nature genetics

SUPPLEMENTARY INFORMATION
https://doi.org/10.1038/s41588-019-0465-0

In the format provided by the authors and unedited.

Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture

Kangcheng Hou[1,2,9], Kathryn S. Burch[3,9]*, Arunabha Majumdar[1], Huwenbo Shi[3,4], Nicholas Mancuso[1,5], Yue Wu[6], Sriram Sankararaman[3,6,7,8] and Bogdan Pasaniuc[1,3,7,8]*

[1]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [2]College of Computer Science and Technology, Zhejiang University, Hangzhou, China. [3]Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [5]Biostatistics Division, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [6]Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA. [7]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [8]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [9]These authors contributed equally: Kangcheng Hou, Kathryn S. Burch. *e-mail: kathrynburch@ucla.edu; pasaniuc@ucla.edu

Supplementary Note

Derivation for $\widehat{h}^2_{\text{GRE}}$ assuming fixed $\boldsymbol{\beta}$ and $N > M$

Recall that $\text{Var}[y_n] = 1$ and $\text{Var}[\mathbf{x}_n^T] = \mathbf{V}$. Our goal is to find an estimator $\widehat{h}^2_{\text{GRE}}$ that satisfies $\text{E}[\widehat{h}^2_{\text{GRE}}] = h_g^2 = \text{Var}[\mathbf{x}_n^T\boldsymbol{\beta}] = \text{E}[\text{Var}[\mathbf{x}_n^T\boldsymbol{\beta}|\boldsymbol{\beta}]] + \text{Var}[\text{E}[\mathbf{x}_n^T\boldsymbol{\beta}|\boldsymbol{\beta}]] = \text{E}[\boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}]$. If $\boldsymbol{\beta}$ were fixed and we observed $\mathbf{V}$ and $\boldsymbol{\beta}$, we could estimate h_g^2 as $\widehat{h}_g^2 = \boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta}$. However, in reality, we observe noisy estimates of $\boldsymbol{\beta}$ and $\mathbf{V}$ from a GWAS. Given a GWAS of N unrelated individuals and M SNPs, we observe $\mathbf{X}$, the $N \times M$ standardized genotype matrix, and $\mathbf{y}$, the $N \times 1$ standardized phenotype vector. We assume that when $N > M$, $\widehat{\mathbf{V}} \to \mathbf{V}$ as $N \to \infty$ (in practice, the assumption that $N > M$ is untrue; in subsequent sections we show how we partition the genome into K blocks such that $N > p_k$ for each block k). In a typical GWAS, the marginal SNP effects are estimated through ordinary least squares (OLS) regression as $\hat{\boldsymbol{\beta}} = (1/N)\mathbf{X}^T\mathbf{y} = (1/N)\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (1/N)\mathbf{X}^T\boldsymbol{\epsilon} = \widehat{\mathbf{V}}\boldsymbol{\beta} + (1/N)\mathbf{X}^T\boldsymbol{\epsilon}$. Given $\mathbf{X}$ and fixed $\boldsymbol{\beta}$, it follows that

$$
\begin{aligned}
\text{E}[\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}] &= \text{E}[\widehat{\mathbf{V}}\boldsymbol{\beta} + (1/N)\mathbf{X}^T\boldsymbol{\epsilon}|\boldsymbol{\beta}, \mathbf{X}] \qquad (1)\\
&= \widehat{\mathbf{V}}\boldsymbol{\beta} + (1/N)\mathbf{X}^T\text{E}[\boldsymbol{\epsilon}]\\
&= \widehat{\mathbf{V}}\boldsymbol{\beta}\\
\text{Cov}[\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}] &= \text{Cov}[\widehat{\mathbf{V}}\boldsymbol{\beta} + (1/N)\mathbf{X}^T\boldsymbol{\epsilon}|\boldsymbol{\beta}, \mathbf{X}] \qquad (2)\\
&= (\sigma_e^2/N^2)\mathbf{X}^T\mathbf{X}\\
&= \frac{\sigma_e^2}{N}\widehat{\mathbf{V}}
\end{aligned}
$$

Thus, as $N \to \infty$, $\hat{\boldsymbol{\beta}} \to \mathbf{V}\boldsymbol{\beta}$. Substituting $\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}} \approx \boldsymbol{\beta}$ and $\widehat{\mathbf{V}} \approx \mathbf{V}$, we obtain an estimator $\widehat{h}_g^2 = \boldsymbol{\beta}^T\mathbf{V}\boldsymbol{\beta} \approx (\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}})^T\widehat{\mathbf{V}}(\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}}^T\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}}$. The expectation of this estimator is

$$
\begin{aligned}
\text{E}[\hat{\boldsymbol{\beta}}^T\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}] &= \text{E}[tr(\hat{\boldsymbol{\beta}}^T\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}})|\boldsymbol{\beta}, \mathbf{X}]\\
&= \text{E}[tr(\widehat{\mathbf{V}}^{-1}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T)|\boldsymbol{\beta}, \mathbf{X}]\\
&= tr(\widehat{\mathbf{V}}^{-1}\text{E}[\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^T|\boldsymbol{\beta}, \mathbf{X}])\\
&= tr(\widehat{\mathbf{V}}^{-1}\text{Cov}[\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}]) + tr(\widehat{\mathbf{V}}^{-1}\text{E}[\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}]\text{E}[\hat{\boldsymbol{\beta}}|\boldsymbol{\beta}, \mathbf{X}]^T)\\
&= tr((\sigma_e^2/N)\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{V}}) + \boldsymbol{\beta}^T\widehat{\mathbf{V}}\boldsymbol{\beta}\\
&= \frac{M}{N}\sigma_e^2 + \boldsymbol{\beta}^T\widehat{\mathbf{V}}\boldsymbol{\beta} \qquad (3)
\end{aligned}
$$

We define $\widehat{h}^2_{\text{GRE}}$ to be an estimator that satisfies $\text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}] = \boldsymbol{\beta}^T \widehat{\mathbf{V}} \boldsymbol{\beta}$. Substituting into Equation 3, we obtain

$$
\begin{aligned}
\text{E}[\hat{\boldsymbol{\beta}}^T \widehat{\mathbf{V}}^{-1} \hat{\boldsymbol{\beta}} | \boldsymbol{\beta}, \mathbf{X}] &= \frac{M(1 - \text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}])}{N} + \text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}] \\
&= \frac{M}{N} + \frac{N-M}{N} \text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}] \\
\text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}] &= \left(\text{E}[\hat{\boldsymbol{\beta}}^T \widehat{\mathbf{V}}^{-1} \hat{\boldsymbol{\beta}} | \boldsymbol{\beta}, \mathbf{X}] - \frac{M}{N} \right) \frac{N}{N-M} \\
&= \frac{N \text{E}[\hat{\boldsymbol{\beta}}^T \widehat{\mathbf{V}}^{-1} \hat{\boldsymbol{\beta}} | \boldsymbol{\beta}, \mathbf{X}] - M}{N-M} \\
\widehat{h}^2_{\text{GRE}} &= \frac{N \hat{\boldsymbol{\beta}}^T \widehat{\mathbf{V}}^{-1} \hat{\boldsymbol{\beta}} - M}{N-M}
\end{aligned} \tag{4}
$$

Unbiasedness of $\widehat{h}^2_{\text{GRE}}$ under the GRE model when $N > M$

Recall that under the GRE model, $\text{E}[\beta_i] = 0$ and $\text{Var}[\beta_i] = \sigma_i^2$, where $\sigma_i^2 \geq 0$ for all SNPs i. In previous sections, we showed that $\text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}] = \boldsymbol{\beta}^T \widehat{\mathbf{V}} \boldsymbol{\beta}$ and $h_g^2 = \sum_{i=1}^M \sigma_i^2$. Recalling that $\text{Cov}[\beta_i, \beta_j] = 0$ for all $i \neq j$, it follows that

$$
\begin{aligned}
\text{E}[\widehat{h}^2_{\text{GRE}}|\mathbf{X}] &= \text{E}[\text{E}[\widehat{h}^2_{\text{GRE}}|\boldsymbol{\beta}, \mathbf{X}]|\mathbf{X}] \\
&= \text{E}[\boldsymbol{\beta}^T \widehat{\mathbf{V}} \boldsymbol{\beta} | \mathbf{X}] \\
&= \text{E}[tr(\boldsymbol{\beta}^T \widehat{\mathbf{V}} \boldsymbol{\beta}) | \mathbf{X}] \\
&= tr(\widehat{\mathbf{V}} \text{E}[\boldsymbol{\beta} \boldsymbol{\beta}^T]) \\
&= \sum_{i=1}^M \sigma_i^2
\end{aligned} \tag{5}
$$

Therefore, $\text{E}[\widehat{h}^2_{\text{GRE}}] = \text{E}[\text{E}[\widehat{h}^2_{\text{GRE}}|\mathbf{X}]] = \sum_{i=1}^M \sigma_i^2 = h_g^2$. This result implies that $\widehat{h}^2_{\text{GRE}}$ is an unbiased estimator for h_g^2 under any genetic architecture that can be defined by the GRE model.

