Supplementary material for “A simple new approach to variable selection in regression, with application to genetic fine-mapping”

Gao Wang, Abhishek Sarkar, Peter Carbonetto and Matthew Stephens

University of Chicago, Chicago, IL, USA

A. Details of posterior computations for the SER model

A.1. Bayesian simple linear regression

To derive posterior computations for the SER model (2.4–2.8), it helps to start with an even simpler (univariate) linear regression model:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}b + \boldsymbol{e} \\ \boldsymbol{e} &\sim N_n(0, \sigma^2 I_n) \\ b &\sim N_1(0, \sigma_0^2). \end{aligned}$$

Here, $\boldsymbol{y}$ is an n-vector of response data (centered to have mean zero), $\boldsymbol{x}$ is an n-vector containing values of a single explanatory variable (similarly centered), $\boldsymbol{e}$ is an n-vector of independent error terms with variance σ^2, b is the scalar regression coefficient to be estimated, σ_0^2 is the prior variance of b, and I_n is the $n \times n$ identity matrix.

Given σ^2 and σ_0^2, the posterior computations for this model are very simple; they can be conveniently written in terms of the usual least-squares estimate of b, $\hat{b} \coloneqq (\boldsymbol{x}^T\boldsymbol{x})^{-1}\boldsymbol{x}^T\boldsymbol{y}$, its variance $s^2 \coloneqq \frac{\sigma^2}{\boldsymbol{x}^T\boldsymbol{x}}$, and the corresponding z score, $z \coloneqq \hat{b}/s$. The posterior distribution for b is

$$b \,|\, \boldsymbol{y}, \sigma^2, \sigma_0^2 \sim N_1(\mu_1, \sigma_1^2),$$

where

$$\sigma_1^2(\boldsymbol{x}, \boldsymbol{y}; \sigma^2, \sigma_0^2) \coloneqq \frac{1}{1/s^2 + 1/\sigma_0^2} \tag{A.1}$$

$$\mu_1(\boldsymbol{x}, \boldsymbol{y}; \sigma^2, \sigma_0^2) \coloneqq \frac{\sigma_1^2}{s^2} \times \hat{b}, \tag{A.2}$$

and the Bayes Factor (BF) for comparing this model with the null model ($b = 0$) is

$$\begin{aligned} \mathrm{BF}(\boldsymbol{x}, \boldsymbol{y}; \sigma^2, \sigma_0^2) &\coloneqq \frac{p(\boldsymbol{y} \,|\, \boldsymbol{x}, \sigma^2, \sigma_0^2)}{p(\boldsymbol{y} \,|\, \boldsymbol{x}; \sigma^2, b = 0)} \\ &= \sqrt{\frac{s^2}{\sigma_0^2 + s^2}} \exp\left(\frac{z^2}{2} \times \frac{\sigma_0^2}{\sigma_0^2 + s^2}\right). \end{aligned} \tag{A.3}$$

This expression matches the “asymptotic BF” of Wakefield (2009), but here, because we consider linear regression given σ^2, it is an exact expression for the BF, not just asymptotic.

A.2. The single effect regression model

Under the SER model (2.4–2.8), the posterior distribution of $(b_1, \ldots, b_p) = (b\gamma_1, \ldots, b\gamma_p)$ conditioned on $\sigma^2, \sigma_0^2, \boldsymbol{\pi}$ is given in the main text (eqs. 2.9 and 2.10), and is reproduced here for convenience:

$$\begin{aligned} \gamma \,|\, \boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2 &\sim \mathrm{Mult}(1, \boldsymbol{\alpha}) \\ b \,|\, \boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2, \gamma_j = 1 &\sim N_1(\mu_{1j}, \sigma_{1j}^2), \end{aligned}$$

where the vector of posterior inclusion probabilities (PIPs), $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$, can be expressed in terms of the simple linear regression BFs (A.3),

$$\alpha_j = \Pr(\gamma_j = 1 \,|\, \boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2) = \frac{\pi_j \mathrm{BF}(\boldsymbol{x}_j, \boldsymbol{y}; \sigma^2, \sigma_0^2)}{\sum_{j'=1}^{p} \pi_{j'} \mathrm{BF}(\boldsymbol{x}_{j'}, \boldsymbol{y}; \sigma^2, \sigma_0^2)},$$

where μ_{1j} and σ_{1j}^2 are the posterior mean (A.2) and variance (A.1) from the simple regression model of $\boldsymbol{y}$ on $\boldsymbol{x}_j$:

$$\begin{aligned}\mu_{1j} &= \mu_1(\boldsymbol{x}_j, \boldsymbol{y}; \sigma^2, \sigma_0^2)\\ \sigma_{1j} &= \sigma_1(\boldsymbol{x}_j, \boldsymbol{y}; \sigma^2, \sigma_0^2).\end{aligned}$$

Our algorithm requires the first and second moments of this posterior distribution, which are

$$\begin{aligned}\mathrm{E}[b_j \,|\, \boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2] &= \alpha_j \mu_{1j}\\ \mathrm{E}[b_j^2 \,|\, \boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2] &= \alpha_j(\sigma_{1j}^2 + \mu_{1j}^2).\end{aligned}$$

A.3. Computing Credible Sets

As noted in the main text, under the SER model it is straightforward to compute a level-ρ CS (Definition 1), $CS(\boldsymbol{\alpha}; \rho)$. The procedure is given in Maller et al. (2012), and for convenience we describe it here as well.

Given $\boldsymbol{\alpha}$, let $r = (r_1, \ldots, r_p)$ denote the indices of the variables ranked in order of decreasing α_j, so that $\alpha_{r_1} > \alpha_{r_2} > \cdots > \alpha_{r_p}$, and let S_k denote the cumulative sum of the k largest PIPs:

$$S_k \coloneqq \sum_{j=1}^{k} \alpha_{r_j}.$$

Now take

$$CS(\boldsymbol{\alpha}; \rho) \coloneqq \{r_1, \ldots, r_{k_0}\}, \tag{A.4}$$

where $k_0 = \min\{k : S_k \geqslant \rho\}$. This choice of k_0 ensures that the CS is as small as possible while satisfying the requirement that it is a level-ρ CS.

A.4. Estimating hyperparameters

As noted in the main text, it is possible to take an empirical Bayes approach to estimating the hyperparameters σ^2, σ_0^2. The likelihood is

$$\ell_{\mathrm{SER}}(\boldsymbol{y}; \sigma_0^2, \sigma^2) \coloneqq p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma_0^2, \sigma^2) = p_0(\boldsymbol{y} \,|\, \sigma^2) \sum_{j=1}^{p} \pi_j \, \mathrm{BF}(\boldsymbol{x}_j, \boldsymbol{y}; \sigma^2, \sigma_0^2), \tag{A.5}$$

where p_0 denotes the distribution of $\boldsymbol{y}$ under the “null” that $b = 0$ (i.e. $N_n(0, \sigma^2 I_n)$), and $\mathrm{BF}(\boldsymbol{x}, \boldsymbol{y}; \sigma^2, \sigma_0^2)$ is given in eq. A.3. The likelihood (A.5) can be maximized over one or both parameters using available numerical algorithms.

B. Derivation of variational algorithms

B.1. Background: Empirical Bayes and variational approximation

Here we introduce some notation and elementary results which are later applied to our specific application.

B.1.1. Empirical Bayes as a single optimization problem

Consider the following generic model:

$$\begin{aligned} \boldsymbol{y} &\sim p(\boldsymbol{y} \,|\, \boldsymbol{b}, \theta) \\ \boldsymbol{b} &\sim g(\boldsymbol{b}), \end{aligned}$$

where $\boldsymbol{y}$ represents a vector of observed data, $\boldsymbol{b}$ represents a vector of unobserved (latent) variables of interest, $g \in \mathcal{G}$ represents a prior distribution for $\boldsymbol{b}$ (which in the empirical Bayes paradigm is treated as an unknown to be estimated), and $\theta \in \Theta$ represents an additional set of parameters to be estimated. This formulation also includes as a special case situations where g is pre-specified rather than estimated simply by making $\mathcal{G}$ contain a single distribution.

