Manuscript Version: November 30, 2016

SUPPLEMENT TO "BAYESIAN LARGE-SCALE MULTIPLE REGRESSION WITH SUMMARY STATISTICS FROM GENOME-WIDE ASSOCIATION STUDIES"

By Xiang Zhu and Matthew Stephens

University of Chicago

CONTENTS

APPENDIX A: THEORETICAL DERIVATION OF RSS

A.1. Proofs of propositions in main text. We first summarize the assumptions that are made for the propositions in main text.

- The centered genotypes of n individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{x}$, where $\boldsymbol{x} := (x_1, \ldots, x_p)^\intercal$, $\mathrm{E}(\boldsymbol{x}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{x}) = \Sigma_x = \mathrm{diag}(\boldsymbol{\sigma}_x) R \mathrm{diag}(\boldsymbol{\sigma}_x)$ is finite, $|\mathrm{E}(x_j x_k x_l x_m)| < \infty$ for any $j, k, l, m \in [p]$. Note that these moment assumptions are satisfied by default for genotype data.
- The additive errors $\epsilon_1, \ldots, \epsilon_n \overset{\text{i.i.d.}}{\sim} \epsilon$, where $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \tau^{-1} < \infty$.
- The centered phenotypes of n individuals $y_1, \ldots, y_n \overset{\text{i.i.d.}}{\sim} y$, where $y = \boldsymbol{x}^\intercal \boldsymbol{\beta} + \epsilon$. For each individual $i \in [n]$, $y_i = \boldsymbol{x}_i^\intercal \boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{x}_i$, $\boldsymbol{\beta}$ and ϵ_i are mutually independent.

For all the asymptotic results, the convergence is established with $n \to \infty$ and p fixed.

A.1.1. *Proof of Proposition 2.1.* Notice that $\widehat{\boldsymbol{\beta}} = D^{-2} X^\intercal \mathbf{y}$, $\widehat{S} = \sqrt{n^{-1} \mathbf{y}^\intercal \mathbf{y}} \cdot D^{-1}$. If $\tau^{-1} = n^{-1} \mathbf{y}^\intercal \mathbf{y}$ and $\widehat{R} = \widehat{R}^{\mathsf{sam}}$, then $\widehat{S}^{-2} \widehat{\boldsymbol{\beta}} = \tau X^\intercal \mathbf{y}$ and $\widehat{S}^{-1} \widehat{R} \widehat{S}^{-1} = \tau X^\intercal X$. When $n > p$, the matrix X is full column rank and thus $\widehat{R} = \widehat{R}^{\mathsf{sam}}$ is non-singular, the full data and summary data likelihood are given by

$$\begin{aligned} -2 \log L_{\mathsf{mvn}}(\boldsymbol{\beta}; \mathbf{y}, X, \tau) &= p \log(2\pi\tau^{-1}) + \tau \mathbf{y}^\intercal \mathbf{y} - 2\tau \mathbf{y}^\intercal X \boldsymbol{\beta} + \tau \boldsymbol{\beta}^\intercal X^\intercal X \boldsymbol{\beta}, \\ -2 \log L_{\mathsf{rss}}(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}, \widehat{S}, \widehat{R}) &= p \log(2\pi) + \log |\widehat{S} \widehat{R} \widehat{S}| + \widehat{\boldsymbol{\beta}}^\intercal (\widehat{S} \widehat{R} \widehat{S})^{-1} \widehat{\boldsymbol{\beta}} - 2 \widehat{\boldsymbol{\beta}}^\intercal \widehat{S}^{-2} \boldsymbol{\beta} + \boldsymbol{\beta}^\intercal \widehat{S}^{-1} \widehat{R} \widehat{S}^{-1} \boldsymbol{\beta}, \end{aligned}$$

respectively, and their difference does not depend on the parameter of interest $\boldsymbol{\beta}$,

$$-2[\log L_{\mathsf{rss}}(\boldsymbol{\beta}; \widehat{\boldsymbol{\beta}}, S, R) - \log L_{\mathsf{mvn}}(\boldsymbol{\beta}; \mathbf{y}, X)] = \log |D^{-1} \widehat{R} D^{-1}| - \tau \mathbf{y}^\intercal [I - X(X^\intercal X)^{-1} X^\intercal] \mathbf{y}, \tag{A.1}$$

implying that these two likelihoods of $\boldsymbol{\beta}$ are equivalent.

A.1.2. *Proof of Proposition 2.2.* First define the statistic $T_n \in \mathbb{R}^{2p \times 1}$,

$$T_n := n^{-1} \left(\textstyle\sum_{i=1}^n x_{i1} y_i, \ldots, \sum_{i=1}^n x_{ip} y_i, \sum_{i=1}^n x_{i1}^2, \ldots, \sum_{i=1}^n x_{ip}^2 \right)^\intercal. \tag{A.2}$$

The asymptotic distribution of T_n is given by the multivariate Central Limit Theorem

$$\sqrt{n}(T_n - \boldsymbol{\mu}_T) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_T), \tag{A.3}$$

where $\boldsymbol{\mu}_T := \mathrm{E}(\mathbf{t})$, $\Sigma_T := \mathrm{Var}(\mathbf{t})$ and $\mathbf{t} := (x_1 y, \ldots, x_p y, x_1^2, \ldots, x_p^2)^\intercal$. Note that Σ_T has finite entries because τ^{-1}, Σ_x and $\mathrm{E}(x_j x_k x_l x_m)$ are finite.

Next, for any $\boldsymbol{\xi} \in \mathbb{R}^{2p \times 1}$, define the following function $g(\boldsymbol{\xi}) \in \mathbb{R}^{p \times 1}$:

$$g(\boldsymbol{\xi}) := (\xi_1 / \xi_{p+1}, \ldots, \xi_p / \xi_{2p})^\intercal. \tag{A.4}$$

Note that $g(T_n) = \widehat{\boldsymbol{\beta}}$ and $g(\boldsymbol{\mu}_T) = \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x) \boldsymbol{\mu}_{xy} = \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x) R \mathrm{diag}(\boldsymbol{\sigma}_x) \boldsymbol{\beta}$.

Use the multivariate Delta method to show that

$$\sqrt{n}(g(T_n) - g(\boldsymbol{\mu}_T)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla^\intercal g(\boldsymbol{\mu}_T) \Sigma_T \nabla g(\boldsymbol{\mu}_T)) \tag{A.5}$$

where $\nabla g(\boldsymbol{\mu}_T) \in \mathbb{R}^{2p \times p}$ is the gradient matrix of g at $\boldsymbol{\mu}_T$. A straightforward calculation yields that

$$\nabla^\intercal g(\boldsymbol{\mu}_T) \Sigma_T \nabla g(\boldsymbol{\mu}_T) = \sigma_y^2 \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)(R + \Delta(\mathbf{c})) \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x). \tag{A.6}$$

The explicit form of $\Delta(\mathbf{c})$ is given by

$$\Delta(\mathbf{c}) := \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x) \cdot [G_1(\mathbf{c}) + G_2(\mathbf{c}) + G_2^\intercal(\mathbf{c}) + G_3(\mathbf{c})] \cdot \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x), \tag{A.7}$$

where functions $G_i(\mathbf{c}) : \mathbb{R}^{p\times 1} \mapsto \mathbb{R}^{p\times p}$ are defined as follows:

$$\begin{aligned} G_1(\mathbf{c}) &:= -(\mathbf{c}^\intercal R^{-1}\mathbf{c})\Sigma_x - \mathrm{diag}(\boldsymbol{\sigma}_x)\mathbf{c}\mathbf{c}^\intercal\mathrm{diag}(\boldsymbol{\sigma}_x) + \mathrm{E}[(\boldsymbol{x}^\intercal\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)R^{-1}\mathbf{c})^2\boldsymbol{x}\boldsymbol{x}^\intercal], \\ G_2(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\mathrm{diag}(\mathbf{c})W(\mathbf{c}),\ \ [W(\mathbf{c})]_{ij} := \sigma_{x,i}\sigma_{x,j}^2 c_i - \mathbf{c}^\intercal R^{-1}\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\mathrm{E}(x_i x_j^2\boldsymbol{x}), \\ G_3(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\ \mathrm{diag}(\mathbf{c})\ \Sigma_{xx}\ \mathrm{diag}\ (\mathbf{c})\ \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x),\ \ [\Sigma_{xx}]_{ij} := \mathrm{Cov}(x_i^2, x_j^2). \end{aligned}$$

Notice that $G_i(\mathbf{c})$ are continuous functions of $\mathbf{c}$, $G_i(\mathbf{0}) = \mathbf{0}$, and $G_i(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$ for $i = 1, 2, 3$.

A.1.3. *Proof of Proposition 2.3.* First note that $SRS^{-1} = \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)R\mathrm{diag}(\boldsymbol{\sigma}_x)$. Hence,

$$\begin{aligned} &\log\mathcal{N}(\widehat{\boldsymbol{\beta}}; SRS^{-1}\boldsymbol{\beta}, SRS) - \log\mathcal{N}(\widehat{\boldsymbol{\beta}}; \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)R\mathrm{diag}(\boldsymbol{\sigma}_x)\boldsymbol{\beta}, n^{-1}\Sigma) \\ &= \frac{1}{2}\left\{\log|R + \Delta(\mathbf{c})| - \log|R| + \sigma_y^{-2}\boldsymbol{\lambda}^\intercal\mathrm{diag}(\boldsymbol{\sigma}_x)[(R + \Delta(\mathbf{c}))^{-1} - R^{-1}]\mathrm{diag}(\boldsymbol{\sigma}_x)\boldsymbol{\lambda}\right\}, \end{aligned} \tag{A.8}$$

where $\boldsymbol{\lambda} := \sqrt{n}(\widehat{\boldsymbol{\beta}} - SRS^{-1}\boldsymbol{\beta})$. Since the determinant and inverse of a matrix are both continuous, we invoke Proposition 2.2, essentially, $\boldsymbol{\lambda} = \mathcal{O}_p(1)$ and $\Delta(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$, to complete the proof.

A.1.4. *Proof of Proposition 3.1.* Since the matrix X is column-centered,

$$V(X\boldsymbol{\beta}) = n^{-1}\textstyle\sum_{i=1}^n(\boldsymbol{x}_i^\intercal\boldsymbol{\beta})^2 = n^{-1}\mathrm{trace}[(X\boldsymbol{\beta})(X\boldsymbol{\beta})^\intercal] = n^{-1}\boldsymbol{\beta}^\intercal X^\intercal X\boldsymbol{\beta}, \tag{A.9}$$

and therefore,

$$\mathrm{E}[V(X\boldsymbol{\beta})|S, X] = \boldsymbol{\mu}_\beta^\intercal\cdot(n^{-1}X^\intercal X)\cdot\boldsymbol{\mu}_\beta + \mathrm{trace}[(n^{-1}X^\intercal X)\cdot\Sigma_\beta], \tag{A.10}$$

where $\boldsymbol{\mu}_\beta := \mathrm{E}(\boldsymbol{\beta}|S) = \mathbf{0}$ and $\Sigma_\beta := \mathrm{Var}(\boldsymbol{\beta}|S) = (\pi\sigma_B^2 + \sigma_P^2)\cdot I_p$. Hence,

$$\mathrm{E}[V(X\boldsymbol{\beta})] = \mathrm{E}[\mathrm{E}[V(X\boldsymbol{\beta})|S, X]] = (\pi\sigma_B^2 + \sigma_P^2)\cdot\textstyle\sum_{j=1}^p\mathrm{E}[V(X_j)] = \dfrac{h}{\sum_{j=1}^p n^{-1}s_j^{-2}}\cdot\sum_{j=1}^p\mathrm{E}[V(X_j)].$$

From the definition of $\{s_j\}$ we can see that $\mathrm{E}[V(X_j)] = n^{-1}s_j^{-2}\mathrm{E}[V(\mathbf{y})]$, implying that

$$\mathrm{E}[V(X\boldsymbol{\beta})] = \frac{h}{\sum_{j=1}^p n^{-1}s_j^{-2}}\cdot\sum_{j=1}^p n^{-1}s_j^{-2}\mathrm{E}[V(\mathbf{y})] = h\cdot\mathrm{E}[V(\mathbf{y})]. \tag{A.11}$$

A.2. Extension of RSS: data on different individuals. The RSS likelihood assumes that the univariate summary data are computed from the same set of individuals, but this assumption is often violated in GWAS (Section 5.1, main text). Here we modify the RSS likelihood for the summary data generated from different individuals.

Suppose that for each SNP j, its single-SNP summary statistics $\{\hat{\beta}_j, \hat{\sigma}_j^2\}$ are computed on a *predefined, nonempty* subset of individuals $\mathcal{I}_j \subseteq [n]$:

$$\hat{\beta}_j(\mathcal{I}_j; X_j, \mathbf{y}) := \left(\textstyle\sum_{i\in\mathcal{I}_j}x_{ij}^2\right)^{-1}\left(\sum_{i\in\mathcal{I}_j}x_{ij}y_i\right), \tag{A.12}$$

$$\hat{\sigma}_j^2(\mathcal{I}_j; X_j, \mathbf{y}) := \left(|\mathcal{I}_j|\cdot\textstyle\sum_{i\in\mathcal{I}_j}x_{ij}^2\right)^{-1}\left[\sum_{i\in\mathcal{I}_j}\left(y_i - x_{ij}\hat{\beta}_j\right)^2\right], \tag{A.13}$$

where $|\cdot|$ denotes the cardinality of a set. Let $\mathcal{I} := \{\mathcal{I}_1, \ldots, \mathcal{I}_p\}$ and $\widehat{\boldsymbol{\beta}}(\mathcal{I}; X, \mathbf{y}) \in \mathbb{R}^p$, whose jth element is $\hat{\beta}_j(\mathcal{I}_j; X_j, \mathbf{y})$.