Genome-wide approximation

If M significantly exceeds N (which is the case for most GWAS), Equation 4 produces meaningless (negative) estimates and $\widehat{\mathbf{V}}$ is a poor estimator of $\mathbf{V}$ genome-wide; as M/N increases, the eigenstructure of $\widehat{\mathbf{V}}$ becomes increasingly distorted (larger eigenvalues are overestimated and smaller eigenvalues are underestimated) [1]. In addition, it is computationally intractable to invert $\widehat{\mathbf{V}}$ genome-wide. Thus, in practice, we divide the

genome into K approximately independent blocks (i.e. chromosomes) and, following a procedure similar to Equations (1)–(4), we obtain

$$\begin{aligned}
\mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}\right] &= \frac{p_k}{N}\sigma_e^2 + \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k \\
&= \frac{p_k}{N}\left(1 - \mathrm{E}\left[\hat{h}_g^2 | \boldsymbol{\beta}, \mathbf{X}\right]\right) + \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k
\end{aligned} \tag{6}$$

To find an estimator that satisfies $\mathrm{E}\left[\hat{h}_g^2 | \boldsymbol{\beta}, \mathbf{X}\right] = \sum_k \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k$ we sum Equation 6 over $k = 1, \ldots, K$:

$$\begin{aligned}
\sum_{k=1}^{K} \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k &= \sum_{k=1}^{K} \mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}\right] - \frac{1}{N}\sum_{k=1}^{K} p_k + \frac{1}{N}\mathrm{E}\left[\widehat{h}_g^2 | \boldsymbol{\beta}, \mathbf{X}\right] \sum_{k=1}^{K} p_k \\
\mathrm{E}\left[\widehat{h}_g^2 | \boldsymbol{\beta}, \mathbf{X}\right]\left(N - \sum_{k=1}^{K} p_k\right) &= N \sum_{k=1}^{K} \mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}\right] - p_k \\
\mathrm{E}\left[\widehat{h}_g^2 | \boldsymbol{\beta}, \mathbf{X}\right] &= \frac{N \sum_{k=1}^{K} \mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{x}\right] - \sum_{k=1}^{K} p_k}{N - \sum_{k=1}^{K} p_k} \\
\widehat{h}_g^2 &= \frac{\sum_k N \widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k - \sum_k p_k}{N - \sum_k p_k} \\
\widehat{h}_g^2 &= \frac{\sum_k N \widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k - M}{N - M}
\end{aligned} \tag{7}$$

While Equation 7 does circumvent the need to invert the genome-wide LD matrix, it still produces negative estimates if $N < M$, which is the case in all of our genome-wide analyses. We therefore use an approximation which estimates the contribution of block k while ignoring the contributions of the remaining blocks. That is, assuming $\mathbf{y} = \mathbf{X}_k \boldsymbol{\beta}_k + \boldsymbol{\epsilon}_k$, where $\mathrm{Var}\left[\boldsymbol{\epsilon}_k\right] = \sigma_{e_k}^2 \mathbf{I}_N$,

$$\begin{aligned}
\mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}\right] &= \frac{p_k}{N}\left(1 - h_k^2\right) + \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k \\
&= \frac{p_k}{N} - \frac{p_k}{N}\mathrm{E}\left[\widehat{h}_k^2 | \boldsymbol{\beta}, \mathbf{X}\right] + \mathrm{E}\left[\widehat{h}_k^2 | \boldsymbol{\beta}, \mathbf{X}\right] \mathrm{E}\left[\widehat{h}_k^2 | \boldsymbol{\beta}, \mathbf{X}\right] \\
&= \frac{N\mathrm{E}\left[\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}\right] - p_k}{N - p_k}
\end{aligned} \tag{8}$$

An estimator that satisfies Equation 8 is

$$\widehat{h}_k^2 = \frac{N \widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k - p_k}{N - p_k}$$

Finally, we estimate genome-wide SNP-heritability as

$$\hat{h}^2_{\text{GRE}} = \sum_{k=1}^{K} \frac{N\widehat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{-1} \widehat{\boldsymbol{\beta}}_k - p_k}{N - p_k}$$

While this estimator is biased, we find that it robustly estimates genome-wide SNP-heritability as long as $N \gg p_k$ for all k (e.g., Figure 1b).

Extension for rank-deficient LD

It is often the case that two SNPs are perfectly correlated in a genotype block $\mathbf{X}_k$, or that $N < p_k$ for a block k. In this case, $\widehat{\mathbf{V}}_k$ is rank-deficient (i.e. its rank is less than p_k) and $\widehat{\mathbf{V}}_k^{-1}$ does not exist. We therefore compute $\mathbf{V}_k^{\dagger}$, the pseudoinverse (Moore-Penrose inverse) of $\widehat{\mathbf{V}}_k$, which approximates $\widehat{\mathbf{V}}_k^{-1}$ using its truncated eigendecomposition.

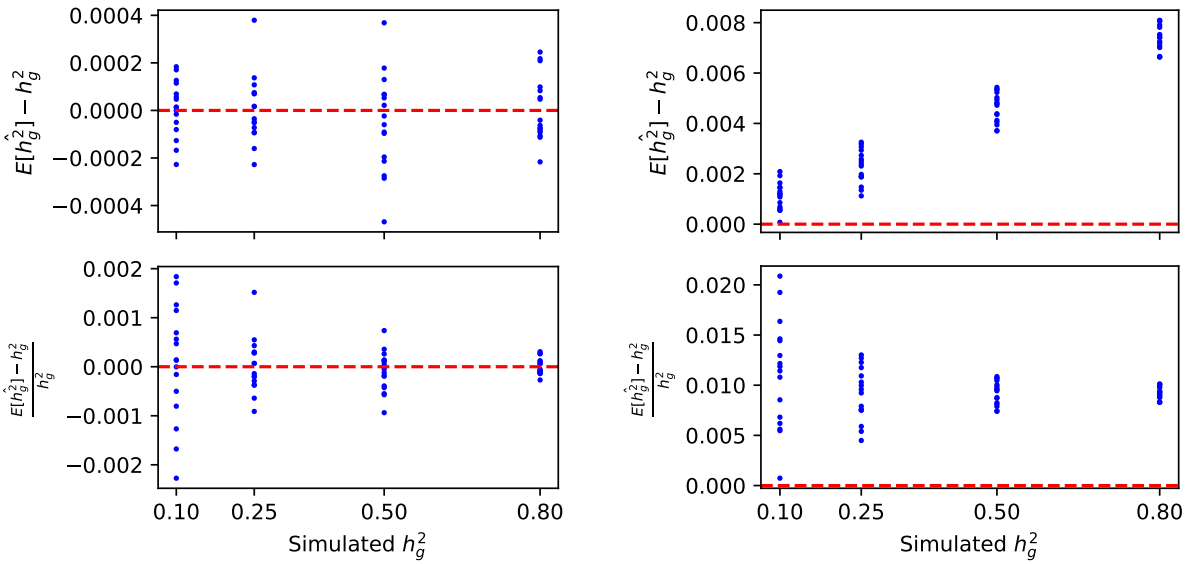
Let $q_k = \text{rank}(\widehat{\mathbf{V}}_k)$ and let $\widehat{\mathbf{V}}_k = \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^T$ be the eigendecomposition of $\widehat{\mathbf{V}}_k$, where $\boldsymbol{\Lambda}_k = \text{diag}(\lambda_1, \ldots, \lambda_{q_k}, 0, \ldots, 0)$. The pseudoinverse of $\widehat{\mathbf{V}}_k$ is $\widehat{\mathbf{V}}_k^{\dagger} = \mathbf{U}_k \boldsymbol{\Lambda}_k^{\dagger} \mathbf{U}_k^T$, where $\boldsymbol{\Lambda}_k^{\dagger} = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_{q_k}^{-1}, 0, \ldots, 0)$. Substituting $\widehat{\mathbf{V}}_k^{\dagger}\hat{\boldsymbol{\beta}}_k \approx \boldsymbol{\beta}_k$ and $\widehat{\mathbf{V}}_k \approx \mathbf{V}_k$, we obtain the following estimator for h_k^2: $\widehat{h}_k^2 = \boldsymbol{\beta}_k^T \mathbf{V}_k \boldsymbol{\beta}_k \approx (\widehat{\mathbf{V}}_k^{\dagger}\hat{\boldsymbol{\beta}}_k)^T \widehat{\mathbf{V}}_k (\widehat{\mathbf{V}}_k^{\dagger}\hat{\boldsymbol{\beta}}_k) = \hat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{\dagger} \hat{\boldsymbol{\beta}}_k$. Let $\mathbf{I}_{q_k}$ be a $p_k \times p_k$ diagonal matrix in which the first q_k diagonal entries are 1 and the rest are 0. The expectation of our estimator given $\boldsymbol{\beta}$ and $\mathbf{X}$ is

$$\begin{aligned}
\text{E}[\hat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{\dagger} \hat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}] &= \text{E}[tr(\widehat{\mathbf{V}}_k^{\dagger} \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^T) | \boldsymbol{\beta}, \mathbf{X}] \\
&= tr(\widehat{\mathbf{V}}_k^{\dagger} \text{E}[\hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k^T | \boldsymbol{\beta}, \mathbf{X}]) \\
&= tr(\widehat{\mathbf{V}}_k^{\dagger} \text{Cov}[\hat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}]) + tr(\widehat{\mathbf{V}}_k^{\dagger} \text{E}[\hat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}] \text{E}[\hat{\boldsymbol{\beta}}_k | \boldsymbol{\beta}, \mathbf{X}]^T) \\
&= tr((\sigma_e^2/N) \widehat{\mathbf{V}}_k^{\dagger} \widehat{\mathbf{V}}_k) + tr(\widehat{\mathbf{V}}_k^{\dagger} (\widehat{\mathbf{V}}_k \boldsymbol{\beta}_k)(\widehat{\mathbf{V}}_k \boldsymbol{\beta}_k)^T) \\
&= tr((\sigma_e^2/N) \mathbf{I}_{q_k}) + tr(\mathbf{U}_k \boldsymbol{\Lambda}_k^{\dagger} \mathbf{U}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^T \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^T) \\
&= \frac{q_k}{N} \sigma_e^2 + tr(\mathbf{U}_k \boldsymbol{\Lambda}_k^{\dagger} \boldsymbol{\Lambda}_k \mathbf{U}_k^T \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^T) \\
&= \frac{q_k}{N} \sigma_e^2 + tr(\mathbf{I}_{q_k} \mathbf{U}_k^T \boldsymbol{\beta}_k \boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k) \\
&= \frac{q_k}{N} \sigma_e^2 + tr(\boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{I}_{q_k} \mathbf{U}_k^T \boldsymbol{\beta}_k) \\
&= \frac{q_k}{N} \sigma_e^2 + \boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{I}_{q_k} \mathbf{U}_k^T \boldsymbol{\beta}_k \\
&= \frac{q_k}{N} \sigma_e^2 + \boldsymbol{\beta}_k^T \mathbf{U}_k \boldsymbol{\Lambda}_k \mathbf{U}_k^T \boldsymbol{\beta}_k \\
&= \frac{q_k}{N} \sigma_e^2 + \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k
\end{aligned}$$

