1 Supplementary Note

1.1 Quantitative Traits

Suppose we sample two cohorts with sample sizes N_1 and N_2. We measure phenotype 1 in cohort 1 and phenotype 2 in cohort 2. We model phenotype vectors for each cohort as $y_1 = Y\beta + \delta$, and $y_2 = Z\gamma + \epsilon$, where Y and Z are matrices of genotypes with columns standardized to mean zero and variance one[1], with dimensions $N_1 \times M$ and $N_2 \times M$, respectively; β and γ are vectors of per-standardized genotype effect sizes, and δ and ϵ are vectors of residuals, representing environmental effects and non-additive genetic effects. In this model, Y and Z are unobserved matrices of all SNPs, including SNPs that are not genotyped.

We treat all of $Y, Z, \beta, \gamma, \delta$ and ϵ as random. We model all of these as independent, except for β, γ, δ, ϵ. Suppose that (β, γ) has mean zero and covariance matrix[2]

$$\mathrm{Var}[(\beta,\gamma)] = \frac{1}{M}\begin{pmatrix} h_1^2 I & \rho_g I \\ \rho_g I & h_2^2 I \end{pmatrix},$$

and (δ, ϵ) has mean zero and covariance matrix

$$\mathrm{Var}[(\delta,\epsilon)] = \begin{pmatrix} (1-h_1^2) I & \rho_e I \\ \rho_e I & (1-h_2^2) I \end{pmatrix}.$$

Let $\rho := \rho_g + \rho_e$. Vectors of genotypes for each individual are drawn *i.i.d.* from a distribution with covariance matrix r (*i.e.,* r is an LD matrix with $r_{jk} = \mathbb{E}[Y_{ij}Y_{ik}]$). There are N_s individuals who are included in both studies.

Lemma 1. *Under this model, the expected genetic covariance (as defined in methods) between phenotypes is ρ_g, justifying our use of the notation ρ_g.*

Proof. Let X denote an $1 \times M$ vector of standardized genotypes for an arbitrary individual. Under the model, the additive genetic component of phenotype1 1 for this individual is $\sum_j X_j\beta_j$, and the additive genetic component of phenotype 1 for this individual is $\sum_j X_j\gamma_j$. Thus, the genetic

[1]We ignore the distinction between normalizing and centering in the population and in the sample, since this introduces only $\mathcal{O}(1/N)$ error.

[2]The assumption that all β is drawn with equal variance for all SNPs hides an implicit assumption that rare SNPs have larger per-allele effect sizes than common SNPs. As discussed in the simulations section of the main text and in our earlier work [1], LD Score regression is robust to moderate violations of this assumption, though it may break down in extreme cases, *e.g.,* if all causal variants are rare. In situations where a different model for $\mathrm{Var}[\beta]$ is more appropriate, all proofs in this note go through with LD Score replaced by weighted LD Scores, $\ell_j = \sum_k \mathrm{Var}[\beta_j] r_{jk}^2$.

covariance between phenotype 1 and phenotype 2 is

$$\begin{aligned}
\mathrm{Cov}\left[\sum_j X_j\beta_j, \sum_j X_j\gamma_j\right] &= \mathbb{E}\left[\left(\sum_j X_j\beta_j\right)\left(\sum_j X_j\gamma_j\right)\right] \\
&= \sum_j \sum_k \mathbb{E}[X_j X_k \beta_j \gamma_k] \\
&= \sum_j \mathbb{E}[X_j^2 \beta_j \gamma_j] \\
&= \sum_j \mathbb{E}[X_j^2]\mathbb{E}[\beta_j \gamma_j] \\
&= \rho_g.
\end{aligned}$$

□

We compute linear regression z-scores $z_{1j} \coloneqq Y_j^\mathsf{T} y_1/\sqrt{N_1}$ and $z_{2j} \coloneqq Y_j^\mathsf{T} y_2/\sqrt{N_2}$ for genotyped SNPs j (where Y_j and Z_j denote the j^{th} columns of Y and Z).

Definition 1. *The LD Score of a variant j is $\ell_j \coloneqq \sum_k r_{jk}^2$, where the sum is taken over all other variants k.*

Proposition 1. *Let j denote a genotyped SNP. Under the model described above,*

$$\mathbb{E}[z_{1j} z_{2j}] = \frac{\sqrt{N_1 N_2}\rho_g}{M}\ell_j + \frac{N_s \rho}{\sqrt{N_1 N_2}}. \tag{1}$$

Proof. By the law of total expectation,

$$\mathbb{E}[z_{1j} z_{2j}] = \mathbb{E}[\mathbb{E}[z_{1j} z_{2j} \,|\, Y, Z]] \tag{2}$$

First we compute the inner expectation from Equation 2, with Z and Y fixed.

$$\begin{aligned}
\mathbb{E}[z_{1j} z_{2j} \,|\, Y, Z] &= \frac{1}{\sqrt{N_1 N_2}}\mathbb{E}[Y_j^\mathsf{T} y_1 y_2^\mathsf{T} Z_j] \\
&= \frac{1}{\sqrt{N_1 N_2}} Y_j^\mathsf{T} \mathbb{E}[(Y\beta + \delta)(Z\gamma + \epsilon)^\mathsf{T}] Z_j \\
&= \frac{1}{\sqrt{N_1 N_2}} Y_j^\mathsf{T} \left(Y\mathbb{E}[\beta^\mathsf{T}\gamma] Z + \mathbb{E}[\delta^\mathsf{T} Z\gamma] + \mathbb{E}[\beta^\mathsf{T} Y^\mathsf{T}\epsilon] + \mathbb{E}[\delta^\mathsf{T}\epsilon]\right) Z_j \\
&= \frac{1}{\sqrt{N_1 N_2}} Y_j^\mathsf{T} \left(Y\mathbb{E}[\beta^\mathsf{T}\gamma] Z + \mathbb{E}[\delta^\mathsf{T}\epsilon]\right) Z_j \\
&= \frac{1}{\sqrt{N_1 N_2}} \left(\frac{\rho_g}{M} Y_j^\mathsf{T} Y Z_j^\mathsf{T} Z + \rho_e Y_j^\mathsf{T} Z_j\right).
\end{aligned} \tag{3}$$

Next, we remove the conditioning on Y and Z.

$$\frac{1}{\sqrt{N_1 N_2}}\mathbb{E}[Y_j^\mathsf{T} Z_j] = \frac{N_s}{\sqrt{N_1 N_2}}, \tag{4}$$

and

$$\frac{1}{\sqrt{N_1 N_2}}\mathbb{E}[Y_j^\mathsf{T} Y Z_j^\mathsf{T} Z] = \ell_j + \frac{M N_s}{\sqrt{N_1 N_2}}. \tag{5}$$

Substituting equations 4 and 5 into Equation 3,

$$\begin{aligned}\mathbb{E}[z_{1j} z_{2j}] &= \frac{\sqrt{N_1 N_2}\rho_g}{M}\ell_j + \frac{N_s\left(\rho_g + \rho_e\right)}{\sqrt{N_1 N_2}} \\ &= \frac{\sqrt{N_1 N_2}\rho_g}{M}\ell_j + \frac{N_s \rho}{\sqrt{N_1 N_2}}. \end{aligned} \tag{6}$$

□

If study 1 and study 2 are the same study, then $N_1 = N_2 = N_s$, $\rho_g = h_g^2$ and $\rho = 1$, so Equation 6 reduces to the LD Score regression equation for a single trait from [1].

1.2 Regression Weights

We can improve the efficiency of LD Score regression by weighting by the reciprocal of the conditional variance function (CVF), $\mathrm{Var}[z_{1j} z_{2j} \,|\, \ell_j]$. The CVF is not uniquely determined by the assumptions about the first and second moments of β and γ used to derive Proposition 1. Therefore we derive the CVF for the case where z_{1j} and z_{2j} are jointly distributed as bivariate normal[3]. From a standard formula for double second moments of the bivariate normal, the CVF is

$$\begin{aligned}\mathrm{Var}[z_{1j} z_{2j} \,|\, \ell_j] &= \mathrm{Var}[z_{1j}]\mathrm{Var}[z_{2j}] + \mathbb{E}[z_{1j} z_{2j}]^2 \\ &= \left(\frac{N_1 h_1^2 \ell_j}{M} + 1\right)\left(\frac{N_2 h_2^2 \ell_j}{M} + 1\right) + \left(\frac{\sqrt{N_1 N_2}\rho_g}{M}\ell_j + \frac{\rho N_s}{\sqrt{N_1 N_2}}\right)^2 \end{aligned} \tag{7}$$

The terms on the left follow from the fact that $\mathrm{Var}[z_{ij}] = \chi_{ij}^2$ and $\mathbb{E}[\chi^2] = N h^2 \ell_j / M + 1$. The term on the right follows from Proposition 1. Note that if $z_1 = z_2$, this reduces to the expression for the CVF of χ^2 statistics from [1] (though there is an error in Equation 3.2 of the supplementary note of [1]; the right side is missing a factor of 2. We thank Peter Visscher for pointing this out).

In cases where the normality assumption does not hold, LD Score regression will remain unbiased, but may be inefficient, because the regression weights will be suboptimal. We also apply a heuristic weighting scheme to avoid overcounting SNPs in high-LD regions, described in the methods.

1.3 Liability Threshold Model

In the liability threshold (probit) model [2], binary traits are determined by an unobserved continuous liability ψ. The observed trait is $y \coloneqq \mathbf{1}[\psi > \tau]$, where τ is the liability threshold. If ψ is normally distributed, then setting $\tau \coloneqq \Phi^{-1}(1 - K)$ (where Φ is the standard normal cdf) yields a population prevalence of K.

For phenotypes generated according to the liability threshold model, we can estimate not only the heritability and genetic covariance of the observed phenotype, but also the heritability and genetic covariance of the unobserved liability.

[3]For instance, it is sufficient but not necessary to assume that β, γ, δ and ϵ are multivariate normal. More generally, the z-scores will be approximately normal if β and γ are reasonably polygenic. If the distribution of effect sizes is heavy-tailed, *e.g.,* if there are few causal SNPs, then the CVF may be larger.

In the next lemma, we derive population case and control allele frequencies in terms of the heritability of liability when liability is generated following the model for quantitative traits from section 1.1. Since we are only modeling additive effects and are willing to assume Hardy-Weinberg equilibrium, we lose no generality and simplify notation considerably by stating the proofs in terms of haploid genotypes.

We state this lemma in terms of marginal per-allele effect sizes, instead of the per-standardized-genotype effect sizes considered in section 1.1. Here marginal means that these are the effect sizes obtained by univariate regression of phenotype against genotype in the infinite data limit. Haploid standardized genotypes are defined $X_{ij} \coloneqq (G_{ij} - p_j)/\sqrt{p_j(1-p_j)}$, where G_{ij} is the 0-1 coded genotype. If β_j is the marginal per-standardized-genotype effect and ζ_j is the marginal per-allele effect, we have $X_j\beta_j = G_j\zeta_j$. Thus, setting $G_{ij} = 1$ yields $\zeta_j = \beta_j\sqrt{(1-p_j)/p_j}$.