Let $\widehat{F}(\mathcal{I}; X, \mathbf{y}) := \operatorname{diag}\left(\widehat{\mathbf{f}}(\mathcal{I}; X, \mathbf{y})\right)$, $\widehat{\mathbf{f}} \in \mathbb{R}^p$, whose jth element is

$$\hat{f}_j(\mathcal{I}_j; X_j, \mathbf{y}) := \sqrt{\left(\textstyle\sum_{i\in\mathcal{I}_j} x_{ij}^2\right)^{-1}\left(\textstyle\sum_{i\in\mathcal{I}_j} y_i^2\right)}. \tag{A.14}$$

Note that $\hat{f}_j^2$ is the estimated ratio of phenotypic and genotypic variance at SNP j (i.e. $\sigma_y^2/\sigma_{x,j}^2$), and it can be computed from the single-SNP summary statistics (A.12, A.13),

$$\hat{f}_j^2(\mathcal{I}_j; X_j, \mathbf{y}) = |\mathcal{I}_j| \cdot \hat{\sigma}_j^2(\mathcal{I}_j; X_j, \mathbf{y}) + \hat{\beta}_j^2(\mathcal{I}_j; X_j, \mathbf{y}). \tag{A.15}$$

We omit the index $\mathcal{I}$ labeling subsets in the following discussion.

Finally we introduce a matrix H to reflect the proportions of sample overlap among different SNPs. Specifically, the (i,j)-entry of H is defined as $H_{ij} := (|\mathcal{I}_i| \cdot |\mathcal{I}_j|)^{-\frac{1}{2}}|\mathcal{I}_i \cap \mathcal{I}_j|$. Note that the diagonals of H are all 1; the other entries are between 0 and 1. For any pair of SNPs (i,j), $H_{ij} = 1$ if and only if $\mathcal{I}_i = \mathcal{I}_j$ (the same set of individuals); $H_{ij} = 0$ if and only if $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ (two disjoint sets of individuals).

With this in place, the modified RSS likelihood of $\boldsymbol{\beta}$ is given by:

$$L_{\text{rss}}^{\text{subset}}(\boldsymbol{\beta}) := \mathcal{N}(\widehat{\boldsymbol{\beta}};\ \widehat{F}\widehat{R}\widehat{F}^{-1}\boldsymbol{\beta},\ N^{-\frac{1}{2}}\widehat{F}(H \circ \widehat{R})\widehat{F}N^{-\frac{1}{2}}). \tag{A.16}$$

where $N := \operatorname{diag}(\mathbf{n})$, $\mathbf{n} := (|\mathcal{I}_1|, \ldots, |\mathcal{I}_p|)^\intercal$, and $H \circ \widehat{R}$ is the Hadamard product. Note that the modified likelihood (A.16) includes the original RSS likelihood as a special case. To see this, when $\mathcal{I}_j = [n]$ for all SNP j, $\widehat{F} = \sqrt{n}\widehat{S}$, $N = nI_p$ and H is an all-one matrix, yielding

	general form	simple form	
(A.17)	$\widehat{F}\widehat{R}\widehat{F}^{-1}\boldsymbol{\beta} = \widehat{S}\widehat{R}\widehat{S}^{-1}\boldsymbol{\beta}$		(mean vector),
(A.18)	$N^{-\frac{1}{2}}\widehat{F}(H \circ \widehat{R})\widehat{F}N^{-\frac{1}{2}} = \widehat{S}\widehat{R}\widehat{S}$		(covariance matrix).

However, the relations (A.17, A.18) do not hold when the summary data are not generated from the same sample. These differences, especially in the mean (A.17), are often omitted by previous work.

The likelihood (A.16) is derived from the Propositions A.1 and A.2 below. Similar to the RSS likelihood, (A.16) is obtained by replacing the nuisance parameters $\{F, R\}$ with the estimates $\{\widehat{F}, \widehat{R}\}$.

PROPOSITION A.1. Let $\boldsymbol{\pi} := (\pi_1, \ldots, \pi_p)^\intercal$, where $\pi_j := |\mathcal{I}_j|/n$. Assume that both H and $\boldsymbol{\pi}$ are non-random and do not depend on n. For any subsets $\mathcal{I} := \{\mathcal{I}_1, \ldots, \mathcal{I}_p\}$,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}(\mathcal{I}; X, \mathbf{y}) - FRF^{-1}\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma^*). \tag{A.19}$$

where $\Sigma^* := (\Pi^{-\frac{1}{2}}F) \cdot [H \circ (R + \Delta(\mathbf{c}))] \cdot (\Pi^{-\frac{1}{2}}F)^\intercal$, $F := \sigma_y \operatorname{diag}^{-1}(\boldsymbol{\sigma}_x)$, $\Pi := \operatorname{diag}(\boldsymbol{\pi})$ and $\Delta(\mathbf{c})$ is defined by (A.7).

PROOF. First define the statistic $T_n^* \in \mathbb{R}^{2p\times 1}$,

$$T_n^* := n^{-1}\left(\textstyle\sum_{i=1}^n m_{i1}x_{i1}y_i, \ldots, \sum_{i=1}^n m_{ip}x_{ip}y_i, \sum_{i=1}^n m_{i1}x_{i1}^2, \ldots, \sum_{i=1}^n m_{ip}x_{ip}^2\right)^\intercal, \tag{A.20}$$

where $m_{ij} := \mathbf{1}\{i \in \mathcal{I}_j\}$, indicating whether the genotype of Individual i is used to compute the summary statistics of SNP j. Here we assume that the subsets $\{\mathcal{I}_j\}$ are *pre-defined* so that the indicators $\{m_{ij}\}$ are *non-random* constants.

Notice that $T_n^* = n^{-1}\sum_{i=1}^n \mathbf{t}_i^*$, where

$$\mathbf{t}_i^* := (m_{i1}, \ldots, m_{ip}, m_{i1}, \ldots, m_{ip})^\intercal \circ \mathbf{t}_i \tag{A.21}$$

$$\mathbf{t}_i := (x_{i1}y_i, \ldots, x_{ip}y_i, x_{i1}^2, \ldots, x_{ip}^2)^\intercal \tag{A.22}$$

From the proof of Proposition 2.2 (Section A.1.2), we know that $\mathbf{t}_i$'s are i.i.d. draws from $\mathbf{t}$ with mean $\boldsymbol{\mu}_T$ and covariance matrix Σ_T. Hence, T_n^* is a sum of independent but non-identical random vectors, and its asymptotic distribution is given by the multivariate Lindeberg-Feller Central Limit Theorem [e.g. Appendix D, Greene (2012)]

$$\sqrt{n}(T_n^* - \boldsymbol{\mu}_T^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma_T^*), \tag{A.23}$$

where the asymptotic mean and covariance matrix are given by

$$\boldsymbol{\mu}_T^* := (I_2 \otimes \Pi) \cdot \boldsymbol{\mu}_T,\ \Sigma_T^* := \left(J_2 \otimes \left(\Pi^{\frac{1}{2}} \cdot H \cdot \Pi^{\frac{1}{2}}\right)\right) \circ \Sigma_T, \tag{A.24}$$

with I_2 and J_2 denoting the 2×2 identity and all-ones matrix respectively, and $\otimes$ denoting their Kronecker product.

Next, use the multivariate Delta method and to show that

$$\sqrt{n}(g(T_n^*) - g(\boldsymbol{\mu}_T^*)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \nabla^\intercal g(\boldsymbol{\mu}_T^*)\Sigma_T^* \nabla g(\boldsymbol{\mu}_T^*)), \tag{A.25}$$

where the function $g(\cdot)$ is defined in (A.4) and $\nabla g(\boldsymbol{\mu}_T^*)$ is the gradient of g at $\boldsymbol{\mu}_T^*$. A straightforward calculation yields that $g(T_n^*) = \widehat{\boldsymbol{\beta}}(\mathcal{I}; X, \mathbf{y})$, $g(\boldsymbol{\mu}_T^*) = FRF^{-1}\boldsymbol{\beta}$ and

$$\nabla^\intercal g(\boldsymbol{\mu}_T^*)\Sigma_T^* \nabla g(\boldsymbol{\mu}_T^*) = (\Pi^{-\frac{1}{2}}F) \cdot [H \circ (R + \Delta(\mathbf{c}))] \cdot (\Pi^{-\frac{1}{2}}F)^\intercal, \tag{A.26}$$

where $\Delta(\mathbf{c})$ is defined by (A.7). □

Proposition A.2. For each $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\log \mathcal{N}(\widehat{\boldsymbol{\beta}}; FRF^{-1}\boldsymbol{\beta}, N^{-\frac{1}{2}}F(H \circ R)FN^{-\frac{1}{2}}) - \log \mathcal{N}(\widehat{\boldsymbol{\beta}}; FRF^{-1}\boldsymbol{\beta}, n^{-1}\Sigma^*) = \mathcal{O}_p(\max_j c_j^2).$$

Proof. A straightforward calculation yields that

$$\begin{aligned} & \log \mathcal{N}(\widehat{\boldsymbol{\beta}}; FRF^{-1}\boldsymbol{\beta}, N^{-\frac{1}{2}}F(H \circ R)FN^{-\frac{1}{2}}) - \log \mathcal{N}(\widehat{\boldsymbol{\beta}}; FRF^{-1}\boldsymbol{\beta}, n^{-1}\Sigma^*) \\ = & \frac{1}{2}\left\{\log|H \circ (R + \Delta(\mathbf{c}))| - \log|H \circ R| + \boldsymbol{\lambda}^\intercal \Pi^{\frac{1}{2}} F^{-1}[(H \circ (R + \Delta(\mathbf{c})))^{-1} - (H \circ R)^{-1}]F^{-1}\Pi^{\frac{1}{2}}\boldsymbol{\lambda}\right\}, \end{aligned}$$

where $\boldsymbol{\lambda} := \sqrt{n}(\widehat{\boldsymbol{\beta}} - FRF^{-1}\boldsymbol{\beta})$. Since the determinant and inverse of a matrix are both continuous, we invoke Proposition A.1, essentially, $\boldsymbol{\lambda} = \mathcal{O}_p(1)$ and $\Delta(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$, to complete the proof. □

A.3. Extension of RSS: imputation error. The RSS likelihood assumes that the GWAS summary data are computed at fully observed genotypes. In GWAS, however, not all SNPs are directly assayed, and the (missing) genotypes of untyped SNPs are obtained by imputation. Here we modify the RSS likelihood for the summary data generated from imputed genotypes.

We first outline the assumptions used in later derivations.

- The true (centered) genotypes of n individuals $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_n^* \overset{\text{i.i.d.}}{\sim} \boldsymbol{x}^*$, where $\boldsymbol{x}^* := (x_1^*, \ldots, x_p^*)^\intercal$, $\mathrm{E}(\boldsymbol{x}^*) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{x}^*) = \Sigma_x^* = \mathrm{diag}(\boldsymbol{\sigma}_x^*)R^*\mathrm{diag}(\boldsymbol{\sigma}_x^*)$, and $\boldsymbol{\sigma}_x^* := (\sigma_{x,1}^*, \ldots, \sigma_{x,p}^*)^\intercal$.
- The imputed (centered) genotypes of n individuals $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{x}$, where $\boldsymbol{x} := (x_1, \ldots, x_p)^\intercal$, $\mathrm{E}(\boldsymbol{x}) = \mathbf{0}$, $\mathrm{Var}(\boldsymbol{x}) = \Sigma_x = \mathrm{diag}(\boldsymbol{\sigma}_x)R\mathrm{diag}(\boldsymbol{\sigma}_x)$, and $\boldsymbol{\sigma}_x := (\sigma_{x,1}, \ldots, \sigma_{x,p})^\intercal$.
- The imputed and true genotypes follow the measurement error model: $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{\eta}$, where $\mathrm{E}(\boldsymbol{\eta}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\eta}) = \Sigma_\eta$. Note that the diagonal elements of Σ_η (i.e. variances of $\boldsymbol{\eta}$) reflect the imputation quality of each SNP. Large variance indicates that the SNP is poorly imputed.

- The (centered) phenotypes of n individuals $y_1, \ldots, y_n \overset{\text{i.i.d.}}{\sim} y$, where $y = (\boldsymbol{x}^*)^\intercal \boldsymbol{\beta} + \epsilon$, $\mathrm{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) = \tau^{-1}$. Note that the coefficients $\boldsymbol{\beta}$ are the effects of each SNP on phenotype based on the true genotypes, *not* the imputed genotypes.
- The true genotype $\boldsymbol{x}^*$, measurement error $\boldsymbol{\eta}$ and residual error ϵ are mutually independent.
- The summary statistics $\{\widehat{\boldsymbol{\beta}}, \widehat{S}\}$ are computed from the imputed genotypes $\{\boldsymbol{x}_i\}$.

With this in place, the modified RSS likelihood of $\boldsymbol{\beta}$ is given by:

$$L_{\mathsf{rss}}^{\mathsf{impute}}(\boldsymbol{\beta}) := \mathcal{N}(\widehat{\boldsymbol{\beta}}; (\widehat{S}\widehat{R}\widehat{S}^{-1} - \mathrm{diag}^{-2}(\widehat{\boldsymbol{\sigma}}_x)\Sigma_\eta)\boldsymbol{\beta}, \widehat{S}\widehat{R}\widehat{S}), \tag{A.27}$$

where $\widehat{R}$ and $\widehat{\boldsymbol{\sigma}}_x$ are the estimates of R and $\boldsymbol{\sigma}_x$ respectively. Note that the modified likelihood (A.27) includes the original RSS likelihood as a special case. This is because when all SNPs are directly genotyped, Σ_η is an all-zero matrix (i.e. the measurement error $\boldsymbol{\eta}$ is zero).

The modified likelihood (A.27) is derived from the Propositions A.3 and A.4 below. Similar to the RSS likelihood, the new likelihood (A.27) is obtained by replacing the nuisance parameters $\{S, R, \boldsymbol{\sigma}_x\}$ with their estimates $\{\widehat{S}, \widehat{R}, \widehat{\boldsymbol{\sigma}}_x\}$.

Proposition A.3. Let $\widetilde{\Sigma} := \sigma_y^2 \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)(R + \widetilde{\Delta}(\mathbf{c}))\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)$.

