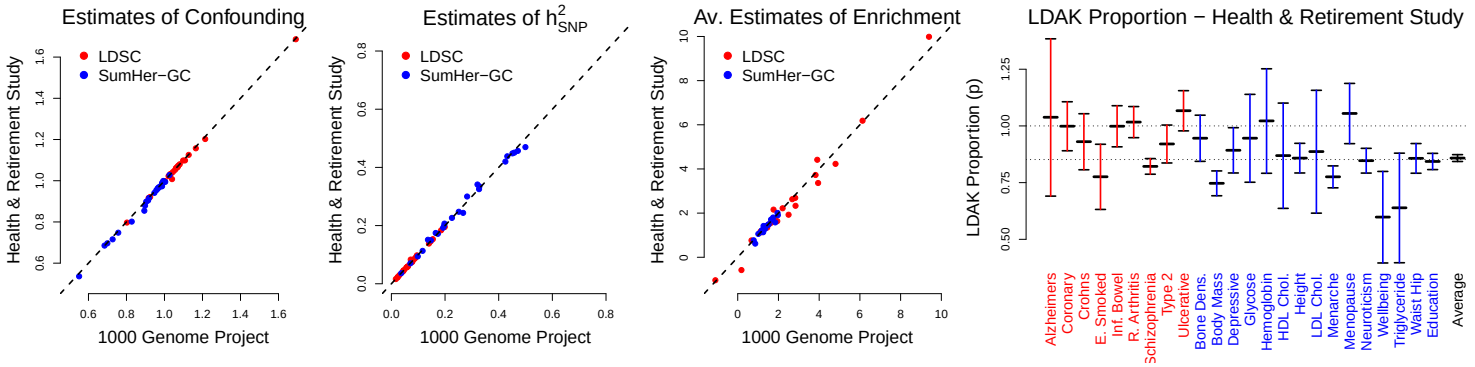


Supplementary Figure 14: Empirical support for the SumHer-GC estimates of confounding bias. Estimates of confounding bias can be used to adjust test statistics. Here we provide empirical evidence for height and body mass index (BMI) that adjusting test statistics based on the SumHer-GC estimates of confounding bias is more accurate than either not adjusting or adjusting based on the LDSC estimates of confounding bias. We choose height and body mass index as for each of these traits we have access to additional results from one related and one independent GWAS: for height, we have so far been using results from the 2014 GIANT Consortium meta-analysis[4] ($\bar{n}_j = 246\,000$), but also have results from the 2010 GIANT Consortium meta-analysis[24] ($\bar{n}_j = 132\,000$), and from an analysis of the UK Biobank ($\bar{n}_j = 336\,000$) by members of the Neale Lab (https://dropbox.com/s/sbfgb6qd5i4cxku/50.assoc.tsv.gz?dl=0); for BMI we have been using results from the 2015 GIANT Consortium meta-analysis[25] ($\bar{n}_j = 230\,000$), but also have results from the 2011 GIANT Consortium meta-analysis[26] ($\bar{n}_j = 122\,000$), and again results from an analysis of the UK Biobank ($\bar{n}_j = 336\,000$) by members of the Neale Lab (https://dropbox.com/s/sweqn7nztyv42zt/21001.assoc.tsv.gz?dl=0).

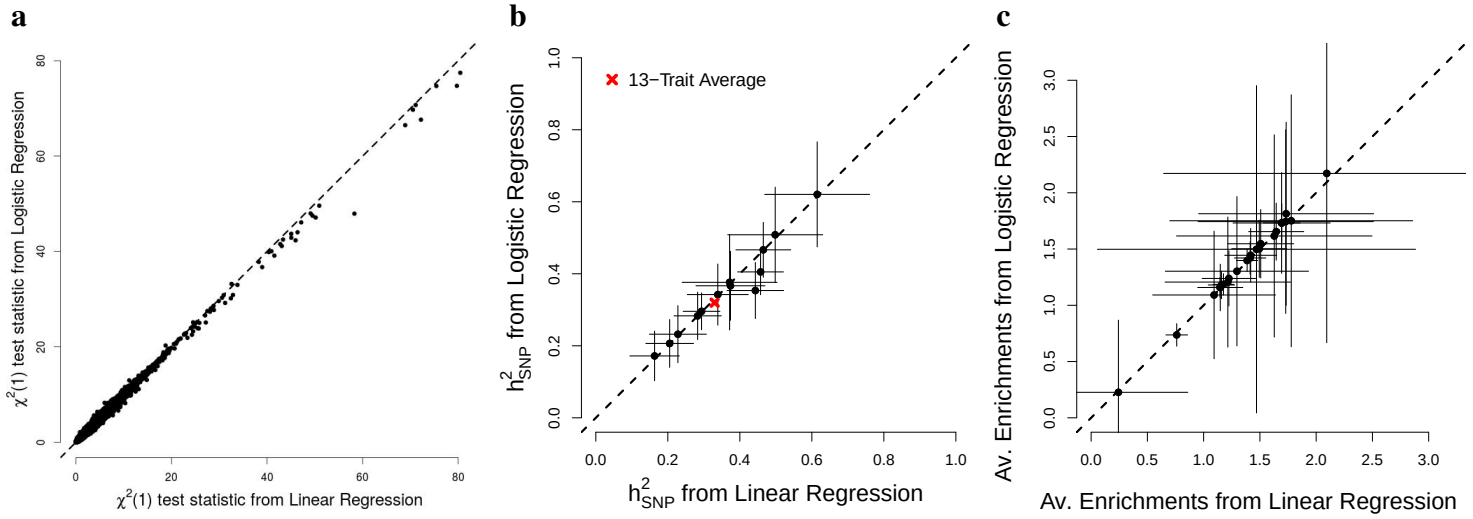
Our basic premise is that for causal SNPs, the effect size estimates from the two pairs of GIANT GWAS should be consistent; for both height and BMI, we only observe this to be the case when we correct test statistics according to the SumHer-GC estimates of confounding bias. To avoid winner's curse, we identified causal SNPs using the (independent) Biobank results; we found 430 genome-wide significant loci ($P < 5 \times 10^{-8}$) for height and 107 for BMI loci. **(a)** compares estimates of (standardized) effect sizes for the height loci calculated from the test statistics as reported (black), after division by the LDSC estimate of confounding bias (blue), and after division by the SumHer-GC estimate of confounding bias (red). Note that the estimated effect size of SNP j is $\sqrt{S_j/(S_j + n_j)}$, where S_j is its test statistic and n_j its sample size (the sign is irrelevant as we align alleles so that the effect is always positive). Without adjustment, effect size estimates from GIANT 2014 are on average 23% (SD 1) higher than those from GIANT 2010, reflecting that the latter used genomic control whereas the former did not; after LDSC adjustment, they are on average 9.5% (SD 1) lower; after SumHer-GC adjustment, they are on average 0.2% higher (SD 1). **(b)** reports the same as (a) except for BMI. Without adjustment, effect size estimates from GIANT 2015 are on average 11% (SD 2) lower than those from GIANT 2010, reflecting that while both GWAS used genomic control, test statistics were corrected more strongly in the former owing to its larger sample size; after LDSC adjustment, they are on average 8.4% (SD 2) lower; after SumHer-GC adjustment, they are on average 1.1% higher (SD 2). **(c)** and **(d)** are the same as (a) and (b), except they consider estimates of the heritability tagged by each causal SNP, $(S_j - 1)/n_j$. In all plots, the dashed black line ($y = x$) is partly or wholly obscured by the red line of best fit.



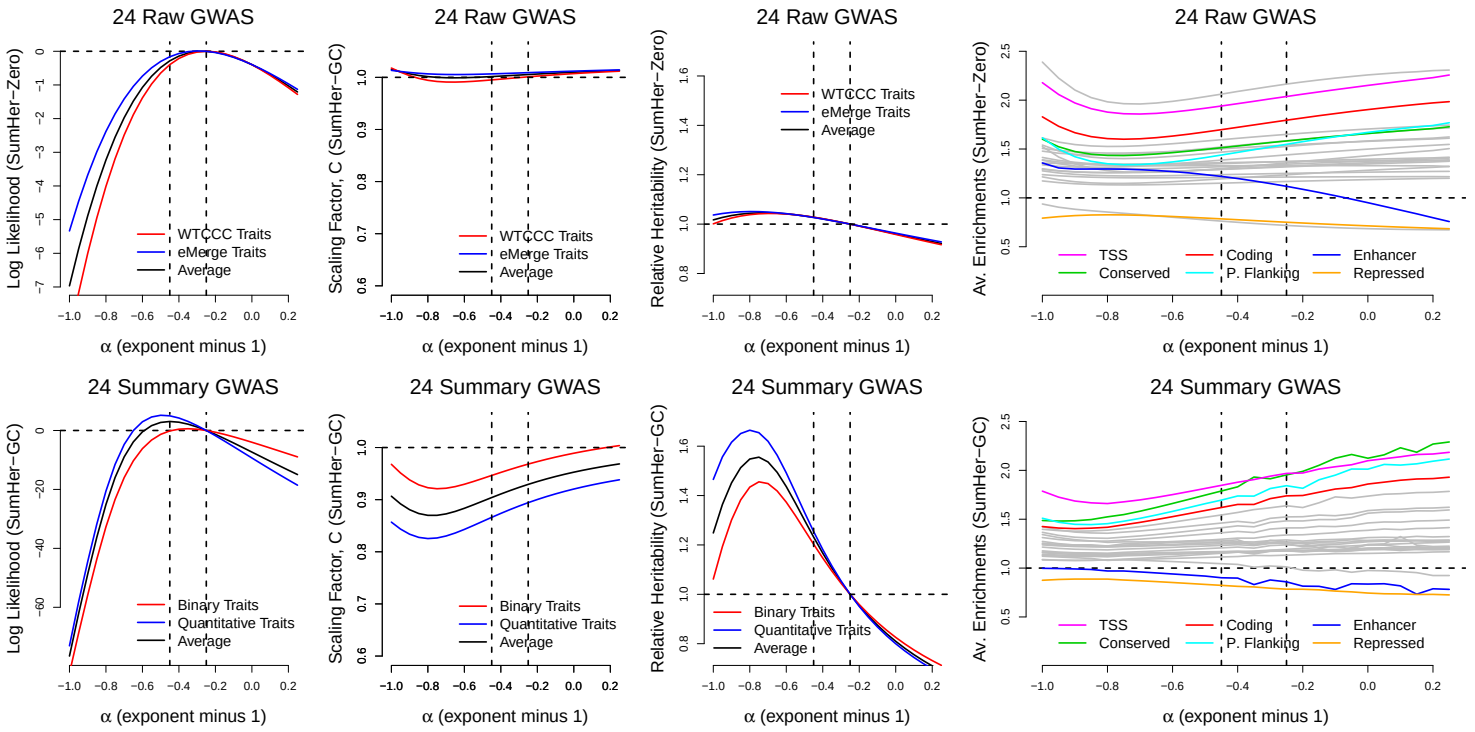
Supplementary Figure 15: Varying the LD window size. Equation (4) in Online Methods uses the approximation $v_j^2 = h_j^2 + \sum_{l \in N_j} r_{jl}^2 \, h_l^2$, where N_j indexes SNPs "near" SNP j. This is based on the assumption that r_{jl}^2 will be negligible for SNPs sufficiently far apart. For the 200 simulated phenotypes described in Supplementary Figure 2, bars report average estimates of h^2_{SNP} from LDSC-Zero and SumHer-Zero using five definitions of N_j (SNPs within 0.1, 0.2, 0.5, 1, 2 or 5 cM). The window size should be sufficiently large to capture the majority of tagging due to LD, but not too large to make computation prohibitively slow. The fact that estimates change little when we increase the cutoff to 2 cM or 5 cM, indicates that our recommended choice, 1 cM, performs well.



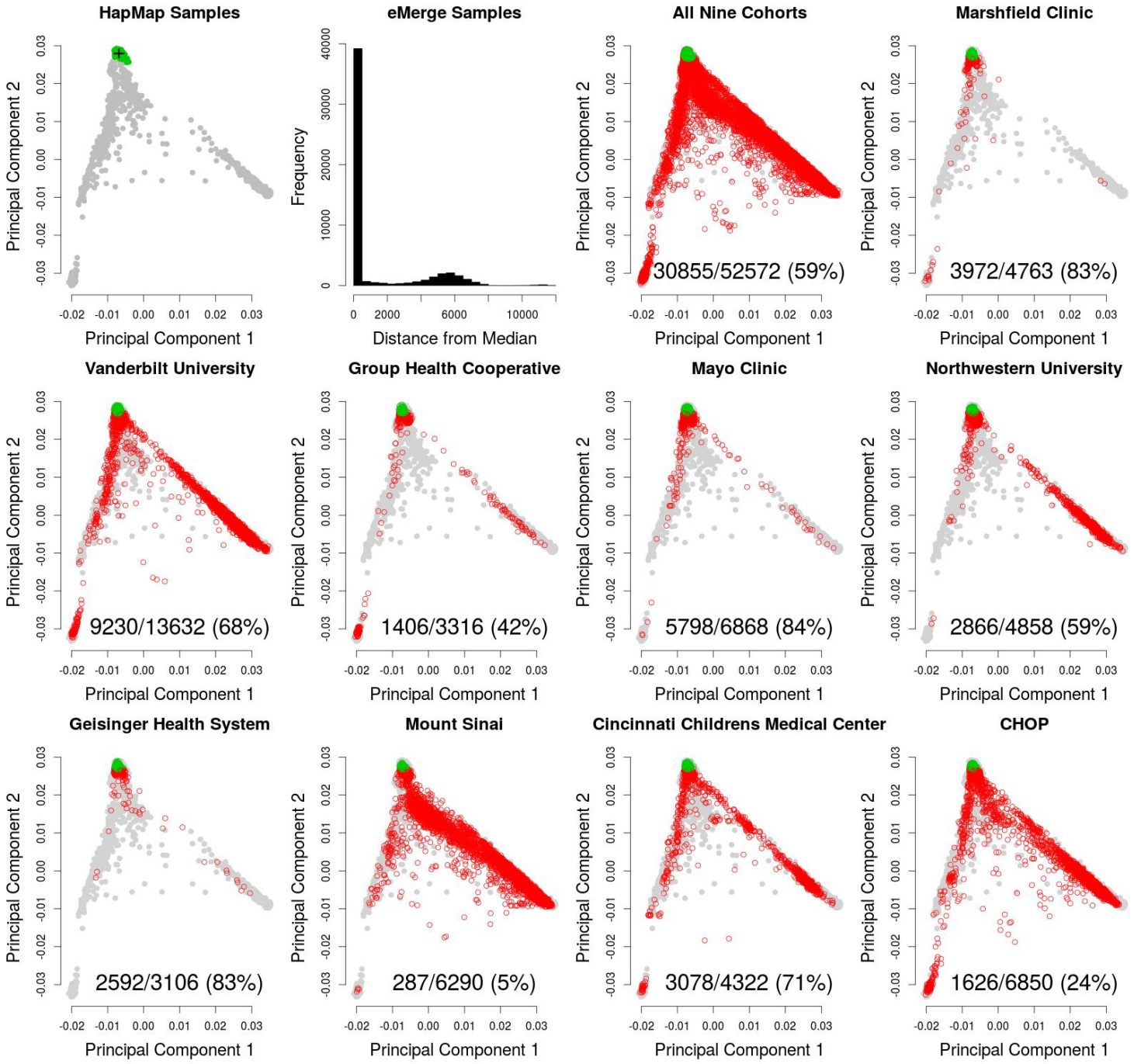
Supplementary Figure 16: Impact of changing the reference panel for the 24 summary GWAS. Our recommended reference panel is 404 non-Finnish Europeans from 1000 Genome Project[1] (1000G). Now we instead use 8 850 unrelated Caucasian individuals from the Health & Retirement Study[27] (HRS). Changing the reference panel has almost no effect on either LDSC or SumHer-GC estimates of confounding bias, h^2_{SNP} and enrichments. Nor does it significantly impact the estimate of p, the LDAK proportion in the hybrid model; the HRS estimate of 0.858 (SD 0.01) is very close to the 1000G estimate of 0.852 (marked by the lower horizontal line). Thus we conclude that our recommended reference panel suffices despite its small sample size (however, a larger reference panel would likely be required if we wished to include rare SNPs in the analyses).