Lemma 2. *Suppose unobserved liabilities ψ, φ for traits y_1, y_2 with thresholds τ_1, τ_2 corresponding to prevalences K_1, K_2 are generated according to the mode for quantitative traits from section 1.1, i.e., $\psi_i = \sum_j X_{ij}\beta_j + \delta$, $\varphi_i = \sum_j X_{ij}\gamma_j + \epsilon$, with*

$$\mathrm{Var}[(\beta, \gamma)] = \frac{1}{M}\begin{pmatrix} h_1^2 I & \rho_g I \\ \rho_g I & h_2^2 I \end{pmatrix},$$

and

$$\mathrm{Var}[(\delta, \epsilon)] = \begin{pmatrix} (1-h_1^2)I & \rho_e I \\ \rho_e I & (1-h_2^2)I \end{pmatrix}.$$

Let ζ_j and ξ_j denote the marginal per-allele effect sizes of SNP j on ψ and φ. Let

$$p_{cas,kj} \coloneqq \mathbb{P}[G_{ij} = 1 \,|\, y_{ik} = 1]$$
$$p_{con,kj} \coloneqq \mathbb{P}[G_{ij} = 1 \,|\, y_{ik} = 0]$$

denote the allele frequencies of SNP j in cases and controls for phenotype k, where y_{ik} denotes the value of phenotype k for individual i and $k = 1, 2$. Then

$$\begin{aligned}
\mathbb{E}[p_{cas,1j} - p_{con,1j}] &= 0, \\
\mathbb{E}[p_{cas,2j} - p_{con,2j}] &= 0, \\
\mathrm{Var}[p_{cas,1j} - p_{con,1j}] &= \frac{p_j(1-p_j)\phi(\tau_1)^2 h_1^2}{MK_1^2(1-K_1)^2}\ell_j, \\
\mathrm{Var}[p_{cas,2j} - p_{con,2j}] &= \frac{p_j(1-p_j)\phi(\tau_2)^2 h_2^2}{MK_2^2(1-K_2)^2}\ell_j, \\
\mathrm{Cov}[p_{cas,1j} - p_{con,1j}, p_{cas,2j} - p_{con,2j}] &= \frac{p_j(1-p_j)\phi(\tau_1)\phi(\tau_2)\rho_g}{MK_1(1-K_1)K_2(1-K_2)}\ell_j,
\end{aligned}$$

where the expectation is taken over where ϕ is the standard normal density. These results apply to population allele frequencies, not allele frequencies in a finite sample. We deal with ascertained finite samples in the next section.

Proof. This proof is accomplished in two steps. First, we compute allele frequencies conditional on the marginal effects on liability. To do this, we reverse the conditional probability using Bayes' theorem, which reduces the problem to a series of [Taylor approximations to] Gaussian integrals.

Second, we remove the conditioning on the marginal effects on liability in order to express the allele frequencies in terms of h_1^2, h_2^2, ρ_g and ℓ_j. Since liability is just a quantitative trait, we need only apply the LD Score regression equation for quantitative traits.

By Bayes' rule,

$$\begin{aligned}\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 1, \zeta_j] &= \frac{\mathbb{P}[y_{i1} = 1 \,|\, G_{ij} = 1, \zeta_j]\mathbb{P}[G_{ij} = 1]}{\mathbb{P}[y_{i1} = 1]} \\ &= \frac{p_j}{K_1}\mathbb{P}[y_{i1} = 1 \,|\, G_{ij} = 1, \zeta_j] \\ &= \frac{p_j}{K_1}\mathbb{P}[\psi_i > \tau_1 \,|\, G_{ij} = 1, \zeta_j]. \end{aligned} \tag{8}$$

The distribution of ψ given G_{ij} and ζ_j is $\psi \,|\, (G_{ij} = 1, \zeta_j) \sim N(\zeta_j, 1 - \zeta_j^2) \approx N(\zeta_j, 1)$ (where the approximation that the variance equals one holds when the marginal heritability explained by j is small, which is the typical case in GWAS). Thus $\mathbb{P}[\psi_i > \tau_1 \,|\, G_{ij} = 1]$ is simply a Gaussian integral. We approximate this probability with a first-order Taylor expansion around τ_1.

$$\begin{aligned}\mathbb{P}[\psi_i > \tau_1 \,|\, G_{ij} = 1, \zeta_j] &= 1 - \Phi(\tau_1 - \zeta_j) \\ &\approx K_1 + \phi(\tau_1)\zeta_j, \end{aligned} \tag{9}$$

Substituting Equation 9 into Equation 8,

$$\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 1, \zeta_j] = \frac{p_j}{K}\left(K_1 + \phi(\tau_1)\zeta_j\right). \tag{10}$$

A similar argument shows that

$$\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 0, \zeta_j] = \frac{p_j}{1 - K_1}\left(1 - K_1 - \phi(\tau_1)\zeta_j\right). \tag{11}$$

Subtracting Equation 11 from Equation 10,

$$\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 1, \zeta_j] - \mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 0, \zeta_j] = p_j \frac{\phi(\tau_1)\zeta_j}{K_1(1 - K_1)}. \tag{12}$$

Similar results hold for trait 2, replacing ζ with ξ and subscript 1 with subscript 2.

We have written the probabilities in question in terms of constants and marginal effects on liability. Since liability is simply a quantitative trait, the means, variances, and covariances of the marginal effects on liability are described by the LD Score regression equation for quantitative traits from Proposition 1. Precisely, $\mathbb{E}[\xi_j] = \mathbb{E}[\zeta_j] = 0$, $\mathrm{Var}[\xi_j] = (1 - p_j)h_1^2\ell_j/p_jM$, $\mathrm{Var}[\zeta_j] = (1 - p_j)h_2^2\ell_j/p_jM$ and $\mathrm{Cov}[\zeta_j, \xi_j] = (1 - p_j)\rho_g\ell_j/p_jM$. If we combine these results with Equation 12, we find that

$$\mathbb{E}[p_{cas,1j} - p_{con,1j}] = 0; \tag{13}$$

$$\begin{aligned}\mathrm{Var}[p_{cas,1j} - p_{con,1j}] &= \mathrm{Var}\left[\frac{p_j\phi(\tau_1)\zeta_j}{K_1(1 - K_1)}\right] \\ &= \frac{p_j(1 - p_j)\phi(\tau_1)^2 h_1^2}{MK_1^2(1 - K_1)^2}\ell_j \end{aligned} \tag{14}$$

(similarly for trait two), and

$$\begin{aligned}\mathrm{Cov}[p_{cas,1j} - p_{con,1j}, p_{cas,2j} - p_{con,2j}] &= \mathrm{Cov}\left[\frac{p_j\phi(\tau_1)\zeta_j}{K_1(1-K_1)}, \frac{p_j\phi(\tau_2)\xi_j}{K_2(1-K_2)}\right] \\ &= \frac{p_j(1-p_j)\phi(\tau_1)\phi(\tau_2)\rho_g}{MK_1(1-K_1)K_2(1-K_2)}\ell_j. \end{aligned} \tag{15}$$

□

1.4 Ascertained Studies of Liability Threshold Traits

In the next proposition, we derive an LD Score regression equation for ascertained case/control studies.

Let P_i denote the sample prevalence of y_i in study i for $i = 1, 2$. We compute z-scores

$$z_j \coloneqq \frac{\sqrt{NP(1-P)}(\hat{p}_{cas} - \hat{p}_{con})}{\sqrt{\hat{p}_j(1-\hat{p}_j)}},$$

where $\hat{p}_j$ denotes allele frequency in the entire sample[4], $\hat{p}_{cas}$ denotes sample case allele frequency and $\hat{p}_{con}$ denotes sample control allele frequency.

We emphasize one subtlety before stating the main proposition. The results in this section allow for study k to select samples based on phenotype l only if $k = l$. If study 1 ascertains on phenotype 2 – for example, if all cases i in study 1 have $y_{i1} = y_{i2} = 1$ — then $\hat{p}_{cas,1j}$ will not be an unbiased estimate of $p_{cas,1j}$. Indeed, in this example, $\mathbb{E}[\hat{p}_{cas,1j}] = \mathbb{P}[G_{ij} = 1 \,|\, y_1 = y_2 = 1]$, which will not equal $p_{cas,1j} = \mathbb{P}[G_{ij} = 1 \,|\, y_1 = 1]$ unless $\rho = 1$ or $\rho = 0$. This follows from the fact that the conditionals and marginals of a bivariate normal are equal iff $\rho = 0$ or $\rho = 1$. We do not derive formulae describing the bias, except to note that the most common scenario, the "healthy controls" model — cases are sampled independently but all controls are controls for both traits — is probably nothing to worry about, so long as cases for both traits are uncommon. In this scenario, $\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 0] \approx \mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = y_{i2} = 0]$. Conditioning on $y_{i2} = 0$ hardly changes the distribution, because $y_{i2} = 0$ most of the time, anyway. In addition, excluding double cases from the analysis (as a conservative defense against spurious comorbidity) is also likely to be safe for pairs of uncommon traits with small excess comorbidity. In this case, $\mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 1] \approx \mathbb{P}[G_{ij} = 1 \,|\, y_{i1} = 1, y_{i2} = 0]$, so long as y_2 is uncommon and not too highly correlated with y_1.

Proposition 2. *Under the liability threshold model from lemma 1.3,*

$$\mathbb{E}[z_{1j}z_{2j}] \approx \frac{\sqrt{N_1N_2}\rho_{g,obs}}{M}\ell_j + \sqrt{N_1N_2P_1(1-P_1)P_2(1-P_2)}\left(\sum_{a,b\in\{cas,con\}} \frac{N_{a,b}(-1)^{1+\mathbf{1}[a=b]}}{N_{a,1}N_{b,2}}\right) \tag{16}$$

where

$$\rho_{g,obs} \coloneqq \rho_g\left(\frac{\sqrt{\phi(\tau_1)\phi(\tau_2)P_1(1-P_1)P_2(1-P_2)}}{K_1(1-K_1)K_2(1-K_2)}\right)$$

denotes observed scale genetic covariance, $N_{a,b}$ denotes the number of individuals with phenotype a in study 1 and b in study two for $a, b \in \{cas, con\}$ (e.g., $N_{cas,con}$ is the number of individuals who

[4]Conditional on the marginal effect of j, the expected value of $\hat{p}_j$ is not equal to p_j unless $P = K$ or the marginal effect of j is zero.

are a case in study 1 but a control in study 2), N_i denotes total sample size in study i and $N_{a,i}$ for $a \in \{cas, con\}$ and $i = 1, 2$ denotes the number of individuals with phenotype a in study i.

Observe that $\rho_{g,obs}/\sqrt{h^2_{1,obs}h^2_{2,obs}} = \rho_g/\sqrt{h^2_1 h^2_2} = r_g$. Put another way, the natural definition for "observed scale genetic correlation" turns out to be the same as regular genetic correlation, because the scale transformation factors in the numerator and denominator cancel. This is convenient: we can compute genetic correlations for binary traits on a sensible scale without having to worry about sample and population prevalences.

Proof. The full form of $z_{1j}z_{2j}$ is

$$z_{1j}z_{2j} = \frac{\sqrt{cN_1N_2}(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j})}{\sqrt{\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j})}},$$

where $c \coloneqq P_1(1-P_1)P_2(1-P_2)$. Our strategy for obtaining the expectation is

$$\mathbb{E}[z_{1j}z_{2j}] \approx \sqrt{cN_1N_2}\frac{\mathbb{E}[(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j})]}{\mathbb{E}[\sqrt{\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j})}]} \tag{17}$$

$$\approx \sqrt{cN_1N_2}\frac{\mathbb{E}[(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j})]}{\sqrt{\mathbb{E}[\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j})]}} \tag{18}$$

$$= \sqrt{cN_1N_2}\frac{\mathbb{E}\left[\mathbb{E}[(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j}) \mid \zeta_j, \xi_j]\right]}{\sqrt{\mathbb{E}\left[\mathbb{E}[\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j}) \mid \zeta_j, \xi_j]\right]}}, \tag{19}$$

where ζ_j and ξ_j denote the marginal per-allele effects of j. Approximation 17 hides $\mathcal{O}(1/N)$ error from moving from the expectation of a ratio to a ratio of expectations. Approximation 18 hides $\mathcal{O}(1/N)$ error from moving from the expectation of a square root to a square root of expectations, and dear reader we admire your perseverance in making it this far. Equality 19 follows from applying of the law of total expectation to the numerator and denominator.

