Supplementary Information for: Detection and interpretation of shared genetic influences on 42 human traits

Joseph K. Pickrell[1,2,†], Tomaz Berisa[1], Jimmy Z. Liu[1], Laure Segurel[3], Joyce Y. Tung[4], David Hinds[4]

[1] New York Genome Center, New York, NY, USA
[2] Department of Biological Sciences, Columbia University, New York, NY, USA
[3] UMR7206 Eco-anthropologie et ethnobiologie, CNRS-MNHN-Paris 7, Paris, France
[4] 23andMe, Inc., Mountain View, CA, USA
[†] Correspondence to: jkpickrell@nygenome.org

April 15, 2016

Contents

1 GWAS data

In this section, we describe the sources of the GWAS data used in the paper and initial analyses of the numbers of associations identified in each. Overall, for each study, we imputed summary statistics or genotypes for all autosomal variants in the March 2012 release of the 1000 Genomes Phase 1 [Abecasis et al., 2010]. Our method uses the Z-scores and standard errors of the estimated effect sizes for each SNP. In studies where standard errors were not provided, we approximated them using the allele frequencies from the European-descent individuals in the 1000 Genomes Phase 1 release and the reported sample size of the study (see Pickrell [2014] for details). Throughout the paper, we report effect sizes of variants as the effect of the non-reference allele in human genome reference hg19.

1.1 Previously described in Pickrell [2014]

For some phenotypes, the data sources and processing were previously described in Pickrell [2014]. These are (using the abbreviations from the main text): FG, LDL, HDL, TG, TC, FNBMD, LSBMD, PLT, MPV, HB, MCHC, RBC, MCV, and PCV.

1.2 Body mass index

The GIANT consortium GWAS data described in Locke et al. [2015] was downloaded from `http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files`. We removed all SNPs typed on less than 200,000 individuals or more than 240,000 individuals, then performed imputation at the level of the summary statistics. After removing poorly-imputed SNPs (at an r^2 threshold less than 0.8), we were left with 6,133,872 summary statistics. To approximate the variance in the effect size estimates, we used a sample size of 240,000 and the allele frequencies estimated from the European-descent individuals in the 1000 Genomes Project.

1.3 Waist-hip ratio

We downloaded summary statistics from the GIANT consortium GWAS of waist-hip ratio corrected for BMI [Shungin et al., 2015] from `http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files`. We used the summary statistics generated using both males and females together. We removed all SNPs genotyped on less than 130,000 individuals or more than 150,000 individuals, then performed imputation at the level of summary statistics. After removing poorly-imputed SNPs (at an r^2 threshold less than 0.8), we were left with 5,859,436 summary statistics. To approximate the variance in the effect size estimates, we used a sample size of 142,762 and the allele frequencies estimated from the European-descent individuals in the 1000 Genomes Project.

1.4 Coronary artery disease

The CARDIoGRAM GWAS data described in Schunkert et al. [2011] was downloaded from `http://www.cardiogramplusc4d.org/downloads/`. We removed all SNPs typed on less than 15,000 cases or 50,000 controls, then imputed summary statistics for all SNPs in the 1000 Genomes Phase 1 release using ImpG [Pasaniuc et al., 2014]. After removing poorly-imputed SNPs (at an r^2 threshold less than 0.8), we were left with 5,768,612 summary statistics on an average of around 15,000 cases and 50,000 controls. To approximate the variance in the effect size estimates we used the allele frequencies from the European-descent individuals in the 1000 Genomes Project.

1.5 Crohn's disease

We obtained updated versions of the summary statistics from the GWAS described in Jostins et al. [2012] from `http://www.broadinstitute.org/~sripke/share_links/KMaiXtod0Alvozgg5oJgl3QK7VVH2N_IBD_0711a/`, and processed them as described previously [Pickrell, 2014]. After imputation, we had 6,065,937 summary statistics.

1.6 Ulcerative colitis

We obtained updated versions of the summary statistics from the GWAS described in Jostins et al. [2012] from `http://www.broadinstitute.org/~sripke/share_links/KMaiXtod0Alvozgg5oJgl3QK7VVH2N_IBD_0711a/`, and processed them as described previously [Pickrell, 2014]. After imputation, we had 6,067,283 summary statistics.

1.7 Educational attainment

Summary statistics from the GWAS described in Okbay et al. [2016] were kindly provided by the Social Science Genetic Association Consortium and 23andMe. The data consist of summary statistics at 9,256,491 variants.

1.8 Type 2 diabetes

The DIAGRAM consortium summary statistics described in Morris et al. [2012] were downloaded from `http://diagram-consortium.org/downloads.html`. We removed SNPs typed on less than 9,000 cases or 50,000 controls, then imputed summary statistics for all variants in the 1000 Genomes Phase 1 release. After imputation, we had 5,824,487 summary statistics. To approximate the variance in the effect size estimates we used the allele frequencies from the European-descent individuals in the 1000 Genomes Project.

1.9 Alzheimer's disease

Summary statistics from the IGAP (International Genomics of Alzheimer's Project) consortium GWAS [Lambert et al., 2013] were downloaded from `http://www.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php`. As per the description of the data:

> International Genomics of Alzheimer's Project (IGAP) is a large two-stage study based upon genome-wide association studies (GWAS) on individuals of European ancestry. In stage 1, IGAP used genotyped and imputed data on 7,055,881 single nucleotide polymorphisms (SNPs) to meta-analyse four previously-published GWAS datasets consisting of 17,008 Alzheimer's disease cases and 37,154 controls (The European Alzheimer's disease Initiative - EADI the Alzheimer Disease Genetics Consortium - ADGC The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium - CHARGE The Genetic and Environmental Risk in AD consortium - GERAD). In stage 2, 11,632 SNPs were genotyped and tested for association in an independent set of 8,572 Alzheimer's disease cases and 11,312 controls. Finally, a meta-analysis was performed combining results from stages 1 & 2.

We used only the stage 1 results in all analysis.

1.10 Schizophrenia

Summary statistics from the psychiatric genomics consortium GWAS [Psychiatric Genomics Consortium, 2014] were downloaded from `http://www.med.unc.edu/pgc/downloads`. The data consist of summary statistics at 9,444,231 variants.

1.11 Height

Summary statistics from the GIANT consortium GWAS [Wood et al., 2014] were downloaded from `http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files`. We removed all SNPs with a sample size less than 230,000, then imputed summary statistics for all variants in the 1000 Genome Phase 1 release. After imputation, we had 6,021,095 summary statistics. To approximate the variance in the effect size estimates, we used a sample size of 230,000 and allele frequencies from the European individuals in the 1000 Genomes Project.

1.12 Age at menarche

Summary statistics from the Reprogen consortium GWAS [Perry et al., 2014] were downloaded from `http://www.reprogen.org/data_download.html`. We imputed summary statistics for all variants in the 1000 Genome Phase 1 release. After removing poorly-imputed SNPs, we had 6,277,050 summary statistics. To approximate the variance in the effect size estimates, we used a sample size of 132,989 and the allele frequencies from the European-descent individuals in the 1000 Genomes Project.

1.13 Rheumatoid arthritis

Summary statistics from `http://plaza.umin.ac.jp/~yokada/datasource/files/GWASMetaResults/`. We used the GWAS conducted in European-descent individuals only. These data consist of 8,747,963 summary statistics.

1.14 23andMe data

GWAS for a number of traits were performed by the personal genomics company 23andMe using a survey design [Eriksson et al., 2010]. Some of these GWAS have been partially described previously [Do et al., 2011; Eriksson et al., 2012a, 2010, 2012b; Kiefer et al., 2013]. Details of each GWAS are described in separate Supplementary Data.

We used filters and summary data described in the Supplementary Data, with two exceptions: for both allergies and Parkinson's disease, we used more stringent thresholds for genotyping batch effects. Specifically, 23andMe reports a P-value for whether the allele frequencies of each SNP are associated to genotyping batch or to imputation batch. We used P-value thresholds of $P = 1 \times 10^{-7}$ and 1×10^{-6}, respectively, for these two studies. This more stringent cutoff was used to account for an imbalance in the enrollment of cases versus controls over time in these two studies. Some small residual batch effect likely explains the small correlation in effect sizes between Parkinson's disease and allergies in Figure 4 in the main text.

1.15 Counting independent numbers of associated variants

For each genome-wide association study, we ran fgwas v.0.3.6 [Pickrell, 2014] with the default settings, except that rather than splitting the genome into blocks with equal numbers of SNPs (as in Pickrell [2014]), we split the genome into approximately independent blocks based on patterns of linkage disequilibrium in the European populations in Phase 1 of the 1000 Genomes Project [Berisa and Pickrell, 2015]. These blocks are available at `https://bitbucket.org/nygcresearch/ldetect-data`. We used fgwas to estimate

the prior probability that any block contains an association. The output of this model is, for each region of the genome, the posterior probability that it contains a variant that influences the trait. We used a threshold of a posterior probability of association of 0.9, as in Pickrell [2014], which can be roughly interpreted as a false discovery rate of 10%. For analyses that use variants identified in these individual GWAS, we extracted the single SNP from each region with the largest posterior probability of being the causal SNP in this model.

1.16 Approximating the correlations in the effect sizes under the null model

For genome-wide association studies of correlated traits performed on overlapping individuals, we expect the observed association statistics of a given variant to both traits to be correlated, even under the null model that the variant influences neither trait. To approximate this expected correlation, for both traits we extracted all genomic regions with a posterior probability of containing an association less than 0.2 (using the method described above). We then extracted all SNPs from these regions, and calculated the correlation in the Z-scores between the two traits (using all SNPs remaining in both studies). This correlation is a function of the number of overlapping samples and the correlation in the phenotypes. Specifically, if $\vec{g}_1$ is a vector of (mean-centered) genotypes at a variant in study 1, $\vec{x}$ is a vector of (standard normally distributed) phenotypes in study 1, $\vec{g}_2$ is a set of is a vector of genotypes at a variant in study 2, and $\vec{y}$ is a vector of phenotypes in study 2, then:

$$Cor(Z_1, Z_2) = E\left[\frac{\sum_i g_{1i}x_i}{\sqrt{\sum_i g_{1i}^2}}\frac{\sum_j g_{2j}y_j}{\sqrt{\sum_j g_{2j}^2}}\right] \quad (1)$$

$$\approx \frac{N_0}{\sqrt{N_1 N_2}}\rho, \quad (2)$$

where N_0 is the number of overlapping individuals in the two studies, N_1 is the number of individuals in study 1, N_2 is the number of individuals in study 2, and ρ is the correlation between the phenotypes. We used this summary statistic-level correlation (later called C) as a correction factor in all pairwise GWAS.

2 Hierarchical model

In this section we describe the hierarchical model used for the main scan for overlapping association signals in two GWAS. Our goal is to write down a model that allows us to estimate the probability that a genomic locus contains a variant that influences two traits. Our approach is to split the genome into non-overlapping regions (we used the same approximately independent blocks as above); each region then falls into one of five categories, following Giambartolomei et al. [2014]:

0. There are no SNPs in the region that influence either trait (denoted RM_0, for regional model 0),
1. There is one causal SNP in the region that influences the first trait (RM_1),
2. There is one causal SNP in the region that influences the second trait (RM_2),
3. There is one causal SNP in the region that influences both traits (RM_3),
4. There are two causal SNPs in the region, one of which influences the first trait and one of which influences the second (RM_4).

We will estimate the proportion of genomic regions in each of these categories with an empirical Bayes approach. Note that we do not consider situations where multiple variants in a region influence a single trait, though in principle the model could be extended to allow this possibility. In what follows, we start by writing

down the model for the simplest case where two phenotypes have been studied in separate cohorts, and then introduce modifications for the more complex situations that arise in real data. Software implementing the model is available at `https://github.com/joepickrell/gwas-pw`.

In the simplest case, consider two separate genome-wide association studies performed on two traits. In this case, the model is a hierarchical version of that in Giambartolomei et al. [2014], which we re-iterate here for completeness. Let there be N_1 individuals in GWAS of the first phenotype and N_2 individuals in the GWAS of the second phenotype. We start by considering a single SNP. Let $\vec{g}$ be the vector of genotypes at the SNP in the first study, $\vec{h}$ be the vector of genotypes at the SNP in the second study, $\vec{x}$ be the vector of phenotype measurements for the first phenotype (assumed to be distributed as a standard normal), and $\vec{y}$ be the vector of phenotype measurements for the second phenotype (also assumed to be distributed as a standard normal). We first need a measure of the evidence that the SNP influences each of the traits.

2.1 Bayes factor calculations

We use a simple linear regression model to relate the phenotypes and the genotypes:

$$E[x_i] = \beta_1 g_i \tag{3}$$
$$E[y_j] = \beta_2 h_j. \tag{4}$$

There are four potential models to consider for this SNP:

0. M_0: the SNP is associated with neither trait
1. M_1: the SNP is associated with the first trait (but not the second)
2. M_2: the SNP is associated with the second trait (but not the first)
3. M_3: the SNP is associated with both traits.

Model M_0 corresponds to the case where $\beta_1 = 0$ and $\beta_2 = 0$, model M_1 corresponds to the case where β_1 is free to vary while $\beta_2 = 0$, and so on. We can thus define three Bayes factors corresponding to the evidence in favor of the three alternative models:

$$BF^{(1)} = \frac{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_1)d\theta_1}{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_0)d\theta_0} \tag{5}$$
$$BF^{(2)} = \frac{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_2)d\theta_2}{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_0)d\theta_0} \tag{6}$$
$$BF^{(3)} = \frac{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_3)d\theta_3}{\int P(\vec{x},\vec{y}|\vec{g},\vec{h},\theta_0)d\theta_0}, \tag{7}$$

where θ_j represents the parameters of model j. To compute these Bayes factors, we use the approximate Bayes factors from Wakefield [2008]. If we let $\hat{\beta}_1$ be the maximum likelihood estimate of β_1 and $\sqrt{V_1}$ be the standard error in that estimate, then $Z_1 = \frac{\hat{\beta}_1}{\sqrt{V_1}}$. If the prior on the true effect size is $\beta_1 \sim N(0, W_1)$ we can write down the Wakefield approximate Bayes factor measuring the evidence that the SNP is associated with the first phenotype:

$$WABF_1 = \sqrt{1-r_1}\exp\left[\frac{Z_1^2}{2}r_1\right], \tag{8}$$

where $r_1 = \frac{W_1}{V_1+W_1}$. $WABF_2$ is defined analogously. In all applications, we averaged over Bayes factors computed with $W = 0.01, W = 0.1$, and $W = 0.5$. To now connect these approximate Bayes factors to the three alternative models for the SNP [Giambartolomei et al., 2014]:

$$BF^{(1)} = WABF_1 \tag{9}$$

$$BF^{(2)} = WABF_2 \tag{10}$$

$$BF^{(3)} = WABF_1 WABF_2, \tag{11}$$

where $BF^{(3)}$ is a consequence of the fact that the two cohorts are independent. This latter Bayes factor is equivalent to that derived under the "maximum heterogeneity" model in Wen and Stephens [2014].

