SUPPLEMENTARY NOTE

1. LD Score in an Unstructured Sample

1.1. **Model.** We model phenotypes as generated from the equation

$$\phi = X\beta + \epsilon \tag{1.1}$$

where ϕ is an $N \times 1$ vector of (quantitative) phenotypes, X is an $N \times M$ matrix of genotypes normalized to mean zero and variance one[1] (we ignore the distinction between normalizing and centering in our sample and in the population since the error so introduced has expectation zero and $\mathcal{O}(1/N)$ variance), β is an $M \times 1$ vector of per-normalized-genotype effect sizes and ϵ is an $N \times 1$ vector of environmental effects. We describe a model where all three variables on the right side of equation 1.1 are random. In this model, $\mathbb{E}[\epsilon] = 0$, $\mathrm{Var}[\epsilon] = (1 - h_g^2)I$, $\mathbb{E}[\beta] = 0$ and $\mathrm{Var}[\beta] = (h_g^2/M)I$.[2]To model genotypes, we assume that the genotype at variant j for individual i is independent of other individuals' genotypes, but we do incorporate linkage disequilibrium into the model: define $r_{jk} := \mathbb{E}[X_{ij}X_{ik}]$, which does not depend on i. Finally, we assume that X, β and ϵ are mutually independent. We will relax the assumption that environmental effects are independent of genotype when we model population stratification in §2.2.

1.2. **Relationship between LD and χ^2-Statistics.** For each variant $j = 1, \ldots, M$, we compute least-squares estimates of effect size $\hat{\beta}_j := X_j^{\mathsf{T}}\phi/N$ (where X_j denotes the $N \times 1$ vector of genotypes at variant j) and χ^2-statistics $\chi_j^2 := N\hat{\beta}_j^2$. In this section, we compute $\mathbb{E}[\chi_j^2]$ with the expectation taken over random X, β, and ϵ.

Proposition 1. *Define the **LD Score** of variant j as*

$$\ell_j := \sum_{k=1}^{M} r_{jk}^2. \tag{1.2}$$

Under the model described in §1.1, the expected χ^2-statistic of variant j is

$$\mathbb{E}[\chi_j^2] \approx \frac{Nh_g^2}{M}\ell_j + 1. \tag{1.3}$$

[1]Note that the normalization to variance hides an implicit assumption that rare SNPs have larger effect sizes. We show via simulation in the main text that LD Score regression, as an inference procedure, is not particularly sensitive to this assumption.

[2]This is almost the same assumption made in [1] (the difference is that they assume *i.i.d.* per-normalized genotype effect sizes for *genotyped* SNPs, and we make this assumption for all SNPs). If one wishes to specify a different variance structure for the per-normalized-genotype effect sizes, *e.g.*, $\mathrm{Var}[\beta_j] = f_j$, then all results presented herein hold with normalized genotypes $(G_{ij} - 2p_j)/\sqrt{f_j}$ replacing the usual $(G_{ij} - 2p_j)/\sqrt{2p_j(1-p_j)}$, where G_{ij} denotes additively coded (0,1,2) genotypes.

Proof. Since $\mathbb{E}[\hat{\beta}_j] = 0$, observe that $\mathbb{E}[\chi_j^2] = N \cdot \mathrm{Var}[\hat{\beta}_j]$. We will obtain the variance of $\hat{\beta}_j$ via the law of total variance:

$$\begin{aligned} \mathrm{Var}[\hat{\beta}_j] &= \mathbb{E}[\mathrm{Var}[\hat{\beta}_j \,|\, X]] + \mathrm{Var}[\mathbb{E}[\hat{\beta}_j \,|\, X]] \\ &= \mathbb{E}[\mathrm{Var}[\hat{\beta}_j \,|\, X]], \end{aligned} \tag{1.4}$$

where the second line follows from the fact that $\mathbb{E}[\hat{\beta}_j \,|\, X] = 0$, irrespective of X. First,

$$\begin{aligned} \mathrm{Var}[\hat{\beta}_j \,|\, X] &= \frac{1}{N^2}\mathrm{Var}[X_j^\mathsf{T}\phi \,|\, X] \\ &= \frac{1}{N^2}X_j^\mathsf{T}\mathrm{Var}[\phi \,|\, X]X_j \\ &= \frac{1}{N^2}\left(\frac{h_g^2}{M}X_j^\mathsf{T}XX^\mathsf{T}X_j + N(1-h_g^2)\right). \end{aligned} \tag{1.5}$$

We can write the term on the left in terms of more familiar quantities as

$$\frac{1}{N^2}X_j^\mathsf{T}XX^\mathsf{T}X_j = \sum_{k=1}^{M}\tilde{r}_{jk}^2, \tag{1.6}$$

where $\tilde{r}_{jk} := \frac{1}{N}\sum_{i=1}^{N} X_{ij}X_{ik}$ denotes the sample correlation between additively-coded genotypes at variants j and k. Since

$$\mathbb{E}[\tilde{r}_{jk}^2] \approx r_{jk}^2 + (1 - r_{jk}^2)/N, \tag{1.7}$$

(where the approximation sign hides terms of order $\mathscr{O}(1/N^2)$ and smaller; one can obtain this approximation via *e.g.,* the δ-method),

$$\mathbb{E}\left[\sum_{k=1}^{M}\tilde{r}_{jk}^2\right] \approx \ell_j + \frac{M - \ell_j}{N}. \tag{1.8}$$

Thus,

$$\begin{aligned} \mathbb{E}[\chi_j^2] &\approx \frac{N(1-1/N)h_g^2}{M}\ell_j + 1 \\ &\approx \frac{Nh_g^2}{M}\ell_j + 1, \end{aligned} \tag{1.9}$$

Values of N (study sample size) considered in the main text generally fall between 10^4 and 10^5, so the approximation $1 - 1/N \approx 1$ is appropriate. □

2. LD Score with Population Stratification

2.1. **Model of Population Structure.** We model population structure induced by genetic drift in a mixture of two populations in equal proportions as follows: we draw a matrix of normalized genotypes X consisting of $N/2$ samples from population 1 and $N/2$ samples from population 2 (we will use the notation $i \in P_m$ for $m \in \{1, 2\}$ to denote that individual i is a member of population m), subject to the following constraints: $\mathrm{Var}[X_{ij}] = 1$, $\mathbb{E}[X_{ij} \,|\, i \in P_1] = f_j$ and $\mathbb{E}[X_{ij} \,|\, i \in P_2] = -f_j$.

We model the drift term f as $f \sim N(0, F_{ST}V)$, where V is a correlation matrix[3] and F_{ST} is Wright's F_{ST} [2]. We postpone discussion of the off-diagonal entries of V (which might depend on LD in the ancestral population or recombination rates) until §2.2. Finally, if $\ell_{j,m}$ denotes the LD Score of variant j in population m, we assume that $\ell_{j,1} \approx \ell_{j,2} =: \ell_j$. The last assumption warrants a brief explanation. Assuming approximately equal LD Scores in both populations is certainly not reasonable for very large values of F_{ST} (*e.g.,* if population 1 and population 2 are from different continents) or in scenarios where one population has passed through a more severe bottleneck than the other (*e.g.,* if population 1 is from Finland and population 2 is from West Africa). However, we are interested in modeling the population stratification that may remain after principal components analysis in GWAS that sample from non-admixed populations, and for this purpose the assumption that $\ell_{j,1} \approx \ell_{j,2}$ seems reasonable, and is supported by the large values of $R^2(\ell_{j,m}, \ell_{j,n})$ that we observe for all pairs (m, n) of 1000 Genomes European subpopulations.

For reference, typical values of F_{ST} for human populations are ≈ 0.1 for populations from different continents, $F_{ST} \approx 0.01$ for populations on the same continent, and $F_{ST} < 0.01$ for subpopulations within the same country.

2.2. **LD in a Mixture of Populations.** Suppose j and k are unlinked variants such that $r_{jk,1} = r_{jk,2} = 0$ and f_j is independent of f_k. In a mixture of populations, it will often hold that j and k will be in LD in the whole population even if they are in equilibrium in both component populations. Let $r_{mix,jk}$ denote the correlation between SNPs j and k in such a mixture of populations. Conditional on f,

$$\begin{aligned} \mathbb{E}[r_{mix,jk} \mid f] &= \mathbb{E}[X_{ij}X_{ik} \mid f] \\ &= \frac{1}{2}\left(\mathbb{E}[X_{ij}X_{ik} \mid f, i \in P_1] + \mathbb{E}[X_{ij}X_{ik} \mid f, i \in P_2]\right) \\ &= f_j f_k. \end{aligned} \tag{2.1}$$

If we take the expectation over random f_j and f_j, then $\mathbb{E}[r_{mix,jk}] = 0$, because f_j and f_k are independent with expectation zero. We can use equation 2.1 to compute the variance,

$$\begin{aligned} \mathrm{Var}[r_{mix,jk}] &= \mathrm{Var}[\mathbb{E}[r_{mix,k} \mid f]] + \mathbb{E}[\mathrm{Var}[r_{mix,jk} \mid f]] \\ &= \mathbb{E}[f_j^2 f_k^2] + 0 \\ &= \mathbb{E}[f_j^2]\mathbb{E}[f_k^2] \\ &= F_{ST}^2. \end{aligned} \tag{2.2}$$

Observe that since $\mathbb{E}[r_{mix,jk}] = 0$, $\mathrm{Var}[r_{mix,jk}] = \mathbb{E}[r_{mix,jk}^2]$. By equation 1.7, in a finite sample,

$$\mathbb{E}[\tilde{r}_{mix,jk}^2] \approx F_{ST}^2 + (1 - F_{ST}^2)/N. \tag{2.3}$$

[3]In particular, we assume that the diagonal entries of V are all equal, or at least uncorrelated with LD Score. This assumption is unlikely to hold exactly: some parts of the genome drift faster than others, and the rate of drift may be correlated with LD Score (*e.g.,* as a result of linked selection). Nevertheless, our simulations with real population stratification show that this is not likely to be a severe confounder in LD Score regression.

Thus, the sample LD Score is approximately

$$\begin{aligned}\mathbb{E}[\tilde{\ell}_j] &\approx \ell_j + MF_{ST}^2 + \frac{M(1-F_{ST}^2)}{N} \\ &\approx \ell_j + MF_{ST}^2 + \frac{M}{N}.\end{aligned} \tag{2.4}$$

Note that we have ignored the case where j and k are linked and $V_{jk} \neq 0$. In this case, $\mathbb{E}[f_j^2 f_k^2] = F_{ST}^2 + 2F_{ST}^2 V_{jk}^2$ (from the formula for the double second moments of a multivariate normal distribution). Even if for some variants j, the number of variants k such that $V_{jk} > 0$ is $\approx 10^3$, this will make a negligible difference in $\mathbb{E}[\tilde{\ell}_j]$, because $\sum_{k:V_{jk}>0} 2F_{ST}^2 V_{jk}^2 < 2000F_{ST}^2 \ll MF_{ST}^2$ when $M \approx 10^7$.

2.3. **Model of Stratified Phenotype.** To model population stratification, we model phenotypes as generated by the equation

$$\phi = X\beta + S + \epsilon, \tag{2.5}$$

where X is as described in §2.1, β is as described in §1.1 and where S is an environmental stratification[4] term defined by

$$S_i := \begin{cases} \sigma_s/2, & i \in P_1 \\ -\sigma_s/2, & i \in P_2. \end{cases} \tag{2.6}$$

Finally, ϵ is as described in §1.1, except $\mathrm{Var}[\epsilon] = (1 - h_g^2 - \sigma_s^2)$, which assures that the variance of ϕ in the population is 1[5]. We compute χ^2-statistics as defined in §1.1. In this section, we compute $\mathbb{E}[\chi_j^2]$ with the expectation taken over random X, β, ϵ, f but with S fixed to ensure population stratification.

2.4. **Relationship between LD and Stratified χ^2-Statistics.**

Proposition 2. *Under the model described in §2.3, the expected χ^2-statistic of variant j is*

$$\mathbb{E}[\chi_j^2] = \frac{Nh_g^2}{M}\ell_j + 1 + aNF_{ST}, \tag{2.7}$$

*where a is the expectation of squared difference in mean phenotypes between population*1 *and population* 2*.*

Proof. Since $\mathbb{E}[\hat{\beta}_j] = 0$, observe that $\mathbb{E}[\chi_j^2] = N \cdot \mathrm{Var}[\hat{\beta}_j]$. We will obtain the variance of $\hat{\beta}_j$ via the law of total variance:

$$\mathrm{Var}[\hat{\beta}_j] = \mathbb{E}[\mathrm{Var}[\hat{\beta}_j \mid X]] + \mathrm{Var}[\mathbb{E}[\hat{\beta}_j \mid X]]. \tag{2.8}$$

Note that one can calculate f from X, so by conditioning on X we also implicitly condition on f. Unlike in equation 1.4, $\mathbb{E}[\hat{\beta}_j \mid X] \neq 0$, because of confounding from population stratification. The inner portion of the first term on the right side of equation 2.8 is the same as in equation 1.5,

[4] *Environmental* population stratification occurs when environmental effects are correlated with ancestry. *Genetic* population stratification occurs when the alelle frequency of trait-increasing alleles is correlated with ancestry. Our model includes environmental stratification and a small amount of genetic stratification from drift. Stronger genetic stratification requires the action of natural selection on the phenotype in question (or a related phenotype). For a more thorough discussion of genetic stratification, see [3].

[5] Note that we implicitly require $1 - h_g^2 - \sigma_s^2 \geq 0$.

$$\text{Var}[\hat{\beta}_j \mid X] = \frac{1}{N^2}\left(\frac{h_g^2}{M} X_j^\intercal X X^\intercal X_j + N(1-h_g^2)\right). \tag{2.9}$$

We can take the expectation over random X (and therefore over random f) using the result from equation 2.4. Thus,

$$\begin{aligned} \mathbb{E}[\text{Var}[\hat{\beta}_j \mid X]] &= \frac{1}{N^2}\left(\frac{h_g^2}{M}\mathbb{E}[X_j^\intercal X X^\intercal X_j] + N(1-h_g^2)\right) \\ &\approx \frac{h_g^2}{M}\ell_j + h_g^2 F_{ST}^2 + \frac{1}{N}. \end{aligned} \tag{2.10}$$

Next, the inner portion of the second term on the right side of equation 2.8 is

$$\begin{aligned} \mathbb{E}[\hat{\beta}_j \mid X] &= \frac{1}{N}\mathbb{E}[X_j^\intercal X\beta + X_j^\intercal S + X_j^\intercal \epsilon] \\ &= \frac{1}{N} X_j^\intercal S \\ &= f\sigma_s. \end{aligned} \tag{2.11}$$

Since f has variance F_{ST}, $\text{Var}[f\sigma_s] = \sigma_s^2 F_{ST}$. Thus,

$$\begin{aligned} \mathbb{E}[\chi_j^2] &= N \cdot \text{Var}[\hat{\beta}_j] \\ &\frac{Nh_g^2}{M}\ell_j + 1 + NF_{ST}(\sigma_s^2 + h_g^2 F_{ST}). \end{aligned} \tag{2.12}$$

We can interpret the final term, $NF_{ST}(\sigma_S^2 + h_g^2 F_{ST})$, as NF_{ST} times the expected squared mean difference in phenotype between populations, which has environmental component σ_s^2 and genetic component $h_g^2 F_{ST}$ (if we model X, β and f as random, there is zero genetic stratification on expectation, but with some small variance about zero). Precisely, if we let $\bar{\phi}_m$ denote the mean phenotype in population $m \in \{1, 2\}$, then

$$\begin{aligned} \mathbb{E}[(\bar{\phi}_1 - \bar{\phi}_2)] &= \sigma_s^2 + \sum_{j=1}^{M}\left[\mathbb{E}[\beta_j^2]\left(\sum_{i\in P_1}\mathbb{E}[X_{ij}^2 \mid i \in P_1] + \sum_{i\in P_2}\mathbb{E}[X_{ij}^2 \mid i \in P_2]\right)\right] \\ &= \sigma_s^2 + h_g^2 F_{ST}. \end{aligned} \tag{2.13}$$

Set $a := \mathbb{E}[(\bar{\phi}_1 - \bar{\phi}_2)^2]$. Then we have

$$\mathbb{E}[\chi_j^2] = \frac{Nh_g^2}{M}\ell_j + 1 + aNF_{ST}, \tag{2.14}$$

as desired. □

3. Variance

The results in §2.4 suggest a method for estimating the confounding term aNF_{ST} from summary statistics: if we regress χ^2 against LD Score, then the intercept minus one is an estimate of aNF_{ST}. Because the variance of χ^2 increases with LD Score, we can improve the efficiency of this estimator by weighting the regression by the reciprocal of the conditional variance function $\text{Var}[\chi_j^2 \mid \ell_j]$. We have derived the conditional expectation $\mathbb{E}[\chi_j^2 \mid \ell_j]$ without making distributional assumptions on β

or ϵ; however, we need stronger assumptions[6] in order to derive the conditional variance: we assume that N is large and $\beta \sim N(0, h_g^2 I)$, and $\epsilon \sim N(0, (1-h_g^2)I)$[7]. Then,

$$\begin{aligned} \hat{\beta}_j &= \frac{1}{N}(X_j^\mathsf{T} X\beta + X_j^\mathsf{T}\epsilon) \\ &\sim N(0, h_g^2 \ell_j/M + h_g^2/N) + N(0, (1-h_g^2)/N) \\ &\sim N(0, h_g^2 \ell_j/M + 1/N), \end{aligned} \tag{3.1}$$

where the second line follows from a central limit theorem argument and §.1.2. Thus, $\chi_j^2 = N\hat{\beta}_j^2$ follows a scaled χ^2 distribution with scale factor $Nh_g^2\ell_j/M$, and the conditional variance function is

$$\mathrm{Var}[\chi_j^2 \,|\, \ell_j] = \left(1 + \frac{Nh_g^2}{M}\ell_j\right)^2. \tag{3.2}$$

Note that this is the correct conditional variance function for GWAS with no confounding bias. Since most published GWAS have taken steps to control for population stratification, the most likely use case will be a GWAS with at most a small amount of population stratification.

4. Meta-Analysis

Consider a GWAS for a quantitative trait consisting of t sub-studies (all of which sample from the same population) with sample sizes $N_1, \ldots, N_k$ and total sample size N. For a SNP j, we compute z-scores $z_{j1}, \ldots, z_{jt}$ (via linear regression: $z_{j,s} := X_{j,s}^\mathsf{T}\phi_s/\sqrt{N_s}$, where $X_{j,s}$ is a vector of genotypes for SNP j in study s and ϕ_s is a vector of phenotypes for study s), then perform sample size weighted meta-analysis with single genomic control to obtain test statistics

$$z_{j,meta} := \sum_{s=1}^{t} z_{js}\sqrt{\frac{N_s}{\lambda_s N}}, \tag{4.1}$$

and

$$\chi_{j,meta}^2 = z_{j,meta}^2. \tag{4.2}$$

For meta-analyses without genomic control, set $\lambda_s = 1$ for all s.

[6]Note that the regression weights do not affect the expectation of the parameter estimates, only the standard error. Therefore, if the distributional assumptions that we make in order to derive the conditional variance are violated, it will only increase the standard error. Concretely, if there are very few causal SNPs, or if the distribution of effect sizes is particularly leptokurtotic, then $\mathrm{Var}[\chi_j^2 \,|\, \ell_j]$ will increase with ℓ_j faster than the function that we derive in this section, and our estimates will be inefficient.

[7]Normality is a stronger assumption than is necessary. We only need that $\hat{\beta}_j$ be normally distributed, which can hold even if β is not normal, so long as N is large and there are sufficiently many causal SNPs so that $\frac{1}{N}X_j^\mathsf{T} X\beta$ is approximately normal (which would follow from a CLT argument).

Proposition 3. *Under the model of meta-analysis with genomic control described above, the expected meta-analysis χ^2-statistic of variant j is*[8]

$$\frac{h_g^2}{MN}\left(\sum_{r=1}^{t}\sum_{s=1}^{t}\frac{N_r N_s}{\sqrt{\lambda_r \lambda_s}}\right)\ell_j + \frac{1}{N}\sum_{s=1}^{t}\frac{N_s}{\lambda_s}. \tag{4.3}$$

Proof. First, we need to compute $\mathbb{E}[z_{jr}z_{js}]$. Under a model where genotypes, environmental effects and SNP-effects are random, $\mathbb{E}[z_{js}] = 0$ for all s, so $\mathbb{E}[z_{jr}z_{js}]$ is equal to $\mathrm{Cov}[z_{jr}, z_{js}]$. Let X denote the matrix of normalized genotypes in study r and Y the matrix of normalized genotypes in study s. Let δ denote the vector of environmental effects in study r and ϵ the vector of environmental effects in study s. Suppose that there is no sample overlap, and more generally no cryptic relatedness within or between studies. Further suppose that there is no population stratification. Finally, assume that (to a reasonable approximation), the LD structure in all population from which samples were drawn is approximately equal (though this allows for sampling variance around the population parameter in each study). Then,

$$\begin{aligned}
\mathrm{Cov}[z_{jr}, z_{js}] &= \frac{1}{N}\mathrm{Cov}[(X_j^\mathsf{T} X\beta + X_j^\mathsf{T}\delta), (Y_j^\mathsf{T} Y\beta + Y_j^\mathsf{T}\epsilon)] \\
&= \frac{1}{N}\left(\mathrm{Cov}[X_j^\mathsf{T} X\beta, Y_j^\mathsf{T} Y\beta] + \mathrm{Cov}[X_j^\mathsf{T}\delta, Y_j^\mathsf{T}\epsilon]\right). \qquad (4.4)
\end{aligned}$$

We can evaluate these covariances with the law of total covariance. The term on the right is zero if we assume no sample overlap (because then δ and ϵ are independent). The term on the left is

$$\begin{aligned}
\mathrm{Cov}[X_j^\mathsf{T} X\beta, Y_j^\mathsf{T} Y\beta \,|\, X, Y] &= X_j^\mathsf{T} X \mathrm{Cov}[\beta, \beta] Y^\mathsf{T} Y_j \\
&= \frac{h_g^2}{M} X_j^\mathsf{T} X Y^\mathsf{T} Y_j. \qquad (4.5)
\end{aligned}$$

Removing the conditioning on X and Y,

$$\begin{aligned}
\mathrm{Cov}[X_j^\mathsf{T} X\beta, Y_j^\mathsf{T} Y\beta] &= \mathbb{E}[\mathrm{Cov}[X_j^\mathsf{T} X\beta, Y_j^\mathsf{T} Y\beta \,|\, X, Y]] + \mathrm{Cov}[\mathbb{E}[X_j^\mathsf{T} X\beta, \,|\, X], \mathbb{E}[Y_j^\mathsf{T} Y\beta \,|\, Y]] \\
&= \mathbb{E}[\mathrm{Cov}[X_j^\mathsf{T} X\beta, Y_j^\mathsf{T} Y\beta \,|\, X, Y]] \\
&= \frac{h_g^2}{M}\mathbb{E}[X_j^\mathsf{T} X Y^\mathsf{T} Y_j] \\
&= \frac{h_g^2}{M}\sum_{k=1}^{M}\sqrt{N_r N_s}\,\mathbb{E}[\hat{r}_{jk,r}\hat{r}_{jk,s}] \\
&= \frac{\sqrt{N_r N_s}\,h_g^2}{M}\ell_j, \qquad (4.6)
\end{aligned}$$

where $\hat{r}_{jk,r}$ denotes the sample correlation between SNPs j and k in study r. Since we have assumed that the samples in studies r and s are independent, $\hat{r}_{jk,r}$ and $\hat{r}_{jk,s}$ are independent estimates of the parameter r_{jk}, so $\mathbb{E}[\hat{r}_{jk,r}\hat{r}_{jk,s}] = \mathbb{E}[\hat{r}_{jk,r}]\mathbb{E}[\hat{r}_{jk,s}] =$

[8]Note that the intercept term will generally be less than one. On the other hand, if application of GC correction in each study was warranted (*i.e.,* if, for each s, all inflation in λ_s reflects confounding), and there is no between-study cryptic relatedness (*e.g.,* sample overlap) then the intercept should be 1.