First, we compute the numerator. By linearity of expectation,

$$\begin{aligned}\mathbb{E}[(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j})] \mid \zeta_j, \xi_j] = \mathbb{E}[\hat{p}_{cas,1j}\hat{p}_{cas,2j} \mid \zeta_j, \xi_j] - \mathbb{E}[\hat{p}_{cas,1j}\hat{p}_{con,2j} \mid \zeta_j, \xi_j] \\ - \mathbb{E}[\hat{p}_{con,1j}\hat{p}_{cas,2j} \mid \zeta_j, \xi_j] + \mathbb{E}[\hat{p}_{con,1j}\hat{p}_{con,2j} \mid \zeta_j, \xi_j]\end{aligned} \tag{20}$$

After conditioning on the marginal effects ζ_j and ξ_j, the only source of variance in the sample allele frequencies $\hat{p}_{cas,1}, \hat{p}_{con,1}, \hat{p}_{cas,2}, \hat{p}_{con,2}$ is sampling error. Write $\hat{p}_{cas,1j}\hat{p}_{cas,2j} = (p_{cas,1j}+\eta)(p_{cas,2j}+\nu)$, where η and ν denote sampling error. If study 1 and study 2 share samples, ν and η will be correlated:

$$\begin{aligned}\mathbb{E}[\hat{p}_{cas,1j}\hat{p}_{cas,2j} \mid \zeta_j, \xi_j] &= p_{cas,1j}p_{cas,2j} + \mathbb{E}[\eta\nu] \\ &\approx p_{cas,1j}p_{cas,2j} + \frac{N_{cas,cas}\sqrt{p_{cas,1j}(1-p_{cas,1j})p_{cas,2j}(1-p_{cas,2j})}}{N_{cas,1}N_{cas,2}} \qquad (21) \\ &\approx p_{cas,1j}p_{cas,2j}\left(1 + \frac{N_{cas,cas}}{N_{cas,1}N_{cas,2}}\right), \qquad (22)\end{aligned}$$

where approximation 21 is the (bivariate) central limit theorem, and approximation 22 comes from ignoring the difference between $\sqrt{p_{cas,1j}(1-p_{cas,1j})p_{cas,2j}(1-p_{cas,2j})}$ and $p_j(1-p_j)$. This step is justified in the derivation of the denominator. Similar relationships hold for the other terms in Equation 20.

If we combine equations 22 and 15, we obtain

$$\mathbb{E}[(\hat{p}_{cas,1j} - \hat{p}_{con,1j})(\hat{p}_{cas,2j} - \hat{p}_{con,2j})] \approx p_j(1-p_j)\left(\frac{\phi(\tau_1)\phi(\tau_2)\rho_g}{c'M}\ell_j + \sum_{a,b\in\{cas,con\}} \frac{N_{a,b}(-1)^{1+\mathbf{1}[\mathbf{a}=\mathbf{b}]}}{N_{a,1}N_{b,2}}\right), \tag{23}$$

where $c' \coloneqq K_1(1-K_1)K_2(1-K_2)$.

Next, we derive the expectation of the denominator. Conditional on ζ_j and ξ_j, $\hat{p}_{1j}(1-\hat{p}_{1j})$ is $P_1 p_{cas,1j} + (1-P_1)p_{con,1j}$ plus $\mathcal{O}(1/N)$ sampling variance. If studies 1 and 2 share samples, the $\mathcal{O}(1/N)$ sampling variance in $\hat{p}_{1j}(1-\hat{p}_{1j})$ and $\hat{p}_{2j}(1-\hat{p}_{2j})$ will be correlated, but this still only amounts to $\mathcal{O}(N_s/N_1N_2)$ error. If we remove the conditioning on ζ_j and ξ_j, then $P_1 p_{cas,1j} + (1-P_1)p_{con,1j}$ is equal to $p_j(1-p_j)$ plus $\mathcal{O}(h^2_{1,obs}\ell_j/M)$ error from uncertainty in ζ_j. The covariance between uncertainty in ζ_j and uncertainty in ξ_j is driven by $\rho_{g,obs}$, so the expectation of the denominator is $\mathbb{E}\left[\sqrt{\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j})}\right] = p_j(1-p_j)\left(1 + \mathcal{O}(N_s/N_1N_2) + \mathcal{O}(\rho_{g,obs}\ell_j/M)\right)$. We make the approximation[5] that

$$\mathbb{E}\left[\sqrt{\hat{p}_{1j}(1-\hat{p}_{1j})\hat{p}_{2j}(1-\hat{p}_{2j})}\right] \approx p_j(1-p_j). \tag{24}$$

We obtain the desired result by dividing $\sqrt{cN_1N_2}$ times Equation 23 by Equation 24. □

Corollary 1. *If study 1 is an ascertained study of a binary trait, and study 2 is a non-ascertained quantitative study, then proposition 2 holds, except with genetic covariance on the half-observed scale*

$$\rho_{g,obs} \coloneqq \rho_g\left(\frac{\sqrt{\phi(\tau_1)P_1(1-P_1)}}{K_1(1-K_1)}\right).$$

Corollary 2. *For a single binary trait,*

$$\mathbb{E}[\chi^2_j] = \frac{Nh^2_{obs}}{M}\ell_j + 1, \tag{25}$$

where $h^2_{obs} = h^2\phi(\tau)^2P(1-P)/K^2(1-K)^2$.

Proof. This follows from proposition 2 if we set study 1 equal to study 2 and note that the observed scale genetic covariance between a trait and itself is observed scale heritability. To show that the intercept is one, observe that if study 1 and study 2 are the same, then

$$\begin{aligned}\sqrt{cN_1N_2}\left(\sum_{a,b\in\{cas,con\}}\frac{N_{a,b}(-1)^{1+\mathbf{1}[\mathbf{a}=\mathbf{b}]}}{N_{a,1}N_{b,2}}\right) &= NP(1-P)\left(\frac{1}{N_{cas}} + \frac{1}{N_{con}}\right)\\ &= \frac{NP(1-P)(N_{cas}+N_{con}))}{N_{cas}N_{con}}\\ &= \frac{N^2P(1-P)}{N_{cas}N_{con}}.\end{aligned} \tag{26}$$

But $NP = N_{cas}$ and $N(1-P) = N_{con}$, so Equation 26 simplifies to 1. □

[5]For $\ell_j = 100$ (roughly the median 1kG LD Score), $M = 10^7$ and $\rho_{g,obs} = 1$, we get $\rho_{g,obs}\ell_j/M = 10^{-5}$. A worst-case value for N_s/N_1N_2 might be $N_s = N_1 = N_2 = 10^3$, in which case $N_s/N_1N_2 = 10^{-3}$. Thus, $\rho_{g,obs}\ell_j/M$ and N_s/N_1N_2 will generally be at least 3 orders of magnitude smaller than 1.

1.5 Flavors of Heritability and Genetic Correlation

The heritability parameter estimated by `ldsc` is subtly different than the heritability parameter h_g^2 estimated by `GCTA`. If g denotes the set of all genotyped SNPs in some GWAS, define $\beta_{\texttt{GCTA}} \coloneqq \text{argmax}_{\alpha \in \mathbb{R}^{|g|}} \text{Cor}\left[y_1, X_g \alpha\right]$, where X_g is a random vector of standardized genotypes for SNPs in g. Then the heritability parameter estimated by `GCTA` is defined

$$h_g^2 \coloneqq \sum_{j \in g} \beta_{\texttt{GCTA},j}^2.$$

Let S denote the set of SNPs used to compute LD Scores (*i.e.*, $\ell_j = \sum_{k \in S} r_{jk}^2$), and let $\beta_S \coloneqq \text{argmax}_{\alpha \in \mathbb{R}^{|S|}} \text{Cor}\left[y_1, X_S \alpha\right]$. Generally $\beta_{S,j} \neq \beta_{\texttt{GCTA},j}$ unless all SNPs in $S \setminus g$ are not in LD with SNPs in g. Define

$$h_S^2 \coloneqq \sum_{j \in S} \beta_{S,j}^2.$$

Let S' denote the set of SNPs in S with MAF above 5%. Define

$$h_{\text{5-50\%}}^2 \coloneqq \sum_{j \in S'} \beta_{S,j}^2. \tag{27}$$

The default setting in `ldsc` is to report $h_{\text{5-50\%}}^2$, estimated as the slope from LD Score regression times $M_{\text{5-50\%}}$, the number of SNPs with MAF above 5%.

The reason for this is the following: suppose that h^2 per SNP is not constant as a function of MAF. Then the slope of LD Score regression will represent some sort of weighted average of the values of h^2 per SNP, with more weight given to classes of SNPs that are well-represented among the regression SNPs. In a typical GWAS setting, the regression SNPs are mostly common SNPs, so multiplying the slope from LD Score regression by M (which includes rare SNPs) amounts to extrapolating that h^2 per SNP among common variants is the same as h^2 per SNP among rare variants. This extrapolation is particularly risky, because there are many more rare SNPs than common SNPs.

It is probably reasonable to treat h^2 per SNP as a constant function of MAF for SNPs with MAF above 5%, but we have very little information about h^2 per SNP for SNPs with MAF below 5%. Therefore we report $h_{\text{5-50\%}}^2$ instead of h_S^2 to avoid excessive extrapolation error. This lower bound can be pushed lower with larger sample sizes and better rare variant coverage, either from sequencing or imputation.

There are two main distinctions between $h_{\text{5-50\%}}^2$ and h_g^2. First, h_g^2 does not include the effects of common SNPs that are not tagged by the set of genotyped SNPs g. Second, the effects of causal 4% SNPs are not counted towards $h_{\text{5-50\%}}^2$. In practice, neither of these distinctions makes a large difference, since most GWAS arrays focus on common variation and manage to assay or tag almost all common variants, which is why we do not emphasize this distinction in the main text.

The relationship between the genetic covariance parameter estimated by LD Score regression and the genetic covariance parameter estimated by `GCTA` is similar to the relationship between $h_{\text{5-50\%}}^2$ and h_g^2. Choice of M is not important for genetic correlation, because the factors of M in the numerator and denominator cancel.

2 References

[1] Brendan Bulik-Sullivan, Po-Ru Loh, Hilary Finucane, Stephan Ripke, Jian Yang, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 2015.

[2] Karl Pearson and Alice Lee. On the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society of London, A (195)*, pages 79–150, 1901.

[3] Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*.

[4] Pamela Sklar, Stephan Ripke, Laura J Scott, Ole A Andreassen, Sven Cichon, Nick Craddock, Howard J Edenberg, John I Nurnberger, Marcella Rietschel, Douglas Blackwood, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near odz4. *Nature genetics*.

[5] Stephan Ripke, Naomi R Wray, Cathryn M Lewis, Steven P Hamilton, Myrna M Weissman, Gerome Breen, Enda M Byrne, Douglas HR Blackwood, Dorret I Boomsma, Sven Cichon, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry*.

[6] Vesna Boraska, Christopher S Franklin, James AB Floyd, Laura M Thornton, Laura M Huckins, Lorraine Southam, N William Rayner, Ioanna Tachmazidou, Kelly L Klump, Janet Treasure, et al. A genome-wide association study of anorexia nervosa. *Molecular psychiatry*, 2014.

[7] Cross-Disorder Group of the Psychiatric Genomics Consortium et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*.

[8] Tobacco, Genetics Consortium, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics*.