We note that in the Wakefield approximate Bayes factor the effect size of a SNP enters only through the Z-score. As a consequence, if we consider the "reverse" regression model where we swap the genotypes and phenotypes:

$$E[g_i] = \beta_1' x_i \tag{12}$$

$$E[h_j] = \beta_2' y_j, \tag{13}$$

then the Bayes factors from this model are identical to the previous model (as long as the ratios r_1 and r_2 remain constant). In fact it is this latter "reverse" regression that we will use going forward, though it is simpler to interpret parameters like the prior on the effect size in the traditional parameterization.

2.2 Regional Bayes factor

We now consider a Bayes factor measuring the support for an association in a given genomic region r. To do this, now consider the matrix $\mathbf{G}_r$ of genotypes in the region in the first study (with N_1 rows of individuals and K columns of SNPs) and the matrix $\mathbf{H}_r$ of genotypes in the region in the second study (with N_2 rows and K columns). The vectors of phenotypes remain $\vec{x}$ and $\vec{y}$. We now want to write down Bayes factors measuring the evidence in favor of the four alternative models discussed at the beginning of Section 2 relative to the null model of no associations in the region. For regional model 1 (there is a single SNP casually associated with the first phenotype and none with the second phenotype):

$$RBF_r^{(1)} = \sum_{i=1}^{K} \frac{\pi_i^{(1)} P(\mathbf{G}_r, \mathbf{H}_r | \text{SNP } i \text{ is causal}, RM_1)}{P(\mathbf{G}_r, \mathbf{H}_r | RM_0)} \tag{14}$$

$$= \sum_{i=1}^{K} \pi_i^{(1)} BF_i^{(1)}, \tag{15}$$

where $\pi_i^{(1)}$ is the prior probability that SNP i is the causal one under model 1. Note that the probabilities of all genotypes at the non-causal sites cancel out because they are identical once we have conditioned on the genotype at a causal site [Maller et al., 2012].

Analogously,

$$RBF_r^{(2)} = \sum_{i=1}^{K} \pi_i^{(2)} BF_i^{(2)} \tag{16}$$

$$RBF_r^{(3)} = \sum_{i=1}^{K} \pi_i^{(3)} BF_i^{(3)} \tag{17}$$

$$RBF_r^{(4)} = \sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i^{(1)} \pi_j^{(2)} BF_i^{(1)} BF_j^{(2)} I[i \neq j], \tag{18}$$

$$\tag{19}$$

where $I[i \neq j]$ is an indicator that evaluates to 1 if i and j are different and 0 otherwise, $\pi_i^{(2)}$ is the prior probability that SNP i is the causal SNP under model 2, and $\pi_i^{(3)}$ is the prior probability that SNP i is the causal SNP under model 3. For model 4 (where there are two causal SNPs, one of which only influences the first phenotype and one of which only influences the second phenotype), we assume that the prior probabilities that SNP i influences the first or second phenotype are identical to those under model 1 and 2, respectively. In all applications we set $\pi_i^{(1)} = \pi_i^{(2)} = \pi_i^{(3)} = \frac{1}{K}$.

2.3 Likelihood

We now turn to the model for the whole genome. We denote the full matrix of genotypes in the first study as $\mathbf{G}$ and the full matrix of genotypes in the second study as $\mathbf{H}$. We split the genome into M approximately independent blocks. Under the assumption that all blocks are independent, the probability of all the genotypes is:

$$P(\mathbf{G}, \mathbf{H} | \vec{x}, \vec{y}) = \prod_{i=1}^{M} P(\mathbf{G}_i, \mathbf{H}_i | \vec{x}, \vec{y}) \tag{20}$$

$$= \prod_{i=1}^{M} \left[\sum_{j=0}^{4} \Pi_j P(\mathbf{G}_i, \mathbf{H}_i | \vec{x}, \vec{y}, RM_j) \right], \tag{21}$$

where Π_j is the prior probability of regional model j and RM_j is regional model j. These are the probabilities we would like to learn. We can do so by maximizing the log-likelihood:

$$l(\theta | \mathbf{G}, \mathbf{H}) = \sum_{i=1}^{M} \ln \left[\Pi_0 + \sum_{j=1}^{4} \Pi_j RBF_i^{(j)} \right], \tag{22}$$

where θ is the set of parameters in the model (the prior probabilities and all of the parameters that go into the construction of the Bayes factors). We maximized this likelihood with the approach described in Section 2.5.

2.4 Bayes factors for overlapping cohorts

The above model makes the key assumption that the two phenotypes in question have been measured on two separate sets of individuals. In practice, the cohorts we will use are often overlapping or partially overlapping. This causes two problems for the model. First, if the two phenotypes are correlated, we may overestimate the evidence in favor of regional model 3 (where a single variant influences both phenotypes). Second, the patterns of linkage disequilibrium in the population can no longer be ignored when considering the evidence in favor of regional model 4, and we may overestimate the evidence in favor of this model. In practice, we were most concerned with the first of these; see Section 2.8 for discussion of the second.

The degree to which we may overestimate the evidence in favor of regional model 3 depends on the number of overlapping samples in the two studies and the correlation in the phenotypes. We consider the case where there is a single cohort of individuals. Let the vector of genotypes at the SNP be $\vec{g}$, and let the two vectors of phenotypes be $\vec{x}$ and $\vec{y}$. Let the phenotypes be bivariate-normally distributed with mean zero, variance one, and correlation coefficient C. In this case the correlation between the phenotypes is identical the correlation between the summary statistics.

As before, we first want to calculate the Bayes factors measuring the evidence in favor of the three alternative models from Section 2.1. We use a multivariate linear regression model:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} = g_i \begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix} + \begin{bmatrix} \varepsilon_{xi} \\ \varepsilon_{yi} \end{bmatrix}, \tag{23}$$

where β_x is the effect of the SNP on phenotype x, β_y is the effect of the SNP on phenotype y, and ε_{xi} and ε_{yi} are error terms that are multivariate normally distributed with mean zero and covariance matrix Σ (though in all that follows we assume the effects of any SNP are small, so this residual covariance matrix is approximated by the covariance matrix of the phenotypes).

To compute the Bayes factors, we use a multivariate extension of the approximate Bayes factor from Wakefield [2008]. Instead of working directly with the phenotype and genotype vectors, we instead consider $\hat{\beta}_x$ and $\hat{\beta}_y$, the estimated effect sizes from each individual regression. We also use V_x and V_y, the respective variances in the estimates of each regression coefficient, and $Z_x = \frac{\hat{\beta}_x}{\sqrt{V_x}}$ and $Z_y = \frac{\hat{\beta}_y}{\sqrt{V_y}}$. We let:

$$\begin{bmatrix} \hat{\beta}_x \\ \hat{\beta}_y \end{bmatrix} \sim MVN\left(\begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix}, \begin{bmatrix} V_x & C\sqrt{V_x V_y} \\ C\sqrt{V_x V_y} & V_y \end{bmatrix} \right). \tag{24}$$

We now place a multivariate normal prior on β_x and β_y. The form we choose is:

$$\begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} W_x & C\sqrt{(V_x+W_x)(V_y+W_y)} - C\sqrt{V_x V_y} \\ C\sqrt{(V_x+W_x)(V_y+W_y)} - C\sqrt{V_x V_y} & W_y \end{bmatrix} \right). \tag{25}$$

This prior has the somewhat odd property that it depends to a small extent on the variances of the effect size estimates (i.e. on the minor allele frequency of the SNP in question), such that rarer SNPs, or those with a large amount of missing data, have larger prior covariances. The benefit of this prior is that now the posterior predictive distribution of the estimated effect sizes has a simple form:

$$\begin{bmatrix} \hat{\beta}_x \\ \hat{\beta}_y \end{bmatrix} | H_1 \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_x+W_x & C\sqrt{(V_x+W_x)(V_y+W_y)} \\ C\sqrt{(V_x+W_x)(V_y+W_y)} & V_y+W_y \end{bmatrix} \right). \tag{26}$$

The other models are similar. With these assumptions, we can analytically compute the Bayes factors:

$$BF^{(1)} \approx \sqrt{(1-r_x)} \exp\left[\frac{1}{2(1-C^2)} [Z_x^2 r_x - 2CZ_x Z_y \left(1 - \sqrt{(1-r_x)}\right)] \right] \tag{27}$$

$$BF^{(2)} \approx \sqrt{(1-r_y)} \exp\left[\frac{1}{2(1-C^2)} [Z_y^2 r_y - 2CZ_x Z_y \left(1 - \sqrt{(1-r_y)}\right)] \right] \tag{28}$$

$$BF^{(3)} \approx \sqrt{(1-r_x)(1-r_y)} \exp\left[\frac{1}{2(1-C^2)} [Z_x^2 r_x + Z_y^2 r_y - 2CZ_x Z_y \left(1 - \sqrt{(1-r_x)(1-r_y)}\right)] \right], \tag{29}$$

where $r_x = \frac{W_x}{V_x+W_x}$ and $r_y = \frac{W_y}{V_y+W_y}$. Note that if the two phenotypes are uncorrelated all three Bayes factors are identical to those in Section 2.1.

For all of the pairwise GWAS, we used these Bayes factors instead of those in Section 2.1. We used C estimated from the summary statistics, as described in Section 1.16. Note that when the cohorts are only partially overlapping, the C we calculate is a function of the amount of overlap between the cohorts as well as the correlation in the phenotypes. In principle some knowledge of the true correlation between two phenotypes could be obtained from external data and incorporated into the prior here, but we have chosen not to do this, and so as the overlap in the cohorts goes to zero, these Bayes factors tend to the prior assumption that the two phenotypes are uncorrelated (i.e. the maximum heterogeneity Bayes factor from Wen and Stephens [2014]). This may be suboptimal in situations where good external information is available.

2.5 Fitting the model

The natural approach to fitting this model would be to maximize the log-likelihood in Equation 22. However, in a small subset of cases (generally in pairs of GWAS with small numbers of associated variants), we found that this maximization was numerically unstable. To fix this, we placed a weak logistic normal prior on the Π parameters. Specifically, we define hyperparameters:

$$\alpha_0 \sim N(2,9) \tag{30}$$
$$\alpha_1 \sim N(-2,9) \tag{31}$$
$$\alpha_2 \sim N(-2,9) \tag{32}$$
$$\alpha_3 \sim N(-2,9) \tag{33}$$
$$\alpha_4 \sim N(-2,9) \tag{34}$$

and then define:

$$\Pi_i = \frac{e^{\alpha_i}}{\sum_j e^{\alpha_j}}. \tag{35}$$

Instead of maximizing the likelihood, we maximized the function:

$$f(\theta) = l(\theta) + g(\theta), \tag{36}$$

where θ is the set of five α parameters, $l(\theta)$ is the log-likelihood from Equation 22 and $g(\theta)$ is the log of the prior density described above. We maximized this function using the Nelder-Mead algorithm implemented in the GNU Scientific Library. The estimates of the parameters are maximum *a posteriori* estimates rather than maximum likelihood estimates. In practice, this serves to prevent estimates of the Π parameters from going all the way to zero.

2.6 Comparison to Stephens's Bayes factors

There are a number of existing approaches to testing for an association between a genetic variants and multiple traits [Ferreira and Purcell, 2009; Korte et al., 2012; O'Reilly et al., 2012; Stephens, 2013; Zhang et al., 2014; Zhou and Stephens, 2014]. The most similar approach to ours is that of Stephens [2013]. We compared the qualitative performance of the two (SNP-level) Bayes factors using simulations. Recall that there are three possible non-null models for each SNP: either it influences 1) the first trait alone, 2) the second trait alone, or 3) both traits. The simulation algorithm was as follows:

1. Simulate an effect size: $\beta \sim N(0, 0.09)$ and an allele frequency $f \sim Beta(2,2)$.

2. Simulate 5,000 genotypes: $\vec{g} = [g_1, g_2, ...g_{5000}]$, where $g_i \sim Bin(2, f)$.

3. Simulate a pair of phenotypes for each simulated genotype according to one of the models described above: $[x_i, y_i] \sim MVN(\vec{\mu}, \Sigma)$, where $\mu_1 = g_i\beta$ if simulating from models 1 or 3 (and zero otherwise), $\mu_2 = g_i\beta$ if simulating from models 2 or 3 (and zero otherwise) and Σ is a variance/covariance matrix with the diagonal set to one and the off-diagonal terms set to a correlation C.

4. Calculate the Bayes factors measuring the evidence in favor of each of the three alternative models using their the Bayes factors as described in the main text or the code in Supplementary Section 3 of Stephens [2013]. We used a prior variance of 0.5, and estimated the correlation in the phenotypes directly from the simulated phenotypes (rather than setting it to the simulated value).

Results. We considered two correlation coefficients (0 and -0.4) and all three possible alternative models, for a total of six simulation conditions. In each condition, we repeated the procedure above 1,000 times and calculated all three possible Bayes factors. For example, in the condition where the "truth" is two uncorrelated phenotypes and a genetic variant that influences only the first, we calculated Bayes factors measuring the support for the true model (model 1: only the first phenotype is associated with the genotype) as well as the Bayes factor supporting the model 2 (only the second phenotype is associated with the genotype) and the Bayes factor measuring the support for model 3 (both phenotypes are associated with the genotype).

In Figure 1, we show our Bayes factors and the Stephens Bayes factors for the situation where the two phenotypes are uncorrelated. The two Bayes factors are similar in all situations. In Figure 2, we show our Bayes factors and the Stephens Bayes factors for the situation where the two phenotypes are (negatively) correlated. The two Bayes factors disagree considerably in two situations. Specifically, in simulations where a genetic variant is associated with only a single phenotype and the test is for association with the *other* phenotype, our Bayes factors show little-to-no evidence for association, while the Stephens Bayes factors sometimes show substantial evidence for association. On reflection, this appears to be due to a small but apparently important difference in underlying models. Specifically, the Stephens [2013] model is equivalent for testing for independence between the genotype and the second phenotype *conditional* on the first phenotype, while we are testing for *unconditional* independence between the genotype and the phenotype.

2.7 Simulations

We wanted to evaluate the performance of the hierarchical model in moderately realistic simulations. To do this, we simulated GWAS using haplotypes from the HapMap Project [Frazer et al., 2007]. We predefined 111 genomic regions to simulate data from, and assigned two genetic variants in each region as the potentially causal variants (see Figure 1 in the main text). Each simulation then had the following steps:

1. Simulate 10,000 haplotypes in the 111 genomic regions using hapgen2 [Su et al., 2011] and reference haplotypes from the HapMap Project.