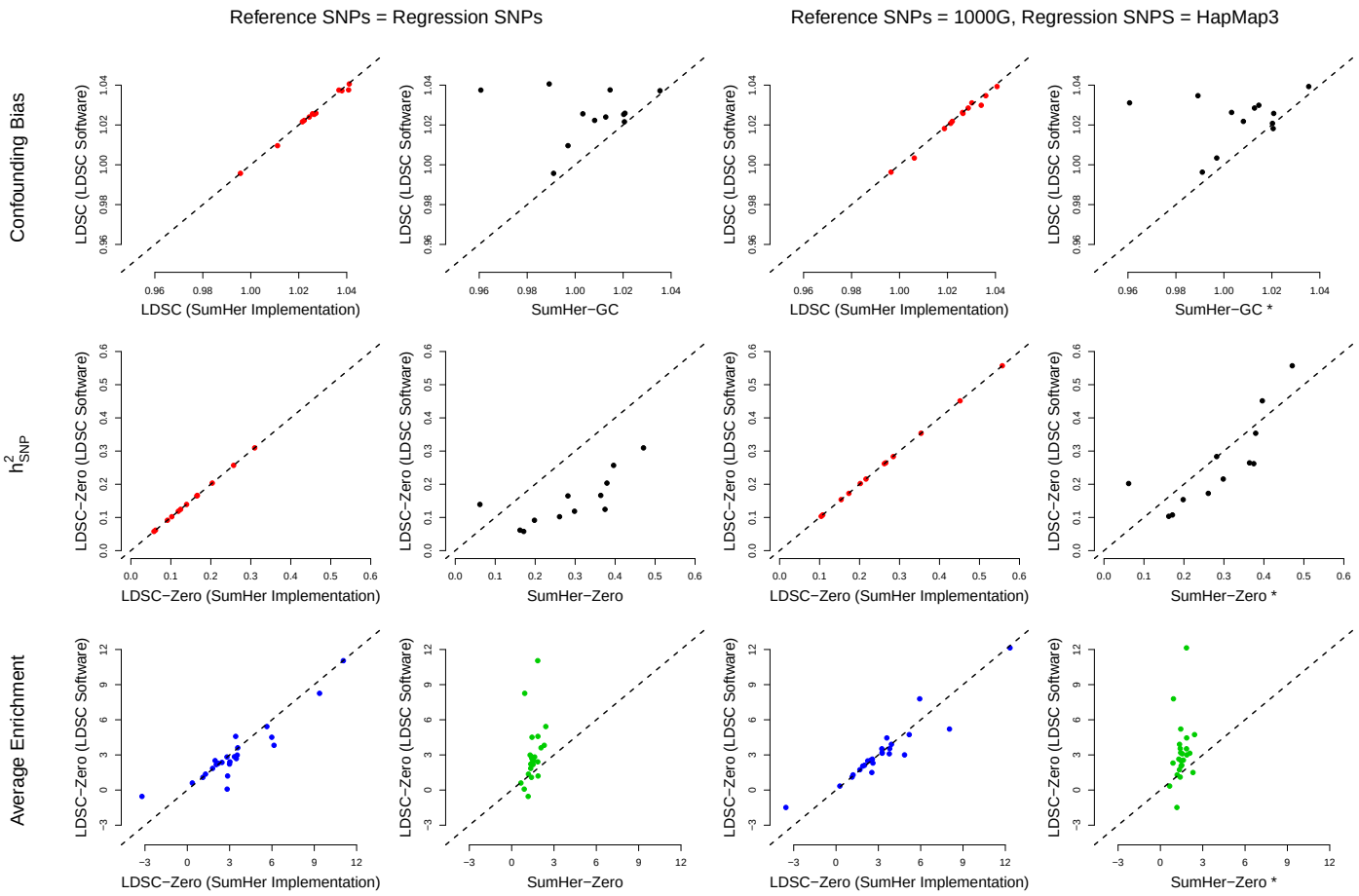
Supplementary Figure 17: Comparing estimates from linear and logistic regression for the 13 WTCCC GWAS. SumHer is designed to be used with summary statistics from (classical) linear regression. However, for binary traits (e.g., case-control studies), its estimates will in general be very similar if instead summary statistics from logistic regression are used. This is because for SNPs with small or moderate effect there will be high concordance between test statistics from linear and logistic regression (while the two methods can produce contrasting test statistics for large-effect SNPs, we recommend excluding these when running SumHer). **(a)** Points compare test statistics from logistic regression to those from linear regression; we focus on approximately 50 000 SNPs per trait, obtained by first thinning (within 1 cM and $r^2_{jl} > 0.1$), then excluding SNPs which individually explain $>1\%$ of phenotypic variation. **(b)** Points compare (liability-scale) estimates of h^2_{SNP} from SumHer-Zero using test statistics from logistic regression to those using test statistics from linear regression. **(c)** For the 24 functional categories, points compare the average estimates of enrichments from SumHer-Zero using test statistics from logistic regression to those using test statistics from linear regression.



Supplementary Figure 18: Investigating the relationship between heritability and MAF for the 25 raw and 24 summary GWAS. The LDAK Model assumes $\mathbb{E}[h_j^2] \propto [f_j(1-f_j)]^{1+\alpha} w_j$, where h_j^2, f_j and w_j are, respectively, the heritability contributed by SNP j, its MAF and its LDAK weighting.[8,11] The parameter α specifies the assumed relationship between heritability and MAF. We recommend $\alpha = -0.25$ based on a previous REML analysis of 42 human traits, where we found that, of the seven values considered (from -1.25 to 0.25 with spacing 0.25), using -0.25 resulted in highest average likelihood.[8] For Column 1, we perform the equivalent analysis using SumHer, now considering 26 alternative values for α (from -1 to 0.25 with spacing 0.05). Across the 25 raw GWAS (top), we again find highest support for $\alpha = -0.25$ (even if we restrict to the 12 eMERGE GWAS, which were not in our previous study); across the 24 summary GWAS (bottom), the most supported value is $\alpha = -0.45$ (using this value increases the average log likelihood by 3.1 nats). Whenever performing heritability analysis, we advise checking whether results are sensitive to the choice of α. Columns 2 & 4 show that for both the 25 raw and 24 summary GWAS, estimates of confounding bias and functional enrichments are relatively stable for plausible values of α (e.g., $-0.5 \leq \alpha \leq 0$). Column 3 shows that for the 24 summary GWAS, estimates of h^2_{SNP} do vary; for example, those based on $\alpha = -0.45$ are on average 23% (SD 3) higher than those based on $\alpha = -0.25$. However, we note that this change is small compared to the 65% (SD 1) reduction if we switched to using the GCTA Model, and only adds support to our conclusion that LDSC tends to substantially under-estimate h^2_{SNP}.



Supplementary Figure 19: Sample quality control for the eMERGE data. The post-imputed eMERGE data contains 52 572 individuals.[28] To filter based on ancestry, we perform principal component analysis (PCA) on the 2 404 individuals from the 1000 Genome Project. **Panel 1** plots the first two PCs for the 1000 Genome Project individuals. The 404 non-Finnish Europeans are marked in green; the center of the cross indicates their median (M_x, M_y), while the horizontal and vertical line segments mark 95% confidence intervals (with widths $3.96\,S_x$ and $3.96\,S_y$, respectively). If (P_{xi}, P_{yi}) denotes the projection of the ith eMERGE individual onto these top two PCs, then we compute $D_i^2 = (P_{xi} - M_x)^2/S_x^2 + (P_{yi} - M_y)^2/S_y^2$, the square of the standardized distance between the individual and the median. We will exclude individuals with $D_i^2 > 5.99$, the 95th percentile of the $\chi^2(2)$ distribution. **Panel 2** shows the distribution of D_i^2 for the eMERGE individuals. **Panel 3** plots D_i^2 for all eMERGE individuals; the numbers indicate how many of the individuals have $D_i^2 < 5.99$ (these individuals are marked in green). **Panels 4-12** plot the same for each of the nine cohorts separately; we decided to exclude individuals from Mount Sinai and CHOP (Children's Hospital of Philadelphia), as for these two cohorts, fewer than 25% of individuals pass our test. In addition to filtering based on D_i^2, we excluded 125 individuals reported as having "Hispanic or Latino" ethnicity, then filtered individuals so that no pair remained with allelic correlation[14] >0.05, (which left 25 875 individuals).



Supplementary Figure 20: Comparing implementations of LDSC for the 12 eMERGE GWAS. In general, we obtained all LDSC-Zero and LDSC estimates using SumHer instead of the LDSC software;[21] doing so allowed us to ensure consistency between comparisons, compute model likelihoods, and was more convenient when running multiple analyses for tens of GWAS. Here, we show that LDSC-Zero and LDSC estimates from SumHer closely match those from the LDSC software (LDSC is equivalent to running the LDSC software with the default options, LDSC-Zero is equivalent to adding the option `--intercept-h2 1`). For computational reasons we focus on the eMERGE traits (as these use the same SNPs we need only compute LD Scores once). When performing an analysis using either SumHer or the LDSC software, it is necessary to decide the sets of reference and regression SNPs (the reference SNPs are used for computing the expected heritability tagged by each SNP, but only the regression SNPs are used when estimating coefficients). We recommend that the two SNP sets are the same, achieved by reducing the reference panel to SNPs for which summary statistics are available. LDSC recommends that the reference panel is as large as possible, but that only HapMap 3[22] SNPs with MAF $\geq$0.05 are used when performing the regression.[10,21]

Here we report estimates of confounding bias (top row), h^2_{SNP} (middle row) and functional enrichments averaged across the 12 traits (bottom row). For Columns 1 & 2, the reference SNPs match the reference SNPs; for Columns 3 & 4, the reference SNPs are all those with MAF $\geq$0.01 in the 1000 Genome Project data,[1] while the regression SNPs are those in HapMap 3 with MAF $\geq$0.05. Columns 1 & 3 compare LDSC-Zero and LDSC estimates from the SumHer implementation to those from the LDSC software. When estimating confounding bias and h^2_{SNP} (plots with red points) we see near-perfect concordance (correlations $>$0.999, mean absolute differences $<$0.001). When estimating enrichments (plots with blue points), we observe some discordance (the correlations are 0.92 and 0.94, the mean absolute differences 0.74 and 0.57), indicating differences between how SumHer and the LDSC software estimate coefficients for multi-part models (we suspect these arise when dealing with extreme values, and because for multi-part models the LDSC software uses one-step weighted least-squares rather than iterative weighted least-squares[10]). However, we see that the discordance is slight compared to that between estimates from the LDSC software and those from our favored method, SumHer-GC (correlations 0.35 and 0.25, mean absolute differences 1.8 and 2.0; plots with green points), and does not affect our overall conclusion that SumHer produces far more modest estimates of enrichment. *Note that the SumHer-Zero and SumHer-GC estimates in Column 4 are obtained using matching reference and regression SNPs (i.e., the same as those in Column 2), as it is not clear how to apply the LDAK Model when the reference and regression SNPs differ.

	1000 Genomes (8 598 885 SNPs)				eMERGE Data (4 555 718 SNPs)			
	Average	Expected share		Percent	Average	Expected share		Percent
Category Annotation	**LD**	**GCTA**	**LDAK**	**Coding**	**LD**	**GCTA**	**LDAK**	**Coding**
Coding[29,30]	124.8	0.015	0.020	100	121.9	0.011	0.014	100
Conserved[31,32]	104.0	0.027	0.032	19.0	109.7	0.025	0.029	14.6
CTCF[33]	112.9	0.024	0.028	2.8	107.8	0.021	0.025	2.1
Digital Genomic Footprint[30,34]	111.9	0.137	0.150	3.3	108.0	0.128	0.148	2.4
DNase I Hypersensitive Site[34–36]	108.6	0.166	0.175	2.8	107.1	0.162	0.184	2.0
FANTOM5 Enhancer[37]	105.8	0.004	0.006	0.0	95.1	0.004	0.005	0.0
Enhancer[33]	106.5	0.042	0.050	3.4	99.7	0.036	0.047	2.4
Fetal DHS[34–36]	104.5	0.084	0.093	3.4	103.2	0.079	0.095	2.5
H3K27ac (Hnisz)[35,38]	108.8	0.391	0.438	2.5	101.3	0.362	0.439	1.9
H3K27ac (PGC2)[35,39]	114.3	0.272	0.314	2.8	103.6	0.248	0.301	2.1
H3K4me1[35,36]	110.3	0.425	0.458	2.6	106.5	0.410	0.476	2.0
H3K4me3[35,36]	138.7	0.136	0.157	4.3	127.9	0.119	0.140	3.2
H3K9ac[35,36]	105.8	0.127	0.150	4.6	102.4	0.109	0.140	3.5
Intronic[29,30]	120.7	0.391	0.405	0.5	119.0	0.397	0.408	0.4
Promoter Flanking[33]	133.4	0.009	0.009	2.4	122.5	0.008	0.009	1.6
Promoter[29,30]	142.6	0.032	0.036	17.5	131.5	0.026	0.030	15.7
Repressed[33]	139.9	0.456	0.425	0.5	133.4	0.475	0.450	0.4
Super Enhancer[38]	104.1	0.169	0.200	2.9	90.7	0.147	0.197	2.2
Transcription Factor Binding Site[30,34]	115.9	0.132	0.147	3.7	107.2	0.120	0.143	2.7
Transcribed[33]	132.1	0.350	0.353	2.7	129.5	0.355	0.343	2.0
Transcription Start Site[33]	132.4	0.018	0.021	9.8	110.9	0.014	0.018	8.2
3' Untranslated Region[29,30]	134.3	0.012	0.014	68.0	114.7	0.009	0.012	67.2
5' Untranslated Region[29,30]	129.9	0.006	0.007	50.3	118.4	0.004	0.005	48.6
Weak Enhancer[33]	94.8	0.021	0.026	2.3	95.8	0.018	0.025	1.6

Supplementary Table 1: Details of the 24 functional annotations. When estimating enrichments of SNP categories, we use the same annotations as Finucane *et al.*, which can be downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/baseline_bedfiles.tgz. This table summarizes the 24 categories, first based on all autosomal SNPs with MAF $\geq$0.01 across the 404 non-Finnish Europeans in the 1000 Genome Project[1], then based on just those SNPs also present in the eMERGE data. We measure the LD of SNP j by $\sum_{l \in N_j} r_{jl}^2$, where N_j indexes the SNPs within 1 cM and r_{jl}^2 is (an estimate of) the squared correlation between SNPs j and l. Given a heritability model, the expected proportion of h_{SNP}^2 contributed by Category k is $(\sum_j I_{jk} q_j)/(\sum_j q_j)$, where I_{jk} indicates whether SNP j belongs to that category (under the GCTA Model, this fraction represents the proportion of SNPs in the category). To demonstrate the gene-centric nature of many of the annotations, for each category we report the proportion of SNPs also in coding regions. The majority of categories (those marked in red) contain SNPs whose average LD is lower than the average LD of all SNPs (136.9 if considering all 1000 Genome Project SNPs or 130.1 if restricting to those in the eMERGE data), which is why most categories are expected to contribute more under the LDAK Model than under the GCTA Model. However, for no category is the expectation under the LDAK Model more than 40% higher than that under the GCTA Model (for the majority, the difference is less than 20%), and so these differences are not the primary reason why estimates of enrichments under the two models are so different (Supplementary Tables 8 & 14).