[9] Jean-Charles Lambert, Carla A Ibrahim-Verbaas, Denise Harold, Adam C Naj, Rebecca Sims, Céline Bellenguez, Gyungah Jun, Anita L DeStefano, Joshua C Bis, Gary W Beecham, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for alzheimer's disease. *Nature genetics*, 2013.

[10] Cornelius A Rietveld, Sarah E Medland, Jaime Derringer, Jian Yang, Tõnu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal, et al. Gwas of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*.

[11] Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, 2014.

[12] Adam E Locke, Bratati Kahali, Sonja I Berndt, Anne E Justice, Tune H Pers, Felix R Day, Corey Powell, Sailaja Vedantam, Martin L Buchkovich, Jian Yang, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*.

[13] Heribert Schunkert, Inke R König, Sekar Kathiresan, Muredach P Reilly, Themistocles L Assimes, Hilma Holm, Michael Preuss, Alexandre FR Stewart, Maja Barbalic, Christian Gieger, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*.

[14] Tanya M Teslovich, Kiran Musunuru, Albert V Smith, Andrew C Edmondson, Ioannis M Stylianou, Masahiro Koseki, James P Pirruccello, Samuli Ripatti, Daniel I Chasman, Cristen J Willer, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*.

[15] Andrew P Morris, Benjamin F Voight, Tanya M Teslovich, Teresa Ferreira, Ayellet V Segre, Valgerdur Steinthorsdottir, Rona J Strawbridge, Hassan Khan, Harald Grallert, Anubha Mahajan, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*.

[16] Alisa K Manning, Marie-France Hivert, Robert A Scott, Jonna L Grimsby, Nabila Bouatia-Naji, Han Chen, Denis Rybin, Ching-Ti Liu, Lawrence F Bielak, Inga Prokopenko, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature genetics*.

[17] Early Growth Genetics (EGG) Consortium et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nature genetics*.

[18] Ralf JP van der Valk, Eskil Kreiner-Møller, Marjolein N Kooijman, Mònica Guxens, Evangelia Stergiakouli, Annika Sääf, Jonathan P Bradfield, Frank Geller, M Geoffrey Hayes, Diana L Cousminer, et al. A novel common variant in dcst2 is associated with length in early life and height in adulthood. *Human molecular genetics*, page ddu510, 2014.

[19] Momoko Horikoshi, Hanieh Yaghootkar, Dennis O Mook-Kanamori, Ulla Sovio, H Rob Taal, Branwen J Hennig, Jonathan P Bradfield, Beate St Pourcain, David M Evans, Pimphen Charoen, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature genetics*.

[20] H Rob Taal, Beate St Pourcain, Elisabeth Thiering, Shikta Das, Dennis O Mook-Kanamori, Nicole M Warrington, Marika Kaakinen, Eskil Kreiner-Møller, Jonathan P Bradfield, Rachel M Freathy, et al. Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nature genetics*.

[21] John RB Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tõnu Esko, Gudmar Thorleifsson, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*.

[22] Luke Jostins, Stephan Ripke, Rinse K Weersma, Richard H Duerr, Dermot P McGovern, Ken Y Hui, James C Lee, L Philip Schumm, Yashoda Sharma, Carl A Anderson, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*.

[23] Eli A Stahl, Soumya Raychaudhuri, Elaine F Remmers, Gang Xie, Stephen Eyre, Brian P Thomson, Yonghong Li, Fina AS Kurreeman, Alexandra Zhernakova, Anne Hinks, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature genetics*.

[24] Beben Benyamin, BSt Pourcain, Oliver S Davis, Gail Davies, Narelle K Hansell, M-JA Brion, RM Kirkpatrick, RAM Cents, S Franić, MB Miller, et al. Childhood intelligence is heritable, highly polygenic and associated with fnbp1l. *Molecular Psychiatry*.

[25] Elizabeth K Speliotes, Cristen J Willer, Sonja I Berndt, Keri L Monda, Gudmar Thorleifsson, Anne U Jackson, Hana Lango Allen, Cecilia M Lindgren, Jian'an Luan, Reedik Mägi, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*.

[26] Hana Lango Allen, Karol Estrada, Guillaume Lettre, Sonja I Berndt, Michael N Weedon, Fernando Rivadeneira, Cristen J Willer, Anne U Jackson, Sailaja Vedantam, Soumya Raychaudhuri, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*.

[27] Shashaank Vattikuti, Juen Guo, and Carson C Chow. Heritability and genetic correlations explained by common snps for metabolic syndrome traits. *PLoS genetics*.

[28] Ole A Andreassen, Srdjan Djurovic, Wesley K Thompson, Andrew J Schork, Kenneth S Kendler, Michael C O?Donovan, Dan Rujescu, Thomas Werge, Martijn van de Bunt, Andrew P Morris, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics*.

[29] Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium et al. Genome-wide association study identifies five new schizophrenia loci. *Nature genetics*.

gencor

Collaborators

Psychiatric Genomics Consiortium

Collaborators from the Psychiatric Genomics Consortium were, in alphabetical order: Devin Absher, Rolf Adolfsson, Ingrid Agartz, Esben Agerbo, Huda Akil, Margot Albus, Madeline Alexander, Farooq Amin, Ole A Andreassen, Adebayo Anjorin, Richard Anney, Dan Arking, Philip Asherson, Maria H Azevedo, Silviu A Bacanu, Lena Backlund, Judith A Badner, Tobias Banaschewski, Jack D Barchas, Michael R Barnes, Thomas B Barrett, Nicholas Bass, Michael Bauer, Monica Bayes, Martin Begemann, Frank Bellivier, Judit Bene, Sarah E Bergen, Thomas Bettecken, Elizabeth Bevilacqua, Joseph Biederman, Tim B Bigdeli, Elisabeth B Binder, Donald W Black, Douglas HR Blackwood, Cinnamon S Bloss, Michael Boehnke, Dorret I Boomsma, Anders D Borglum, Elvira Bramon, Gerome Breen, Rene Breuer, Richard Bruggeman, Nancy G Buccola, Randy L Buckner, Jan K Buitelaar, Brendan Bulik-Sullivan, William E Bunner, Margit Burmeister, Joseph D Buxbaum, William F Byerley, Sian Caesar, Wiepke Cahn, Guiqing Cai, Murray J Cairns, Dominique Campion, Rita M Cantor, Vaughan J Carr, Noa Carrera, Miquel Casas, Stanley V Catts, Aravinda Chakravarti, Kimberley D Chambert, Raymond CK Chan, Eric YH Chen, Ronald YL Chen, Wei Cheng, Eric FC Cheung, Siow Ann Chong, Khalid Choudhury, Sven Cichon, David St Clair, C Robert Cloninger, David Cohen, Nadine Cohen, David A Collier, Edwin Cook, Hilary Coon, Bru Cormand, Paul Cormican, Aiden Corvin, William H Coryell, Nicholas Craddock, David W Craig, Ian W Craig, Benedicto Crespo-Facorro, James J Crowley, David Curtis, Darina Czamara, Mark J Daly, Ariel Darvasi, Susmita Datta, Michael Davidson, Kenneth L Davis, Richard Day, Franziska Degenhardt, Lynn E DeLisi, Ditte Demontis, Bernie Devlin, Dimitris Dikeos, Timothy Dinan, Srdjan Djurovic, Enrico Domenici, Gary Donohoe, Alysa E Doyle, Elodie Drapeau, Jubao Duan, Frank Dudbridge, Naser Durmishi, Howard J Edenberg, Hannelore Ehrenreich, Peter Eichhammer, Amanda Elkin, Johan Eriksson, Valentina Escott-Price, Tonu Esko, Laurent Essioux, Bruno Etain, Ayman H Fanous, Stephen V Faraone, Kai-How Farh, Anne E Farmer, Martilias S Farrell, Jurgen Del Favero, Manuel A Ferreira, I Nicol Ferrier, Matthew Flickinger, Tatiana Foroud, Josef Frank, Barbara Franke, Lude Franke, Christine Fraser, Robert Freedman, Nelson B Freimer, Marion Friedl, Joseph I Friedman, Louise Frisen, Menachem Fromer, Pablo V Gejman, Giulio Genovese, Lyudmila Georgieva, Elliot S Gershon, Eco J De Geus, Ina Giegling, Michael Gill, Paola Giusti-Rodriguez, Stephanie Godard, Jacqueline I Goldstein, Vera Golimbet, Srihari Gopal, Scott D Gordon, Katherine Gordon-Smith, Jacob Gratten, Elaine K Green, Tiffany A Greenwood, Gerard Van Grootheest, Magdalena Gross, Detelina Grozeva, Weihua Guan, Hugh Gurling, Omar Gustafsson, Lieuwe de Haan, Hakon Hakonarson, Steven P Hamilton, Christian Hammer, Marian L Hamshere, Mark Hansen, Thomas F Hansen, Vahram Haroutunian, Annette M Hartmann, Martin Hautzinger, Andrew C Heath, Anjali K Henders, Frans A Henskens, Stefan Herms, Ian B Hickie, Maria Hipolito, Joel N Hirschhorn, Susanne Hoefels, Per Hoffmann, Andrea Hofman, Mads V Hollegaard, Peter A Holmans, Florian Holsboer, Witte J Hoogendijk, Jouke Jan Hottenga, David M Hougaard, Hailiang Huang, Christina M Hultman, Masashi Ikeda, Andres Ingason, Marcus Ising, Nakao Iwata, Assen V Jablensky, Stephane Jamain, Inge Joa, Edward G Jones, Ian Jones, Lisa Jones, Erik G Jonsson, Milan Macek Jr, Richard A Belliveau Jr, Antonio Julia, Tzeng Jung-Ying, Anna K Kahler, Rene S Kahn, Luba Kalaydjieva, Radhika Kandaswamy, Sena Karachanak-Yankova, Juha Karjalainen, David Kavanagh, Matthew C Keller, Brian J Kelly, John R Kelsoe, Kenneth S Kendler, James L Kennedy, Elaine Kenny, Lindsey Kent, Jimmy Lee Chee Keong, Andrey Khrunin, Yunjung Kim, George K Kirov, Janis Klovins, Jo Knight, James A Knowles, Martin