2. Assign each region to one of the five models $RM_{[0-4]}$ randomly (in general, we evenly distributed the regions to all five models).

3. Sum the number of causal alleles for both phenotypes for each individual (combining two simulated haplotypes to make an individual).

4. Simulate phenotypes for each individual as $MVN(\mu, \Sigma)$ (where μ and Σ were set as described below) and quantile normalize the phenotypes.

5. Run linear regression of each SNP against both phenotypes (in simulations of separate cohorts, we used half the individuals for the association study of the first phenotype, and half for the association study of the second).

6. Run the model on the summary statistics generated from these linear regressions. For all runs, we used a prior variance on the effects sizes of 0.5.

2.7.1 Simulated traits

For uncorrelated traits, we simulated the phenotypes of each individual as independent normally distributed variables with $\mu = 0.3n_i$, where n_i is the counts of causal alleles for phenotype i, and $\sigma^2 = 0.1$. For simulations of correlated traits, we used the same structure, except that we set the covariance between the two phenotypes as $0.1 * \rho$, where ρ was set according to the simulation.

2.7.2 Results

We simulated four situations: uncorrelated traits (both overlapping cohorts and non-overlapping cohorts) and correlated traits (both overlapping cohorts and non-overlapping cohorts) with a correlation coefficient of 0.3. Shown in Supplementary Figure 3 are the parameter estimates obtained in the simulations. For comparison, we also used a simple P-value model where a region was called a "hit" for each phenotype if a SNP had a P-value of less than 5×10^{-8} for the phenotype. If a single variant had a P-value less than 5×10^{-8} for one phenotype and a P-value less than 5×10^{-5} for the other phenotype, we called this as fitting model 3 (of co-localized signals).

We see that in the case of separate cohorts (Supplementary Figure 3A,C), the hierarchical model gets close to the truly simulated proportions, while the P-value method dramatically overestimates Π_0. There is a slight overestimation of Π_3, which we attribute the fact that the evidence against the null for variants that truly influence both phenotypes is stronger than for variants that influence only a single trait. In the case of perfectly overlapping cohorts (Supplementary Figure 3B,D), the hierarchical model additionally overestimates Π_4, likely for the reason discussed below. In this case the P-value method performs well, though we attribute this to the fact that we have simulated a GWAS where all variants have the same effect size and the sample size is large enough to have excellent power to detect this effect size. In real GWAS, this is unlikely to be the case.

2.8 Accounting for linkage disequilibrium in overlapping cohorts

Up to this point, we have ignored linkage disequilibrium when fitting the hierarchical model. Here we describe an approach that accounts for linkage disequilibrium as well. This is only relevant for the calculation of the Bayes factor measuring the support for regional model 4, where there are two causal variants in a region, one of which influences the first phenotype and one of which influences the second phenotype. This section is included for completeness, but apart from the explorations described in Section 2.8.3 we did not use this modification as we did not find it useful in practice.

2.8.1 Conditional Bayes factors.

In this situation, it is necessary to approximate a conditional regression analysis–that is, to estimate the effect size of one genetic variant conditional on the estimated effect size of another genetic variant. We redefine some notation for simplicity. Consider a single phenotype and two genetic variants. Let $\vec{y}$ be a vector of (standard normally-distributed) phenotypes, $\vec{g}_1$ be the mean-centered vector of genotypes at SNP 1, $\vec{g}_2$ be the mean-centered vector of genotypes at SNP 2, $\hat{\beta}_1$ be the (non-conditional) estimate of the effect of SNP

1 on the phenotype from a simple linear regression of the phenotype on the genotypes at SNP 1, and $\hat{\beta}_2$ be the analogous estimate for SNP 2. We consider the conditional regression model where:

$$y_i - \hat{\beta}_2 g_{i2} = \beta_1' g_{i1} + \varepsilon, \tag{37}$$

where β_1' is the conditional estimate of the effect size of SNP 1 and i indexes individuals. The least squares estimate of this effect size is:

$$\hat{\beta}_1' = \frac{\sum_i (y_i - \hat{\beta}_2 g_{i2}) g_{i1}}{\sum_i g_{i1}^2} \tag{38}$$

$$\hat{\beta}_1' = \hat{\beta}_1 - \frac{Cov(g_1, g_2)}{Var(g_1)} \hat{\beta}_2, \tag{39}$$

In order to compute our approximate Bayes factors, we need the variance in this estimate. If we let $R = \frac{Cov(g_1,g_2)}{Var(g_1)}$:

$$Var(\hat{\beta}_1') = Var(\hat{\beta}_1 - R\hat{\beta}_2) \tag{40}$$

$$= Var(\hat{\beta}_1) + Var(R\hat{\beta}_2) - 2Cov(\hat{\beta}_1, R\hat{\beta}_2). \tag{41}$$

If we assume that there is no error in our estimate of R, this reduces to:

$$Var(\hat{\beta}_1') \approx Var(\hat{\beta}_1) - R^2 Var(\hat{\beta}_2), \tag{42}$$

which is equivalent to Equation 20 from Yang et al. [2012] (assuming effect sizes are small, such that the residual variance is equivalent to the phenotypic variance). We can then define $Z' = \frac{\hat{\beta}_1'}{\sqrt{Var(\hat{\beta}_1')}}$ and use any of the approximate Bayes factors as defined previously.

Propagating error in R. It has been suggested that if R is estimated from around 2,000 individuals then it can be treated as a scalar rather than a random variable [Yang et al., 2012]. In general, however, we do not have access to the genotype data underlying a study, so we will estimate R from a separate panel of publicly-available haplotypes from individuals of the same ancestry as the individuals in the association study. These panels generally have hundreds rather than thousands of individuals, so we need to propagate the error in R. Under the null, the second term in Equation 41 is:

$$Var(R\hat{\beta}_2) = E[R]^2 Var(\hat{\beta}_2) + 2Var(\hat{\beta}_2) Var(R), \tag{43}$$

and the third is:

$$Cov(\hat{\beta}_1, R\hat{\beta}_2) = E[R\hat{\beta}_1\hat{\beta}_2] \tag{44}$$

$$= RCov(\hat{\beta}_1, \hat{\beta}_2) \tag{45}$$

$$= R^2 Var(\hat{\beta}_2), \tag{46}$$

so substituting and simplifying we get:

$$Var(\hat{\beta}_1') = Var(\hat{\beta}_1) + Var(\hat{\beta}_2)[2Var(R) - R^2]. \tag{47}$$

Finally, we need to approximate the variance in R. We assume this is estimated from a reference panel consisting of $2M$ phased haplotypes from M diploid individuals. Let the count of haplotypes carrying the non-reference alleles at both sites be c_{11}, the count of haplotypes carrying the reference allele at the first site and the non-reference allele at the second site be c_{10}, and so on for c_{01} and c_{00}. We define the analogous frequencies f_{11}, f_{10}, f_{01}, and f_{00}. With this notation, $\hat{R} = \frac{f_{11}-(f_{10}+f_{11})(f_{01}+f_{11})}{(f_{10}+f_{11})(1-(f_{10}+f_{11}))}$. Using the delta method [Agresti, 2002]:

$$Var(\hat{R}) \approx \Delta \Sigma \Delta^T, \tag{48}$$

where Δ is the vector of partial derivatives of $\hat{R}$ with respect to f_{11}, f_{10}, and f_{01}, and Σ is the covariance matrix of these frequencies. Explicitly, Δ is:

$$\frac{\partial \hat{R}}{\partial f_{11}} = \frac{f_{10}(1-f_{1.})^2 - f_{01}f_{1.}^2}{f_{1.}^2(1-f_{1.})^2} \tag{49}$$

$$\frac{\partial \hat{R}}{\partial f_{10}} = \frac{-f_{01}f_{1.}^2 - f_{11}(1-f_{1.})^2}{f_{1.}^2(1-f_{1.})^2} \tag{50}$$

$$\frac{\partial \hat{R}}{\partial f_{01}} = \frac{1}{f_{1.}-1}, \tag{51}$$

where $f_{1.} = f_{11} + f_{10}$. Additionally, Σ is:

$$Cov(f_i, f_j) = \begin{cases} \frac{f_i(1-f_i)}{2M} & \text{if } i = j \\ \frac{-f_i f_j}{2M} & \text{if } i \neq j \end{cases}. \tag{52}$$

We can thus calculate $Var(\hat{R})$, and plug this into Equation 47 to estimate the variance in the estimate of the conditional effect size.

Multivariate conditional Bayes factors. We note that the above calculation are for a single phenotype. For two phenotypes, we simply perform the above procedure for each individually. This gives a corrected Z-score for each of the two phenotypes, which can then be plugged into any of the calculations for our approximate Bayes factors.

2.8.2 Regional Bayes factors

We can now write down the Bayes factors measuring the support for an association in a given genomic region r. As in the situation where the two traits have been measured in separate cohorts,

$$RBF_r^{(1)} = \sum_{i=1}^{K} \pi_i^{(1)} BF_i^{(1)} \tag{53}$$

$$RBF_r^{(2)} = \sum_{i=1}^{K} \pi_i^{(2)} BF_i^{(2)} \tag{54}$$

$$RBF_r^{(3)} = \sum_{i=1}^{K} \pi_i^{(3)} BF_i^{(3)} \tag{55}$$

$$RBF_r^{(4)} = \sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i^{(1)} \pi_j^{(2)} BF_i^{(1)} BF_j'^{(2)} I[i \neq j], \tag{56}$$

$$\tag{57}$$

where $I[i \neq j]$ is an indicator that evaluates to 1 if i and j are different and 0 otherwise, $\pi_i^{(1)}$ is the prior probability that SNP i is the causal SNP under model 1, $\pi_i^{(2)}$ is the prior probability that SNP i is the causal SNP under model 2, and $\pi_i^{(3)}$ is the prior probability that SNP i is the causal SNP under model 3. Additionally, $BF_j'^{(2)}$ is the Bayes factor at SNP j conditional on the observed effects at SNP i, calculated using the procedure in Section 2.8.1. Note the only difference to the method that does not account for LD is the calculation of $RBF_r^{(4)}$.

2.8.3 Application

We applied this method to several genome-wide association studies that we know were performed on the same cohorts. We calculated R_{ij} between all pairs of SNPs using the haplotypes from the European-descent individuals in the 1000 Genomes Project as $\frac{\hat{V}_{ij}}{\hat{V}_{ii}}$, where both of these were estimated using the shrinkage estimator of Wen and Stephens [2010] and the recombination rate estimates from the HapMap Project as in Berisa and Pickrell [2015]. We ran the model in three modes: 1) assuming the cohorts were completely separate, 2) accounting for the correlation in the summary statistics under the null but not linkage disequilibrium, and 3) accounting for both the correlation in the phenotypes and linkage disequilibrium. Shown in Supplementary Figure 5 are the parameter estimates in these three situations for four pairs of GWAS. In general, correcting for LD made little difference in practice. For all of the main analyses, we only corrected for the correlation in summary statistics under the null.

3 Causal inference

The observation that a genetic variant influences two traits can be interpreted in the "Mendelian randomization" framework as evidence that one trait causally influences the other. However, using this framework requires the strong assumption that the variant does not influence the two traits via independent mechanisms.

Our goal was to develop a robust method for measuring the evidence in favor of a causal relationship between two traits using data from many genetic associations, while recognizing that strong conclusions are likely impossible in this setting. Our aim was not to estimate the magnitude of a causal effect (should one exist), but rather to simply determine if such an effect exists.

Our motivating example comes from LDL cholesterol and heart disease risk–if we identify variants that influence LDL levels, these variants have correlated effects on heart disease risk (Figure 5 in the main text). However, if we identify genetic variants that influence heart disease, these variants do not have correlated effects on LDL levels (Figure 5 in the main text). The intuition is as follows: if a trait X causally influences trait Y, then to a first approximation every genetic variant that influences trait X should also influence trait Y, and the effect sizes of these variants on the two traits should be correlated. The reverse, however, is not true: genetic variants that influence trait Y do not necessarily influence trait X, since Y can be influenced by mechanisms independent of X. (Assume that X is one of a large number of factors that causally influence Y, such that most of the variants that influence Y do not act through X).

To scan through all pairs of traits, we aimed to formulate this intuition in a manner that allows for automation. Related work has been done on Mendelian randomization with multiple genetic variants [Davey Smith and Hemani, 2014; Do et al., 2013; Evans et al., 2013] and "reciprocal" Mendelian randomization [Timpson et al., 2011]. We assume we have identified a set of N_X genetic variants that influence X (without using information about Y). Assume we have also identified a set of N_Y genetic variants that influence Y (without using information about X). Let $\vec{\beta}_{XX}$ be the vector of effect sizes on trait X for the set of variants ascertained through the association study of X, and $\vec{\beta}_{XY}$ be the vector of the effect sizes of these variants on trait Y. Define $\vec{\beta}_{YY}$ and $\vec{\beta}_{YX}$ analogously. Now let $\hat{\rho}_X$ be the rank correlation between $\vec{\beta}_{XX}$ and $\vec{\beta}_{XY}$, and let $\hat{\rho}_Y$ be the rank correlation between $\vec{\beta}_{YY}$ and $\vec{\beta}_{YX}$. Using Fisher's Z-transformation, we can

approximate the sampling distributions of $\hat{\rho}_X$ and $\hat{\rho}_Y$. If we let $\hat{Z}_X = \text{arctanh}(\hat{\rho}_X)$:

$$\hat{Z}_X \sim N\left(Z_X, \frac{1}{N_X - 3}\right) \tag{58}$$

$$\hat{Z}_Y \sim N\left(Z_Y, \frac{1}{N_Y - 3}\right). \tag{59}$$

We can thus define an approximate likelihood for the two correlation coefficients:

$$L(Z_X, Z_Y; \hat{Z}_X, \hat{Z}_Y, N_X, N_Y) = N\left(\hat{Z}_X; Z_X, \frac{1}{N_X - 3}\right) N\left(\hat{Z}_Y; Z_Y, \frac{1}{N_Y - 3}\right), \tag{60}$$

where $N(x; \mu, \sigma^2)$ is the density of a normal distribution with mean μ and variance σ^2 evaluated at x.

Now we define the models we would like to compare:

1. M1: if trait X causes Y, then we estimate Z_X and set $Z_Y = 0$.
2. M2: If trait Y causes X, then we estimate Z_Y and set $Z_X = 0$.
3. M3: If there are no relationships between the traits, then $Z_X = Z_Y = 0$.
4. M4: If the correlation does not depend on how the variants were ascertained, $Z_X = Z_Y$.

The first two models on this list we think of as causal models, by the line of reasoning outlined at the beginning of this section. The third model we obviously interpret as a non-causal model. The fourth we also interpret as non-causal, though this perhaps merits some discussion (see below). We fit each model by maximizing the corresponding approximate likelihood.