r_{jk}^2. Then,

$$\begin{aligned}
\mathbb{E}[\chi^2_{j,meta}] &= \mathbb{E}\left[\left(\sum_{s=1}^{t} z_{js}\sqrt{\frac{N_s}{\lambda_s N}}\right)^2\right] \\
&= \frac{1}{N}\sum_{r=1}^{t}\sum_{s=1}^{t}\sqrt{\frac{N_r N_s}{\lambda_r\lambda_s}}\mathbb{E}[z_{js}z_{jr}] \\
&= \frac{1}{N}\sum_{r=1}^{t}\sum_{s=1}^{t}\frac{N_r N_s}{\sqrt{\lambda_r\lambda_s}}\left(\frac{h_g^2}{M}\ell_j\right) + \frac{1}{N}\sum_{s=1}^{t}\frac{N_s}{\lambda_s} \\
&= \frac{h_g^2}{MN}\left(\sum_{r=1}^{t}\sum_{s=1}^{t}\frac{N_r N_s}{\sqrt{\lambda_r\lambda_s}}\right)\ell_j + \frac{1}{N}\sum_{s=1}^{t}\frac{N_s}{\lambda_s}. \qquad (4.7)
\end{aligned}$$

□

Schizophrenia Working Group of the Psychiatric Genomics Consortium

Stephan Ripke[1,2], Benjamin M. Neale[1,2,3,4], Aiden Corvin[5], James T. R. Walters[6], Kai-How Farh[1], Peter A. Holmans[6,7], Phil Lee[1,2,4], Brendan Bulik-Sullivan[1,2], David A. Collier[8,9], Hailiang Huang[1,3], Tune H. Pers[3,10,11], Ingrid Agartz[12,13,14], Esben Agerbo[15,16,17], Margot Albus[18], Madeline Alexander[19], Farooq Amin[20,21], Silviu A. Bacanu[22], Martin Begemann[23], Richard A Belliveau Jr[2], Judit Bene[24,25], Sarah E. Bergen [2,26], Elizabeth Bevilacqua[2], Tim B Bigdeli [22], Donald W. Black[27], Richard Bruggeman[28], Nancy G. Buccola[29], Randy L. Buckner[30,31,32], William Byerley[33], Wiepke Cahn[34], Guiqing Cai[35,36], Murray J. Cairns[39,120,170], Dominique Campion[37], Rita M. Cantor[38], Vaughan J. Carr[39,40], Noa Carrera[6], Stanley V. Catts[39,41], Kimberly D. Chambert[2], Raymond C. K. Chan[42], Ronald Y. L. Chen[43], Eric Y. H. Chen[43,44], Wei Cheng[45], Eric F. C. Cheung[46], Siow Ann Chong[47], C. Robert Cloninger[48], David Cohen[49], Nadine Cohen[50], Paul Cormican[5], Nick Craddock[6,7], Benedicto Crespo-Facorro[210], James J. Crowley[51], David Curtis[52,53], Michael Davidson[54], Kenneth L. Davis[36], Franziska Degenhardt[55,56], Jurgen Del Favero[57], Lynn E. DeLisi[128,129] , Ditte Demontis[17,58,59], Dimitris Dikeos[60], Timothy Dinan[61], Srdjan Djurovic[14,62], Gary Donohoe[5,63], Elodie Drapeau[36], Jubao Duan[64,65], Frank Dudbridge[66], Naser Durmishi[67], Peter Eichhammer[68], Johan Eriksson[69,70,71], Valentina Escott-Price[6], Laurent Essioux[72], Ayman H. Fanous[73,74,75,76], Martilias S. Farrell[51], Josef Frank[77], Lude Franke[78], Robert Freedman[79], Nelson B. Freimer[80], Marion Friedl[81], Joseph I. Friedman[36], Menachem Fromer[1,2,4,82], Giulio Genovese[2], Lyudmila Georgieva[6], Elliot S. Gershon[209], Ina Giegling[81,83], Paola Giusti-Rodríguez[51], Stephanie Godard[84], Jacqueline I. Goldstein[1,3], Vera Golimbet[85], Srihari Gopal[86], Jacob Gratten[87], Lieuwe de Haan[88], Christian Hammer[23], Marian L. Hamshere[6], Mark Hansen[89], Thomas Hansen[17,90], Vahram Haroutunian[36,91,92], Annette M. Hartmann[81], Frans A. Henskens[39,93,94], Stefan Herms[55,56,95], Joel N. Hirschhorn[3,11,96], Per Hoffmann[55,56,95], Andrea Hofman[55,56], Mads V. Hollegaard[97], David M. Hougaard[97], Masashi Ikeda[98], Inge Joa[99], Antonio Julià[100], René S. Kahn[34], Luba Kalaydjieva[101,102], Sena Karachanak-Yankova[103], Juha Karjalainen[78], David Kavanagh[6], Matthew C. Keller[104], Brian J. Kelly[120], James L. Kennedy[105,106,107], Andrey Khrunin[108], Yunjung Kim[51], Janis Klovins[109], James A. Knowles[110], Bettina Konte[81], Vaidutis Kucinskas[111], Zita Ausrele Kucinskiene[111], Hana Kuzelova-Ptackova[112], Anna K. Kähler[26], Claudine Laurent[19,113], Jimmy Lee Chee Keong[47,114], S. Hong Lee[87], Sophie E. Legge[6], Bernard Lerer[115], Miaoxin Li[43,44,116] Tao Li[117], Kung-Yee Liang[118], Jeffrey Lieberman[119], Svetlana Limborska[108], Carmel M. Loughland[39,120], Jan Lubinski[121], Jouko Lönnqvist[122], Milan Macek Jr[112], Patrik K. E. Magnusson[26], Brion S. Maher[123], Wolfgang Maier[124], Jacques Mallet[125], Sara Marsal[100], Manuel Mattheisen[17,58,59,126], Morten Mattingsdal[14,127], Robert W. McCarley[128,129], Colm McDonald[130], Andrew M. McIntosh[131,132], Sandra Meier[77], Carin J. Meijer[88], Bela Melegh[24,25], Ingrid Melle[14,133], Raquelle I. Mesholam-Gately[128,134], Andres Metspalu[135], Patricia T. Michie[39,136], Lili Milani[135], Vihra Milanova[137], Younes Mokrab[8], Derek W. Morris[5,63], Ole Mors[17,58,138], Kieran C. Murphy[139], Robin M. Murray[140], Inez Myin-Germeys[141], Bertram Müller-Myhsok[142,143,144], Mari Nelis[135], Igor Nenadic[145], Deborah A. Nertney[146], Gerald Nestadt[147], Kristin K. Nicodemus[148], Liene Nikitina-Zake[109], Laura Nisenbaum[149], Annelie Nordin[150], Eadbhard O'Callaghan[151], Colm O'Dushlaine[2], F. Anthony O'Neill[152], Sang-Yun Oh[153], Ann Olincy[79], Line

Olsen[17,90], Jim Van Os[141,154], Psychosis Endophenotypes International Consortium[155], Christos Pantelis[39,156], George N. Papadimitriou[60], Sergi Papiol[23], Elena Parkhomenko[36], Michele T. Pato[110], Tiina Paunio[157,158], Milica Pejovic-Milovancevic[159], Diana O. Perkins[160], Olli Pietiläinen[158,161], Jonathan Pimm[53], Andrew J. Pocklington[6], John Powell[140], Alkes Price[3],[162], Ann E. Pulver[147], Shaun M. Purcell[82], Digby Quested[163], Henrik B. Rasmussen[17,90], Abraham Reichenberg[36], Mark A. Reimers[164], Alexander L. Richards[6], Joshua L. Roffman[30,32], Panos Roussos[82,165], Douglas M. Ruderfer[6,82], Veikko Salomaa[71], Alan R. Sanders[64,65], Ulrich Schall[39,120], Christian R. Schubert[166], Thomas G. Schulze[77,167], Sibylle G. Schwab[168], Edward M. Scolnick[2], Rodney J. Scott[39,169,170], Larry J. Seidman[128,134], Jianxin Shi[171], Engilbert Sigurdsson[172], Teimuraz Silagadze[173], Jeremy M. Silverman[36,174], Kang Sim[47], Petr Slominsky[108], Jordan W. Smoller[2,4], Hon-Cheong So[43], Chris C. A. Spencer[175], Eli A. Stahl[3,82], Hreinn Stefansson[176], Stacy Steinberg[176], Elisabeth Stogmann[177], Richard E. Straub[178], Eric Strengman[179,34], Jana Strohmaier[77], T. Scott Stroup[119], Mythily Subramaniam[47], Jaana Suvisaari[122], Dragan M. Svrakic[48], Jin P. Szatkiewicz[51], Erik Söderman[12], Srinivas Thirumalai[180], Draga Toncheva[103], Paul A. Tooney[39,120,170] , Sarah Tosato[181], Juha Veijola[182,183], John Waddington[184], Dermot Walsh[185], Dai Wang[86], Qiang Wang[117], Bradley T. Webb[22], Mark Weiser[54], Dieter B. Wildenauer[186], Nigel M. Williams[6], Stephanie Williams[51], Stephanie H. Witt[77], Aaron R. Wolen[164], Emily H. M. Wong[43], Brandon K. Wormley[22], Jing Qin Wu[39,170], Hualin Simon Xi[187], Clement C. Zai[105,106], Xuebin Zheng[188], Fritz Zimprich[177], Naomi R. Wray[87], Kari Stefansson[176], Peter M. Visscher[87], Wellcome Trust Case-Control Consortium 2[189], Rolf Adolfsson[150], Ole A. Andreassen[14,133], Douglas H. R. Blackwood[132], Elvira Bramon[190], Joseph D. Buxbaum[35,36,91,191], Anders D. Børglum[17,58,59,138], Sven Cichon[55,56,95,192], Ariel Darvasi[193], Enrico Domenici[194], Hannelore Ehrenreich[23], Tõnu Esko[3,11,96,135], Pablo V. Gejman[64,65], Michael Gill[5], Hugh Gurling[53], Christina M. Hultman[26], Nakao Iwata[98], Assen V. Jablensky[39,102,186,195], Erik G. Jönsson[12,14], Kenneth S. Kendler[196], George Kirov[6], Jo Knight[105,106,107], Todd Lencz[197,198,199], Douglas F. Levinson[19], Qingqin S. Li[86], Jianjun Liu[188,200], Anil K. Malhotra[197,198,199], Steven A. McCarroll[2,96], Andrew McQuillin[53], Jennifer L. Moran[2], Preben B. Mortensen[15,16,17], Bryan J. Mowry[87,201], Markus M. Nöthen[55,56], Roel A. Ophoff[38,80,34], Michael J. Owen[6,7], Aarno Palotie[2,4,161], Carlos N. Pato[110], Tracey L. Petryshen[2,128,202], Danielle Posthuma[203,204,205], Marcella Rietschel[77], Brien P. Riley[196], Dan Rujescu[81,83], Pak C. Sham[43,44,116] Pamela Sklar[82,91,165], David St Clair[206], Daniel R. Weinberger[178,207], Jens R. Wendland[166], Thomas Werge[17,90,208], Mark J. Daly[1,2,3], Patrick F. Sullivan[26,51,160] & Michael C. O'Donovan[6,7]

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

[3]Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

[4]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[5]Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Dublin 8, Ireland.

[6]MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, CF24 4HQ, UK.

[7]National Centre for Mental Health, Cardiff University, Cardiff, CF24 4HQ, UK.

[8]Eli Lilly and Company Limited, Erl Wood Manor, Sunninghill Road, Windlesham, Surrey, GU20 6PH, UK. [9]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, SE5 8AF, UK.

[10]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, DK-2800, Denmark.

[11]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, Massachusetts, 02115USA.

[12]Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, SE-17176 Stockholm, Sweden. [13]Department of Psychiatry, Diakonhjemmet Hospital, 0319 Oslo, Norway.

[14]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, 0424 Oslo, Norway.

[15]Centre for Integrative Register-based Research, CIRRAU, Aarhus University, DK-8210 Aarhus, Denmark.

[16]National Centre for Register-based Research, Aarhus University, DK-8210 Aarhus, Denmark.

[17]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark.

[18]State Mental Hospital, 85540 Haar, Germany.

[19]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California 94305, USA.

[20]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Atlanta, Georgia 30033, USA.

[21]Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta Georgia 30322, USA.

[22]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[23]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen 37075, Germany.

[24]Department of Medical Genetics, University of Pécs, Pécs H-7624, Hungary.

[25]Szentagothai Research Center, University of Pécs, Pécs H-7624, Hungary.

[26]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden.

[27]Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, Iowa 52242, USA.

[28]University Medical Center Groningen, Department of Psychiatry, University of Groningen NL-9700 RB, The Netherlands.

[29]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, Louisiana 70112, USA.

[30]Athinoula A. Martinos Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA.

[31]Center for Brain Science, Harvard University, Cambridge, Massachusetts, 02138 USA.

[32]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, 02114 USA.

[33]Department of Psychiatry, University of California at San Francisco, San Francisco, California, 94143 USA.

[34]University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, 3584 Utrecht, The Netherlands.

[35]Department of Human Genetics, Icahn School of Medicine at Mount Sinai, New York, New York 10029 USA.

[36]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029 USA.

[37]Centre Hospitalier du Rouvray and INSERM U1079 Faculty of Medicine, 76301 Rouen, France.

[38]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

[39]Schizophrenia Research Institute, Sydney NSW 2010, Australia.

[40]School of Psychiatry, University of New South Wales, Sydney NSW 2031, Australia.

[41]Royal Brisbane and Women's Hospital, University of Queensland, Brisbane, St Lucia QLD 4072, Australia.

[42]Institute of Psychology, Chinese Academy of Science, Beijing 100101, China.

[43]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

[44]State Key Laboratory for Brain and Cognitive Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

[45]Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27514, USA.

[46]Castle Peak Hospital, Hong Kong, China.

[47]Institute of Mental Health, Singapore 539747, Singapore.

[48]Department of Psychiatry, Washington University, St. Louis, Missouri 63110, USA.

[49]Department of Child and Adolescent Psychiatry, Assistance Publique Hopitaux de Paris, Pierre and Marie Curie Faculty of Medicine and Institute for Intelligent Systems and Robotics, Paris, 75013, France.

[50] Blue Note Biosciences, Princeton, New Jersey 08540, USA

[51]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599-7264, USA.

[52]Department of Psychological Medicine, Queen Mary University of London, London E1 1BB, UK.

[53]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, London WC1E 6JJ, UK.

[54]Sheba Medical Center, Tel Hashomer 52621, Israel.

[55]Department of Genomics, Life and Brain Center, D-53127 Bonn, Germany.

[56]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany.

[57]Applied Molecular Genomics Unit, VIB Department of Molecular Genetics, University of Antwerp, B-2610 Antwerp, Belgium.

[58]Centre for Integrative Sequencing, iSEQ, Aarhus University, DK-8000 Aarhus C, Denmark.

[59]Department of Biomedicine, Aarhus University, DK-8000 Aarhus C, Denmark.

[60]First Department of Psychiatry, University of Athens Medical School, Athens 11528, Greece.

[61]Department of Psychiatry, University College Cork, Co. Cork, Ireland.

[62]Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway.

[63]Cognitive Genetics and Therapy Group, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Co. Galway, Ireland.

[64]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois 60637, USA.

[65]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, Illinois 60201, USA.

[66]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.

[67]Department of Child and Adolescent Psychiatry, University Clinic of Psychiatry, Skopje 1000, Republic of Macedonia.

[68]Department of Psychiatry, University of Regensburg, 93053 Regensburg, Germany.

[69]Department of General Practice, Helsinki University Central Hospital, University of Helsinki P.O. Box 20, Tukholmankatu 8 B, FI-00014, Helsinki, Finland

[70]Folkhälsan Research Center, Helsinki, Finland, Biomedicum Helsinki 1, Haartmaninkatu 8, FI-00290, Helsinki, Finland.

[71]National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland.

[72]Translational Technologies and Bioinformatics, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.

[73]Department of Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA.

[74]Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, California 90033, USA.

[75]Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, Virginia 23298, USA.

[76]Mental Health Service Line, Washington VA Medical Center, Washington DC 20422, USA.

[77]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg , D-68159 Mannheim, Germany.

[78]Department of Genetics, University of Groningen, University Medical Centre Groningen, 9700 RB Groningen, The Netherlands.

[79]Department of Psychiatry, University of Colorado Denver, Aurora, Colorado 80045, USA.

[80]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California 90095, USA.

[81]Department of Psychiatry, University of Halle, 06112 Halle, Germany.

[82]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[83]Department of Psychiatry, University of Munich, 80336, Munich, Germany.

[84]Departments of Psychiatry and Human and Molecular Genetics, INSERM, Institut de Myologie, Hôpital de la Pitiè-Salpêtrière, Paris, 75013, France.

[85]Mental Health Research Centre, Russian Academy of Medical Sciences, 115522 Moscow, Russia.

[86]Neuroscience Therapeutic Area, Janssen Research and Development, Raritan, New Jersey 08869, USA.

[87]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, QLD 4072, Australia.

[88]Academic Medical Centre University of Amsterdam, Department of Psychiatry, 1105 AZ Amsterdam, The Netherlands.

[89]Illumina, La Jolla, California, California 92122, USA.

[90]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, DK-4000, Denmark.

[91]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[92]J. J. Peters VA Medical Center, Bronx, New York, New York 10468, USA.

[93]Priority Research Centre for Health Behaviour, University of Newcastle, Newcastle NSW 2308, Australia.

[94]School of Electrical Engineering and Computer Science, University of Newcastle, Newcastle NSW 2308, Australia.

[95]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, CH-4058, Switzerland.

[96]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

[97]Section of Neonatal Screening and Hormones, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, DK-2300, Denmark.

[98]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Aichi, 470-1192, Japan.

[99]Regional Centre for Clinical Research in Psychosis, Department of Psychiatry, Stavanger University Hospital, 4011 Stavanger, Norway.

[100]Rheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, 08035, Spain.

[101]Centre for Medical Research, The University of Western Australia, Perth, WA 6009, Australia.

[102]The Perkins Institute for Medical Research, The University of Western Australia, Perth, WA 6009, Australia.

[103]Department of Medical Genetics, Medical University, Sofia1431, Bulgaria.

[104]Department of Psychology, University of Colorado Boulder, Boulder, Colorado 80309, USA.

[105]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, M5T 1R8, Canada.

[106]Department of Psychiatry, University of Toronto, Toronto, Ontario, M5T 1R8, Canada.

[107]Institute of Medical Science, University of Toronto, Toronto, Ontario, M5S 1A8, Canada.

[108]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow123182, Russia.

[109]Latvian Biomedical Research and Study Centre, Riga, LV-1067, Latvia.

[110]Department of Psychiatry and Zilkha Neurogenetics Institute, Keck School of Medicine at University of Southern California, Los Angeles, California 90089, USA.

[111]Faculty of Medicine, Vilnius University, LT-01513 Vilnius, Lithuania.

[112] Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic.

[113] Department of Child and Adolescent Psychiatry, Pierre and Marie Curie Faculty of Medicine, Paris 75013, France.

[114]Duke-NUS Graduate Medical School, Singapore 169857, Singapore.

[115]Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel.

[116]Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China.

[117]Mental Health Centre and Psychiatric Laboratory, West China Hospital, Sichuan University, Chengdu, 610041, Sichuan, China.

[118]Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205, USA.

[119]Department of Psychiatry, Columbia University, New York, New York 10032, USA.

[120]Priority Centre for Translational Neuroscience and Mental Health, University of Newcastle, Newcastle NSW 2300, Australia.

[121]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, 70-453 Szczecin, Poland.

[122]Department of Mental Health and Substance Abuse Services; National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland

[123]Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.

[124]Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany.

[125]Centre National de la Recherche Scientifique, Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, Hôpital de la Pitié Salpêtrière, 75013, Paris, France.

[126]Department of Genomics Mathematics, University of Bonn, D-53127 Bonn, Germany.

[127]Research Unit, Sørlandet Hospital, 4604 Kristiansand, Norway.

[128]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA.

[129]VA Boston Health Care System, Brockton, Massachusetts 02301, USA.

[130]Department of Psychiatry, National University of Ireland Galway, Co. Galway, Ireland.

[131]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH16 4SB, UK.

[132]Division of Psychiatry, University of Edinburgh, Edinburgh EH16 4SB, UK.

[133]Division of Mental Health and Addiction, Oslo University Hospital, 0424 Oslo, Norway.

[134]Massachusetts Mental Health Center Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, Massachusetts 02114, USA.

[135]Estonian Genome Center, University of Tartu, Tartu 50090, Estonia.

[136]School of Psychology, University of Newcastle, Newcastle NSW 2308, Australia.

[137]First Psychiatric Clinic, Medical University, Sofia 1431, Bulgaria.

[138]Department P, Aarhus University Hospital, DK-8240 Risskov, Denmark.

[139]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

[140]King's College London, London SE5 8AF, UK.

[141]Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, EURON, 6229 HX Maastricht, The Netherlands.

[142]Institute of Translational Medicine, University of Liverpool, Liverpool L69 3BX, UK.

[143]Max Planck Institute of Psychiatry, 80336 Munich, Germany.

[144]Munich Cluster for Systems Neurology (SyNergy), 80336 Munich, Germany.

[145]Department of Psychiatry and Psychotherapy, Jena University Hospital, 07743 Jena, Germany.

[146]Department of Psychiatry, Queensland Brain Institute and Queensland Centre for Mental Health Research, University of Queensland, Brisbane, Queensland, St Lucia QLD 4072, Australia.

[147]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

[148]Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland.

[149]Eli Lilly and Company, Lilly Corporate Center, Indianapolis, 46285 Indiana, USA.

[150]Department of Clinical Sciences, Psychiatry, Umeå University, SE-901 87 Umeå, Sweden.

[151]DETECT Early Intervention Service for Psychosis, Blackrock, Co. Dublin, Ireland.

[152]Centre for Public Health, Institute of Clinical Sciences, Queen's University Belfast, Belfast BT12 6AB, UK.

[153]Lawrence Berkeley National Laboratory, University of California at Berkeley, Berkeley, California 94720, USA.

[154]Institute of Psychiatry, King's College London, London SE5 8AF, UK.

[155]A list of authors and affiliations appear in the Supplementary Information.

[156]Melbourne Neuropsychiatry Centre, University of Melbourne & Melbourne Health, Melbourne, Vic 3053, Australia.

[157]Department of Psychiatry, University of Helsinki, P.O. Box 590, FI-00029 HUS, Helsinki, Finland.

[158]Public Health Genomics Unit, National Institute for Health and Welfare, P.O. BOX 30, FI-00271 Helsinki, Finland

[159]Medical Faculty, University of Belgrade, 11000 Belgrade, Serbia.

[160]Department of Psychiatry, University of North Carolina, Chapel Hill, North Carolina 27599-7160, USA.

[161]Institute for Molecular Medicine Finland, FIMM, University of Helsinki, P.O. Box 20

FI-00014, Helsinki, Finland

[162]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA.

[163]Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK.

[164]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[165]Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[166]PharmaTherapeutics Clinical Research, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.

[167]Department of Psychiatry and Psychotherapy, University of Gottingen, 37073 Göttingen, Germany.

[168]Psychiatry and Psychotherapy Clinic, University of Erlangen, 91054 Erlangen, Germany.

[169]Hunter New England Health Service, Newcastle NSW 2308, Australia.

[170]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan NSW 2308, Australia.

[171]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA.

[172]University of Iceland, Landspitali, National University Hospital, 101 Reykjavik, Iceland.

[173]Department of Psychiatry and Drug Addiction, Tbilisi State Medical University (TSMU), **N33, 0177** Tbilisi, Georgia.

[174]Research and Development, Bronx Veterans Affairs Medical Center, New York, New York 10468, USA.

[175]Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK.

[176]deCODE Genetics, 101 Reykjavik, Iceland.

[177]Department of Clinical Neurology, Medical University of Vienna, 1090 Wien, Austria.

[178]Lieber Institute for Brain Development, Baltimore, Maryland 21205, USA.

[179]Department of Medical Genetics, University Medical Centre Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands.

[180]Berkshire Healthcare NHS Foundation Trust, Bracknell RG12 1BQ, UK.

[181]Section of Psychiatry, University of Verona, 37134 Verona, Italy.

[182]Department of Psychiatry, University of Oulu, P.O. BOX 5000, 90014, Finland

[183]University Hospital of Oulu, P.O.BOX 20, 90029 OYS, Finland.

