Fast and flexible joint fine-mapping of multiple traits via the Sum of Single Effects model

Yuxin Zou

Department of Statistics, University of Chicago, Chicago, IL, USA
Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA

Peter Carbonetto

Department of Human Genetics and the Research Computing Center, University of Chicago, Chicago, IL, USA

Dongyue Xie

Department of Statistics, University of Chicago, Chicago, IL, USA

Gao Wang*

Department of Neurology and the Gertrude. H. Sergievsky Center, Columbia University, New York, NY, USA

Matthew Stephens*

Departments of Statistics and Human Genetics, University of Chicago, Chicago, IL, USA
**e-mail: wang.gao@columbia.edu; mstephens@uchicago.edu*

Abstract.
We introduce mvSuSiE, a multi-trait fine-mapping method for identifying putative causal variants from genetic association data (individual-level or summary data). mvSuSiE learns patterns of shared genetic effects from data, and exploits these patterns to improve power to identify causal SNPs. Comparisons on simulated data show that mvSuSiE is competitive in speed, power and precision with existing multi-trait methods, and uniformly improves on single-trait fine-mapping (SuSiE) in each trait separately. We applied mvSuSiE to jointly fine-map 16 blood cell traits using data from the UK Biobank. By jointly analyzing the traits and modeling heterogeneous effect sharing patterns, we discovered a much larger number of causal SNPs (>3,000) compared with single-trait fine-mapping, and with narrower credible sets. mvSuSiE also more comprehensively characterized the ways in which the genetic variants affect one or more blood cell traits; 68% of causal SNPs showed significant effects in more than one blood cell type.

Introduction

Genome-wide association analyses (GWAS) have been performed for thousands of traits and have identified many genomic regions associated with diseases and complex traits [2, 14, 18, 62, 71, 85, 91]. Many statistical fine-mapping methods have been developed to prioritize putative causal SNPs for a single trait [9, 20, 35, 42, 46, 51, 52, 54, 56, 59, 67, 75, 78, 79, 95, 96, 99, 101, 106]. A simple strategy to fine-map multiple traits is to fine-map each trait separately, then integrate the results *post hoc*. However, integration of results is not straightforward; for example, it is difficult to say whether signals identified in different single-trait analyses likely correspond to the same underlying causal SNP. Further, analyzing each trait independently is inefficient in that it cannot exploit the potential for increased power of a multivariate analysis [81]. Therefore, it is desirable to fine-map the traits simultaneously—that is, to perform *multi-trait fine-mapping*.

Although several methods have been developed for multi-trait fine-mapping (Table 1), these methods have important practical limitations. Many of them are computationally impractical for more than a small numbers of traits (e.g., 2 traits) or for more than 1 or 2 causal SNPs in the fine-mapping region. Most methods also make restrictive assumptions about how SNPs affect the traits, such as that the effects of causal SNPs are uncorrelated among traits [e.g., 50, 55, 74]. These assumptions are easily violated in fine-mapping applications; for example, in the blood cell traits considered in this paper, some genetic effects are specific to subsets of the traits (e.g., red blood cell traits). There are also

Table 1. Overview of available statistical methods for multi-trait fine-mapping.

method	upper limit on number of causal SNPs	data accepted		CSs	allows correlated traits	models effect sharing	sample runtimes		software	version
		summary	sufficient				2 traits	20 traits		
mvSuSiE	user-specified	yes	yes	yes	yes	yes	41 s	2 min	R	9f28916
flashfm† [39]	10	yes	yes	yes	yes	yes	5 min	–	R	0.0.0.9000
MT-HESS [55]	no limit	no	no	no	yes	yes	>1 day	–	R	1.99
BayesSUR [103]	no limit	no	no	no	yes	yes	7 h	–	R	2.0-1
msCAVIAR [53]	user-specified	yes	no	no	no	yes	>1 day	–	command-line	0.1
CAFEH [3]	user-specified	yes	yes	yes	no	no	20 s	37 s	Python	1.0
PAINTOR [50]	user-specified	yes	no	no	no	no	30 min	–	command-line	3.1
MFM§ [5]	user-specified	no	no	yes	no	no	–	–	R	0.2-1
HyPrColoc [29]	1	yes	no	yes‡	no	no	<1 s	<1 s	R	1.0
moloc [34]	1	yes	no	yes‡	no	no	<1 s	–	R	0.1.0

Sample runtimes were obtained by running on data sets with J = 5,000 SNPs, N = 250,000 individuals (relevant for methods that do not accept summary data), and R = 2 traits or 20. When possible, the upper limit on the number of causal SNPs, L, was set to 10. In our tests, PAINTOR ran for a very long time when allowing 3 or more causal SNPs, so we set L = 2. (This was without the "MCMC" option, because at the time of our experiments the "MCMC" option produced unreasonable results.) moloc was computationally impractical with more than 4 traits [29]. See the Supplementary Note for more details. §MFM is specific to multiple case-control traits with a shared set of controls. †flashfm's properties depend on the single-trait fine-mapping method; to illustrate, the properties shown here are for FINEMAP [8, 9]. flashfm with FINEMAP was limited to at most 5 traits. ‡Calculation of CSs is trivial when limiting to at most 1 causal SNP.

several methods developed for the problem of colocalization of two traits [22, 30, 33, 43, 45, 94, 100], which has different analysis aims, but overlaps with multi-trait fine-mapping.

Here we introduce mvSuSiE, a fast and flexible method for multi-trait fine-mapping. The name "mvSuSiE" evokes its origins as an extension of the "Sum of Single Effects" (SuSiE) model [96] to the multivariate analysis setting. In particular, mvSuSiE combines the SuSiE model with ideas from [89] to learn, in a flexible way, the patterns of shared genetic effects among traits. In this way, mvSuSiE automatically adapts to the patterns of effect sharing in the particular traits being analyzed, making it widely applicable to fine-mapping any set of related traits (and potentially even unrelated traits). We also leverage ideas from [106] to allow the analysis of summary statistics generated from a genetic association study, which are often more readily accessible than individual-level data [49, 70]. mvSuSiE is computationally practical for jointly fine-mapping many traits in "biobank scale" data. We demonstrate its effectiveness compared with existing methods using simulations, and by fine-mapping 16 blood-cell traits in 248,980 UK Biobank samples.