To compare the models, we calculate the Akaike information criterion (AIC) for each, where the numbers of parameters is 1, 1, 0, and 1, respectively, for the four models above. We then choose the smallest AIC from the two causal models (AIC_{causal}) and the smallest AIC from the two non-causal models ($AIC_{noncausal}$). We then calculate the relative likelihood of these two models:

$$r = \exp\left(\frac{AIC_{causal} - AIC_{noncausal}}{2}\right). \tag{61}$$

This is the relative likelihood of the best non-causal model compared to the best causal model. In Figure 5 in the main text, we show the four pairs of traits where this ratio is less than 0.01, and in Supplementary Figure 12, we show 10 additional pairs of traits where this ratio is less than 0.05. As we can see visually, this model successfully identifies patterns that look similar to our motivating example of LDL and heart disease.

A key caveat in interpretation of this method is that we may not have measured the truly causal phenotype, but rather some proxy for it. For example, if it is not BMI *per se* that causally influences risk of type 2 diabetes, but rather some other measure of adiposity that is highly correlated to BMI (and shares the same underlying genetic basis), then we have no way of detecting this. We suspect that more detailed phenotyping will identify "clusters" of highly correlated traits that will be difficult to disentangle.

3.1 Implications of looking explicitly for asymmetry.

We have set up this model in the context where a causal trait is one of many factors that influences a downstream trait. This induces the asymmetry we try to detect. However, it is possible that the "causal" trait is the *major* factor that influences the "caused" trait. For example, consider type 2 diabetes and fasting glucose levels. Clearly any factor that increases fasting glucose increases risk of type 2 diabetes (and vice versa), just by virtue of the definition of the disease. This type of causal relationship will be missed by

this approach. On the other hand, consider the two measures of bone mineral density. Any factor that increases one will also almost certainly increase the other, because the two phenotypes are closely related at a molecular level. We would not consider this a causal relationship between the traits, but rather that the two traits are measurements of a single underlying variable (namely overall bone density). We prefer to miss causal relationships of the first kind in order to avoid the interpretive difficulties of the second case.

3.2 Simulations

To test the behavior of this method, we performed simulations. Each simulation was set up as follows: we simulate two sets of bivariate normally distributed points, then use the procedure described above to detect an asymmetry in the Spearman correlation coefficients in the two sets. Note that this does not simulate all aspects of a genetic association study, rather just the estimated effect sizes.

3.2.1 Simulations under the null.

To test the distribution of our test statistic under the null model, we performed simulations of a situation with no causal relationship between two traits. Specifically, we simulated N_1 bivariate normally distributed random variables X_1 and X_2, and then N_2 bivariate normally distributed random variables Y_1 and Y_2. The means of all simulated random variables were set to zero and the variances to 1. We then applied the method from the main text to obtain the relative likelihood in favor of a "non-causal" model versus a "causal" model. In Supplementary Figure 15, we show the distribution of this ratio for different values of N_1 and N_2. Note that when N_1 is small (for example, five identified variants) the normal approximation breaks down; this is reflected in a number of simulations where the test statistic is unreasonably small (that is, the support for causality is unreasonably large).

3.2.2 Power simulations.

To test the power of our approach, we then performed simulations under a model that approximates the situation where one trait is causally upstream of another. Specifically, we simulated N_1 bivariate normally distributed random variables X_1 and X_2 with means of zero, variances of one, and correlation coefficient C. We then simulated N_2 bivariate normally distributed random variables Y_1 and Y_2 with means zero, variances one, and correlation coefficient zero. This approximates the asymmetric situation we expect if trait 1 causes trait 2. We then applied the method from the main text to obtain the relative likelihood of a "non-causal" model versus a "causal" model, and calculated the power as the fraction of simulations where this test statistic was less than 0.01. Shown in Supplementary Figure 14 is this estimate of the power for different numbers of variants and correlation coefficients. It is important to note that the correlation we estimate (either in the simulations or in the analysis of real data) is not corrected for sampling error in the effect sizes, and so will be biased downwards from the "true" genetic correlation.

3.3 Expanded analysis of putative causally-related traits

As noted in the main text, in our initial scan for putative causally-related traits, we identified five pairs of traits. One of these was a putative causal relationship between risk of coronary artery disease and risk of rheumatoid arthritis (Supplementary Figure 18). In the data used, variants that increase risk of CAD tend to decrease risk of RA (Supplementary Figure 18A), while variants that influence RA appear to have no correlated effects on CAD (Supplementary Figure 18B). This echoes the patterns seem for the other phenotypes in Figure 5 in the main text.

However, three issues caused us to worry this correlation might be spurious. First, the number of variants identified in the GWAS of CAD was small–eleven, compared to 30 for BMI, 30 for hypothyroidism, and

41 for LDL cholesterol. Second, there appears to be no overall genome-wide genetic correlation between risk of CAD and risk of RA [Bulik-Sullivan et al., 2015]. Third, this observed correlation is in the opposite direction to the known epidemiological correlation where individuals with rheumatoid arthritis are at higher risk of coronary artery disease (e.g. Maradit-Kremers et al. [2005]).

We thus identified an expanded set of loci that influence CAD risk using a larger GWAS performed using a targeted genotyping array [CARDIoGRAMplusC4D Consortium et al., 2013]. We downloaded the 79,129 summary statistics from the CARDIoGRAMplusC4D Metabochip study [CARDIoGRAMplusC4D Consortium et al., 2013] from `http://www.cardiogramplusc4d.org/downloads/`, and ran fgwas as on the original GWAS (in this case without performing any imputation). This analysis identified 70 loci associated to CAD risk at a false discovery rate of 10%. We then compared the effect sizes of these variants on CAD risk with their effect sizes on RA risk (Supplementary Figure 18C). The negative correlation from Supplementary Figure 18A ($\rho = -0.91, P = 5.6 \times 10^{-5}$) is no longer apparent in Supplementary Figure 18C ($\rho = -0.21, P = 0.07$). We conclude that this apparent causal relationship between risk of CAD and risk of RA was likely a false positive driven by small numbers of variants associated with CAD.

We next wished to test the robustness of the other inferred causal relationships. To do this, we used the fact that we discarded variants typed on the Metabochip from Locke et al. [2015] for our main analysis. This was done to allow for imputation at the level of summary statistics, but resulted in a significant loss of power to identify variants since many individuals were typed using only this genotyping array. We thus re-processed the data from Locke et al. [2015] as described in the main text, except that we 1) did not perform imputation, but 2) included variants typed on the Metabochip. This increased the number of loci identified at a false discovery rate of 10% from 30 to 75. We plotted the effects of these variants on BMI and triglyceride levels (Supplementary Figure 17A), and on BMI and type 2 diabetes (Supplementary Figure 17A). The evidence for a correlation in the effect sizes of the set of 30 variants on BMI and triglyceride levels ($\rho = 0.74, P = 5.9 \times 10^{-6}$) remained strong in the set of 75 loci ($\rho = 0.59, P = 7.1 \times 10^{-8}$). Likewise, the evidence for a correlation in the effect sizes of the set of 30 variants on BMI and risk of type 2 diabetes ($\rho = 0.78, P = 1.6 \times 10^{-6}$) remained strong in the set of 75 loci ($\rho = 0.64, P = 6.2 \times 10^{-9}$).

Interestingly, in both cases a single locus appeared as an outlier–in the case of BMI and triglycerides, a variant near APOE was associated with decreased BMI but increased levels of triglycerides (Supplementary Figure 17A), while in the case of BMI and type 2 diabetes, a variant near TCF7L2 was associated with decreased BMI but highly increased risk of type 2 diabetes (Supplementary Figure 17A, see also Helgason et al. [2007]). This suggests that both the APOE locus and the TCF7L2 locus may harbor variants that influence the downstream phenotypes (triglycerides and type 2 diabetes risk, respectively) through mechanisms that are independent of BMI. More generally, variants near TCF7L2 have been used in a number of Mendelian randomization studies of type 2 diabetes (e.g. Pfister et al. [2011]; Song et al. [2012]; Villareal et al. [2010]); if this locus affects multiple phenotypes via different molecular mechanisms then it may not be an appropriate "instrument" for this type of analysis.

4 23andMe GWAS Methods

23andMe GWAS Methods

23andMe, Inc.

Introduction

This document describes the methods used by 23andMe to conduct GWAS in the 23andMe participant cohort.

Methods

All 23andMe participants included in the analyses provided informed consent to take part in this research and answered surveys online under a human subjects protocol reviewed and approved by Ethical & Independent Review Services, an AAHRPP-accredited institutional review board (http://www.eandireview.com).

National Genetics Institute (NGI), a CLIA licensed clinical laboratory and a subsidiary of Laboratory Corporation of America extracted DNA from saliva samples and preformed genotyping for 23andMe. Genotyping was conducted on one of four genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550+ BeadChip, including about 25,000 custom SNPs selected by 23andMe, with a total of about 560,000 SNPs. The V3 platform was based on the Illumina OmniExpress+ BeadChip, with custom content to improve the overlap with our V2 array, with a total of about 950,000 SNPs. The V4 platform in current use is a fully custom array, including a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation, and about 570,000 SNPs. Samples that failed to reach 98.5% call rate were re-analyzed. 23andMe participants whose analyses failed repeatedly were re-contacted by 23andMe customer service to provide additional samples, as is done for all 23andMe customers.

For our standard GWAS, we restricted participants to a set of individuals who have >97% European ancestry, as determined through an analysis of local ancestry[1]. Briefly, our algorithm first partitioned phased genomic data into short windows of about 100 SNPs. Within each window, we used a support vector machine (SVM) to classify individual haplotypes into one of 31 reference populations. The SVM classifications were then fed into a hidden Markov model (HMM) that accounted for switch errors and incorrect assignments, and gave probabilities for each reference population in each window. Finally, we used simulated admixed individuals to recalibrate the HMM probabilities so that the reported assignments are consistent with the simulated admixture proportions. The reference population data was derived from public datasets (the Human Genome Diversity Project, HapMap, and 1000 Genomes), as well as 23andMe participants who have reported having four grandparents from the same country.

A maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent (IBD) estimation algorithm[2]. Individuals were defined as related if they shared more than 700 cM IBD, including regions where the two individuals share either one or both genomic segments identical-by-descent. This level of relatedness (roughly 20% of the genome) corresponds approximately to the minimal expected sharing between first cousins in an outbred population.

Participant genotype data were imputed against the March 2012 "v3" release of 1000 Genomes reference haplotypes[3]. We phased and imputed data for each genotyping platform separately. First, we used Beagle[4] (version 3.3.1) to phase batches of 8000-9000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with Hardy-Weinberg equilibrium $P<10^{-20}$, call rate < 95%, or with large allele frequency discrepancies compared to European 1000 Genomes reference data. Frequency discrepancies were identified by computing a 2x2 table of allele counts for European 1000 Genomes samples and 2000 randomly sampled 23andMe participants with European ancestry, and identifying SNPs with a chi squared $P<10^{-15}$. We imputed each phased segment against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using Minimac2[5], using 5 rounds and 200 states for parameter estimation.

For the non-pseudoautosomal region of the X chromosome, males and females were phased together in segments, treating the males as already phased; the pseudoautosomal regions were phased separately. We then imputed males and females together using minimac, as with the autosomes, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region.

For case control comparisons, we computed association test results by logistic regression assuming additive allelic effects. For tests using imputed data, we used the imputed dosages rather than best-guess genotypes. We typically include covariates for age, gender, and the top five principal components to account for residual population structure. The association test *P* value we report was computed using a likelihood ratio test, which in our experience is better behaved than a Wald test on the regression coefficient. For quantitative traits, association tests were performed by linear regression. Results for the X chromosome were computed similarly, with male genotypes coded as if they were homozygous diploid for the observed allele.

HLA allele dosages have been imputed from SNP genotype data using HIBAG[6]. We imputed alleles for HLA-A, B, C, DPB1, DQA1, DQB1, and DRB1 loci at four-digit resolution. To test associations between HLA allele dosages and phenotypes, we performed logistic or linear regression using the same set of covariates used in the SNP based GWAS for that phenotype. We performed separate association tests for each imputed allele.

Association Test Results

For reporting associations of SNPs, when choosing between imputed and genotyped GWAS results, if either the imputed test passes quality control, or a genotyped test is unavailable, we report the imputed result; otherwise, we report the genotyped result.

For quality control of genotyped GWAS results, we flagged SNPs that were only genotyped on our "V1" platform due to small sample size, and SNPs on chrM or chrY because many of these are not currently called reliably. Using trio data, we flagged SNPs that failed a test for parent-offspring transmission; specifically, we regressed the child's allele count against the mean parental allele count and flagged SNPs with fitted $\beta<0.6$ and $P<10^{-20}$ for a test of $\beta<1$.

We flagged SNPs with a Hardy-Weinberg $P<10^{-20}$ in Europeans; or a call rate of <90%. We also tested genotyped SNPs for genotype date effects, and flagged SNPs with $P<10^{-50}$ by ANOVA of SNP genotypes against a factor dividing genotyping date into 20 roughly equal-sized buckets.

For imputed GWAS results, we flagged SNPs with avg.rsq<0.5 or min.rsq<0.3 in any imputation batch, as well as SNPs that had strong evidence of an imputation batch effect. The batch effect test is an F test from an ANOVA of the SNP dosages against a factor representing imputation batch; we flagged results with $P<10^{-50}$. Prior to GWAS, we identified, for each SNP, the largest subset of the data passing these criteria, based on their original genotyping platform -- either v2+v3+v4, v3+v4, v3, or v4 only -- and computed association test results for whatever was the largest passing set. As a result, there are no imputed results for SNPs that fail these filters.

Across all results, we flagged logistic regression results that did not converge due to complete separation, identified by abs(effect)>10 or stderr>10 on the log odds scale. We have also flagged linear regression results for SNPs with MAF < 0.1% because tests of low frequency variants can be sensitive to violations of the regression assumption of normally distributed residuals.

GWAS Reports

The GWAS reports include a detailed description of how each phenotype was constructed from 23andMe participant responses, and a summary of findings from the GWAS. Results reported have been adjusted for genomic control[7]. The genomic control procedure adjusts test statistics and standard errors of effect sizes to compensate for variance inflation due to residual population stratification that has not been effectively controlled through use of principal components in the regression models. The genomic control inflation factor used is included in the reports, and is computed from the median *P* value for results that passed QC.

Phenotype Statistics

These tables show participant demographics for this phenotype, computed across all 23andMe customers who have consented to participate in research, and across just the European, unrelated participants included in the GWAS.

Null model with covariates

This section shows model fitting results for covariates included in the GWAS – normally, age, gender, and the top five principal components. The form of the table depends on the model type. The table will include effect sizes and standard errors, Wald test statistics for the null hypothesis that a term's effect size is 0 (z-value) and the corresponding p-value (Pr(>|z|)), and for logistic regression, likelihood ratio test (LRT) results for dropping this term from the model and the corresponding p-value (Pr(>Chi)).