	Trait (Disease Prevalence, %)	n	m	GIF	GIF'	LDSC-Zero h^2_{SNP} (SD)	LDSC h^2_{SNP} (SD)	LDSC $1+A$ (SD)	SumHer-Zero h^2_{SNP} (SD)	SumHer-GC h^2_{SNP} (SD)	SumHer-GC C (SD)
Wellcome Trust Case Control Consortium	Bipolar Disorder (0.5)	4788	2220776	1.111	1.037	0.24 (0.03)	0.13 (0.03)	1.048 (0.008)	0.47 (0.04)	0.54 (0.08)	0.988 (0.013)
	Coronary Artery Disease (6)	4857	2230330	1.068	1.023	0.21 (0.04)	0.12 (0.05)	1.021 (0.007)	0.37 (0.07)	0.55 (0.13)	0.982 (0.011)
	Crohn's Disease (0.5)	4653	2213243	1.064	1.041	0.14 (0.03)	0.03 (0.04)	1.047 (0.009)	0.34 (0.04)	0.34 (0.09)	1.000 (0.013)
	Hypertension (6)	4865	2230864	1.056	1.046	0.21 (0.04)	0.10 (0.04)	1.029 (0.009)	0.50 (0.07)	0.64 (0.11)	0.985 (0.010)
	Rheumatoid Arthritis (0.5)	4781	2226840	1.048	1.027	0.06 (0.02)	0.00 (0.02)	1.024 (0.007)	0.16 (0.04)	0.20 (0.07)	0.993 (0.011)
	Type 1 Diabetes (0.5)	4890	2225102	1.046	1.021	0.09 (0.02)	0.04 (0.03)	1.024 (0.009)	0.23 (0.04)	0.38 (0.08)	0.971 (0.013)
	Type 2 Diabetes (8)	4841	2225452	1.074	1.039	0.29 (0.05)	0.15 (0.05)	1.032 (0.007)	0.61 (0.07)	0.69 (0.14)	0.994 (0.011)
	Barrett's Oesophagus (1.6)	7049	3400133	1.054	1.039	0.16 (0.03)	0.06 (0.04)	1.031 (0.006)	0.37 (0.05)	0.34 (0.09)	1.004 (0.010)
	Celiac Disease (1)	9946	2064172	1.106	1.028	0.13 (0.02)	0.07 (0.02)	1.041 (0.009)	0.28 (0.03)	0.37 (0.06)	0.980 (0.013)
	Ischaemic Stroke (2)	8982	3363337	1.078	1.076	0.20 (0.02)	-0.00 (0.02)	1.079 (0.006)	0.46 (0.03)	0.07 (0.06)	1.069 (0.010)
	Parkinson's Disease (0.2)	6871	3386927	1.039	1.017	0.11 (0.02)	0.09 (0.04)	1.008 (0.008)	0.21 (0.03)	0.28 (0.07)	0.987 (0.010)
	Psoriasis (0.5)	7474	3380977	1.101	1.074	0.20 (0.03)	0.04 (0.02)	1.064 (0.007)	0.44 (0.04)	0.28 (0.07)	1.029 (0.010)
	Ulcerative Colitis (0.2)	8020	3541951	1.066	1.025	0.13 (0.02)	0.04 (0.02)	1.049 (0.007)	0.29 (0.03)	0.31 (0.05)	0.997 (0.010)
eMERGE Network	Age-related Macular Disease (2.5)	8475	4555346	1.022	1.006	0.13 (0.04)	-0.02 (0.07)	1.027 (0.007)	0.38 (0.07)	0.16 (0.14)	1.017 (0.009)
	Heart Failure (2)	9005	4555718	1.019	1.028	0.10 (0.03)	-0.04 (0.03)	1.036 (0.006)	0.26 (0.04)	-0.02 (0.08)	1.032 (0.008)
	Peripheral Arterial Disease (4)	10541	4555718	1.043	1.038	0.17 (0.03)	0.03 (0.04)	1.029 (0.007)	0.36 (0.05)	0.10 (0.10)	1.024 (0.009)
	Shingles (Herpes Zoster) (25)	13961	4555718	1.025	1.006	0.14 (0.08)	0.19 (0.10)	0.996 (0.007)	0.08 (0.13)	0.25 (0.24)	0.993 (0.009)
	Venous Thromboembolism (5)	11966	4555718	1.035	1.023	0.12 (0.04)	0.07 (0.05)	1.012 (0.008)	0.29 (0.06)	0.31 (0.11)	0.998 (0.010)
	Triglyceride	12137	4555339	1.046	1.023	0.17 (0.03)	0.08 (0.04)	1.025 (0.007)	0.29 (0.04)	0.28 (0.09)	1.002 (0.010)
	LDL Cholesterol	13420	4555718	1.033	1.029	0.09 (0.02)	0.02 (0.03)	1.026 (0.007)	0.20 (0.03)	0.09 (0.06)	1.018 (0.009)
	HDL Cholesterol	13788	4555684	1.061	1.058	0.21 (0.03)	0.08 (0.03)	1.042 (0.008)	0.39 (0.04)	0.29 (0.06)	1.015 (0.010)
	Systolic Blood Pressure	15058	4555718	1.028	1.025	0.06 (0.02)	0.00 (0.02)	1.024 (0.007)	0.17 (0.03)	0.12 (0.06)	1.010 (0.009)
	Diastolic Blood Pressure	15062	4555718	1.018	1.030	0.06 (0.02)	0.01 (0.03)	1.022 (0.006)	0.17 (0.03)	0.06 (0.05)	1.020 (0.009)
	Height	18152	4555718	1.103	1.046	0.31 (0.03)	0.22 (0.04)	1.036 (0.009)	0.48 (0.03)	0.68 (0.08)	0.961 (0.012)
	Body Mass Index	19309	4555718	1.110	1.062	0.26 (0.02)	0.17 (0.02)	1.040 (0.008)	0.40 (0.03)	0.48 (0.06)	0.984 (0.010)
	Average	**9716**	**3575117**	**1.058**	**1.035**	**0.15 (0.01)**	**0.05 (0.01)**	**1.033 (0.001)**	**0.31 (0.01)**	**0.27 (0.02)**	**1.005 (0.002)**
	Relative					**1**	**0.39 (0.04)**		**1.94 (0.05)**	**1.81 (0.09)**	
	Relative					**0.49 (0.02)**	**0.17 (0.02)**		**1**	**0.87 (0.05)**	

Supplementary Table 2: Details of the 25 raw GWAS, and estimates of h^2_{SNP} and confounding bias. n denotes the number of samples, m the number of SNPs, GIF the (naïve) genomic inflation factor and GIF' the genomic inflation factor computed across a thinned subset of SNPs (within 1 cM and $r^2_{jl} > 0.1$). For each trait, we report estimates of h^2_{SNP} from LDSC-Zero, LDSC, SumHer-Zero and SumHer-GC, as well as estimates of confounding bias from LDSC and SumHer-GC (LDSC measures bias via the intercept, $1 + A$, while SumHer-GC estimates the scaling factor, C). For binary traits, estimates of h^2_{SNP} have been converted to the liability scale assuming the stated prevalence.[40,41] The 22 traits with significant h^2_{SNP} ($P < 0.05/25$) from both LDSC-Zero and SumHer-Zero are marked in red (these are the ones we restricted to when estimating genetic correlations).

Wellcome Trust Case Control Consortium data were applied for and downloaded from https://ebi.ac.uk/ega; the accession codes are EGAD00000000001, EGAD00000000002, EGAD00000000003, EGAD00000000004, EGAD00000000005, EGAD00000000006, EGAD00000000007, EGAD00000000008, EGAD00000000009 (WTCCC 1 studies) and EGAD00000000021, EGAD00000000022, EGAD00000000023, EGAD00000000024, EGAD00000000025, EGAD00000000057, EGAD00010000124, EGAD00010000264, EGAD00010000506, EGAD00010000634, EGAS00001000672 (WTCCC 2 studies). eMERGE Network data were applied for and downloaded from https://ncbi.nlm.nih.gov/gap; the accession codes are phs000888.v1.p1.c1, phs000888.v1.p1.c3, phs000888.v1.p1.c4, phs000888.v1.p1.c5.

Trait	Type	$\bar{n}_j$	m	MM	Correction of test statistics	GIF	GIF'	$1 + A$ (SD)	C (SD)
Alzheimer's Diseases[2]	Meta	54162	4347171		GC within cohorts (average GIF 1.03)	1.09	1.07	1.07 (0.02)	1.03 (0.01)
Coronary Artery Disease[42]	Meta	79409	1683418		GC within cohorts (average GIF 1.06)	1.10	1.07	1.06 (0.01)	0.99 (0.01)
Crohn's Disease[43]	Mega	20883	4737260		None	1.14	1.09	1.08 (0.01)	0.97 (0.02)
Ever Smoked?[44]	Meta	74053	1685473		GC within cohorts (average GIF 1.03)	1.11	1.05	1.02 (0.01)	0.96 (0.01)
Inflammatory Bowel[43]	Mega	34652	4906109		None	1.17	1.13	1.13 (0.01)	0.98 (0.01)
Rheumatoid Arthritis[45]	Meta	58284	4544987		GC within cohorts and after meta-analysis	1.05	1.00	1.00 (0.01)	0.90 (0.02)
Schizophrenia[39]	Meta	82315	5292675		None	1.57	1.30	1.16 (0.01)	0.91 (0.01)
Type 2 Diabetes[46]	Meta	157328	4563431		GC within cohorts (average GIF unknown)	1.17	1.08	1.07 (0.01)	0.95 (0.01)
Ulcerative Colitis[43]	Mega	27432	5037937		None	1.12	1.08	1.10 (0.01)	0.99 (0.01)
Bone Mineral Density[47]	Meta	32965	4140623	✓	None	1.11	1.06	1.07 (0.01)	1.00 (0.01)
Body Mass Index[25]	Meta	229902	1721589		GC within cohorts and after meta-analysis	1.13	0.90	0.80 (0.01)	0.55 (0.02)
Depressive Symptoms[48]	Meta	161460	4236802		Corrected using LDSC	1.12	1.06	1.03 (0.01)	0.96 (0.01)
Fasting Glucose[49]	Meta	58074	1750529		GC within cohorts (average GIF 1.06)	1.08	1.05	1.04 (0.01)	0.99 (0.01)
Glycated Hemoglobin[50]	Meta	46368	1721586		GC within cohorts (average GIF 1.03)	1.04	1.04	1.03 (0.01)	0.99 (0.01)
HDL Cholesterol[51]	Meta	95671	1686373	✓	GC within cohorts and after meta-analysis	1.03	0.95	1.04 (0.07)	0.68 (0.03)
Height[4]	Meta	245753	1718207		GC within cohorts (average GIF 1.03)	2.09	1.69	1.69 (0.06)	0.98 (0.04)
LDL Cholesterol[51]	Meta	90946	1684750	✓	GC within cohorts and after meta-analysis	1.03	0.94	1.00 (0.04)	0.73 (0.04)
Menarche Age[52]	Meta	252514	4563412		GC within cohorts (average GIF 1.03)	1.66	1.33	1.21 (0.02)	0.89 (0.02)
Menopause Age[53]	Meta	69360	1684637	✓	Unclear whether GC was performed	1.10	1.05	1.06 (0.02)	0.92 (0.02)
Neuroticism[48]	Meta	170911	4236700		Corrected using LDSC	1.26	1.13	1.06 (0.01)	0.90 (0.02)
Subjective Well-Being[48]	Meta	298420	1625391		Corrected using LDSC	1.16	1.08	1.03 (0.01)	0.97 (0.02)
Triglyceride Levels[51]	Meta	92130	1685101	✓	GC within cohorts and after meta-analysis	1.02	0.96	0.92 (0.03)	0.70 (0.04)
Waist-Hip Ratio[54]	Meta	142286	4548618		GC within cohorts and after meta-analysis	1.05	0.95	0.92 (0.01)	0.76 (0.01)
Years Education[3]	Meta	328917	1716059		GC within cohorts (average GIF 1.02)	1.54	1.22	1.11 (0.01)	0.83 (0.01)
Average		**121008**	**3146618**			**1.21**	**1.09**	**1.04 (0.00)**	**0.93 (0.00)**

Supplementary Table 3: Details of the 24 summary GWAS and estimates of confounding bias. Type reports whether the GWAS performed a mega-analysis (all individuals analyzed together) or meta-analysis (cohorts analyzed separately, then their results combined). $\bar{n}_j$ denotes average sample size; numbers in red mark the eight traits for which per-SNP sample sizes were available (else we set $n_j = n$, the total sample size). m denotes the number of SNPs after filtering; for the four traits marked in red, we excluded SNPs with imputation info score < 0.95; for the remaining traits, for which info scores were not available, we instead restricted to the 4 555 718 SNPs present in the eMERGE data, on the basis that these are SNPs likely to be well-imputed in Caucasian GWAS (Supplementary Table 4). MM indicates traits where mixed-model association analysis was used for one or more analyses (this list is not exhaustive, as for some of the meta-analysis GWAS, it was unclear what analyses were used within cohorts). Column 6 summarizes any correction of test statistics, either using genomic control (GC) or dividing by the LDSC intercept. The 11 traits marked in red are those where correction was lowest; specifically, those where classical linear regression was performed, correction was only within cohorts, and the average correction was 1.03 or less (we decided to include bone mineral density even though mixed-model association analysis was performed, because classical linear regression was used for the majority of cohorts). GIF and GIF' are the genomic inflation factors computed, respectively, from all SNPs and from a thinned subset of SNPs (within 1 cM and $r^2_{jl} > 0.1$). $1 + A$ and C are estimates of the intercept from LDSC and the scaling factor from SumHer-GC, respectively. The low estimates of C for Menarche Age (0.89) and Years Education (0.83) indicate that although genomic control had a relatively small impact on test statistics for individual cohorts, its impact on the combined test statistics was substantial.

	Crohn's Disease			Inflammatory Bowel Disease			Ulcerative Colitis			Schizophrenia		
Filtering	**Number SNPs**	**Average Info**	**Percent <0.95**	**Number SNPs**	**Average Info**	**Percent <0.95**	**Number SNPs**	**Average Info**	**Percent <0.95**	**Number SNPs**	**Average Info**	**Percent <0.95**
All SNPs	11002658	0.863	52	11555662	0.860	52	11113952	0.870	49	9444230	0.942	35
1000 Genome Project	8125415	0.909	41	8139841	0.913	39	8129688	0.916	37	7709782	0.941	31
Wellcome Trust Controls	4108338	0.973	15	4108554	0.975	13	4108472	0.977	11	4096361	0.983	8
HRS (Info $\geq$0.8)	7195748	0.937	34	7196494	0.940	32	7196184	0.943	30	7136612	0.953	26
HRS (Info $\geq$0.99)	3459954	0.975	13	3460058	0.977	11	3459986	0.979	10	3445395	0.985	6
eMerge Network	4573808	0.980	10	4573830	0.982	8	4573817	0.984	6	4572847	0.987	3
HapMap 3	1176981	0.962	22	1177105	0.966	19	1177004	0.969	17	1171395	0.977	13

Supplementary Table 4: Proxies for info scores. When possible, we advise restricting analyses to only high-quality SNPs, for example, those with imputation info score $\geq$0.95. However, info scores were only available for four of our 24 summary GWAS (Crohn's Disease, inflammatory bowel disease, schizophrenia and ulcerative colitis), so for the remaining traits, we restricted to SNPs well-imputed in alternative datasets (in addition to excluding SNPs not in our reference panel, 404 non-Finnish Europeans from the 1000 Genome Project[1] data). We considered five proxies: SNPs with info score $\geq$0.99 across the 7 906 individuals in the three Wellcome Trust[55] control datasets; SNPs with info score $\geq$0.8 across the 10 563 Caucasian individuals in the Health and Retirement Study[27] (HRS); SNPs with info score $\geq$0.99 across the 10 563 Caucasian HRS individuals; SNPs with info score $\geq$0.95 across the 52 572 (multi-ethnic) individuals in the eMERGE data; SNPs present in HapMap 3.[22] For the four traits for which info scores were provided, we assessed how effective each of these filterings based on the number of SNPs which remained, their average info score and the proportion of these with info score $<$0.95 (note that the high correlation between the results for Crohn's Disease, inflammatory bowel disease and ulcerative colitis reflects that these three traits were analyzed by the same authors; Supplementary Table 3). Of the five proxies considered, we find that restricting to SNPs present in the eMERGE data is most effective; it produces the highest average info score, the lowest proportion of SNPs with info $<$0.95, and results in the removal of only 25-35% of 1000 Genome Project SNPs. The authors of LDSC have recommended restricting to SNPs in HapMap 3.[10,21,23,56,57] These results indicate that this is an inefficient proxy; although restricting to HapMap 3 SNPs does improve the average info score, the improvement is less than restricting to high-quality Wellcome Trust, HRS or eMERGE SNPs, and leads to the exclusion of over 80% of 1000 Genome Project SNPs. While it appears that reducing to eMERGE SNPS is an effective way of performing quality control for GWAS which do not provide info scores, it is important to note that it will not be perfect and some poorly-genotyped SNPs will not be filtered out (evidenced by the fact that 3-10% of the remaining SNPs have info score $<$0.95).

Trait	$L(S\|\widehat{h^2_{\text{SNP}}}, D^0)$ GCTA	$L(S\|\widehat{h^2_{\text{SNP}}}, D^{\widehat{h^2_{\text{SNP}}}}) - L(S\|0, D^{\widehat{h^2_{\text{SNP}}}})$ GCTA	LDAK	Proportion of LDAK, p (SD)
	LDAK			
Bipolar Disorder	-89184.8	-89153.5	41.0	69.1	0.91 (0.07)
Coronary Artery Disease	-86637.0	-86630.6	12.1	16.3	0.89 (0.13)
Crohn's Disease	-88431.9	-88408.3	15.3	36.6	1.08 (0.13)
Hypertension	-87332.9	-87317.7	14.0	30.4	1.03 (0.10)
Rheumatoid Arthritis	-86830.9	-86823.3	3.4	10.8	1.22 (0.11)
Type 1 Diabetes	-87808.4	-87796.7	8.5	21.3	1.07 (0.14)
Type 2 Diabetes	-87862.3	-87849.7	18.2	33.0	0.94 (0.10)
Barrett's Oesophagus	-125123.3	-125108.0	14.4	30.9	1.08 (0.11)
Celiac Disease	-95275.2	-95250.5	26.2	48.3	0.98 (0.09)
Ischaemic Stroke	-126967.2	-126923.1	39.0	88.7	1.08 (0.06)
Parkinson's Disease	-124082.7	-124087.5	13.1	20.4	0.66 (0.29)
Psoriasis	-128981.6	-128938.6	38.7	81.6	1.08 (0.06)
Ulcerative Colitis	-127867.0	-127823.5	32.9	69.3	1.15 (0.06)
Age-related Macular Disease	-136043.5	-136037.9	3.6	13.9	1.27 (0.20)
Heart Failure	-135762.6	-135755.1	5.8	15.4	1.30 (0.15)
Peripheral Arterial Disease	-135768.7	-135762.8	10.1	20.1	1.05 (0.15)
Shingles (Herpes Zoster)	-133490.8	-133491.7	1.5	0.2	-1.30 (5.37)
Venous Thromboembolism	-135590.7	-135587.2	6.4	15.8	1.02 (0.25)
Triglyceride	-140091.2	-140084.9	19.6	28.4	0.88 (0.16)
LDL Cholesterol	-136366.2	-136360.2	8.7	18.8	1.10 (0.16)
HDL Cholesterol	-141591.3	-141569.0	35.3	59.3	1.04 (0.10)
Systolic Blood Pressure	-134966.3	-134960.7	4.7	17.5	1.20 (0.18)
Diastolic Blood Pressure	-135031.6	-135027.4	5.3	17.2	1.14 (0.22)
Height	-143344.5	-143305.7	111.0	139.1	0.85 (0.08)
Body Mass Index	-141368.6	-141326.2	94.0	119.4	0.89 (0.07)
Average	**-119672.0**	**-119655.2**	**23.3**	**40.9**	**1.03 (0.02)**
Difference	**0**	**16.9**			

Supplementary Table 5: Comparing heritability models for the 25 raw GWAS using SumHer. $L(S|h^2_{\text{SNP}}, D)$ is the (weighted) log likelihood, defined in Online Methods. SumHer provides two likelihood-based metrics for comparing heritability models. Our preferred metric is $L(S|\widehat{h^2_{\text{SNP}}}, D^0)$, where $\widehat{h^2_{\text{SNP}}}$ is the final estimate of h^2_{SNP}, and D^0 is the initial weight matrix (obtained by setting $1/D_{jj} = \sum_{l \in N_j} r^2_{jl}$). The second metric is $L(S|\widehat{h^2_{\text{SNP}}}, D^{\widehat{h^2_{\text{SNP}}}}) - L(S|0, D^{\widehat{h^2_{\text{SNP}}}})$, where $D^{\widehat{h^2_{\text{SNP}}}}$ is the weight matrix after the final iteration; while it might seem natural to compare heritability models based on $L(S|\widehat{h^2_{\text{SNP}}}, D^{\widehat{h^2_{\text{SNP}}}})$, this is not valid because $D^{\widehat{h^2_{\text{SNP}}}}$ depends on the heritability model, hence why we normalize by subtracting $L(S|0, D^{\widehat{h^2_{\text{SNP}}}})$, the null likelihood calculated using the same weights (the likelihood ratio test statistic reported by SumHer is twice this metric). We see that regardless of which metric is used, the relative performances of the GCTA and LDAK Models are very similar to when we compare models using the log likelihood from REML (Supplementary Table 6). An alternative way to compare the GCTA and LDAK Models is using Hybrid-Zero, which assumes the heritability model $q_j = (1-p) \times 1/m + p \times [f_j(1-f_j)]^{0.75}\ w_j/Q'$, where $Q' = \sum_j [f_j(1-f_j)]^{0.75}\ w_j$. The final column reports estimates of p, the LDAK proportion. The imprecise estimate for Shingles reflects that this trait does not have significant h^2_{SNP} (Supplementary Table 2).