We wish to find an estimator that satisfies $\mathrm{E}[\widehat{h}^2_{\mathrm{GRE}}|\boldsymbol{\beta}, \mathbf{X}] = \boldsymbol{\beta}^T \widehat{\mathbf{V}} \boldsymbol{\beta} = \sum_{k=1}^{K} \boldsymbol{\beta}_k^T \widehat{\mathbf{V}}_k \boldsymbol{\beta}_k$. Substituting into the above equation, we obtain

$$\widehat{h}^2_{\mathrm{GRE}} = \sum_{k=1}^{K} \frac{N \hat{\boldsymbol{\beta}}_k^T \widehat{\mathbf{V}}_k^{\dagger} \hat{\boldsymbol{\beta}}_k - q_k}{N - q_k}$$

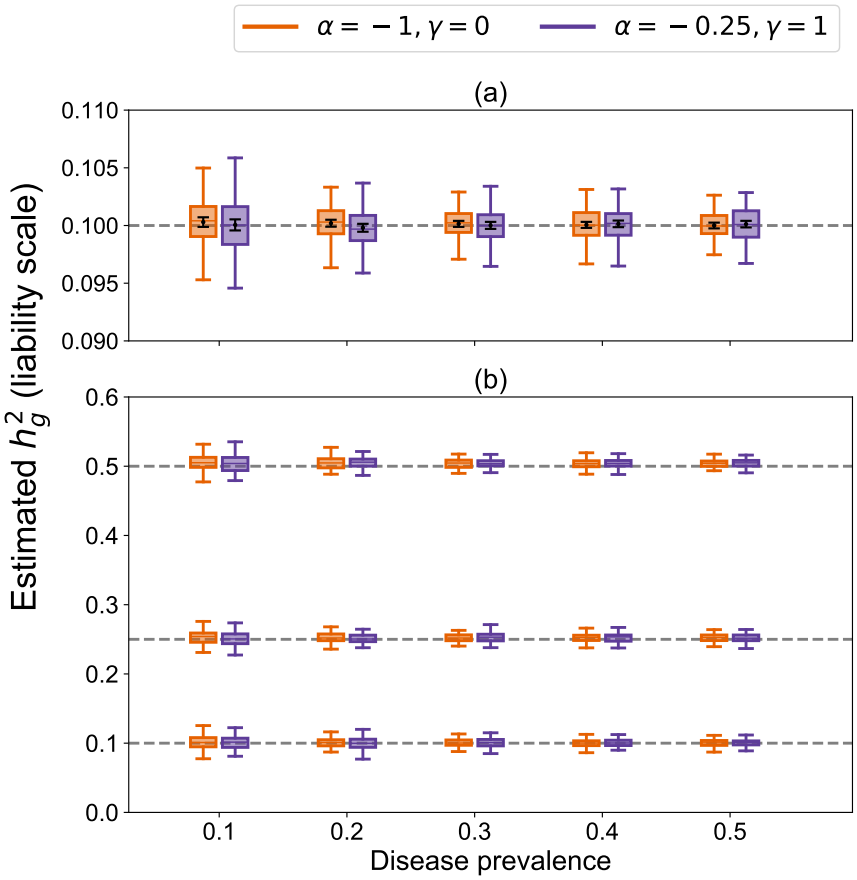
Supplementary Figures



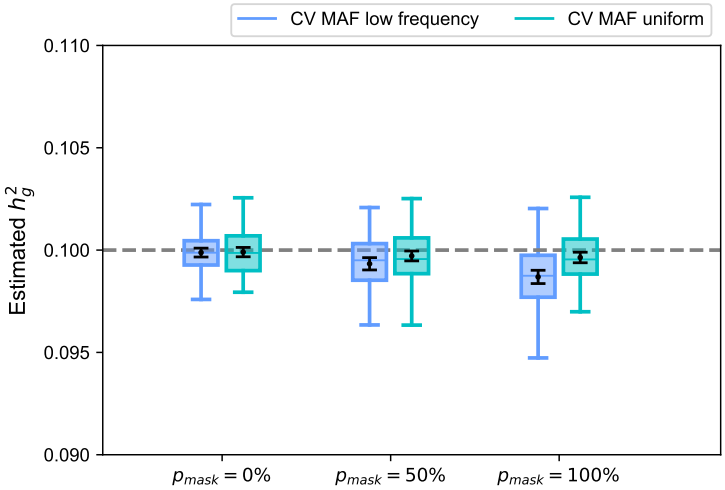
(a) chromosome 22 simulations ($M = 9654$)

(b) genome-wide simulations ($M = 593300$)

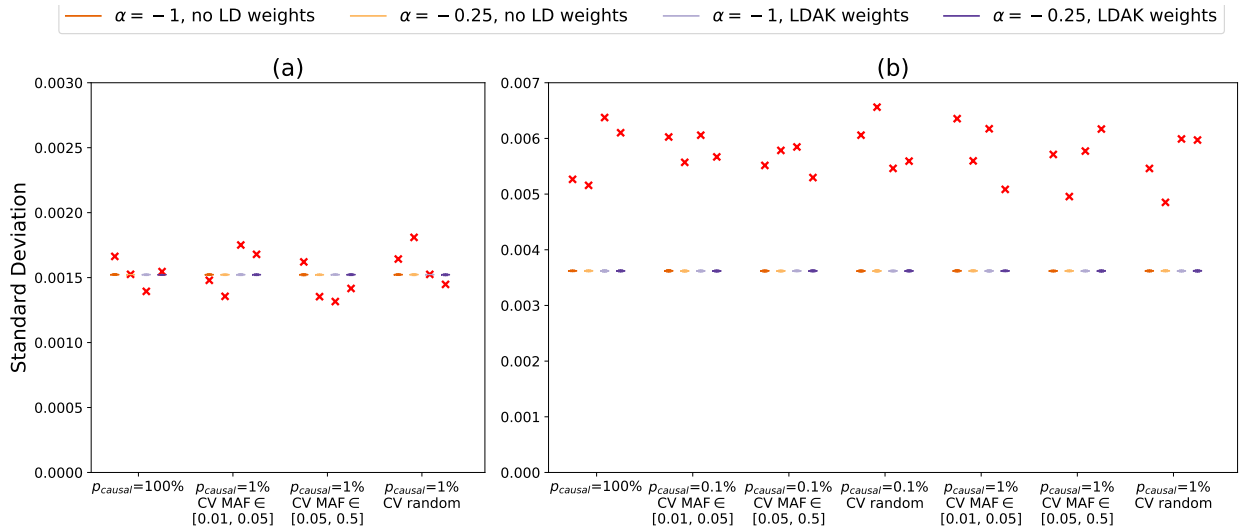
Supplementary Figure 1: Bias and relative bias of $\widehat{h}^2_{\text{GRE}}$ in simulations under 64 MAF- and LDAK-LD-dependent architectures ($N = 337$K). (a) Phenotypes were drawn from $M = 9654$ SNPs on chromosome 22; h_g^2 was estimated with a single LD block spanning chromosome 22. (b) Phenotypes were drawn from $M = 593300$ SNPs genome-wide; h_g^2 was estimated using 22 chromosome-wide LD blocks. Each point represents the magnitude of the bias of $\widehat{h}^2_{\text{GRE}}$ (top row) or the bias of $\widehat{h}^2_{\text{GRE}}$ relative to the simulated h_g^2 (bottom row) estimated from 100 simulations under a single genetic architecture.