A Kohli, Daniel L Koller, Bettina Konte, Ania Korszun, Robert Krasucki, Vaidutis Kucinskas, Zita Ausrele Kucinskiene, Jonna Kuntsi, Hana Kuzelova-Ptackova, Phoenix Kwan, Mikael Landen, Niklas Langstrom, Mark Lathrop, Claudine Laurent, Jacob Lawrence, William B Lawson, Marion Leboyer, Phil Hyoun Lee, S Hong Lee, Sophie E Legge, Todd Lencz, Bernard Lerer, Klaus-Peter Lesch, Douglas F Levinson, Cathryn M Lewis, Jun Li, Miaoxin Li, Qingqin S Li, Tao Li, Kung-Yee Liang, Paul Lichtenstein, Jeffrey A Lieberman, Svetlana Limborska, Danyu Lin, Chunyu Liu, Jianjun Liu, Falk W Lohoff, Jouko Lonnqvist, Sandra K Loo, Carmel M Loughland, Jan Lubinski, Susanne Lucae, Donald MacIntyre, Pamela AF Madden, Patrik KE Magnusson, Brion S Maher, Pamela B Mahon, Wolfgang Maier, Anil K Malhotra, Jacques Mallet, Sara Marsal, Nicholas G Martin, Manuel Mattheisen, Keith Matthews, Morten Mattingsdal, Robert W McCarley, Steven A McCarroll, Colm McDonald, Kevin A McGhee, James J McGough, Patrick J McGrath, Peter McGuffin, Melvin G McInnis, Andrew M McIntosh, Rebecca McKinney, Alan W McLean, Francis J McMahon, Andrew McQuillin, Helena Medeiros, Sarah E Medland, Sandra Meier, Carin J Meijer, Bela Melegh, Ingrid Melle, Fan Meng, Raquelle I Mesholam-Gately, Andres Metspalu, Patricia T Michie, Christel M Middeldorp, Lefkos Middleton, Lili Milani, Vihra Milanova, Philip B Mitchell, Younes Mokrab, Grant W Montgomery, Jennifer L Moran, Gunnar Morken, Derek W Morris, Ole Mors, Preben B Mortensen, Valentina Moskvina, Bryan J Mowry, Pierandrea Muglia, Thomas W Muehleisen, Walter J Muir, Bertram Mueller-Myhsok, Kieran C Murphy, Robin M Murray, Richard M Myers, Inez Myin-Germeys, Benjamin M Neale, Michael C Neale, Mari Nelis, Stan F Nelson, Igor Nenadic, Deborah A Nertney, Gerald Nestadt, Kristin K Nicodemus, Caroline M Nievergelt, Liene Nikitina-Zake, Ivan Nikolov, Vishwajit Nimgaonkar, Laura Nisenbaum, Willem A Nolen, Annelie Nordin, Markus M Noethen, John I Nurnberger, Evaristus A Nwulia, Dale R Nyholt, Eadbhard O'Callaghan, Michael C O'Donovan, Colm O'Dushlaine, F Anthony O'Neill, Robert D Oades, Sang-Yun Oh, Ann Olincy, Line Olsen, Edwin JCG van den Oord, Roel A Ophoff, Jim Van Os, Urban Osby, Hogni Oskarsson, Michael J Owen, Aarno Palotie, Christos Pantelis, George N Papadimitriou, Sergi Papiol, Elena Parkhomenko, Carlos N Pato, Michele T Pato, Tiina Paunio, Milica Pejovic-Milovancevic, Brenda P Penninx, Michele L Pergadia, Diana O Perkins, Roy H Perlis, Tune H Pers, Tracey L Petryshen, Hannes Petursson, Benjamin S Pickard, Olli Pietilainen, Jonathan Pimm, Joseph Piven, Andrew J Pocklington, Porgeir Porgeirsson, Danielle Posthuma, James B Potash, John Powell, Alkes Price, Peter Propping, Ann E Pulver, Shaun M Purcell, Vinay Puri, Digby Quested, Emma M Quinn, Josep Antoni Ramos-Quiroga, Henrik B Rasmussen, Soumya Raychaudhuri, Karola Rehnstrom, Abraham Reichenberg, Andreas Reif, Mark A Reimers, Marta Ribases, John Rice, Alexander L Richards, Marcella Rietschel, Brien P Riley, Stephan Ripke, Joshua L Roffman, Lizzy Rossin, Aribert Rothenberger, Guy Rouleau, Panos Roussos, Douglas M Ruderfer, Dan Rujescu, Veikko Salomaa, Alan R Sanders, Susan Santangelo, Russell Schachar, Ulrich Schall, Martin Schalling, Alan F Schatzberg, William A Scheftner, Gerard Schellenberg, Peter R Schofield, Nicholas J Schork, Christian R Schubert, Thomas G Schulze, Johannes Schumacher, Sibylle G Schwab, Markus M Schwarz, Edward M Scolnick, Laura J Scott, Rodney J Scott, Larry J Seidman, Pak C Sham, Jianxin Shi, Paul D Shilling, Stanley I Shyn, Engilbert Sigurdsson, Teimuraz Silagadze, Jeremy M Silverman, Kang Sim, Pamela Sklar, Susan L Slager, Petr Slominsky, Susan L Smalley, Johannes H Smit, Erin N Smith, Jordan W Smoller, Hon-Cheong So, Erik Soderman, Edmund Sonuga-Barke, Chris C A Spencer, Eli A Stahl, Matthew State, Hreinn Stefansson, Kari Stefansson, Michael Steffens, Stacy Steinberg, Hans-Christoph Steinhausen, Elisabeth Stogmann, Richard E Straub, John Strauss, Eric Strengman, Jana Strohmaier, T Scott Stroup, Mythily Subramaniam, Patrick F Sullivan, James Sutcliffe, Jaana Suvisaari, Dra-

gan M Svrakic, Jin P Szatkiewicz, Peter Szatmari, Szabocls Szelinger, Anita Thapar, Srinivasa Thirumalai, Robert C Thompson, Draga Toncheva, Paul A Tooney, Sarah Tosato, Federica Tozzi, Jens Treutlein, Manfred Uhr, Juha Veijola, Veronica Vieland, John B Vincent, Peter M Visscher, John Waddington, Dermot Walsh, James TR Walters, Dai Wang, Qiang Wang, Stanley J Watson, Bradley T Webb, Daniel R Weinberger, Mark Weiser, Myrna M Weissman, Jens R Wendland, Thomas Werge, Thomas F Wienker, Dieter B Wildenauer, Gonneke Willemsen, Nigel M Williams, Stephanie Williams, Richard Williamson, Stephanie H Witt, Aaron R Wolen, Emily HM Wong, Brandon K Wormley, Naomi R Wray, Adam Wright, Jing Qin Wu, Hualin Simon Xi, Wei Xu, Allan H Young, Clement C Zai, Stan Zammit, Peter P Zandi, Peng Zhang, Xuebin Zheng, Fritz Zimprich, Frans G Zitman, and Sebastian Zoellner.

Genetic Consortium for Anorexia Nervosa (GCAN)

Vesna Boraska Perica, Christopher S Franklin, James A B Floyd, Laura M Thornton, Laura M Huckins, Lorraine Southam, N William Rayner, Ioanna Tachmazidou, Kelly L Klump, Janet Treasure, Cathryn M Lewis, Ulrike Schmidt, Federica Tozzi, Kirsty Kiezebrink, Johannes Hebebrand, Philip Gorwood, Roger A H Adan, Martien J H Kas, Angela Favaro, Paolo Santonastaso, Fernando Fernández-Aranda, Monica Gratacos, Filip Rybakowski, Monika Dmitrzak-Weglarz, Jaakko Kaprio, Anna Keski-Rahkonen, Anu Raevuori-Helkamaa, Eric F Van Furth, Margarita C T Slof-Op't Landt, James I Hudson, Ted Reichborn-Kjennerud, Gun Peggy S Knudsen, Palmiero Monteleone, Allan S Kaplan, Andreas Karwautz, Hakon Hakonarson, Wade H Berrettini, Yiran Guo, Dong Li, Nicholas J Schork, Gen Komaki, Tetsuya Ando, Hidetoshi Inoko, Tõnu Esko, Krista Fischer, Katrin Männik, Andres Metspalu, Jessica H Baker, Roger D Cone, Jennifer Dackor, Janiece E DeSocio, Christopher E Hilliard, Julie K O'Toole, Jacques Pantel, Jin P Szatkiewicz, Chrysecolla Taico, Stephanie Zerwas, Sara E Trace, Oliver S P Davis, Sietske Helder, Katharina Bühren, Roland Burghardt, Martina de Zwaan, Karin Egberts, Stefan Ehrlich, Beate Herpertz-Dahlmann, Wolfgang Herzog, Hartmut Imgart, André Scherag, Susann Scherag, Stephan Zipfel, Claudette Boni, Nicolas Ramoz, Audrey Versini, Marek K Brandys, Unna N Danner, Carolien de Kove, Judith Hendriks, Bobby P C Koeleman, Roel A Ophoff, Eric Strengman, Annemarie A van Elburg, Alice Bruson, Maurizio Clementi, Daniela Degortes, Monica Forzan, Elena Tenconi, Elisa Docampo, Geòrgia Escaramí, Susana Jiménez-Murcia, Jolanta Lissowska, Andrzej Rajewski, Neonila Szeszenia-Dabrowska, Agnieszka Slopien, Joanna Hauser, Leila Karhunen, Ingrid Meulenbelt, P Eline Slagboom, Alfonso Tortorella, Mario Maj, George Dedoussis, Dimitris Dikeos, Fragiskos Gonidakis, Konstantinos Tziouvas, Artemis Tsitsika, Hana Papezova, Lenka Slachtova, Debora Martaskova, James L Kennedy, Robert D Levitan, Zeynep Yilmaz, Julia Huemer, Doris Koubek, Elisabeth Merl, Gudrun Wagner, Paul Lichtenstein, Gerome Breen, Sarah Cohen-Woods, Anne Farmer, Peter McGuffin, Sven Cichon, Ina Giegling, Stefan Herms, Dan Rujescu, Stefan Schreiber, H-Erich Wichmann, Christian Dina, Rob Sladek, Giovanni Gambaro, Nicole Soranzo, Antonio Julia, Sara Marsal, Raquel Rabionet, Valerie Gaborieau, Danielle M Dick, Aarno Palotie, Samuli Ripatti, Elisabeth Widén, Ole A Andreassen, Thomas Espeseth, Astri Lundervold, Ivar Reinvang, Vidar M Steen, Stephanie Le Hellard, Morten Mattingsdal, Ioanna Ntalla, Vladimir Bencko, Lenka Foretova, Vladimir Janout, Marie Navratilova, Steven Gallinger, Dalila Pinto, Stephen W Scherer, Harald Aschauer, Laura Carlberg, Alexandra Schosser, Lars Alfredsson, Bo Ding, Lars Klareskog, Leonid Padyukov, Chris Finan, Gursharan Kalsi, Marion Roberts, Darren W Logan, Leena Peltonen, Graham R S Ritchie, Jeff C Barrett, Xavier Estivill, Anke Hinney, Patrick F Sullivan, David A Collier, Eleftheria Zeggini, and Cynthia M Bulik.

Wellcome Trust Case Control Consortium 3 (WTCCC3)

Carl A Anderson, Jeffrey C Barrett, James A B Floyd, Christopher S Franklin, Ralph McGinnis, Nicole Soranzo, Eleftheria Zeggini, Jennifer Sambrook, Jonathan Stephens, Willem H Ouwehand, Wendy L McArdle, Susan M Ring, David P Strachan, Graeme Alexander, Cynthia M Bulik, David A Collier, Peter J Conlon, Anna Dominiczak, Audrey Duncanson, Adrian Hill, Cordelia Langford, Graham Lord, Alexander P Maxwell, Linda Morgan, Leena Peltonen, Richard N Sandford, Neil Sheerin, Frederik O Vannberg, Hannah Blackburn, Wei-Min Chen, Sarah Edkins, Mathew Gillman, Emma Gray, Sarah E Hunt, Suna Nengut-Gumuscu, Simon Potter, Stephen S Rich, Douglas Simpkin, and Pamela Whittaker.