Note that when comparing heritability models based on $L(S|\widehat{h^2_{\text{SNP}}}, D^0)$, we must ensure that the same SNPs are used for each model (so that D^0 is constant). When calculating the tagfile (which contains $q_j + \sum_{l \in N_j} q_l r^2_{jl}$ for each SNP), by default SumHer excludes SNPs with $q_j = 0$. This is not relevant for the GCTA Model ($q_j = 1$), but is for the LDAK Model ($q_j = w_j\ [f_j(1-f_j)]^{0.75}$), as many SNPs will have $w_j = 0$ (these are SNPs whose variation is perfectly tagged by their neighbors). To prevent SumHer ignoring SNPs with $q_j = 0$, use the option `-reduce NO`.

	GCTA Model		LDAK Model	
Trait (Disease prevalence, %)	**h^2_{SNP} (SD)**	**Log Likelihood**	**h^2_{SNP} (SD)**	**Log Likelihood**
Bipolar Disorder (0.5)	0.21 (0.02)	-3247.4	0.35 (0.03)	-3221.0
Coronary Artery Disease (6)	0.19 (0.04)	-3147.2	0.33 (0.06)	-3142.9
Crohn's Disease (0.5)	0.13 (0.02)	-3143.9	0.28 (0.03)	-3121.7
Hypertension (6)	0.20 (0.04)	-3365.7	0.41 (0.05)	-3348.8
Rheumatoid Arthritis (0.5)	0.05 (0.02)	-3184.2	0.15 (0.03)	-3175.7
Type 1 Diabetes (0.5)	0.09 (0.02)	-3417.8	0.18 (0.03)	-3406.6
Type 2 Diabetes (8)	0.26 (0.04)	-3338.5	0.51 (0.06)	-3324.8
Barrett's Oesophagus (1.6)	0.16 (0.03)	-3873.2	0.31 (0.04)	-3857.8
Celiac Disease (1)	0.13 (0.02)	-5503.9	0.25 (0.02)	-5478.3
Ischaemic Stroke (2)	0.18 (0.02)	-5611.3	0.35 (0.03)	-5562.1
Parkinson's Disease (0.2)	0.11 (0.02)	-3882.5	0.19 (0.03)	-3878.0
Psoriasis (0.5)	0.17 (0.02)	-4434.4	0.33 (0.03)	-4388.7
Ulcerative Colitis (0.2)	0.13 (0.01)	-5225.7	0.24 (0.02)	-5187.3
Age-related Macular Disease (2.5)	0.12 (0.04)	-2988.6	0.35 (0.06)	-2976.0
Heart Failure (2)	0.10 (0.03)	-4767.5	0.23 (0.04)	-4756.4
Peripheral Arterial Disease (4)	0.17 (0.04)	-4593.9	0.38 (0.05)	-4572.9
Shingles (Herpes Zoster) (25)	0.14 (0.08)	-2394.3	0.10 (0.12)	-2395.5
Venous Thromboembolism (5)	0.12 (0.03)	-5636.3	0.30 (0.05)	-5621.8
Triglyceride	0.17 (0.02)	-71355.4	0.26 (0.03)	-71348.0
LDL Cholesterol	0.09 (0.02)	-64285.5	0.20 (0.03)	-64271.8
HDL Cholesterol	0.20 (0.02)	-56028.7	0.35 (0.03)	-56000.5
Systolic Blood Pressure	0.06 (0.02)	-60424.0	0.17 (0.03)	-60404.8
Diastolic Blood Pressure	0.06 (0.02)	-52669.7	0.17 (0.03)	-52652.8
Height	0.29 (0.02)	-60490.5	0.42 (0.02)	-60464.1
Body Mass Index	0.23 (0.02)	-64084.7	0.35 (0.02)	-64062.6
Average	**0.15 (0.00)**	**-20043.8**	**0.28 (0.01)**	**-20024.8**
Difference		**0**		**19.0**

Supplementary Table 6: Comparing heritability models for the 25 raw GWAS using REML. For each trait, we report estimates of h^2_{SNP} and log likelihoods from REML assuming the GCTA and LDAK Models;[8] for consistency with our SumHer analyses, when calculating kinship matrices we excluded SNPs not in our reference panel (even though that is not used for this analysis) and those in the major histocompatibility complex (Chromosome 6: 25-34 Mb), as well as SNPs which individually explain >1% of phenotypic variation and SNPs in LD with these (within 1 cM and $r^2_{jl} > 0.1$). For binary traits, estimates of h^2_{SNP} have been converted to the liability scale assuming the stated prevalence.[40,41] For each trait, the higher likelihood is marked in red.

	log likelihood $L(S\|\widehat{h^2_{\text{SNP}}}, D^0)$				Proportion of LDAK, p (SD)	
Trait	**LDSC**	**SumHer-GC**	**LDSC-Zero**	**SumHer-Zero**	**Hybrid-GC**	**Hybrid-Zero**
Alzheimer's Disease	-155870	-155861	-155900	-155863	1.05 (0.18)	1.05 (0.14)
Coronary Artery Disease	-126004	-125985			0.99 (0.06)	
Crohn's Disease	-156340	-156288	-156374	-156293	0.95 (0.06)	0.93 (0.07)
Ever Smoked?	-120324	-120323	-120328	-120326	0.74 (0.08)	0.62 (0.10)
Inflammatory Bowel	-166746	-166678	-166824	-166679	1.00 (0.04)	1.00 (0.05)
Rheumatoid Arthritis	-156691	-156638			1.02 (0.04)	
Schizophrenia	-186224	-185879	-186497	-185924	0.81 (0.02)	0.77 (0.02)
Type 2 Diabetes	-170951	-170904			0.91 (0.05)	
Ulcerative Colitis	-155970	-155920	-156039	-155919	1.07 (0.04)	1.07 (0.04)
Bone Mineral Density	-145355	-145337	-145396	-145337	0.93 (0.06)	0.93 (0.06)
Body Mass Index	-168672	-168574			0.73 (0.03)	
Depressive Symptoms	-127467	-127445			0.89 (0.05)	
Fasting Glucose	-150936	-150929			0.93 (0.11)	
Glycated Hemoglobin	-122114	-122110	-122119	-122110	1.02 (0.11)	1.01 (0.13)
HDL Cholesterol	-203754	-203713			0.91 (0.15)	
Height	-219673	-219485	-220089	-219494	0.86 (0.03)	0.85 (0.04)
LDL Cholesterol	-192567	-192537			0.89 (0.13)	
Menarche Age	-216546	-216351	-216750	-216377	0.76 (0.03)	0.71 (0.03)
Menopause Age	-149661	-149617			1.06 (0.06)	
Neuroticism	-137431	-137355			0.82 (0.03)	
Subjective Well-Being	-115635	-115647			0.61 (0.10)	
Triglyceride	-201576	-201571			0.64 (0.13)	
Waist-Hip Ratio	-128195	-128097			0.84 (0.04)	
Years Education	-170237	-169873	-170404	-169959	0.84 (0.02)	0.75 (0.03)
Average	**-160206**	**-160130**	**-165156**	**-164935**	**0.85 (0.01)**	**0.82 (0.01)**

Supplementary Table 7: Comparing heritability models for the 24 summary GWAS using SumHer. Columns 2 to 5 report the log likelihood $L(S|\widehat{h^2_{\text{SNP}}}, D^0)$ (see Supplementary Table 5), first calculated from LDSC and SumHer-GC, then, for the 11 traits least impacted by genomic control (Supplementary Table 3), calculated from LDSC-Zero and SumHer-Zero (the higher value in each pair is marked in red). The final two columns report estimates of p, the LDAK proportion, for each trait from Hybrid-GC and, for the 11 traits least impacted by genomic control, from Hybrid-Zero.

	LDSC-Zero (53-part model)			SumHer-Zero (25-part model)		
Category	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**
Coding	0.052 (0.020)	0.012	4.54 (1.72)	0.027 (0.005)	0.014	1.84 (0.32)
Conserved	0.247 (0.039)	0.025	9.74 (1.54)	0.044 (0.007)	0.030	1.46 (0.23)
CTCF	0.026 (0.033)	0.020	1.29 (1.65)	0.013 (0.006)	0.024	0.53 (0.25)
Digital Genomic Footprint	0.337 (0.083)	0.122	2.77 (0.68)	0.205 (0.014)	0.147	1.40 (0.09)
DNase I Hypersensitive Site	0.342 (0.089)	0.156	2.21 (0.57)	0.234 (0.015)	0.186	1.25 (0.08)
FANTOM5 Enhancer	0.038 (0.015)	0.003	11.29 (4.39)	0.006 (0.003)	0.005	1.28 (0.56)
Enhancer	0.118 (0.034)	0.033	3.55 (1.01)	0.075 (0.008)	0.045	1.69 (0.17)
Fetal DHS	0.139 (0.064)	0.075	1.87 (0.84)	0.130 (0.011)	0.095	1.37 (0.12)
H3K27ac (Hnisz)	0.606 (0.031)	0.345	1.76 (0.09)	0.583 (0.015)	0.428	1.36 (0.04)
H3K27ac (PGC2)	0.511 (0.053)	0.237	2.17 (0.22)	0.413 (0.016)	0.294	1.40 (0.05)
H3K4me1	0.757 (0.065)	0.396	1.92 (0.16)	0.607 (0.017)	0.471	1.29 (0.04)
H3K4me3	0.283 (0.045)	0.114	2.50 (0.39)	0.204 (0.013)	0.135	1.52 (0.10)
H3K9ac	0.271 (0.045)	0.102	2.72 (0.44)	0.235 (0.013)	0.133	1.78 (0.10)
Intronic	0.507 (0.024)	0.402	1.25 (0.06)	0.477 (0.013)	0.410	1.16 (0.03)
Promoter Flanking	-0.009 (0.020)	0.008	-1.25 (2.64)	0.013 (0.004)	0.008	1.58 (0.44)
Promoter	0.023 (0.022)	0.025	1.02 (0.89)	0.044 (0.006)	0.029	1.57 (0.22)
Repressed	0.348 (0.066)	0.478	0.73 (0.14)	0.334 (0.018)	0.455	0.73 (0.04)
Super Enhancer	0.314 (0.018)	0.135	2.29 (0.13)	0.302 (0.011)	0.188	1.60 (0.06)
Transcription Factor Binding Site	0.277 (0.065)	0.114	2.43 (0.57)	0.182 (0.014)	0.141	1.30 (0.10)
Transcribed	0.347 (0.059)	0.363	0.95 (0.16)	0.430 (0.017)	0.347	1.23 (0.05)
Transcription Start Site	0.043 (0.021)	0.012	3.68 (1.69)	0.031 (0.005)	0.016	1.91 (0.31)
3' Untranslated Region	0.046 (0.015)	0.009	4.74 (1.61)	0.019 (0.004)	0.012	1.65 (0.35)
5' Untranslated Region	0.001 (0.013)	0.004	0.09 (2.85)	0.011 (0.003)	0.005	2.11 (0.57)
Weak Enhancer	0.060 (0.032)	0.016	3.62 (1.97)	0.036 (0.006)	0.023	1.58 (0.25)

Supplementary Table 8: Average estimates of functional enrichments across the 25 raw GWAS. We estimate enrichments using either LDSC-Zero with a 53-part model or SumHer-Zero with a 25-part model. For the 25-part model, we divide the genome into a set for each of the 24 categories, plus a set containing all SNPs, while for the 53-part model, we divide the genome into a set for each of the 24 categories, a set for each of 28 buffer regions (see Finucane *et al.*,[10]), plus a set containing all SNPs. For each trait, we calculate the estimated enrichment of each category, obtained by dividing the estimated share of h^2_{SNP} contributed by the category by its expected share; for each category, we then report the (inverse-variance-weighted) average estimated enrichment across the 25 traits. Average estimated enrichments significantly different from one ($P < 0.05$) are marked in red.

Trait 1	Trait 2	Genetic Correlation (SD)			
		LDSC-Zero	**LDSC**	**SumHer-Zero**	**SumHer-GC**
Bipolar Disorder	Crohn's Disease	0.21 (0.10)	0.51 (0.36)	0.41 (0.14)	**0.41 (0.12)**
Bipolar Disorder	Hypertension	0.20 (0.11)	0.38 (0.20)	**0.51 (0.14)**	**0.43 (0.10)**
Bipolar Disorder	Type 1 Diabetes	0.11 (0.13)	0.21 (0.24)	0.27 (0.15)	0.22 (0.11)
Bipolar Disorder	Type 2 Diabetes	0.11 (0.10)	0.25 (0.18)	0.29 (0.13)	0.27 (0.11)
Bipolar Disorder	Ulcerative Colitis	0.12 (0.10)	0.29 (0.25)	**0.47 (0.11)**	**0.43 (0.10)**
Coronary Artery Disease	Hypertension	0.15 (0.14)	0.25 (0.24)	0.41 (0.19)	0.33 (0.13)
Coronary Artery Disease	Type 1 Diabetes	0.37 (0.16)	0.66 (0.29)	0.32 (0.21)	0.23 (0.14)
Coronary Artery Disease	Type 2 Diabetes	0.20 (0.15)	0.38 (0.22)	0.36 (0.19)	0.30 (0.14)
Coronary Artery Disease	Psoriasis	0.14 (0.14)	0.28 (0.30)	**0.66 (0.18)**	**0.66 (0.17)**
Coronary Artery Disease	Ulcerative Colitis	0.12 (0.14)	0.24 (0.35)	0.45 (0.19)	0.39 (0.15)
Crohn's Disease	Hypertension	-0.10 (0.16)	-0.25 (0.60)	0.38 (0.15)	0.36 (0.13)
Crohn's Disease	Ulcerative Colitis	0.26 (0.12)	1.01 (0.51)	**0.86 (0.14)**	**0.87 (0.14)**
Hypertension	Type 1 Diabetes	0.14 (0.16)	0.30 (0.29)	0.37 (0.16)	0.27 (0.11)
Hypertension	Type 2 Diabetes	0.13 (0.15)	0.24 (0.26)	0.37 (0.17)	0.32 (0.13)
Hypertension	Ulcerative Colitis	0.09 (0.13)	0.18 (0.35)	**0.55 (0.14)**	**0.48 (0.12)**
Type 1 Diabetes	Celiac Disease	0.52 (0.16)	1.51 (1.23)	0.43 (0.18)	0.37 (0.15)
Type 1 Diabetes	Parkinson's Disease	0.38 (0.24)	0.82 (0.49)	0.72 (0.26)	0.53 (0.16)
Type 1 Diabetes	Psoriasis	0.21 (0.15)	0.62 (0.53)	0.49 (0.17)	0.51 (0.17)
Type 1 Diabetes	Ulcerative Colitis	0.21 (0.15)	0.64 (0.45)	0.39 (0.18)	0.31 (0.14)
Type 2 Diabetes	Psoriasis	0.10 (0.12)	0.30 (0.28)	0.33 (0.13)	0.38 (0.14)
Type 2 Diabetes	Ulcerative Colitis	0.14 (0.12)	0.29 (0.30)	0.27 (0.13)	0.26 (0.12)
Type 2 Diabetes	Body Mass Index	0.16 (0.09)	0.27 (0.18)	**0.45 (0.13)**	0.42 (0.13)
Celiac Disease	Parkinson's Disease	0.08 (0.13)	0.15 (0.21)	0.39 (0.21)	0.28 (0.14)
Celiac Disease	Psoriasis	0.04 (0.09)	0.13 (0.20)	0.34 (0.13)	0.34 (0.13)
Parkinson's Disease	Ulcerative Colitis	0.19 (0.14)	0.46 (0.29)	0.36 (0.16)	0.31 (0.12)
Psoriasis	Ulcerative Colitis	0.15 (0.10)	0.58 (0.42)	0.32 (0.12)	0.40 (0.14)
Ulcerative Colitis	Venous Thromboembolism	0.16 (0.16)	0.35 (0.43)	0.45 (0.20)	0.38 (0.17)
Venous Thromboembolism	Height	0.24 (0.12)	0.39 (0.22)	0.31 (0.15)	0.23 (0.12)
Triglyceride	HDL Cholesterol	-0.35 (0.12)	-0.69 (0.23)	-0.42 (0.15)	-0.49 (0.16)
Triglyceride	Body Mass Index	0.30 (0.10)	0.47 (0.18)	0.40 (0.14)	0.37 (0.14)
LDL Cholesterol	HDL Cholesterol	0.18 (0.13)	0.58 (0.48)	0.35 (0.15)	0.59 (0.28)
LDL Cholesterol	Body Mass Index	-0.16 (0.11)	-0.40 (0.33)	-0.36 (0.15)	-0.48 (0.23)
HDL Cholesterol	Body Mass Index	**-0.27 (0.08)**	**-0.52 (0.15)**	**-0.50 (0.11)**	**-0.55 (0.12)**

Supplementary Table 9: Estimates of genetic correlation for the 25 raw GWAS. In general, it is only possible to get a meaningful estimate of genetic correlation when both traits have substantial h^2_{SNP}, so for this analysis we only use the 22 traits for which both LDSC-Zero and SumHer-Zero find significant h^2_{SNP} ($P < 0.05/25$); this filtering excludes rheumatoid arthritis, age-related macular disease and shingles (Supplementary Table 2). Of the ${}^{22}C_2 = 231$ pairs of traits we considered, this table reports genetic correlation for the 33 pairs with significant correlation $P < 0.05$) from either LDSC or SumHer-GC. Nominally significant estimates ($P < 0.05$) are marked in red, while Bonferroni significant estimates ($P < 0.05/231$) are also in **bold**.