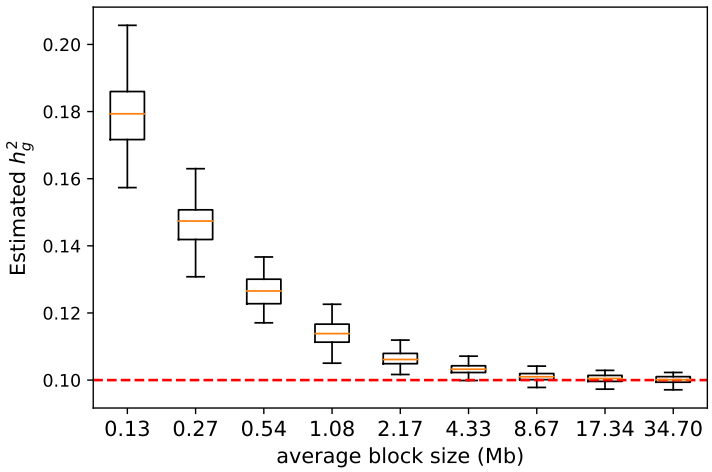
Supplementary Figure 2: $\widehat{h}^2_{\text{GRE}}$ for case-control GWAS with no ascertainment ($N = 337$K). Each boxplot represents estimates from 100 independent simulations at the specified disease prevalence. In all simulations, $p_{\text{causal}} = 1$ and causal variants were drawn uniformly. (a) Each individual's liability was drawn from $M = 9654$ SNPs on chromosome 22 and converted to a binary case-control status; h_g^2 was estimated with a single block. Black points and error bars represent the mean and ± 2 s.e.m. (b) Each individual's liability was drawn from $M = 593300$ SNPs genome-wide and converted to a binary case-control status; h_g^2 was estimated with 22 chromosome-wide LD blocks.



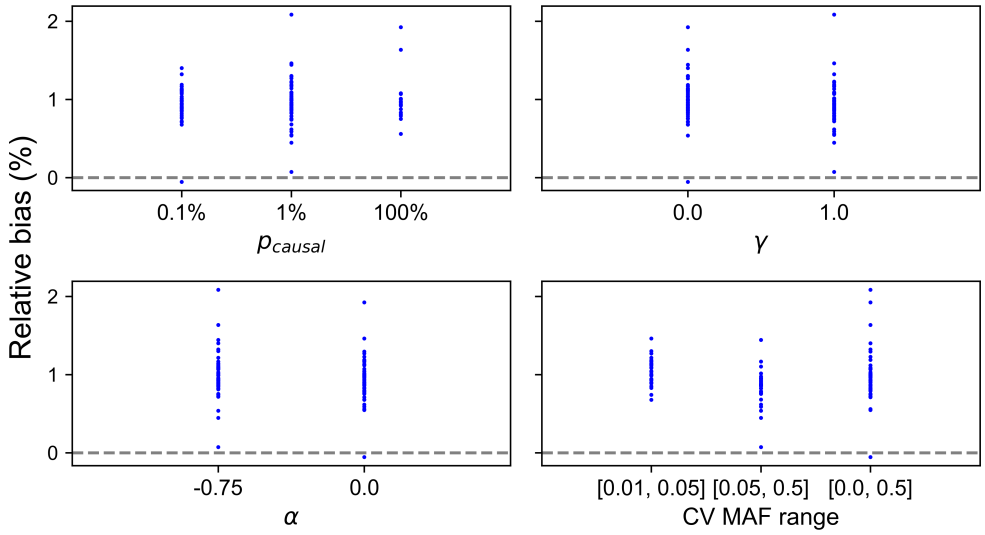
Supplementary Figure 3: $\widehat{h}^2_{\mathrm{GRE}}$ in simulations on chromosome 22 where a percentage of causal SNPs are masked from the observed summary statistics (p_{mask} = 0%, 50%, or 100%). "CV MAF low frequency" refers to CV MAF = $[0.01, 0.05]$. "CV MAF uniform" means causal variants were drawn uniformly from the chromosome 22 typed SNPs.



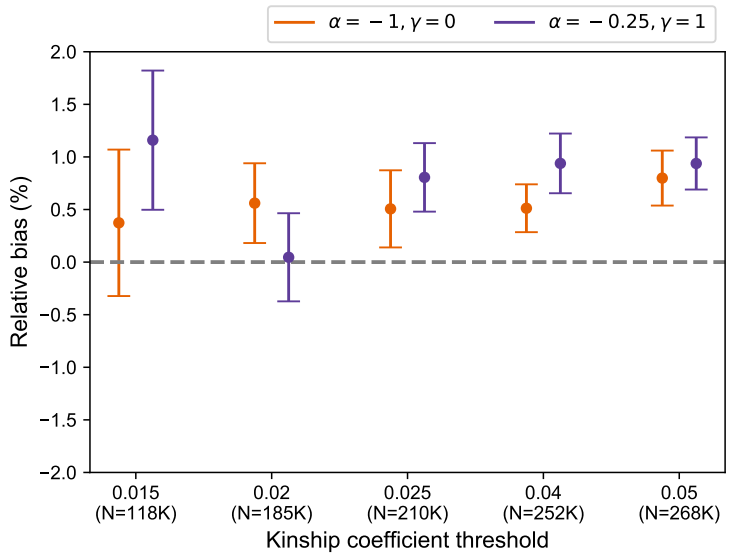
Supplementary Figure 4: Comparison of the analytical standard error of $\widehat{h}^2_{\mathrm{GRE}}$ with the standard deviation of $\widehat{h}^2_{\mathrm{GRE}}$ computed from 100 simulations ($h^2_g = 0.25$). (a) Phenotypes were simulated from SNPs on chromosome 22 ($N = 337205$, $M = 9564$ array SNPs) under one of 16 LDAK-LD- and/or MAF-dependent architectures and $\widehat{h}^2_{\mathrm{GRE}}$ was computed with a single chromosome-wide LD block. (b) Phenotypes were simulated from all genome-wide SNPs ($N = 337205$, $M = 593300$ array SNPs) under one of 28 LDAK-LD- and/or MAF-dependent architectures and $\widehat{h}^2_{\mathrm{GRE}}$ was computed with 22 chromosome-wide LD blocks. The colored bars represent the distribution of standard error estimates from 100 simulations. The red crosses mark the empirical standard deviation of the 100 estimates of h^2_g.



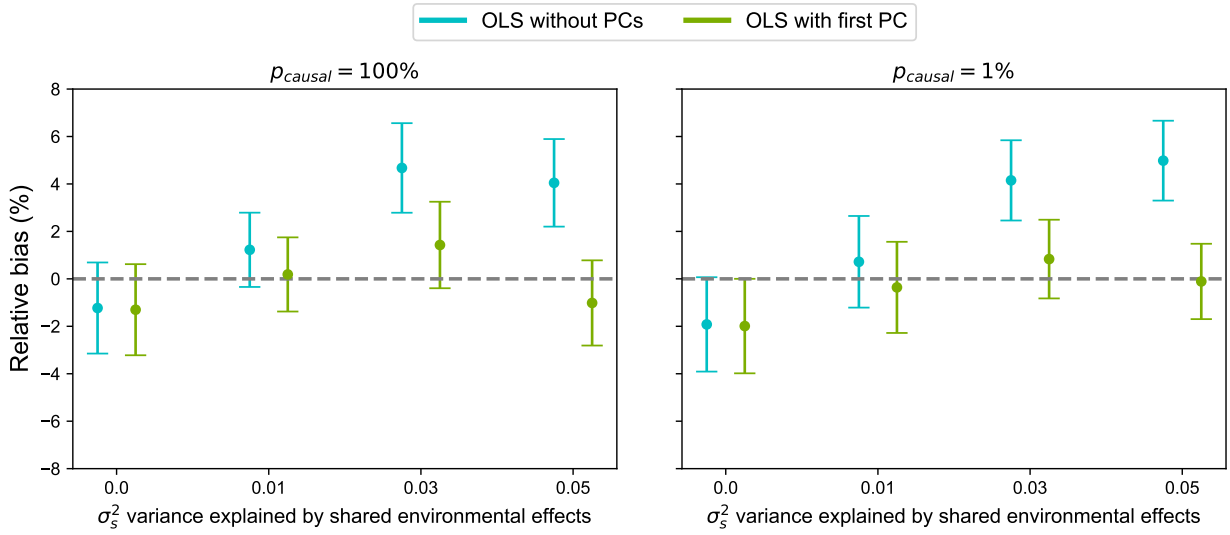
Supplementary Figure 5: Distribution of $\widehat{h}^2_{\text{GRE}}$ in simulations on chromosome 22 ($N = 337205$, $M = 9564$ array SNPs) as a function of the average size (Mb) of the LD blocks that were used to compute $\widehat{h}^2_{\text{GRE}}$. The largest block size (34.70 Mb) corresponds to using a single chromosome-wide LD block. All simulations were performed $h^2_g = 0.1$, $p_{\text{causal}} = 0.01$, $\alpha = -1$, and $\gamma = 0$ (no LD weights). Each boxplot represents 100 estimates.