[184]Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin 2, Ireland.

[185]Health Research Board, Dublin 2, Ireland.

[186]School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth WA6009, Australia.

[187]Computational Sciences CoE, Pfizer Worldwide Research and Development, Cambridge, Massachusetts 02139, USA.

[188]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore 138672, Singapore.

[189]A list of authors and affiliations appear in the Supplementary Information.

[190]University College London, London WC1E 6BT, UK.

[191]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

[192]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, 52428 Juelich, Germany.

[193]Department of Genetics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel.

[194]Neuroscience Discovery and Translational Area, Pharma Research and Early Development, F. Hoffman-La Roche, CH-4070 Basel, Switzerland.

[195]Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Medical Research Foundation Building, Perth WA 6000, Australia.

[196]Virginia Institute for Psychiatric and Behavioral Genetics, Departments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia 23298, USA.

[197]The Feinstein Institute for Medical Research, Manhasset, New York, 11030 USA.

[198]The Hofstra NS-LIJ School of Medicine, Hempstead, New York, 11549 USA.

[199]The Zucker Hillside Hospital, Glen Oaks, New York,11004 USA.

[200]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore.

[201]Queensland Centre for Mental Health Research, University of Queensland, Brisbane 4076, Queensland, Australia.

[202]Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

[203]Department of Child and Adolescent Psychiatry, Erasmus University Medical Centre, Rotterdam 3000, The Netherlands.

[204]Department of Complex Trait Genetics, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam 1081, The Netherlands.

[205]Department of Functional Genomics, Center for Neurogenomics and Cognitive Research, Neuroscience Campus Amsterdam, VU University, Amsterdam 1081, The Netherlands.

[206]University of Aberdeen, Institute of Medical Sciences, Aberdeen, AB25 2ZD, UK.

[207]Departments of Psychiatry, Neurology, Neuroscience and Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA.

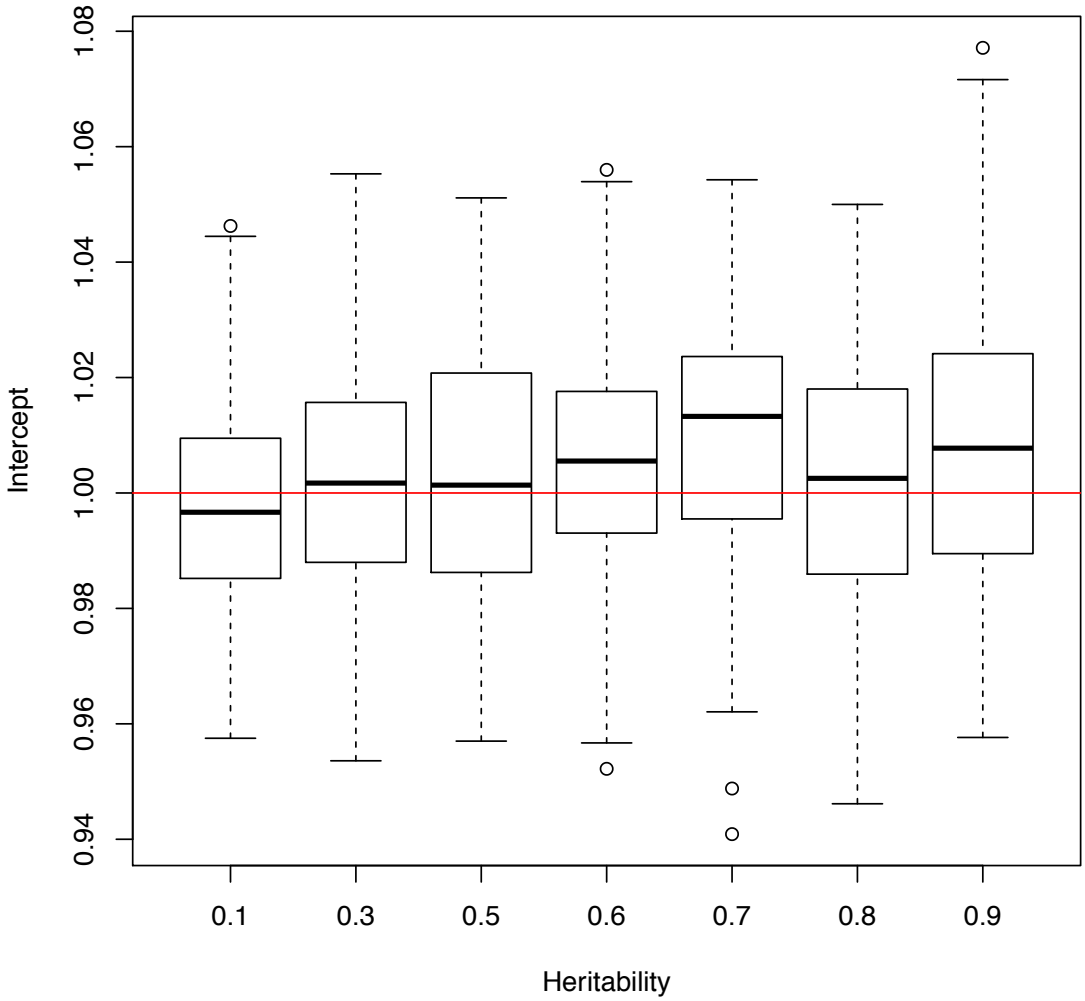
[208]Department of Clinical Medicine, University of Copenhagen, Copenhagen 2200, Denmark.

[209]Departments of Psychiatry and Human Genetics, University of Chicago, Chicago, Illinois 60637, USA.

[210]University Hospital Marqués de Valdecilla, Instituto de Formación e Investigación Marqués de Valdecilla, University of Cantabria, E-39008 Santander, Spain

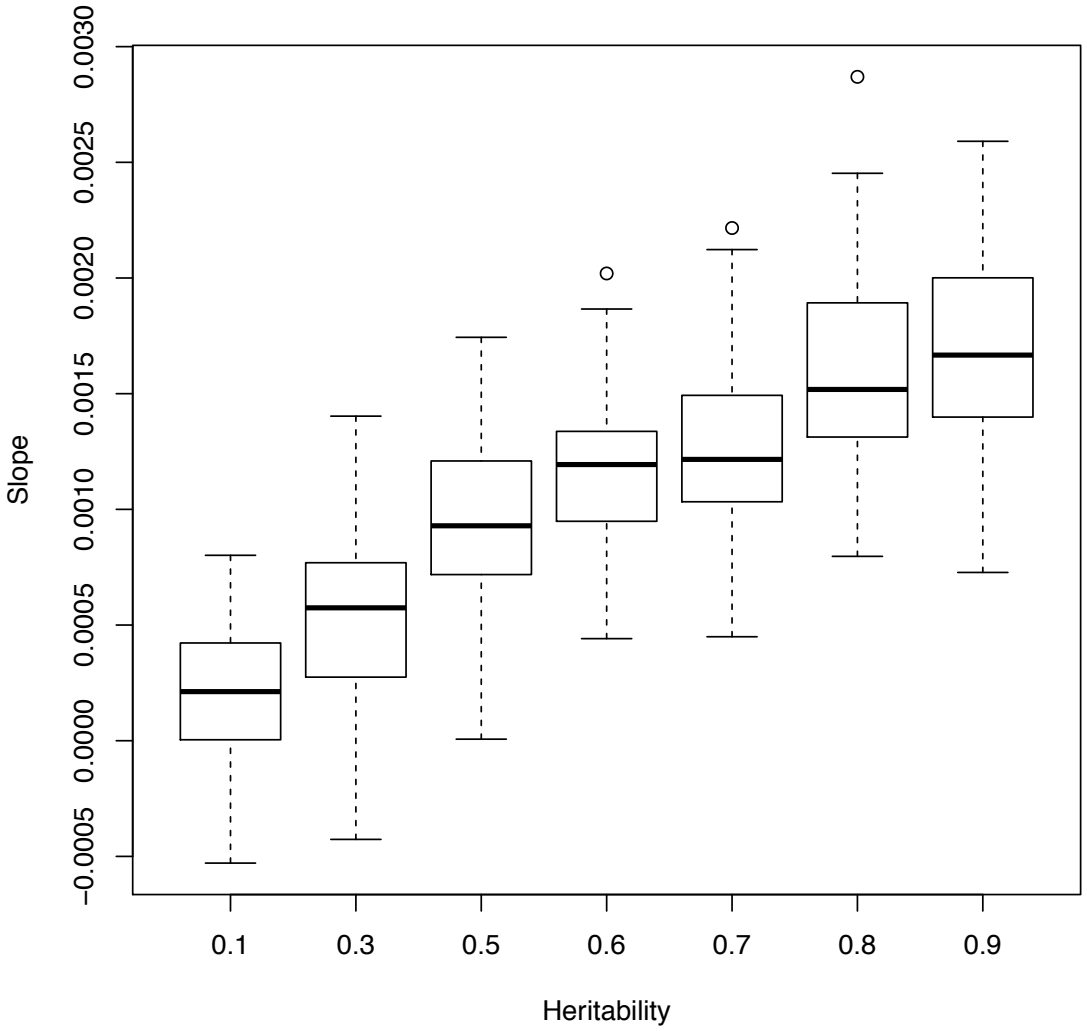
Supplemental Figures

Supplementary Figure 1: Intercepts from simulations with varying heritability



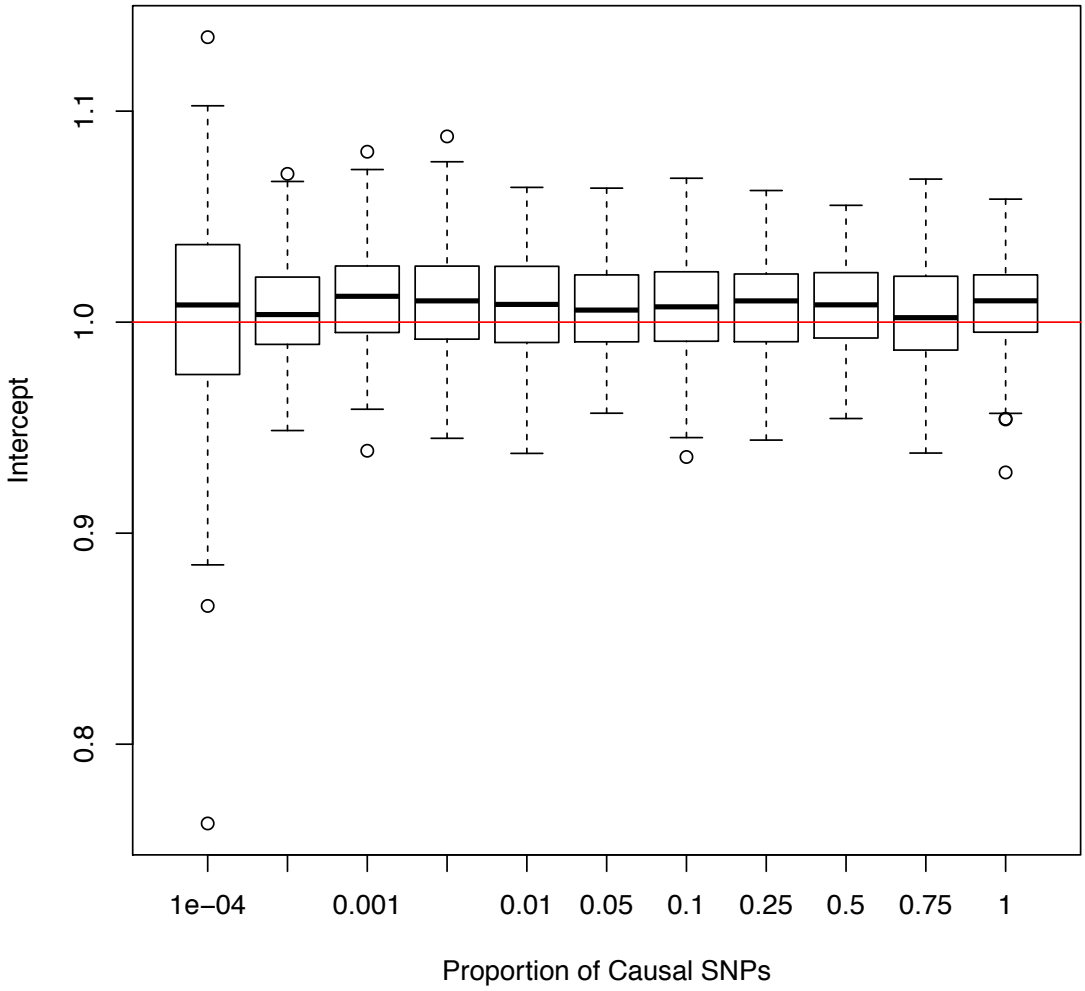
The *x*-axis displays different heritabilities specified for simulations, and the *y*-axis displays LD Score regression intercepts from 100 simulation replicates for each value of heritability. The red line shows the expected LD Score regression intercept in the absence of confounding bias. For all simulations, 1% of SNPs were causal.

Supplementary Figure 2: Slopes from simulations with varying heritability



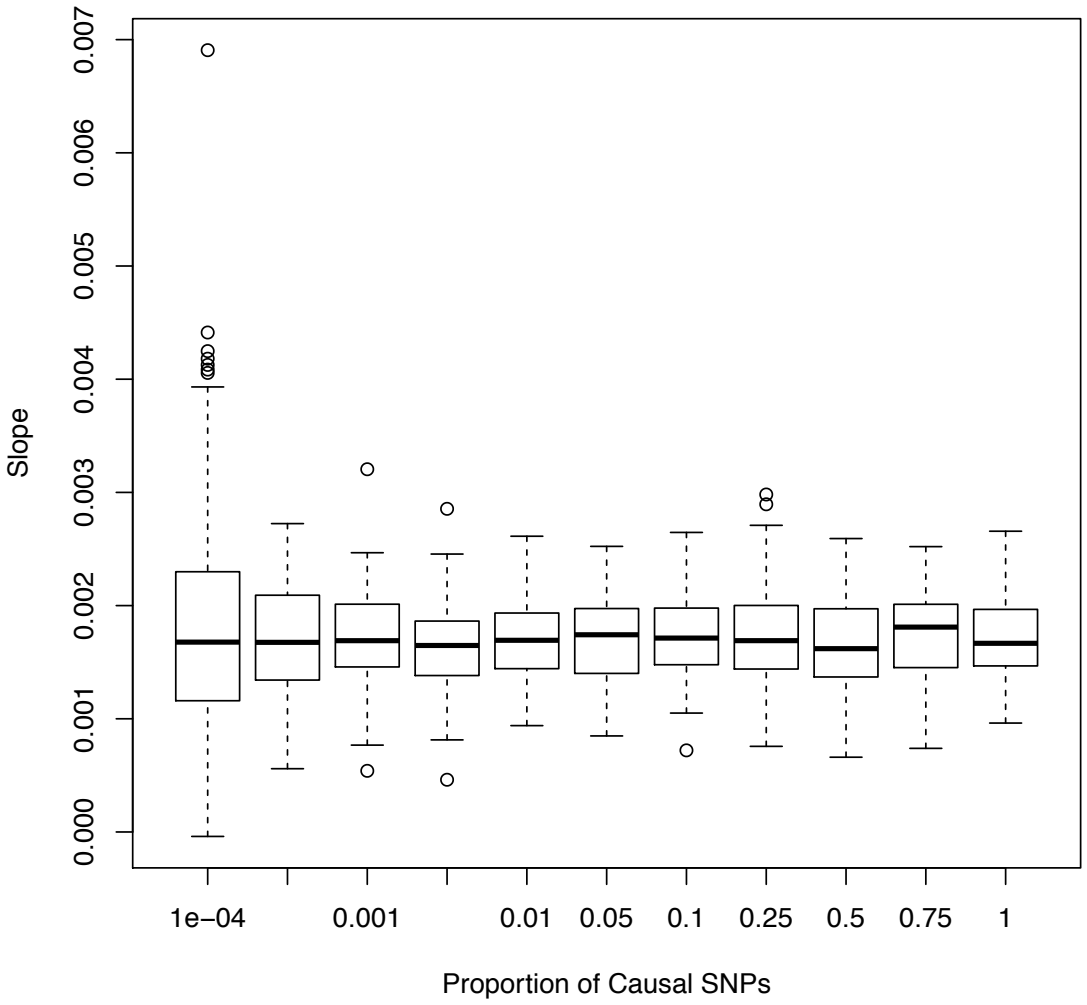
The *x*-axis displays different heritabilities specified for simulations, and the *y*-axis displays LD Score regression slopes from 100 simulation replicates for each value of heritability. For all simulations, 1% of SNPs were causal.

Supplementary Figure 3: Intercepts from simulations with various proportions of causal SNPs



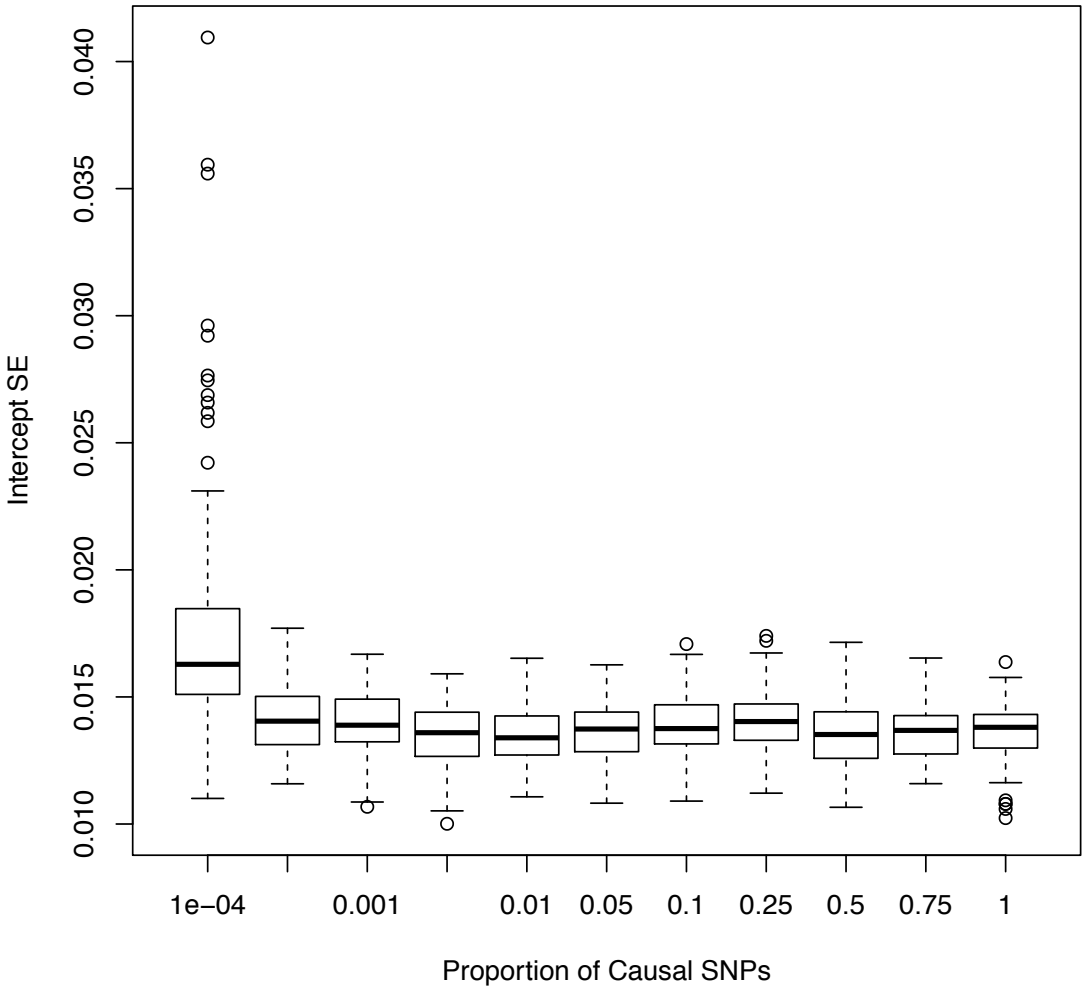
The *x*-axis displays different proportions of causal SNPs specified for simulations, and the *y*-axis displays LD Score regression intercepts from 100 simulation replicates for each value of the proportion of causal SNPs. For all simulations, the heritability was 0.9.

Supplementary Figure 4: Slopes from simulations with various proportions of causal SNPs



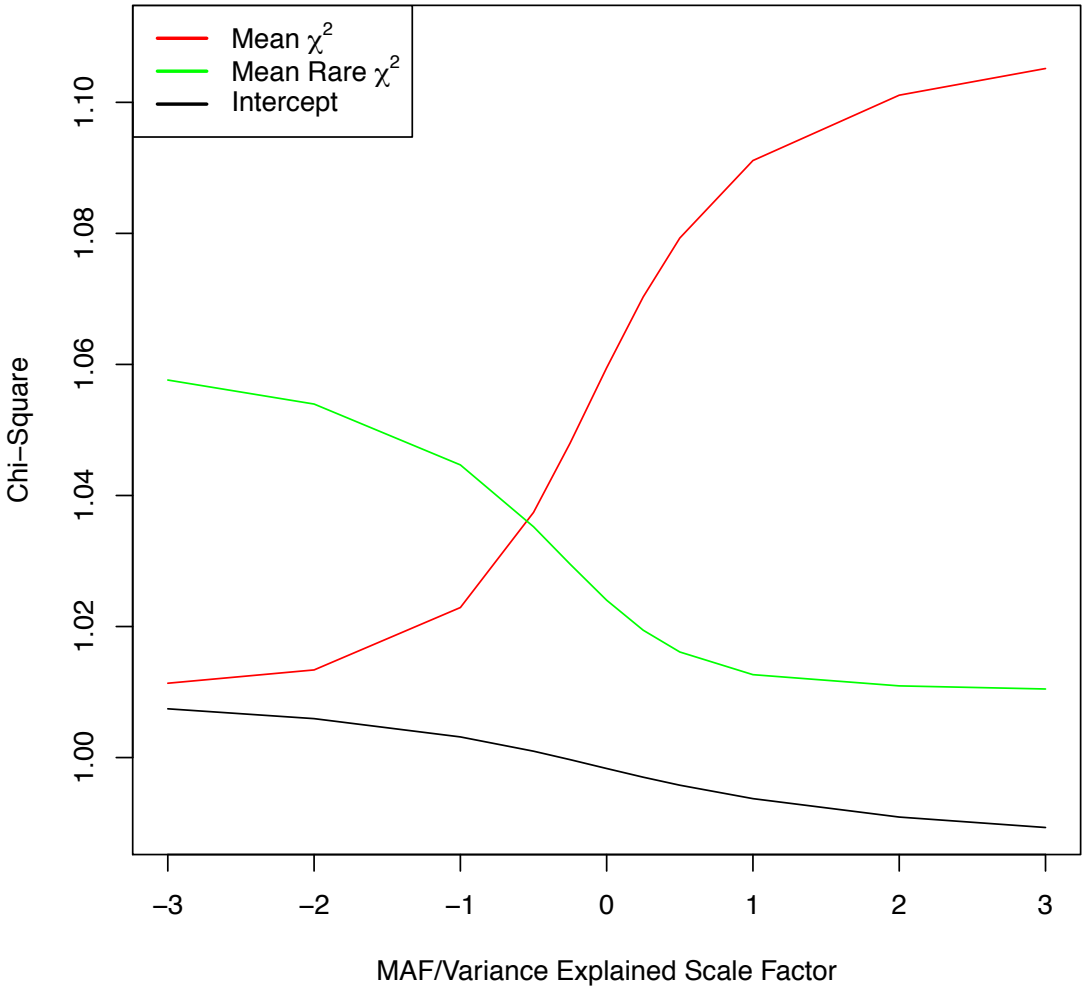
The *x*-axis displays different proportions of causal SNPs specified for simulations, and the *y*-axis displays LD Score regression slopes from 100 simulation replicates for each value of the proportion of causal SNPs. For all simulations, the heritability was 0.9.