SNP-level QC information

This section includes several tables showing filters that were applied to the association test results as described above to create a cleaned result set. For genotyped data, these include filters on sample size or reliability (V1-only, chrM, chrY), parent-offspring test, minor allele frequency (MAF), Hardy Weinberg equilibrium (HWE), genotype call rate (gt.rate), and batch effects. For imputed data, these include filters on imputation quality and imputation batch effects. A final table shows counts of genotyped and imputed SNP association results that passed the filters, and how these were merged together.

Manhattan Plot

The Manhattan plot depicts the distribution of association test statistics versus genomic position, with chromosomes 1 to 22, X, and Y arranged along the X axis. The Y axis represents log-scaled *P* values. Positions with $P<5\times10^{-8}$ (a score of about 7.3) are shown in red. If the results include loci with $P<10^{-25}$ then the vertical scale is adjusted nonlinearly to preserve detail for signals near the genome-wide threshold. Up to 25 loci with smallest $P<10^{-6}$ are labeled with the name of the nearest gene. A "good" Manhattan plot should show towers of SNPs with small P values supporting most signals that pass the genome wide threshold.

Q-Q Plot

The Q-Q plot depicts observed versus expected quantiles for the GWAS P values, where the expected distribution of *P* values is uniform under the null hypothesis, plotted on a log scale. A solid red line is shown with a slope of 1, and dashed red lines represent a 95% confidence envelope under the assumption that the test results are independent. A "good" Q-Q plot follows the null distribution for larger *P* values ($P>0.01$) then diverges from the null distribution for small P values. The test statistics in the Q-Q plot have already been adjusted using genomic control.

Index SNPs for Strongest Associations

The table of index SNPs shows information for the most-associated SNP in each associated region, for at least 5 and at most 50 regions for each phenotype. We define regions by identifying SNPs with $P<10^{-5}$, then grouping these into intervals separated by gaps of at least 250 kb, and choosing the SNP with smallest *P* within each interval. The source (src) of each SNP, genotyped (G) or imputated (I), is indicated.

Column name	Definition
assay.name	dbSNP build 137 rsID when available; a small number of SNPs without good matches in dbSNP have 23andMe names like 'i12345'.
scaffold	Chromosome name, NCBI Build 37
position	Chromosomal position, NCBI Build 37
alleles	The two SNP alleles, A/B, in alphabetical order
src	Source of SNP – genotyped (G) or imputed (I)

The gene.context field is constructed using the HG19 release of the UCSC Known Genes tables, and has the following interpretations:

Format of gene.context	Interpretation
[Gene1,Gene2,...]	The SNP is contained within the transcripts of the specified gene(s)
Gene1--[]---Gene2	The SNP is flanked by Gene1 and Gene2. Dashes indicate distance: '' = <1kb, '-' = <10kb, '--' = <100kb, '---' = <1000kb
Gene1--[]	The SNP is flanked by Gene1 on the left, but there is no gene within 1000kb on the right.
[]---Gene2	The SNP is flanked by Gene2 on the right, but there is no gene within 1000kb on the left
[]	There are no genes within 1000kb of the SNP

Quality Statistics for Index SNPs

These quality statistics describe the index SNPs in the set of all 23andMe participants with European ancestry -- not just the ones included in this GWAS. Columns indicate which 23andMe genotyping array(s) include each variant. For genotyped SNPs, the statistics include call rate, Hardy Weinberg *P* values, and allele frequency. For imputed SNPs, the statistics include average and minimum r^2 across imputation batches, and a *P* value for a dosage dependence on batch.

Column name	Definition
is.v1, is.v2, is.v3, is.v4	Flags indicating whether this SNP is present on the 23andMe V1, V2, V3, and/or V4 genotyping platforms (0 = no, 1 = yes)
freq.b	Average dosages of the B allele in Europeans
avg.rsqr	The average imputation r^2 across all batches of imputation results, a measure of overall imputation quality
min.rsqr	The minimum imputation r^2 in any one batch of imputation results, a measure of consistency of imputation quality
p.batch	A test for an imputation batch effect: An F test from an ANOVA of the SNP genotypes against a factor representing imputation batch
qc.mask	The set of platforms for which this variant was successfully imputed (joint avg.rsqr>0.5, min.rsqr>0.3, and p.batch>1e-50): one of 'v2v3v4', 'v3v4', 'v3', or 'v4'.

SNP Statistics in the GWAS Sample

These statistics describe the index SNPs in the set of 23andMe participants included in this GWAS. For binary phenotypes, results are shown for controls and cases separately. For genotyped SNPs, the statistic includes counts for each observed genotype. For imputed SNPs, we report mean dosages.

Column name	Definition
im.num.0	Number of controls imputed for this variant

dose.b.0	Average imputed dosage of the B allele in controls
im.num.1	Number of cases imputed for this variant
dose.b.1	Average imputed dosage of the B allele in cases
AA.0, AB.0, BB.0	Counts of AA, AB, BB genotypes in controls
AA.1, AB.1, BB.1	Counts of AA, AB, BB genotypes in cases

For quantitative traits, **im.num.0**, **dose.b.0**, **AA.0**, **AB.0**, and **BB.0** describe all individuals included in the GWAS; and **im.num.1**, **dose.b.1**, **AA.1**, **AB.1**, and **BB.1** are set to NA.

Annotations from NHGRI GWAS Catalog

This table shows results of looking up our association index SNPs in the NHGRI GWAS Catalog[8], to identify other reports of associations, for any phenotype, with SNPs that are within 500 kb and in moderate to high linkage disequilibrium (r^2>0.5) with one of our index SNPs.

Replication of GWAS Catalog Results

In this table, we take all previous reports of associations with closely related phenotypes from the NHGRI GWAS Catalog, and show association results for our best proxy, defined as the SNP with highest r^2, within 100 kb and with r^2>0.8. This can identify loci with support in the 23andMe analysis that did not make it into the list of top loci.

Nearby Nonsynonymous SNPs

In this table, we show cases where an index SNP is within 500 kb and r^2>0.5 with a SNP that leads to an alteration in a protein's amino acid sequence. Coding SNP annotations were taken from the UCSC Genome Browser's snp138CodingDbSnp table.

Nearby Expression QTLs

Here, we show cases where an index SNP is within 500 kb and r^2>0.5 with a SNP that has been reported to be associated with expression of a nearby gene (an expression QTL or eQTL). The following datasets were included in the analysis:

Publication	Tissue(s)
Stranger et al., 2007	Lymphoblastoid
Schadt et al., 2008	Liver
Dimas et al., 2009	Lymphoblastoid, fibroblast, T cell
Montgomery et al., 2010	Lymphoblastoid
Gibbs et al., 2010	Cerebellum, frontal cortex, pons, temporal cortex
Zeller et al., 2010	Monocyte
Fairfax et al., 2012	Monocyte, B cell
Lappalainen et al., 2013	Lymphoblastoid

Regional Association Plots

The regional association plots show association test statistics versus position in the vicinity of the strongest associations. The plots are generated with LocusZoom[9], using linkage

disequilibrium data from the March 2012 release of 1000 Genomes data. To preserve detail, results with $P<10^{-100}$ are set to 10^{-100}. In the plots, a 'o' symbol indicates a genotyped SNP and a '+' indicates an imputed SNP. Color indicates strength of linkage disequilibrium with the index SNP.

Privacy Considerations

It is widely acknowledged that aggregated genomic datasets such as GWAS results pose privacy risks and the possibility of re-identification of individual participants[10]. We use several strategies to reduce the likelihood of a privacy breach. Similar to policies applied by dbGAP, we require that genome-wide association test results and genome-wide platform statistics can only be used under the terms of a data transfer agreement that does not permit use of the data to re-identify individuals and restricts sharing of the data with third parties. Statistics for up to 10,000 SNPs can be published since this amount of data is considered to be insufficient to enable a re-identification attack.

References

1. Durand, E.Y., Do, C.B., Mountain, J.L., Macpherson, J.M. Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. http://biorxiv.org/content/early/2014/10/18/010512.

2. Henn, B.M., Hon, L., Macpherson, J.M., Eriksson, N., Saxonov, S., Pe'er, I. & Mountain, J.L. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* **7(4)**: e34267.

3. Durbin, R.M. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).

4. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).

5. Fuchsberger, C., Abecasis, GR, & Hinds, DA. Minimac2: faster genotype imputation. *Bioinformatics* doi: 10.1093/bioinformatics/btu704 (2014).

6. Zheng, X., Shen, J., Cox, C., Wakefield, J., Ehm, M., Nelson, M., Weir, B.S.. HIBAG – HLA Genotype Imputation with Attribute Bagging. *Pharmacogenomics J*. **14**, 192-200 (2013). http://students.washington.edu/zhengx/HIBAG/

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

8. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. **42**(Database issue): D1001-6 (2014).

   http://www.genome.gov/gwastudies/

9. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., & Willer, C.J. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26**: 2336-2337 (2010).

10. Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., & Craig, D.W. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet*. **4**, e1000167.

5 Supplementary Tables

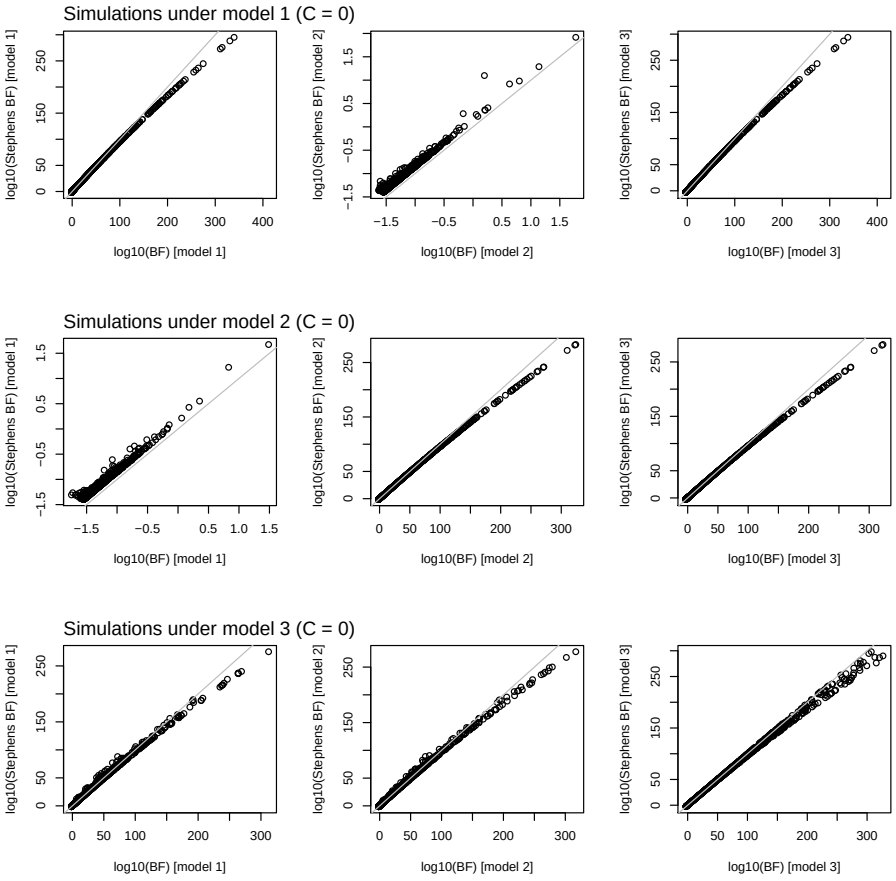
See file table_S1.tab

Supplementary Table 1. **Genomic regions that contain a variant that influences more than one trait.** We identified all genomic regions that contain a variant that influences more than one trait, at a threshold of a posterior probability of association greater than 0.9. Listed are the position of each locus, the phenotypes associated with the locus, and the lead SNPs for all phenotypes.

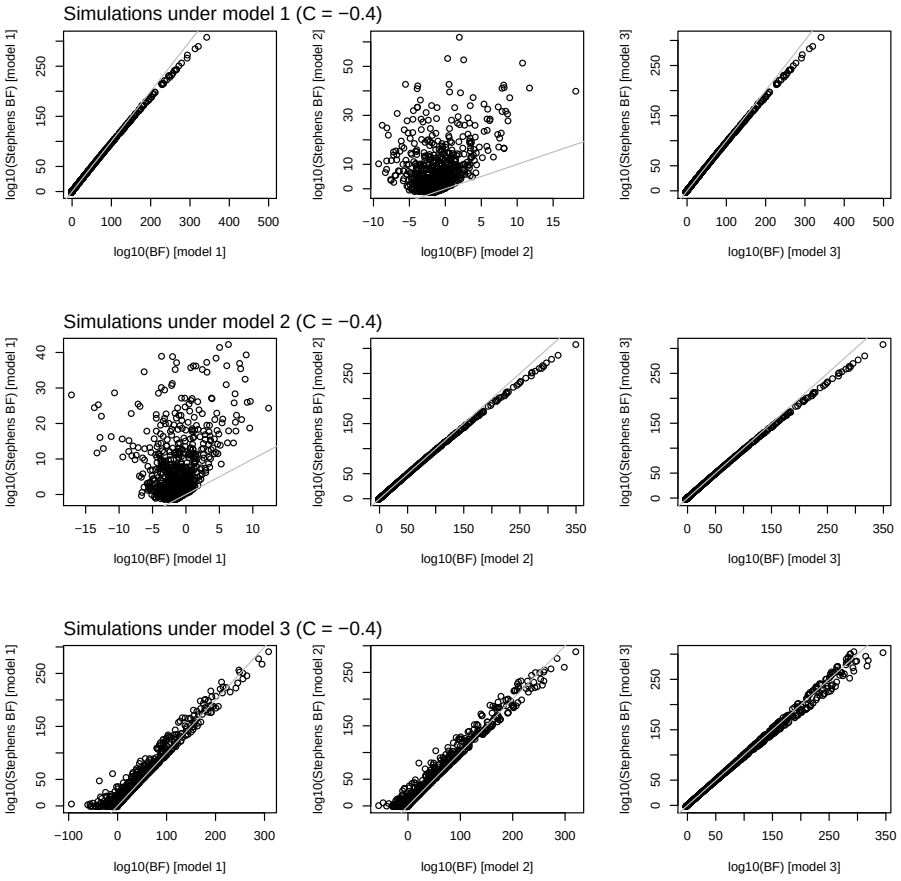
chromosome	start	stop	putative causal gene	phenotypes	notes
chr12	110336719	113263518	SH2B3	ALL, PLT, FNBMD, PCV, TS, CD, RA, HB, HTHY, LDL, HDL, RBC, TC, CAD	often led by nonsynonymous polymorphism rs3184504, maybe two signals
chr9	135298842	137041122	ABO	ALL, PCV, TS, ATH, LDL, MIGR, CEI, HB, RBC, TC, CAD	putatively driven by an eQTL at ABO or type A1 blood, see Figure in main text
chr2	26894985	28598777	GCKR	PLT, CUP, LDL, CD, FG, TG, HEIGHT, NST, TC	often led by nonsynonymous SNP rs1260326
chr16	27445755	29036613	none	ALL, BMI, TS, ATH, CD, RA, PD, NST, EDU	large LD block at 28-29Mb
chr11	58780549	62223771	none	PLT, HEIGHT, LDL, CD, FG, TG, HDL, TC	LD block covers FADS1, FADS2
chr6	89973052	91843196	none	ALL, ATH, CD, DIMP, SCZ, HTHY, UC	covers BACH2
chr6	33236497	35455756	none	CUP, BMI, LDL, MIGR, HDL, TC, CAD	LD block from 34.5-35Mb
chr6	125424383	127540461	none	UB, AAM, HEIGHT, EDU, MIGR, MPB	covers CENPW
chr4	100678360	103221356	SLC39A8	ALL, HEIGHT, CD, PD, SCZ, HDL, NST	often led by nonsynonymous SNP rs13107325
chr22	19912358	22357325	none	HEIGHT, CD, RA, MCV, HDL, TC, UC	covers UBE2L3, lots of gaps in SNP coverage
chr19	44744108	46102697	APOE	AD, LDL, WHR, TG, HDL, TC, NST	covers APOE
chr1	7247335	9365199	none	ALL, AAM, FNBMD, ATH, PD, SCZ, UC	LD block covers RERE, SLC45A1
chr16	53382572	55903774	IRX3/IRX5	CUP, AAM, BMI, T2D, AVD, HDL	intron of FTO
chr10	63341695	65794114	none	PLT, BMI, HEIGHT, MPV, TG, HDL , EDU	large LD block from 65Mb-65.5Mb
chr8	79132861	81956395	none	TS, ATH, RA, EDU, ALL, UC	covers ZBTB10
chr15	58441366	59694116	none	PCV, HB, TG, HDL, RBC, TC	nearest gene is LIPC
chr4	87534648	89238028	none	PCV, HB, TG, HDL, RBC, TC	large LD block 87.5-88Mb
chr3	139954597	141339097	none	ALL, MPB, CD, MCV, HEIGHT, NST	LD block covers ACFL2, ZBTB38, RASA2