ReproGen Consortium

John RB Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tnu Esko, Gudmar Thorleifsson, Eva Albrecht, Wei Q Ang, Tanguy Corre, Diana L Cousminer, Bjarke Feenstra, Nora Franceschini, Andrea Ganna, Andrew D Johnson, Sanela Kjellqvist, Kathryn L Lunetta, George McMahon, Ilja M Nolte, Lavinia Paternoster, Eleonora Porcu, Albert V Smith, Lisette Stolk, Alexander Teumer, Natalia T?ernikova, Emmi Tikkanen, Sheila Ulivi, Erin K Wagner, Najaf Amin, Laura J Bierut, Enda M Byrne, JoukeJan Hottenga, Daniel L Koller, Massimo Mangino, Tune H Pers, Laura M YergesArmstrong, Jing Hua Zhao, Irene L Andrulis, Hoda AntonCulver, Femke Atsma, Stefania Bandinelli, Matthias W Beckmann, Javier Benitez, Carl Blomqvist, Stig E Bojesen, Manjeet K Bolla, Bernardo Bonanni, Hiltrud Brauch, Hermann Brenner, Julie E Buring, Jenny ChangClaude, Stephen Chanock, Jinhui Chen, Georgia ChenevixTrench, J. Margriet Colle, Fergus J Couch, David Couper, Andrea D Coveillo, Angela Cox, Kamila Czene, Adamo Pio D'adamo, George Davey Smith, Immaculata De Vivo, Ellen W Demerath, Joe Dennis, Peter Devilee, Aida K Dieffenbach, Alison M Dunning, Gudny Eiriksdottir, Johan G Eriksson, Peter A Fasching, Luigi Ferrucci, Dieter FleschJanys, Henrik Flyger, Tatiana Foroud, Lude Franke, Melissa E Garcia, Montserrat GarcaClosas, Frank Geller, Eco EJ de Geus, Graham G Giles, Daniel F Gudbjartsson, Vilmundur Gudnason, Pascal Gunel, Suiqun Guo, Per Hall, Ute Hamann, Robin Haring, Catharina A Hartman, Andrew C Heath, Albert Hofman, Maartje J Hooning, John L Hopper, Frank B Hu, David J Hunter, David Karasik, Douglas P Kiel, Julia A Knight, VeliMatti Kosma, Zoltan Kutalik, Sandra Lai, Diether Lambrechts, Annika Lindblom, Reedik Mgi, Patrik K Magnusson, Arto Mannermaa, Nicholas G Martin, Gisli Masson, Patrick F McArdle, Wendy L McArdle, Mads Melbye Kyriaki Michailidou, Evelin Mihailov, Lili Milani, Roger L Milne, Heli Nevanlinna, Patrick Neven, Ellen A Nohr, Albertine J Oldehinkel, Ben A Oostra, Aarno Palotie,, Munro Peacock, Nancy L Pedersen, Paolo Peterlongo, Julian Peto, Paul DP Pharoah, Dirkje S Postma, Anneli Pouta, Katri Pylks, Paolo Radice, Susan Ring, Fernando Rivadeneira, Antonietta Robino, Lynda M Rose, Anja Rudolph, Veikko Salomaa, Serena Sanna, David Schlessinger, Marjanka K Schmidt, Mellissa C Southey, Ulla Sovio Meir J Stampfer, Doris Stckl Anna M Storniolo, Nicholas J Timpson Jonathan Tyrer, Jenny A Visser, Peter Vollenweider, Henry Vlzke, Gerard Waeber, Melanie Waldenberger, Henri Wallaschofski, Qin Wang, Gonneke Willemsen, Robert Winqvist, Bruce HR Wolffenbuttel, Margaret J Wright, Australian Ovarian Cancer Study The GENICA Network, kConFab, The LifeLines Cohort Study, The InterAct Consortium, Early Growth Genetics (EGG) Consortium, Dorret I Boomsma, Michael J Econs, KayTee Khaw, Ruth JF Loos, Mark I McCarthy, Grant W Montgomery, John P Rice, Elizabeth A Streeten, Unnur Thorsteinsdottir, Cornelia M van Duijn, Behrooz Z Alizadeh, Sven Bergmann,

Eric Boerwinkle, Heather A Boyd, Laura Crisponi, Paolo Gasparini, Christian Gieger, Tamara B Harris, Erik Ingelsson, MarjoRiitta Jrvelin, Peter Kraft, Debbie Lawlor, Andres Metspalu, Craig E Pennell, Paul M Ridker, Harold Snieder, Thorkild IA Srensen, Tim D Spector, David P Strachan, Andr G Uitterlinden, Nicholas J Wareham, Elisabeth Widen, Marek Zygmunt, Anna Murray, Douglas F Easton, Kari Stefansson, Joanne M Murabito, Ken K Ong.

Supplementary Tables

Simulations with one Binary Trait and one Quantitative Trait

Prevalence	$\hat{h}^2$	$\hat{h}^2_{liab}$	$\hat{r}_g$
0.01	0.72 (0.1)	0.59 (0.04)	0.51 (0.4)
0.05	0.72 (0.12)	0.59 (0.07)	0.45 (0.17)
0.2	0.72 (0.11)	0.6 (0.08)	0.46 (0.14)
0.5	0.73 (0.11)	0.59 (0.08)	0.42 (0.17)

Supplementary Table 1: Simulations with one binary trait and one quantitative trait. The prevalence column describes the population prevalence of the binary trait. We ran 100 simulations for each prevalence. The $\hat{h^2}$ column shows the mean heritability estimate for the quantitative trait. The $\hat{h}^2_{liab}$ column shows the mean liability-scale heritability estimate for the binary trait. The $\hat{r}_g$ column shows the mean genetic correlation estimate. Standard deviations across 100 simulations in parentheses. The true parameter values were $r_g = 0.46$, $h^2 = 0.7$ for the quantitative trait and $h^2_{liab} = 0.6$ for the binary trait. For all simulations, the quantitative trait sample size was 1000, the binary trait sample size was 1000 cases and 1000 controls, and there were 500 overlapping samples. There were 1000 effective independent SNPs. The environmental covariance was 0.2. We simulated case/control ascertainment using simulated LD block genotypes and a rejection sampling model of ascertainment. This is the same strategy used to simulate case/control ascertainment in [1].

Simulations with MAF- and LD-Dependent Genetic Architecture

LD Score	h^2(5-50%)	ρ_g(5-50%)	r_g(5-50%)
Truth	0.83	0.42	0.5
HM3	0.53 (0.08)	0.28 (0.07)	0.52 (0.1)
PSG	0.36 (0.08)	0.18 (0.06)	0.5 (0.13)
30 Bins	0.81 (0.12)	0.41 (0.08)	0.51 (0.09)
60 Bins	0.81 (0.12)	0.41 (0.09)	0.51 (0.09)

Supplementary Table 2: Simulations with MAF- and LD-dependent genetic architecture. Effect sizes were drawn from normal distributions such that the variance of per-allele effect sizes was uncorrelated with MAF, and variants with LD Score below 100 were fourfold enriched for heritability. Sample size was 2062 with complete overlap between studies; causal SNPs were about 600,000 best-guess imputed 1kG SNPs on chr 2, and the SNPs retained for the LD Score regression were the subset of about 100,000 of these SNPs that were included in HM3. True parameter values are shown in the top line of the table. Estimates are averages across 100 simulations. Standard deviations (in parentheses) are standard deviations across 100 simulations. LD Scores were estimated using in-sample LD and a 1cM window. HM3 means LD Score with sum taken over SNPs in HM3. PSG (per-standardized-genotype) means LD Score with the sum taken over all SNPs in 1kG as in [1]. 30 bins means per-allele LD Score binned on a MAF by LD Score grid with MAF breaks at 0.05, 0.1, 0.2, 0.3 and 0.4 and LD Score breaks at 35, 75, 150 and 400. 60 bins means per-allele LD Score binned on a MAF by LD Score grid with MAF breaks at 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4 and 0.45 and LD Score breaks at 30, 60, 120, 200 and 300, These simulations demonstrate that naive (HM3, PSG) LD Score regression gives correct genetic correlation estimates even when heritability and genetic covariance estimates are biased, so long as genetic correlation does not depend on LD.

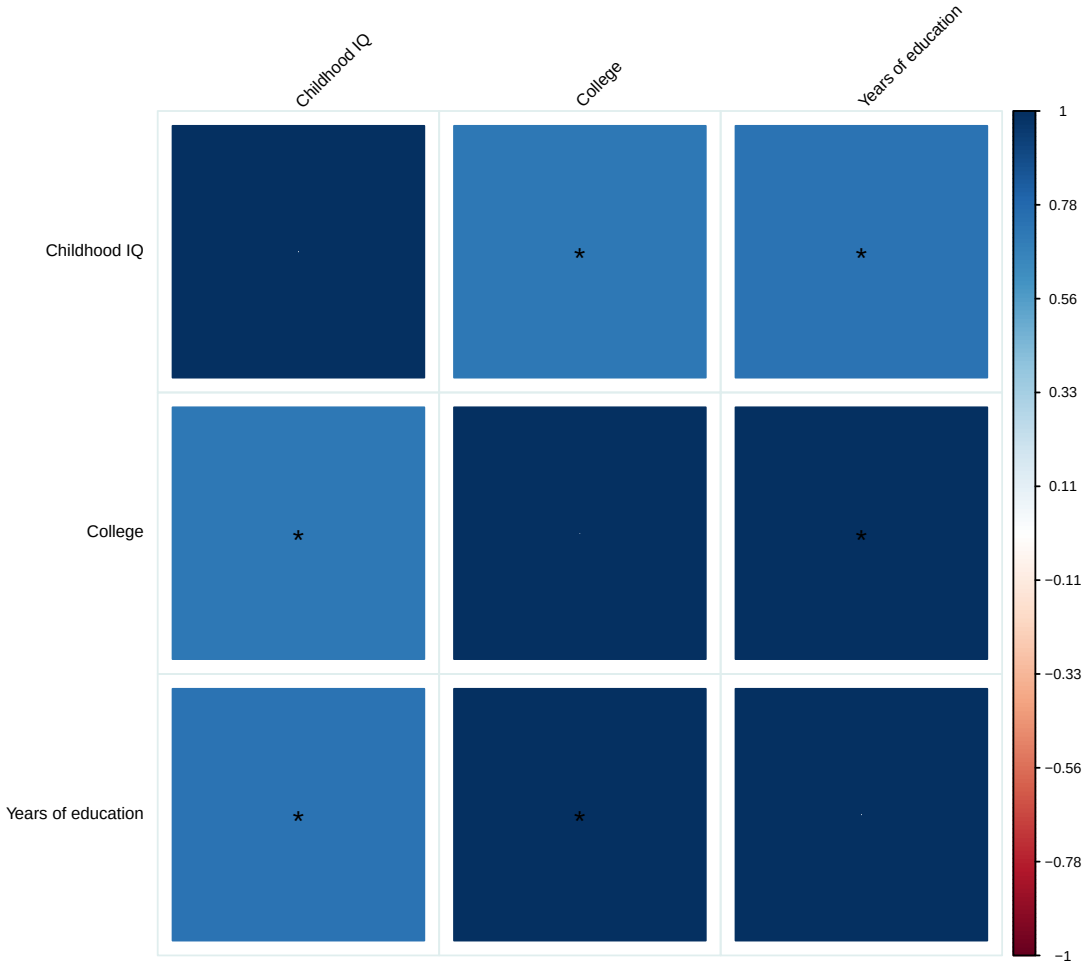
Sample Sizes and References

Trait	Reference	Sample Size
Schizophrenia	PGC Schizophrenia Working Group, *Nature,* 2014 [3]	70,100
Bipolar disorder	PGC Bipolar Working Group, *Nat Genet,* 2011 [4]	16,731
Major depression	PGC MDD Working Group, *Mol Psych,* 2013 [5]	18,759
Anorexia Nervosa	Boraska, *et al., Mol Psych,* 2014 [6]	17,767
Autism Spectrum Disorder	PGC Cross-Disorder Group, *Lancet,* 2013 [7]	10,263
Ever/Never Smoked	TAG Consortium, 2010 *Nat Genet,* [8]	74,035
Alzheimer's	Lambert, *et al., Nat Genet,* 2013 [9]	54,162
College	Rietveld, *et al., Science,* 2013 [10]	101,069
Height	Wood *et al., Nat Genet* 2014 [11]	253,288
BMI	Locke, *et al., Nature* 2015 [12]	236,231
Coronary Artery Disease	Schunkert, *et al., Nat Genet,* 2011 [13]	86,995
Triglycerides	Teslovich, *et al., Nature,* 2010 [14]	96,598
LDL Cholesterol	Teslovich, *et al., Nature,* 2010 [14]	95,454
HDL Cholesterol	Teslovich, *et al., Nature,* 2010 [14]	99,900
Type-2 Diabetes	Morris, *et al., Nat Genet,* 2012 [15]	69,033
Fasting Glucose	Manning, *et al., Nat Genet,* 2012 [16]	46,186
Childhood Obesity	EGG Consortium, *Nat Genet,* 2012 [17]	13,848
Birth Length	van der Valk, *et al., HMG,* 2014 [18]	22,263
Birth Weight	Horikoshi, *et al., Nat Genet,* 2013 [19]	26,836
Infant Head Circumference	Taal, *et al., Nat Genet,* 2012 [20]	10,767
Age at Menarche	Perry, *et al., Nature,* 2014 [21]	132,989
Crohn's Disease	Jostins, *et al., Nature,* 2012 [22]	20,883
Ulcerative Colitis	Jostins, *et al., Nature,* 2012 [22]	27,432
Rheumatoid Arthritis	Stahl, *et al., Nat Genet,* 2010 [23]	25,708

Supplementary Table 3: Sample sizes and references for traits analyzed in the main text.