Supplementary Figure 5: Estimated standard error from simulations with various proportions of causal SNPs



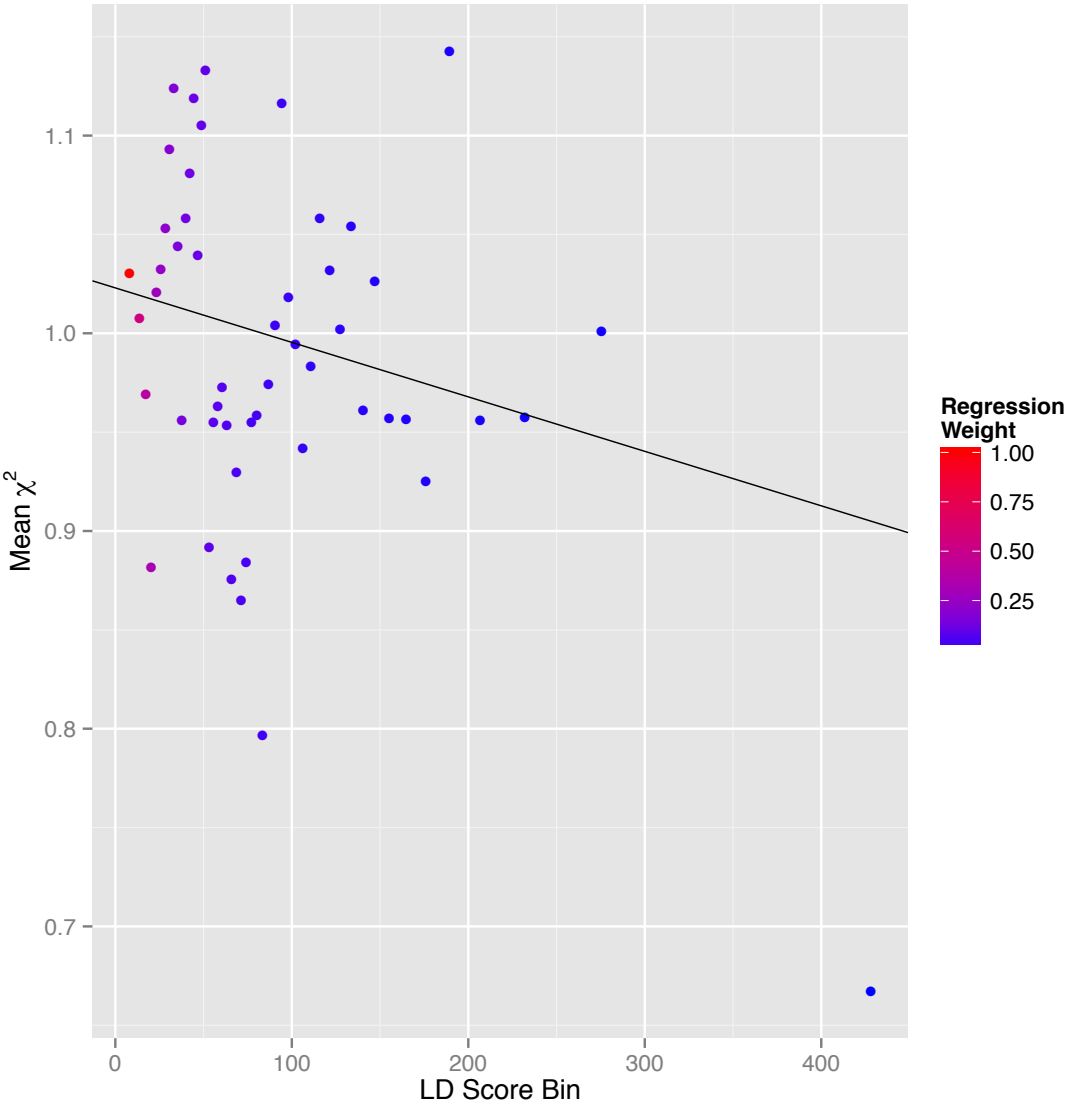
The *x*-axis displays different proportions of causal SNPs specified for simulations, and the *y*-axis displays block jackknife estimates of the standard error of the intercept from each of 100 simulation replicates for each proportion of causal SNPs. For all simulations, the heritability was 0.9.

Supplementary Figure 6: Simulations with frequency-dependent architecture



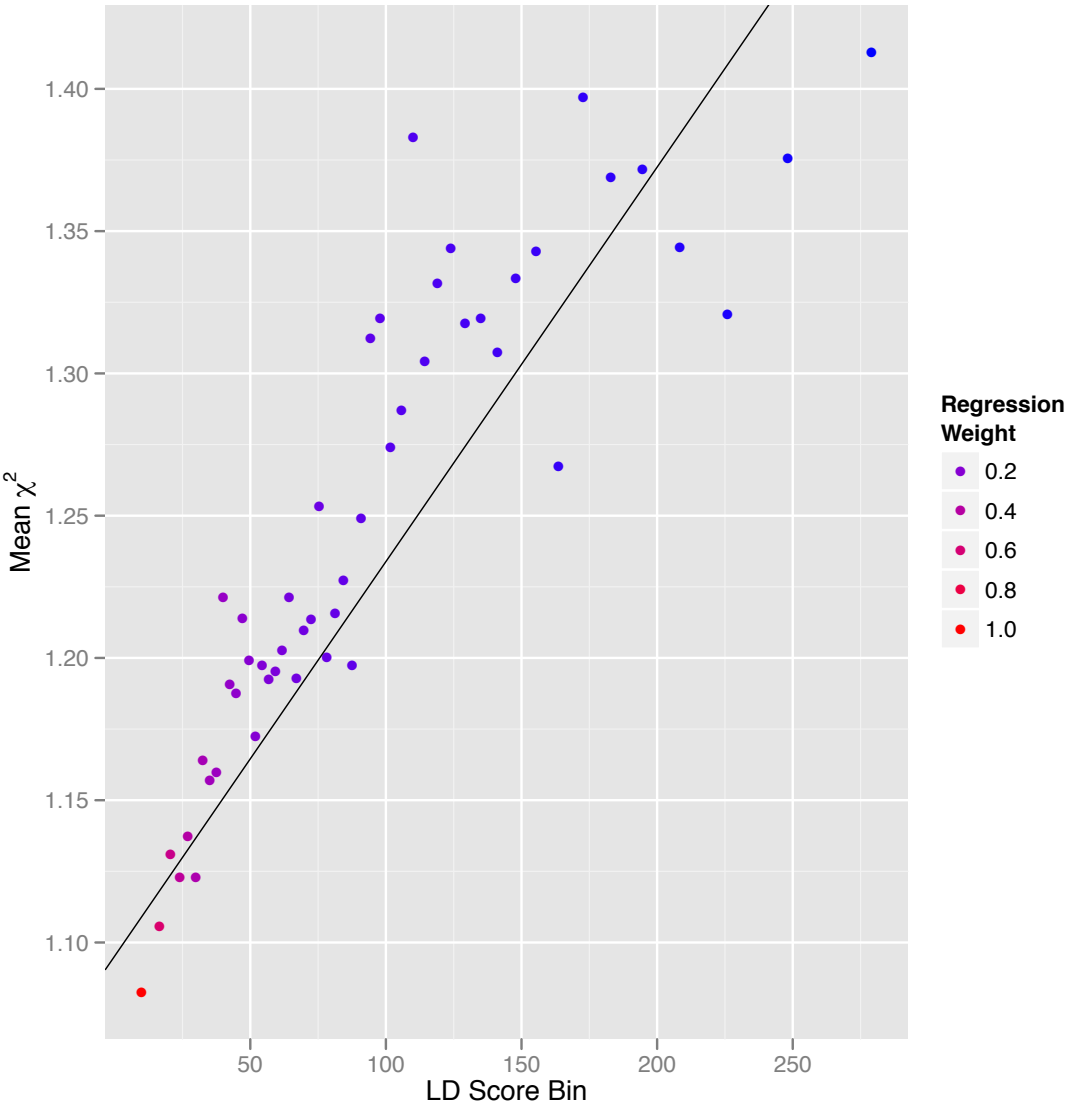
The *x*-axis describes the simulate relationship between minor allele frequency and effect size. Precisely, per-normalized genotype effects for 10,000 causal variants were drawn from $N(0, (p(1-p))^x)$, where *p* is MAF and *x* is the *x*-coordinate. To prevent singleton and doubleton variants from having extreme effects for large negative values of *x*, we drew the effect sizes for variants with MAF < 1% from $N(0,0.0099^x)$. The red line is the mean χ^2 among the common HapMap 3[4] variants retained for LD Score regression. The green line is the mean χ^2 among variants with MAF < 1%. The black line is the LD Score regression intercept. Each data point is the average across 10 simulation replicates with randomly chosen effects. Our model holds when *x=0*, which corresponds to moderate negative selection on the phenotype in question, similar to a typical disease phenotype. *x=1* is an appropriate model for a selectively neutral phenotype. Values of *x* outside the range [0,1] represent extreme genetic architectures.

Supplementary Figure 7: Simulation where all causal variants are rare



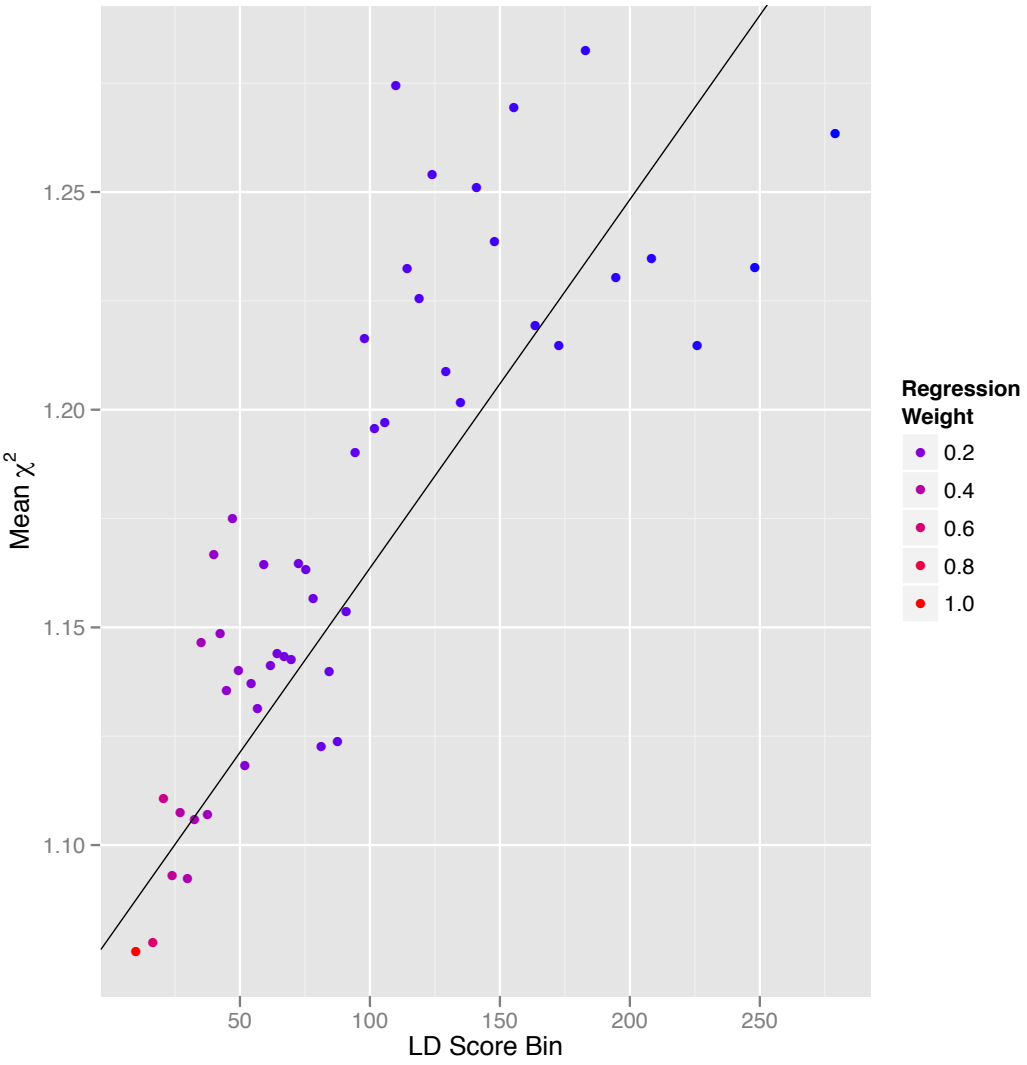
LD Score regression plot for a simulation with 1000 Swedish samples and ~700,000 SNPs on chromosome 1 where all causal variants had MAF < 1%. Each point represents an LD Score quantile, where the *x*-coordinate of the point is the mean LD Score of variants in that quantile and the *y*-coordinate is the mean χ^2 of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line. The slope of the LD Score regression line is -3.2E-4, which is statistically significantly less than zero (block jackknife p=0.013).

Supplementary Figure 8a: LD Score plot for IBD



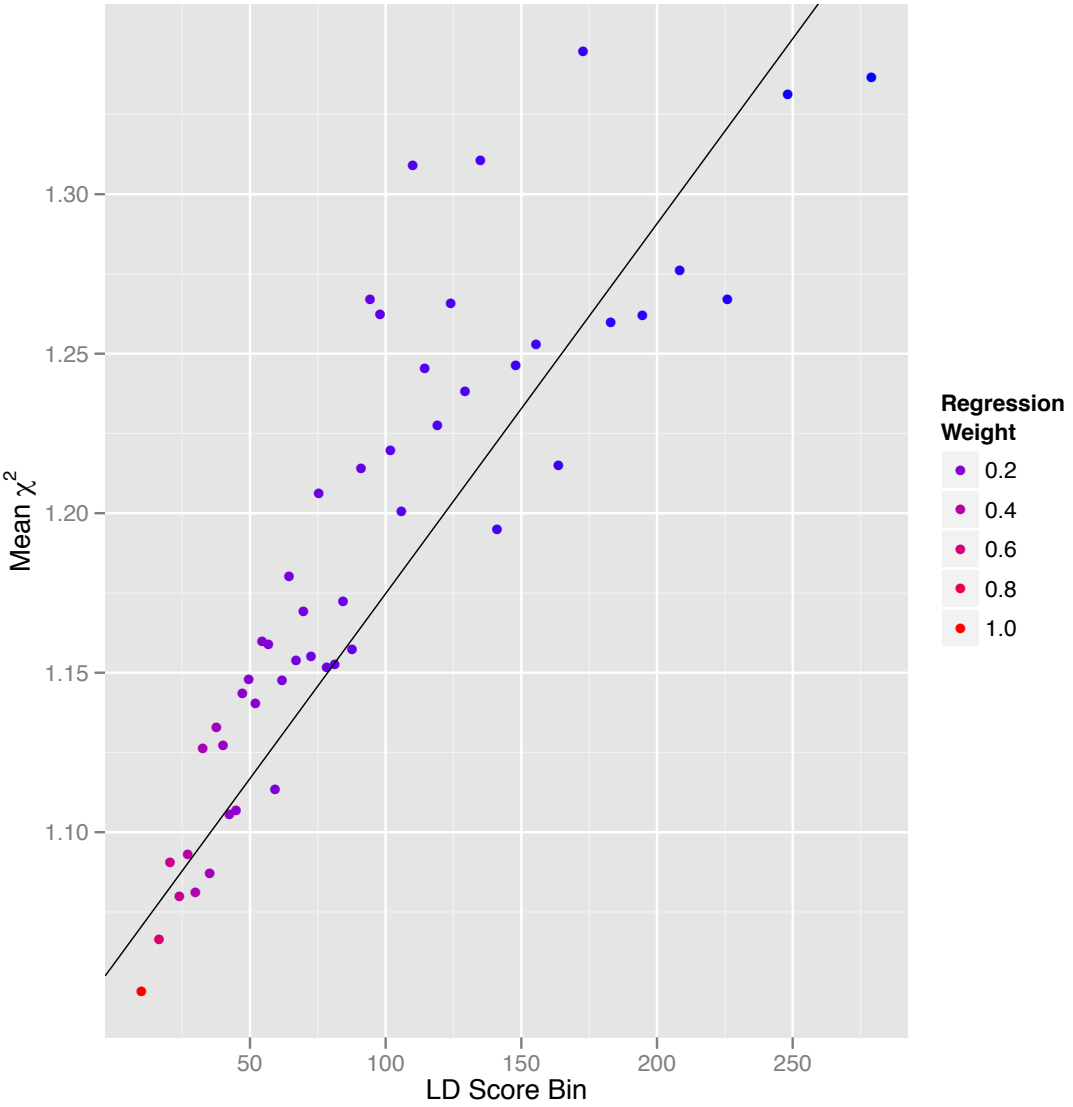
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8b: LD Score plot for Ulcerative Colitis



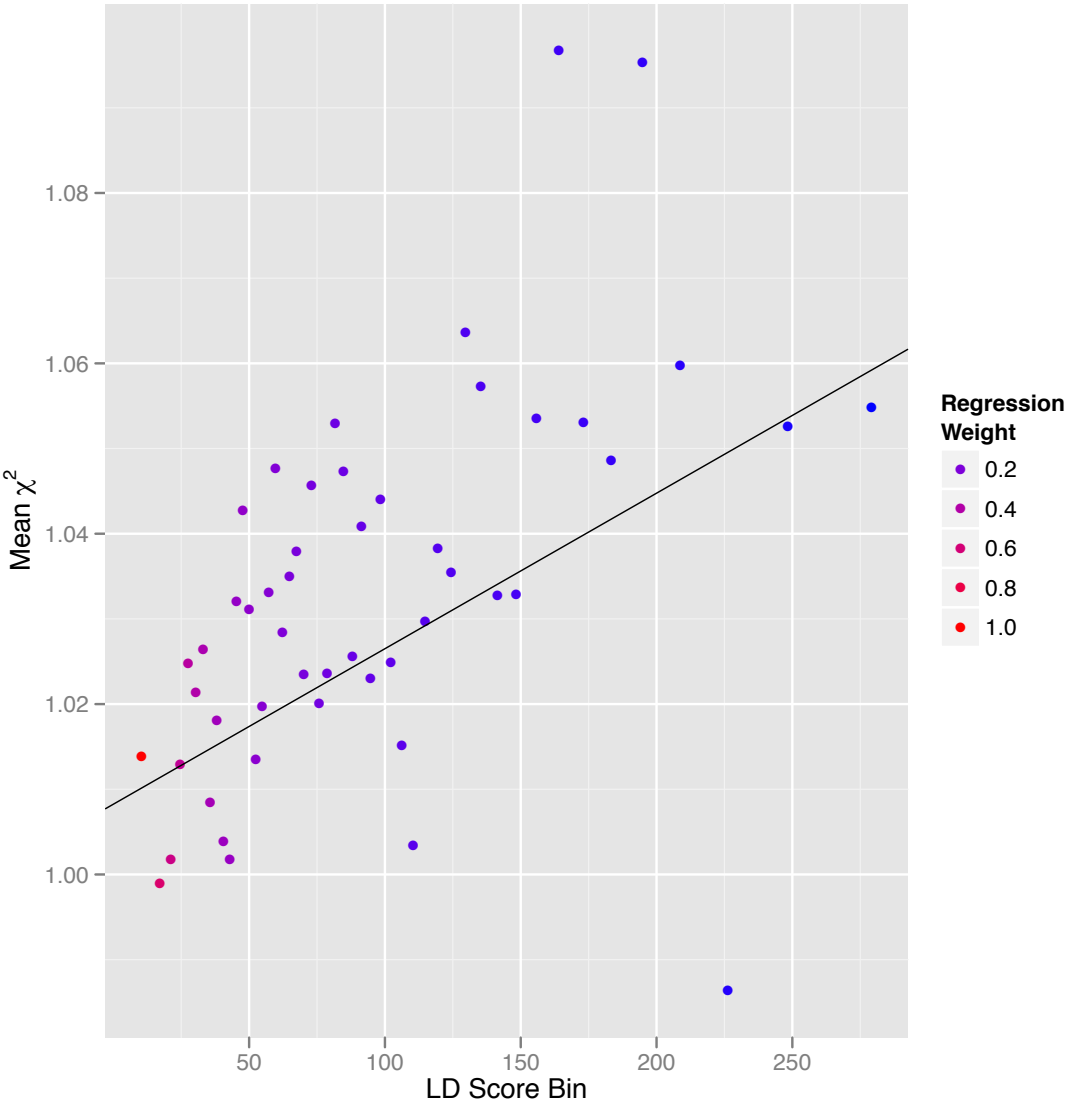
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8c: LD Score plot for Crohn's Disease



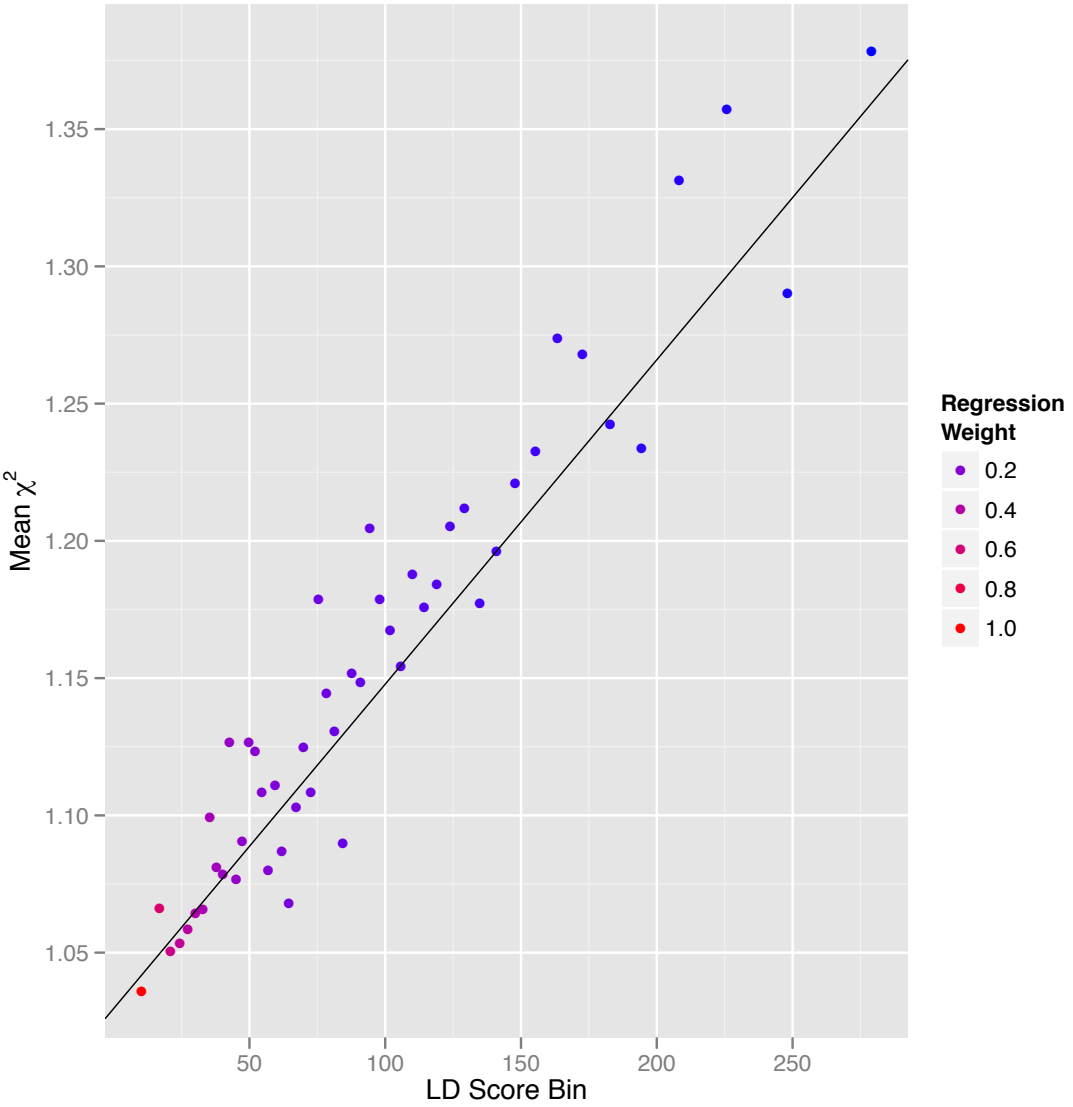
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8d: LD Score plot for ADHD