Supplementary Table 2. **Genomic regions with a variant that influences a large number of phenotypes in these data.** We sorted all genomic regions in Supplementary Table 1 (excluding the MHC region from 26-34Mb on chromosome 6, and merging the two GWAS of age at menarche) based on the number of phenotypes a single variant was predicted to influence. Shown are all regions with a variant that influences six or more traits. The "putative causal gene" is listed only when there is a nonsynonymous SNP or otherwise functionally well-characterized allele among the strongest associations in the region.

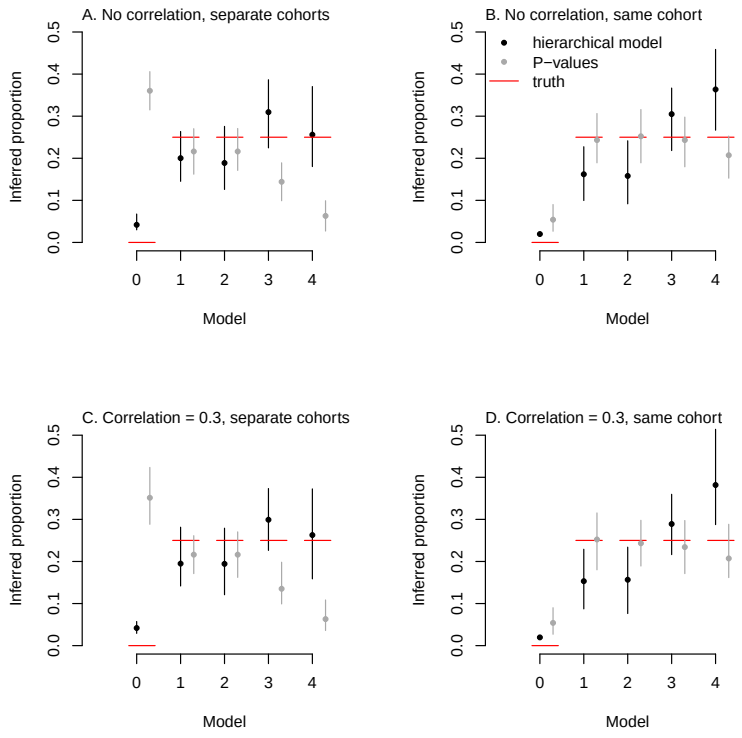
6 Supplementary Figures



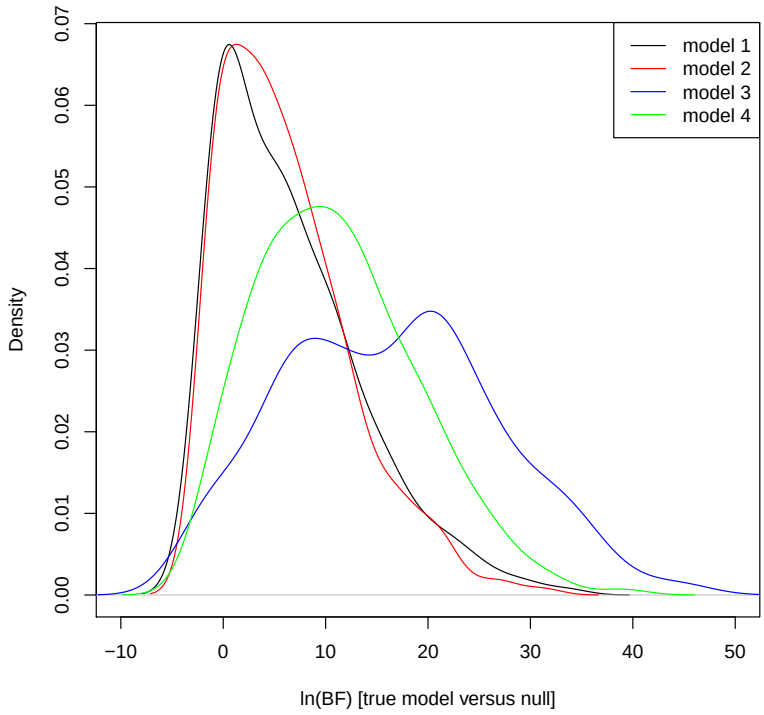
Supplementary Figure 1. **Comparison to Stephens' Bayes factor for uncorrelated traits.** We simulated genotypes and phenotypes for 5,000 individuals, and compared the approximate Bayes factor used here to that from Stephens [2013]. In each panel, each point is a single simulation.



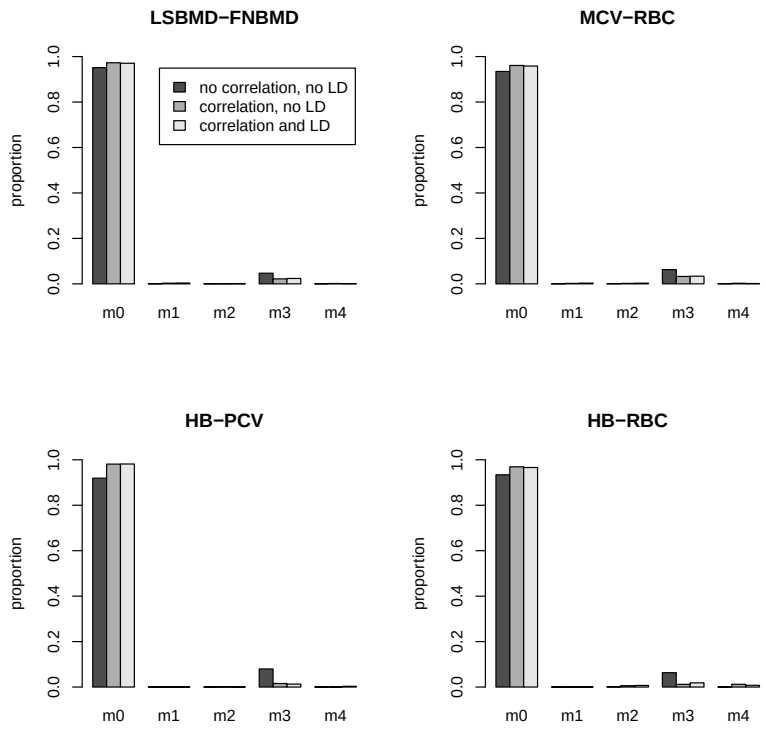
Supplementary Figure 2. **Comparison to Stephens' Bayes factor for correlated traits.** We simulated genotypes and phenotypes for 5,000 individuals, and compared the approximate Bayes factor used here to that from Stephens [2013]. In each panel, each point is a single simulation.



Supplementary Figure 3. **Simulations.** We simulated GWAS under four different assumptions, and then compared our model to a simple P-value based model. In each panel, we show the simulated values of $\Pi_{[0-4]}$ and the estimated values. Each point is the mean over 100 simulations, and the bars show the 5-95% range of parameter values obtained over these 100 simulations. **A.** Simulations of uncorrelated phenotypes in separate cohorts. **B.** Simulations of uncorrelated phenotypes in overlapping cohorts. **C.** Simulations of correlated phenotypes in separate cohorts. **D.** Simulations of correlated phenotypes in overlapping cohorts.

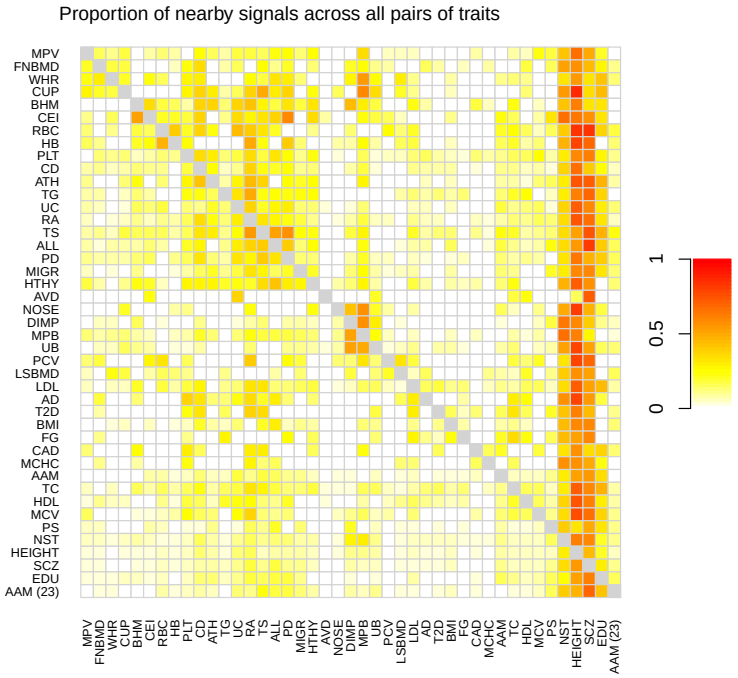


Supplementary Figure 4. **Regional Bayes factors in simulations.** For ten simulations (recall that each simulation consists of 111 regions) of separate cohorts and uncorrelated traits, we extracted all regional Bayes factors, and grouped regions according to the model under which each was simulated. We then plot the distribution of the regional Bayes factors under the "true" model. That is, for regions simulated under regional model 1, we show the distribution of RBF_1, for regions simulated under regional model 2, we show the distribution of RBF_2, and so on. Note that the distribution is shifted to the right for regions simulated under model 3 (a single genetic variant that influences both traits), indicating that there is more evidence against the null model of no associations in this situation.

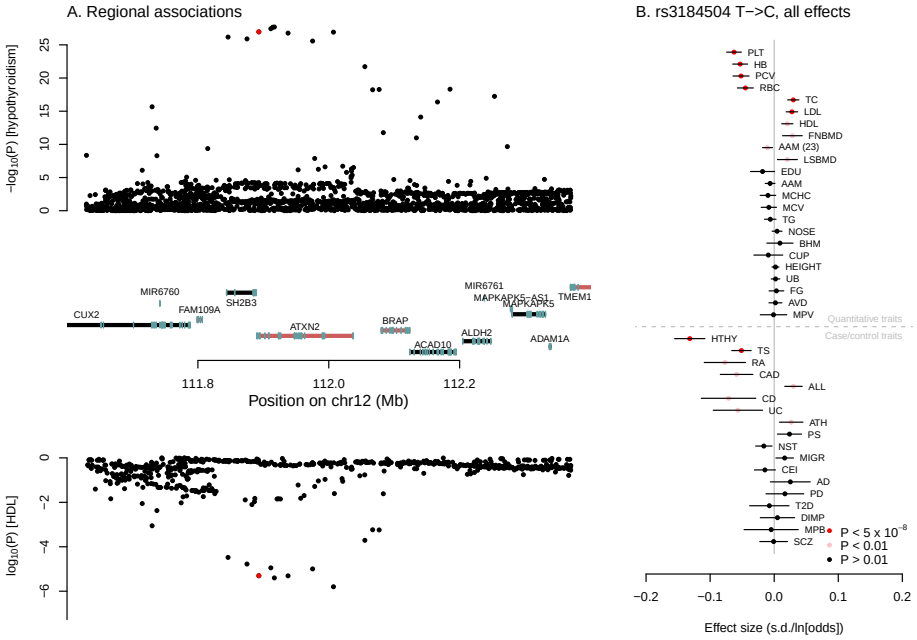


Supplementary Figure 5. **Parameter estimates in select GWAS accounting for different factors.** For the noted pairs of GWAS (which we know to have been performed on overlapping sets of individuals), we ran the model in modes where we either 1) corrected for neither the correlation in the summary statistics under the null nor for LD, 2) corrected for the correlation in the summary statistics alone, or 3) corrected for both the correlation in the summary statistics and LD. Shown are the parameter estimates for these three situations.