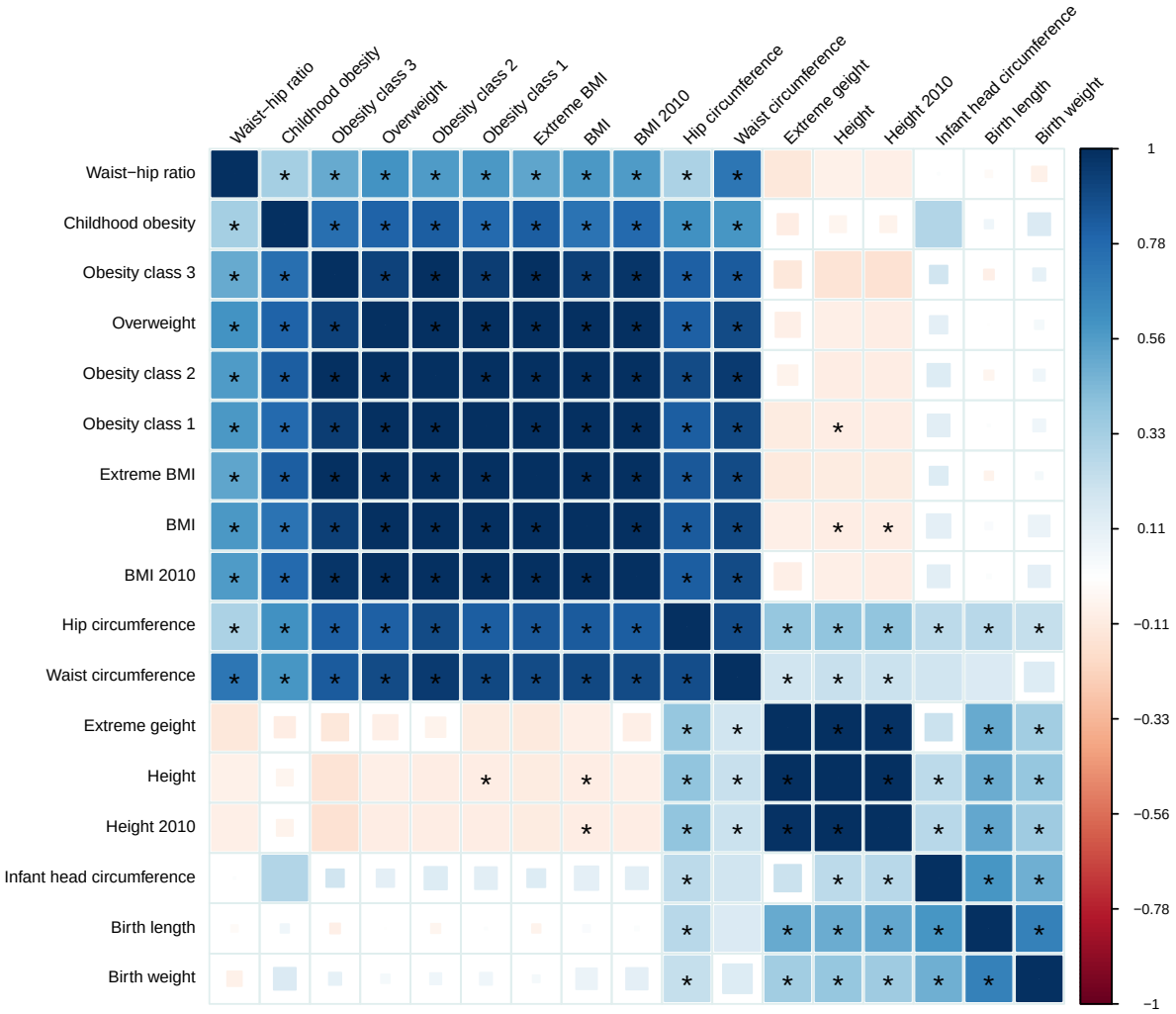
Supplementary Figures

Genetic Correlation between Educational Attainment and IQ Phenotypes



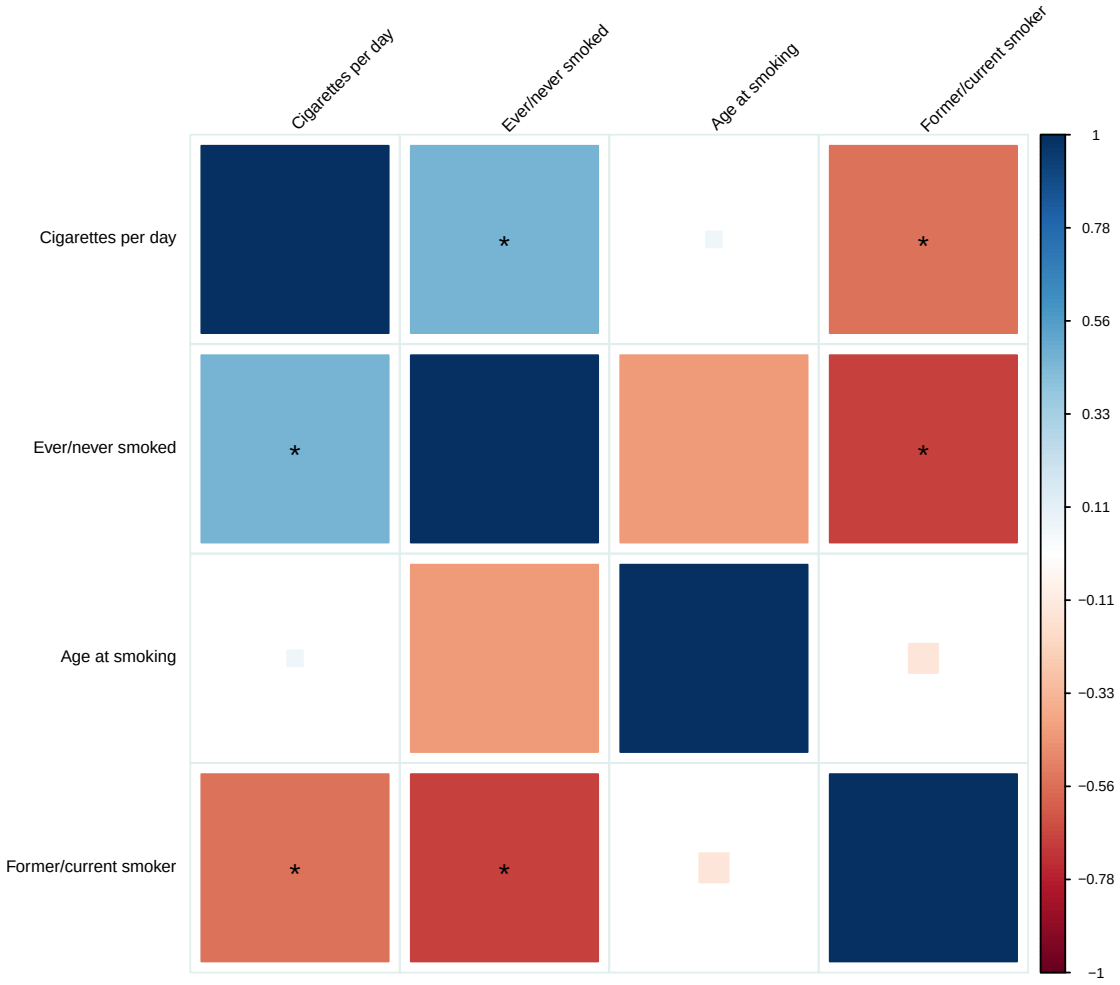
Supplementary Figure 1: Genetic correlation between the two educational attainment phenotypes from Rietveld, et al. [10] and childhood IQ from [24]. The structure of the figure is the same as Figure 2 in the main text: blue corresponds to positive genetic correlations; red corresponds to negative genetic correlation. Larger squares correspond to more significant p-values. Genetic correlations that are different from zero at 1% FDR are shown as full-sized squares. Genetic correlations that are significantly different from zero at significance level 0.05 after Bonferroni correction are given an asterisk.

Genetic Correlations among Anthropometric Traits



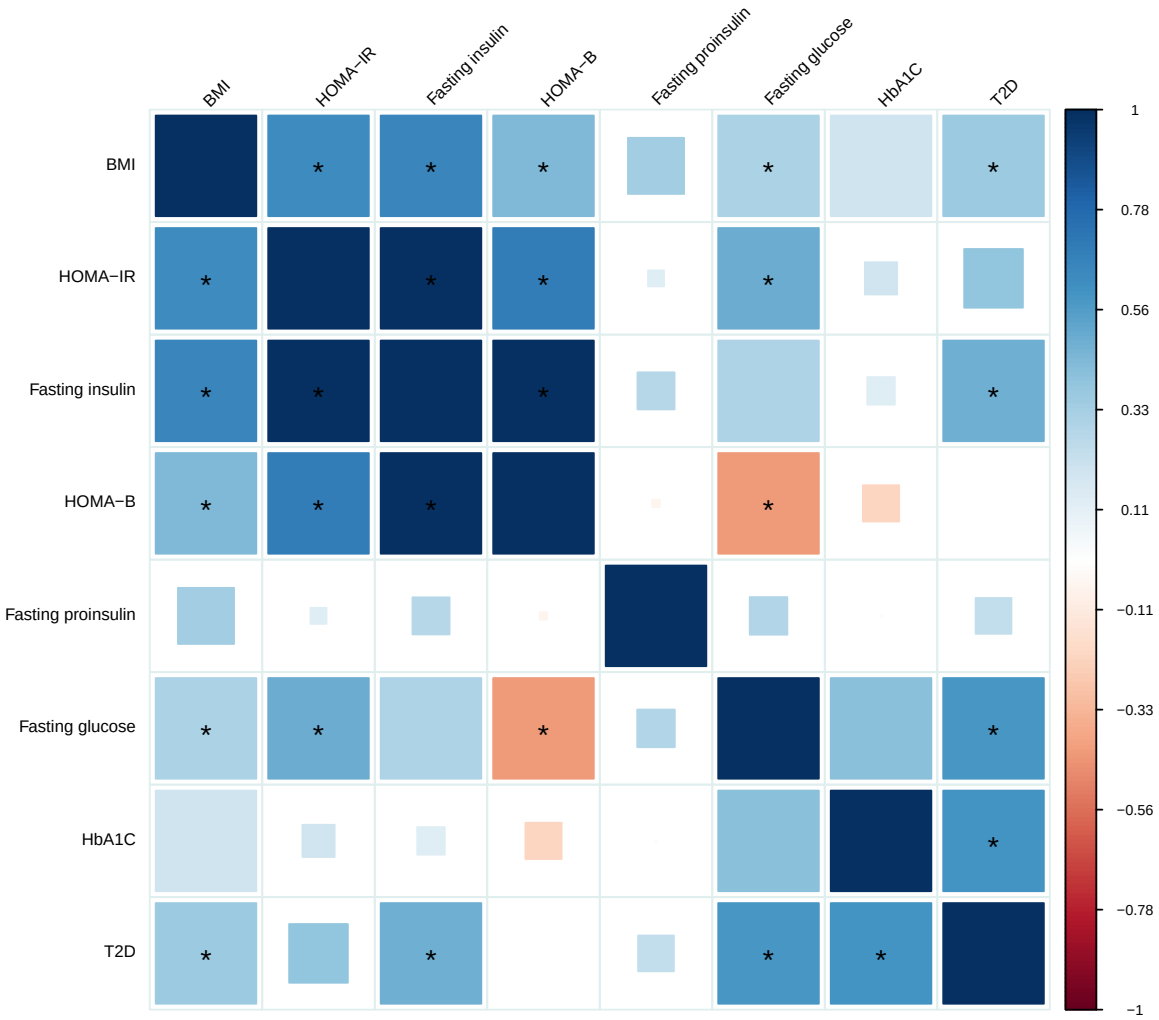
Supplementary Figure 2: Genetic correlations among anthropometric traits from studies by the GIANT and EGG consortia. The structure of the figure is the same as Figure 2 in the main text: blue corresponds to positive genetic correlations; red corresponds to negative genetic correlation. Larger squares correspond to more significant p-values. Genetic correlations that are different from zero at 1% FDR are shown as full-sized squares. Genetic correlations that are significantly different from zero at significance level 0.05 after Bonferroni correction are given an asterisk. BMI 2010 and Height 2010 refer to the results from [25] and [26], respectively.