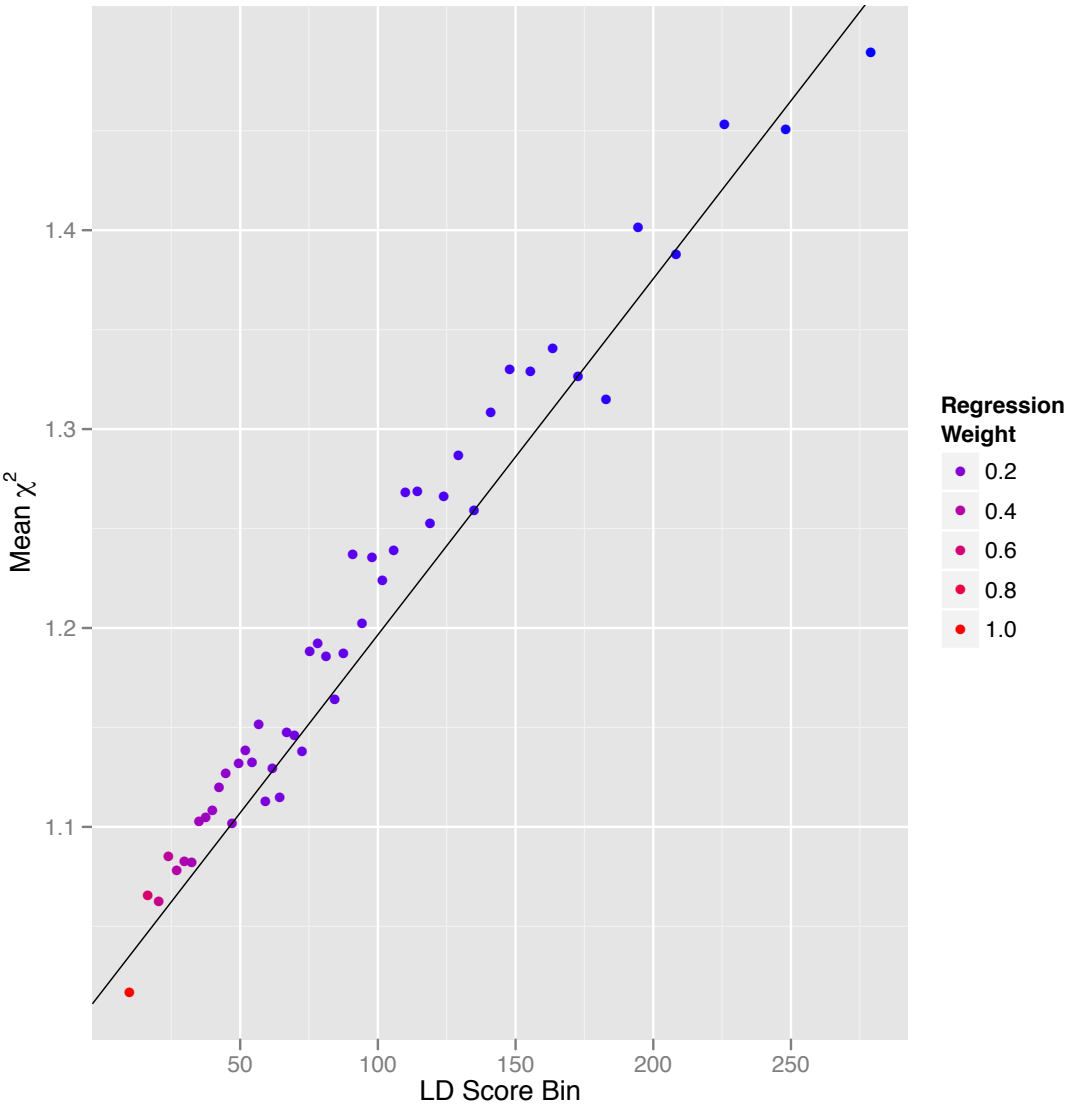
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8e: LD Score plot for Bipolar Disorder



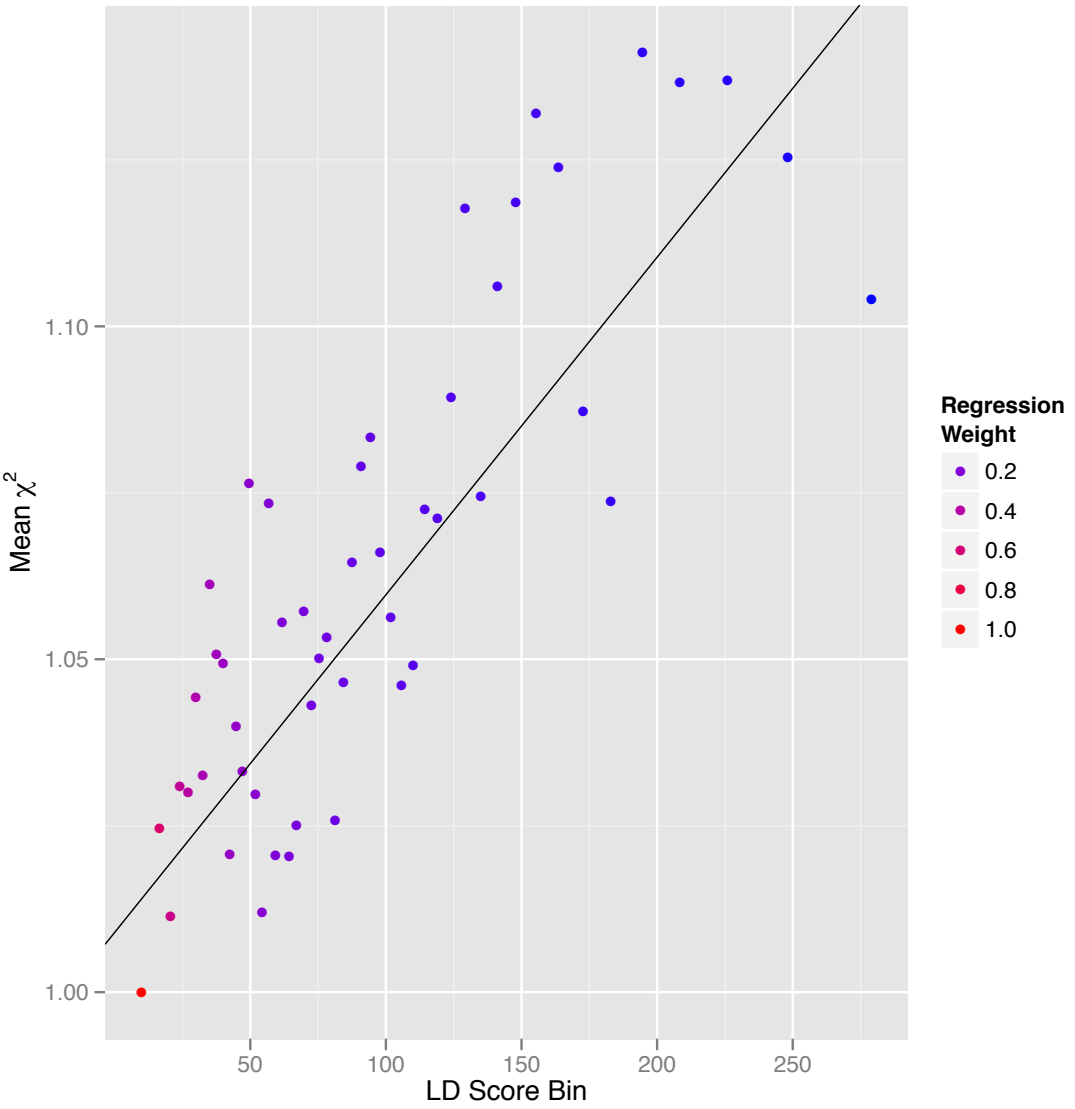
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8f: LD Score plot for PGC Cross-Disorder Analysis



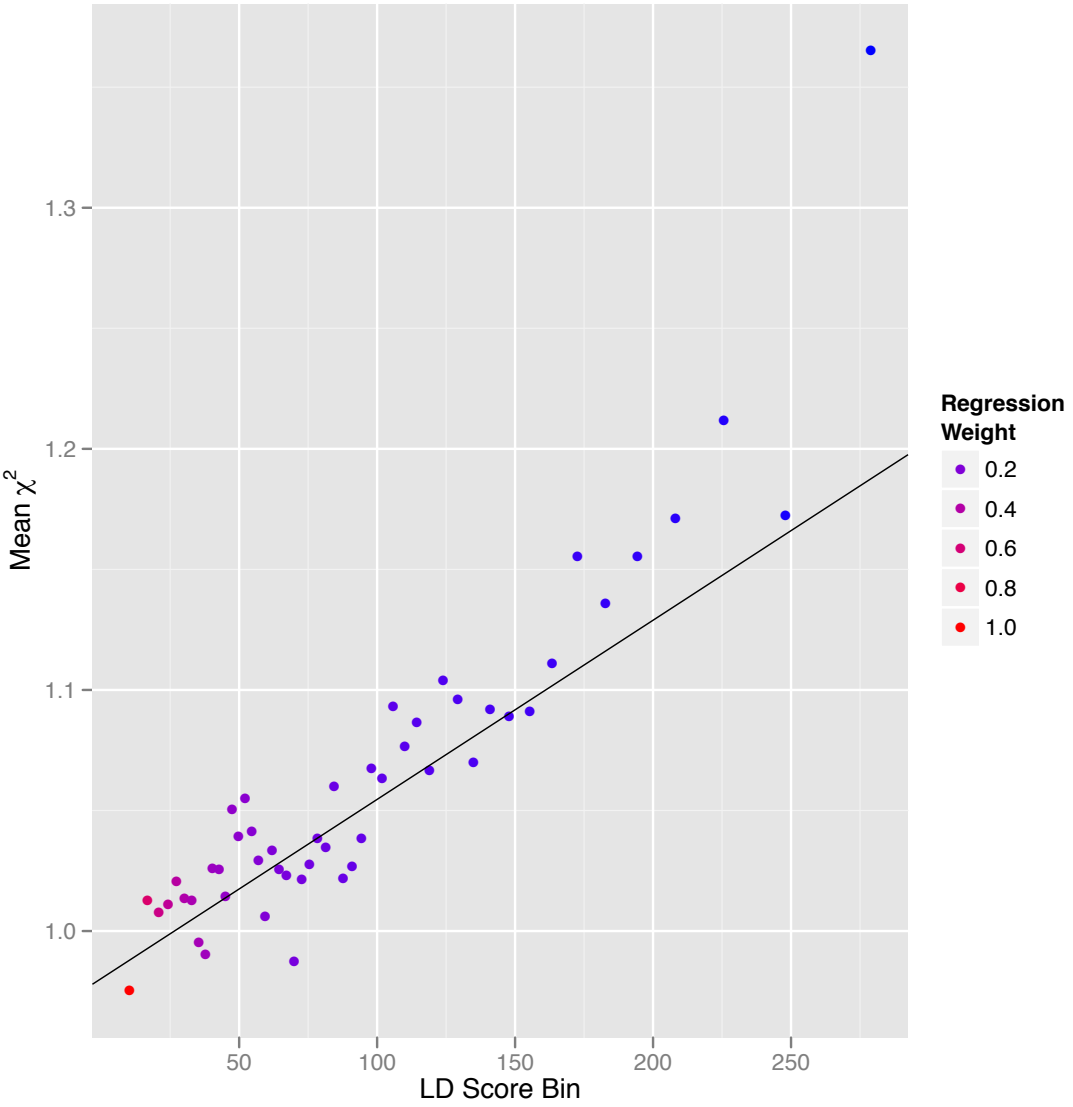
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8g: LD Score plot for Major Depression



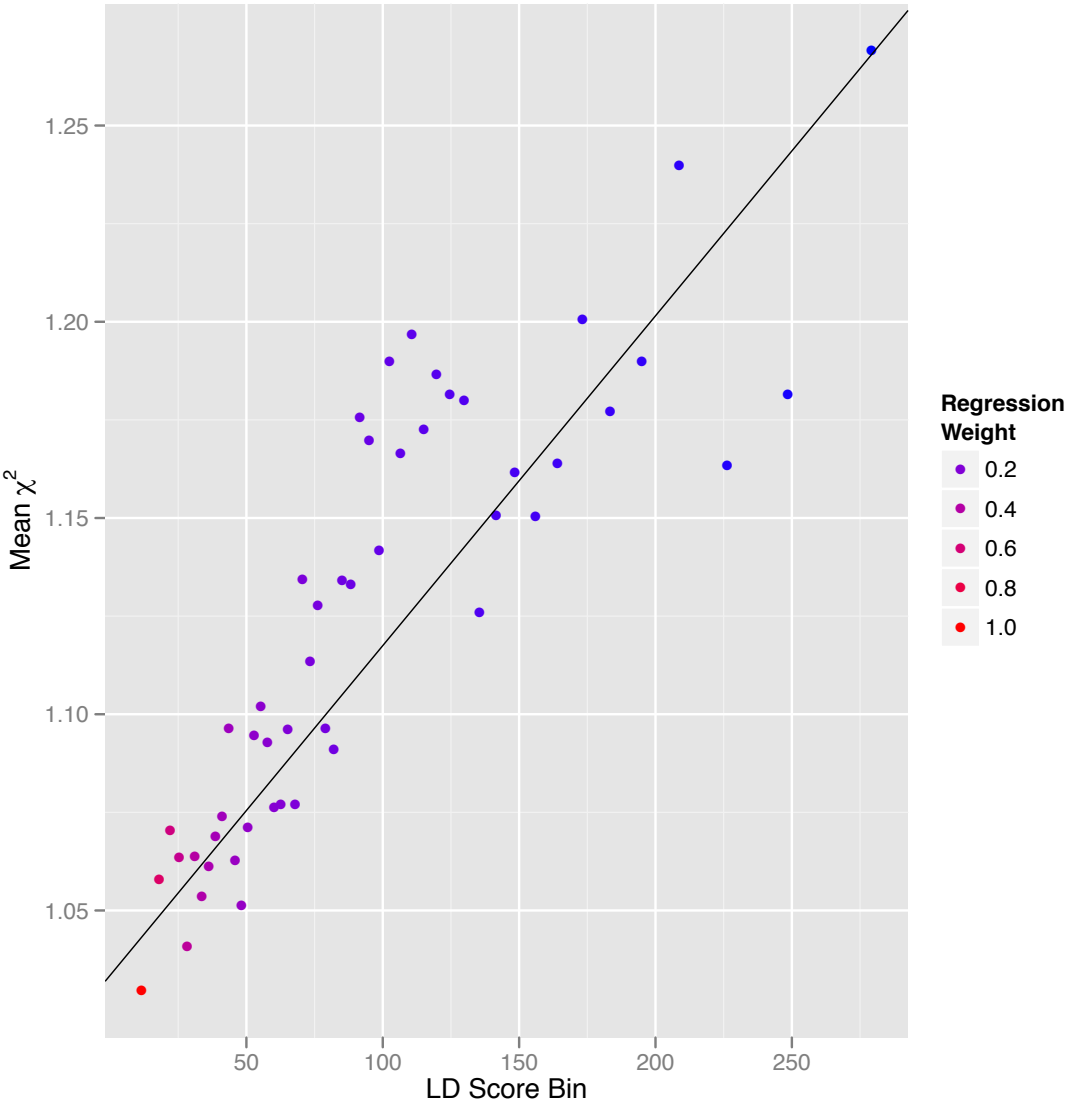
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8h: LD Score plot for Rheumatoid Arthritis



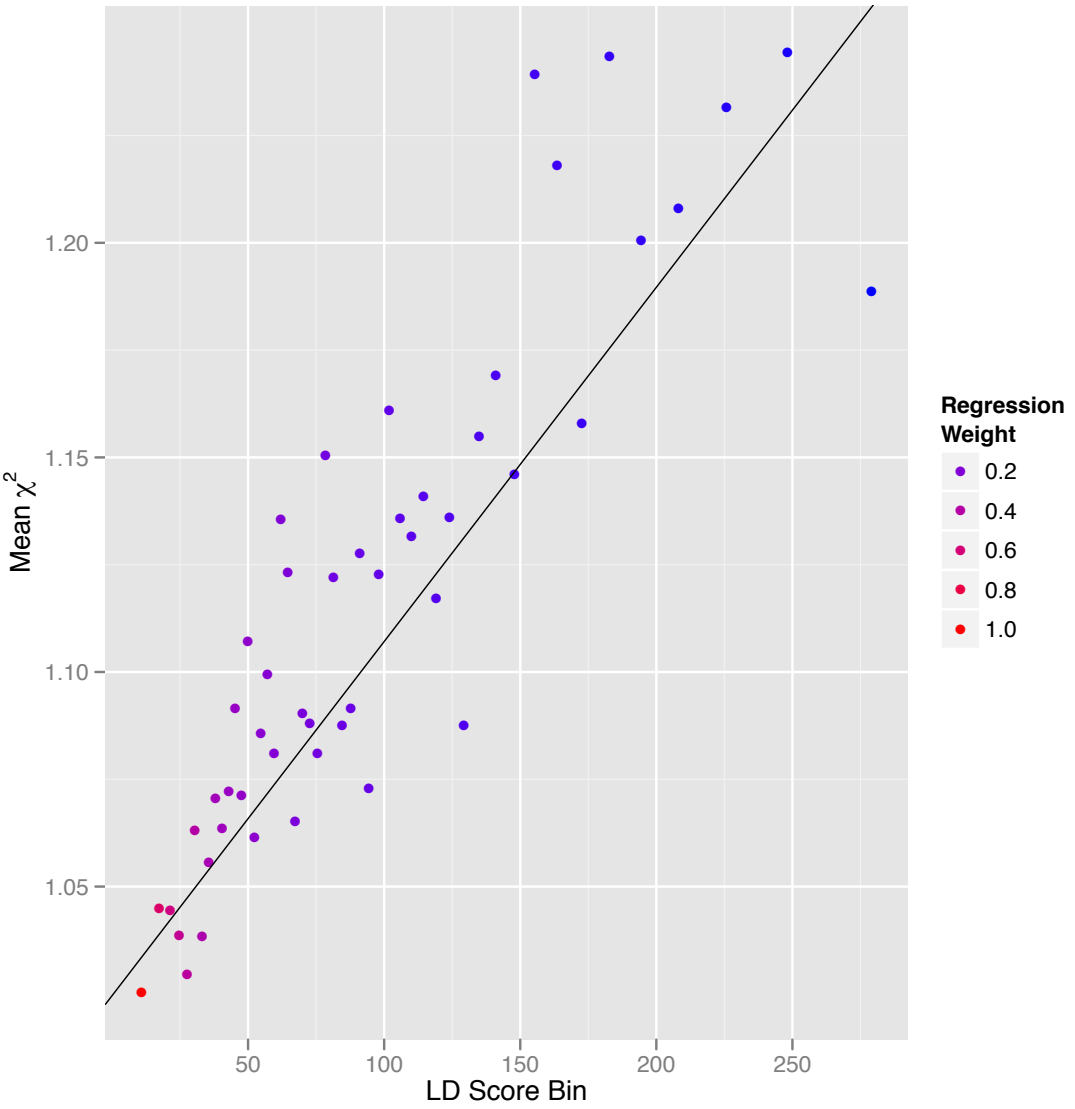
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8i: LD Score plot for Coronary Artery Disease



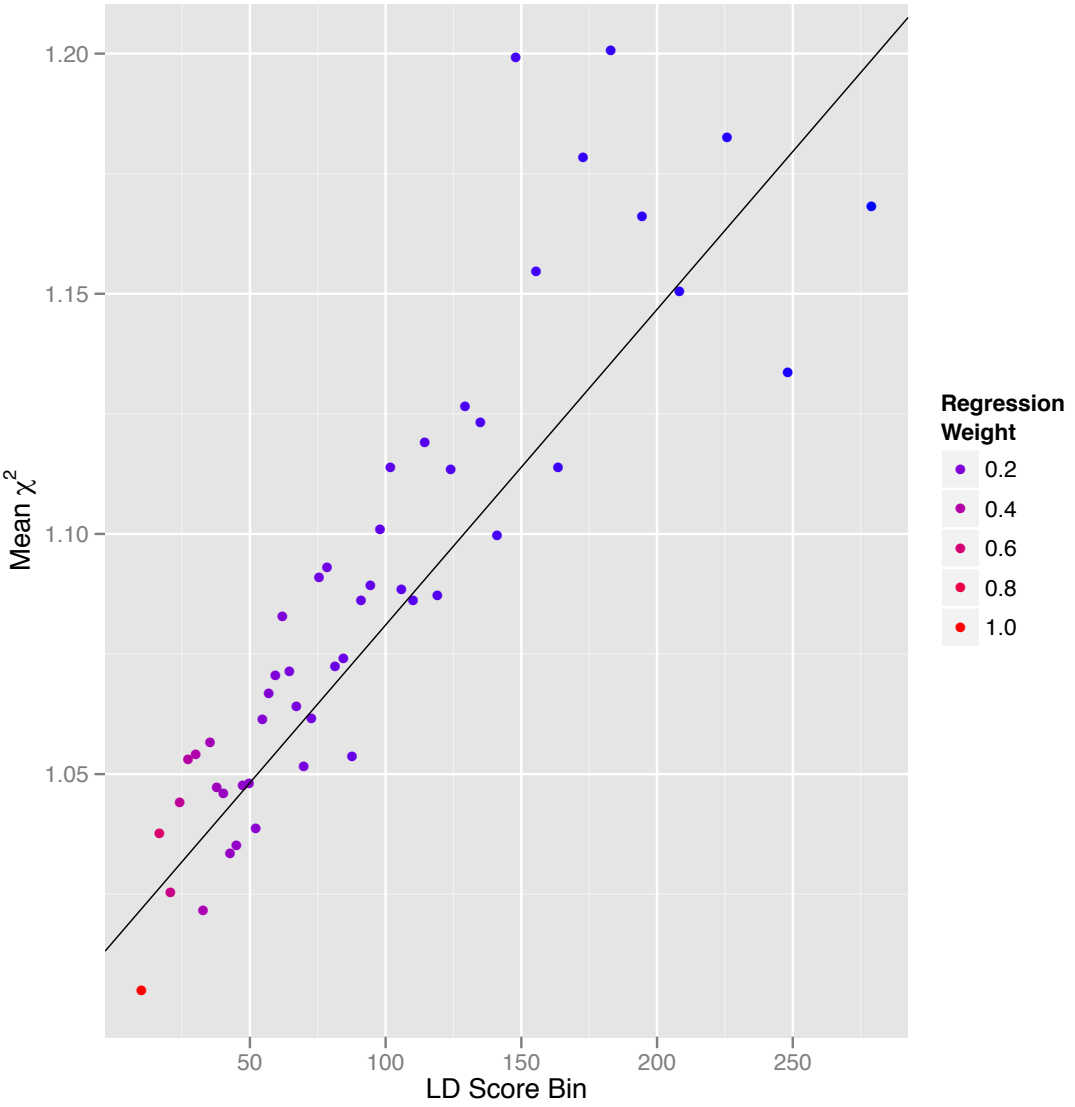
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8j: LD Score plot for Type-2 Diabetes