Fitting this model by empirical Bayes typically involves the following two steps:

(a) Obtain estimates $(\hat{g}, \hat{\theta})$ of (g, θ) by maximizing the log-likelihood:

$$(\hat{g}, \hat{\theta}) := \underset{g \in \mathcal{G},\, \theta \in \Theta}{\operatorname{argmax}} \; \ell(g, \theta; \boldsymbol{y}),$$

where

$$\ell(\boldsymbol{y}; g, \theta) := \log \textstyle\int p(\boldsymbol{y} \,|\, \boldsymbol{b}, \theta)\, g(\boldsymbol{b})\, d\boldsymbol{b}.$$

(b) Given these estimates, $\hat{g}$ and $\hat{\theta}$, compute the posterior distribution for $\boldsymbol{b}$,

$$\hat{p}_{\text{post}}(\boldsymbol{b}) := p_{\text{post}}(\boldsymbol{b}; \boldsymbol{y}, g, \theta) = p(\boldsymbol{b} \,|\, \boldsymbol{y}, g, \theta) \propto p(\boldsymbol{y} \,|\, \boldsymbol{b}, \theta)\, g(\boldsymbol{b}).$$

This two-step procedure can be conveniently expressed as a single optimization problem:

$$(\hat{p}_{\text{post}}, \hat{g}, \hat{\theta}) = \underset{g \in \mathcal{G},\, \theta \in \Theta,\, q}{\operatorname{argmax}} \; F(q, g, \theta; \boldsymbol{y}), \tag{B.1}$$

with

$$F(q, g, \theta; \boldsymbol{y}) := \ell(g, \theta; \boldsymbol{y}) - D_{\text{KL}}(q \,\|\, \hat{p}_{\text{post}}), \tag{B.2}$$

and where

$$D_{\text{KL}}(q \,||\, p) := \int q(\boldsymbol{b}) \, \log \frac{q(\boldsymbol{b})}{p(\boldsymbol{b})} \, d\boldsymbol{b}$$

is the Kullback-Leibler (KL) divergence from q to p, and the optimization of q in (B.1) is over *all possible distributions on* $\boldsymbol{b}$. The function F (B.2) is often called the “evidence lower bound”, or ELBO, because it is a lower bound for the “evidence” (the marginal log-likelihood). (This follows from the fact that KL divergence is always non-negative.)

This optimization problem (B.1) is equivalent to the usual two-step EB procedure. This equivalence follows from two observations:

(a) Since the marginal log-likelihood, ℓ, does not depend on q, we have

$$\underset{q}{\operatorname{argmax}} \, F(q, g, \theta; \boldsymbol{y}) = \underset{q}{\operatorname{argmin}} \, D_{\text{KL}}(q \,\|\, \hat{p}_{\text{post}}) = \hat{p}_{\text{post}}.$$

(b) Since the minimum of D_{KL} with respect to q is zero for any (θ, g), we have that $\max_q F(q, g, \theta; \boldsymbol{y}) = \ell(\boldsymbol{y}; g, \theta)$, and as a result

$$(\hat{g}, \hat{\theta}) = \underset{g \in \mathcal{G},\, \theta \in \Theta}{\operatorname{argmax}} \; \ell(\boldsymbol{y}; g, \theta) = \underset{g \in \mathcal{G},\, \theta \in \Theta, q}{\operatorname{argmax}} \; \max_{q} \; F(q, g, \theta; \boldsymbol{y}).$$

B.1.2. Variational approximation

The optimization problem (B.1) is often intractable. The idea of variational approximation is to adjust the problem to make it tractable, simply by restricting the optimization over all possible distributions on $\boldsymbol{b}$ to $q \in \mathcal{Q}$, where $\mathcal{Q}$ denotes a suitably chosen class of distributions. Therefore, we seek to solve B.1 subject to the additional constraint that $q \in \mathcal{Q}$:

$$(\hat{p}_{\text{post}}, \hat{g}, \hat{\theta}) = \underset{g \in \mathcal{G}, \theta \in \Theta, q \in \mathcal{Q}}{\operatorname{argmax}} F(q, g, \theta; \boldsymbol{y}). \tag{B.3}$$

From the definition of F, it follows that optimizing F over $q \in \mathcal{Q}$ (for a given g and θ) corresponds to minimizing the KL divergence from q to the posterior distribution, and so can be interpreted as finding the "best" approximation to the posterior distribution for $\boldsymbol{b}$ among distributions in the class $\mathcal{Q}$. And the optimization of F over (g, θ) can be thought of as replacing the optimization of the log-likelihood with optimization of a lower bound to the log-likelihood (the ELBO).

We refer to solutions of the general problem (B.1), in which q is unrestricted, as "empirical Bayes (EB) solutions," and we refer to solutions of the restricted problem (B.3) as "variational empirical Bayes (VEB) solutions."

B.1.3. Form of ELBO

It is helpful to note that, by simple algebraic manipulations, the ELBO (B.2) can be decomposed as

$$\begin{aligned} F(q, g, \theta; \boldsymbol{y}) &= \mathrm{E}_q\left[\log \frac{p(\boldsymbol{y}, \boldsymbol{b} \mid g, \theta)}{q(\boldsymbol{b})}\right] \\ &= \mathrm{E}_q\left[\log p(\boldsymbol{y} \mid \boldsymbol{b}, \theta)\right] + \mathrm{E}_q\left[\log \frac{g(\boldsymbol{b})}{q(\boldsymbol{b})}\right]. \end{aligned} \tag{B.4}$$

B.2. The additive effects model

We now apply the above results to fitting an additive model, $\mathcal{M}$, that includes the SuSiE model (3.1–3.6) as a special case:

$$\begin{aligned} \boldsymbol{y} &= \sum_{l=1}^{L} \boldsymbol{\mu}_l + \boldsymbol{e} \\ \boldsymbol{e} &\sim N_n(0, \sigma^2 I_n) \\ \boldsymbol{\mu}_l &\sim g_l, \qquad \text{independently for } l = 1, \ldots, L, \end{aligned}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n), \boldsymbol{e} = (e_1, \ldots, e_n), \boldsymbol{\mu}_l = (\mu_{l1}, \ldots, \mu_{ln}) \in \mathbb{R}^n$. We let $\mathcal{M}_l$ denote the simpler model that is derived from $\mathcal{M}$ by setting $\boldsymbol{\mu}_{l'} = 0$ for all $l' \neq l$ (*i.e.*, $\mathcal{M}_l$ is the model that includes only the lth additive term), and we use ℓ_l to denote the marginal log-likelihood for this simpler model:

$$\ell_l(\boldsymbol{y}; g_l, \sigma^2) := \log p(\boldsymbol{y} \mid \mathcal{M}_l, g_l, \sigma^2). \tag{B.5}$$

The SuSiE model corresponds to the special case of $\mathcal{M}$ where $\boldsymbol{\mu}_l = \boldsymbol{X}\boldsymbol{b}_l$, for $l = 1, \ldots, L$, and each g_l is the "single effect prior" in (2.6–2.8). Further, in this special case each $\mathcal{M}_l$ is a "single effect regression" (SER) model (2.4–2.8).