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x)(\Sigma_x - \Sigma_\eta)\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \widetilde{\Sigma}), \tag{A.28}$$

where $\widetilde{\Delta}(\mathbf{c}) \in \mathbb{R}^{p\times p}$ is a continuous function of $\mathbf{c}$ and $\widetilde{\Delta}(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$.

Proof. The proof is almost identical to the proof of Proposition 2.2 (Section A.1.2). Here we only highlight the differences.

First, $g(\boldsymbol{\mu}_T)$ is different from Proposition 2.2. Specifically,

$$g(\boldsymbol{\mu}_T) = \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x)\Sigma_x^*\boldsymbol{\beta} = \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x)(\Sigma_x - \Sigma_\eta)\boldsymbol{\beta}, \tag{A.29}$$

where the last equation holds because $\boldsymbol{x}^*$ and $\boldsymbol{\eta}$ are mutually independent.

Second, $\nabla^\intercal g(\boldsymbol{\mu}_T)\Sigma_T \nabla g(\boldsymbol{\mu}_T)$ also has a different analytic form:

$$\nabla^\intercal g(\boldsymbol{\mu}_T)\Sigma_T \nabla g(\boldsymbol{\mu}_T) = \sigma_y^2 \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)(R + \widetilde{\Delta}(\mathbf{c}))\mathrm{diag}^{-1}(\boldsymbol{\sigma}_x). \tag{A.30}$$

The explicit form of $\widetilde{\Delta}(\mathbf{c})$ is given by

$$\widetilde{\Delta}(\mathbf{c}) := \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x) \cdot \left[\widetilde{G}_1(\mathbf{c}) + \widetilde{G}_2(\mathbf{c}) + \widetilde{G}_2^\intercal(\mathbf{c}) + \widetilde{G}_3(\mathbf{c})\right] \cdot \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x), \tag{A.31}$$

where functions $\widetilde{G}_i(\mathbf{c}) : \mathbb{R}^{p\times 1} \mapsto \mathbb{R}^{p\times p}$ are defined as follows:

$$\begin{aligned}
\widetilde{G}_1(\mathbf{c}) &:= -(\mathbf{c}^\intercal \mathrm{diag}(\boldsymbol{\sigma}_x)(\Sigma_x^*)^{-1}\mathrm{diag}(\boldsymbol{\sigma}_x)\mathbf{c})\Sigma_x - \mathrm{diag}(\boldsymbol{\sigma}_x)\mathbf{c}\mathbf{c}^\intercal \mathrm{diag}(\boldsymbol{\sigma}_x) + \mathrm{E}[((\boldsymbol{x}^*)^\intercal(\Sigma_x^*)^{-1}\mathrm{diag}(\boldsymbol{\sigma}_x)\mathbf{c})^2 \boldsymbol{x}\boldsymbol{x}^\intercal], \\
\widetilde{G}_2(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\mathrm{diag}(\mathbf{c})\widetilde{W}(\mathbf{c}), \quad [\widetilde{W}(\mathbf{c})]_{ij} := \sigma_{x,i}\sigma_{x,j}^2 c_i - \mathbf{c}^\intercal \mathrm{diag}(\boldsymbol{\sigma}_x)(\Sigma_x^*)^{-1}\mathrm{E}(x_i x_j^2 \boldsymbol{x}^*), \\
\widetilde{G}_3(\mathbf{c}) &:= \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)\ \mathrm{diag}(\mathbf{c})\ \Sigma_{xx}\ \mathrm{diag}\ (\mathbf{c})\ \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x), \quad [\Sigma_{xx}]_{ij} := \mathrm{Cov}(x_i^2, x_j^2).
\end{aligned}$$

Notice that $\widetilde{G}_i(\mathbf{c})$ are continuous functions of $\mathbf{c}$, $\widetilde{G}_i(\mathbf{0}) = \mathbf{0}$, and $\widetilde{G}_i(\mathbf{c}) = \mathcal{O}(\max_j c_j^2)$ for $i = 1, 2, 3$. □

Proposition A.4. Let $S := n^{-\frac{1}{2}}\sigma_y \mathrm{diag}^{-1}(\boldsymbol{\sigma}_x)$. For each $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$\log \mathcal{N}(\widehat{\boldsymbol{\beta}}; (SRS^{-1} - \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x)\Sigma_\eta)\boldsymbol{\beta}, SRS) - \log \mathcal{N}(\widehat{\boldsymbol{\beta}}; \mathrm{diag}^{-2}(\boldsymbol{\sigma}_x)(\Sigma_x - \Sigma_\eta)\boldsymbol{\beta}, n^{-1}\widetilde{\Sigma}) = \mathcal{O}_p(\max_j c_j^2).$$

Proof. The proof is the same as the proof of Proposition 2.3; see Section A.1.3. □

APPENDIX B: DETAILS OF POSTERIOR SAMPLING SCHEME

We describe the Markov chain Monte Carlo (MCMC) algorithms in terms of $\{S, R\}$, and then replace the unknown $\{S, R\}$ with their estimates $\{\widehat{S}, \widehat{R}\}$ in practice. This is similar to the likelihood derivation and prior specification in main text.

B.1. Rank-based strategy. When locally updating the SNP-specific parameters (e.g. genetic effect β_j and sparsity indicator γ_j for each SNP j) in the MCMC algorithms, we allocate more computational resources to SNPs with larger marginal association signals, using the rank-based strategy (Guan and Stephens, 2011). In particular, we first rank all the variants based on the single-SNP p-values and draw one SNP to update according to some probability distributions with decreasing probability. In our current implementation, we use a mixture distribution $q_p = 0.3u_p + 0.7g_p$, where u_p is a discrete uniform distribution and g_p is a geometric distribution truncated to $1, \ldots, p$ with its parameter chosen to give a mean of 2000.

Based on q_p, we introduce $Q(\cdot|\gamma)$, a proposal for the indicator γ. To propose a new value γ^* given the current value γ, we start by setting $\gamma^* = \gamma$ and then randomly choose one of the following:

1. With probability P_a, draw SNP r according to q_p until $\gamma_r = 0$ and set $\gamma_r^* = 1$.
2. With probability P_r, draw SNP r uniformly from $\{j : \gamma_j = 1\}$ and set $\gamma_r^* = 0$.
3. With probability P_e, sample two SNPs by the above two steps and switch their indicators.

The default setting in our software is $P_a = P_r = 0.4, P_e = 0.2$.

B.2. BVSR prior. For RSS with BVSR prior, we use Metropolis-Hastings (MH) algorithm to obtain posterior samples of (γ, π, h) on the product space of $\{0,1\}^p \times (0,1) \times (0,1)$,

$$p(\gamma, \pi, h|\widehat{\boldsymbol{\beta}}, S, R) \propto p(\widehat{\boldsymbol{\beta}}|S, R, \gamma, \pi, h)p(\gamma|\pi)p(\pi)p(h). \tag{B.1}$$

Here we are exploiting the fact that $\boldsymbol{\beta}$ can be integrated out analytically to compute $p(\widehat{\boldsymbol{\beta}}|S, R, \gamma, \pi, h)$:

$$\widehat{\boldsymbol{\beta}}|S, R, \gamma, \pi, h \sim \mathcal{N}(\mathbf{0}, SRS + \sigma_B^2 M_\gamma M_\gamma^\intercal), \tag{B.2}$$

where $M := SRS^{-1}$ and M_γ denotes the sub-matrix of M restricted to those columns j for which $\gamma_j = 1$. We update γ using the rank-based proposal $Q(\cdot|\gamma)$. We update $\log \pi$ by adding a random number from $\mathcal{U}(-0.05, 0.05)$ to the current value, and update h by adding a random number from $\mathcal{U}(-0.1, 0.1)$ to the current value. New values of $\log \pi$ and h outside boundaries are reflected back.

For each simulated posterior draw of (γ, π, h), we sample $\boldsymbol{\beta}$ according to its conditional distributions given (γ, π, h) and $(\widehat{\boldsymbol{\beta}}, S, R)$:

$$\boldsymbol{\beta}_\gamma|\widehat{\boldsymbol{\beta}}, S, R, \gamma, \pi, h \sim \mathcal{N}(\boldsymbol{\mu}, \Omega^{-1}), \tag{B.3}$$

$$\boldsymbol{\beta}_{-\gamma}|\widehat{\boldsymbol{\beta}}, S, R, \gamma, \pi, h \sim \boldsymbol{\delta}_0, \tag{B.4}$$

where $\boldsymbol{\beta}_\gamma$ and $\boldsymbol{\beta}_{-\gamma}$ denote the subsets of $\boldsymbol{\beta}$ corresponding to the entries that $\gamma_j = 1$ and 0 respectively, $\boldsymbol{\delta}_0$ denotes the point mass at zero and,

$$\Omega := M_\gamma^\intercal (SRS)^{-1} M_\gamma + \sigma_B^{-2}(\gamma, \pi, h) I_{|\gamma|}, \tag{B.5}$$

$$\boldsymbol{\mu} := \Omega^{-1} M_\gamma^\intercal (SRS)^{-1} \widehat{\boldsymbol{\beta}}. \tag{B.6}$$

The marginal likelihood (B.2), up to some constant, can be written in terms of $(\Omega, \boldsymbol{\mu})$,

$$p(\widehat{\boldsymbol{\beta}}|S, R, \gamma, \pi, h) \propto \sigma_B^{-|\gamma|} |\Omega|^{-1/2} \exp\{\boldsymbol{\mu}^\intercal \mathbf{q}_\gamma / 2\}, \tag{B.7}$$

where $\mathbf{q}_\gamma$ denotes the subset of $\mathbf{q} := S^{-1}\boldsymbol{\beta}$ corresponding to the entries that $\gamma_j = 1$. The matrix computation in a single step of the MCMC algorithm above involves one Cholesky decomposition of

Ω and three triangular linear systems. Hence, the computational cost for each iteration of MCMC is $\mathcal{O}(|\gamma|^3 + 3|\gamma|^2)$, where $|\gamma|$ denotes the number of non-zero entries in γ.

To improve precision, we can use Rao-Blackwellized estimates (Casella and Robert, 1996; Guan and Stephens, 2011). For SPIP, we have

$$\Pr(\gamma_j = 1|\widehat{\beta}, S, R) = \mathrm{E}(\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1}\textstyle\sum_{i=1}^{M}\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)})$$

where ξ_{-j} stands for $\{\beta_{-j}, \gamma_{-j}, \pi, h\}$, γ_{-j} and β_{-j} denote the vectors γ and β excluding the jth coordinate and $\xi_{-j}^{(i)}$ denotes the ith MCMC sample from the posterior distribution of ξ_{-j}. For the posterior mean of the multiple-SNP effect at SNP j, we have

$$\mathrm{E}(\beta_j|\widehat{\beta}, S, R) = \mathrm{E}(\mathrm{E}(\beta_j|\widehat{\beta}, S, R, \xi_{-j})) \approx M^{-1}\textstyle\sum_{i=1}^{M}\mathrm{E}(\beta_j|\widehat{\beta}, S, R, \gamma_j = 1, \xi_{-j}^{(i)})\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j}^{(i)}).$$

To obtain the Rao-Blackwellized estimates, we only need $p(\gamma_j|\widehat{\beta}, S, R, \xi_{-j})$ and $p(\beta_j|\widehat{\beta}, S, R, \gamma_j, \xi_{-j})$:

$$\begin{aligned}
\frac{\Pr(\gamma_j = 1|\widehat{\beta}, S, R, \xi_{-j})}{\Pr(\gamma_j = 0|\widehat{\beta}, S, R, \xi_{-j})} &= \frac{\pi}{1-\pi}\sqrt{\frac{s_j^2}{s_j^2 + \sigma_B^2}}\exp\left\{\frac{1}{2(\sigma_B^{-2} + s_j^{-2})}\left(\frac{\hat{\beta}_j}{s_j^2} - \sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right)^2\right\} \\
\beta_j|\widehat{\beta}, S, R, \gamma_j = 1, \xi_{-j} &\sim \mathcal{N}\left(\frac{1}{\sigma_B^{-2} + s_j^{-2}}\left(\frac{\hat{\beta}_j}{s_j^2} - \sum_{i\neq j}\frac{r_{ij}\beta_i}{s_i s_j}\right), \frac{1}{\sigma_B^{-2} + s_j^{-2}}\right) \\
\beta_j|\widehat{\beta}, S, R, \gamma_j = 0, \xi_{-j} &\sim \delta_0
\end{aligned}$$

where r_{ij} is the (i,j)-th entry of R.

B.3. BSLMM prior. We propose a component-wise MCMC algorithm for RSS with BSLMM prior. First, we re-parameterize the multiple-SNP effect sizes β_j as follows

$$\beta_j|\gamma_j = 1, \pi, h, \rho, S = \sqrt{\sigma_B^2 + \sigma_P^2}\cdot\tilde{\beta}_j \tag{B.8}$$

$$\beta_j|\gamma_j = 0, \pi, h, \rho, S = \sigma_P\cdot\tilde{\beta}_j \tag{B.9}$$

where the standardized effect sizes $\tilde{\beta}_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, for $j \in \{1, \ldots, p\}$. Equivalently,

$$\beta = B\widetilde{\beta}, \quad \widetilde{\beta} \sim \mathcal{N}(\mathbf{0}, I_p) \tag{B.10}$$

where the scaling matrix B is diagonal with the jth diagonal b_j defined as

$$b_j = \sigma_P\mathbf{1}\{\gamma_j = 0\} + \sqrt{\sigma_B^2 + \sigma_P^2}\mathbf{1}\{\gamma_j = 1\}. \tag{B.11}$$

The new parameterization could help speed up the convergence of MCMC, since $\tilde{\beta}$ are independent with (γ, π, h, ρ) *a priori*. We then draw posterior samples of $(\tilde{\beta}, \gamma, \pi, h, \rho)$ iteratively.