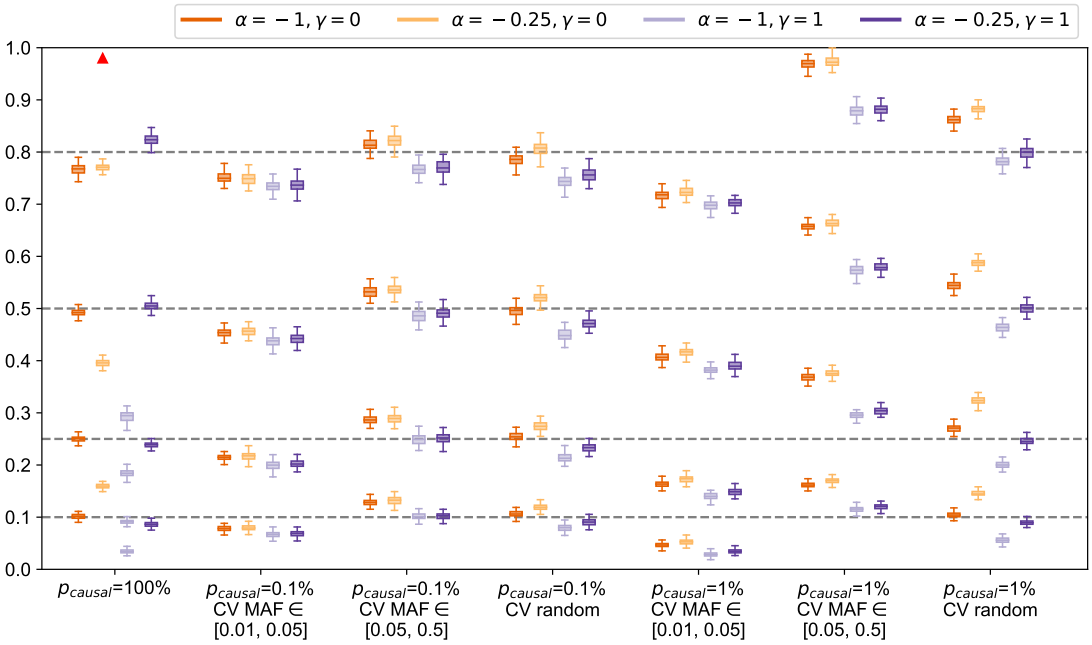
Supplementary Figure 6: Relative bias of $\widehat{h}^2_{\text{GRE}}$ in genome-wide simulations ($N = 337$K, $M = 593$K) with respect to different values of p_{causal}, α, γ, and CV MAF. Each point is the estimated relative bias of GRE (as a percentage of the simulated h^2_g) for a single architecture. Each plot contains the results from the same 64 architectures shown in Figure 1b and Supplementary Table S1b.



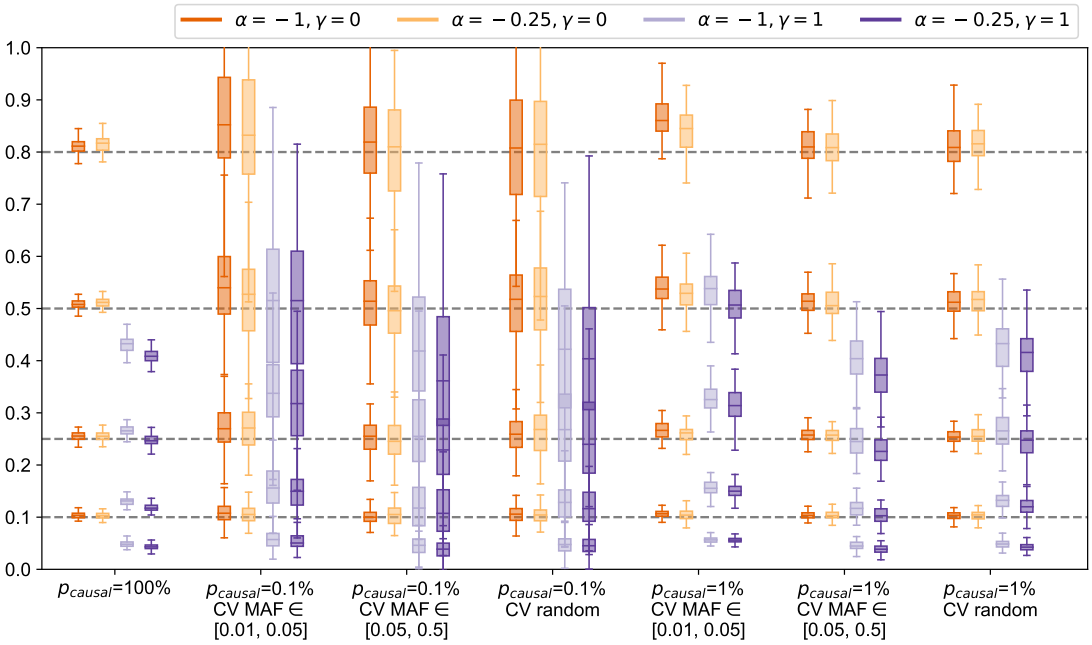
Supplementary Figure 7: Relative bias of $\widehat{h}^2_{\mathrm{GRE}}$ in genome-wide simulations ($M = 593$K) in which individuals were filtered at different kinship coefficient thresholds. Kinship matrix is defined as $\mathbf{X}\mathbf{X}^T/M$. Each point marks the relative bias (as a percentage of h^2_g) estimated from 100 independent simulations; bars represent ± 2 s.e.m. In all simulations, $h^2_g = 0.25$, $p_{\mathrm{causal}} = 1$, causal variants are drawn uniformly, and $\widehat{h}^2_{\mathrm{GRE}}$ is computed with 22 chromosome-wide blocks.



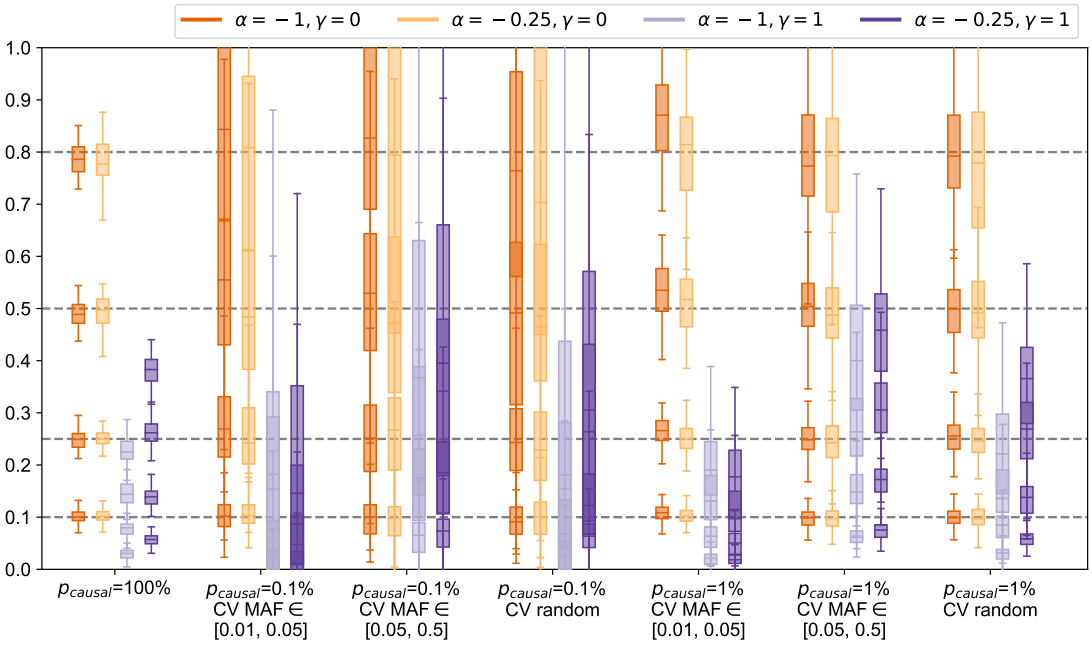
Supplementary Figure 8: Relative bias of $\widehat{h}^2_{\mathrm{GRE}}$ in genome-wide simulations ($N = 8430$, $M = 14821$) with population stratification (see Methods). σ^2_s is the proportion of total phenotypic variance explained by the covariate (i.e. the first genetic PC). Other simulation parameters are fixed ($h^2_g = 0.25$, $\alpha = -1$, $\gamma = 0$) and causal variants are drawn uniformly.



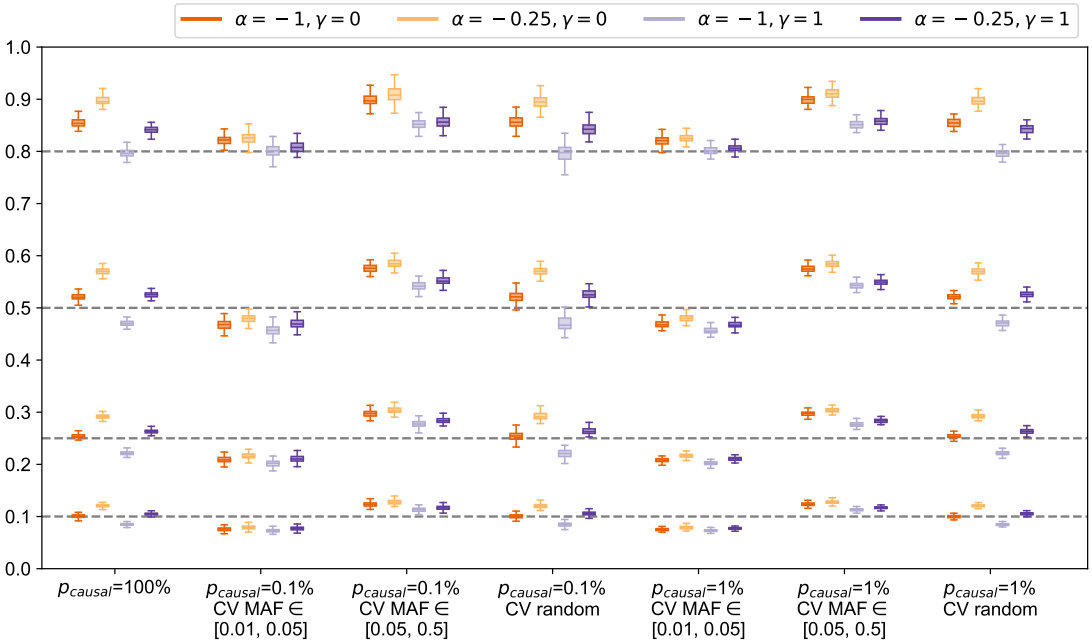
Supplementary Figure 9: Distribution of h_g^2 estimates from LDSC (no annotations) in simulations across 112 LDAK-LD- and/or MAF-dependent architectures ($N = 337205$ individuals, $M = 593300$ array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within $1.5 \times$ IQR from the first and third quartiles, respectively.