The key idea introduced in this section is that we can fit $\mathcal{M}$ by variational empirical Bayes (VEB) provided we can fit each simpler model $\mathcal{M}_l$ by EB. To expand on this, consider fitting the model $\mathcal{M}$ by VEB, where the restricted family $\mathcal{Q}$ is the class of distributions on $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_L)$ that factorize over $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_L$; that is, for any $q \in \mathcal{Q}$,

$$q(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_L) = \prod_{l=1}^{L} q_l(\boldsymbol{\mu}_l).$$

For $q \in \mathcal{Q}$, using expression (B.4), we obtain the following expression for the ELBO, F:

$$F(q, g, \sigma^2; \boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathrm{E}_q\big[\|\boldsymbol{y} - \textstyle\sum_{l=1}^{L}\boldsymbol{\mu}_l\|^2\big] + \displaystyle\sum_{l=1}^{L} \mathrm{E}_{q_l}\left[\log \frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right], \tag{B.6}$$

Algorithm 2 Coordinate ascent for fitting additive model $\mathcal{M}$ by VEB (outline)

```
for $t$ in $0, 1, 2, \ldots$ do
    for $l$ in $1, \ldots, L$ do
        $(q_l, g_l) \leftarrow \operatorname{argmax}_{q_l, g_l} F(q, g, \sigma^2; \boldsymbol{y})$
    $\sigma^2 \leftarrow \operatorname{argmax}_{\sigma^2} F(q, g, \sigma^2; \boldsymbol{y})$
```

Algorithm 3 Coordinate ascent for fitting additive model $\mathcal{M}$ by VEB

```
Require: Initial settings of $\sigma^2$ and $g_l, \bar{\boldsymbol{\mu}}_l$, for $l = 1, \ldots, L$.
for $t$ in $0, 1, 2, \ldots$ do
    $\bar{\boldsymbol{r}} \leftarrow \boldsymbol{y} - \sum_{l=1}^{L} \boldsymbol{\mu}_l$        ▷ Compute expected residuals.
    for $l$ in $1, \ldots, L$ do
        $\bar{\boldsymbol{r}}_l \leftarrow \bar{\boldsymbol{r}} + \bar{\boldsymbol{\mu}}_l$        ▷ Disregard $l$th effect in residuals.
        $g_l \leftarrow \operatorname{argmax} \ell_l(\bar{\boldsymbol{r}}_l; g_l, \sigma^2)$        ▷ EB update of $g_l$ (optional).
        Compute posterior distribution $q_l(\boldsymbol{\mu}_l) = p(\boldsymbol{\mu}_l \mid \bar{\boldsymbol{r}}_l, \mathcal{M}_l, g_l, \sigma^2)$.
        $\bar{\boldsymbol{\mu}}_l \leftarrow \mathrm{E}_{q_l}[\boldsymbol{\mu}_l]$
        $\overline{\boldsymbol{\mu}_l^2} \leftarrow \mathrm{E}_{q_l}[\boldsymbol{\mu}_l^2]$
        $\bar{\boldsymbol{r}} \leftarrow \bar{\boldsymbol{r}}_l - \bar{\boldsymbol{\mu}}_l$        ▷ Update expected residuals.
    $\sigma^2 \leftarrow \mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}^2})/n$        ▷ Update $\sigma^2$ (optional); see (B.9).
```

in which $\|\cdot\|$ denotes the Euclidean norm, and g denotes the collection of priors $(g_1, \ldots, g_L)$. The expected value in the second term of (B.6) is the expected residual sum of squares (ERSS) under the variational approximation q, and depends on q only through its first and second moments. Indeed, if we denote the posterior first and second moments by

$$\bar{\mu}_{li} \coloneqq \mathrm{E}_{q_l}[\mu_{li}] \tag{B.7}$$

$$\overline{\mu_{li}^2} \coloneqq \mathrm{E}_{q_l}[\mu_{li}^2], \tag{B.8}$$

and we define $\bar{\boldsymbol{\mu}}_l \coloneqq (\bar{\mu}_{l1}, \ldots, \bar{\mu}_{ln})$, $\overline{\boldsymbol{\mu}_l^2} \coloneqq (\overline{\mu_{l1}^2}, \ldots, \overline{\mu_{ln}^2})$, $\bar{\boldsymbol{\mu}} \coloneqq (\bar{\boldsymbol{\mu}}_1, \ldots, \bar{\boldsymbol{\mu}}_L)$, $\overline{\boldsymbol{\mu}^2} \coloneqq (\overline{\boldsymbol{\mu}_1^2}, \ldots, \overline{\boldsymbol{\mu}_L^2})$, then we have that

$$\mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}^2}) = \mathrm{E}_q\big[\|\boldsymbol{y} - \textstyle\sum_{l=1}^{L} \boldsymbol{\mu}_l\|^2\big] = \|\boldsymbol{y} - \textstyle\sum_{l=1}^{L} \bar{\boldsymbol{\mu}}_l\|^2 + \displaystyle\sum_{l=1}^{L}\sum_{i=1}^{n} \mathrm{Var}[\mu_{li}], \tag{B.9}$$

where $\mathrm{Var}[\mu_{li}] = \overline{\mu_{li}^2} - \bar{\mu}_{li}^2$. This expression follows from the definition of the expected residual sum of squares, and from independence across $l = 1, \ldots, L$, after some algebraic manipulation; see Section B.7.

Fitting $\mathcal{M}$ by VEB involves optimizing F in (B.6) over q, g, σ^2. Our strategy is to update each (q_l, g_l) for $l = 1, \ldots, L$ while keeping σ^2 and other elements of q, g fixed, and with a separate optimization step for σ^2 with q, g fixed. This strategy is summarized in Algorithm 2.

The update for σ^2 in Algorithm 2 is easily obtained by taking partial derivative of (B.6), setting to zero, and solving for σ^2, giving

$$\hat{\sigma}^2 \coloneqq \frac{\mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}^2})}{n}. \tag{B.10}$$

The update for q_l, g_l corresponds to finding the EB solution for the simpler (single effect) model $\mathcal{M}_l$ in which the data $\boldsymbol{y}$ are replaced with the expected residuals,

$$\bar{\boldsymbol{r}}_l \coloneqq \mathrm{E}_q[\boldsymbol{r}_l] \coloneqq \mathrm{E}_q\big[\boldsymbol{y} - \textstyle\sum_{l' \neq l} \boldsymbol{\mu}_{l'}\big] = \boldsymbol{y} - \textstyle\sum_{l' \neq l} \bar{\boldsymbol{\mu}}_{l'}.$$

The proof of this result is given below in Proposition A1.

Substituting these ideas into Algorithm 2 yields Algorithm 3, which generalizes the IBSS algorithm (Algorithm 1) given in the main text.