Results

Methods overview

Consider fine-mapping R traits in a region containing J SNPs (or other biallelic loci). For each individual $i = 1, \ldots, N$, let y_{ir} denote trait r measured individual i, and let x_{ij} denote the genotype of individual i at SNP j, encoded as the number of copies of the minor allele. We perform multi-trait fine-mapping using the following multivariate linear regression model:

$$y_{ir} = \mu_r + \sum_{j=1}^{J} x_{ij} b_{jr} + e_{ir}, \tag{1}$$

where μ_r reflects the mean of trait r, b_{jr} is the effect of SNP j on trait r, and the e_{ir}'s are normally-distributed error terms (which may be correlated among the R traits). Within this regression model, we frame fine-mapping as a "variable selection problem" [32, 63]: most SNPs are assumed to have no effect on any trait—that is, most effects b_{jr} are zero—and the goal of multi-trait fine-mapping is to identify which SNPs have a non-zero effect on which traits, and to assess uncertainty in these inferences. (For brevity, we use the term "causal SNP" to mean a SNP with non-zero effect.)

Our new mvSuSiE method achieves this goal by extending the *Sum of Single Effects* (SuSiE) model [96] to the multivariate setting. By using ideas from [106], it can perform fine-mapping using either

individual-level data (genotypes and phenotypes) or summary data (e.g., LD matrix and marginal z-scores); see Online methods for details.

Among existing approaches to fine-mapping, mvSuSiE is most closely related to CAFEH [3], which also extends SuSiE to perform multi-trait fine-mapping. Both CAFEH and mvSuSiE inherit much of the simplicity and benefits of single-trait SuSiE. Like SuSiE, both mvSuSiE and CAFEH require the user to specify an upper bound, L, on the number of causal SNPs in a region, and are robust to this upper bound being larger than needed. And both methods exploit SuSiE's simple fitting procedure, Iterative Bayesian Stepwise Selection [96], which is similar to simple forward stepwise selection, but improves on it by (i) using Bayesian computations to take into account uncertainty in which SNPs are selected at each step; and (ii) iterating through selection events to allow errors in initial selections to be corrected as fitting progresses.

However, mvSuSiE improves on CAFEH in two key ways:

(a) mvSuSiE uses a flexible prior distribution—a mixture of multivariate normal distributions, as in [89]—to model effect sharing patterns across traits. Further, the parameters of this prior are estimated from the data, allowing mvSuSiE to adapt to each data set. This flexible approach allows that different causal SNPs may show different patterns of association; for example, in analyses of blood cell traits shown later, mvSuSiE learns that some SNPs may affect primarily red blood cell (erythrocyte) traits, some may affect primarily white blood cell (leukocyte) traits, and some may affect both, or a subset of one or the other. In contrast, CAFEH assumes a less flexible and less adaptive prior in which causal effects are independent across traits.

(b) mvSuSiE allows for correlations in measurements among traits, with these correlations again being estimated from the data. In contrast, CAFEH assumes measurements are independent across traits, which is often not the case because association studies often involve correlated traits.

For (a), estimating the prior distribution from the data involves combining information across many causal SNPs from many regions, which is an additional step compared with standard single-trait fine-mapping analyses. This additional step can be avoided by using a simpler fixed prior (see Online methods) but at potential loss of power.

We also introduce novel ways to summarize the inferences from multi-trait fine-mapping. Again, this builds on SuSiE, which summarizes single-trait results by reporting, for each SNP, a "posterior inclusion probability" (PIP) quantifying the probability that the SNP is causal, and by reporting "credible sets" (CSs) [59, 96] that are designed to capture, with high probability, at least one causal SNP. Informally, each CS represents an independent association signal in the data, and the size of a CS (*i.e.,* the number of SNPs in the CS) indicates how precisely one can pinpoint the causal SNP underlying this signal. For multi-trait analyses, it may seem natural to report PIPs and CSs separately for each trait. However, this raises tricky issues; for example, if the reported CSs for two traits overlap, do these represent the "same" signal, with a single underlying causal SNP, or different signals with multiple causal SNPs? To avoid this problem, we separate inference into two questions.

First question: Which SNPs are causal for *at least one* trait? This question is answered by *cross-trait PIPs and CSs* that summarize the inferences across all traits.

Second question: For each causal SNP (*i.e.*, CS) identified, which traits does it affect? This is answered by computing a *trait-specific* measure of significance, the *local false sign rate* (*lfsr*) [82, 89], for each SNP in each trait (with small *lfsr* indicating high confidence in the *sign* of the effect). Because SNPs in a CS are typically in high LD, their trait-specific *lfsr* values are typically similar, and it is convenient to use a single number, the *average lfsr*, as a trait-specific measure of significance of each CS. If the average *lfsr* for trait r is small, this indicates high confidence in the sign of the effect, and we say the CS is "significant for trait r."

In summary, the reported results from a mvSuSiE analysis are the *cross-trait* PIPs and CSs together with *trait-specific* measures of significance (*lfsr*) for each SNP and each CS in each trait. Fig. 1 summarizes the mvSuSiE analysis workflow for a typical genetic association study.