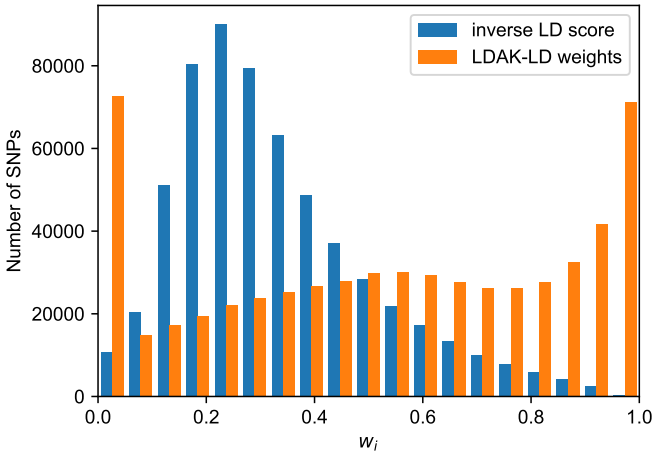
Supplementary Figure 10: Distribution of h_g^2 estimates from S-LDSC (10 MAF bins) in simulations across 112 LDAK-LD- and/or MAF-dependent architectures ($N = 337205$ individuals, $M = 593300$ array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within $1.5 \times$ IQR from the first and third quartiles, respectively.



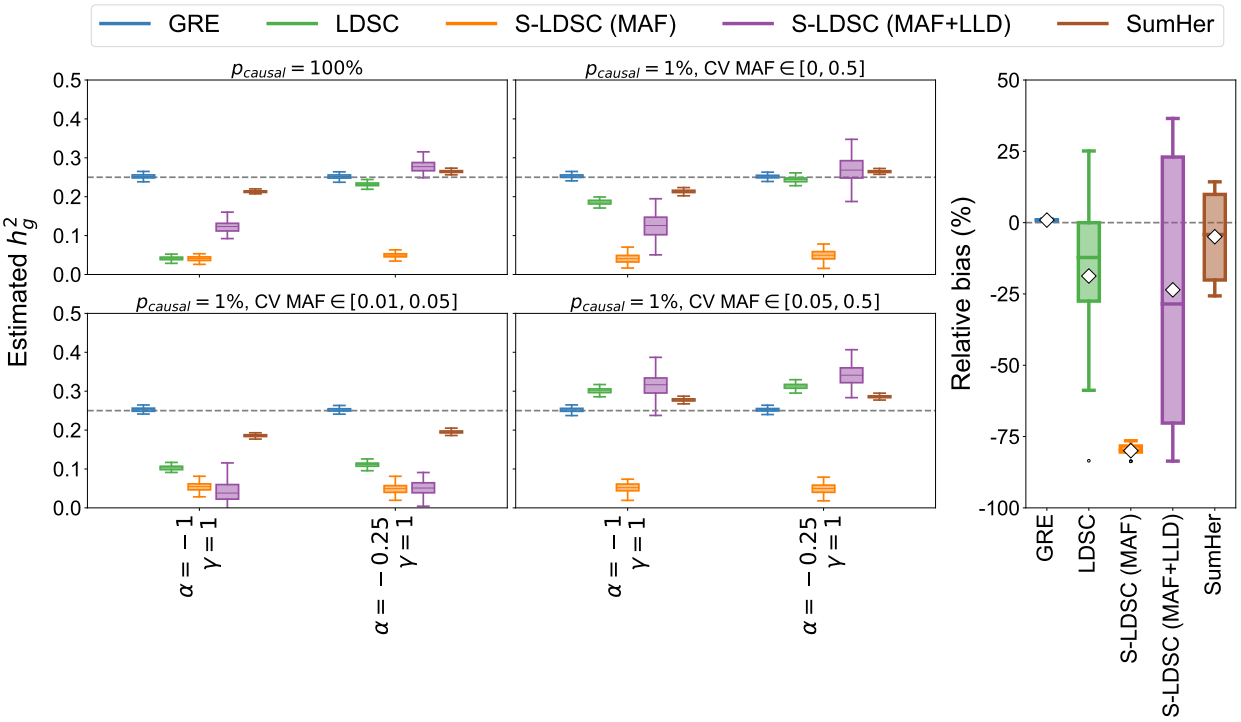
Supplementary Figure 11: Distribution of h_g^2 estimates from S-LDSC (10 MAF bins + LLD) in simulations across 112 LDAK-LD- and/or MAF-dependent architectures (N = 337205 individuals, M = 593300 array SNPs). Each boxplot shows the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



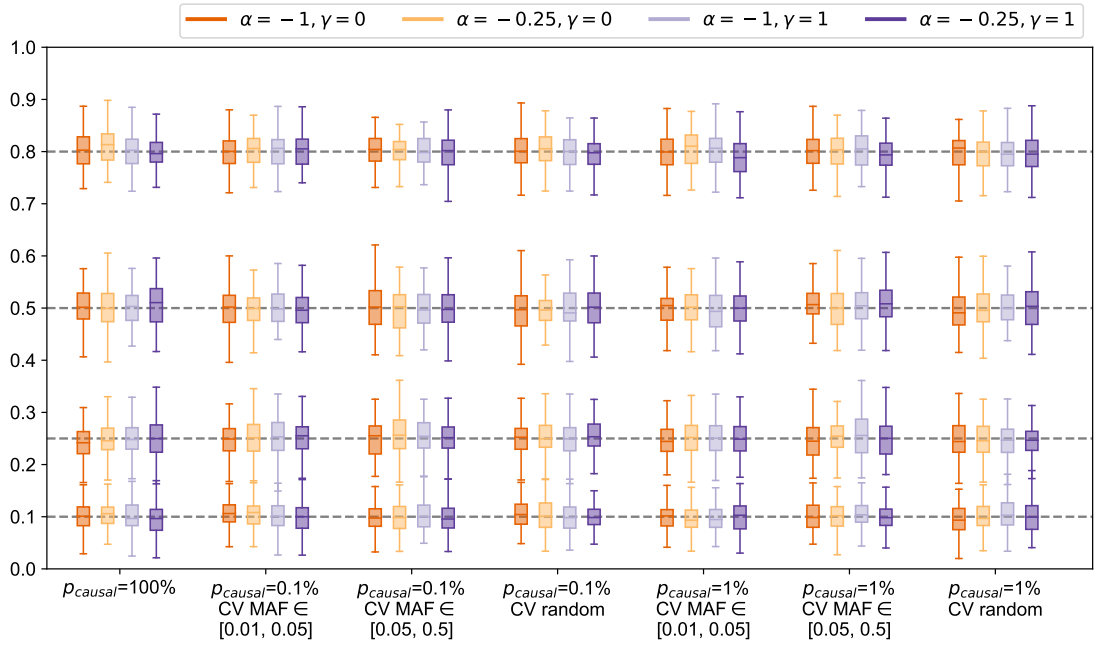
Supplementary Figure 12: Distribution of h_g^2 estimates from SumHer in simulations across 112 LDAK-LD- and/or MAF-dependent architectures (N = 337205 individuals, M = 593300 array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



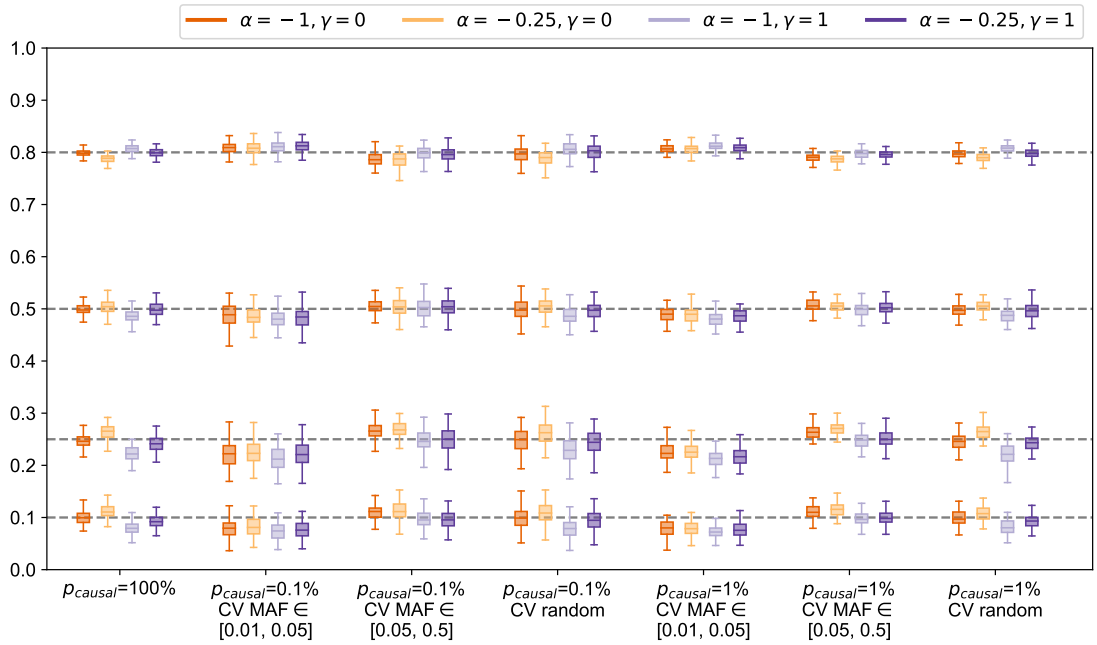
Supplementary Figure 13: Histograms of LDAK weights and inverse LD score weights used in genome-wide simulations (M = 593K SNPs).