Algorithm 4 Iterative Bayesian stepwise selection (extended version)

```
Require: Data $\boldsymbol{X}, \boldsymbol{y}$.
Require: Number of effects, $L$; initial estimates of hyperparameters $\sigma^2, \boldsymbol{\sigma_0^2}$.
Require: A function $\mathrm{SER}(\boldsymbol{X}, \boldsymbol{y}; \sigma^2, \sigma_0^2) \rightarrow (\boldsymbol{\alpha}, \boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$ that computes the posterior distribution for $\boldsymbol{b}_l$
    under the SER model; see (2.11).
Require: Initial setting of $\bar{\boldsymbol{b}}_l$, an estimate of the posterior mean of $\boldsymbol{b}_l$, for $l = 1, \ldots, L$.
repeat
    $\bar{\boldsymbol{r}} \leftarrow \boldsymbol{y} - \boldsymbol{X} \sum_{l=1}^{L} \bar{\boldsymbol{b}}_l$.                 ▷ Compute expected residuals.
    for $l$ in $1, \ldots, L$ do
        $\bar{\boldsymbol{r}}_l \leftarrow \bar{\boldsymbol{r}} + \boldsymbol{X} \bar{\boldsymbol{b}}_l$                 ▷ Disregard $l$th single effect in residuals.
        $\sigma_{0l}^2 \leftarrow \mathrm{argmax}\, \ell_{\mathrm{SER}}(\bar{\boldsymbol{r}}_l; \sigma_{0l}^2, \sigma^2)$                 ▷ EB update of $\sigma_{0l}^2$ (optional); see (A.5).
        $(\boldsymbol{\alpha}_l, \boldsymbol{\mu}_{1l}, \boldsymbol{\sigma}_{1l}^2) \leftarrow \mathrm{SER}(\boldsymbol{X}, \bar{\boldsymbol{r}}_l; \sigma^2, \sigma_{0l}^2)$                 ▷ Fit SER to residuals.
        $\underline{\bar{\boldsymbol{b}}_l} \leftarrow \boldsymbol{\alpha}_l \circ \boldsymbol{\mu}_{1l}$                 ▷ "$\circ$" denotes elementwise multiplication.
        $\overline{\boldsymbol{b}_l^2} \leftarrow \boldsymbol{\alpha}_l \circ (\boldsymbol{\sigma_{1l}^2} + \boldsymbol{\mu}_{1l}^2)$                 ▷ Compute posterior second moments.
        $\bar{\boldsymbol{r}} \leftarrow \bar{\boldsymbol{r}}_l - \boldsymbol{X} \bar{\boldsymbol{b}}_l$                 ▷ Update expected residuals.
    $\sigma^2 \leftarrow \mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{b}}, \overline{\boldsymbol{b}^2})/n$.                 ▷ Update $\sigma^2$ (optional).
until convergence criterion satisfied
return $\sigma^2, \boldsymbol{\sigma_0^2}, \boldsymbol{\alpha}_1, \boldsymbol{\mu}_{11}, \boldsymbol{\sigma}_{11}^2, \ldots, \boldsymbol{\alpha}_L, \boldsymbol{\mu}_{1L}, \boldsymbol{\sigma}_{1L}^2$.
```

B.3. Special case of SuSiE model

The SuSiE model is a special case of the above additive effects model when $\boldsymbol{\mu}_l = \boldsymbol{X}\boldsymbol{b}_l$. In this case, $\mathcal{M}_l$ is the SER model, and the first and second moments of $\boldsymbol{\mu}_l$ are easily found from the first and second moments of $\boldsymbol{b}_l$:

$$\mathrm{E}[\mu_{li}] = \mathrm{E}\big[\textstyle\sum_{j=1}^{p} x_{ij} b_{lj}\big] = \sum_{j=1}^{p} x_{ij} \mathrm{E}[b_{lj}]$$
$$\mathrm{E}[\mu_{li}^2] = \mathrm{E}\big[(\textstyle\sum_{j=1}^{p} x_{ij} b_{lj})^2\big] = \sum_{j=1}^{p} x_{ij}^2 \mathrm{E}[b_{lj}^2].$$

The expression for the second moment simplifies because only one element of $\boldsymbol{b}_l$ is non-zero under the SER model, and so $b_{lj} b_{lj'} = 0$ for any $j \neq j'$. Because of this, we can easily formulate $\mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}^2})$ as a function of the first and second moments of $\boldsymbol{b}_l$ — denoting this as $\mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{b}}, \overline{\boldsymbol{b}^2})$ — and Algorithm 3 can be implemented using posterior distributions of $\boldsymbol{b}$ instead of posterior distributions of $\boldsymbol{\mu}$.

For completeness, we give this algorithm, which is Algorithm 4. This algorithm is the same as the IBSS algorithm in the main text (Algorithm 1), with additional steps for fitting the hyperparameters σ^2 and $\boldsymbol{\sigma_0^2}$. This is the algorithm implemented in the `susieR` software. The step to update σ_{0l}^2 is a one-dimensional optimization problem; we implemented this step using the R function `optim`, which finds a stationary point of the likelihood surface with respect to σ_{0l}^2. The algorithm terminates when the increase in the ELBO between successive iterations is smaller than a small non-negative number, δ (set to 0.001 unless otherwise stated). This is a commonly used stopping criterion in algorithms for fitting variational approximations.

B.4. Update for q_l, g_l in additive effects model is EB solution for simpler model, $\mathcal{M}_l$

Here we establish that the update to q_l, g_l in Algorithm 2 can be implemented as the EB solution for $\mathcal{M}_l$ (Steps 5 and 6 in Algorithm 3). This result is formalized in the following proposition, which generalizes Proposition 1 in the main text.

Proposition A1. *The q_l, g_l that maximizes F in* (B.6)*, the ELBO for the additive model, $\mathcal{M}$, can be found by maximizing the ELBO for the simpler model, $\mathcal{M}_l$, in which the observed responses $\boldsymbol{y}$ are replaced by the expected residuals, $\bar{\boldsymbol{r}}_l$:*

$$\underset{q_l, g_l}{\mathrm{argmax}}\ F(q, g, \sigma^2; \boldsymbol{y}) = \underset{q_l, g_l}{\mathrm{argmax}}\ F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l),$$

where $\bar{\boldsymbol{\mu}}_l$ is the vector of posterior mean effects defined above (see eq. B.7), and we define

$$F_l(q_l, g_l, \sigma^2; \boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathrm{E}_{q_l}\left[\|\boldsymbol{y}-\boldsymbol{\mu}_l\|^2\right] + \mathrm{E}_{q_l}\left[\log\frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right]. \tag{B.11}$$

Proof. Omitting terms in the expression for F (from eq. B.6) that do not depend on q_l, g_l (these terms are captured by "const"), we have

$$\begin{aligned}
F(q, g, \sigma^2; \boldsymbol{y}) &= -\frac{1}{2\sigma^2}\mathrm{E}_q\left[(\boldsymbol{r}_l-\boldsymbol{\mu}_l)^T(\boldsymbol{r}_l-\boldsymbol{\mu}_l)\right] + \mathrm{E}_{q_l}\left[\log\frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right] + \text{const} \\
&= -\frac{1}{2\sigma^2}\mathrm{E}_q\left[-2\boldsymbol{r}_l^T\boldsymbol{\mu}_l + \boldsymbol{\mu}_l^T\boldsymbol{\mu}_l\right] + \mathrm{E}_{q_l}\left[\log\frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right] + \text{const} \\
&= -\frac{1}{2\sigma^2}\mathrm{E}_{q_l}\left[-2\bar{\boldsymbol{r}}_l^T\boldsymbol{\mu}_l + \boldsymbol{\mu}_l^T\boldsymbol{\mu}_l\right] + \mathrm{E}_{q_l}\left[\log\frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right] + \text{const} \\
&= F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l) + \text{const}.
\end{aligned}$$

Further note that the optimization of F_l does not restrict q_l, so the maximum yields the exact EB solution for M_l (refer to Section B.1.1); that is, $q_l(\boldsymbol{\mu}_l) = p(\boldsymbol{\mu}_l \,|\, \bar{\boldsymbol{r}}_l, \mathcal{M}_l, g_l, \sigma^2) \propto p(\bar{\boldsymbol{r}}_l \,|\, \mathcal{M}_l, g_l, \sigma^2)\, g_l(\boldsymbol{\mu}_l)$ at the maximum.

B.5. Convergence of IBSS algorithm

B.5.1. Proof of Corollary 1

Proof. Step 5 of Algorithm 1 is simply computing the right-hand side of (3.9), in which the posterior distribution is determined by parameters $\boldsymbol{\alpha}_l, \boldsymbol{\mu}_{1l}, \boldsymbol{\sigma}_{1l}^2$. Therefore, by Proposition 1, it is a coordinate ascent step for optimizing the lth coordinate of $F(q_1, \ldots, q_L; \sigma^2, \boldsymbol{\sigma_0^2})$ in which q_l is determined by the parameters $\boldsymbol{\alpha}_l, \boldsymbol{\mu}_{1l}, \boldsymbol{\sigma}_{1l}^2$.