- Given $(\widetilde{\beta}, \pi, h, \rho)$, we update γ by a standard MH algorithm, where the proposal is $Q(\cdot|\gamma)$.
- Given (γ, π, h, ρ), we update $\widetilde{\beta}$ by a mixture of global and local moves. With probability P_{g}, we draw a new value of $\widetilde{\beta}$ from its full conditional,

  $$\widetilde{\beta}|\widehat{\beta}, S, R, \gamma, \pi, h, \rho \sim \mathcal{N}((BS^{-1}RS^{-1}B + I)^{-1}BS^{-2}\widehat{\beta}, (BS^{-1}RS^{-1}B + I)^{-1}). \tag{B.12}$$

  With probability $1 - P_{\mathrm{g}}$, we randomly pick a SNP j according to the distribution q_p and draw $\tilde{\beta}_j$ from its full conditional

  $$\tilde{\beta}_j|\widehat{\beta}, S, R, \tilde{\beta}_{-j}, \gamma, \pi, h, \rho \sim \mathcal{N}\left(\frac{b_j s_j \ell_j}{s_j^2 + b_j^2}, \frac{s_j^2}{s_j^2 + b_j^2}\right), \ \ell_j := \frac{\hat{\beta}_j}{s_j} - \sum_{i\neq j}\frac{r_{ij}b_i\tilde{\beta}_i}{s_i}. \tag{B.13}$$

- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, h, \rho)$, we update π by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log((\pi - p^{-1})/(1-\pi))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, \pi, \rho)$, we update h by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(h/(1-h))$.
- Given $(\widetilde{\boldsymbol{\beta}}, \gamma, \pi, h)$, we update ρ by a Metropolis algorithm, where the proposal is symmetric Gaussian random walks on $\log(\rho/(1-\rho))$.

The most computationally intensive step is drawing $\widetilde{\boldsymbol{\beta}}$ from a p-dimensional multivariate normal distribution (B.12). For each draw, one Cholesky decomposition of $BS^{-1}RS^{-1}B + I$ and two triangular linear systems are required. Since matrix R is banded with some bandwidth w (Wen and Stephens, 2010), the matrix $BS^{-1}RS^{-1}B + I$ also has the same bandwidth and therefore, the per-iteration cost of the algorithm above is at most $\mathcal{O}(pw^2 + 2p^2)$. For all the simulations, we set $P_{\text{g}} = 0.05$[1]. For the analysis of height data, we set $P_{\text{g}} = 0.001$ (the default value in our software).

B.4. Small world proposal. To improve the convergence rate of the MCMC schemes, we use the "small-world" proposal (Guan and Krone, 2007) as an add-on for every Metropolis step in our main algorithms above. Specifically, with probability 0.3 in each iteration, a long-range move is made by compounding randomly many (from 2 to 20) local proposals.

APPENDIX C: CONNECTION WITH LD SCORE REGRESSION

The LD score regression model (Bulik-Sullivan et al., 2015) is given by,

$$\mathrm{E}(\chi_j^2 | \ell_j) = nh^2\ell_j/p + na + 1, \tag{C.1}$$

where n is the sample size, p is the number of SNPs, h^2/p is the heritability per SNP, a is the contribution of confounding biases per individual, $\chi_j^2 := (\widehat{\beta}_j/s_j)^2$ is the single-SNP association χ^2 statistic and $\ell_j := \sum_{k=1}^p r_{jk}^2$ is the "LD score" of SNP j (r_{jk} is the pairwise LD between SNP j and k).

To draw the connection between the LD score regression and RSS, we consider

$$\widehat{\boldsymbol{\beta}} | S, R, \boldsymbol{\beta} \sim \mathcal{N}(SRS^{-1}\boldsymbol{\beta}, SRS + na \cdot S^2), \tag{C.2}$$

which is a generalization of RSS accounting for possible over-dispersion in real data. When $a = 0$, model (C.2) becomes the original RSS. Let $\mathbf{z} = (\mathrm{z}_1, \ldots, \mathrm{z}_p)^\intercal$, where $\mathrm{z}_j := \widehat{\beta}_j/s_j$ is the single-SNP z-score of SNP j and $z_j^2 = \chi_j^2$. Noting that $\mathbf{z} = S^{-1}\widehat{\boldsymbol{\beta}}$, we rewrite (C.2) in terms of z-scores,

$$\mathbf{z} | S, R, \boldsymbol{\beta} \sim \mathcal{N}(RS^{-1}\boldsymbol{\beta}, R + na \cdot I_p). \tag{C.3}$$

Next, we specify the following prior on $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}|S,R) = \textstyle\prod_{j=1}^p p(\beta_j|S,R),\ \mathrm{E}(\beta_j|S,R) = 0,\ \mathrm{Var}(\beta_j|S,R) = nh^2 s_j^2/p. \tag{C.4}$$

Since $s_j := (\sqrt{n}\sigma_{x,j})^{-1}\sigma_y$, the the prior variance of β_j is $(p\sigma_{x,j}^2)^{-1}(h^2\sigma_y^2)$ and thus prior (C.4) does not depend on the sample size n.

Integrating out $\boldsymbol{\beta}$ under prior (C.4), we obtain the LD score regression model:

$$\begin{aligned}
\mathrm{E}(\mathrm{z}_j^2|S,R) &= \mathrm{E}(\mathrm{Var}(\mathrm{z}_j|S,R,\boldsymbol{\beta})) + \mathrm{E}(\mathrm{E}^2(\mathrm{z}_j|S,R,\boldsymbol{\beta})) \\
&= 1 + na + \textstyle\sum_{k=1}^p r_{jk}^2 s_k^{-2}\mathrm{E}(\beta_k^2|S,R) + \sum_{k\neq\ell} r_{jk}r_{j\ell}s_k^{-1}s_\ell^{-1}\mathrm{E}(\beta_k\beta_\ell|S,R) \\
&= 1 + na + (nh^2/p)\textstyle\sum_{k=1}^p r_{jk}^2. \qquad \text{(C.5)}
\end{aligned}$$

[1]The large value of P_{g} in simulations increases the computation time of RSS-BSLMM; see Supplementary Figure 5.

APPENDIX D: LEAVE-ONE-OUT RESIDUAL IMPUTATION

We define the marginally standardized error of $\widehat{\boldsymbol{\beta}}$ as $\mathbf{e} := S^{-1}(\widehat{\boldsymbol{\beta}} - SRS^{-1}\boldsymbol{\beta})$. When the RSS likelihood is correctly specified, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, R)$. For each $i \in [p]$, the univariate complete conditional distribution of the ith entry of $\mathbf{e}$ is also normal:

$$e_i|\mathbf{e}_{-i} \sim \mathcal{N}\left(-v_{ii}^{-1}\textstyle\sum_{i\neq j} v_{ij}e_j, v_{ii}^{-1}\right), \tag{D.1}$$

where v_{ij} is the (i,j)-entry of matrix V, $V := R^{-1}$. The conditional distribution (D.1) provides us a way to impute the error of SNP i based on the errors of other SNPs. Furthermore, we can evaluate the quality of imputation using the z-score:

$$z_i(\mathbf{e}) := \frac{e_i - \mathrm{E}(e_i|\mathbf{e}_{-i})}{\sqrt{\mathrm{Var}(e_i|\mathbf{e}_{-i})}} = \sqrt{v_{ii}}\left(e_i + v_{ii}^{-1}\textstyle\sum_{i\neq j} v_{ij}e_j\right) \sim \mathcal{N}(0,1). \tag{D.2}$$

The error $\mathbf{e}$ is not observed because of the unknown true effect $\boldsymbol{\beta}$. Instead, we can only calculate the marginally standardized residual of $\widehat{\boldsymbol{\beta}}$, $\widetilde{\mathbf{e}} := S^{-1}(\widehat{\boldsymbol{\beta}} - SRS^{-1}\widetilde{\boldsymbol{\beta}})$, where $\widetilde{\boldsymbol{\beta}}$ is the posterior estimate of $\boldsymbol{\beta}$ obtained from the MCMC. We perform the leave-one-out imputation (D.1) on the residual $\widetilde{\mathbf{e}}$. The corresponding z-scores $\{z_i(\widetilde{\mathbf{e}})\}$ empirically measure the goodness of fit, and thus can be used to filter out SNPs that may be misspecified in the RSS likelihood.

APPENDIX E: SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1. Full names, abbreviations and the corresponding references of the GWAS phenotypes that are listed in main text (Table 1).

Phenotype (abbreviation)	Reference
Adult human height	Lango Allen et al. (2010)
Adult human height	Wood et al. (2014)
Body mass index (BMI)	Locke et al. (2015)
Waist-to-hip ratio adjusted for BMI (WHRadjBMI)	Shungin et al. (2015)
High-density lipoprotein (HDL)	Teslovich et al. (2010)
HDL	Global Lipids Genetics Consortium (2013)
Low-density lipoprotein (LDL)	Teslovich et al. (2010)
LDL	Global Lipids Genetics Consortium (2013)
Total cholesterol (TC)	Teslovich et al. (2010)
TC	Global Lipids Genetics Consortium (2013)
Triglycerides (TG)	Teslovich et al. (2010)
TG	Global Lipids Genetics Consortium (2013)
Cigarettes per day	Tobacco and Genetics Consortium (2010)
Smoking age of onset	Tobacco and Genetics Consortium (2010)
Ever versus never smoked	Tobacco and Genetics Consortium (2010)
Current versus former smoker	Tobacco and Genetics Consortium (2010)
Years of educational attainment	Rietveld et al. (2013)
College completion or not	Rietveld et al. (2013)
Depressive	Okbay et al. (2016)
Neuroticism	Okbay et al. (2016)
Schizophrenia	Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014)
Alzheimer	Lambert et al. (2013)
Coronary artery disease (CAD)	Schunkert et al. (2011)
Type 2 diabetes (T2D)	Morris et al. (2012)
Haemoglobin	van der Harst et al. (2012)
Mean cell haemoglobin (MCH)	van der Harst et al. (2012)
Mean cell haemoglobin concentration (MCHC)	van der Harst et al. (2012)
Mean cell volume (MCV)	van der Harst et al. (2012)
Packed cell volume (PCV)	van der Harst et al. (2012)
Red blood cell count (RBC)	van der Harst et al. (2012)
Fasting glucose adjusted for BMI (FGadjBMI)	Manning et al. (2012)
Fasting insulin adjusted for BMI (FIadjBMI)	Manning et al. (2012)
Heart rate	Den Hoed et al. (2013)
Serum urate	Köttgen et al. (2013)
Gout	Köttgen et al. (2013)
Rheumatoid arthritis (RA)	Okada et al. (2014)
Inflammatory bowel disease (IBD)	Liu et al. (2015)
Crohn's disease (CD)	Liu et al. (2015)
Ulcerative colitis (UC)	Liu et al. (2015)
CAD	Nikpay et al. (2015)
Myocardial infarction (MI)	Nikpay et al. (2015)
Age at natural menopause (ANM)	Day et al. (2015)

Supplementary Table 2. Linear relationship between the estimated PVE (SNP heritability) of each chromosome and the chromosome length (unit: Mb) for adult human height (Wood et al., 2014). Shown are the simple linear regression analyses with and without intercept.

	Estimate	**Std. Error**	***t* value**	***p* value**
Intercept	-6.505×10^{-3}	5.022×10^{-3}	-1.295	0.21
Length	2.379×10^{-4}	3.581×10^{-5}	6.644	1.81×10^{-6}

(a) RSS-BVSR

	Estimate	**Std. Error**	***t* value**	***p* value**
Intercept	2.189×10^{-4}	2.639×10^{-3}	0.083	0.94
Length	1.854×10^{-4}	1.882×10^{-5}	9.853	4.06×10^{-9}

(b) RSS-BSLMM

	Estimate	**Std. Error**	***t* value**	***p* value**
Length	1.961×10^{-4}	1.574×10^{-5}	12.460	3.62×10^{-11}

(c) RSS-BVSR

	Estimate	**Std. Error**	***t* value**	***p* value**
Length	1.868×10^{-4}	7.943×10^{-6}	23.520	$< 2 \times 10^{-16}$

(d) RSS-BSLMM

Supplementary Table 3. Estimated PVE (SNP heritability) of each chromosome for human adult height (Wood et al., 2014). The chromosome length is defined as the distance between the first and the last analyzed SNPs on each chromosome, in Megabases (Mb). The restricted maximum likelihood (REML) estimates h^2_C were obtained from the individual-level data of three GWAS of height (number of SNPs: 593,521-687,398; sample size: 6,293-15,792); see Supplementary Table 2 of Yang et al. (2011). The RSS results were summarized as the posterior median and 95% credible intervals (C.I.s).