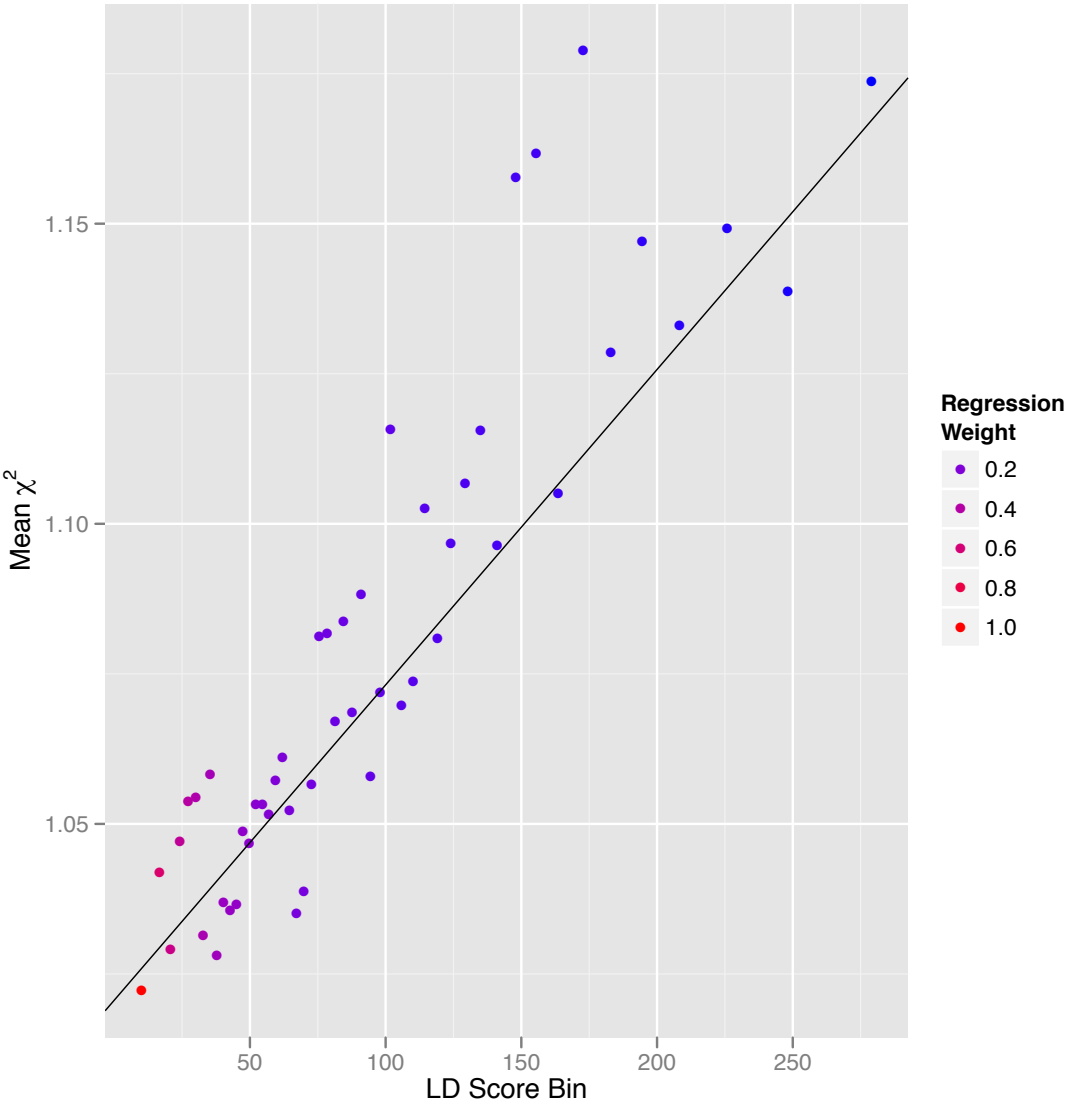
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8k: LD Score plot for BMI-Adjusted Fasting Insulin



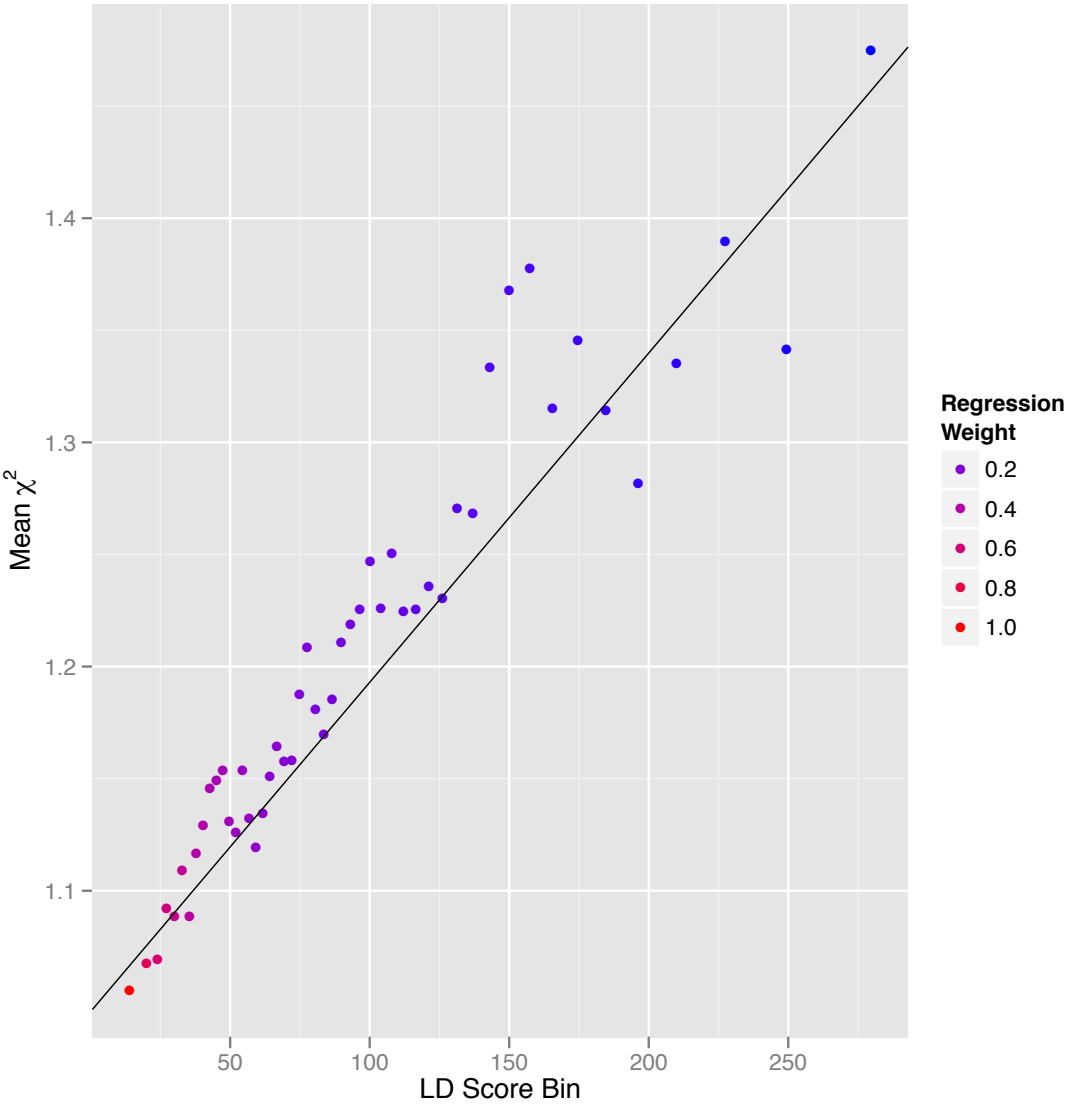
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8l: LD Score plot for Fasting Insulin



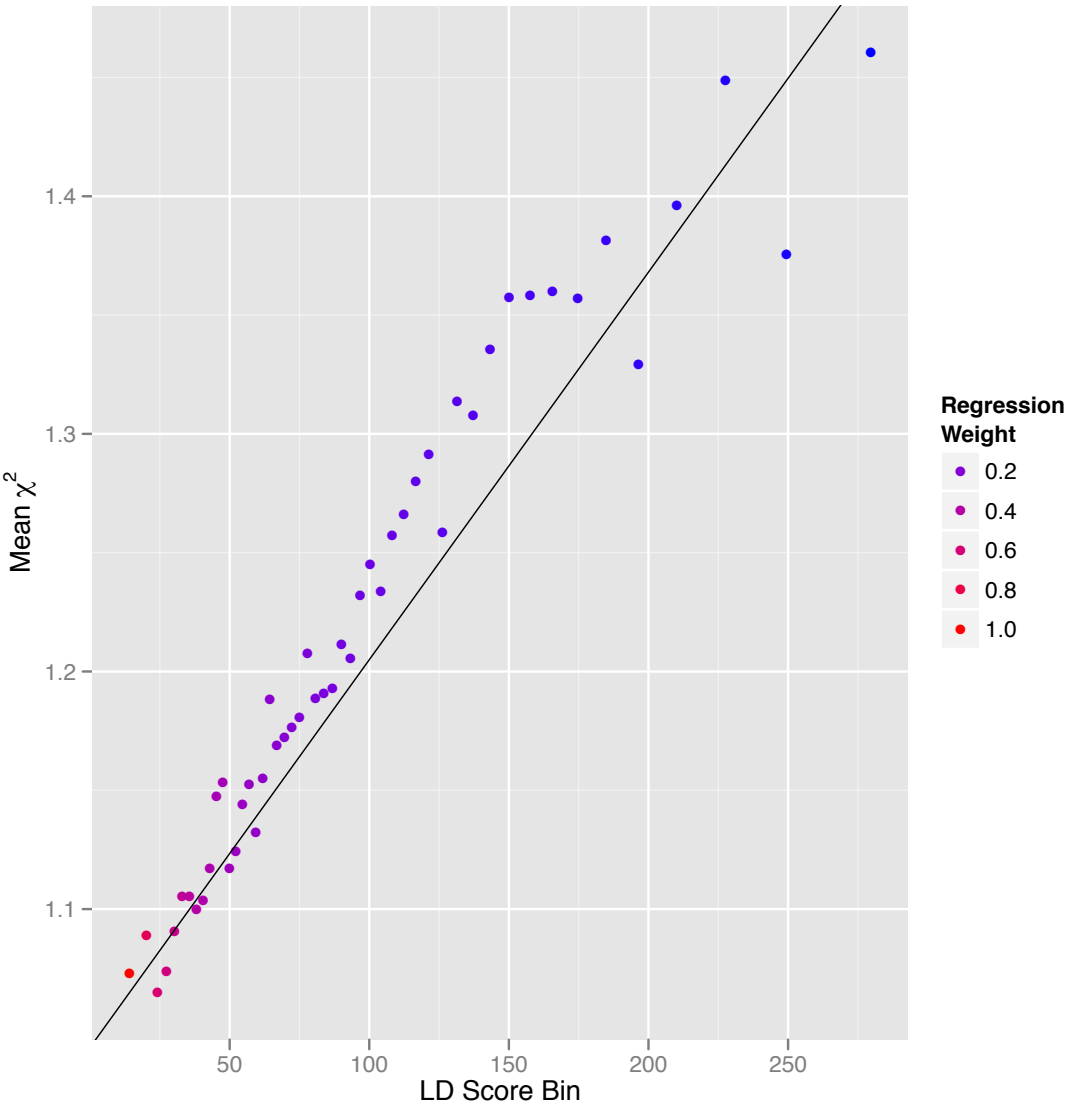
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8m: LD Score plot for College



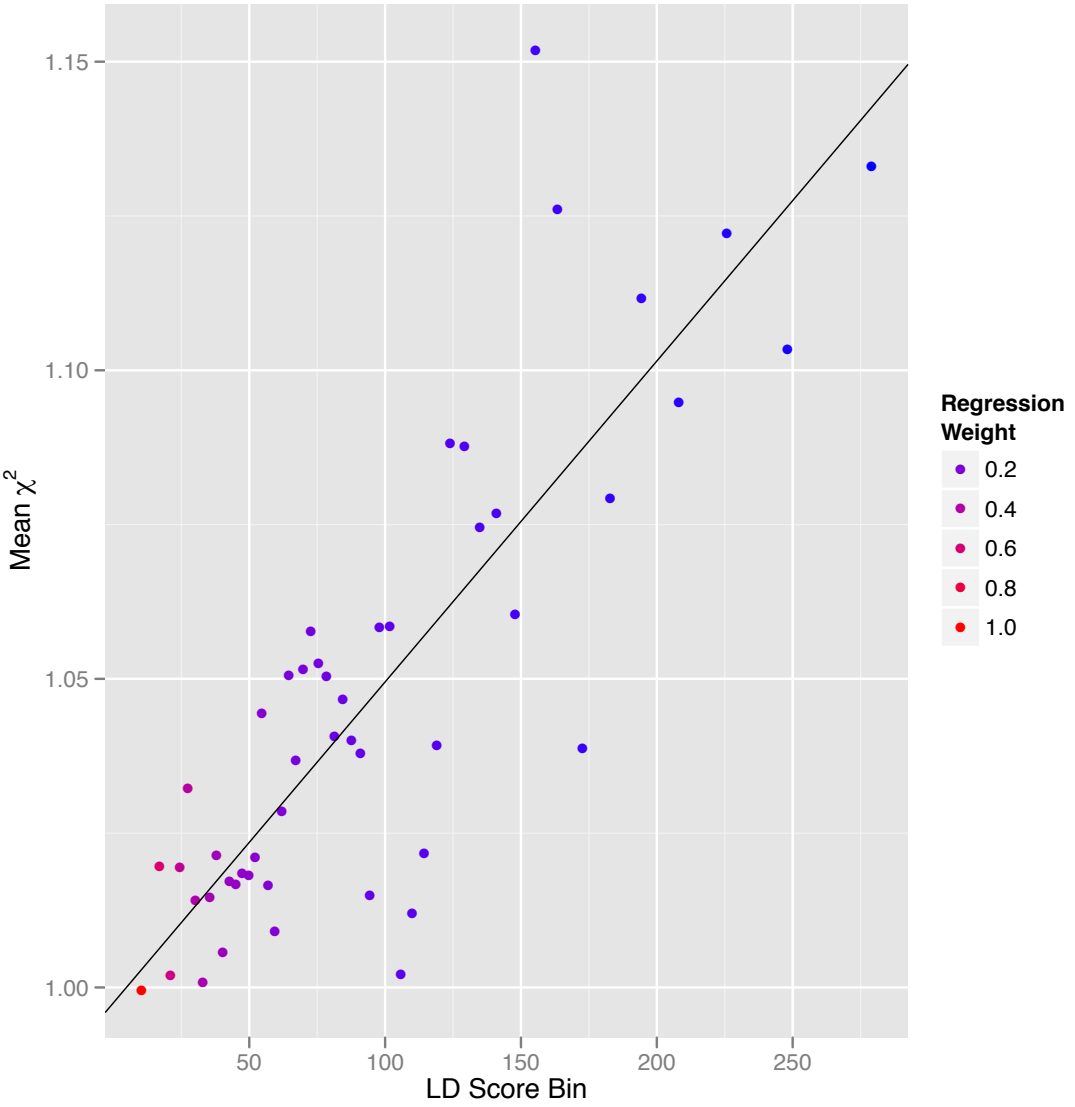
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8n: LD Score plot for Years of Education



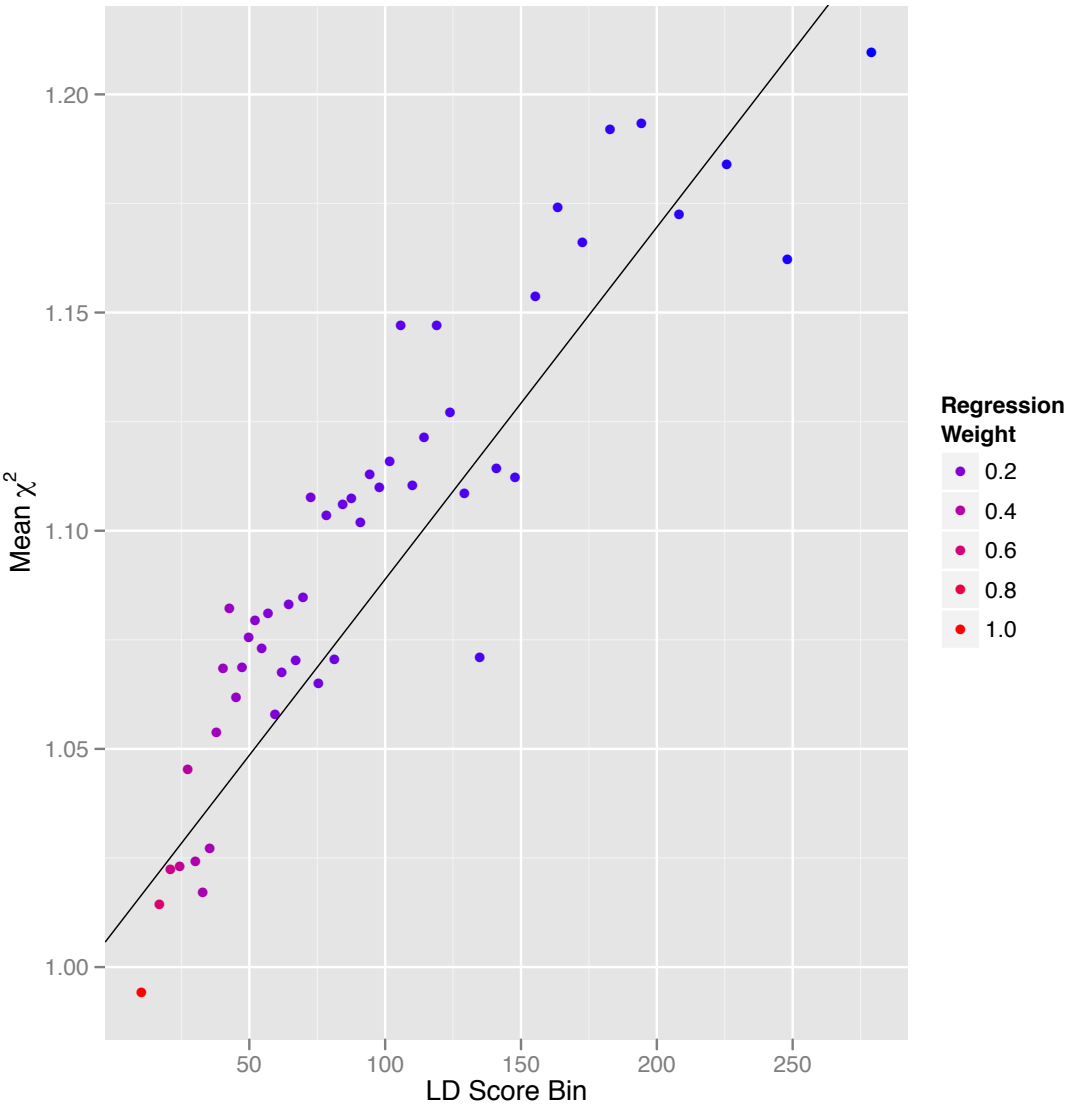
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 80: LD Score plot for Cigarettes Per Day



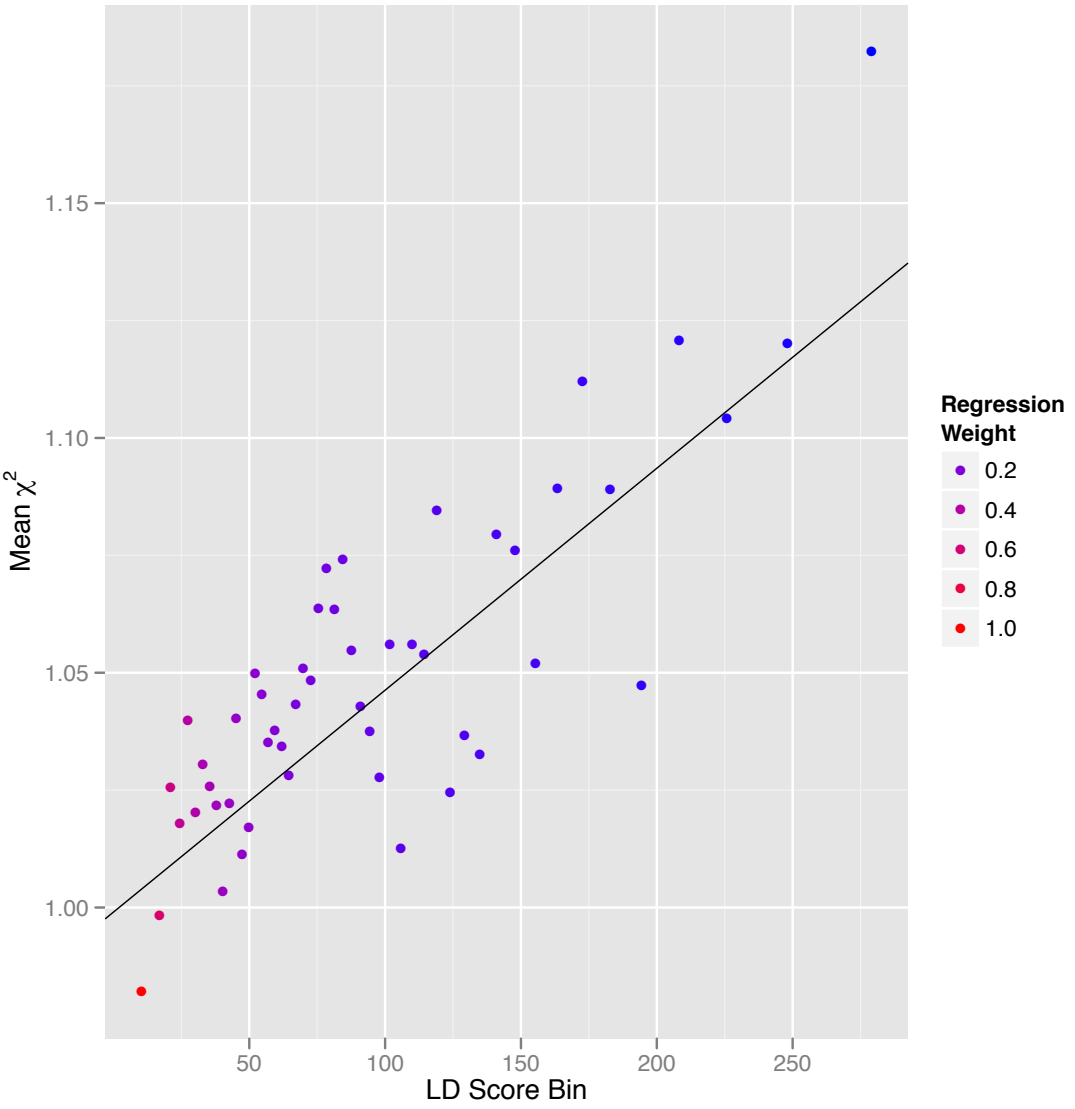
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8p: LD Score plot for Ever Smoked?



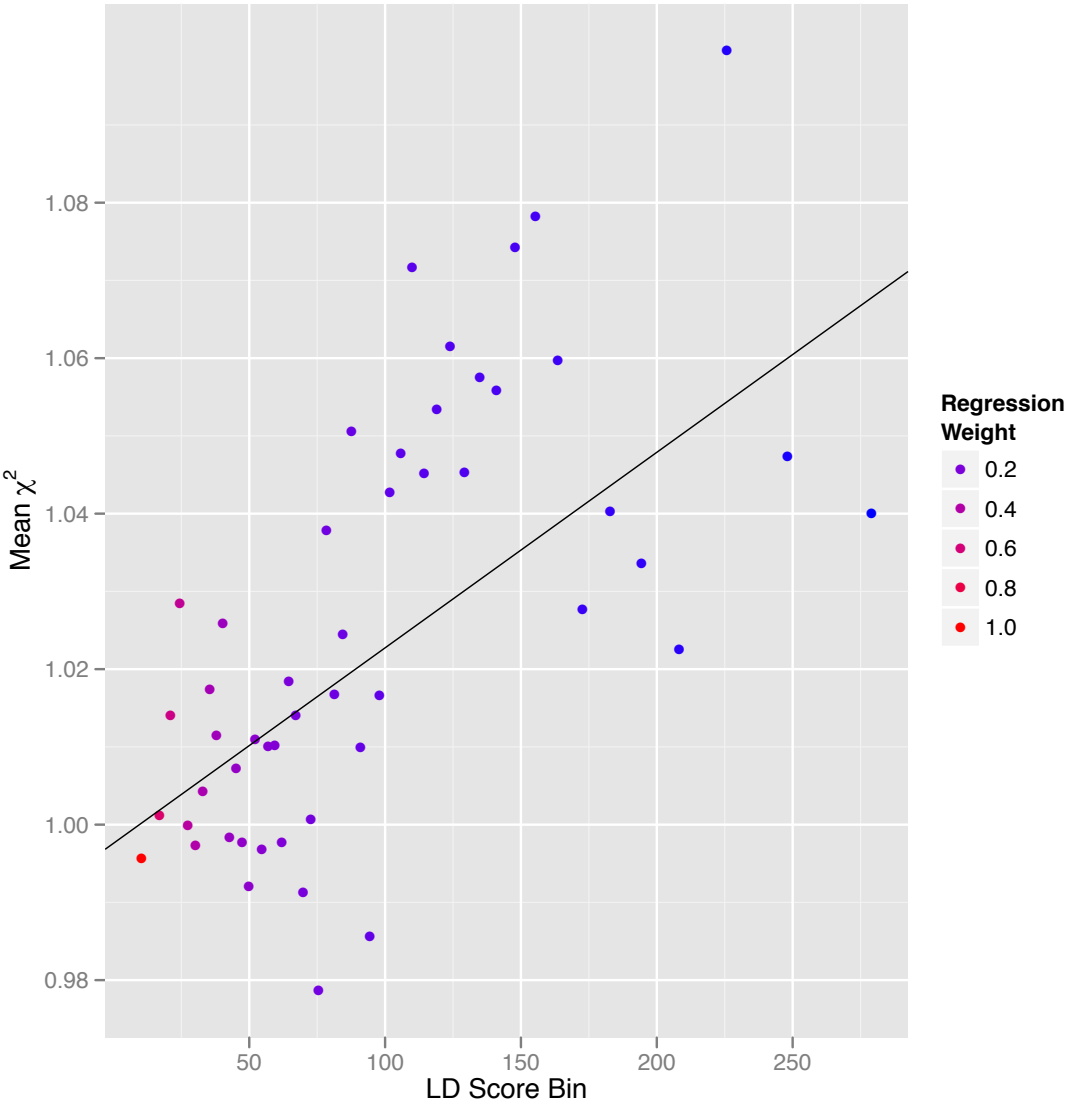
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8q: LD Score plot for Former Smoker?