B.5.2. Proof of Proposition 2

Proof. By Proposition 2.7.1 of Bertsekas (1999), the sequence of iterates q converges to a stationary point of F provided that $\operatorname{argmax}_{q_l, g_l} F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l)$ is uniquely attained for each l. When $\mathcal{M}_l$ is the SER model and $\boldsymbol{\mu}_l = \boldsymbol{X}\boldsymbol{b}_l$, the lower bound F_l (B.11) is

$$\begin{aligned}
F_l(q_l, g_l, \sigma^2; \boldsymbol{y}) = &-\frac{n}{2}\log(2\pi\sigma^2) - \frac{\|\boldsymbol{y}-\boldsymbol{X}\bar{\boldsymbol{b}}\|^2}{2\sigma^2} + \frac{\|\boldsymbol{X}\bar{\boldsymbol{b}}\|^2}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{j=1}^{p}\boldsymbol{x}_j^T\boldsymbol{x}_j\alpha_j(\mu_{1j}^2+\sigma_{1j}^2) \\
&+ \sum_{j=1}^{p}\frac{\alpha_j}{2}\left[1 + \log\frac{\sigma_{1j}^2}{\sigma_0^2} - \frac{\mu_{1j}^2+\sigma_{1j}^2}{\sigma_0^2}\right] + \sum_{j=1}^{p}\alpha_j\log\frac{\pi_j}{\alpha_j},
\end{aligned}$$

To lighten notation in the above expression, the l subscript was omitted from the quantities $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$, $\boldsymbol{\mu}_1 = (\mu_{11}, \ldots, \mu_{1p})$ and $\sigma_1^2 = (\sigma_{11}^2, \ldots, \sigma_{1p}^2)$ specifying the SER approximate posterior, q_l, and likewise for the vector of posterior means, $\bar{\boldsymbol{b}} \coloneqq \bar{\boldsymbol{b}}_l$ with elements $\bar{b}_j = \alpha_j\mu_{1j}$. Taking partial derivatives of this expression with respect to the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\mu}_1$ and $\boldsymbol{\sigma}_1^2$, the maximum can be expressed as the solution to the following system of equations:

$$\alpha_j\left[\frac{1}{\sigma_{1j}^2} - \left(\frac{\boldsymbol{x}_j^T\boldsymbol{x}_j}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\right] = 0 \tag{B.12}$$

$$\alpha_j\left[\frac{\mu_{1j}}{\sigma_{1j}^2} - \frac{(\boldsymbol{X}^T\boldsymbol{y})}{\sigma^2}\right] = 0 \tag{B.13}$$

$$\log\frac{\alpha_j}{\pi_j} - \log\frac{\sigma_{1j}}{\sigma_0} - \frac{\mu_{1j}^2}{2\sigma_{1j}^2} + \lambda = 0, \tag{B.14}$$

where $\lambda \in \mathbb{R}$ is an additional unknown, set so that $\alpha_1 + \cdots + \alpha_p = 1$ is satisfied. The solution to this set of equations is finite and unique if $0 < \sigma, \sigma_0 < \infty$ and $\pi_j > 0$ for all $j = 1, \ldots, p$. Also note that the solution to (B.12–B.14) recovers the posterior expressions for the SER model.

B.6. Computing the evidence lower bound

Although not strictly needed to implement Algorithms 3 and 4, it can be helpful to compute the objective function, F (e.g., to monitor the algorithm's progress, or to compare solutions). Here we outline a practical approach to computing F for the SuSiE model.

Refer to the expression for the ELBO, F, given in (B.6). Computing the first term is straightforward. The second term is the ERSS (B.9). The third term can be computed from the marginal log-likelihoods ℓ_l in (B.5), and computing this is straightforward for the SER model, involving a sum over the p possible single effects (see eq. A.5). This is shown by the following lemma:

LEMMA A1. *Let* $\hat{q}_l \coloneqq \operatorname{argmax}_q F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l)$*. Then*

$$\mathrm{E}_{\hat{q}_l}\left[\log \frac{g_l(\boldsymbol{\mu}_l)}{\hat{q}_l(\boldsymbol{\mu}_l)}\right] = \ell_l(\bar{\boldsymbol{r}}_l; g_l, \sigma^2) + \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\mathrm{E}_{\hat{q}_l}\|\bar{\boldsymbol{r}}_l - \boldsymbol{\mu}_l\|^2. \tag{B.15}$$

PROOF. Rearranging (B.11), and replacing $\boldsymbol{y}$ with $\bar{\boldsymbol{r}}_l$, we have

$$\mathrm{E}_{q_l}\left[\log \frac{g_l(\boldsymbol{\mu}_l)}{q_l(\boldsymbol{\mu}_l)}\right] = F_l(q_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l) + \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\mathrm{E}_{q_l}\|\bar{\boldsymbol{r}}_l - \boldsymbol{\mu}_l\|^2. \tag{B.16}$$

The result then follows from observing that F_l is equal to ℓ_l at the maximum, $q_l = \hat{q}_l$; that is, $F_l(\hat{q}_l, g_l, \sigma^2; \bar{\boldsymbol{r}}_l) = \ell_l(\bar{\boldsymbol{r}}_l; g_l, \sigma^2)$.

B.7. Expression for the expected residual sum of squares (ERSS)

The expression (B.9) is derived as follows:

$$\begin{aligned}
\mathrm{ERSS}(\boldsymbol{y}, \bar{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}^2}) &= \mathrm{E}_q\big[\|\boldsymbol{y} - \textstyle\sum_{l=1}^L \boldsymbol{\mu}_l\|^2\big] \\
&= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{y}^T\sum_{l=1}^L \bar{\boldsymbol{\mu}}_l + \sum_{l=1}^L\sum_{l'=1}^L \bar{\boldsymbol{\mu}}_l^T\bar{\boldsymbol{\mu}}_{l'} - \sum_{l=1}^L \bar{\boldsymbol{\mu}}_l^T\bar{\boldsymbol{\mu}}_l + \sum_{l=1}^L \mathrm{E}_{q_l}[\boldsymbol{\mu}_l^T\boldsymbol{\mu}_l] \\
&= \|\boldsymbol{y} - \textstyle\sum_{l=1}^L \bar{\boldsymbol{\mu}}_l\|^2 + \displaystyle\sum_{l=1}^L\sum_{i=1}^n \mathrm{Var}[\mu_{li}],
\end{aligned}$$

where $\mathrm{Var}[\mu_{li}] = \overline{\mu^2_{li}} - \bar{\mu}^2_{li}$.

C. Connecting SuSiE to standard BVSR

When $L \ll p$, the SuSiE model (3.1–3.6) is closely related to a standard BVSR model in which a subset of L regression coefficients are randomly chosen to have non-zero effects.

To make this precise, consider the following regression model:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e} \\
\boldsymbol{e} &\sim N_n(0, \sigma^2 I_n)
\end{aligned}$$

with n observations and p variables, so that $\boldsymbol{b}$ is a p-vector. Let $\Pi_{L,p}^{\mathrm{standard}}(\,\cdot\,; \sigma_0^2)$ denote the prior distribution on $\boldsymbol{b}$ that first randomly selects a subset $S \subset \{1, \ldots, p\}$ uniformly among all $\binom{p}{L}$ subsets of cardinality $|S| = L$, and then randomly samples the non-zero values $\boldsymbol{b}_S \coloneqq \{b_j : j \in S\}$ independently from $N_1(0, \sigma_0^2)$, setting the other values $\boldsymbol{b}_{\bar{S}} \coloneqq \{b_j : j \notin S\}$ to 0. (This is a version of the prior considered by Castillo et al. 2015, with $|S| = L$.) Further, let $\Pi_{L,p}^{\mathrm{susie}}(\,\cdot\,; \sigma_0^2)$ denote the prior distribution on $\boldsymbol{b}$ induced by the SuSiE model (3.1–3.6) with identical prior variances, $\sigma_{l0}^2 = \sigma_0^2$, for all $l = 1, \ldots, L$.