Chr.	**Length (Mb)**	**REML**		**RSS-BVSR**		**RSS-BSLMM**	
		h^2_C	se(h^2_C)	Median	95% C.I.	Median	95% C.I.
1	246.42	0.0377	0.0088	0.0633	[0.0600, 0.0678]	0.0489	[0.0395, 0.0511]
2	242.56	0.0513	0.0094	0.0438	[0.0417, 0.0475]	0.0459	[0.0408, 0.0583]
3	199.30	0.0354	0.0084	0.0334	[0.0308, 0.0402]	0.0362	[0.0294, 0.0394]
4	191.11	0.0310	0.0079	0.0687	[0.0656, 0.0716]	0.0322	[0.0305, 0.0338]
5	180.54	0.0233	0.0078	0.0254	[0.0191, 0.0289]	0.0270	[0.0249, 0.0336]
6	170.64	0.0314	0.0079	0.0334	[0.0311, 0.0361]	0.0363	[0.0298, 0.0383]
7	158.67	0.0147	0.0069	0.0386	[0.0309, 0.0414]	0.0345	[0.0328, 0.0363]
8	146.11	0.0166	0.0068	0.0178	[0.0153, 0.0197]	0.0240	[0.0199, 0.0257]
9	140.15	0.0160	0.0067	0.0186	[0.0153, 0.0312]	0.0318	[0.0292, 0.0336]
10	135.19	0.0196	0.0071	0.0146	[0.0112, 0.0172]	0.0205	[0.0185, 0.0225]
11	134.25	0.0181	0.0064	0.0147	[0.0117, 0.0165]	0.0191	[0.0170, 0.0207]
12	132.26	0.0199	0.0067	0.0332	[0.0294, 0.0361]	0.0319	[0.0281, 0.0339]
13	96.18	0.0139	0.0061	0.0098	[0.0075, 0.0112]	0.0120	[0.0109, 0.0131]
14	87.01	0.0183	0.0060	0.0157	[0.0141, 0.0198]	0.0144	[0.0130, 0.0160]
15	81.88	0.0284	0.0064	0.0239	[0.0194, 0.0319]	0.0245	[0.0225, 0.0260]
16	88.66	0.0129	0.0058	0.0113	[0.0089, 0.0132]	0.0131	[0.0120, 0.0143]
17	78.61	0.0190	0.0060	0.0195	[0.0169, 0.0211]	0.0253	[0.0198, 0.0270]
18	76.11	0.0080	0.0054	0.0046	[0.0039, 0.0055]	0.0069	[0.0060, 0.0079]
19	63.57	0.0067	0.0045	0.0109	[0.0095, 0.0120]	0.0150	[0.0136, 0.0162]
20	62.37	0.0185	0.0058	0.0098	[0.0082, 0.0109]	0.0111	[0.0100, 0.0155]
21	36.88	0.0000	0.0037	0.0036	[0.0029, 0.0045]	0.0044	[0.0038, 0.0051]
22	35.13	0.0080	0.0040	0.0042	[0.0033, 0.0049]	0.0057	[0.0044, 0.0067]
Total		0.4487	0.0290	0.5238	[0.5035, 0.5449]	0.5209	[0.5027, 0.5390]

Supplementary Table 4. Summary of results of analyzing human height summary data (Wood et al., 2014) via RSS methods.

(a) Total PVE (SNP heritability) estimates and 95% credible intervals.

	RSS-BVSR	**RSS-BSLMM**
All SNPs	52.4%, [50.4%, 54.5%]	52.1%, [50.3%, 53.9%]
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 2	34.0%, [32.9%, 35.0%]	45.3%, [44.7%, 46.0%]
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 3	35.3%, [34.2%, 36.3%]	48.2%, [47.5%, 48.9%]

(b) The number of $\pm$40-kb regions around the genome-wide significant SNPs (GWAS hits) reported in Wood et al. (2014) that are identified by RSS-BVSR (estimated ENS ≥ 1).

	All 697 GWAS hits	**Included 384 GWAS hits**
All SNPs	531	371
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 2	532	373
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 3	540	370

(c) The number of $\pm$40-kb regions in the whole genome that are identified by RSS-BVSR (estimated ENS ≥ 1), and the putatively new regions (estimated ENS ≥ 1, and at least 1 Mb away from the 697 previously reported GWAS hits).

	All regions	**Putatively new regions**
All SNPs	5194	2138
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 2	6426	2798
Filtered SNPs, LOO $\lvert z\rvert$-score ≤ 3	6848	3079

Supplementary Table 5. Putatively new loci identified by RSS-BVSR analyses that were associated with adult human height (estimated ENS > 3). Table columns from left to right are: (1) chromosome number; (2) starting position of the $\pm$40-kb region; (3) ending position of the region; (4) estimated ENS; (5) the nearest genome-wide significant SNP reported by Wood et al. (2014); (6) the physical distance to the nearest GWAS hit, in Megabases (Mb); (6) the nearest neighbor genes; (7) the relationship between the region and the nearest gene. The nearest genes to genomic regions are found and annotated by the function `matchGenes` in the package `bumphunter` (Jaffe et al., 2012). All SNP information and genomic positions are based on Human Genome Assembly 19 (Genome Reference Consortium GRCh37).

(a) Using summary data of all SNPs (1,064,575).

Chr.	Start	End	ENS	Nearest Hit	Distance (Mb)	Nearest Gene	Annotation
5	86116344	86196344	5.22	rs6894139	2.13	*COX7C*	downstream
5	86156344	86236344	4.74	rs6894139	2.09	*MIR4280*	downstream
16	10715041	10795041	4.03	rs1659127	3.59	*TEKT5*	covers
16	78795041	78875041	3.83	rs4243206	2.71	*WWOX*	inside intron
16	78835041	78915041	3.83	rs4243206	2.67	*WWOX*	inside intron
22	43637135	43717135	3.78	rs11090631	2.13	*SCUBE1*	covers exon(s)
22	43597135	43677135	3.71	rs11090631	2.17	*SCUBE1*	overlaps 3'
12	85911619	85991619	3.67	rs17783015	4.24	*RASSF9*	downstream
19	57923127	58003127	3.55	rs2059877	9.73	*ZNF419*	overlaps 5'
8	6364984	6444984	3.54	rs4875421	1.54	*MCPH1*	inside intron
19	15723127	15803127	3.50	rs8103068	1.72	*CYP4F12*	overlaps 5'
20	821795	901795	3.41	rs7273787	3.20	*FAM110A*	covers
16	73755041	73835041	3.39	rs11640018	1.49	*LINC01568*	downstream
16	19275041	19355041	3.33	rs2023693	1.52	*CLEC19A*	covers
16	80315041	80395041	3.31	rs4243206	1.19	*DYNLRB2*	upstream
12	85871619	85951619	3.31	rs17783015	4.28	*ALX1*	downstream
20	50181795	50261795	3.31	rs6020202	1.55	*ATP9A*	overlaps 3'
17	14612467	14692467	3.25	rs8069300	2.63	*CDRT7*	upstream
19	52003127	52083127	3.16	rs2059877	3.81	*SIGLEC6*	covers
19	57963127	58043127	3.10	rs2059877	9.77	*ZNF419*	covers
12	30791619	30871619	3.05	rs10843390	1.29	*CAPRIN2*	overlaps 3'
16	19235041	19315041	3.04	rs2023693	1.56	*CLEC19A*	overlaps 5'
16	55075041	55155041	3.02	rs8058684	1.56	*IRX5*	downstream

(b) Using summary data of filtered SNPs based on LOO residual imputation.

Using the 938,798 SNPs with absolute LOO imputation $|z|$-score ≤ 2

Chr.	Start	End	ENS	Nearest Hit	Distance (Mb)	Nearest Gene	Annotation
22	43637135	43717135	4.23	rs11090631	2.13	*SCUBE1*	covers exon(s)
16	73755041	73835041	3.93	rs11640018	1.49	*LINC01568*	downstream
22	43597135	43677135	3.91	rs11090631	2.17	*SCUBE1*	overlaps 3'
19	54683127	54763127	3.61	rs2059877	6.49	*LILRA6*	overlaps 3'
19	723127	803127	3.58	rs11880992	1.37	*PTBP1*	overlaps 5'
15	91561372	91641372	3.45	rs2238300	1.71	*VPS33B*	overlaps 5'
16	79275041	79355041	3.43	rs4243206	2.23	*WWOX*	downstream
17	10852467	10932467	3.13	rs8069300	1.05	*PIRT*	upstream
21	41206282	41286282	3.09	rs2211866	1.52	*PCP4*	overlaps 5'
17	50252467	50332467	3.09	rs4605213	1.01	*CA10*	upstream
16	79315041	79395041	3.08	rs4243206	2.19	*WWOX*	downstream
17	71092467	71172467	3.07	rs10083886	1.17	*SSTR2*	covers
19	15723127	15803127	3.07	rs8103068	1.72	*CYP4F12*	overlaps 5'
16	55075041	55155041	3.04	rs8058684	1.56	*IRX5*	downstream
17	17052467	17132467	3.02	rs4640244	4.15	*MPRIP*	covers

Using the 1,018,617 SNPs with absolute LOO imputation $|z|$-score ≤ 3

Chr.	Start	End	ENS	Nearest Hit	Distance (Mb)	Nearest Gene	Annotation
22	43597135	43677135	4.57	rs11090631	2.17	*SCUBE1*	overlaps 3'
22	43637135	43717135	4.44	rs11090631	2.13	*SCUBE1*	covers exon(s)
16	73755041	73835041	4.08	rs11640018	1.49	*LINC01568*	downstream
19	52123127	52203127	4.02	rs2059877	3.93	*SIGLEC5*	overlaps 5'
17	75492467	75572467	3.97	rs1552173	1.15	*SEPT9*	overlaps 3'
19	1043127	1123127	3.97	rs11880992	1.05	*POLR2E*	covers
19	15723127	15803127	3.93	rs8103068	1.72	*CYP4F12*	overlaps 5'
19	54683127	54763127	3.85	rs2059877	6.49	*LILRA6*	overlaps 3'
16	78835041	78915041	3.82	rs4243206	2.67	*WWOX*	inside intron
19	52163127	52243127	3.80	rs2059877	3.97	*MIR99B*	covers
16	78795041	78875041	3.75	rs4243206	2.71	*WWOX*	inside intron
8	3564984	3644984	3.70	rs4875421	1.18	*CSMD1*	inside intron
17	1612467	1692467	3.59	rs870183	1.01	*MIR22HG*	covers
21	41246282	41326282	3.58	rs2211866	1.56	*PCP4*	overlaps 3'
10	5978481	6058481	3.57	rs4332428	1.01	*FBXO18*	overlaps 3'
14	94785431	94865431	3.45	rs7154721	2.36	*SERPINA6*	overlaps 5'
17	9092467	9172467	3.42	rs8067165	1.06	*STX8*	overlaps 3'
19	14923127	15003127	3.42	rs8103068	2.52	*OR7A10*	covers
16	79035041	79115041	3.40	rs4243206	2.47	*WWOX*	inside intron
17	3932467	4012467	3.40	rs870183	3.33	*ZZEF1*	overlaps 3'
19	51283127	51363127	3.39	rs2059877	3.09	*ACPT*	covers
19	1083127	1163127	3.37	rs11880992	1.01	*POLR2E*	covers
17	5692467	5772467	3.34	rs9217	1.59	*LOC339166*	covers exon(s)
15	96121372	96201372	3.30	rs7181724	1.57	*LINC00924*	downstream
19	15763127	15843127	3.25	rs8103068	1.68	*CYP4F12*	covers
16	12635041	12715041	3.22	rs1659127	1.67	*SNX29*	overlaps 3'
16	12675041	12755041	3.18	rs1659127	1.63	*CPPED1*	overlaps 3'
8	3524984	3604984	3.16	rs4875421	1.22	*CSMD1*	covers exon(s)
21	41206282	41286282	3.15	rs2211866	1.52	*PCP4*	overlaps 5'
17	35172467	35252467	3.15	rs2338115	1.68	*LHX1*	upstream
17	6252467	6332467	3.13	rs9217	1.03	*AIPL1*	overlaps 3'
17	1652467	1732467	3.10	rs870183	1.05	*SERPINF2*	overlaps 3'
22	34197135	34277135	3.09	rs2413143	1.14	*LARGE*	covers exon(s)
17	75532467	75612467	3.08	rs1552173	1.11	*LOC100507351*	covers
17	14612467	14692467	3.06	rs8069300	2.63	*CDRT7*	upstream
17	75012467	75092467	3.05	rs1552173	1.63	*SCARNA16*	covers
16	65875041	65955041	3.04	rs1966913	1.43	*LINC00922*	upstream
17	52932467	53012467	3.04	rs11867943	1.22	*TOM1L1*	overlaps 5'
17	55852467	55932467	3.00	rs1401795	1.01	*MRPS23*	covers

Supplementary Table 6. Computation time (hour:minute:second) of RSS-BVSR and RSS-BSLMM in the analyses of adult human height data (Wood et al., 2014). Computations were performed on a single core of Intel E5-2670 2.6GHz or AMD Opteron 6386 SE, with 2 million MCMC iterations per chromosome.

Chr.	# of SNPs	RSS-BVSR	RSS-BSLMM
1	86924	08:50:41	18:49:15
2	94042	16:27:26	31:33:12
3	76481	01:34:58	34:30:41
4	67627	05:42:59	15:02:51
5	67452	15:01:51	29:39:41
6	60268	03:39:18	24:32:09
7	59740	02:59:43	17:06:43
8	58361	14:59:04	28:19:17
9	52633	11:28:05	20:57:16
10	58236	28:16:28	24:40:29
11	52180	21:12:16	21:43:08
12	51123	02:02:10	18:35:34
13	43464	07:45:23	20:33:17
14	37540	01:02:52	16:32:27
15	34726	08:31:56	15:47:45
16	32260	08:43:07	10:44:12
17	25533	15:33:12	09:04:38
18	31596	05:24:35	13:50:00
19	17507	16:50:13	04:18:35
20	25983	05:58:52	08:31:22
21	15300	01:51:53	04:30:42
22	15599	02:04:01	05:55:55

(a) All SNPs (1,064,575).