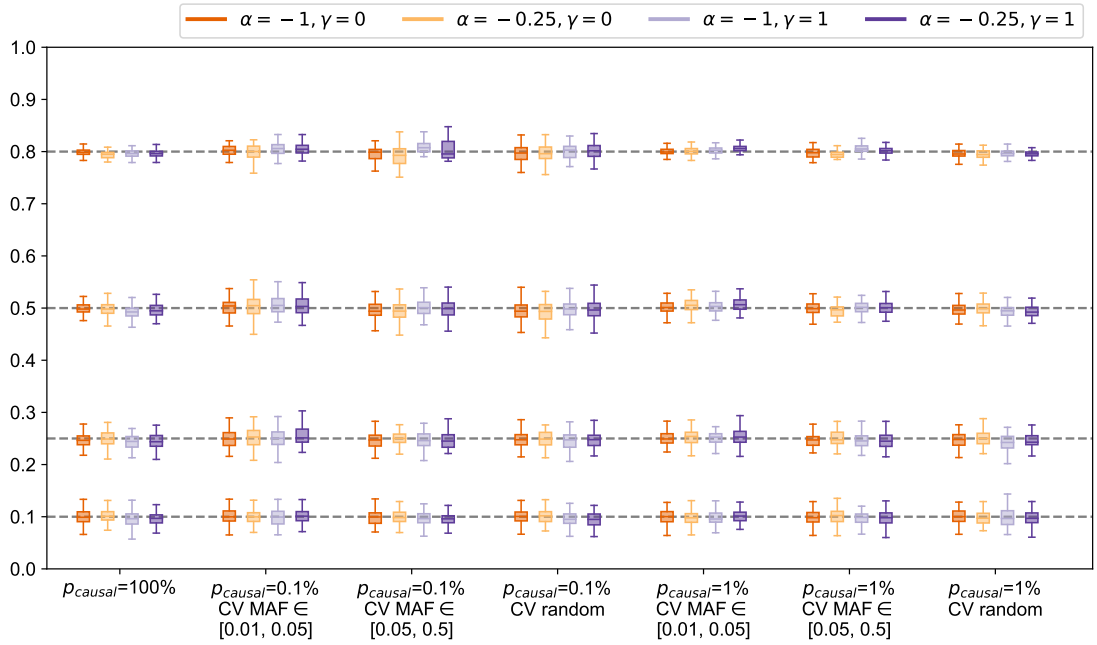
Supplementary Figure 14: Comparison of methods across 14 MAF- and LD-score-dependent architectures ($N = 337205$ individuals, $M = 593300$ array SNPs, $h_g^2 = 0.25$). LD-score-dependent architectures are simulated by coupling the variance of each SNP to the inverse of its LD score (Methods). **Left:** Each boxplot represents 100 estimates under a single architecture; results are shown for $p_{\text{causal}} = 100\%$ and 1%. **Right:** Each boxplot represents the distribution of the relative bias across all 14 LD-score-dependent architectures. White diamonds mark the average of each distribution. All boxplot whiskers mark the minimum and maximum estimates located within $1.5 \times$ IQR from the first and third quartiles, respectively.



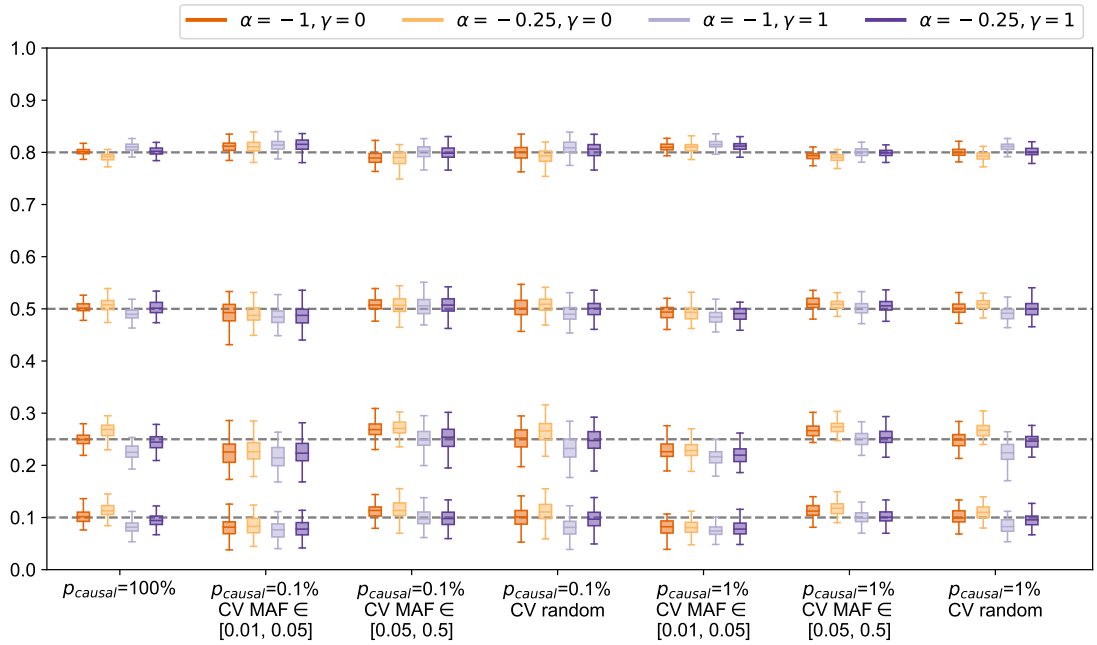
Supplementary Figure 15: Distribution of h_g^2 estimates from GRE in simulations across 112 LDAK-LD- and/or MAF-dependent architectures (N = 8430 individuals, M = 14821 array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



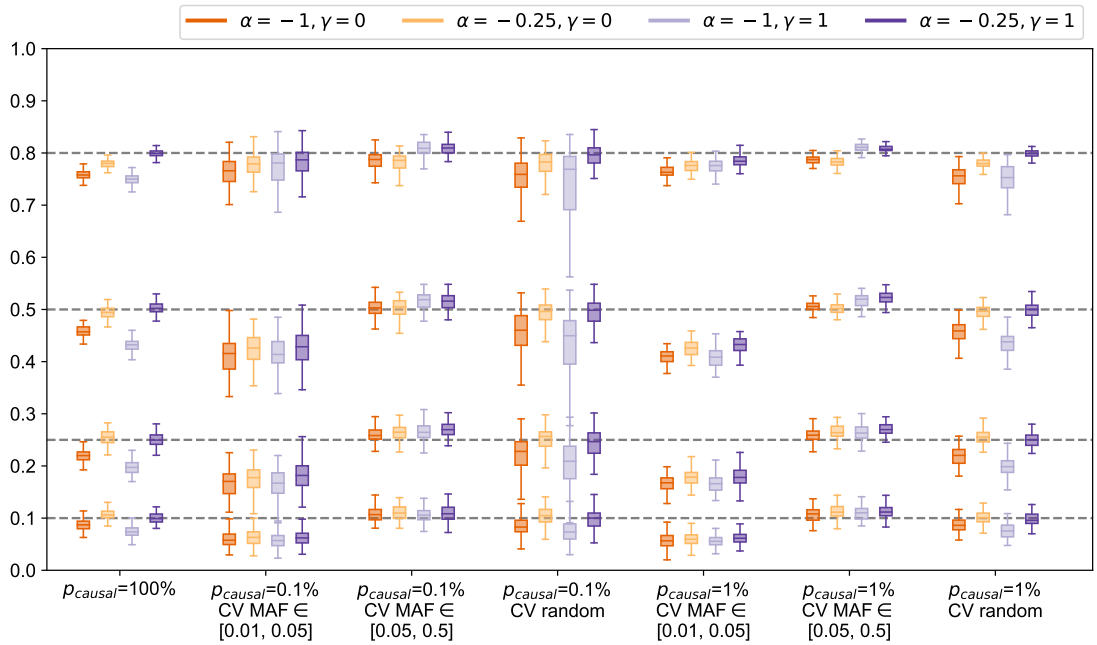
Supplementary Figure 16: Distribution of h_g^2 estimates from single-component GREML in simulations across 112 LDAK-LD- and/or MAF-dependent architectures (N = 8430 individuals, M = 14821 array SNPs). Each boxplot shows the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