PROPOSITION A2. *With L fixed, letting $p \to \infty$, the SuSiE prior is equivalent to the standard prior. Specifically, for any event A,*

$$\lim_{p\to\infty}\left(\Pi_{L,p}^{susie}(A; \sigma_0^2) - \Pi_{L,p}^{standard}(A; \sigma_0^2)\right) = 0.$$

Proof. Fix L and p, and let B denote the event that the L vectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_L$ in the SuSiE model are distinct from one another. Conditional on B, it is clear from symmetry that the SuSiE prior exactly matches the standard prior; that is, $\Pi_{L,p}^{\text{susie}}(A \mid B) = \Pi_{L,p}^{\text{standard}}(A)$, dropping notational dependence on σ_0^2 for simplicity. Thus,

$$\begin{aligned}\Pi_{L,p}^{\text{susie}}(A) - \Pi_{L,p}^{\text{standard}}(A) &= \Pi_{L,p}^{\text{susie}}(A) - \Pi_{L,p}^{\text{susie}}(A \mid B) \\ &= \Pi_{L,p}^{\text{susie}}(A \mid B)\text{Pr}_{L,p}(B) + \Pi_{L,p}^{\text{susie}}(A \mid \bar{B})\text{Pr}_{L,p}(\bar{B}) - \Pi_{L,p}^{\text{susie}}(A \mid B),\end{aligned}$$

where the last line follows from the law of total probability. The result then follows from the fact that the probability $\text{Pr}_{L,p}(B) \rightarrow 1$ as $p \rightarrow \infty$:

$$\text{Pr}_{L,p}(B) = [p/p][(p-1)/p][(p-2)/p] \cdots [(p-L+1)/p] \rightarrow 1 \text{ as } p \rightarrow \infty.$$

D. Simulation details

D.1. Simulated data

For the numerical simulations of eQTL fine mapping in Section 4, we used $n = 574$ human genotypes collected as part of the Genotype-Tissue Expression (GTEx) project (GTEx Consortium, 2017). Specifically, we obtained genotype data from whole-genome sequencing, with imputed genotypes, under dbGaP accession `phs000424.v7.p2`. In our analyses, we only included SNPs with minor allele frequencies 1% or greater. All reported SNP base-pair positions were based on Genome Reference Consortium (GRC) human genome assembly 38.

To select SNPs nearby each gene, we considered two SNP selection schemes in our simulations: (i) in the first scheme, we included all SNPs within 1 Megabase (Mb) of the gene's transcription start site (TSS); (ii) in the second, we used the $p = 1,000$ SNPs closest to the TSS. Since the GTEx data contain a very large number of SNPs, the 1,000 closest SNPs are never more than 0.4 Mb away from the TSS. Selection scheme (i) yields genotype matrices $\boldsymbol{X}$ with at least $p =$ 3,022 SNPs and at most $p =$ 11,999 SNPs, and an average of 7,217 SNPs.

For illustration, correlations among the SNPs for one of the data sets are shown in Fig. A1 (see also Fig. 1).

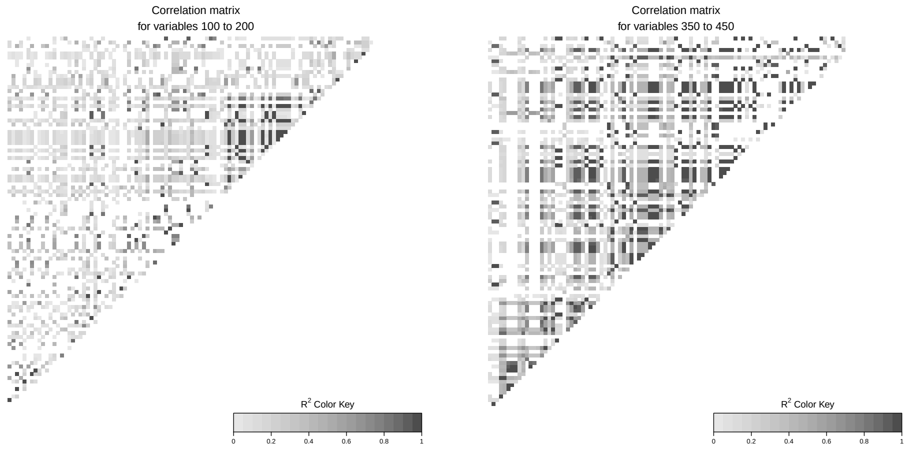

Fig. A1. Correlations among variables (SNPs) in an example data set used in the fine mapping comparisons. Left-hand panel shows correlations among variables shown at positions 100–200 in Fig. 1; right-hand panel shows correlations among variables shown at positions 350–450. For more details on this example data set, see Section 4.1 in the main text.

D.2. Software and hardware specifications for numerical comparisons study

In CAVIAR, we set all prior inclusion probabilities to $1/p$ to match the default settings used in other methods. In CAVIAR and FINEMAP, we set the maximum number of effect variables to the value of S

that was used to simulate the gene expression data. The maximum number of iterations in FINEMAP was set to 100,000 (this is the FINEMAP default). We estimate σ^2 in SuSiE for all simulations.

All computations were performed on Linux systems with Intel Xeon E5-2680 v4 (2.40 GHz) processors. We ran SuSiE in R 3.5.1, with optimized matrix operations provided by the dynamically linked OpenBLAS libraries (version 0.3.5). DAP-G and CAVIAR were compiled from source using GCC 4.9.2, and pre-compiled binary executables, available from the author's website, were used to run FINEMAP.

E. Functional enrichment of splice QTL fine mapping

To strengthen results of Section 5, here we provide evidence that splice QTLs identified by SuSiE are enriched in functional genomic regions, thus likely to contain true causal effects. To perform this analysis, we labelled one CS at each intron the "primary CS." We chose the CS with highest purity at each intron as the primary CS; any additional CSs at each intron were labelled as "secondary CSs." We then tested both primary and secondary CSs for enrichment of biological annotations by comparing the SNPs inside these CSs (those with PIP > 0.2) against random "control" SNPs outside all primary and secondary CSs.

We tested for enrichment of the same generic variant annotations used in Li et al. (2016). These include LCL-specific histone marks (H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me2, H3K9ac, H3K9me3, H4K20me1), DNase I hypersensitive sites, transcriptional repressor CTCF binding sites, RNA polymerase II (PolII) binding sites, extended splice sites (defined as 5 base-pairs upstream and downstream of an intron start site, and 15 base-pairs upstream and downstream of an intron end site), as well as intron and coding annotations. In total, 16 variant annotations were tested for enrichment.

Fig. A2 shows the enrichments in both primary and secondary CSs for the 12 out of 16 annotations that were significant at p-value $< 10^{-4}$ in the primary signals (Fisher's exact test, two-sided, no p-value adjustment for multiple comparisons). The strongest enrichment in both primary and secondary signals was for extended splice sites (odds ratio ≈ 5 in primary signals), which is reassuring given that these results are for splice QTLs. Other significantly enriched annotations in primary signals include PolII binding, several histone marks, and coding regions. The only annotation showing a significant depletion was introns. Results for secondary signals were qualitatively similar to those for primary, though all enrichments are less significant, which is most likely explained by the much smaller number of secondary CSs.