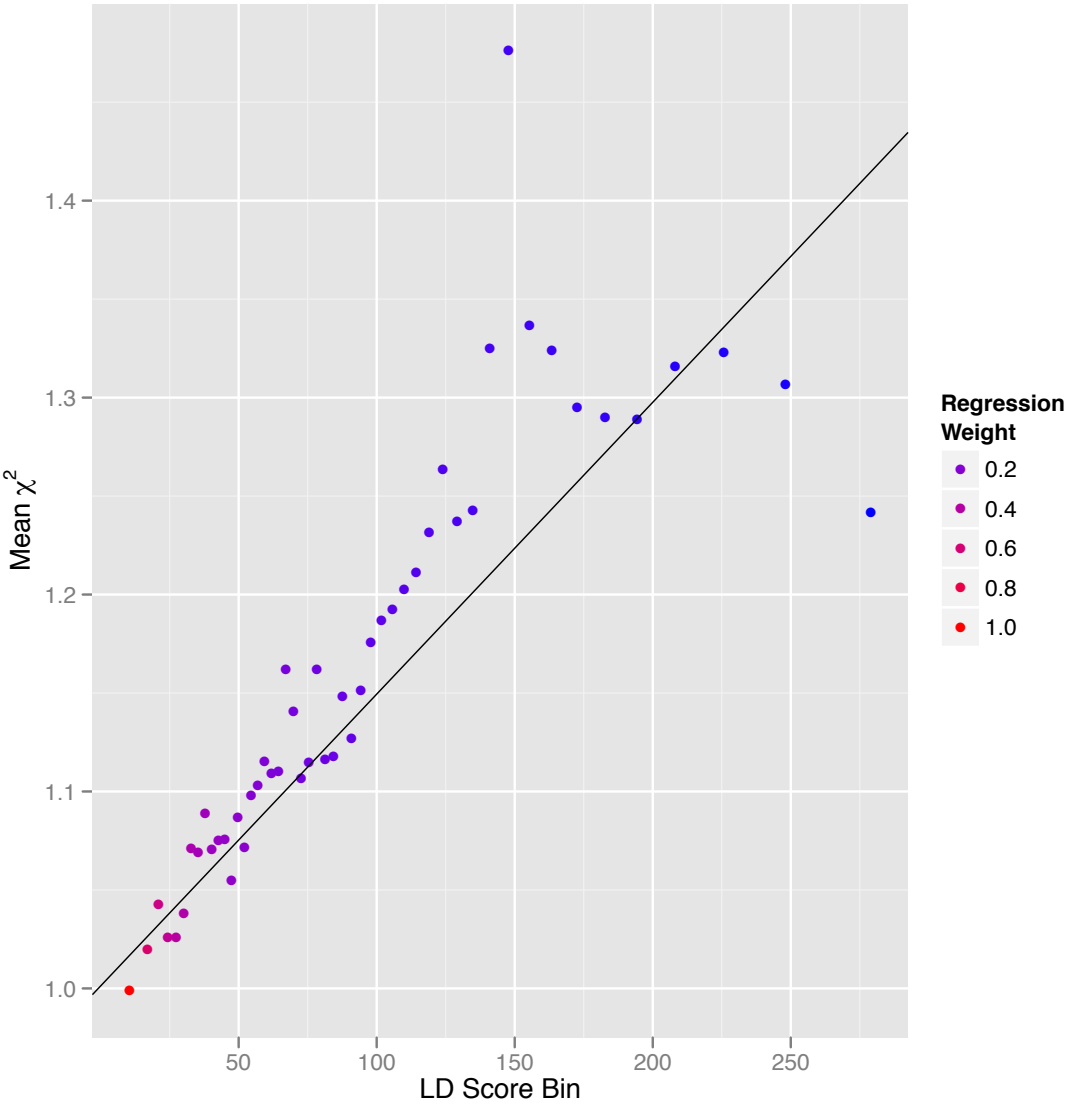
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8r: LD Score plot for Smoking Age of Onset



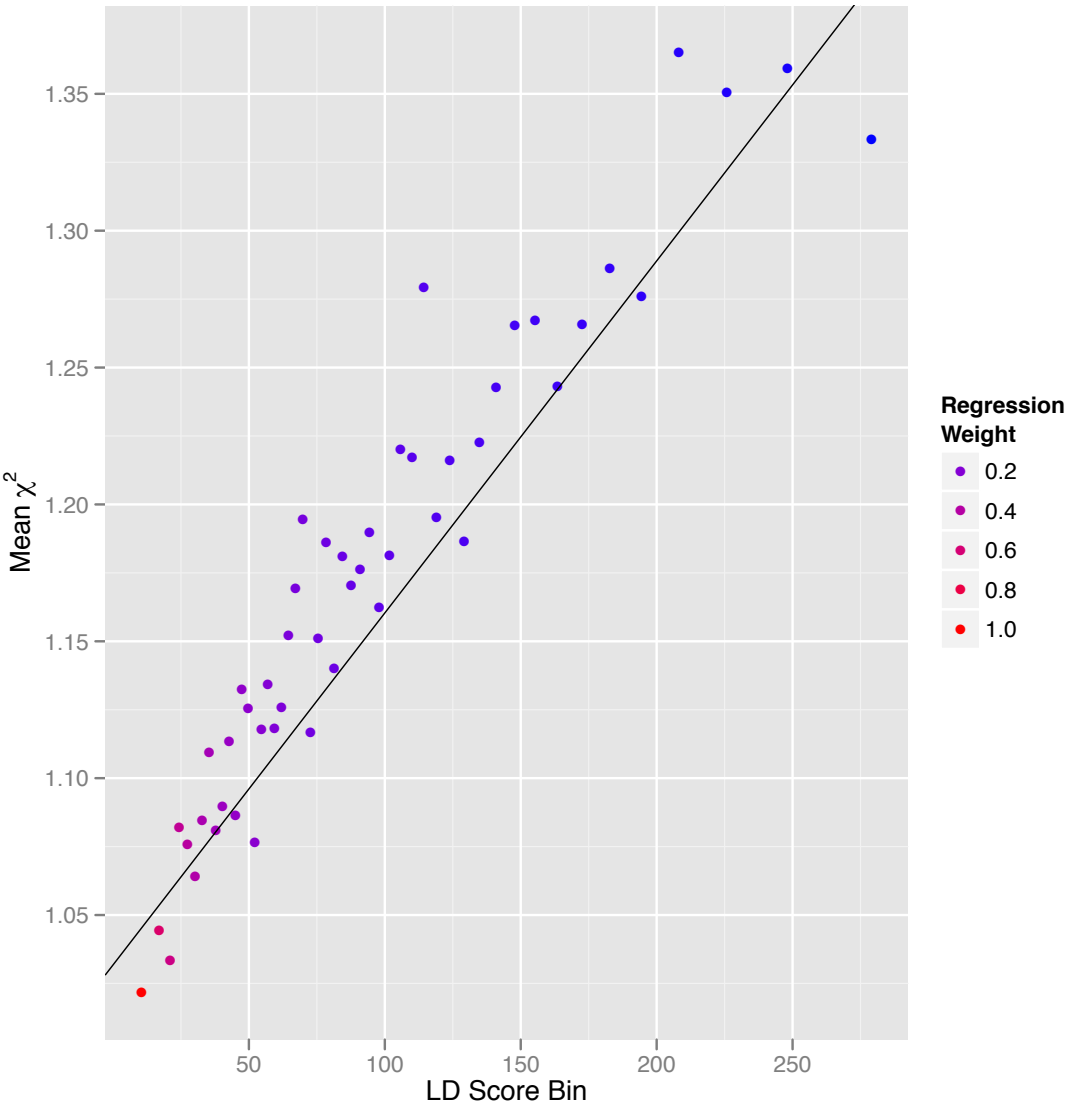
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8s: LD Score plot for Femoral Neck Bone Mineral Density



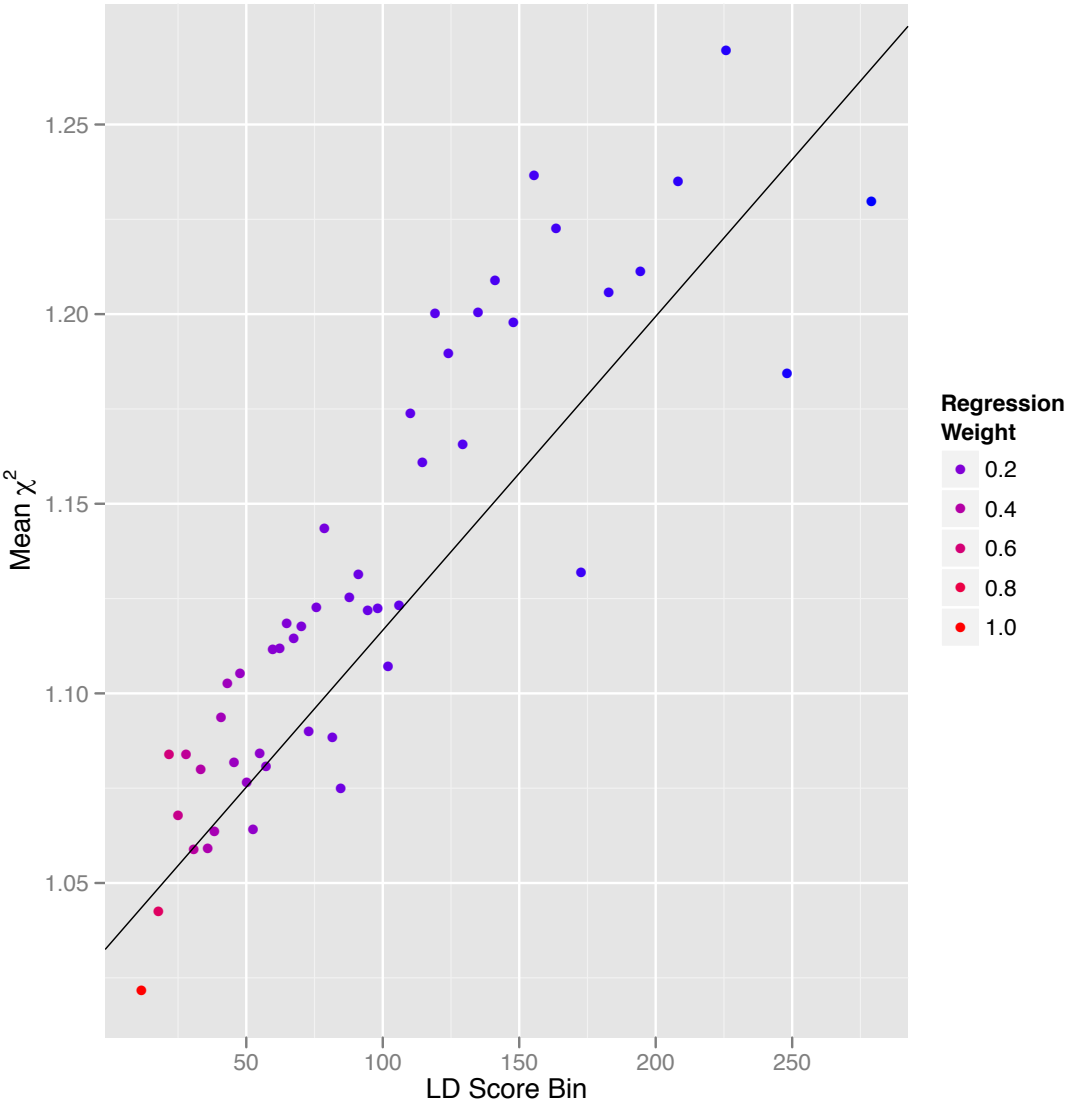
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8t: LD Score plot for Lumbar Spine Bone Mineral Density



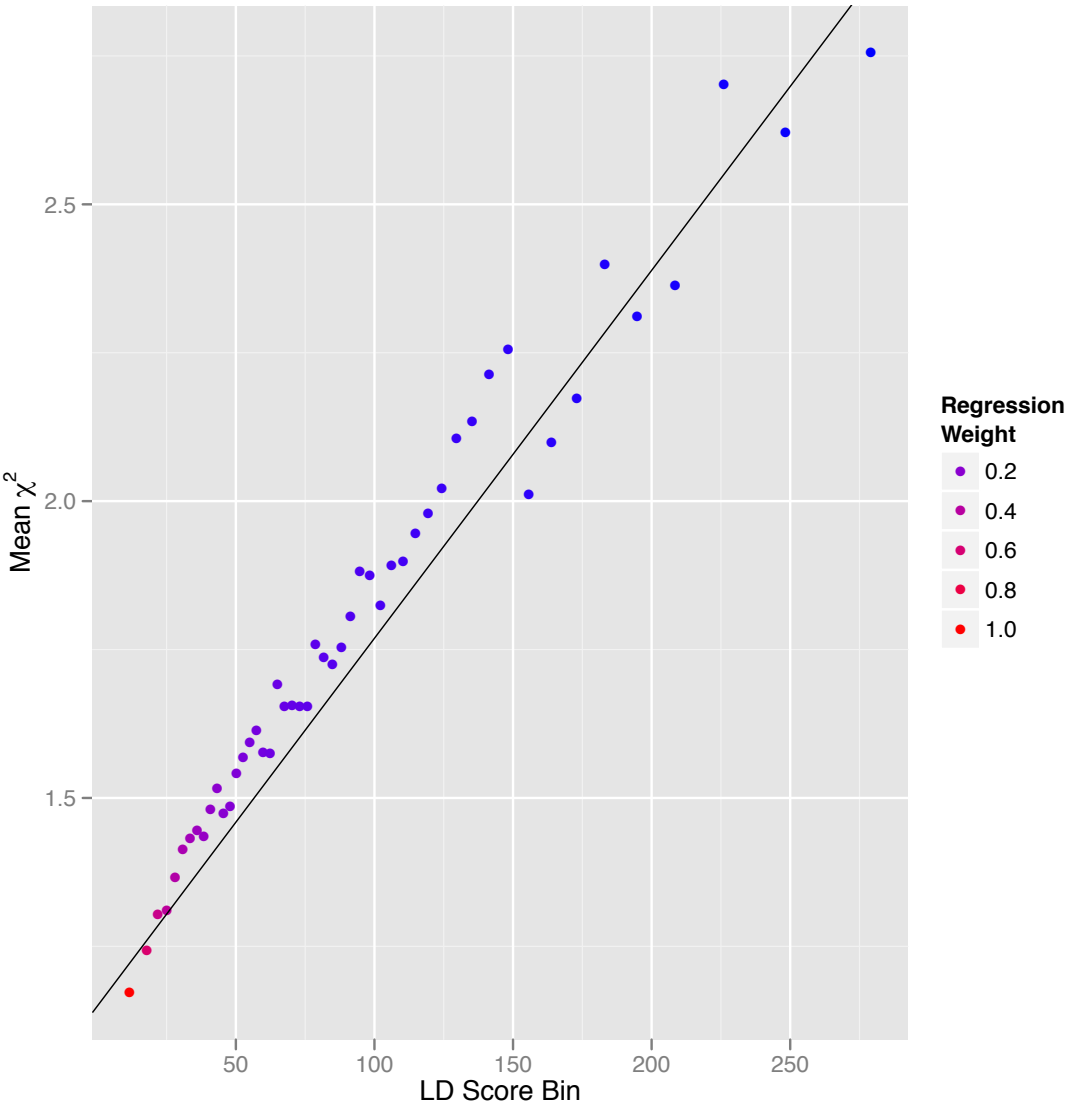
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8u: LD Score plot for Waist-Hip Ratio



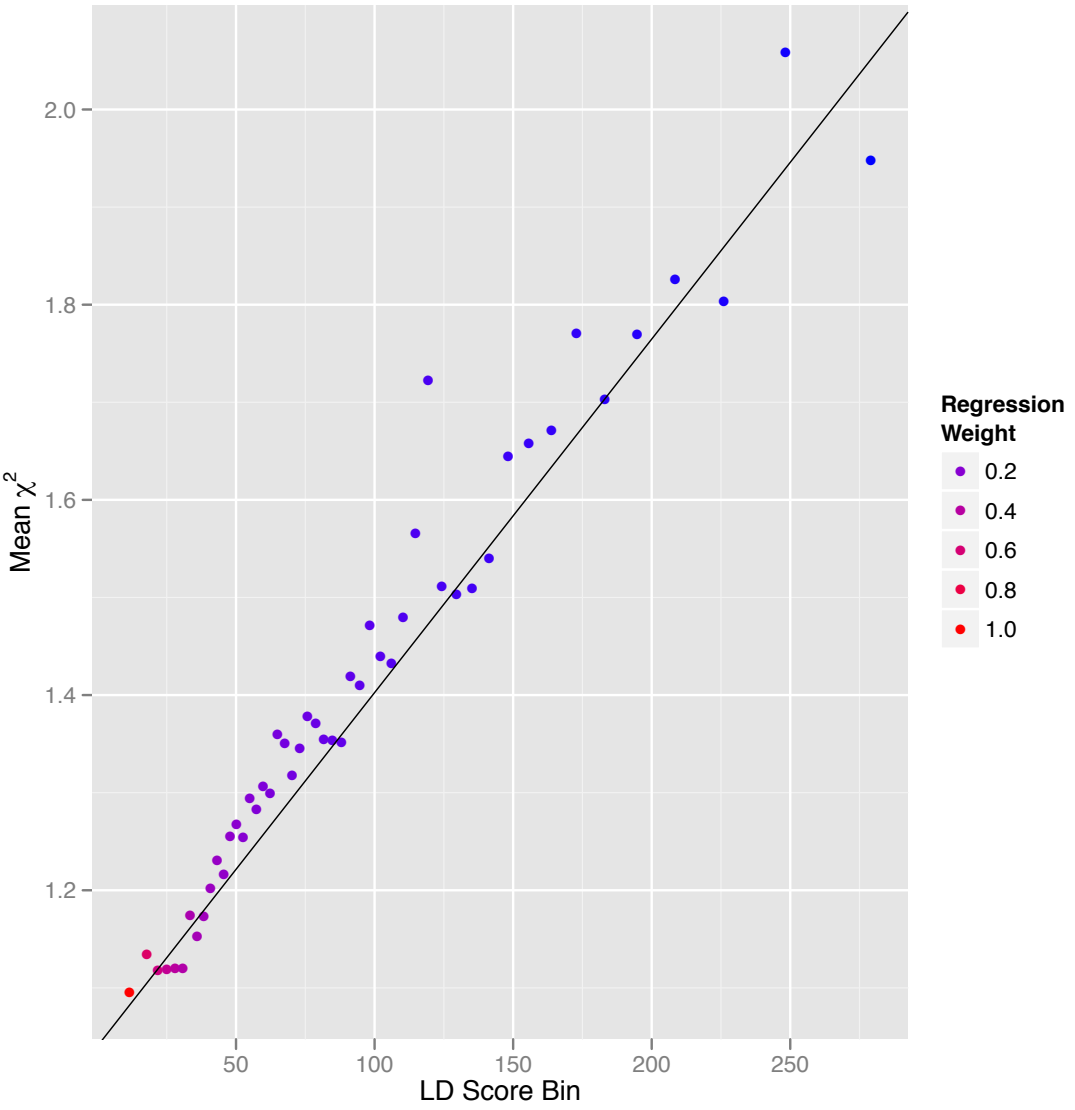
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8v: LD Score plot for Height



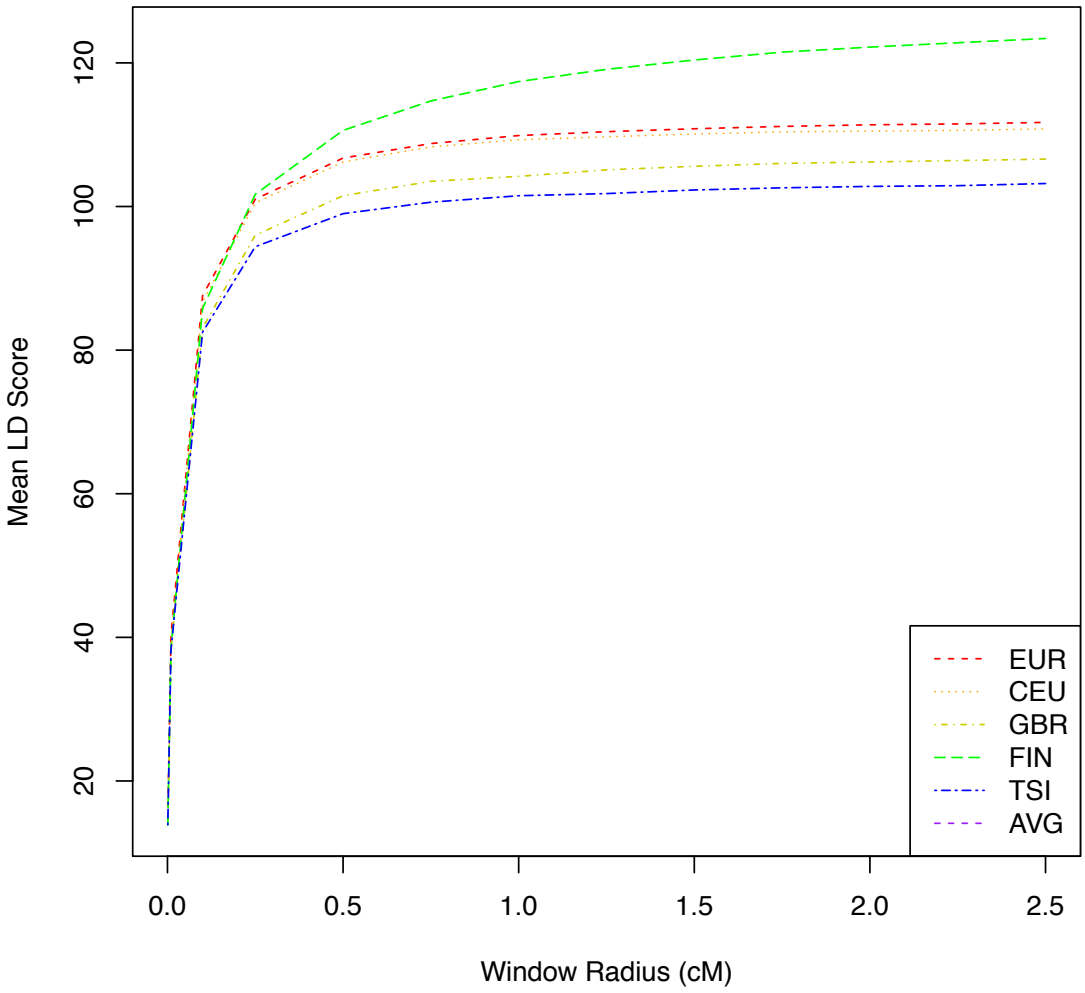
Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 8w: LD Score plot for Body Mass Index



Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line.