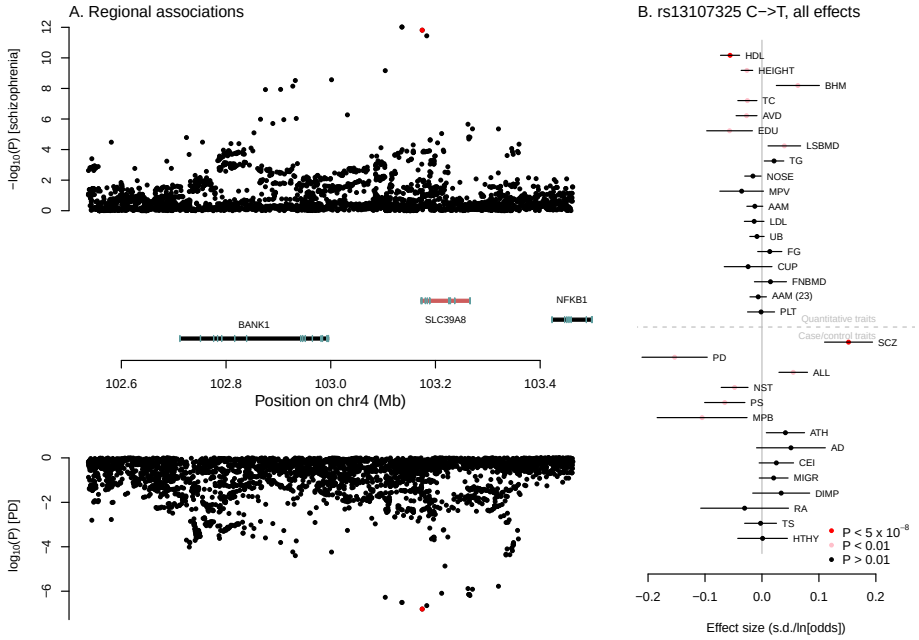
Supplementary Figure 6. **Correlations in Z-scores.** For each pair of GWAS, we identified regions with low posterior probability of having a causal variant for either trait, and then calculated the correlation in the Z-scores for SNPs in these regions. Shown are the estimated correlations for all pairs of traits, ordered from lowest to highest. Some pairs of note are in color.



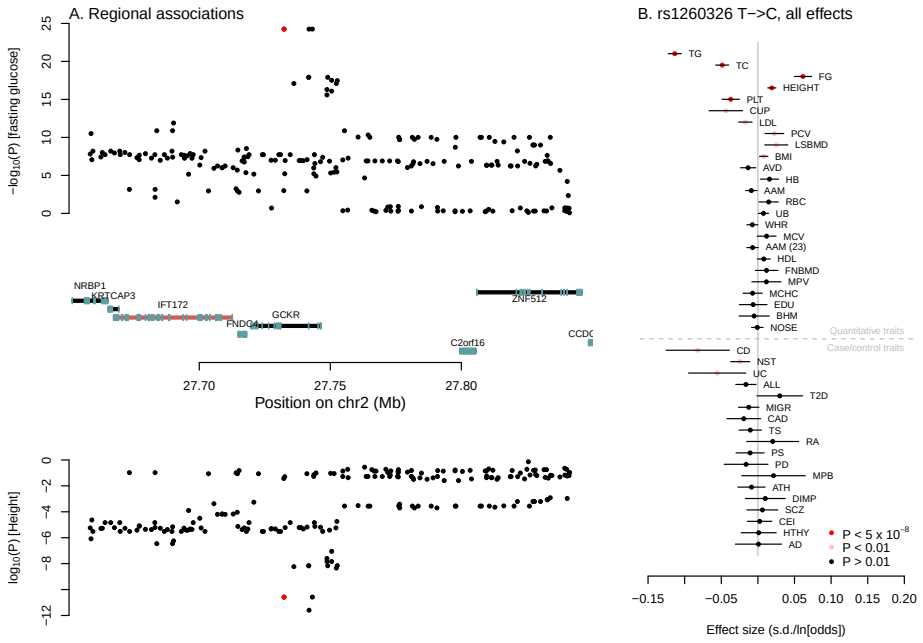
Supplementary Figure 7. **Heatmap showing how often loci that influence two traits fall near each other in the genome.** Each square $[i, j]$ shows the maximum *a posteriori* estimate of the proportion of genetic variants that influence trait i that fall near a variant that influence trait j (in the notation from the main text, this is $\frac{\hat{\Pi}_4}{\hat{\Pi}_1+\hat{\Pi}_3+\hat{\Pi}_4}$). Note that this is not symmetric. Darker colors represent larger proportions. Colors are shown for all pairs of traits that have at least one region with a posterior probability of model 4 greater than 0.9.



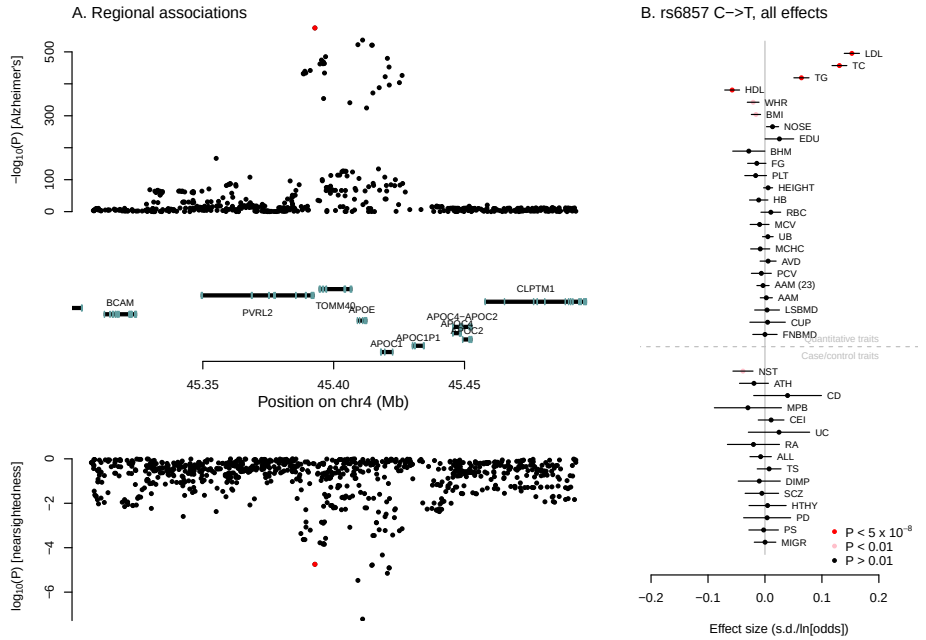
Supplementary Figure 8. **SH2B3 region. A. Regional association signals.** Shown are the P-values for association in the region for hypothyroidism and HDL cholesterol levels. In red is the nonsynonymous SNP rs3184504. **B. Effect of rs3184504 on all phenotypes.** Shown are the effects of the C allele on all phenotypes where this variant was successfully genotyped or imputed. Bars represent 95% confidence intervals.



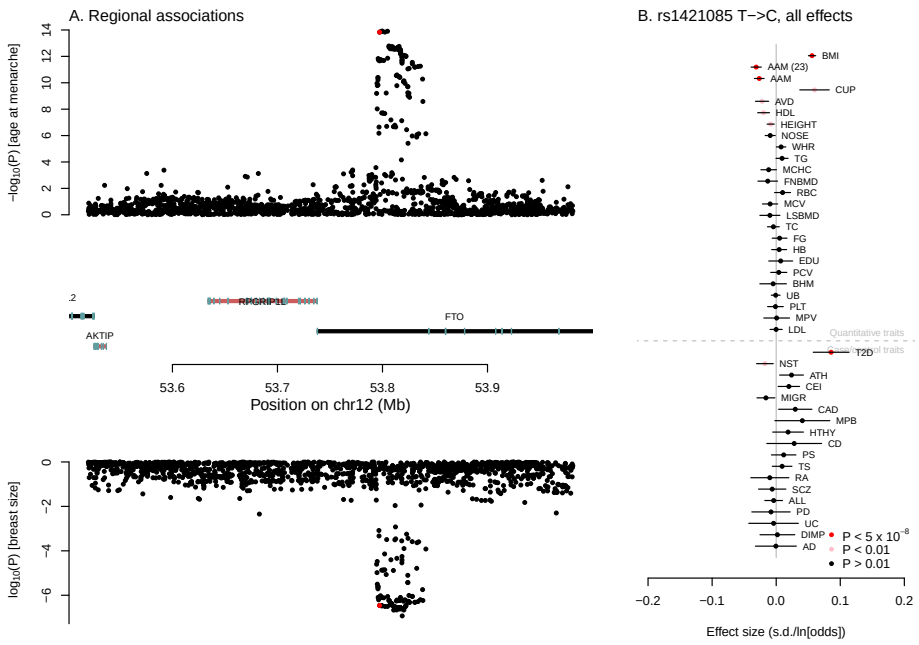
Supplementary Figure 9. **SLC39A8 region. A. Regional association signals.** Shown are the P-values for association in the region for schizophrenia and Parkinson's disease. In red is the nonsynonymous SNP rs13107325. **B. Effect of rs13107325 on all phenotypes.** Shown are the effects of the T allele on all phenotypes where this variant was successfully genotyped or imputed. Bars represent 95% confidence intervals.



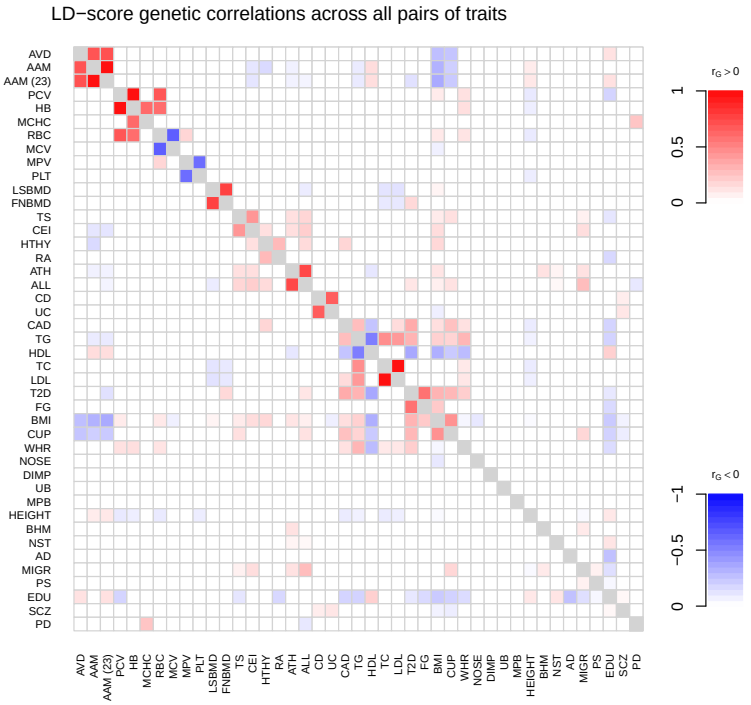
Supplementary Figure 10. **GCKR region. A. Regional association signals.** Shown are the P-values for association in the region for fasting glucose levels and height. In red is the nonsynonymous SNP rs1260326. **B. Effect of rs1260326 on all phenotypes.** Shown are the effects of the C allele on all phenotypes where this variant was successfully genotyped or imputed. Bars represent 95% confidence intervals.



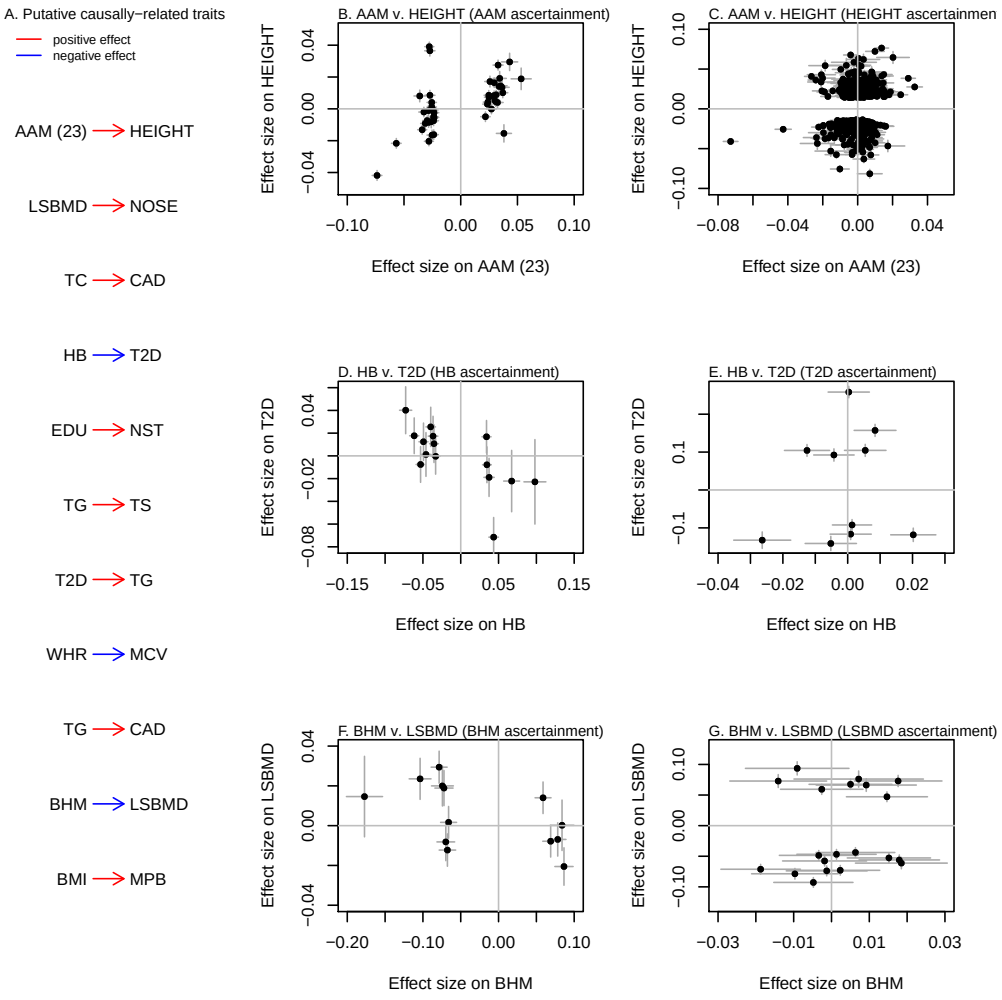
Supplementary Figure 11. **APOE region. A. Regional association signals.** Shown are the P-values for association in the region for Alzheimer's disease and nearsightedness. In red is the SNP rs6857, which tags the APOE4 allele. **B. Effect of rs6857 on all phenotypes.** Shown are the effects of the T allele on all phenotypes where this variant was successfully genotyped or imputed. Bars represent 95% confidence intervals. For ease of display we do not show is the effect size on Alzheimer's disease, which is 1.2 (on a log-odds scale).