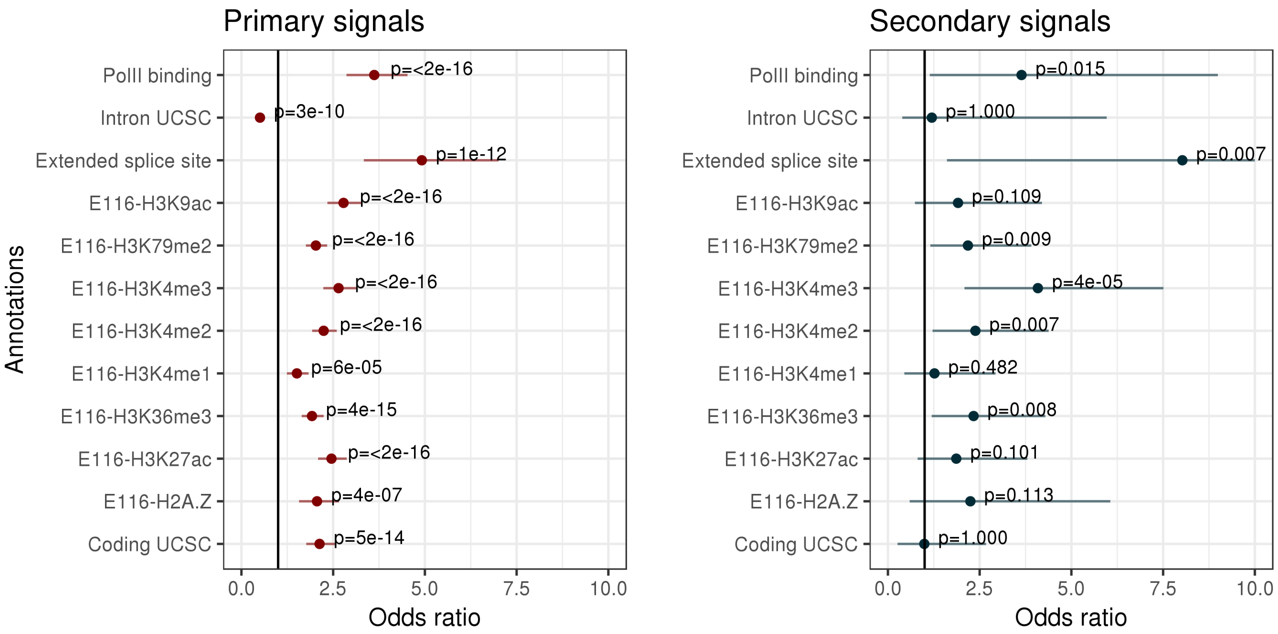

Fig. A2. Splicing QTL enrichment analysis results. Estimated odds ratios, and $\pm$ 2 standard errors, for each variant annotation, obtained by comparing the annotations of SNPs inside primary/secondary CSs against random "control" SNPs outside CSs. The p-values are from two-sided Fisher's exact test, without multiple testing correction. The vertical line in each plot is posited at odds ratio = 1.

Supplementary Figures

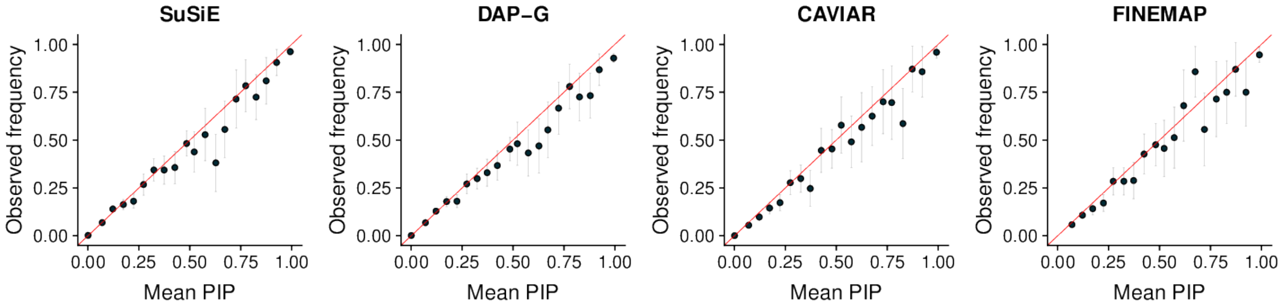

Fig. S1. Assessment of PIP calibration. Variables across all simulations were grouped into bins according to their reported PIP (using 20 equally spaced bins, from 0 to 1). The plots show the average PIP for each bin against the proportion of effect variables in that bin. A well calibrated method should produce points near the $x = y$ line (the diagonal red lines). Gray error bars show ± 2 standard errors.

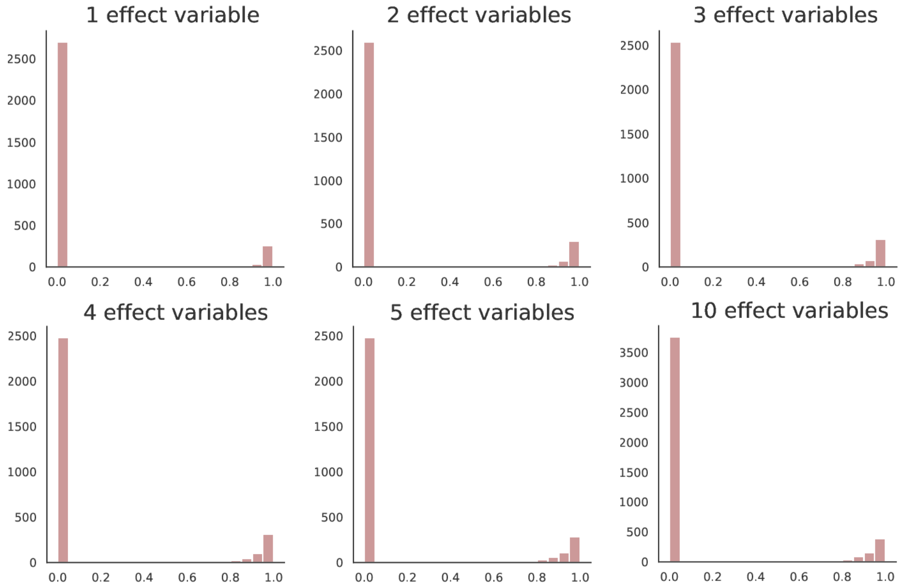

Fig. S2. Distribution of purity for 95% credible sets for different numbers of effect variables. Histograms for 1–5 effect variables are obtained from all 95% credible sets produced by SuSiE in the first simulation scenario, with $S = 1, \dots, 5$, as described in Section 4 of the main text, and the 10 effect variables histogram is obtained from all 95% credible sets produced by SuSiE in the second simulation scenario, with $S = 10$.

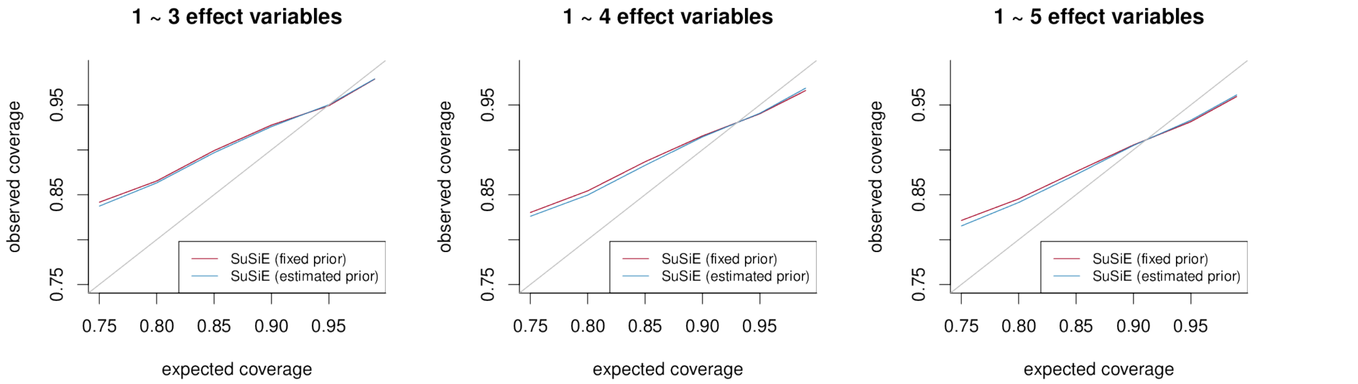

Fig. S3. Additional assessment of SuSiE CS coverage. These three plots show coverage of SuSiE credible sets as ρ (the probability that the credible set contains at least one effect variable; see Definition 1 in the main text) is varied from 75% to 99%. Proportions shown in the vertical axis are based on all credible sets generated by SuSiE in simulations from simulation scenario 1, with different simulation settings for S, the number of effect variables. Consistent with Fig. 3, coverage decreases with the inclusion of weaker signals.