Supplementary Figure 9: LD Score estimates with varying window size



The *x*-axis displays the window radius used for estimating LD Score. The *y*-axis displays the mean LD Score among variants with sample MAF > 1% in all four 1000 Genomes European subpopulations. Each colored line represents one of the four 1000 Genomes European subpopulations: Europeans (EUR, 378 individuals), Utah Residents with Northern and Western European Ancestry (CEU, 85 individuals), British in England and Scotland (GBR, 88 individuals), Finnish in Finland (FIN, 93 individuals) and Toscani in Italia (TSI, 98 individuals). The line labeled AVG is the mean of the four subpopulation LD Scores, and is almost entirely obscured by the EUR line.

Supplementary Tables

Supplementary Table 1: Heritability Estimates

Phenotype	h^2_{obs}	h^2_{obs} SE	Sample Prevalence	Population Prevalence	h^2_{liab}	h^2_{liab} SE
Crohn's	0.763	0.043	0.285	0.0032[5]	0.405	0.023
Ulcerative Colitis	0.416	0.042	0.254	0.0025[5]	0.227	0.023
Schizophrenia	0.994	0.015	0.447	0.010[6]	0.555	0.008
Major Depression	0.342	0.028	0.493	0.150[6]	0.409	0.033
Bipolar	0.951	0.039	0.443	0.010[6]	0.531	0.022

LD Score regression heritability estimates for phenotypes that did not employ any GC correction (GC correction at the individual study level will bias the heritability estimate downwards). Standard errors were obtained as in table 1. These are estimates of the heritability explained by all 1000 Genomes SNPs (h2(1kG)), obtained by multiplying the regression slope by M/N, where N:=sample size and M is about 15 million. If the average rare SNP in 1000 Genomes explains less phenotypic variance than the average common SNP, then a smaller value of M would be more appropriate, and the estimates in this table will be biased upwards. Relaxing these assumptions in order to obtain a robust estimate h2(1kG) is a direction for further research; however, we note that the LD Score regression intercept is robust to these assumptions. We report heritability on the observed scale (h^2_{obs}) and also transformed to the liability scale (h^2_{liab}) using the prevalence estimates listed in the prevalence column. For some phenotypes (*e.g.,* Crohn's disease, which has been increasing in prevalence[5]), it is difficult to obtain accurate prevalence estimates, so liability scale heritability estimates should be interpreted cautiously.

Supplementary Table 2: Descriptions of cohorts for simulations with population stratification

Abbreviation	Origin	Principal Investigator	Controls
clo3	Cardiff, UK	Walters, J	945
cou3	UK	O'Donovan, M	544
egcu	Estonia	Esko, T	1,177
swe5	Sweden	Sullivan, PF	2,617
swe6	Sweden	Sullivan, PF	1,219
umeb	Umeå, Sweden	Adolfsson, R	584
umes	Umeå, Sweden	Adolfsson, R	713

Supplementary table 1 describes the seven PGC Schizophrenia control cohorts used for simulation with population stratification. All cohorts were genotyped on the Illumina Omni Express array; only unaffected individuals (controls) and directly genotyped SNPs post-QC (between approximately 600,000 and 700,000 SNPs, depending on cohort) were retained for simulations. In total genotypes for 9,135 individuals were incorporated into the simulations with pure population stratification

Supplementary Table 3a: Performance of genomic control and LD Score regression intercept in simulations with continental-scale population stratification

Population 1	Population 2	λ_{GC}	Intercept	$(Intercept - 1)/(\lambda_{GC} - 1)$
cou3	clo3	1.093	1.07	0.748
egcu	clo3	9.413	8.508	0.892
egcu	cou3	6.604	6.045	0.900
swe5	clo3	3.582	3.445	0.947
swe5	cou3	2.767	2.635	0.925
swe5	egcu	9.384	8.767	0.926
swe6	clo3	3.744	3.600	0.947
swe6	cou3	3.267	4.238	1.428
swe6	egcu	6.880	6.413	0.920
swe6	swe5	1.712	1.703	0.987
umeb	clo3	3.635	3.614	0.992
umeb	cou3	2.633	2.879	1.151
umeb	egcu	4.603	4.764	1.045
umeb	swe5	2.145	2.313	1.147
umeb	swe6	1.278	1.880	3.162*
umes	clo3	7.316	7.389	1.012
umes	cou3	4.959	5.572	1.155
umes	egcu	9.304	9.742	1.053
umes	swe5	6.900	7.328	1.072
umes	swe6	4.172	4.443	1.085
umes	umeb	3.192	3.218	1.012
	Mean (SD)			**1.017 (0.14)***

This table compares the performance of λ_{GC} and the LD Score regression intercept in simulations with continental-scale population stratification. In each simulation, individuals from population 1 were labeled cases and N_2 individuals from population 2 were labeled controls. We then computed association statistics for variants in the intersection of the subset of HapMap 3 variants used for LD Score regressions on real data (Online Methods) and variants on the Illumina Omni Express array (approx. 450,000 variants in each simulation).

The conclusion is that the LD Score regression intercept gives approximately the same answer as $\boldsymbol{\lambda_{GC}}$ in simulations with pure population stratification, and so would be appropriately conservative if used as a correction factor.

* The mean and SD are computed with the umeb/swe6 outlier removed.

Supplementary Table 3b. Correlation between LD Score and FST in simulations with continental-scale population stratification

Population 1	Population 2	Signed R-squared
cou3	clo3	5.00e-05
egcu	clo3	8.28e-04
egcu	cou3	7.41e-04
swe5	clo3	1.61e-04
swe5	cou3	2.02e-04
swe5	egcu	4.29e-04
swe6	clo3	1.81e-04
swe6	cou3	-1.12e-05
swe6	egcu	4.32e-04
swe6	swe5	5.95e-05
umeb	clo3	1.07e-04
umeb	cou3	5.05e-05
umeb	egcu	2.55e-04
umeb	swe5	1.92e-06
umeb	swe6	-6.60e-05
umes	clo3	5.22e-06
umes	cou3	6.79e-09
umes	egcu	7.26e-05
umes	swe5	-2.19e-06
umes	swe6	-5.23e-06
umes	umeb	-7.52e-06
	Mean	**2.51e-4**

Column descriptions. Population 1 and population 2 are the two populations involved in the simulations. Signed R-squared is the squared Pearson correlation coefficient between F_{ST} (between the two populations in the simulation) and LD Score multiplied by the negative one if the (non-squared) correlation is negative.

Supplementary Table 3c Heritability and intercept for a confounded GWAS with continental-scale population stratification

Population 1	Population2	Heritability	Intercept	Lambda
cou3	clo3	0.140	1.397	1.531
egcu	clo3	0.063	1.454	1.509
swe6	clo3	0.033	1.476	1.502
swe5	clo3	0.034	1.475	1.502
umeb	clo3	0.027	1.480	1.484
umes	clo3	0.009	1.493	1.487
swe5	cou3	0.044	1.468	1.506
umes	cou3	0.007	1.495	1.429
egcu	cou3	0.063	1.454	1.504
swe6	cou3	-0.096	1.570	1.399
umeb	cou3	0.026	1.481	1.418
swe5	egcu	0.048	1.465	1.502
umes	egcu	0.027	1.480	1.456
swe6	egcu	0.051	1.463	1.503
umeb	egcu	0.051	1.463	1.443
swe6	swe5	0.031	1.477	1.483
umes	swe5	0.002	1.498	1.465
umeb	swe5	0.007	1.495	1.431
umeb	swe6	-0.213	1.656	1.208
umes	swe6	-0.001	1.500	1.461
umes	umeb	-0.005	1.504	1.498
	Mean	**0.017**	**1.488**	**1.463**

This table puts the slopes from the simulations with continental-scale population stratification on an interpretable scale by transforming all parameters to the scale of a GWAS with 100,000 samples and mean chi-square of 1.5, where all inflation in the mean chi-square comes from population stratification. All estimates of h2(1kG) use M=15 million. For comparison, the aggregate LD Score estimator of h2(1kG), $\widehat{h^2} = \frac{M(\overline{\chi^2}-1)}{\overline{N\ell}}$, which is representative of heritability estimators that are highly susceptible to population stratification, would give a heritability estimate of **0.68** in all cases, assuming mean LD Score = 110. The reason why the LD Score regression slope is not equal to zero is likely because linked selection introduces a small correlation between LD Score and F_{ST}. The conclusion is that in a worst-case scenario (pure population stratification), LD Score regression misattributes on average a small proportion of stratification to heritability, but nevertheless performs many times better than existing estimators (upward bias of 0.017 for LD Score regression vs. approximately 0.68 for other methods).

Supplementary Table 4a: Performance of genomic control and LD Score regression intercept in simulations with national-scale population stratification

Population	PC	λ_{GC}	Intercept	$(Intercept - 1)/(\lambda_{GC} - 1)$
clo3	1	2.001	2.821	1.818
clo3	2	1.277	1.318	1.151
clo3	3	1.293	1.297	1.014
cou3	1	1.079	1.062	0.781
cou3	2	1.062	1.043	0.697
cou3	3	1.065	1.046	0.711
egcu	1	1.814	1.763	0.937
egcu	2	1.525	1.478	0.911
egcu	3	1.395	1.476	1.203
swe5	1	2.704	2.700	0.998
swe5	2	1.409	1.369	0.904
swe5	3	1.327	1.336	1.028
swe6	1	2.735	2.686	0.972
swe6	2	2.468	2.426	0.971
swe6	3	1.511	1.489	0.957
umeb	1	1.88	1.838	0.953
umeb	2	1.847	1.845	0.997
umeb	3	1.435	1.410	0.943
umes	1	2.039	1.958	0.922
umes	2	1.583	1.540	0.926
umes	3	1.328	1.294	0.896
Mean (SD)				**0.985 (0.224)**

This table compares the performance of λ_{GC} and the LD Score regression intercept in simulations with national-scale population stratification. We LD-pruned the SNPs so that no SNPs on the same chromosome had $R^2 > 0.02$, then computed the top three principal components. We then used these principal components as phenotypes and computed association statistics for the same set of variants as in the simulations described in supplementary table 2.

The conclusion is that the LD Score regression intercept gives approximately the same answer as $\boldsymbol{\lambda_{GC}}$ in simulations with pure population stratification, and so would be appropriately conservative if used as a correction factor.

Supplementary Table 4b. Correlation between LD Score and FST in simulations with national-scale population stratification

Population	PC	Signed R-Squared
clo3	1	-2.96e-04
clo3	2	1.85e-05
clo3	3	5.03e-06
cou3	1	2.71e-05
cou3	2	4.75e-05
cou3	3	4.73e-05
egcu	1	3.74e-05
egcu	2	1.17e-04
egcu	3	3.23e-05
swe5	1	1.94e-04
swe5	2	8.46e-05
swe5	3	8.79e-08
swe6	1	3.37e-05
swe6	2	1.30e-04
swe6	3	4.49e-05
umeb	1	4.49e-05
umeb	2	1.44e-05
umeb	3	2.25e-05
umes	1	1.44e-04
umes	2	2.18e-05
umes	3	9.09e-05
	Mean:	**4.10e-05**

Column descriptions. Population is the population and PC is the principal component used to simulate population stratification in the simulations and population 2. Signed R-squared is the squared Pearson correlation coefficient between F_{ST} (between the two populations in the simulation) and LD Score multiplied by the negative one if the (non-squared) correlation is negative.

Supplementary Table 4c: Heritability and intercept for a confounded GWAS with national-scale population stratification

Population	PC	Heritability	Intercept	Lambda
clo3	1	-0.178	1.630	1.347
clo3	2	0.048	1.465	1.404
clo3	3	0.031	1.477	1.471
cou3	1	0.144	1.395	1.505
cou3	2	0.254	1.313	1.450
cou3	3	0.229	1.332	1.467
egcu	1	0.035	1.474	1.506
egcu	2	0.068	1.450	1.494
egcu	3	0.043	1.468	1.389
swe5	1	0.039	1.472	1.473
swe5	2	0.076	1.444	1.491
swe5	3	0.024	1.483	1.469
swe6	1	0.017	1.488	1.502
swe6	2	0.036	1.474	1.488
swe6	3	0.047	1.466	1.487
umeb	1	0.031	1.477	1.501
umeb	2	0.014	1.490	1.491
umeb	3	0.042	1.469	1.498
umes	1	0.047	1.466	1.505
umes	2	0.037	1.473	1.510
umes	3	0.082	1.440	1.490
	Mean	**0.056**	**1.459**	**1.473**

This table is similar to supplementary table 2c it puts the slopes from the simulations with national-scale population stratification on an interpretable scale by transforming all parameters to the scale of a GWAS with 100,000 samples and mean chi-square of 1.5 where all inflation in the mean chi-square comes from population stratification along the relevant principal component. As in supplementary table 4, the aggregate estimator would give a h2(1kG) estimate of 0.68, which is similar to the result that one would obtain with Haseman-Elston regression or linear mixed models (using M=15 million). The conclusions are similar to supplementary table 4.

Supplementary Table 5: Simulations with both bias and polygenicity

Bias	Intercept (SD)	Null $\bar{\chi}^2$ (SD)	Null $\bar{\chi}^2$/Intercept (SD)
Relatedness	1.46 (0.02)	1.45 (0.02)	1.00 (0.00)
Stratification	1.53 (0.17)	1.48 (0.15)	0.97 (0.01)

Column descriptions. The column labeled bias identifies the source of bias, either cryptic relatedness (from the Framingham Heart Study) or population stratification (from introducing an environmental stratification term correlated with the first PC of the WTCCC2 data). Intercept is LD Score regression intercept, with the standard deviation (SD) across five simulations in parentheses. Null $\bar{\chi}^2$ is the mean χ^2 among SNPs on the opposite halves of chromosomes from causal SNPs, with SD across five simulations in parentheses. Since null SNPs are not in LD with causal SNPs, the mean χ^2 among null SNPs precisely quantifies the mean inflation in χ^2-statistics that results from bias. Null $\bar{\chi}^2$/Intercept is equal to the mean χ^2 among null SNPs divided by the LD Score regression intercept, with the SD across five simulations in parentheses. Null $\bar{\chi}^2$/Intercept should be approximately equal to one if the LD Score regression intercept is accurately estimating the mean inflation in test statistics that results from bias.

Supplementary Table 6: Simulations with Ascertained Binary Phenotypes

Sample Size	Prevalence	$\widehat{h}_l^2$ (SD)	Intercept (SD)	λ_{GC} (SD)	$\overline{\chi}^2$ (SD)
10000	0.01	0.804 (0.027)	0.995 (0.048)	2.253 (0.063)	2.452 (0.025)
10000	0.1	0.793 (0.041)	1.006 (0.04)	1.688 (0.031)	1.761 (0.015)
1000	0.01	0.772 (0.121)	1.005 (0.019)	1.139 (0.014)	1.145 (0.015)
1000	0.1	0.729 (0.226)	1.007 (0.027)	1.083 (0.03)	1.076 (0.013)

This table displays results from simulations with ascertained binary phenotypes following the liability threshold model. In all simulation replicates, the true heritability (of liability, in the population) was 0.8, the effective number of independent SNPs (defined as $M_{eff} := M/\bar{\ell}$) was 10,000 and the proportion of cases in the sample was 0.5. All SNPs were causal, with effect sizes (precisely, per-normalized genotype effects on liability) drawn *i.i.d.* from a normal distribution. Each entry in the table represents 20 simulation replicates. The column labeled $\widehat{h}_l^2$ lists the estimated heritability of liability in the population from the LD Score regression slope. The column labeled intercept lists LD Score regression intercepts. There was no population stratification in these simulations, so the intercept should be close to one. The columns λ_{GC} and $\overline{\chi}^2$ list the genomic control inflation factor and mean χ^2 computed from a perfectly LD-pruned set of variants.

Supplementary Table 7: Simulations with frequency-dependent genetic architecture

Exponent	Intercept (SD)	$\overline{\chi}^2$ (SD)
-3	1.007 (0.013)	1.011 (0.008)
-2	1.006 (0.014)	1.013 (0.008)
-1	1.003 (0.014)	1.023 (0.009)
-0.5	1.001 (0.013)	1.037 (0.009)
-0.25	1.000 (0.012)	1.048 (0.008)
0	0.998 (0.011)	1.059 (0.007)
0.25	0.997 (0.011)	1.070 (0.006)
0.5	0.996 (0.011)	1.079 (0.006)
1	0.994 (0.012)	1.091 (0.007)
2	0.991 (0.013)	1.101 (0.009)
3	0.989 (0.013)	1.105 (0.010)

Supplementary table 5 describes simulations in which per-normalized genotype effects (precisely, if X denotes a matrix of genotypes normalized to mean zero and variance one, the per-normalized genotype effects are a vector β such that $X\beta$ is equal to the additive genetic component of the phenotype) for 10,000 randomly chosen causal variants were drawn from $N(0, (p(1-p))^x)$, where *p* is MAF and *x* is the entry in the column labeled exponent. To prevent singleton and doubleton variants from having extreme effects for large negative values of *x*, we drew the effect sizes for variants with MAF < 1% from $N(0, 0.0099^x)$. Our model holds when *x=0*, which corresponds to moderate negative selection on the phenotype in question, similar to a typical disease phenotype. *x=1* is an appropriate model for a selectively neutral phenotype. Values of *x* outside the range [0,1] represent extreme genetic architectures. Standard errors are empirical standard errors across 10 replicates with randomly chosen causal variants and effect sizes.

Supplementary Table 8a: Summary Statistic Metadata, Quantitative Trait

Citation	Trait	N	Public	Ref
Heid, *et. al.*, Nat Genet, 2010	Waist-Hip Ratio	113,636	Yes	7
Lango Allen, *et. al.*, Nature, 2010	Height	183,727	Yes	8
Speliotes, *et. al.*, Nat Genet, 2010	Body Mass Index	249,796	Yes	9
TAG Consortium, Nat Genet, 2010	Smoking	74,053	Yes	10
International Consortium for Blood Pressure GWAS, Nature, 2011	Diastolic / Systolic Blood Pressure	69,395	Yes	11
Estrada *et. al.*, Nat Genet, 2011	Bone Mineral Density	32,961	Yes	12
Manning *et. al.*, Nat Genet, 2012	Fasting Insulin	51,750	Yes	13
Rietveld, *et. al.*, Science, 2013	Years of Education	126,559	Yes	14

Column descriptions. All columns are self-explanatory, except the column labeled N counts the number of individuals in the discovery phase of the GWAS, not including replication samples. The column labeled public indicates whether the summary statistics are publicly available for download (see URLs).

Supplementary Table 8b: Summary Statistic Metadata, Case/Control

Citation	Trait	Cases	Controls	Public	Ref
Neale, *et. al.*, J Am Acad Adolesc Psychiatry, 2010	ADHD	896	2455	Yes	15
Stahl, *et. al.*, Nat Genet, 2010	Rheumatoid Arthritis	5,539	20,169	Yes	16
PGC Bipolar Working Group, Nat Genet, 2011	Bipolar Disorder	7,481	9,250	Yes	17
Schunkert *et. al.*, Nat Genet, 2011	Coronary Artery Disease	22,233	64,762	Yes	18
Jostins, *et. al.*, Nature, 2012	Inflammatory Bowel Disease	12,882	21,770	No	19
Jostins, *et. al.*, Nature, 2012	Crohn's Disease	5,956	14,927	Yes*	19
Jostins, *et. al.,* Nature, 2012	Ulcerative Colitis	6,968	20,464	Yes*	19
Morris, *et. al.*, Nat Genet, 2012	Type 2 Diabetes	12,171	56,862	Yes	20
Cross-Disorder Group, Lancet, 2013	PGC Cross-Disorder	33,332	27,888	Yes	21
Ripke, *et. al.*, Mol Psych, 2013	Major Depression	9,240	9,519	Yes	22
O'Donovan, *et. al.*, in preparation	Schizophrenia	31,335**	38,765**	Yes	23
Rietveld, *et. al.*, Science, 2013	College	22,044***	73,383	Yes	14

Column descriptions. All columns are self-explanatory, except the columns labeled cases and controls note the number of cases and controls in the discovery phase of the GWAS, not including replication samples. The column labeled public indicates whether the summary statistics are publicly available for download (see URLs)

* These summary statistics may be meta-analyzed with Immunochip data, which is not appropriate for LD Score regression.

** This figure counts only European samples. The full GWAS includes several thousand Asian samples, which were excluded from the LD Score regression, because the 1000 Genomes European LD Score is not representative of LD patterns in Asian populations.

*** Here cases are individuals with college education, controls those without.

Supplementary Table 9: Simulation with intergenic GC correction

Annotation	Mean χ^2	Lambda
Null (chromosome 2)	1.0098	1.0082
Within 100 kB of a coding exon on chromosome 1	1.4592	1.2511
More than 100 kB from a coding exon on chromosome 1	1.2505	**1.0817**

This table describes a simulation with 1000 Swedish samples and ~700,000 best-guess imputed genotypes on chromosome 1. We simulated phenotypes by assigning causal effects to only SNPs within coding exons on chromosome 1. We then computed association statistics for variants within 100 kB of a gene, more than 100 kB from a gene and for null SNPs on chromosome 2. Because of long-range linkage disequilibrium, lambda (*i.e.,* λ_{GC}) is significantly elevated for intergenic SNPs even though there is no bias in the test statistics, as can be seen from the fact that the test statistics of null SNPs are not inflated.

Supplementary Table 10: R^2 matrix of LD Scores with varying window sizes

cM	0.01	0.1	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
0.001	0.6677	0.4249	0.3769	0.3642	0.3543	**0.3505**	0.3504	0.3494	0.3489	0.3485	0.3485	0.3479
0.01		0.7553	0.6929	0.6732	0.6603	**0.6530**	0.6523	0.6502	0.6487	0.6476	0.6479	0.6463
0.1			0.9651	0.9538	0.9424	**0.9359**	0.9347	0.9328	0.9314	0.9303	0.9300	0.9289
0.25				0.9899	0.9856	**0.9810**	0.9801	0.9786	0.9773	0.9763	0.9764	0.9751
0.5					0.9955	**0.9934**	0.9930	0.9920	0.9911	0.9904	0.9904	0.9895
0.75						**0.9981**	**0.9980**	**0.9973**	**0.9965**	**0.9960**	**0.9961**	**0.9952**
1							0.9995	0.9993	0.9990	0.9986	0.9985	0.9980
1.25								0.9996	0.9993	0.9990	0.9990	0.9986
1.5									0.9997	0.9995	0.9994	0.9992
1.75										0.9997	0.9995	0.9995
2											0.9996	0.9997
2.25												0.9995

Each entry is the squared Pearson correlation between the LD Score estimated from the 1000 Genomes Project European reference panel with the window radii listed across the top row and the leftmost column in units of centiMorgans (cM). We chose to use the 1 cM LD Score for all LD Score regressions applied to real data, so the squared correlations with the 1 cM LD Score are in bold

References

1. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-9 (2010).
2. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A.L. Estimating and interpreting FST: the impact of rare variants. *Genome Res* **23**, 1514-21 (2013).
3. Turchin, M.C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* **44**, 1015-9 (2012).
4. International HapMap, C. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
5. Molodecky, N.A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46-54 e42; quiz e30 (2012).
6. Cross-Disorder Group of the Psychiatric Genomics, C. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* **45**, 984-94 (2013).
7. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949-60 (2010).
8. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-8 (2010).
9. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937-48 (2010).
10. Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* **42**, 441-7 (2010).
11. International Consortium for Blood Pressure Genome-Wide Association, S. *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103-9 (2011).
12. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* **44**, 491-501 (2012).
13. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-69 (2012).
14. Rietveld, C.A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-71 (2013).
15. Neale, B.M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**, 884-97 (2010).
16. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* **42**, 508-14 (2010).
17. Psychiatric, G.C.B.D.W.G. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* **43**, 977-83 (2011).
18. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333-8 (2011).
19. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-24 (2012).

20. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
21. Cross-Disorder Group of the Psychiatric Genomics, C. & Genetic Risk Outcome of Psychosis, C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371-9 (2013).
22. Major Depressive Disorder Working Group of the Psychiatric, G.C. *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* **18**, 497-511 (2013).
23. Consortium., S.W.G.o.t.P.G. Biological Insights from 108 Schizophrenia-Associated Genetic Loci. *Nature* **511**, 421-7 (2014).