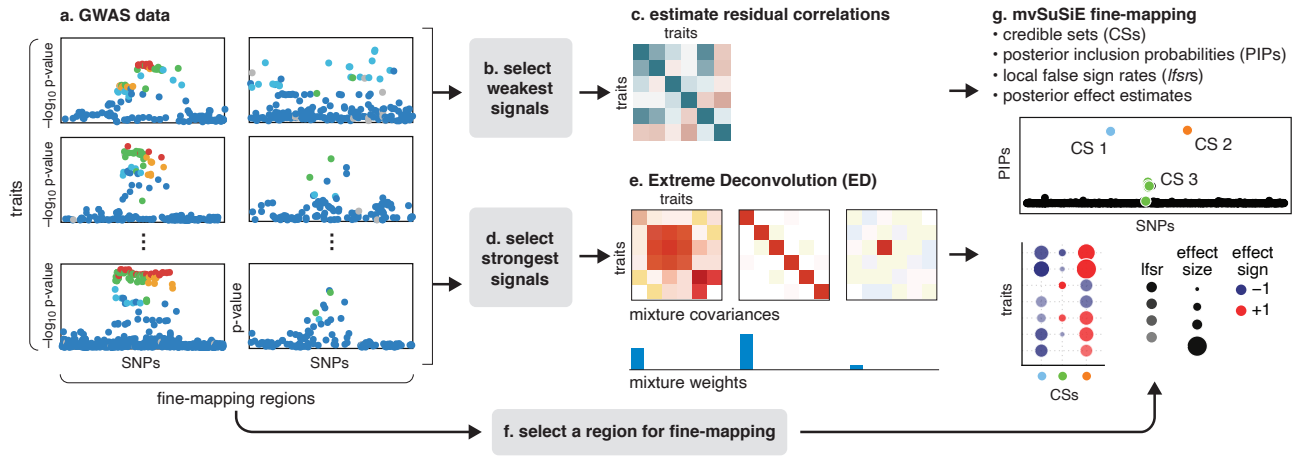


Fig. 1. Overview of multivariate fine-mapping using mvSuSiE. mvSuSiE accepts as input traits and SNP genotypes measured in N individuals, R traits and M target fine-mapping regions. Alternatively, mvSuSiE-RSS accepts SNP-level summary statistics computed from these data (a). The weakest SNP association signals are extracted from these data (b), which are used to estimate correlations in the trait residuals (c). Separately, the strongest association signals are extracted (d) to estimate effect sharing patterns using Extreme Deconvolution (ED) [12] (e). Finally, the effect sharing patterns estimated by ED, together with the estimated weights, are used to construct a prior for the unknown multivariate effects, and this prior is used in mvSuSiE to perform multivariate fine-mapping simultaneously for all SNPs in a selected region (g). Steps f and g are repeated for each fine-mapping region of interest. The key mvSuSiE outputs are: (i) a list of credible sets (CSs), each of which is intended to capture a distinct causal SNP; (ii) a posterior inclusion probability (PIP) for each SNP giving the probability that the SNP is causal for at least one trait; (iii) average local false sign rates (*lfsr*s) summarizing significance of each CS in each trait; and (iv) posterior estimates of SNP effects on each trait. For example, if a region contains 3 distinct causal SNPs, mvSuSiE will, ideally, output 3 CSs, each containing a true causal SNP, with the average *lfsr*s indicating which traits are significant for each CS.

Evaluation in simulations using UK Biobank genotypes

We compared mvSuSiE with existing multi-trait fine-mapping methods and a single-trait fine-mapping method, SuSiE [96, 106], in simulations. Among available multi-trait fine-mapping methods (Table 1), MT-HESS [55] and BayesSUR [11, 103, 104] are similar to mvSuSiE in features and modeling assumptions, but are computationally impractical for large fine-mapping data sets. msCAVIAR [53] shares the ability of mvSuSiE to model effect sharing, but is designed for analyzing data from multiple studies, and therefore makes modeling assumptions that are less appropriate for analyzing multiple traits. MFM [5] is another multi-trait fine-mapping method, but is specific to multiple case-control traits with a shared set of controls. Therefore, we focussed our comparisons on CAFEH [3], which can handle large multi-trait fine-mapping data sets. We also compared with flashfm [39] and PAINTOR [50] on smaller fine-mapping data sets with two traits.

To make our simulations reflective of current large-scale genomic data sets, we obtained imputed genotype data from the UK Biobank [15, 84] and simulated quantitative traits with 1–5 simulated causal SNPs in each fine-mapping region. We simulated from a variety of effect sharing patterns, with effect sizes scaled to roughly reproduce the distributions of z-scores observed in genome-wide association analyses of complex traits from UK Biobank data. The fine-mapping regions were drawn from autosomal chromosomes and varied in size (0.4–1.6 Mb), number of SNPs (1,000–5,000 SNPs) and LD patterns.

We simulated traits under two scenarios:

(a) "Trait-specific + Shared Effects," in which the SNP effects on 20 independent traits were either specific to one trait, or shared among traits in simple ways (e.g., equal effects on a pair of traits and no effect on the remaining traits);

(b) "Complex Shared Effects," in which the SNP effects on 16 correlated traits were generated from a variety of sharing patterns derived from the UK Biobank blood cell traits.