	Original Controls — Confounding (SD)			2 000 POBI Controls — Confounding (SD)			2 000 1000G Controls — Confounding (SD)		
Trait	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**
Bipolar Disorder	1.092	1.029 (0.008)	0.976 (0.013)	1.095	1.056 (0.009)	1.010 (0.013)	0.984	0.951 (0.008)	0.650 (0.012)
Coronary Artery	1.074	1.016 (0.008)	0.990 (0.012)	1.079	1.044 (0.009)	1.026 (0.012)	0.988	0.935 (0.009)	0.682 (0.011)
Crohn's Disease	1.047	1.044 (0.009)	1.003 (0.013)	1.052	1.070 (0.009)	1.026 (0.014)	0.992	0.970 (0.008)	0.697 (0.011)
Hypertension	1.062	1.028 (0.009)	0.985 (0.012)	1.052	1.064 (0.008)	1.024 (0.013)	0.966	0.960 (0.009)	0.670 (0.011)
Rheumatoid Arthritis	1.035	1.020 (0.008)	0.995 (0.013)	1.038	1.062 (0.008)	1.034 (0.013)	0.963	0.948 (0.008)	0.683 (0.011)
Type 1 Diabetes	1.029	1.024 (0.010)	0.977 (0.014)	1.070	1.073 (0.009)	1.027 (0.013)	0.945	0.945 (0.007)	0.674 (0.011)
Type 2 Diabetes	1.062	1.026 (0.007)	0.988 (0.012)	1.076	1.037 (0.009)	0.994 (0.013)	0.987	0.953 (0.008)	0.657 (0.010)
Barrett's Oesophagus	1.046	1.036 (0.007)	1.015 (0.010)	1.048	1.034 (0.007)	1.019 (0.010)	1.019	0.984 (0.007)	0.796 (0.010)
Celiac Disease	1.073	1.041 (0.010)	0.980 (0.013)	1.094	1.045 (0.009)	0.984 (0.014)	1.062	1.009 (0.009)	0.829 (0.012)
Ischaemic Stroke	1.071	1.065 (0.006)	1.058 (0.010)	1.074	1.067 (0.007)	1.063 (0.010)	1.375	1.312 (0.009)	1.081 (0.012)
Parkinson's Disease	1.046	1.010 (0.008)	0.998 (0.011)	1.034	1.010 (0.007)	0.996 (0.010)	0.985	0.946 (0.007)	0.751 (0.010)
Psoriasis	1.078	1.054 (0.007)	1.026 (0.010)	1.079	1.074 (0.007)	1.044 (0.010)	1.107	1.078 (0.008)	0.831 (0.011)
Ulcerative Colitis	1.060	1.047 (0.008)	0.997 (0.012)	1.062	1.057 (0.007)	1.006 (0.011)	1.016	0.993 (0.007)	0.761 (0.010)
Average	**1.060**	**1.036 (0.002)**	**1.003 (0.003)**	**1.066**	**1.052 (0.002)**	**1.022 (0.003)**	**1.030**	**0.993 (0.002)**	**0.746 (0.003)**

	Original Controls — h^2_{SNP} (SD)				2 000 POBI Controls — h^2_{SNP} (SD)				2 000 1000G Controls — h^2_{SNP} (SD)			
Trait	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**
Bipolar Disorder	0.18 (0.02)	0.12 (0.03)	0.34 (0.04)	0.47 (0.08)	0.22 (0.02)	0.09 (0.03)	0.43 (0.04)	0.38 (0.08)	0.08 (0.02)	0.18 (0.03)	0.29 (0.03)	3.16 (0.16)
Coronary Artery	0.20 (0.04)	0.14 (0.05)	0.32 (0.06)	0.42 (0.12)	0.31 (0.04)	0.14 (0.05)	0.56 (0.07)	0.33 (0.11)	0.09 (0.04)	0.34 (0.05)	0.30 (0.07)	4.73 (0.24)
Crohn's Disease	0.12 (0.03)	0.03 (0.03)	0.29 (0.04)	0.28 (0.08)	0.17 (0.02)	0.01 (0.03)	0.40 (0.04)	0.26 (0.08)	0.06 (0.02)	0.12 (0.02)	0.28 (0.04)	2.67 (0.14)
Hypertension	0.22 (0.04)	0.12 (0.04)	0.48 (0.06)	0.62 (0.11)	0.26 (0.04)	0.01 (0.04)	0.61 (0.06)	0.40 (0.12)	0.02 (0.03)	0.17 (0.04)	0.41 (0.06)	4.94 (0.25)
Rheumatoid Arthritis	0.06 (0.02)	0.02 (0.02)	0.15 (0.03)	0.18 (0.07)	0.13 (0.02)	-0.01 (0.02)	0.33 (0.03)	0.16 (0.07)	0.00 (0.02)	0.10 (0.02)	0.16 (0.03)	2.55 (0.14)
Type 1 Diabetes	0.09 (0.02)	0.05 (0.03)	0.22 (0.04)	0.33 (0.08)	0.17 (0.03)	0.01 (0.03)	0.38 (0.04)	0.25 (0.07)	-0.01 (0.02)	0.09 (0.02)	0.12 (0.03)	2.52 (0.14)
Type 2 Diabetes	0.27 (0.05)	0.16 (0.05)	0.52 (0.07)	0.65 (0.14)	0.28 (0.04)	0.13 (0.05)	0.64 (0.07)	0.70 (0.16)	0.06 (0.05)	0.23 (0.05)	0.45 (0.08)	6.00 (0.29)
Barrett's Oesophagus	0.16 (0.03)	0.04 (0.03)	0.36 (0.05)	0.24 (0.09)	0.17 (0.03)	0.06 (0.04)	0.36 (0.05)	0.21 (0.09)	0.13 (0.03)	0.18 (0.04)	0.43 (0.05)	2.66 (0.15)
Celiac Disease	0.13 (0.02)	0.07 (0.02)	0.28 (0.03)	0.36 (0.06)	0.14 (0.02)	0.07 (0.02)	0.30 (0.03)	0.36 (0.07)	0.12 (0.02)	0.10 (0.02)	0.29 (0.03)	1.16 (0.08)
Ischaemic Stroke	0.17 (0.02)	0.01 (0.02)	0.37 (0.03)	0.07 (0.06)	0.20 (0.02)	0.04 (0.02)	0.42 (0.03)	0.08 (0.05)	1.25 (0.04)	0.20 (0.03)	2.22 (0.05)	1.62 (0.09)
Parkinson's Disease	0.10 (0.02)	0.08 (0.04)	0.19 (0.03)	0.20 (0.07)	0.10 (0.02)	0.08 (0.03)	0.19 (0.03)	0.21 (0.07)	0.06 (0.03)	0.17 (0.03)	0.19 (0.03)	2.11 (0.11)
Psoriasis	0.17 (0.02)	0.05 (0.02)	0.37 (0.04)	0.22 (0.06)	0.20 (0.02)	0.02 (0.02)	0.45 (0.04)	0.21 (0.07)	0.37 (0.03)	0.16 (0.03)	0.86 (0.04)	2.18 (0.12)
Ulcerative Colitis	0.12 (0.02)	0.04 (0.02)	0.26 (0.03)	0.28 (0.05)	0.13 (0.02)	0.03 (0.02)	0.29 (0.03)	0.27 (0.05)	0.09 (0.02)	0.11 (0.02)	0.28 (0.03)	1.58 (0.08)
Average	**0.14 (0.01)**	**0.05 (0.01)**	**0.29 (0.01)**	**0.28 (0.02)**	**0.17 (0.01)**	**0.04 (0.01)**	**0.36 (0.01)**	**0.25 (0.02)**	**0.11 (0.01)**	**0.14 (0.01)**	**0.39 (0.01)**	**2.11 (0.03)**
Relative	**1**	**0.38 (0.05)**			**1.19 (0.05)**	**0.31 (0.05)**			**0.89 (0.05)**	**0.99 (0.05)**		
Relative	**1.91 (0.10)**	**1**			**2.24 (0.10)**	**0.74 (0.11)**			**1.11 (0.10)**	**1.92 (0.11)**		
Relative			**1**	**0.90 (0.07)**			**1.23 (0.03)**	**0.80 (0.07)**			**1.52 (0.04)**	**7.02 (0.11)**
Relative			**0.88 (0.03)**	**1**			**1.11 (0.03)**	**0.88 (0.07)**			**0.93 (0.03)**	**7.12 (0.11)**

Supplementary Table 10: Introducing population structure for the 13 WTCCC GWAS. For each GWAS, we replace 2 000 randomly picked controls with either 2 000 individuals from People of the British Isles[12] (POBI) or 2 000 non-European individuals from the 1000 Genome Project[1] (1000G). The top table compares estimates of confounding bias, measured using the genomic inflation factor (GIF), the LDSC intercept ($1 + A$) or the SumHer-GC scaling factor (C); the bottom table compares (liability-scale) estimates of h^2_{SNP} from LDSC-Zero, LDSC, SumHer-Zero and SumHer-GC. Note that the values in the first block (the analyses prior to switching out controls) are slightly different from those in Supplementary Table 2, because here we restricted to SNPs common to the POBI and 1000G datasets.

In our original analyses of the 25 raw GWAS (Supplementary Table 2), SumHer-GC found that average confounding bias was slight, whereas LDSC found that average bias was substantial. This analysis confirms that SumHer-GC can report substantial confounding bias (i.e., that it does always find that bias is slight). We note that while SumHer-GC appears to cope well with country-level population structure (that introduced by switching in POBI individuals), it (like LDSC) fares poorly with continental-level structure (switching in 1000G individuals); however, as it is standard to identify and exclude ancestral outliers prior to performing association analysis,[13] in practice such severe confounding is unlikely to be encountered.

	Original Filtering — Confounding (SD)			Excluding Closely Related — Confounding (SD)			Using All Samples — Confounding (SD)		
Trait	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**	**GIF**	**LDSC (SD)**	**SumHer-GC (SD)**
AMD	1.022	1.027 (0.007)	1.017 (0.009)	1.017	1.030 (0.007)	1.027 (0.009)	1.022	1.038 (0.007)	1.031 (0.009)
Heart Failure	1.019	1.036 (0.006)	1.032 (0.008)	1.023	1.039 (0.006)	1.034 (0.008)	1.022	1.041 (0.007)	1.035 (0.008)
Peripheral Arterial	1.043	1.029 (0.007)	1.024 (0.009)	1.053	1.028 (0.007)	1.024 (0.009)	1.059	1.035 (0.007)	1.033 (0.009)
Shingles	1.025	0.996 (0.007)	0.993 (0.009)	1.025	1.005 (0.006)	1.000 (0.008)	1.023	1.009 (0.006)	1.006 (0.008)
VTE	1.035	1.012 (0.008)	0.998 (0.010)	1.033	1.011 (0.008)	0.996 (0.010)	1.034	1.020 (0.008)	1.005 (0.010)
Triglyceride	1.046	1.025 (0.007)	1.002 (0.010)	1.044	1.026 (0.008)	1.004 (0.010)	1.051	1.034 (0.008)	1.013 (0.011)
LDL Cholesterol	1.033	1.026 (0.007)	1.018 (0.009)	1.034	1.035 (0.007)	1.024 (0.009)	1.044	1.054 (0.007)	1.042 (0.009)
HDL Cholesterol	1.061	1.042 (0.008)	1.015 (0.010)	1.068	1.045 (0.009)	1.015 (0.010)	1.085	1.062 (0.009)	1.033 (0.010)
Systolic BP	1.028	1.024 (0.007)	1.010 (0.009)	1.036	1.024 (0.007)	1.007 (0.010)	1.040	1.031 (0.007)	1.016 (0.009)
Diastolic BP	1.018	1.022 (0.006)	1.020 (0.009)	1.021	1.025 (0.006)	1.020 (0.010)	1.026	1.028 (0.006)	1.021 (0.010)
Height	1.103	1.036 (0.009)	0.961 (0.012)	1.110	1.041 (0.009)	0.961 (0.012)	1.147	1.069 (0.009)	0.982 (0.013)
Body Mass Index	1.110	1.040 (0.008)	0.984 (0.010)	1.115	1.043 (0.008)	0.981 (0.010)	1.136	1.060 (0.008)	0.999 (0.010)
Average	**1.045**	**1.026 (0.002)**	**1.009 (0.003)**	**1.048**	**1.028 (0.002)**	**1.011 (0.003)**	**1.057**	**1.037 (0.002)**	**1.020 (0.003)**

	Original Filtering — h^2_{SNP} (SD)				Excluding Closely Related — h^2_{SNP} (SD)				Using All Samples — h^2_{SNP} (SD)			
Trait	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**	**LDSC-Zero**	**LDSC**	**SH-Zero**	**SH-GC**
AMD	0.13 (0.04)	-0.02 (0.07)	0.38 (0.07)	0.16 (0.14)	0.14 (0.04)	-0.01 (0.07)	0.38 (0.07)	0.05 (0.13)	0.18 (0.04)	-0.01 (0.06)	0.45 (0.06)	0.09 (0.12)
Heart Failure	0.10 (0.03)	-0.04 (0.03)	0.26 (0.04)	-0.02 (0.08)	0.10 (0.02)	-0.05 (0.03)	0.26 (0.04)	-0.00 (0.07)	0.11 (0.02)	-0.04 (0.03)	0.28 (0.04)	0.02 (0.07)
Peripheral Arterial	0.17 (0.03)	0.03 (0.04)	0.36 (0.05)	0.10 (0.10)	0.19 (0.03)	0.06 (0.04)	0.38 (0.05)	0.14 (0.10)	0.21 (0.03)	0.05 (0.04)	0.43 (0.05)	0.10 (0.10)
Shingles	0.14 (0.08)	0.19 (0.10)	0.08 (0.13)	0.25 (0.24)	0.13 (0.07)	0.07 (0.09)	0.16 (0.12)	0.16 (0.22)	0.17 (0.07)	0.08 (0.09)	0.24 (0.12)	0.11 (0.21)
VTE	0.12 (0.04)	0.07 (0.05)	0.29 (0.06)	0.31 (0.11)	0.11 (0.03)	0.07 (0.05)	0.26 (0.05)	0.30 (0.10)	0.14 (0.03)	0.06 (0.04)	0.31 (0.05)	0.27 (0.09)
Triglyceride	0.17 (0.03)	0.08 (0.04)	0.29 (0.04)	0.28 (0.09)	0.17 (0.03)	0.09 (0.03)	0.28 (0.04)	0.25 (0.08)	0.19 (0.03)	0.09 (0.03)	0.31 (0.04)	0.23 (0.08)
LDL Cholesterol	0.09 (0.02)	0.02 (0.03)	0.20 (0.03)	0.09 (0.06)	0.10 (0.02)	0.00 (0.02)	0.23 (0.03)	0.09 (0.05)	0.13 (0.02)	-0.01 (0.02)	0.30 (0.03)	0.07 (0.05)
HDL Cholesterol	0.21 (0.03)	0.08 (0.03)	0.39 (0.04)	0.29 (0.06)	0.22 (0.03)	0.09 (0.03)	0.39 (0.04)	0.31 (0.06)	0.24 (0.03)	0.08 (0.03)	0.45 (0.04)	0.26 (0.06)
Systolic BP	0.06 (0.02)	0.00 (0.02)	0.17 (0.03)	0.12 (0.06)	0.06 (0.02)	0.00 (0.02)	0.16 (0.03)	0.12 (0.05)	0.07 (0.02)	0.00 (0.02)	0.18 (0.02)	0.10 (0.05)
Diastolic BP	0.06 (0.02)	0.01 (0.03)	0.17 (0.03)	0.06 (0.05)	0.07 (0.02)	0.01 (0.02)	0.19 (0.03)	0.08 (0.05)	0.08 (0.02)	0.02 (0.02)	0.20 (0.03)	0.10 (0.05)
Height	0.31 (0.03)	0.22 (0.04)	0.48 (0.03)	0.68 (0.08)	0.31 (0.03)	0.22 (0.04)	0.47 (0.03)	0.67 (0.08)	0.37 (0.03)	0.22 (0.04)	0.58 (0.03)	0.67 (0.07)
Body Mass Index	0.26 (0.02)	0.17 (0.02)	0.40 (0.03)	0.48 (0.06)	0.25 (0.02)	0.16 (0.02)	0.39 (0.03)	0.48 (0.06)	0.29 (0.02)	0.16 (0.02)	0.45 (0.03)	0.45 (0.05)
Average	**0.14 (0.01)**	**0.06 (0.01)**	**0.29 (0.01)**	**0.22 (0.02)**	**0.14 (0.01)**	**0.05 (0.01)**	**0.29 (0.01)**	**0.22 (0.02)**	**0.16 (0.01)**	**0.05 (0.01)**	**0.34 (0.01)**	**0.21 (0.02)**
Relative	**1**	**0.51 (0.06)**			**1.00 (0.04)**	**0.50 (0.05)**			**1.16 (0.04)**	**0.48 (0.05)**		
Relative	**1.52 (0.08)**	**1**			**1.49 (0.07)**	**0.96 (0.09)**			**1.72 (0.07)**	**0.94 (0.09)**		
Relative			**1**	**0.86 (0.07)**			**1.00 (0.03)**	**0.85 (0.07)**			**1.16 (0.03)**	**0.81 (0.06)**
Relative			**0.87 (0.03)**	**1**			**0.86 (0.03)**	**0.99 (0.07)**			**1.02 (0.03)**	**0.94 (0.07)**