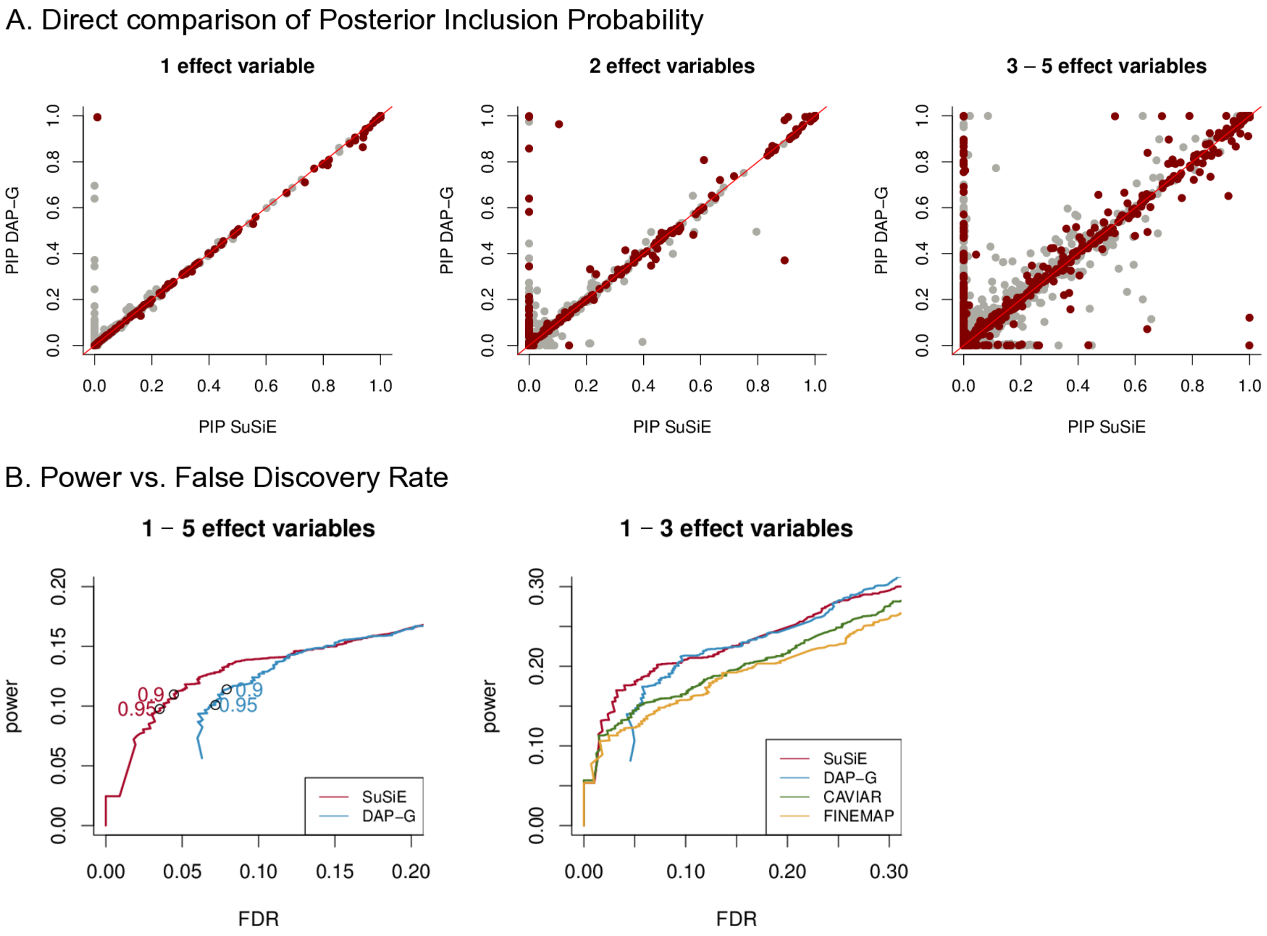

Fig. S4. Comparison of posterior inclusion probabilities (PIPs) computed by SuSiE, in which the prior variances σ^2 are estimated rather than fixed to 0.1, against PIPs computed by DAP-G, and by other methods. The results shown here from methods other than SuSiE are the same as the results in Fig. 2. For an explanation of the individual plots, see Fig. 2.

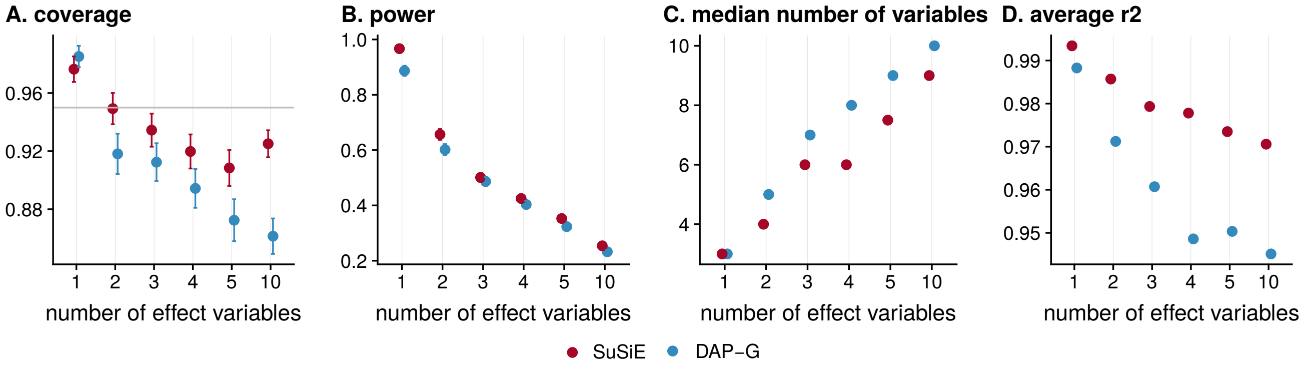

Fig. S5. Comparison of 95% credible sets (CS) from SuSiE, in which the prior variances σ^2 are estimated rather than fixed to 0.1, and DAP-G: (A) coverage, (B) power, (C) median size, and (D) average squared correlation among variables in each credible set. The DAP-G results shown here are the same as the DAP-G results shown in Fig. 3. For an explanation of the individual plots, see Fig. 3.

References

Bertsekas, D. P. (1999). *Nonlinear programming* (2nd ed.). Belmont, MA: Athena Scientific.

Castillo, I., Schmidt-Hieber, J., and van der Vaart, A. (2015). Bayesian linear regression with sparse priors. *Annals of Statistics 43*(5), 1986–2018.

GTEx Consortium (2017). Genetic effects on gene expression across human tissues. *Nature 550*(7675), 204–213.

Li, Y. I., van de Geijn, B., Raj, A., et al. (2016). RNA splicing is a primary link between genetic variation and disease. *Science 352*(6285), 600–604.

Maller, J. B., McVean, G., Byrnes, J., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics 44*(12), 1294–1301.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. *Genetic Epidemiology 33*(1), 79–86.