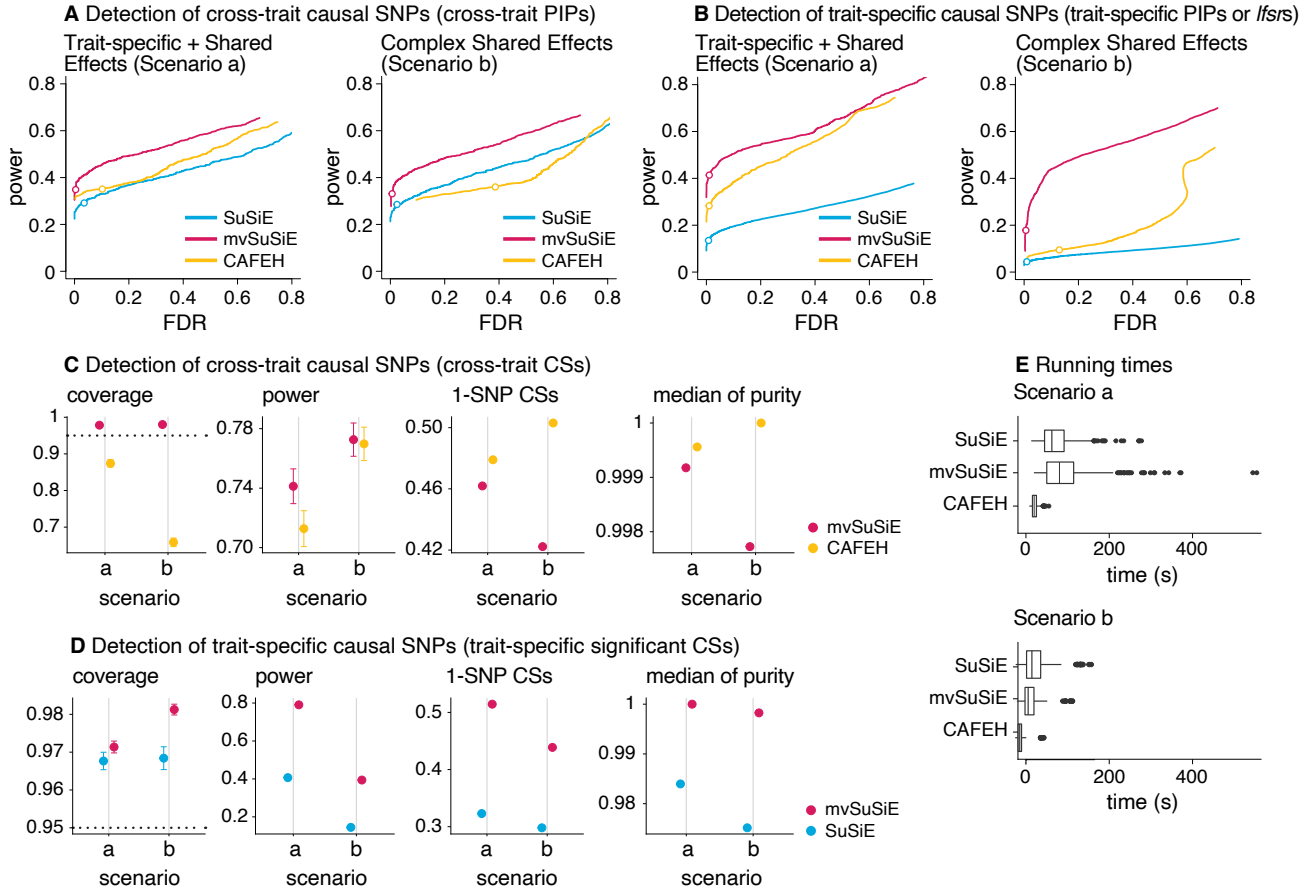


Fig. 2. Comparison of methods in simulated data. Panels A and B show power vs. FDR (*i.e.*, precision-recall with flipped x-axis) in identifying causal SNPs, either cross-trait (A) or trait-specific (B), using SNP-wise measures. In each scenario, FDR and power were calculated from 600 simulations as the measure threshold is varied from 0 to 1. Open circles are drawn at a PIP threshold of 0.95 or *lfsr* threshold of 0.05. The specific SNP-wise measures used in A are PIP (mvSuSiE, CAFEH), max-PIP (SuSiE); in B, PIP (SuSiE), *min-lfsr* (mvSuSiE) and "study PIP" (CAFEH). Power and FDR were calculated as FDR $:= \frac{\text{FP}}{\text{TP}+\text{FP}}$ and power $:= \frac{\text{TP}}{\text{TP}+\text{FN}}$, where FP, TP, FN, TN denote, respectively, the number of false positives, true positives, false negatives and true negatives. Panels C and D evaluate the estimated CSs using the following metrics: *coverage*, the proportion of CSs containing a true causal SNP; *power*, the proportion of true causal SNPs included in at least one CS; the proportion of CSs that contain a single SNP ("1-SNP CSs"); and *median purity*, where "purity" is defined as the smallest absolute correlation among all SNP pairs in a CS. Target coverage (95%) is shown as a dotted horizontal line. In D, mvSuSiE detected a trait-specific effect if the average *lfsr* in the CS was less than 0.05. Error bars show 2 times the empirical standard error from the results in all simulations. Panel E summarizes runtimes; the SuSiE runtimes are for running SuSiE independently on all traits.

To compare with PAINTOR and flashfm we also simulated a third smaller set of data with 2 independent traits and shared effects.

We compared methods in their detection of cross-trait causal SNPs—in which we define a cross-trait causal SNP as one that affects at least one trait—and trait-specific causal SNPs. We assessed the performance of both SNP-wise measures (e.g., PIPs) and Credible Sets (CSs) for these tasks. Since PIPs and CSs are typically used in combination in a fine-mapping analysis—CSs identify the independent causal signals, then PIPs identify the strongest causal SNP candidates within each CS—this experiment evaluates the two main aspects of a typical fine-mapping analysis.

In all our comparisons, mvSuSiE improved power, coverage and resolution (purity and proportion of 1-SNP CSs) over the SuSiE single-trait analyses (Fig. 2A, B, D). The greatest gains were in Scenario b, where mvSuSiE had the advantage that it accounted for correlations among traits. Comparing CAFEH and single-trait SuSiE in SNP-wise inferences (Fig. 2A, B), CAFEH improved performance

in Scenario a, but performed slightly less well for detecting causal SNPs in Scenario b, where it produced poorly calibrated PIPs (Fig. 2A, B, Supplementary Fig. 5). Our explanation for this is that Scenario b contradicts CAFEH's assumptions of independent traits and independent causal effects. In support of this, when we forced mvSuSiE to make the same independence assumptions as CAFEH, the performance gains were greatly reduced and the PIPs were also poorly calibrated (Supplementary Figures 1, 2, 5). These results illustrate the benefits of having a flexible model that can adapt to different fine-mapping scenarios by learning effect sharing patterns from the data (Supplementary Figures 1, 6–9). This flexibility comes at a computational cost—CAFEH was consistently faster than mvSuSiE (Fig. 2E)—but mvSuSiE was fast enough to easily handle the largest fine-mapping data sets we considered.