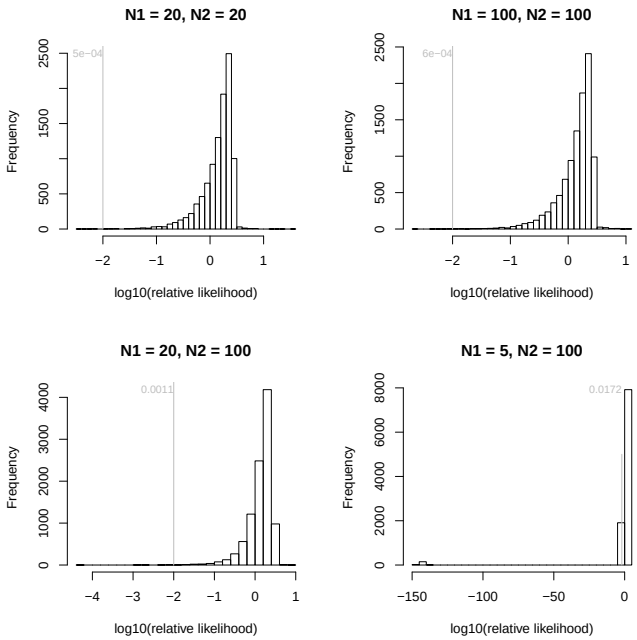
Supplementary Figure 12. **FTO region. A. Regional association signals.** Shown are the P-values for association in the region for age at menarche and breast size. In red is the SNP rs1421085. **B. Effect of rs1421085 on all phenotypes.** Shown are the effects of the C allele on all phenotypes where this variant was successfully genotyped or imputed. Bars represent 95% confidence intervals.



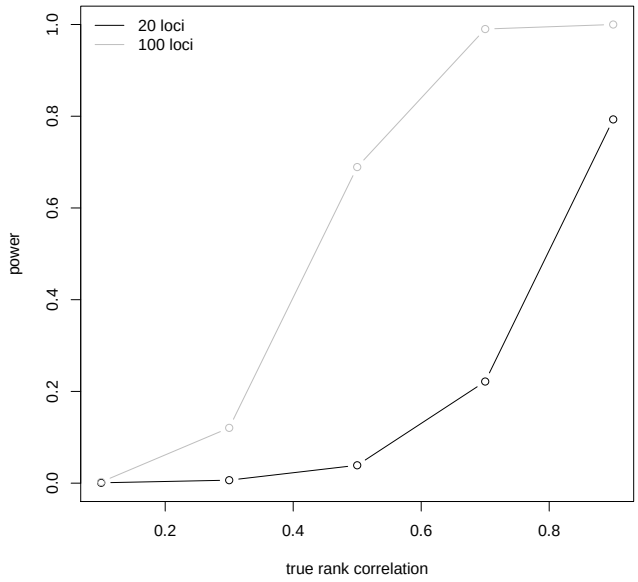
Supplementary Figure 13. **Heatmap of genetic correlations**. Each square $[i, j]$ shows the estimate of the genetic correlation between traits i and j, using the method from Bulik-Sullivan et al. [2015]. Note that this is symmetric. Darker colors represent stronger genetic correlations. Colors are shown for all pairs of traits where the genetic correlation is "significant" at a P-value threshold of 0.005.



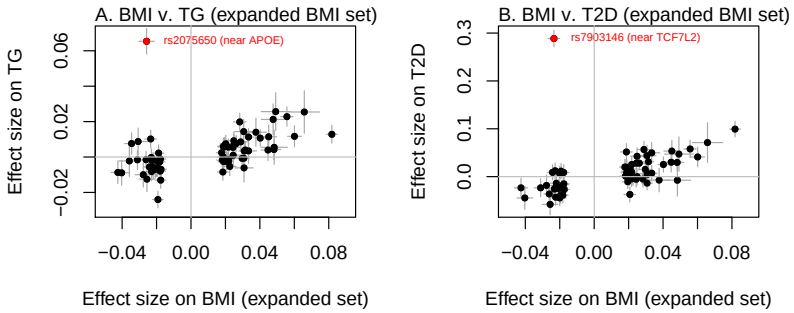
Supplementary Figure 14. **Additional pairs of traits with some evidence of a causal relationship A. List of all pairs of phenotypes.** Shown is the list of pairs of phenotypes with a relative likelihood greater than 20 (but less than 100) in favor of a causal model, ordered with the largest relative likelihood at the top. Examples of SNP effect sizes for select pairs (as in Figure 5 in the main text) are shown for age at menarche and height (**B.** and **C.**), hemoglobin levels and type 2 diabetes (**D.** and **E.**), and Beighton hypermobility and lumbar spine bone density (**F.** and **G.**)



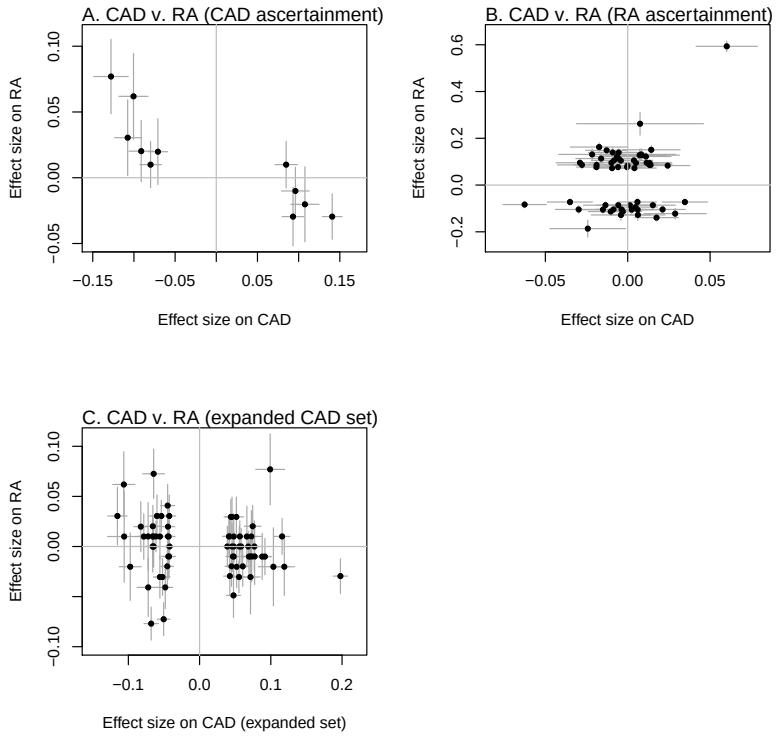
Supplementary Figure 15. **Distribution of the relative likelihood under the null.** We simulated data under the null model where the effects of a variant on two phenotypes are independent, and fit the model described in the main text. Shown is the distribution of the relative likelihood using simulations of different numbers of genetic variants. In each panel $N1$ represents the number of simulated loci that influence trait 1, and $N2$ represents the number of simulated loci that influence trait 2. Each distribution is over 10,000 simulations. The grey line shows a relative likelihood of 0.01 (the threshold used in the main text) and the grey number is the fraction of simulations with a relative likelihood less than 0.01.



Supplementary Figure 16. **Power to detect a causal relationship.** We simulated data under a model where one phenotype causes another, and fit the model described in the main text. Shown is the fraction of simulations where the relative likelihood in favor of a causal model is less than 0.01, as a function of the simulated correlation in effect sizes. Each point is a summary over 10,000 simulations.



Supplementary Figure 17. **Effect sizes of genetic variants on BMI and A. TG or B. type 2 diabetes in a larger dataset.** Shown are the effect sizes of genetic variants on BMI and TG/T2D for variants identified in a larger GWAS for BMI, see Supplementary Text for details. In red are individual SNPs of note.



Supplementary Figure 18. **Effect sizes of genetic variants on CAD and RA.** Lines represent one standard error. **A. and B. CAD and RA, initial analysis.** The effect sizes of genetic variants on CAD and RA for variants identified in the GWAS for CAD (**A.**) or RA (**B.**). **C. Effect sizes of genetic variant on CAD and RA, expanded analysis.** Shown are the effect sizes of genetic variants on CAD and RA for variants identified in a larger GWAS for CAD, see Supplementary Text for details.

References

Abecasis, G., Altshuler, D., Auton, A., Brooks, L., Durbin, R., Gibbs, R. A., Hurles, M. E., McVean, G. A., Bentley, D., Chakravarti, A., *et al.*, 2010. A map of human genome variation from population-scale sequencing. *Nature*, **467**(7319):1061–1073.

Agresti, A., 2002. *Categorical Data Analysis (Wiley Series in Probability and Statistics)*. Wiley Interscience.

Berisa, T. and Pickrell, J. K., 2015. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics*, .

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., *et al.*, 2015. An atlas of genetic correlations across human diseases and traits. *Nat Genet*, **47**(11):1236–41.

CARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R., Ingelsson, E., Saleheen, D., Erdmann, J., *et al.*, 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*, **45**(1):25–33.

Davey Smith, G. and Hemani, G., 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, **23**(R1):R89–98.

Do, C. B., Tung, J. Y., Dorfman, E., Kiefer, A. K., Drabant, E. M., Francke, U., Mountain, J. L., Goldman, S. M., Tanner, C. M., Langston, J. W., *et al.*, 2011. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet*, **7**(6):e1002141.

Do, R., Willer, C. J., Schmidt, E. M., Sengupta, S., Gao, C., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., *et al.*, 2013. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat Genet*, **45**(11):1345–52.

Eriksson, N., Benton, G. M., Do, C. B., Kiefer, A. K., Mountain, J. L., Hinds, D. A., Francke, U., and Tung, J. Y., 2012a. Genetic variants associated with breast size also influence breast cancer risk. *BMC Med Genet*, **13**:53.

Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I., and Mountain, J., *et al.*, 2010. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet*, **6**(6):e1000993.

Eriksson, N., Tung, J. Y., Kiefer, A. K., Hinds, D. A., Francke, U., Mountain, J. L., and Do, C. B., 2012b. Novel associations for hypothyroidism include known autoimmune risk loci. *PLoS One*, **7**(4):e34442.

Evans, D. M., Brion, M. J. A., Paternoster, L., Kemp, J. P., McMahon, G., Munafò, M., Whitfield, J. B., Medland, S. E., Montgomery, G. W., GIANT Consortium, *et al.*, 2013. Mining the human phenome using allelic scores that index biological intermediates. *PLoS Genet*, **9**(10):e1003919.

Ferreira, M. A. R. and Purcell, S. M., 2009. A multivariate test of association. *Bioinformatics*, **25**(1):132–3.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., *et al.*, 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**(7164):851–861.

Giambartolomei, C., Vukcevic, D., Schadt, E. E., Franke, L., Hingorani, A. D., Wallace, C., and Plagnol, V., 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*, **10**(5):e1004383.

Helgason, A., Pálsson, S., Thorleifsson, G., Grant, S. F. A., Emilsson, V., Gunnarsdottir, S., Adeyemo, A., Chen, Y., Chen, G., Reynisdottir, I., *et al.*, 2007. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet*, **39**(2):218–25.

Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., Lee, J. C., Schumm, L. P., Sharma, Y., Anderson, C. A., *et al.*, 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**(7422):119–124.

Kiefer, A. K., Tung, J. Y., Do, C. B., Hinds, D. A., Mountain, J. L., Francke, U., and Eriksson, N., 2013. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. *PLoS Genet*, **9**(2):e1003299.

Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M., 2012. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet*, **44**(9):1066–71.

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., DeStafano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., *et al.*, 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*, **45**(12):1452–8.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., *et al.*, 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, **518**(7538):197–206.

Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M., Auton, A., Myers, S., Morris, A., *et al.*, 2012. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature genetics*, **44**(12):1294–1301.

Maradit-Kremers, H., Crowson, C. S., Nicola, P. J., Ballman, K. V., Roger, V. L., Jacobsen, S. J., and Gabriel, S. E., 2005. Increased unrecognized coronary heart disease and sudden deaths in rheumatoid arthritis: a population-based cohort study. *Arthritis Rheum*, **52**(2):402–11.

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., *et al.*, 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*, **44**(9):981–90.

Okbay, A., Beauchamp, J., Fontana, M., Lee, J., Pers, T., et al., 2016. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, **In press**.

O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C. F., Elliott, P., Jarvelin, M.-R., and Coin, L. J. M., 2012. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One*, **7**(5):e34861.

Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D. P., Patterson, N., and Price, A. L., *et al.*, 2014. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*, **30**(20):2906–14.

Perry, J. R. B., Day, F., Elks, C. E., Sulem, P., Thompson, D. J., Ferreira, T., He, C., Chasman, D. I., Esko, T., Thorleifsson, G., *et al.*, 2014. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, **514**(7520):92–7.

Pfister, R., Sharp, S., Luben, R., Welsh, P., Barroso, I., Salomaa, V., Meirhaeghe, A., Khaw, K.-T., Sattar, N., Langenberg, C., *et al.*, 2011. Mendelian randomization study of B-type natriuretic peptide and type 2 diabetes: evidence of causal association from population studies. *PLoS Med*, **8**(10):e1001112.

Pickrell, J. K., 2014. Joint Analysis of Functional Genomic Data and Genome-wide Association Studies of 18 Human Traits. *Am J Hum Genet*, **94**(4):559–73.

Psychiatric Genomics Consortium, 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**(7510):421–7.

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F. R., Barbalic, M., Gieger, C., *et al.*, 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet*, **43**(4):333–8.

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., *et al.*, 2015. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, **518**(7538):187–96.

Song, Y., Yeung, E., Liu, A., Vanderweele, T. J., Chen, L., Lu, C., Liu, C., Schisterman, E. F., Ning, Y., and Zhang, C., *et al.*, 2012. Pancreatic beta-cell function and type 2 diabetes risk: quantify the causal effect using a Mendelian randomization approach based on meta-analyses. *Hum Mol Genet*, **21**(22):5010–8.

Stephens, M., 2013. A unified framework for association analysis with multiple related phenotypes. *PLoS One*, **8**(7):e65245.

Su, Z., Marchini, J., and Donnelly, P., 2011. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*, **27**(16):2304–5.

Timpson, N. J., Nordestgaard, B. G., Harbord, R. M., Zacho, J., Frayling, T. M., Tybjærg-Hansen, A., and Smith, G. D., 2011. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond)*, **35**(2):300–8.

Villareal, D. T., Robertson, H., Bell, G. I., Patterson, B. W., Tran, H., Wice, B., and Polonsky, K. S., 2010. TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action. *Diabetes*, **59**(2):479–85.

Wakefield, J., 2008. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol*, **33**(1):79–86.

Wen, X. and Stephens, M., 2010. Using linear predictors to impute allele frequencies from summary or pooled genotype data. *Ann Appl Stat*, **4**(3):1158–1182.

Wen, X. and Stephens, M., 2014. Bayesian methods for genetic association analysis with heterogeneous subgroups: From meta-analyses to gene–environment interactions. *The Annals of Applied Statistics*, **8**(1):176–203.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z., *et al.*, 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*, **46**(11):1173–86.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., *et al.*, 2012. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*, **44**(4):369–75, S1–3.

Zhang, Q., Feitosa, M., and Borecki, I. B., 2014. Estimating and testing pleiotropy of single genetic variant for two quantitative traits. *Genet Epidemiol*, **38**(6):523–30.

Zhou, X. and Stephens, M., 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods*, **11**(4):407–9.