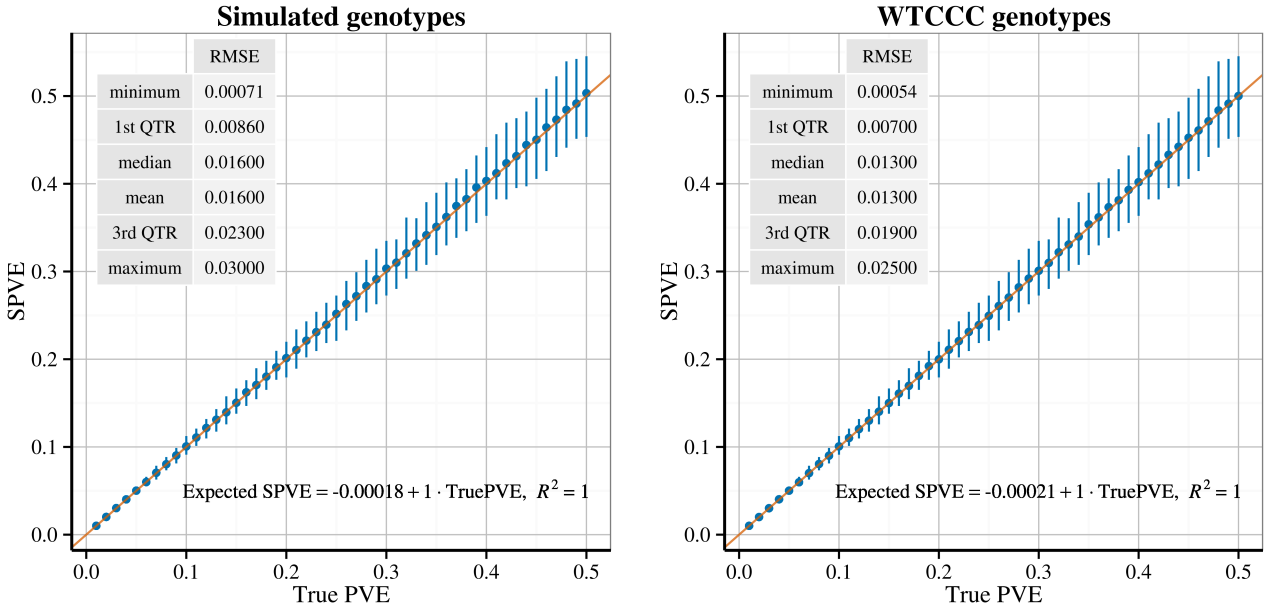
Chr.	# of SNPs	RSS-BVSR	RSS-BSLMM
1	75746	05:07:03	19:26:49
2	83175	05:54:35	24:53:23
3	67258	05:42:02	24:44:39
4	59391	18:44:43	15:51:33
5	59886	04:35:35	18:27:03
6	52539	05:00:59	15:41:28
7	52739	27:20:38	13:40:14
8	52067	18:32:13	35:38:15
9	46720	05:49:21	14:34:32
10	51038	25:59:36	35:41:12
11	46036	23:03:39	35:40:52
12	44721	03:59:24	28:02:50
13	38644	04:31:00	13:58:33
14	33118	18:32:42	12:22:30
15	30644	28:49:42	10:03:44
16	28770	16:10:05	13:06:48
17	25533	16:50:20	05:57:31
18	22337	04:40:37	08:58:16
19	15267	05:49:31	03:00:28
20	23086	03:31:40	05:51:34
21	13663	02:52:47	04:40:02
22	13674	02:14:44	05:20:14

(b) Filtered SNPs (absolute LOO imputation $|z|$-score ≤ 2).

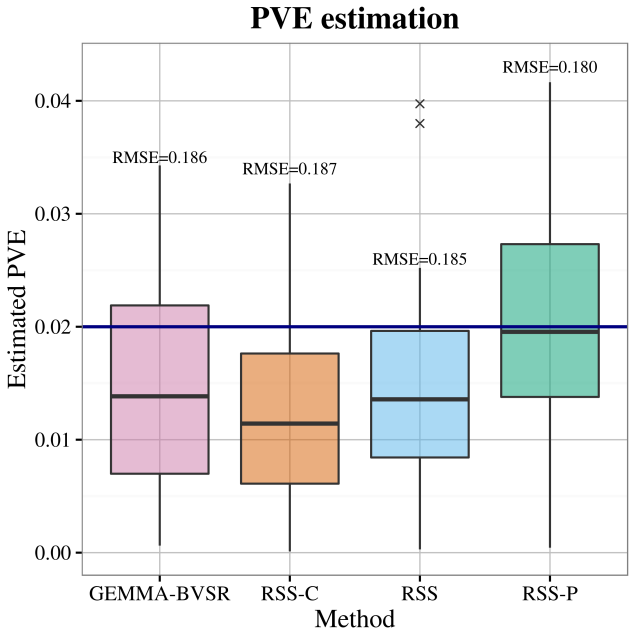
Chr.	# of SNPs	RSS-BVSR	RSS-BSLMM
1	82625	05:20:16	26:37:00
2	90263	03:40:14	29:57:33
3	73042	04:58:29	35:41:18
4	64605	20:10:46	19:51:18
5	64869	05:22:55	27:45:49
6	57241	03:05:23	18:44:27
7	57243	04:43:04	16:08:54
8	56139	33:10:26	35:35:08
9	50555	05:53:44	15:32:51
10	55544	28:07:26	35:36:17
11	49893	24:21:42	35:36:44
12	48770	05:26:07	35:36:09
13	41685	15:52:05	35:36:03
14	36012	22:36:37	27:39:27
15	33203	18:27:45	29:00:33
16	31008	22:47:40	29:01:19
17	24322	19:59:12	07:01:32
18	30449	03:54:09	21:05:19
19	16595	10:26:02	03:26:06
20	24956	04:17:32	06:15:45
21	14755	02:13:41	05:24:09
22	14843	03:00:16	06:36:39

(c) Filtered SNPs (absolute LOO imputation $|z|$-score ≤ 3).

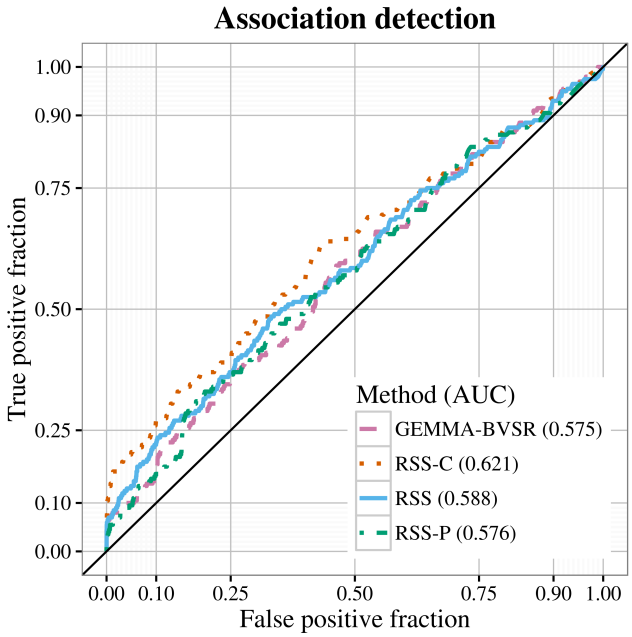
Supplementary Figure 1. Comparison of true PVE and Summary PVE (SPVE) given the true β. The true PVE is computed from the true values of $\{\beta, \tau\}$ and the individual-levle data $\{X, \mathbf{y}\}$. The SPVE is computed from the true β, the summary-level data $\{\hat{\beta}_j, \hat{\sigma}_j^2\}$ and the estimated LD matrix $\widehat{R}$. The simulated genotypes consist of 10,000 independent SNPs from 1000 individuals, so $\widehat{R}$ is set as identity matrix; The real genotypes are 10,000 correlated SNPs randomly drawn from Chromosome 16 (WTCCC UK Blood Service control group, 1458 individuals), and $\widehat{R}$ is estimated from WTCCC 1958 British Birth Cohort (1480 individuals) and HapMap CEU genetic maps using the shrinkage method in Wen and Stephens (2010). Solid dots indicate sample means of 200 replicates; vertical bars indicate symmetric 95% intervals; orange line indicates the reference line with intercept 0 and slope 1. The tables summarize the RMSEs between SPVE and true PVE.



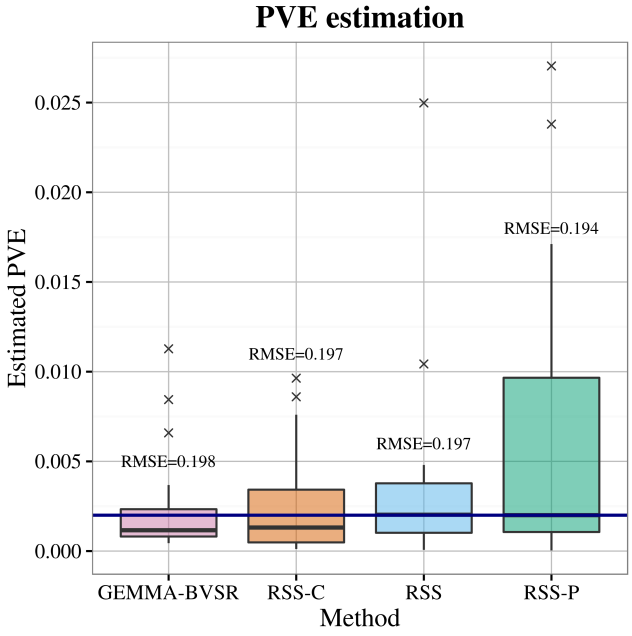
Supplementary Figure 2. Comparison of PVE estimation and association detection on three types of LD matrix: cohort sample LD (RSS-C), shrinkage panel sample LD (RSS) and panel sample LD (RSS-P). The simulation schemes and statistical methods are the same as Figure 1 in main text, except that the true PVE is 0.02 and 0.002 respectively.



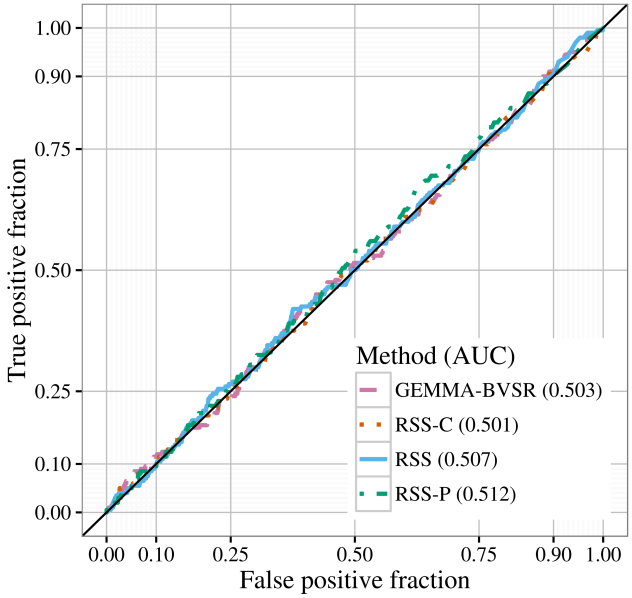
(a) True PVE: 0.02



(b) True PVE: 0.02

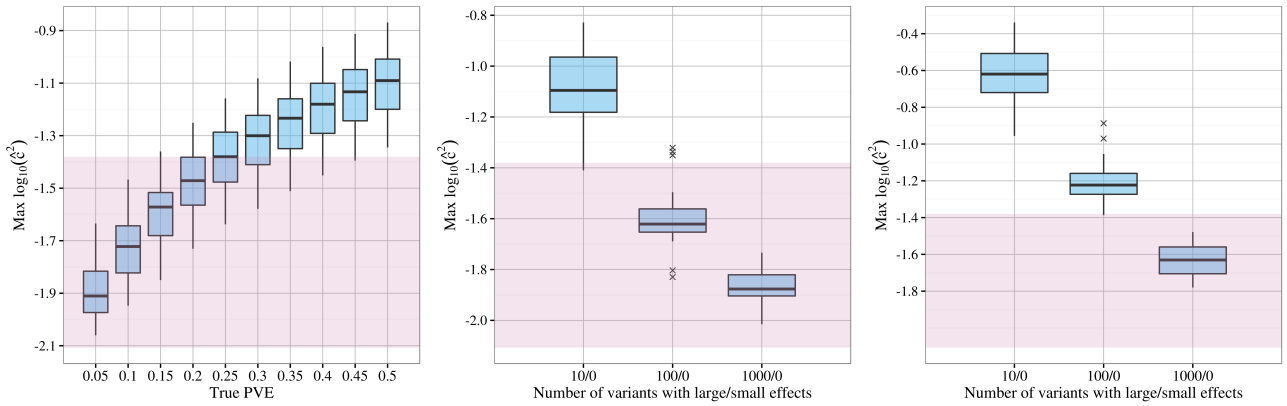


(c) True PVE: 0.002

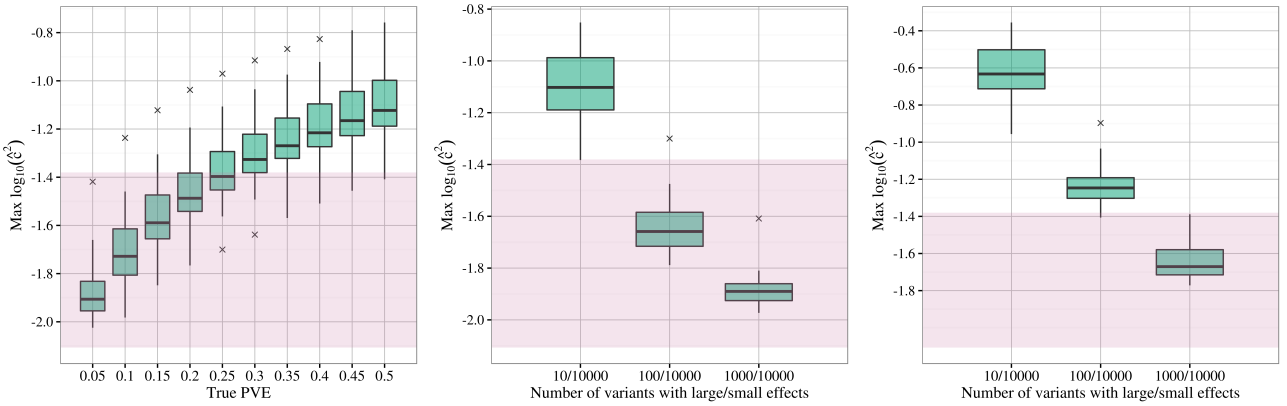


(d) True PVE: 0.002

Supplementary Figure 3. Distribution of $\max_j \log_{10}(\hat{c}_j^2)$ in all the simulated datasets used in main text. For each SNP $j \in [p]$, $\hat{c}_j := (||\mathbf{y}|| \cdot ||X_j||)^{-1}(X_j^\intercal \mathbf{y})$ is the sample correlation between phenotype ($\mathbf{y}$) and genotype of SNP j (X_j), and it can be computed from the single-SNP summary data, $\hat{c}_j^2 = (n\hat{\sigma}_j^2 + \hat{\beta}_j^2)^{-1}\hat{\beta}_j^2$. The simulations use the real genotypes of 12,758 (p) SNPs on Chromosome 16 from 1,458 (n) individuals. The shaded area in the following plots corresponds to the 60%-90% quantile of $\max_j \log_{10}(\hat{c}_j^2)$ across 42 complex traits listed in main text (Table 1). This helps us identify which simulations have "realistic" $\max_j \log_{10}(\hat{c}_j^2)$ values that are close to real GWAS datasets.