We also compared mvSuSiE with PAINTOR and flashfm in smaller scale fine-mapping data sets with 2 traits, focusing on SNP-wise fine-mapping measures (PAINTOR does not provide CSs). Even though the two traits were simulated independently to conform with PAINTOR's modeling assumptions, PAINTOR still had much lower power to detect causal SNPs than both SuSiE and mvSuSiE (Supplementary Fig. 3A). Both flashfm and mvSuSiE improved power over the SuSiE single-trait analysis, but mvSuSiE achieved the greater gains in power (Supplementary Fig. 3B). mvSuSiE also had considerably lower computational cost than PAINTOR and flashfm (Supplementary Fig. 4).

In summary, these simulations demonstrate the benefits of mvSuSiE as an efficient and flexible multi-trait fine-mapping method. In particular, mvSuSiE consistently increased power to detect causal SNPs, and improved precision (reduced CS size) compared with fine-mapping each trait separately.

Multi-trait fine-mapping of blood cell traits from UK Biobank

To illustrate mvSuSiE in a substantive application, we fine-mapped blood cell traits using data from the UK Biobank [15, 84]. Previous analyses of these data include association analyses [6, 48] and single-trait fine-mapping [88, 92], but multi-trait fine-mapping using mvSuSiE has the potential to improve power and precision of fine-mapping. Multi-trait fine-mapping is also better to answer questions about shared genetic effects (which SNPs affect which traits) and hence provide insights into the underlying biology.

Focusing on a subset of 16 blood cell traits (Supplementary Table 1 in Online Methods), we performed standard PLINK association analyses [19] with 248,980 UK Biobank samples for which all 16 traits and imputed genotypes were available (Online Methods). We included covariates such as sex and age, as well as genotype principal components to limit spurious associations due to population structure. From the results of these association analyses, we obtained 975 candidate genomic regions for fine-mapping (Supplementary Table 2) with the property that all SNPs within 250 Kb of each significant genetic association (p-value $< 5 \times 10^{-8}$) belonged to exactly one of the candidate regions (except for the MHC, which we did not fine-map). We then applied the mvSuSiE analysis pipeline to these 975 candidate regions (Online Methods). To understand the benefits of a multi-trait fine-mapping, we also ran SuSiE on the same regions, separately for each trait.

Genetic relationships among blood traits inform discovery of multi-trait causal SNPs. From the 975 candidate regions, mvSuSiE identified 3,396 independent causal signals (95% cross-trait CSs). The median size of a CS was 7 SNPs. Among these CSs, 726 contained just one SNP ("1-SNP CS"), therefore mvSuSiE identified 726 high-probability (95%) candidate causal SNPs. The number of CSs significant in each trait ($lfsr < 0.01$) ranged from 370 (basophil percentage) to 1,423 (platelet count), and the number of 1-SNP CSs ranged from 108 to 335 (Fig. 3D). (Note that 10 of the 3,396 CSs were not significant in any traits at $lfsr < 0.01$.) By combining information across traits, mvSuSiE greatly increased fine-mapping discovery and resolution compared to the SuSiE single-trait fine-mapping; the number of trait-specific significant CSs increased, on average, 2.2-fold compared with SuSiE, and the number of trait-specific significant 1-SNP CSs increased, on average, 3.5-fold (Fig. 3D).

A key way in which mvSuSiE increases discovery and resolution is by learning about patterns of shared (and non-shared) genetic effects from the data. In these data, the most prominent learned patterns involved strong sharing of effects amongst traits for the same blood cell type (Fig. 3E).