Supplementary Table 11: Introducing relatedness for the 12 eMERGE GWAS. For our original analyses of the eMERGE data, we restricted to 25 875 unrelated Caucasian individuals (obtained by filtering the Caucasian individuals so that no pair remained with allelic correlation[14] $\geq$0.05). Here we repeat the analysis twice, first restricting to 27 575 Caucasian individuals (obtained by filtering so that no pair of individuals remains with allelic correlation $\geq$0.175), then using all 28 803 Caucasian individuals. As shown in Supplementary Figure 5, the latter results in the introduction of approximately 1 650 pairs of relatives. The top table compares estimates of confounding bias, measured using the genomic inflation factor (GIF), the LDSC intercept ($1 + A$) or the SumHer-GC scaling factor (C); the bottom table compares estimates of h^2_{SNP} from LDSC-Zero, LDSC, SumHer-Zero and SumHer-GC (for binary traits, estimates are on the liability scale).

Just as the analyses in Supplementary Table 10 demonstrate that SumHer-GC will detect confounding bias in the presence of population structure, these analyses show that SumHer-GC will detect bias when there is familial relatedness. As it is standard to identify and exclude closely related individuals (e.g., twins, full-sibs and half-sibs) prior to performing association analysis,[13] these analyses also indicate that for a carefully-performed GWAS, confounding bias due to relatedness is likely to be slight.

	SNPs $>$ 1 Mb or $r^2_{jl}<0.2$				SNPs $>$ 1 Mb or $r^2_{jl}<0.1$				SNPs $>$ 3 Mb or $r^2_{jl}<0.1$			
Trait	**Raw**	**GC**	**LDSC**	**SumHer-GC**	**Raw**	**GC**	**LDSC**	**SumHer-GC**	**Raw**	**GC**	**LDSC**	**SumHer-GC**
Alzheimer's Disease	25	22	22	25	21	19	19	21	21	19	19	21
Coronary Artery Disease	10	6	7	10	10	6	7	10	10	6	7	10
Crohn's Disease	76	62	69	77	64	52	58	64	64	52	58	64
Ever Smoked?	0	0	0	0	0	0	0	0	0	0	0	0
Inflammatory Bowel	90	69	74	92	78	59	65	80	78	59	65	80
Rheumatoid Arthritis	163	157	163	181	109	104	109	123	109	104	109	123
Schizophrenia	117	26	68	159	105	23	63	140	104	22	62	138
Type 2 Diabetes	43	30	37	47	38	25	32	42	38	25	32	42
Ulcerative Colitis	47	38	38	47	38	31	31	38	38	31	31	38
Bone Mineral Density	24	21	23	24	19	18	18	19	19	18	18	19
Body Mass Index	81	59	151	378	69	52	135	336	69	52	135	336
Depressive Symptoms	0	0	0	1	0	0	0	1	0	0	0	1
Fasting Glucose	28	25	25	29	22	20	20	23	22	20	20	23
Glycated Hemoglobin	13	13	13	13	10	10	10	10	10	10	10	10
HDL Cholesterol	158	151	149	261	130	122	121	216	130	122	121	217
Height	895	239	356	935	720	196	288	754	720	196	288	754
LDL Cholesterol	133	128	133	201	101	96	101	155	101	96	101	155
Menarche Age	345	121	223	425	289	111	190	354	289	111	190	354
Menopause Age	54	40	42	62	49	39	39	55	49	39	39	55
Neuroticism	11	4	8	19	10	4	7	18	8	3	5	15
Subjective Well-Being	0	0	0	0	0	0	0	0	0	0	0	0
Triglyceride	114	111	123	188	82	82	91	152	81	81	90	150
Waist-Hip Ratio	28	25	35	69	26	23	33	66	26	23	33	66
Years Education	80	14	53	162	70	13	46	148	70	13	46	148
Total	**2535**	**1361**	**1812**	**3405**	**2060**	**1105**	**1483**	**2825**	**2056**	**1102**	**1479**	**2819**
Average	**105.6**	**56.7**	**75.5**	**141.9**	**85.8**	**46.0**	**61.8**	**117.7**	**85.7**	**45.9**	**61.6**	**117.5**
Relative	**1**	**0.54**	**0.71**	**1.34**								
Relative					**1**	**0.54**	**0.72**	**1.37**				
Relative									**1**	**0.54**	**0.72**	**1.37**

Supplementary Table 12: Number of significant loci for the 24 summary GWAS after correction for confounding bias. Values report the number of independent loci with $P < 5 \times 10^{-8}$ (equivalently, $\chi^2(1)$ test statistic > 29.72), either based on the reported test statistics, or after correction using genomic control (dividing test statistics by the genomic inflation factor), LDSC (dividing them by the intercept, $1 + A$) or SumHer-GC (dividing them by the scaling factor, C). We consider increasingly strict definitions of independent: first we define two SNPs as independent if they are either $>$1 cM apart or have correlation squared $>$0.2; then as independent if either $>$1 cM apart or have correlation squared $>$0.1, and finally as independent if either $>$3 cM apart or have correlation squared $>$0.1 (for all GWAS, we estimate correlations between SNPs using our reference panel).

	log likelihood $L(S\|\widehat{h^2_{\text{SNP}}}, D^0)$			Confounding, $1+A$ or C (SD)		h^2_{SNP} (SD)			
Trait (Disease Prevalence, %)	**SH-Zero**	**SH-Cept**	**SH-GC**	**SH-CEPT**	**SH-GC**	**SH-Zero**	**SH-CEPT**	**SH-GC**	**24 Raw**
Alzheimer's Disease (7.5)	-198436	-198434	-198434	1.026 (0.013)	1.026 (0.013)	0.17 (0.02)	0.12 (0.03)	0.12 (0.03)	
Coronary Artery Disease (6)	-154977	-154977	-154977	0.987 (0.013)	0.987 (0.013)	0.13 (0.01)	0.15 (0.02)	0.15 (0.02)	0.37 (0.07)
Crohn's Disease (0.5)	-198886	-198883	-198883	0.971 (0.016)	0.971 (0.016)	0.39 (0.03)	0.46 (0.05)	0.47 (0.06)	0.34 (0.04)
Ever Smoked? (56)	-148860	-148857	-148857	0.964 (0.011)	0.964 (0.011)	0.13 (0.01)	0.19 (0.02)	0.19 (0.02)	
Inflammatory Bowel (0.7)	-213135	-213135	-213135	0.982 (0.015)	0.982 (0.015)	0.30 (0.02)	0.32 (0.03)	0.33 (0.03)	
Rheumatoid Arthritis (0.5)	-201067	-201041	-201041	0.904 (0.017)	0.904 (0.017)	0.07 (0.01)	0.16 (0.03)	0.17 (0.03)	0.16 (0.04)
Schizophrenia (1)	-240313	-240272	-240272	0.914 (0.014)	0.914 (0.014)	0.33 (0.01)	0.39 (0.01)	0.42 (0.02)	
Type 2 Diabetes (8)	-213748	-213743	-213742	0.947 (0.014)	0.947 (0.013)	0.16 (0.01)	0.21 (0.02)	0.23 (0.02)	0.61 (0.07)
Ulcerative Colitis (0.2)	-203378	-203378	-203378	0.993 (0.013)	0.993 (0.013)	0.25 (0.02)	0.27 (0.03)	0.27 (0.03)	0.29 (0.03)
Bone Mineral Density	-185834	-185834	-185834	1.004 (0.011)	1.004 (0.011)	0.29 (0.02)	0.28 (0.04)	0.28 (0.04)	
Body Mass Index	-198606	-198385	-198291	0.581 (0.016)	0.551 (0.016)	0.05 (0.01)	0.18 (0.01)	0.33 (0.03)	0.40 (0.03)
Depressive Symptoms	-164947	-164943	-164943	0.957 (0.012)	0.957 (0.012)	0.05 (0.00)	0.07 (0.01)	0.07 (0.01)	
Fasting Glucose	-191499	-191499	-191499	0.987 (0.014)	0.987 (0.014)	0.12 (0.02)	0.13 (0.03)	0.14 (0.03)	
Glycated Hemoglobin	-152262	-152262	-152262	0.987 (0.012)	0.987 (0.012)	0.08 (0.01)	0.10 (0.02)	0.10 (0.02)	
HDL Cholesterol	-255266	-255260	-255217	0.888 (0.052)	0.685 (0.026)	0.12 (0.03)	0.20 (0.03)	0.50 (0.09)	0.39 (0.04)
Height	-263932	-263930	-263929	0.988 (0.046)	0.981 (0.038)	0.44 (0.02)	0.45 (0.03)	0.46 (0.04)	0.48 (0.03)
LDL Cholesterol	-237414	-237406	-237364	0.863 (0.036)	0.727 (0.040)	0.11 (0.03)	0.22 (0.04)	0.43 (0.10)	0.20 (0.03)
Menarche Age	-266406	-266378	-266378	0.894 (0.021)	0.894 (0.021)	0.25 (0.01)	0.29 (0.01)	0.32 (0.02)	
Menopause Age	-179168	-179158	-179158	0.919 (0.017)	0.919 (0.017)	0.15 (0.01)	0.23 (0.03)	0.25 (0.03)	
Neuroticism	-175327	-175295	-175295	0.898 (0.022)	0.898 (0.022)	0.10 (0.01)	0.15 (0.02)	0.17 (0.02)	
Subjective Well-Being	-142252	-142248	-142248	0.965 (0.015)	0.965 (0.015)	0.03 (0.00)	0.04 (0.00)	0.04 (0.00)	
Triglyceride	-235250	-235231	-235174	0.825 (0.044)	0.700 (0.041)	0.10 (0.03)	0.24 (0.05)	0.45 (0.11)	0.29 (0.04)
Waist-Hip Ratio	-154484	-154364	-154277	0.792 (0.014)	0.757 (0.014)	0.03 (0.01)	0.14 (0.01)	0.20 (0.02)	
Years Education	-214581	-214485	-214485	0.828 (0.015)	0.828 (0.015)	0.12 (0.00)	0.16 (0.01)	0.20 (0.01)	
Average	**-199584**	**-199558**	**-199545**	**0.936 (0.003)**	**0.929 (0.003)**	**0.07 (0.00)**	**0.12 (0.00)**	**0.12 (0.00)**	
Eight-Trait Average	**-214209**	**-214162**	**-214121**						

Supplementary Table 13: Empirical support for modeling inflation due to confounding as multiplicative. SumHer allows the user to specify not only the heritability model, but also the model used for estimating confounding bias; whereas LDSC assumes that confounding inflation is additive (using the model $\mathbb{E}[S_j] = 1 + an_j + n_j v_j^2$, where S_j, n_j and v_j^2 are the test statistic, sample size and heritability tagged by SNP j), we prefer to assume it is multiplicative (using the model $\mathbb{E}[S_j] = C(1 + n_j v_j^2)$). While it is straightforward to compare different heritability models based on likelihood (Supplementary Table 7), it is more difficult to test different confounding models; this is because when n_j is constant (as is typically the case), whether we model confounding inflation as additive or multiplicative does not affect the model likelihood (to appreciate why, note that $C(1 + n_j v_j^2) = 1 + A + C n_j v_j^2$, where $A = an_j = C - 1$).

Here we empirically assess the two confounding models in two ways. First, we compare model likelihoods for the eight traits (those marked in red) for which n_j does vary across SNPs; on average the log likelihood is 41 nats higher if we assume multiplicative inflation (SumHer-GC) rather than additive inflation (SumHer-CEPT). Second, we compare estimates of h^2_{SNP} for the ten traits also present in the 25 raw GWAS (for the five binary traits, estimates have been converted to the liability scale assuming the stated prevalence[40,41]). If we assume additive inflation, 6 out of 10 pairs of estimates are consistent ($P > 0.05/10$) with the corresponding estimate from the 25 raw GWAS (obtained using SumHer-Zero), but if we assume multiplicative inflation, 8 out of 10 pairs of estimates are consistent (consistent estimates are marked in red).

While both assessments support using the multiplicative model, we realize that for the 24 summary GWAS, most of the confounding is likely due to genomic control (evidenced by the many estimates of confounding less than one), which by definition affects test statistics multiplicatively. Therefore, it remains uncertain whether inflation due to other causes (i.e., population structure or relatedness) is best modeled as additive or multiplicative. Also, for the second analysis, there are phenotypic differences for the GWAS in common (for example, for the three GWAS by the Global Lipids Genetics Consortium,[51] individuals on lipid-lowering medication were excluded, but this was not possible when analyzing the eMERGE data), and therefore we should expect some differences in h^2_{SNP}. Finally, we remind readers that (when n_j is constant) estimates of h^2_{SNP} using an additive model of inflation will be $C = 1 + A$ higher than those using a multiplicative model. Therefore, when confounding is slight ($C = 1 + A$ is close to 1) the choice of confounding model is no longer important.

	LDSC (53-part model)			SumHer-GC (25-part model)		
Category	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**
Coding	0.098 (0.010)	0.015	6.13 (0.66)	0.030 (0.002)	0.018	1.74 (0.09)
Conserved	0.323 (0.020)	0.033	9.39 (0.60)	0.071 (0.002)	0.036	1.95 (0.07)
CTCF	-0.024 (0.015)	0.022	-1.10 (0.68)	0.027 (0.002)	0.027	1.01 (0.06)
Digital Genomic Footprint	0.312 (0.037)	0.142	2.20 (0.26)	0.203 (0.004)	0.165	1.26 (0.03)
DNase I Hypersensitive Site	0.257 (0.038)	0.189	1.32 (0.20)	0.245 (0.004)	0.214	1.18 (0.02)
FANTOM5 Enhancer	0.005 (0.006)	0.004	1.45 (1.61)	0.005 (0.001)	0.005	0.86 (0.13)
Enhancer	0.109 (0.016)	0.039	2.84 (0.42)	0.066 (0.002)	0.050	1.34 (0.05)
Fetal DHS	0.166 (0.033)	0.094	1.76 (0.36)	0.135 (0.003)	0.111	1.27 (0.03)
H3K27ac (Hnisz)	0.599 (0.015)	0.375	1.61 (0.04)	0.525 (0.005)	0.446	1.18 (0.01)
H3K27ac (PGC2)	0.466 (0.027)	0.262	1.81 (0.10)	0.381 (0.005)	0.310	1.24 (0.02)
H3K4me1	0.778 (0.030)	0.443	1.76 (0.07)	0.592 (0.005)	0.503	1.18 (0.01)
H3K4me3	0.328 (0.023)	0.125	2.68 (0.19)	0.207 (0.004)	0.148	1.41 (0.03)
H3K9ac	0.324 (0.023)	0.117	2.82 (0.20)	0.221 (0.004)	0.148	1.52 (0.03)
Intronic	0.447 (0.013)	0.403	1.12 (0.03)	0.462 (0.004)	0.408	1.13 (0.01)
Promoter Flanking	0.001 (0.009)	0.008	0.18 (1.10)	0.017 (0.001)	0.009	1.84 (0.12)
Promoter	0.052 (0.010)	0.026	1.94 (0.39)	0.045 (0.002)	0.031	1.48 (0.06)
Repressed	0.322 (0.033)	0.476	0.68 (0.07)	0.353 (0.005)	0.451	0.78 (0.01)
Super Enhancer	0.255 (0.010)	0.150	1.71 (0.07)	0.232 (0.003)	0.197	1.17 (0.02)
Transcription Factor Binding Site	0.324 (0.034)	0.130	2.50 (0.26)	0.196 (0.004)	0.155	1.29 (0.03)
Transcribed	0.413 (0.030)	0.351	1.17 (0.09)	0.405 (0.005)	0.339	1.19 (0.01)
Transcription Start Site	0.052 (0.010)	0.013	3.91 (0.75)	0.036 (0.002)	0.018	1.97 (0.09)
3' Untranslated Region	0.053 (0.008)	0.011	4.80 (0.70)	0.022 (0.001)	0.014	1.64 (0.09)
5' Untranslated Region	0.022 (0.006)	0.005	3.95 (1.16)	0.007 (0.001)	0.006	1.27 (0.13)
Weak Enhancer	0.074 (0.015)	0.020	3.82 (0.77)	0.033 (0.002)	0.027	1.25 (0.06)

Supplementary Table 14: Average estimates of functional enrichments across the 24 summary GWAS. We estimate enrichments using either LDSC with a 53-part model or SumHer-GC with a 25-part model. For the 25-part model, we divide the genome into a set for each of the 24 categories, plus a set containing all SNPs, while for the 53-part model, we divide the genome into a set for each of the 24 categories, a set for each of 28 buffer regions (see Finucane *et al.*,[10]), plus a set containing all SNPs. For each trait, we calculate the estimated enrichment of each category, obtained by dividing the estimated share of h^2_{SNP} contributed by the category by its expected share; for each category, we then report the (inverse-variance-weighted) average estimated enrichment across the 24 traits. Average estimated enrichments significantly different from one ($P < 0.05$) are marked in red.