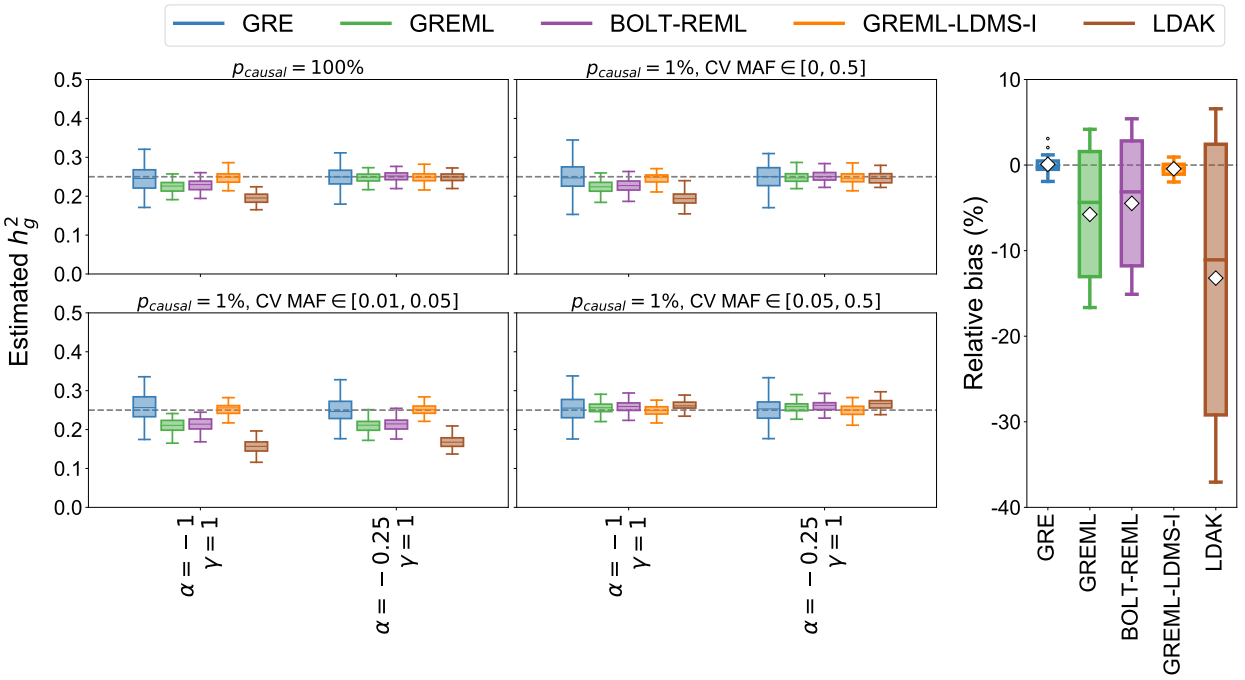
Supplementary Figure 17: Distribution of h_g^2 estimates from GREML-LDMS-I in simulations across 112 LDAK-LD- and/or MAF-dependent architectures ($N = 8430$ individuals, $M = 14281$ array SNPs). Each boxplot shows the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



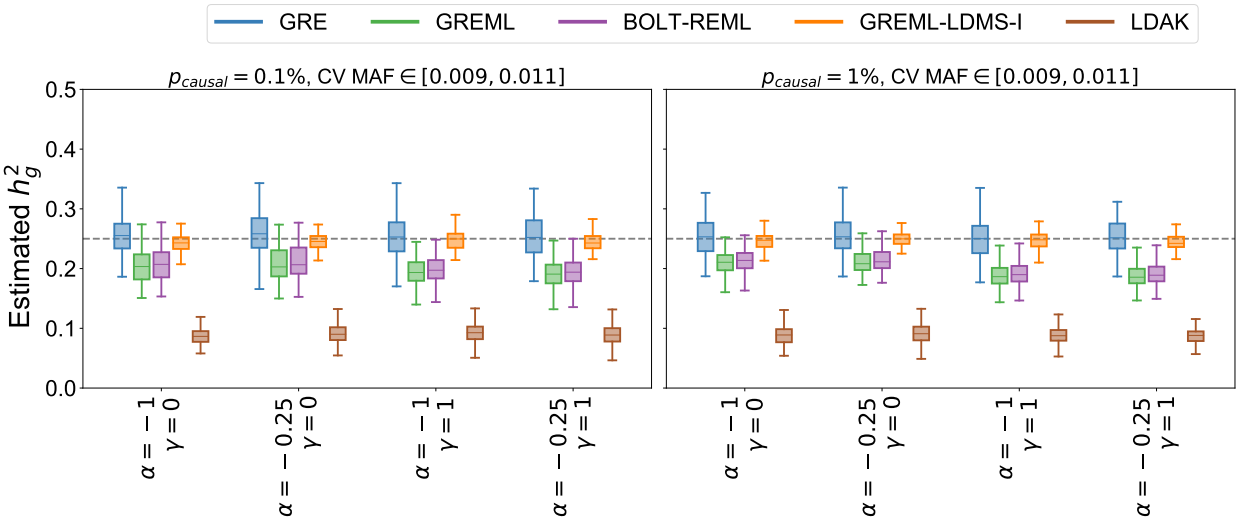
Supplementary Figure 18: Distribution of h_g^2 estimates from BOLT-REML in simulations across 112 LDAK-LD- and/or MAF-dependent architectures ($N = 8430$ individuals, $M = 14281$ array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



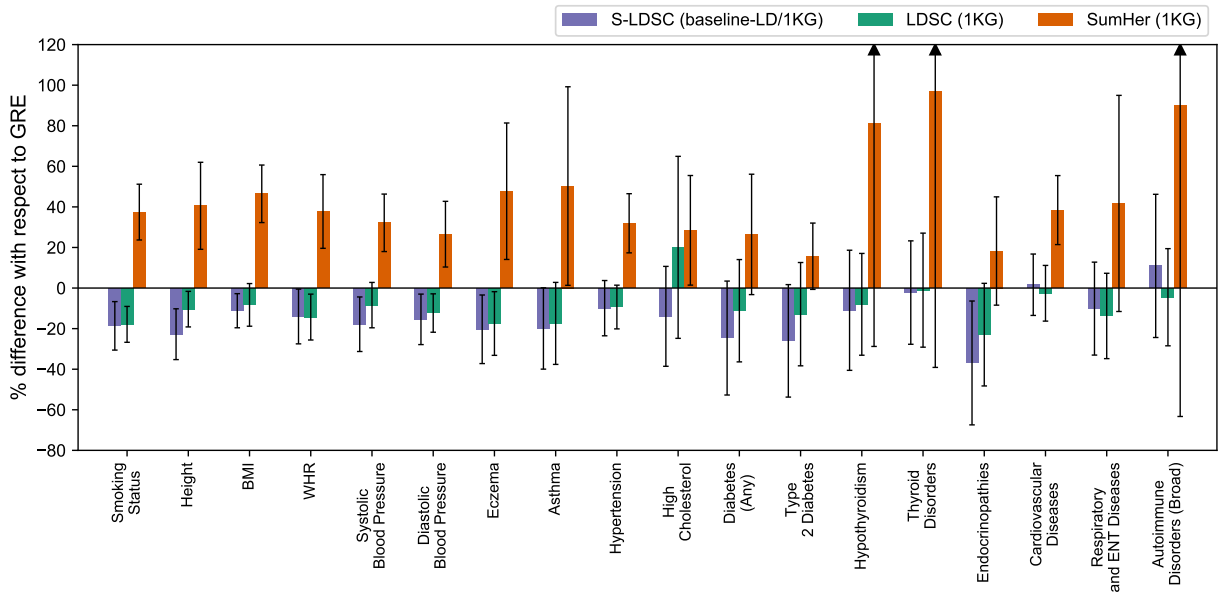
Supplementary Figure 19: Distribution of h_g^2 estimates from LDAK in simulations across 112 LDAK-LD- and/or MAF-dependent architectures ($N = 8430$ individuals, $M = 14281$ array SNPs). Each boxplot represents the distribution of 100 estimates under a single architecture. Boxplot whiskers extend to the minimum and maximum estimates located within 1.5 × IQR from the first and third quartiles, respectively.



Supplementary Figure 20: Comparison of methods across 14 MAF- and LD-score-dependent architectures ($N = 8430$ individuals, $M = 14281$ array SNPs). LD-score-dependent architectures are simulated by coupling the variance of each SNP to the inverse of its LD score (see Methods). **Left:** Each boxplot represents 100 estimates under a single architecture; results are shown for $p_{\text{causal}} = 100\%$ and 1%. **Right:** Each boxplot represents the distribution of the relative bias across all 14 LD-score-dependent architectures. White diamonds mark the average of each distribution. All boxplot whiskers mark the minimum and maximum estimates located within $1.5 \times$ IQR from the first and third quartiles, respectively.



Supplementary Figure 21: Comparison of GRE, GREML, BOLT-REML, GREML-LDMS-I, and LDAK in small-scale simulations ($N = 8430$ individuals, $M = 14821$ array SNPs) under MAF- and/or LDAK-LD-dependent architectures where all causal variants were drawn from the MAF range $[0.009, 0.011]$. Each boxplot contains estimates of h_g^2 from 100 simulations. The GRE estimator was computed with 22 chromosome-wide LD blocks. For GREML-LDMS-I, 8 GRMs were used (2 MAF bins $\times$ 4 LD quartiles). Boxplot whiskers mark the minimum and maximum estimates located within 1.5$\times$ IQR units from the first and third quartiles, respectively.



Supplementary Figure 22: Percent difference of SNP-heritability estimates from LDSC (1KG), S-LDSC (baseline-LD/1KG), and SumHer (1KG) with respect to $\widehat{h}^2_{\text{GRE}}$ for 18 complex traits and diseases in the UK Biobank for which $\widehat{h}^2_{\text{GRE}} > 0.05$ ($N = 290$K unrelated British individuals and $M = 460$K typed SNPs; see Methods). Each bar represents the difference between the estimated SNP-heritability and $\widehat{h}^2_{\text{GRE}}$ as a percentage of $\widehat{h}^2_{\text{GRE}}$. Black bars mark ± 2 standard errors.

References

[1] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004.