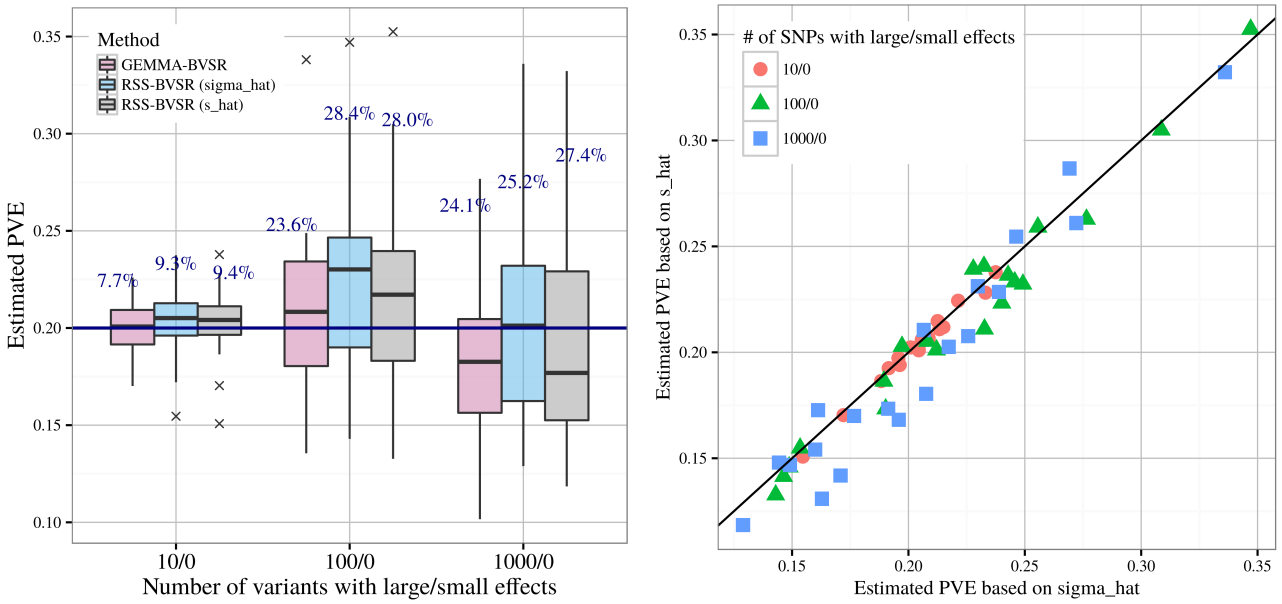


(a) Scenario 1.1 (b) Scenario 2.1, true PVE 0.2 (c) Scenario 2.1, true PVE 0.6

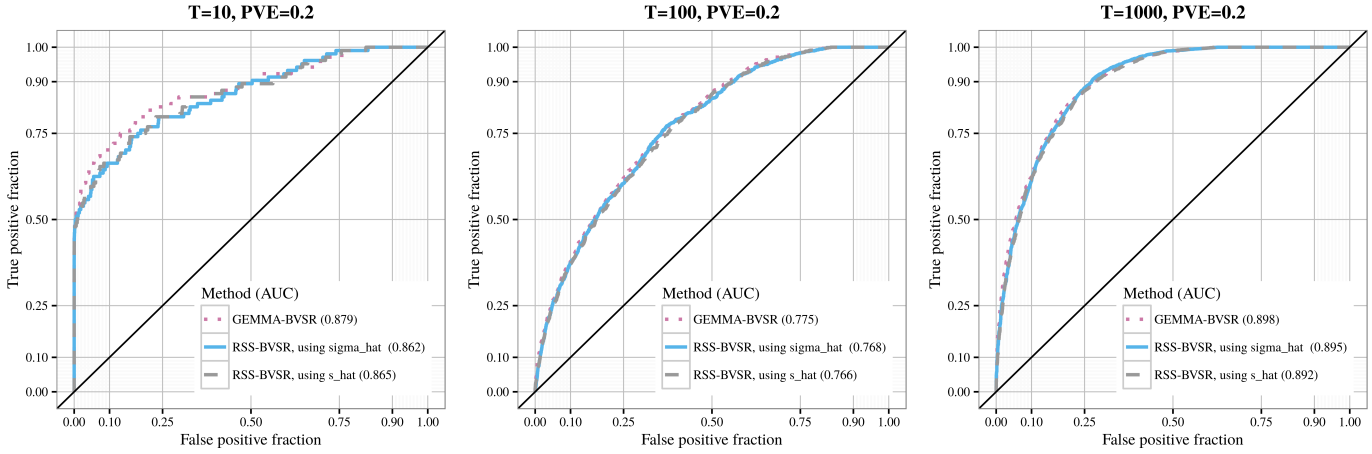


(d) Scenario 1.2 (e) Scenario 2.2, true PVE 0.2 (f) Scenario 2.2, true PVE 0.6

Supplementary Figure 4. Comparison of PVE estimation and association detection based on $\{\hat{\sigma}_j^2\}$ and $\{\hat{s}_j^2\}$ respectively. The RSS-BVSR models are fitted on the Scenario 2.1 simulated datasets in main text, with $\widehat{S}$ defined by $\{\hat{\sigma}_j^2\}$ and $\{\hat{s}_j^2\}$ respectively.

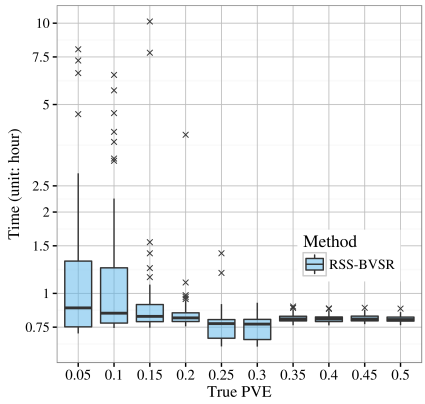


(a) Comparison of PVE estimation. Left panel: Relative RMSE for each method is reported (percentages on top of box plots). The true PVE are shown as the solid horizontal line. Each box plot summarizes results from 20 replicates. Right panel: Each point corresponds to one simulated dataset. The reference line has intercept 0 and slope 1.

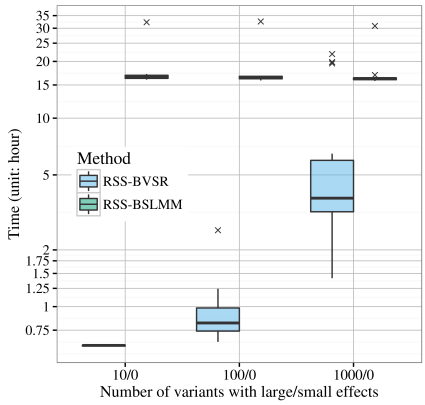


(b) Comparison of association detection. The associations are evaluated at the 200-kb region level. A region is causal if and only if it contains at least one causal SNP.

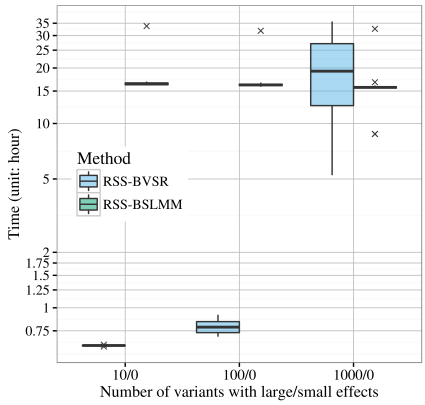
Supplementary Figure 5. Computation time, in hours, of RSS-BVSR and RSS-BSLMM in the simulation studies in main text (Section 4). For each simulated dataset and method, the computation was performed on a single core of Intel E5-2670 2.6GHz, with 2 million MCMC iterations. There are 50 replicates in Scenario 1.1 and 1.2, and 20 replicates in Scenario 2.1 and 2.2. The computation time of RSS-BSLMM in simulations is longer than real data analysis (Supplementary Table 6) because a larger P_g was used; see Appendix B.3 for details.