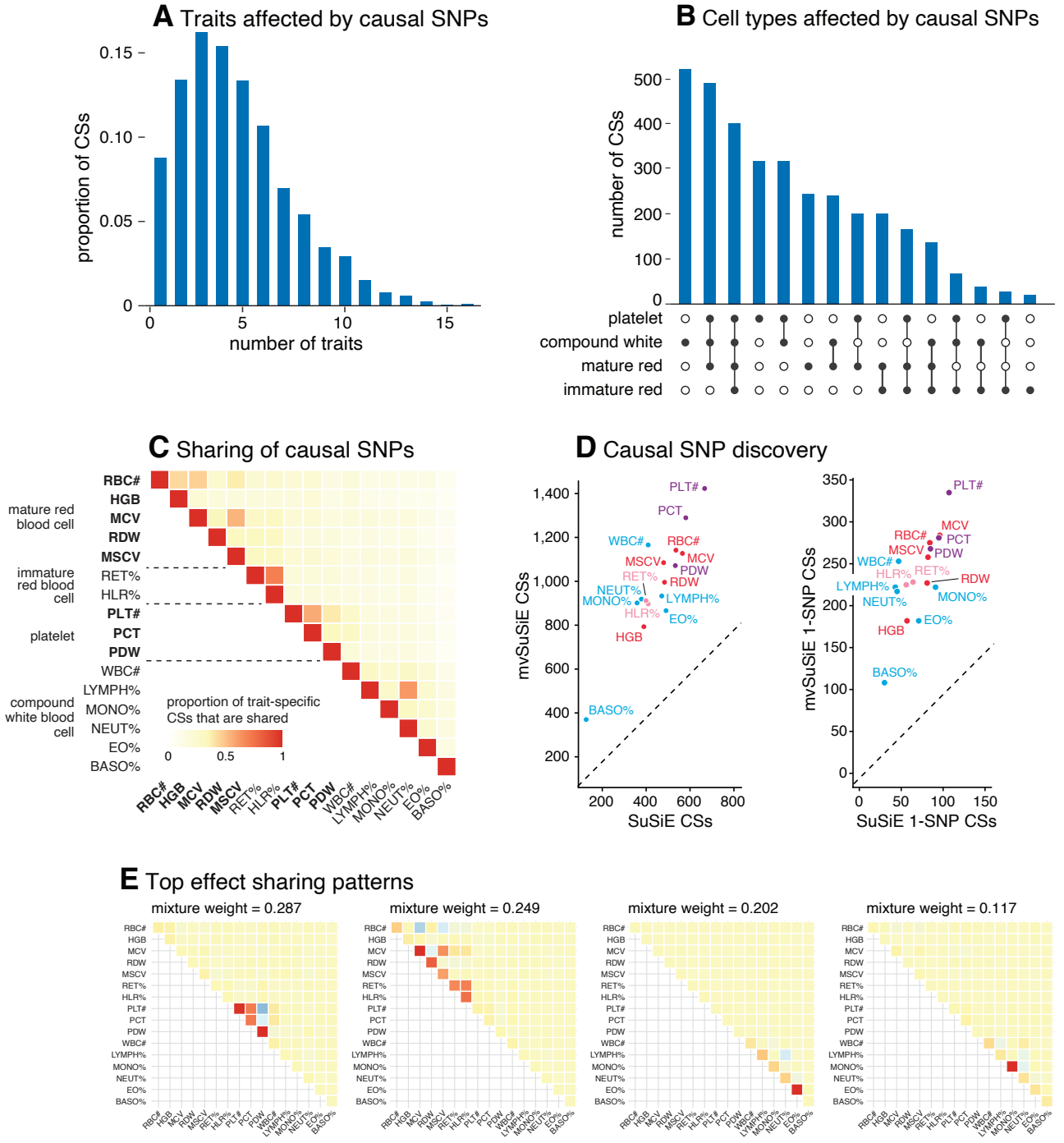


Fig. 3. mvSuSiE fine-mapping and primary effect sharing patterns in UK Biobank blood cell traits. Panels A–C give summaries of the 3,396 mvSuSiE CSs identified from the 975 candidate fine-mapping regions: (A) number of significant (*lfsr* < 0.01) traits in each CS; (B) significant traits in CSs grouped by blood cell type; (C) pairwise sharing of significant CSs among the traits. In C, for each pair of traits we show the ratio of the number of CSs that are significant in both traits to the number of CSs that are significant in at least one trait. (D) Number of CSs and 1-SNP CSs for each trait identified by SuSiE and mvSuSiE (after removing CSs with purity less than 0.5). In D, each mvSuSiE count is the number of mvSuSiE CSs or 1-SNP CSs that are significant for the given trait. (E) Covariance matrices in the mvSuSiE data-driven prior capturing the primary sharing patterns (these are the covariance matrices with the largest mixture weights in the prior). The covariance matrices were scaled separately for each plot so that the plotted values lie between -1 and 1. See Supplementary Fig. 7 for the full prior.

However, many other patterns were also identified (Supplementary Fig. 7), including both trait-specific and broad effects, suggesting that SNPs can affect blood cells in a wide variety of ways, presumably reflecting a wide variety of underlying biological mechanisms. By applying mvSuSiE with a prior that incorporates these learned sharing patterns, we obtain a genome-wide summary that underscores the diversity of genetic effects on blood cell traits (Fig. 3A–C). Genetic effects are most commonly shared among traits within the same blood cell type as one might expect (Fig. 3C), but SNPs affecting multiple blood cell types are also common (Fig. 3B).

Multi-trait fine-mapping reveals highly heterogeneous genetic determination of blood traits. To illustrate the potential for mvSuSiE to help dissect complex genetic association signals, we examine four example regions in more detail (Fig. 4).

Figure 4A shows the mvSuSiE results for the *EXT1–SAMD12* locus. Single-trait association analysis of this region shows only one trait, basophil percentage, with a genome-wide significant association (PLINK p-value $< 5 \times 10^{-8}$). Similarly, single-trait fine-mapping with SuSiE identified a single CS for basophil percentage containing 10 candidate SNPs, and no CSs in other traits. From these results one might conclude that the causal SNP is specific to basophil percentage. However, the mvSuSiE fine-mapping results assess the CS as significant in most traits, suggesting that in fact the causal SNP has broad effects across many traits. (Indeed, all traits had marginal association p-values less than 0.003 with the lead SNP, which in other settings might be considered "significant".) Also, the mvSuSiE CS is smaller than the single-trait CS (8 vs. 10 SNPs), illustrating the improved fine-mapping resolution that comes from combining information across traits.

Figure 4B shows mvSuSiE results for the Tensin 3 (*TNS3*) locus. Vuckovic *et al* [92] used single-trait fine-mapping to identify causal signals for several red and white blood cell traits in this region. However, a single-trait analysis does not tell us whether these signals are due to one or a few causal SNPs affecting many blood cell traits, or many causal SNPs affecting individual traits. The multi-trait mvSuSiE analysis identified three causal signals (cross-trait CSs) with three distinct patterns of genetic effect: one mostly affects red blood cell traits (CS3); another has a detectable effect in HLR% only (CS1); and a third has smaller effects in both white blood cell and platelet traits (CS2). The three very different patterns suggest that the biological effects of these SNPs are also very different, and suggest a multi-faceted role for *TNS3* in affecting blood-cell traits. This example illustrates the flexibility of mvSuSiE, including its ability to capture different patterns of effect sharing even within a single locus, and its ability to extract relatively simple inferences in quite complex situations.

Figure 4C shows a more complex example involving many signals in and around the gene *RUNX1*. SNPs in the *RUNX1* locus have previously been associated with rheumatoid arthritis [76, 87] and other immune-related diseases [4, 13, 38], and colocalization analyses have suggested that the causal SNPs are also associated with eosinophil proportions in blood [92]. Multi-trait fine-mapping results from mvSuSiE suggest a complex picture with 11 signals (cross-trait CSs), each with detectable effects in many different blood-cell traits, and some with no detectable effect on eosinophil proportions. These results suggest that the mechanisms by which this gene affects immune-related diseases is likely more complex than just through eosinophils, possibly involving many platelet, red blood cell and other white blood cell traits.