This table highlights that the differences between LDSC and SumHer-GC estimates of enrichments are primarily due to differences between the estimated shares of h^2_{SNP} contributed by each category. This is most noticeable for coding regions (9.8% versus 3.0%) and conserved regions (32% versus 7.1%). It is not obvious to us why these differences occur, however, we suspect that when assuming the GCTA Model, the estimated shares of h^2_{SNP} for some categories reflect not the amount of heritability the categories contribute, but rather the amount of heritability the categories tag.

	LDSC	Finucane *et al.* (MAF ≥0.05 SNPS)			Finucane *et al.* (All SNPS)			SumHer-GC
Category	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Av. Enrich. (SD)**
Coding	6.13 (0.66)	0.089 (0.007)	0.014	6.22 (0.50)	0.102 (0.008)	0.016	6.36 (0.48)	1.74 (0.09)
Conserved	9.39 (0.60)	0.303 (0.014)	0.026	11.81 (0.53)	0.328 (0.014)	0.029	11.49 (0.49)	1.95 (0.07)
CTCF	-1.10 (0.68)	0.001 (0.011)	0.024	0.03 (0.44)	0.002 (0.010)	0.024	0.09 (0.42)	1.01 (0.06)
Digital Gen. Footprint	2.20 (0.26)	0.291 (0.027)	0.136	2.14 (0.20)	0.299 (0.026)	0.138	2.16 (0.19)	1.26 (0.03)
DNase I Hyper. Site	1.32 (0.20)	0.235 (0.017)	0.125	1.90 (0.14)	0.244 (0.017)	0.127	1.94 (0.14)	1.18 (0.02)
FANTOM5 Enhancer	1.45 (1.61)	0.010 (0.005)	0.004	2.34 (1.08)	0.011 (0.005)	0.004	2.40 (1.03)	0.86 (0.13)
Enhancer	2.84 (0.42)	0.089 (0.008)	0.031	2.94 (0.24)	0.090 (0.007)	0.032	2.93 (0.23)	1.34 (0.05)
Fetal DHS	1.76 (0.36)	0.222 (0.021)	0.084	2.65 (0.26)	0.230 (0.021)	0.086	2.68 (0.24)	1.27 (0.03)
H3K27ac (Hnisz)	1.61 (0.04)	0.600 (0.010)	0.389	1.54 (0.02)	0.608 (0.009)	0.393	1.55 (0.02)	1.18 (0.01)
H3K27ac (PGC2)	1.81 (0.10)	0.470 (0.018)	0.269	1.75 (0.07)	0.479 (0.017)	0.273	1.76 (0.06)	1.24 (0.02)
H3K4me1	1.76 (0.07)	0.760 (0.020)	0.424	1.79 (0.05)	0.766 (0.019)	0.429	1.79 (0.04)	1.18 (0.01)
H3K4me3	2.68 (0.19)	0.336 (0.015)	0.133	2.53 (0.11)	0.347 (0.015)	0.137	2.53 (0.11)	1.41 (0.03)
H3K9ac	2.82 (0.20)	0.347 (0.015)	0.125	2.77 (0.12)	0.357 (0.015)	0.129	2.77 (0.12)	1.52 (0.03)
Intronic	1.12 (0.03)	0.455 (0.008)	0.387	1.17 (0.02)	0.456 (0.008)	0.394	1.16 (0.02)	1.13 (0.01)
Promoter Flanking	0.18 (1.10)	0.001 (0.006)	0.008	0.13 (0.77)	0.002 (0.006)	0.009	0.23 (0.74)	1.84 (0.12)
Promoter	1.94 (0.39)	0.112 (0.010)	0.046	2.42 (0.21)	0.119 (0.010)	0.048	2.48 (0.20)	1.48 (0.06)
Repressed	0.68 (0.07)	0.266 (0.021)	0.461	0.58 (0.05)	0.257 (0.020)	0.453	0.57 (0.04)	0.78 (0.01)
Super Enhancer	1.71 (0.07)	0.280 (0.006)	0.167	1.67 (0.04)	0.286 (0.006)	0.170	1.68 (0.03)	1.17 (0.02)
T. Factor Binding Site	2.50 (0.26)	0.352 (0.023)	0.131	2.68 (0.17)	0.358 (0.022)	0.133	2.69 (0.16)	1.29 (0.03)
Transcribed	1.17 (0.09)	0.415 (0.020)	0.346	1.20 (0.06)	0.424 (0.019)	0.353	1.20 (0.05)	1.19 (0.01)
Transcription Start Site	3.91 (0.75)	0.090 (0.008)	0.018	5.05 (0.45)	0.094 (0.008)	0.019	5.03 (0.43)	1.97 (0.09)
3' Untranslated Region	4.80 (0.70)	0.039 (0.005)	0.011	3.49 (0.48)	0.043 (0.005)	0.012	3.61 (0.45)	1.64 (0.09)
5' Untranslated Region	3.95 (1.16)	0.027 (0.004)	0.005	4.88 (0.78)	0.030 (0.004)	0.006	5.04 (0.75)	1.27 (0.13)
Weak Enhancer	3.82 (0.77)	0.058 (0.010)	0.021	2.76 (0.50)	0.058 (0.010)	0.021	2.72 (0.47)	1.25 (0.06)

Supplementary Table 15: Average estimates of functional enrichments across the 24 summary GWAS from the LDSC software following the recommendations of Finucane *et al.* For each category, values report, averaged across the 24 summary GWAS, its estimated share of h^2_{SNP}, its expected share and its estimated enrichment, obtained using the LDSC software. We follow the recommendations of Finucane *et al.*[10] Therefore, we use a 53-part model (one part for each category, one for each of 28 buffer regions, and one containing all SNPs), compute LD scores using the supplied reference panel (489 European individuals from the 1000 Genomes Project[1]), but for the regression use only HapMap 3 SNPs (for Columns 3-5, we restrict to those with MAF ≥0.05, for Columns 6-8, we omit this restriction by adding `--not-M-5-50`). For comparison, we also report average estimates of enrichments from LDSC (implemented within SumHer) and from our recommended method, SumHer-GC. Average estimated enrichments significantly different from one ($P < 0.05$) are marked in red.

	LDSC	Gazal *et al.* (MAF $\geq$0.05 SNPS)			Gazal *et al.* (All SNPS)			SumHer-GC
Category	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Av. Enrich. (SD)**
Coding	6.13 (0.66)	0.065 (0.006)	0.014	4.54 (0.44)	0.099 (0.010)	0.016	6.13 (0.60)	1.74 (0.09)
Conserved	9.39 (0.60)	0.187 (0.012)	0.026	7.29 (0.46)	0.262 (0.017)	0.029	9.18 (0.60)	1.95 (0.07)
CTCF	-1.10 (0.68)	0.003 (0.010)	0.024	0.11 (0.40)	-0.002 (0.013)	0.024	-0.07 (0.53)	1.01 (0.06)
Digital Gen. Footprint	2.20 (0.26)	0.226 (0.024)	0.136	1.66 (0.17)	0.264 (0.032)	0.138	1.91 (0.23)	1.26 (0.03)
DNase I Hyper. Site	1.32 (0.20)	0.183 (0.015)	0.125	1.49 (0.13)	0.216 (0.021)	0.127	1.71 (0.17)	1.18 (0.02)
FANTOM5 Enhancer	1.45 (1.61)	0.008 (0.004)	0.004	1.91 (0.96)	0.009 (0.006)	0.004	2.06 (1.28)	0.86 (0.13)
Enhancer	2.84 (0.42)	0.070 (0.007)	0.031	2.33 (0.21)	0.086 (0.009)	0.032	2.86 (0.28)	1.34 (0.05)
Fetal DHS	1.76 (0.36)	0.168 (0.019)	0.084	2.00 (0.23)	0.203 (0.026)	0.086	2.37 (0.30)	1.27 (0.03)
H3K27ac (Hnisz)	1.61 (0.04)	0.543 (0.009)	0.389	1.40 (0.02)	0.601 (0.011)	0.393	1.53 (0.03)	1.18 (0.01)
H3K27ac (PGC2)	1.81 (0.10)	0.414 (0.016)	0.269	1.54 (0.06)	0.480 (0.022)	0.273	1.76 (0.08)	1.24 (0.02)
H3K4me1	1.76 (0.07)	0.663 (0.018)	0.424	1.57 (0.04)	0.760 (0.025)	0.429	1.77 (0.06)	1.18 (0.01)
H3K4me3	2.68 (0.19)	0.274 (0.013)	0.133	2.06 (0.10)	0.337 (0.018)	0.137	2.46 (0.13)	1.41 (0.03)
H3K9ac	2.82 (0.20)	0.270 (0.014)	0.125	2.15 (0.11)	0.335 (0.019)	0.129	2.60 (0.15)	1.52 (0.03)
Intronic	1.12 (0.03)	0.445 (0.007)	0.387	1.15 (0.02)	0.457 (0.010)	0.394	1.16 (0.03)	1.13 (0.01)
Promoter Flanking	0.18 (1.10)	0.004 (0.006)	0.008	0.52 (0.69)	0.004 (0.008)	0.009	0.41 (0.89)	1.84 (0.12)
Promoter	1.94 (0.39)	0.079 (0.009)	0.046	1.70 (0.19)	0.097 (0.012)	0.048	2.01 (0.26)	1.48 (0.06)
Repressed	0.68 (0.07)	0.308 (0.019)	0.461	0.67 (0.04)	0.253 (0.025)	0.453	0.56 (0.05)	0.78 (0.01)
Super Enhancer	1.71 (0.07)	0.256 (0.005)	0.167	1.53 (0.03)	0.285 (0.007)	0.170	1.68 (0.04)	1.17 (0.02)
T. Factor Binding Site	2.50 (0.26)	0.268 (0.021)	0.131	2.05 (0.16)	0.326 (0.029)	0.133	2.45 (0.22)	1.29 (0.03)
Transcribed	1.17 (0.09)	0.425 (0.018)	0.346	1.23 (0.05)	0.455 (0.023)	0.353	1.29 (0.06)	1.19 (0.01)
Transcription Start Site	3.91 (0.75)	0.057 (0.007)	0.018	3.22 (0.42)	0.073 (0.010)	0.019	3.88 (0.55)	1.97 (0.09)
3' Untranslated Region	4.80 (0.70)	0.029 (0.005)	0.011	2.63 (0.42)	0.039 (0.007)	0.012	3.25 (0.55)	1.64 (0.09)
5' Untranslated Region	3.95 (1.16)	0.016 (0.004)	0.005	2.86 (0.68)	0.022 (0.005)	0.006	3.67 (0.89)	1.27 (0.13)
Weak Enhancer	3.82 (0.77)	0.041 (0.009)	0.021	1.98 (0.44)	0.050 (0.012)	0.021	2.33 (0.58)	1.25 (0.06)

Supplementary Table 16: Average estimates of functional enrichments across the 24 summary GWAS from the LDSC software following the recommendations of Gazal *et al.* Details are the same as Supplementary Table 15, except when using the LDSC software we follow the recommendations of Gazal *et al.*[23] and use a 75-part model (this is constructed by adding to the 53-part model used by Finucane *et al.*,[10] 3 more functional annotations, 3 extra buffers, 10 MAF tranches and 6 continuous LD-related annotations).

Trait 1	Trait 2	LDSC Correlation (SD)	SumHer-GC Correlation (SD)
Alzheimer's Disease	Years Education	-0.24 (0.08)	**-0.21 (0.06)**
Body Mass Index	Depressive Symptoms	**0.19 (0.05)**	0.10 (0.04)
Body Mass Index	Ever Smoked?	**0.20 (0.04)**	**0.14 (0.04)**
Body Mass Index	Fasting Glucose	**0.33 (0.06)**	**0.24 (0.05)**
Body Mass Index	HDL Cholesterol	-0.48 (0.16)	**-0.22 (0.03)**
Body Mass Index	Menarche Age	**-0.37 (0.03)**	**-0.34 (0.02)**
Body Mass Index	Schizophrenia	**-0.10 (0.03)**	-0.07 (0.02)
Body Mass Index	Triglyceride	**0.24 (0.06)**	**0.17 (0.04)**
Body Mass Index	Type 2 Diabetes	**0.55 (0.05)**	**0.40 (0.04)**
Body Mass Index	Waist-Hip Ratio	**0.66 (0.04)**	**0.53 (0.03)**
Body Mass Index	Years Education	**-0.28 (0.02)**	**-0.24 (0.02)**
Coronary Artery Disease	HDL Cholesterol	-0.49 (0.15)	**-0.27 (0.05)**
Coronary Artery Disease	LDL Cholesterol	0.22 (0.09)	**0.30 (0.05)**
Coronary Artery Disease	Triglyceride	**0.42 (0.07)**	**0.33 (0.05)**
Coronary Artery Disease	Type 2 Diabetes	**0.51 (0.07)**	**0.39 (0.06)**
Coronary Artery Disease	Waist-Hip Ratio	**0.28 (0.07)**	**0.24 (0.05)**
Coronary Artery Disease	Years Education	**-0.26 (0.05)**	**-0.26 (0.05)**
Crohn's Disease	Inflammatory Bowel	**0.92 (0.02)**	**0.87 (0.02)**
Crohn's Disease	Ulcerative Colitis	**0.65 (0.08)**	**0.60 (0.05)**
Depressive Symptoms	Neuroticism	**0.79 (0.04)**	**0.72 (0.04)**
Depressive Symptoms	Schizophrenia	**0.27 (0.05)**	**0.27 (0.04)**
Depressive Symptoms	Subjective Well-Being	**-0.80 (0.07)**	**-0.81 (0.08)**
Depressive Symptoms	Waist-Hip Ratio	**0.23 (0.06)**	0.16 (0.05)
Depressive Symptoms	Years Education	**-0.34 (0.05)**	**-0.38 (0.05)**
Ever Smoked?	Years Education	**-0.33 (0.05)**	**-0.28 (0.04)**
Fasting Glucose	Type 2 Diabetes	**0.60 (0.10)**	**0.57 (0.09)**
Glycated Hemoglobin	Type 2 Diabetes	**0.70 (0.14)**	**0.57 (0.09)**
Glycated Hemoglobin	Waist-Hip Ratio	**0.43 (0.11)**	0.24 (0.07)
HDL Cholesterol	Menarche Age	0.20 (0.06)	**0.13 (0.02)**
HDL Cholesterol	Triglyceride	**-1.02 (0.19)**	**-0.52 (0.08)**
HDL Cholesterol	Type 2 Diabetes	**-0.50 (0.13)**	**-0.33 (0.04)**
HDL Cholesterol	Waist-Hip Ratio	-0.73 (0.21)	**-0.35 (0.04)**
HDL Cholesterol	Years Education	**0.25 (0.06)**	**0.15 (0.03)**
Height	Menarche Age	**0.14 (0.03)**	**0.12 (0.02)**
Height	Triglyceride	**-0.12 (0.03)**	**-0.08 (0.02)**
Height	Years Education	**0.13 (0.02)**	**0.12 (0.02)**
Inflammatory Bowel	Ulcerative Colitis	**0.91 (0.03)**	**0.91 (0.02)**
LDL Cholesterol	Triglyceride	**0.49 (0.11)**	**0.38 (0.05)**
Menarche Age	Triglyceride	**-0.12 (0.03)**	-0.10 (0.03)
Menarche Age	Type 2 Diabetes	**-0.20 (0.03)**	**-0.17 (0.03)**
Menarche Age	Waist-Hip Ratio	**-0.24 (0.04)**	**-0.17 (0.03)**
Menopause Age	Years Education	**0.20 (0.05)**	**0.13 (0.03)**
Neuroticism	Subjective Well-Being	**-0.72 (0.05)**	**-0.74 (0.05)**
Neuroticism	Years Education	**-0.26 (0.04)**	**-0.27 (0.04)**
Rheumatoid Arthritis	Years Education	**-0.26 (0.05)**	**-0.24 (0.04)**
Schizophrenia	Subjective Well-Being	**-0.29 (0.05)**	**-0.33 (0.05)**
Schizophrenia	Years Education	**0.13 (0.03)**	**0.08 (0.02)**
Triglyceride	Type 2 Diabetes	**0.33 (0.06)**	**0.36 (0.06)**
Triglyceride	Waist-Hip Ratio	**0.49 (0.07)**	**0.38 (0.05)**
Triglyceride	Years Education	**-0.19 (0.04)**	**-0.14 (0.03)**
Type 2 Diabetes	Waist-Hip Ratio	**0.61 (0.06)**	**0.50 (0.05)**
Type 2 Diabetes	Years Education	**-0.21 (0.04)**	**-0.18 (0.03)**
Waist-Hip Ratio	Years Education	**-0.35 (0.03)**	**-0.30 (0.03)**

Supplementary Table 17: Estimates of genetic correlation for the 24 summary GWAS. Of the ${}^{24}C_2 = 276$ pairs of traits we considered, this table reports estimates of genetic correlation for the 53 pairs with significant correlation ($P < 0.05/276$) from either LDSC or SumHer-GC. Nominally significant estimates ($P < 0.05$) are marked in red, while Bonferonni significant estimates ($P < 0.05/276$) are also in **bold**.