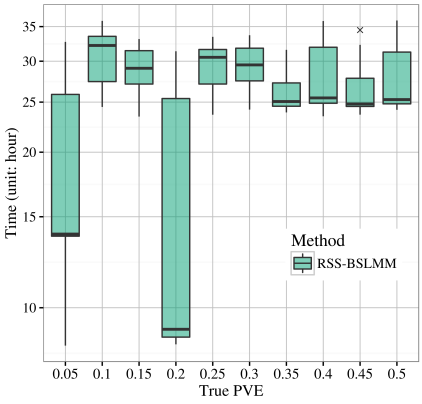
(a) Scenario 1.1



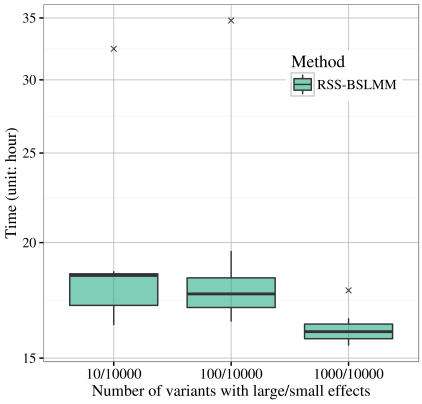
(b) Scenario 2.1, true PVE 0.2



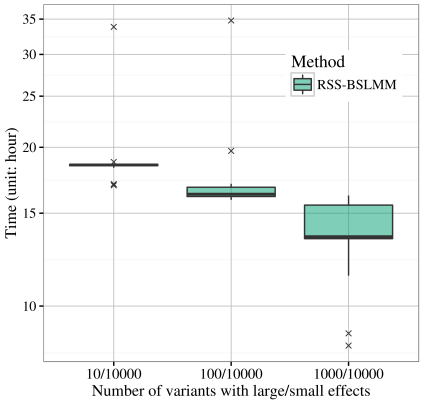
(c) Scenario 2.1, true PVE 0.6



(d) Scenario 1.2

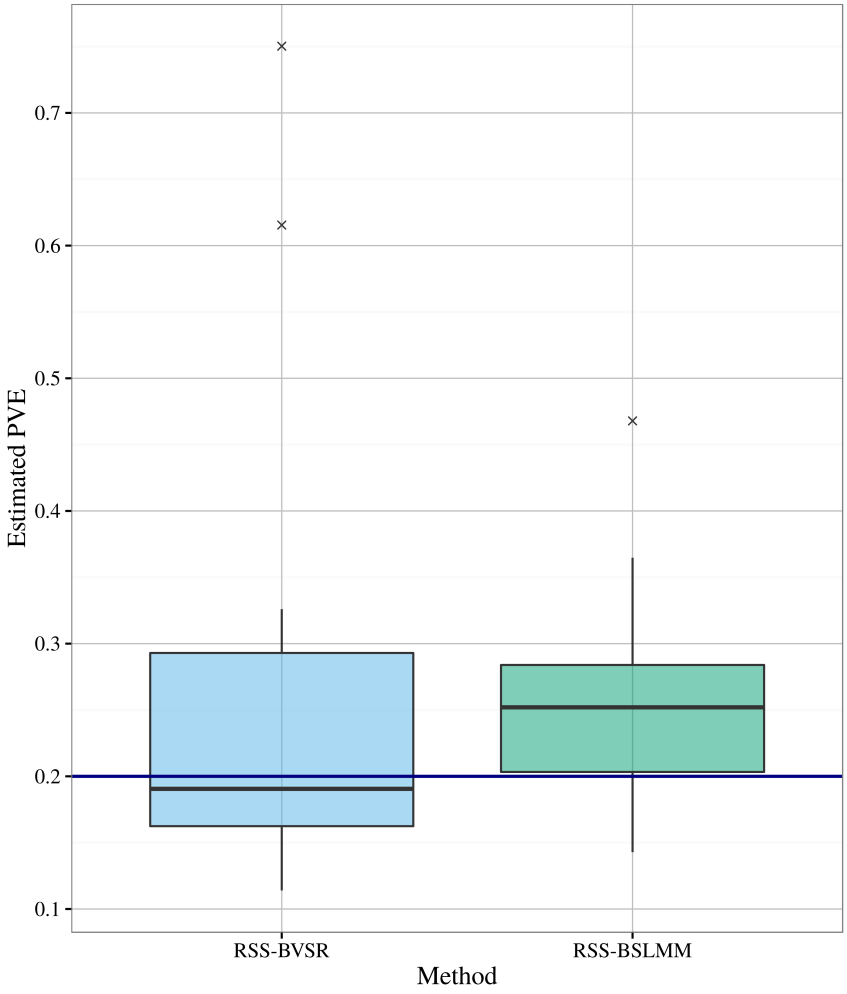


(e) Scenario 2.2, true PVE 0.2

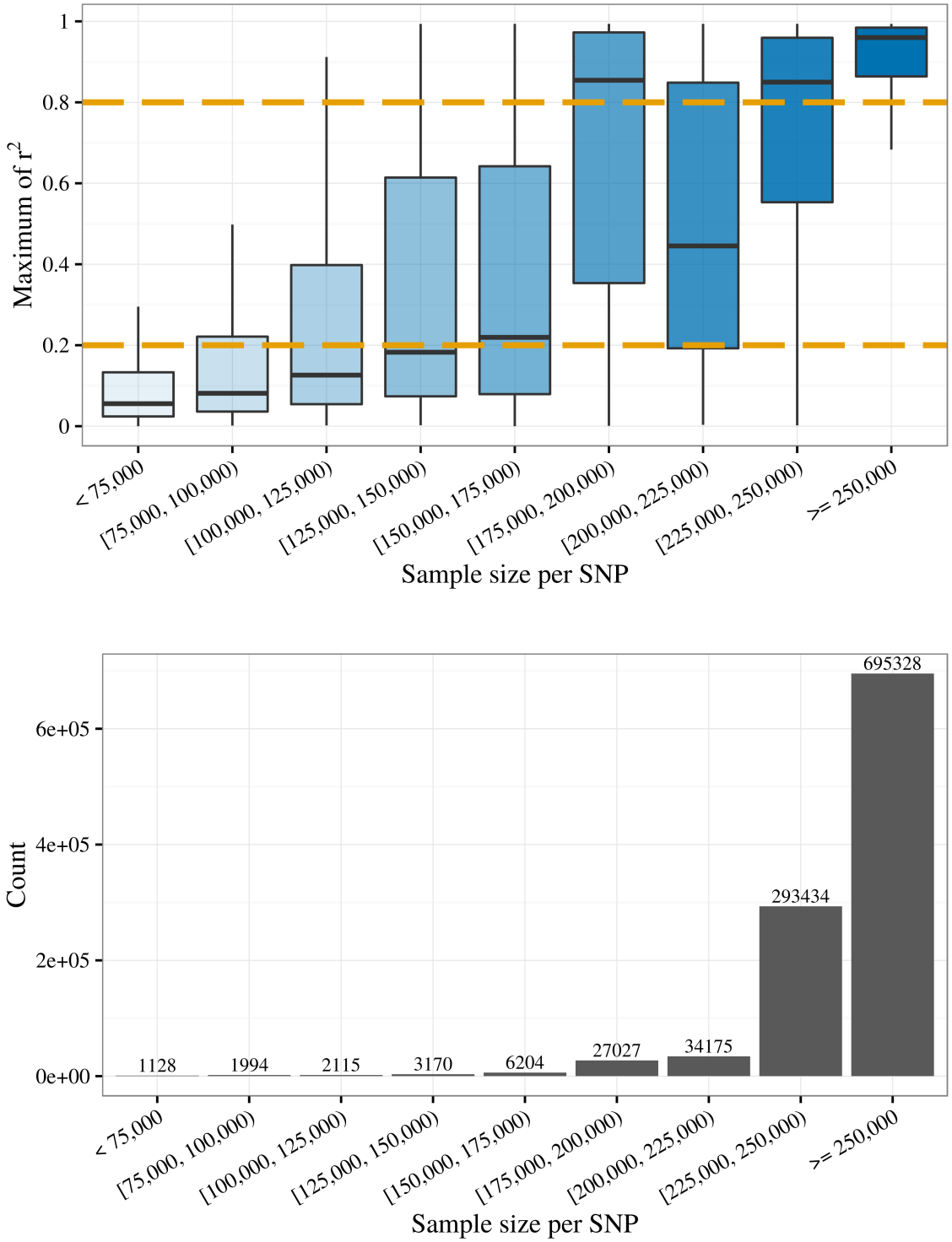


(f) Scenario 2.2, true PVE 0.6

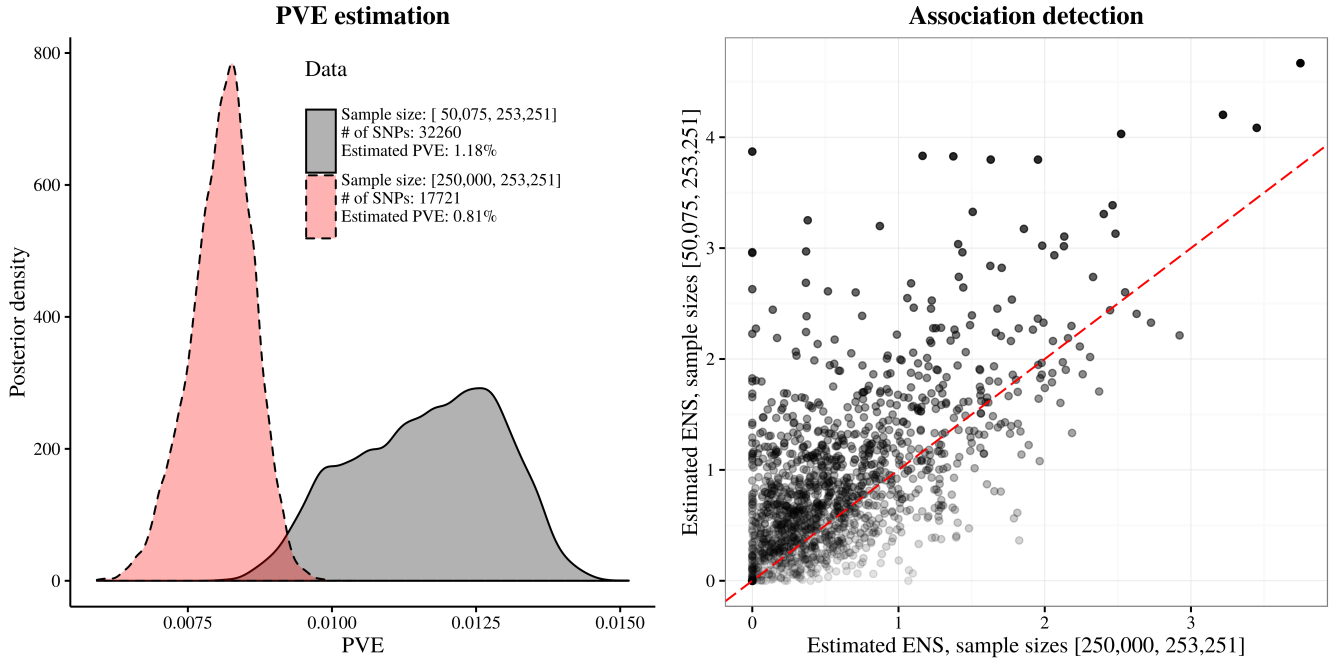
Supplementary Figure 6. Simulations show that PVE estimation can be biased when RSS methods are applied to the summary data that are *not* generated from the same sample. Here the summary data are generated as follows. For each simulated individual-level dataset (Scenario 2.1, true PVE $= 0.2$ and $T = 1000$), we first randomly draw 10% of SNPs. For each of these SNPs, we randomly draw 50% of individuals and use their data to compute the single-SNP summary statistics. For the rest of SNPs, their summary statistics are computed from all individuals.



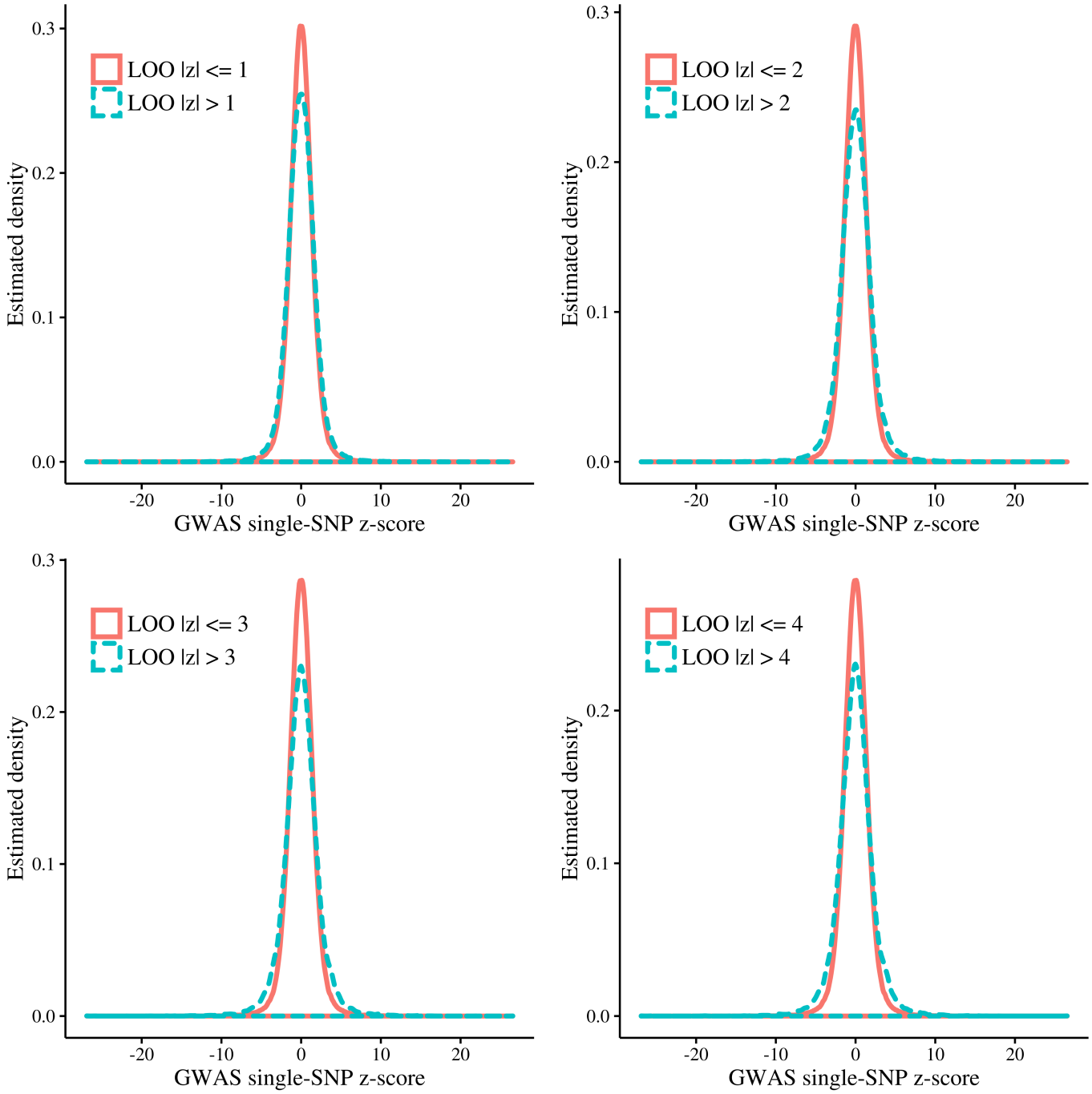
Supplementary Figure 7. Summary of sample sizes and maximum squared correlations (r^2) for the 1,064,575 analyzed SNPs from the human height summary dataset (Wood et al., 2014).



Supplementary Figure 8. SNP filtering based on sample sizes can lead to conservative results if the sample size cut-off is too high. Below are the results of fitting RSS-BVSR to the human height summary data (Wood et al., 2014) on Chromosome 16, using all 32,260 SNPs and the 17,721 SNPs with sample size greater than or equal to 250,000, respectively. The cut-off 250,000 may ensure that the summary data of the filtered SNPs are approximately generated from the same sample, but it removes almost *half* of SNPs on Chromosome 16, which further reduces the PVE estimates and association signals.



Supplementary Figure 9. Distributions of single-SNP z-scores from the human height GWAS (Wood et al., 2014). Each panel below contains the GWAS z-score distribution of SNPs that pass the leave-one-out (LOO) residual diagnostic filter (red solid curve), and the z-score distribution of SNPs that do *not* pass the filter (green dash curve).



REFERENCES

Bulik-Sullivan, B., Loh, P.-R., Finucane, H., Ripke, S., Yang, J., Psychiatric Genomics Consortium, Schizophrenia Working Group, Patterson, N., Daly, M. J., Price, A. L. and Neale, B. M. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47** 291–295.

Casella, G. and Robert, C. P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika* **83** 81–94.

Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipids levels. *Nature Genetics* **45** 1274–1283.

Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B. et al. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics* **47** 1294–1303.

Den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B. J., Peal, D. S., Evans, D. M., Nolte, I. M., Segrè, A. V., Holm, H., Handsaker, R. E. et al. (2013). Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nature Genetics* **45** 621–631.

Greene, W. H. (2012). *Econometric Analysis, 7th Edition*. Pearson Education.

Guan, Y. and Krone, S. (2007). Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing. *The Annals of Applied Probability* **17** 284–304.

Guan, Y. and Stephens, M. (2011). Bayesian variable selection regression for genome-wide association studies, and other large-scale problems. *The Annals of Applied Statistics* **5** 1780–1815.

Jaffe, A. E., Murakami, P., Lee, H., Leek, J. T., Fallin, M. D., Feinberg, A. P. and Irizarry, R. A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *International Journal of Epidemiology* **41** 200-209.

Köttgen, A., Albrecht, E., Teumer, A., Vitart, V., Krumsiek, J., Hundertmark, C., Pistis, G., Ruggiero, D., O'Seaghdha, C. M., Haller, T. et al. (2013). Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nature Genetics* **45** 145–154.

Lambert, J.-C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W. et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45** 1452–1458.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S. et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467** 832–838.

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., Ripke, S., Lee, J. C., Jostins, L., Shah, T. et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47** 979–986.

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

Manning, A. K., Hivert, M.-F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.-T., Bielak, L. F., Prokopenko, I. et al. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* **44** 659–669.

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

Nikpay, M., Goel, A., Won, H.-H., Hall, L. M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C. P., Hopewell, J. C. et al. (2015). A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics* **47** 1121–1130.

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

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506** 376–381.

Okbay, A., Baselmans, B., De Neve, J., Turley, P., Nivard, M., Fontana, M., Meddens, S., Linnér, R., Rietveld, C., Derringer, J. et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48** 624.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A. et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340** 1467–1471.

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., Preuss, M., Stewart, A. F., Barbalic, M., Gieger, C. et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics* **43** 333–338.

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E. et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518** 187–196.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J. et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466** 707–713.

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

van der Harst, P., Zhang, W., Leach, I. M., Rendon, A., Verweij, N., Sehmi, J., Paul, D. S., Elling, U., Allayee, H., Li, X. et al. (2012). Seventy-five genetic loci influencing the human red blood cell. *Nature* **492** 369–375.

Wen, X. and Stephens, M. (2010). Using linear predictors to impute allele frequencies from summary or pooled genotype data. *The Annals of Applied Statistics* **4** 1158–1182.

Wood, A. R., Esko, T., Yang, J., Vedantam, S., Pers, T. H., Gustafsson, S., Chu, A. Y., Estrada, K., Luan, J., Kutalik, Z. et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **46** 1173–1186.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M. G. et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43** 519–525.