Finally, Figure 4D shows a more complex example still, where many causal signals are mapped to a region containing many genes, including *PIEZO1* and *ZFPM1*. This is a gene-dense region with well-studied connections to blood cell traits and blood-related diseases [1, 16, 25, 57, 58, 68, 69, 90]. mvSuSiE identified 14 independent signals (cross-trait CSs) in the region. These 14 signals show a wide variety of effect patterns: for example, some are significant only in a few traits related to mature red blood cells (e.g. CS12, CS14), some are significant across a broader range of red blood cell traits (CS2), and some are significant across most traits (CS13). Regions of this complex may take considerable additional investigation to fully understand. Although this is a complex example, we note that of the 14 CSs identified in this region, 7 contain a single SNP, demonstrating that even in complex regions mvSuSiE can identify high-confidence causal SNPs.

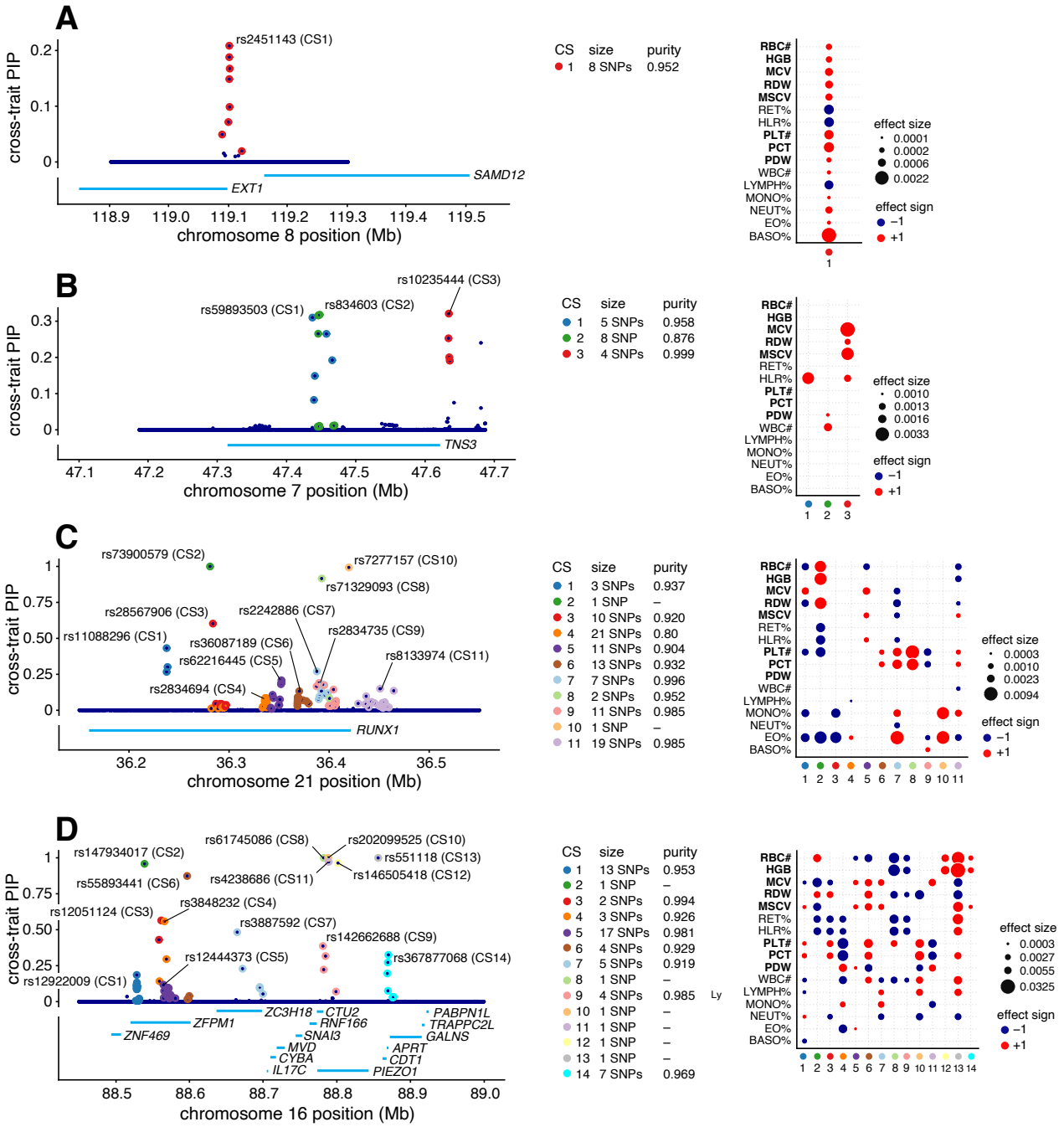


Fig. 4. Examples of blood cell trait loci fine-mapped using mvSuSiE. The left-hand plots are "PIP plots" showing the posterior inclusion probabilities (PIPs) for each SNP analyzed in the given fine-mapping regions. The PIP is an estimate of the probability that the SNP is causal for at least one trait. The labeled SNPs are the "sentinel SNPs", the SNPs with the highest cross-trait PIP in each CS. "Purity" is defined as the minimum absolute pairwise correlation among SNPs in the CS. The right-hand plots show the effect estimates of the sentinel SNPs whenever the CS is significant for the given trait ($lfsr < 0.01$).

Discussion

We have introduced mvSuSiE, a fast and flexible multi-trait fine-mapping method. mvSuSiE outperforms single-trait fine-mapping methods in both power and the resolution to fine-map causal effects. Unlike most available multi-trait fine-mapping methods, mvSuSiE can efficiently analyze dozens of correlated traits, and can model complex patterns of effect size variation via a flexible data-driven prior distribution. The prior model also includes as special cases several simpler models that are

commonly used in meta-analyses, such as the fixed effects model which assumes equal effects in all traits, and the random effects model which allows for different effect sizes among traits [37]. These models can be used in place of the data-driven prior to speed computation if users desire, though at a potential loss of power.