Polygenic Risk Score	h^2_{SNP}	Clump	BMI	Height	HDL	LDL	TG	Mean (SD)	Relative
Classical		YES	0.250	0.273	0.181	0.059	0.166	**0.198 (0.003)**	**0.991 (0.016)**
Classical		NO	0.242	0.283	0.177	0.064	0.178	**0.200 (0.003)**	1
Bayesian: GCTA Model	$\widehat{h^2_{SNP}}$	YES	0.235	0.278	0.185	0.070	0.196	**0.202 (0.003)**	**0.999 (0.016)**
Bayesian: GCTA Model	$\widehat{h^2_{SNP}}$	NO	0.191	0.264	0.128	0.064	0.161	**0.169 (0.003)**	**0.855 (0.016)**
Bayesian: GCTA Model	0.5	YES	0.249	0.279	0.187	0.062	0.186	**0.204 (0.003)**	**1.016 (0.016)**
Bayesian: GCTA Model	0.5	NO	0.227	0.267	0.157	0.063	0.171	**0.187 (0.003)**	**0.936 (0.016)**
Bayesian: GCTA Model	Adjusted	YES	0.239	0.276	0.185	0.070	0.195	**0.201 (0.003)**	**1.001 (0.016)**
Bayesian: GCTA Model	Adjusted	NO	0.201	0.252	0.123	0.064	0.163	**0.169 (0.003)**	**0.850 (0.016)**
Bayesian: Enriched GCTA	$\widehat{h^2_{SNP}}$	YES	0.244	0.296	0.200	0.079	0.198	**0.214 (0.003)**	**1.052 (0.016)**
Bayesian: Enriched GCTA	$\widehat{h^2_{SNP}}$	NO	0.204	0.270	0.153	0.071	0.177	**0.183 (0.003)**	**0.914 (0.016)**
Bayesian: Enriched GCTA	0.5	YES	0.254	0.294	0.193	0.071	0.191	**0.213 (0.003)**	**1.054 (0.016)**
Bayesian: Enriched GCTA	0.5	NO	0.231	0.280	0.167	0.069	0.180	**0.194 (0.003)**	**0.975 (0.016)**
Bayesian: Enriched GCTA	Adjusted	YES	0.246	0.296	0.196	0.083	0.204	**0.214 (0.003)**	**1.058 (0.016)**
Bayesian: Enriched GCTA	Adjusted	NO	0.210	0.276	0.162	0.070	0.164	**0.186 (0.003)**	**0.928 (0.016)**
Bayesian: LDAK Model	$\widehat{h^2_{SNP}}$	YES	0.255	0.282	0.201	0.077	0.194	**0.213 (0.003)**	**1.047 (0.016)**
Bayesian: LDAK Model	$\widehat{h^2_{SNP}}$	NO	0.210	0.270	0.127	0.068	0.165	**0.177 (0.003)**	**0.893 (0.016)**
Bayesian: LDAK Model	0.5	YES	0.257	0.282	0.198	0.070	0.186	**0.210 (0.003)**	**1.041 (0.016)**
Bayesian: LDAK Model	0.5	NO	0.234	0.270	0.156	0.068	0.173	**0.190 (0.003)**	**0.953 (0.016)**
Bayesian: LDAK Model	Adjusted	YES	0.257	0.282	0.198	0.071	0.186	**0.208 (0.003)**	**1.041 (0.016)**
Bayesian: LDAK Model	Adjusted	NO	0.232	0.271	0.157	0.068	0.173	**0.190 (0.003)**	**0.953 (0.016)**
Bayesian: Enriched LDAK	$\widehat{h^2_{SNP}}$	YES	0.258	0.286	0.208	0.076	0.190	**0.216 (0.003)**	**1.061 (0.016)**
Bayesian: Enriched LDAK	$\widehat{h^2_{SNP}}$	NO	0.211	0.273	0.146	0.071	0.176	**0.183 (0.003)**	**0.920 (0.016)**
Bayesian: Enriched LDAK	0.5	YES	0.259	0.287	0.203	0.075	0.190	**0.214 (0.003)**	**1.059 (0.016)**
Bayesian: Enriched LDAK	0.5	NO	0.235	0.274	0.159	0.071	0.176	**0.193 (0.003)**	**0.964 (0.016)**
Bayesian: Enriched LDAK	Adjusted	YES	0.260	0.286	0.204	0.071	0.184	**0.211 (0.003)**	**1.055 (0.016)**
Bayesian: Enriched LDAK	Adjusted	NO	0.233	0.273	0.158	0.069	0.178	**0.192 (0.003)**	**0.960 (0.016)**

Supplementary Table 18: Comparing the predictive performance of Classical and Bayesian PRS. Each polygenic risk score (PRS) uses a model of the form $\sum_j \beta_j X_j$, where the vector X_j contains genotypes for SNP j, and the β_j are trained using the relevant set of summary statistics from the 24 summary GWAS (BMI, height, HDL cholesterol, LDL cholesterol or triglycerides). Values report correlations between predicted and observed phenotypes for the eMERGE data (which are independent of the 24 summary GWAS). For the Classical PRS, the β_j are estimates from classical linear regression. For the Bayesian PRS, the β_j are posterior means, obtained using one of four prior distributions: GCTA Model, Enriched GCTA Model, LDAK Model and Enriched LDAK Model (the enriched versions incorporate average estimates of enrichments for the 24 functional categories). For the Bayesian PRS, it is necessary to provide a value for h^2_{SNP}: first we used the corresponding estimates from LDSC-Zero or SumHer-Zero; second we set $h^2_{SNP} = 0.5$; third we used the corresponding estimates from LDSC or SumHer-GC (in which case we trained the PRS using adjusted summary statistics, obtained by dividing the test statistics by the corresponding estimates of confounding bias). When constructing PRS, it can be beneficial to clump[6] (identify pairs of SNPs within 1 cM with $r^2_{jl} > 0.5$, then discard the one with highest p-value / lowest Bayes Factor); we found that clumping had little impact for the Classical PRS, but always benefited the Bayesian PRS. The PRS marked in red are those reported in the main text.

The accuracy of each Bayesian PRS reflects the accuracy of the corresponding prior distribution. We see that the PRS constructed from the LDAK Model predict significantly better ($P < 0.05/4$) than the Classical PRS, demonstrating that the LDAK Model has value as a heritability model; there is a suggestion that incorporating estimates of enrichment improves accuracy further, although we recognize that the difference (approximately 1.5%) is not significant. By contrast, the performance of the PRS constructed from the GCTA Model is no different from that of the Classical PRS, indicating that the GCTA Model is no better than an agnostic model. Now, incorporating estimates of enrichments does lead to a significant improvement; it is difficult to disentangle how much of this improvement comes from the identification of important categories of SNPs, or because there are systematic differences between the LD and MAF of the functional annotations[8] (Supplementary Table 1).

	LDSC	Hybrid-CEPT (78-part model)			Hybrid-GC (78-part model)			SumHer-GC
Category	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Share (SD)**	**Expected**	**Av. Enrich. (SD)**	**Av. Enrich. (SD)**
Coding	6.13 (0.66)	0.051 (0.006)	0.017	3.33 (0.37)	0.043 (0.005)	0.017	2.54 (0.33)	1.74 (0.09)
Conserved	9.39 (0.60)	0.135 (0.011)	0.035	4.08 (0.34)	0.120 (0.011)	0.035	3.40 (0.32)	1.95 (0.07)
CTCF	-1.10 (0.68)	0.010 (0.009)	0.024	0.42 (0.36)	0.007 (0.008)	0.024	0.29 (0.34)	1.01 (0.06)
Digital Gen. Footprint	2.20 (0.26)	0.243 (0.022)	0.154	1.63 (0.14)	0.208 (0.021)	0.154	1.39 (0.14)	1.26 (0.03)
DNase I Hyper. Site	1.32 (0.20)	0.229 (0.022)	0.201	1.16 (0.11)	0.248 (0.022)	0.201	1.24 (0.11)	1.18 (0.02)
FANTOM5 Enhancer	1.45 (1.61)	0.007 (0.003)	0.005	1.61 (0.73)	0.007 (0.004)	0.005	1.50 (0.80)	0.86 (0.13)
Enhancer	2.84 (0.42)	0.086 (0.009)	0.044	1.99 (0.21)	0.079 (0.010)	0.044	1.83 (0.22)	1.34 (0.05)
Fetal DHS	1.76 (0.36)	0.137 (0.017)	0.102	1.40 (0.18)	0.164 (0.018)	0.102	1.64 (0.18)	1.27 (0.03)
H3K27ac (Hnisz)	1.61 (0.04)	0.538 (0.011)	0.411	1.33 (0.03)	0.525 (0.010)	0.411	1.29 (0.03)	1.18 (0.01)
H3K27ac (PGC2)	1.81 (0.10)	0.401 (0.015)	0.286	1.43 (0.06)	0.395 (0.015)	0.286	1.41 (0.05)	1.24 (0.02)
H3K4me1	1.76 (0.07)	0.619 (0.018)	0.473	1.34 (0.04)	0.631 (0.017)	0.473	1.35 (0.04)	1.18 (0.01)
H3K4me3	2.68 (0.19)	0.318 (0.009)	0.136	2.42 (0.07)	0.226 (0.013)	0.136	1.69 (0.10)	1.41 (0.03)
H3K9ac	2.82 (0.20)	0.224 (0.014)	0.133	1.73 (0.10)	0.224 (0.013)	0.133	1.73 (0.10)	1.52 (0.03)
Intronic	1.12 (0.03)	0.427 (0.009)	0.406	1.06 (0.02)	0.437 (0.008)	0.406	1.08 (0.02)	1.13 (0.01)
Promoter Flanking	0.18 (1.10)	0.010 (0.005)	0.008	1.15 (0.58)	0.009 (0.004)	0.008	1.11 (0.52)	1.84 (0.12)
Promoter	1.94 (0.39)	0.044 (0.006)	0.028	1.53 (0.20)	0.031 (0.006)	0.028	1.08 (0.20)	1.48 (0.06)
Repressed	0.68 (0.07)	0.383 (0.018)	0.463	0.83 (0.04)	0.396 (0.017)	0.463	0.85 (0.04)	0.78 (0.01)
Super Enhancer	1.71 (0.07)	0.223 (0.006)	0.174	1.30 (0.04)	0.233 (0.006)	0.174	1.35 (0.04)	1.17 (0.02)
T. Factor Binding Site	2.50 (0.26)	0.221 (0.017)	0.143	1.59 (0.13)	0.210 (0.018)	0.143	1.50 (0.13)	1.29 (0.03)
Transcribed	1.17 (0.09)	0.367 (0.016)	0.345	1.06 (0.05)	0.358 (0.016)	0.345	1.04 (0.04)	1.19 (0.01)
Transcription Start Site	3.91 (0.75)	0.037 (0.006)	0.016	2.35 (0.36)	0.038 (0.006)	0.016	2.41 (0.36)	1.97 (0.09)
3' Untranslated Region	4.80 (0.70)	0.026 (0.004)	0.013	2.28 (0.35)	0.019 (0.004)	0.013	1.57 (0.33)	1.64 (0.09)
5' Untranslated Region	3.95 (1.16)	0.013 (0.004)	0.005	2.31 (0.68)	0.006 (0.003)	0.005	0.98 (0.59)	1.27 (0.13)
Weak Enhancer	3.82 (0.77)	0.028 (0.008)	0.024	1.29 (0.33)	0.050 (0.008)	0.024	2.16 (0.35)	1.25 (0.06)

Supplementary Table 19: Average estimates of functional enrichments across the 24 summary GWAS from the hybrid model. In order to compare the fit of the GCTA and LDAK Models, we ran SumHer using a hybrid heritability model where the fractions $1-p$ and p indicate the proportions of GCTA and LDAK, respectively. Here we report for each category, averaged across the 24 summary GWAS, its estimated share of h^2_{SNP}, its expected share (calculated assuming $p = 0.5$) and its estimated enrichment, obtained using Hybrid-CEPT (the hybrid model assuming inflation due to confounding is additive) and Hybrid-GC (the hybrid model assuming multiplicative inflation). For comparison, we also report average estimates of enrichments from LDSC and from our recommended method, SumHer-GC; we see that estimates obtained using the hybrid model are closer to those from SumHer-GC than those from LDSC, to be expected considering the average estimate of p is 0.85 (for a visual comparison, see Supplementary Figure 13). Average estimated enrichments significantly different from one ($P < 0.05$) are marked in red. Despite the concordance between estimates from the hybrid model and those from SumHer-GC, we see the latter finds more categories significantly enriched; this is because using the hybrid model results in less precise estimates, as a consequence of including 53 extra coefficients (those corresponding to the GCTA Model).

Raw GWAS	Number of loci (MHC)	Summary GWAS	Number of loci (MHC)
Coronary Artery Disease	1 (0)	Alzheimer's Disease	1 (0)
Crohn's Disease	5 (0)	Crohn's Disease	2 (0)
Rheumatoid Arthritis	10 (10)	Rheumatoid Arthritis	5 (5)
Type 1 Diabetes	16 (16)	HDL Cholesterol	1 (0)
Celiac Disease	18 (18)		
Psoriasis	6 (6)		
Ulcerative Colitis	1 (1)		
Age-related Macular Disease	1 (0)		
Triglyceride	1 (0)		
HDL Cholesterol	1 (0)		

Supplementary Table 20: Numbers of large-effect loci for the 25 raw and 24 summary GWAS. For all analyses, we exclude SNPs within the major histocompatibility complex (Chromosome 6: 25-34 Mb), as well as SNPs which individually explain $>1\%$ of phenotypic variation, and SNPs in LD with these. This table reports for each GWAS the number of large-effect SNPs after thinning (within 1 cM and $r^2_{jl} > 0.1$), and how many of these are in the MHC.

References

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

2. Lambert, J. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).

3. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

4. Wood, A. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

5. Gottesman, O. *et al.* The electronic medical records and genomics (emerge) network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).

6. Euesden, J., Lewis, C. & O'Reilly, P. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).

7. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

8. Speed, D. *et al.* Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).

9. Speed, D. *et al.* Describing the genetic architecture of epilepsy through heritability analysis. *Brain* **137**, 2680âĂŞ2689 (2014).

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

11. Speed, D., Hemani, G., Johnson, M. & Balding, D. Improved heritability estimation from genome-wide SNP data. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).

12. Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* **519**, 309–314 (2015).

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

14. Astle, W. & Balding, D. Population structure and cryptic relatedness in genetic association studies. *Statist. Sci.* **24**, 451–471 (2009).

15. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208 (2006).

16. Yang, J., Zaitlen, N., Goddard, M., Visscher, P. & Price, A. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106 (2014).

17. Loh, P. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

18. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nat. Methods* **9**, 525–526 (2012).

19. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

20. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).

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

22. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

23. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49** (2017).

24. Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).

25. Locke, A. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

26. Speliotes, E. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

27. Juster, F. & Suzman, R. An overview of the Health and Retirement Study. *J. Hum. Resources* **30**, S7–S56 (1995).

28. Verma, S. *et al.* Imputation and quality control steps for combining multiple genome-wide datasets. *Front. Genet.* **5**, 370 (2015).

29. Kent, W. *et al.* The human genome browser at ucsc. *Genome Res.* **12**, 996–1006 (2002).

30. 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).

31. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

32. Ward, L. & Kellis, M. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675–1678 (2012).

33. Hoffman, M. *et al.* Integrative annotation of chromatin elements from encode data. *Nucleic Acids Res.* **41**, 827–841 (2013).

34. The ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. *Nature* **489**, 57–74 (2012).

35. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

36. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).

37. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

38. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

39. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

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

41. 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).

42. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).

43. Liu, J. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).

44. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).

45. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

46. Scott, R. *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).

47. Zheng, H. *et al.* Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature* **526**, 112–117 (2015).

48. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 626–633 (2016).

49. Manning, A. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).

50. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin aâĆĄ(c) levels via glycemic and nonglycemic pathway. *Diabetes* **59**, 3229–3239 (2010).

51. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).

52. Perry, J. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).

53. Day, F. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and brca1-mediated dna repair. *Nat. Genet.* **47**, 1294–1303 (2015).

54. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nat. Genet.* **518**, 187–196 (2015).

55. 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).

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

57. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform ld score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2016).