Genetic Correlations among Smoking Traits



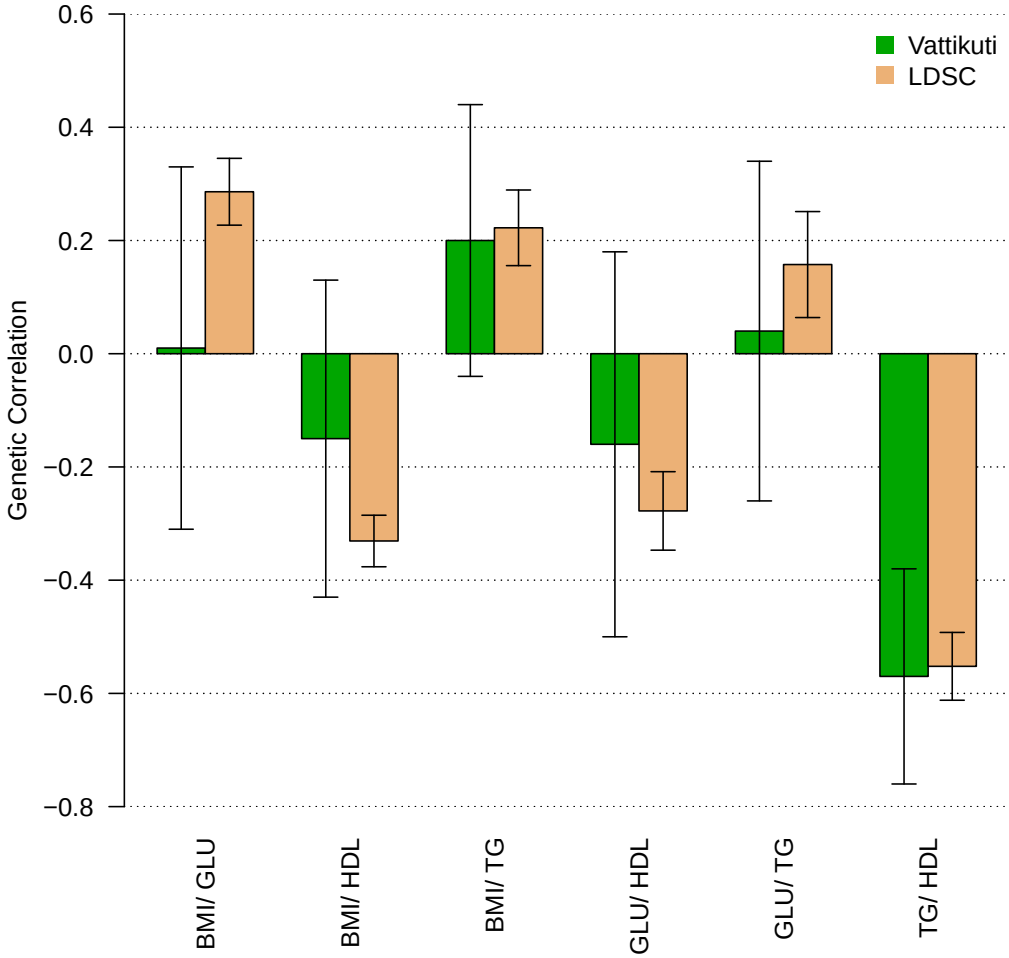
Supplementary Figure 3: Genetic correlations among smoking traits from the Tobacco and Genetics (TAG) consortium [8]. The structure of the figure is the same as Figure 2 in the main text: blue corresponds to positive genetic correlations; red corresponds to negative genetic correlation. Larger squares correspond to more significant p-values. Genetic correlations that are different from zero at 1% FDR are shown as full-sized squares. Genetic correlations that are significantly different from zero at significance level 0.05 after Bonferroni correction are given an asterisk.

Genetic Correlations among Glycemic Traits



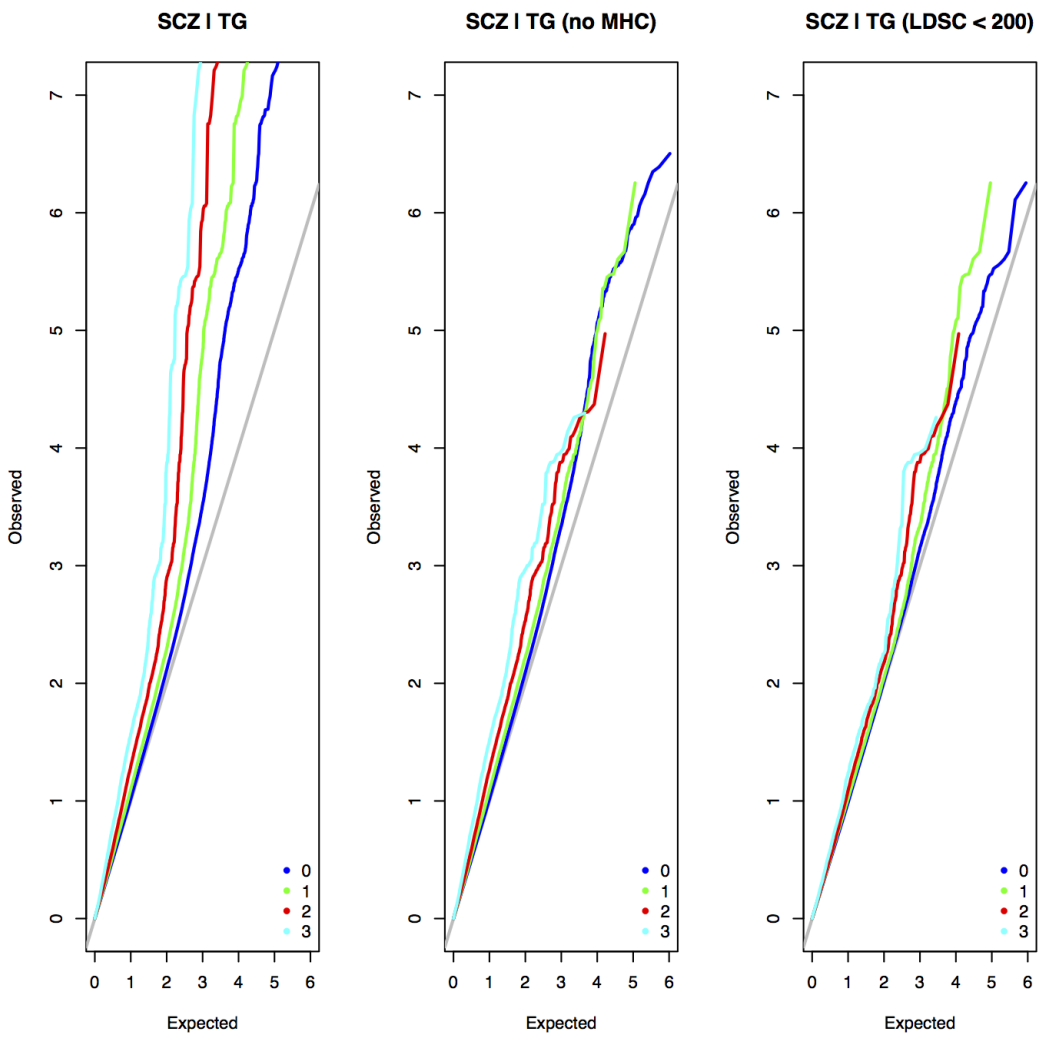
Supplementary Figure 4: Genetic correlations among insulin-related traits from studies by the MAGIC consortium. We have also included BMI data from [12] and T2D data from [15]. The structure of the figure is the same as Figure 2 in the main text: blue corresponds to positive genetic correlations; red corresponds to negative genetic correlation. Larger squares correspond to more significant p-values. Genetic correlations that are different from zero at 1% FDR are shown as full-sized squares. Genetic correlations that are significantly different from zero at significance level 0.05 after Bonferroni correction are given an asterisk.

Metabolic Genetic Correlations from Vattikuti, et al. and LD Score



Supplementary Figure 5: This figure compares estimates of genetic correlations among metabolic traits from table 3 of Vattikuti et al. [27] to estimates from LD Score regression. The LD Score regression estimates used much larger sample sizes. Error bars are standard errors.

Schizophrenia — TG Conditional QQ Plot with and without the MHC



Supplementary Figure 6: At left, we reproduced the conditional QQ plot comparing schizophrenia (SCZ) and triglycerides (TG) from Andreassen et al. [28] using the same data (PGC1 schizophrenia [29] and TG from Teslovich, et al. [14]). Conditional QQ plots show the distribution of p-values for SCZ conditional on the $-\log_{10}(p)$ for TG exceeding different thresholds. The thresholds are indicated by color, as described in the legends. Dark blue corresponds to no threshold, green corresponds to $-\log_{10}(p) > 1$, red corresponds to $-\log_{10}(p) > 2$ and light blue corresponds to $-\log_{10}(p) > 3$. The major histocompatibility complex (MHC, chr6, 25-35 MB) is a genomic region containing SNPs with exceptionally long-range LD and the strongest GWAS association for schizophrenia [3], as well as an association to TG [14]. If we remove the MHC, the signal of enrichment in the conditional QQ plot is substantially attenuated (middle); in particular, the red line falls below the green and blue lines (which correspond to less stringent thresholds for TG). If in addition we remove SNPs with very high LD Scores ($\ell > 200$, roughly the top 15% of SNPs), the signal of enrichment is further attenuated. The most likely explanation for the attenuation is that conditional QQ plots will report pleiotropy if causal SNPs are in LD (even if the causal SNPs for trait 1 are different from the causal SNPs for trait 2), which is more likely to occur in regions with long-range LD.