A limitation of the current mvSuSiE model is that each trait is assumed to be measured on all samples. That is, all y_{ir} must be observed. (In terms of summary statistics, the z-scores should be calculated using the exact same samples for each trait.) A general approach to deal with missing data could expand the method's applicability to other settings such as colocalization and cross-study/cohort fine-mapping.

The mvSuSiE model is "sparse" in that it assumes a small number of causal SNPs. However, the data-driven prior model for the effect sizes at these causal SNPs will not generally be sparse. That is, each causal SNP is effectively assumed to affect all traits, with some exceptions (such as when the prior includes sharing patterns reflecting an effect in one trait and no effects in the others). Instead of inducing trait-specific sparsity on the effects of causal SNPs, we focussed on *estimating* these effects, and assessing their significance by computing the *lfsr*s. This approach simplifies computation, and worked well in our examples. Indeed, in additional simulation studies we found that mixture models constructed using our data-driven approach could capture the predominant sparsity patterns reasonably well (Supplementary Fig. 8), and so mvSuSiE did not suffer from a loss of power to detect such sparse association signals (Fig. 2). That being said, it is possible that explicitly modeling trait-specific sparsity of causal SNPs could be helpful in settings with large numbers of traits that are less related; with a large number of less related traits, the SNP effects may be shared primarily among small subsets of more related traits. This could perhaps be achieved by combining the mvSuSiE prior with indicator variables for each trait, similar to what was done in CAFEH.

URLs

SuSiE R package, `https://github.com/stephenslab/susieR`; mvSuSiE R package, `https://github.com/stephenslab/mvsusieR`; mashr R package, `https://github.com/stephenslab/mashr`; flashr R package, `https://github.com/stephenslab/flashr`; CAFEH Python package, `https://github.com/karltayeb/cafeh`; PAINTOR software, `https://github.com/gkichaev/PAINTOR_V3.0`; BayesSUR R package, `https://cran.r-project.org/package=BayesSUR`; flashfm R package, `https://github.com/jennasimit/flashfm`; FINEMAP software, `http://www.christianbenner.com`; msCAVIAR software, `https://github.com/nlapier2/MsCAVIAR`; hyprcoloc R package, `https://github.com/cnfoley/hyprcoloc`; moloc R package, `https://bogdan.dgsom.ucla.edu/pages/MOLOC`; PLINK, `https://www.cog-genomics.org/plink2`; R, `https://cran.r-project.org`; Python, `https://www.python.org/`; DSC software, `https://github.com/stephenslab/dsc`; UK Biobank, `https://www.ukbiobank.ac.uk`; code for processing of UK Biobank data, `https://github.com/stephenslab/finemap-uk-biobank`; PLINK association test statistics for UK Biobank blood traits, [Zenodo or Data Dryad repository will be created for this and URL will be provided]; code and data resources for fine-mapping simulations and fine-mapping analyses of UK Biobank blood cell traits, `https://github.com/gaow/mvarbvs`; simulation results, `https://github.com/zouyuxin/mmbr-rss-dsc`.

Acknowledgements

We thank Karl Tayeb for helping with CAFEH. We thank the staff at the Research Computing Center and the Center for Research Informatics at the University of Chicago for maintaining the high-performance computing resources used to implement the numerical experiments. This work was supported in part by NIH grant 5R01HG002585 and by a grant from the Gordon and Betty Moore Foundation to MS. GW was supported by NIH grant AG076901 and funding from the Thompson Family Foundation (TAME-AD). This research has been conducted using the UK Biobank Resource under Application Number 27386.

Author contributions

YZ, GW and MS conceived of the project and developed the statistical methods. YZ implemented the methods comparisons in simulations and performed the statistical analyses. YZ and GW wrote the software, with contributions from MS and PC. YZ and GW prepared the online code and data resources. All authors prepared the results for the manuscript and wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Online methods

Multivariate multiple regression

mvSuSiE is based on a basic multivariate multiple regression model for R quantitative traits observed in N individuals,

$$\boldsymbol{Y} \sim MN_{N\times R}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_N, \boldsymbol{V}), \tag{2}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N\times R}$ is a matrix storing N observations of R traits, $\boldsymbol{X} \in \mathbb{R}^{N\times J}$ is a matrix of N genotypes at J SNPs, $\boldsymbol{B} \in \mathbb{R}^{J\times R}$ is a matrix of regression coefficients ("effects") for the J SNPs and R traits, $\boldsymbol{V}$ is an $R \times R$ covariance matrix (we assume $\boldsymbol{V}$ is invertible), $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $MN_{N\times R}(\boldsymbol{M}, \boldsymbol{\Sigma}^{\text{row}}, \boldsymbol{\Sigma}^{\text{col}})$ denotes the matrix normal distribution [26, 36] with mean $\boldsymbol{M} \in \mathbb{R}^{N\times R}$ and covariance matrices $\boldsymbol{\Sigma}^{\text{row}}$, $\boldsymbol{\Sigma}^{\text{col}}$ (of dimension $N \times N$ and $R \times R$, respectively).

Intercept. We do not explicitly include an intercept in (2). Instead, we account for an intercept implicitly by "centering" the columns of $\boldsymbol{X}$ and the columns of $\boldsymbol{Y}$ so that the mean of each column is zero. From a Bayesian perspective, centering the columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is equivalent to integrating with respect to an (improper) uniform prior on the intercept. (This is a multivariate generalization of the result for univariate regression given in [32]. See the Supplementary Note for a more formal proof of this result.) In short, centering eliminates the need to explicitly include an intercept in (2), and we proceed with mvSuSiE assuming that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have been centered.

The mvSuSiE model

mvSuSiE generalizes the "Sum of Single Effects" (SuSiE) model [96] to the multivariate setting:

$$\boldsymbol{B} = \sum_{l=1}^{L} \boldsymbol{B}^{(l)} \tag{3}$$

$$\boldsymbol{B}^{(l)} = \boldsymbol{\gamma}^{(l)} \otimes \boldsymbol{b}^{(l)}, \tag{4}$$

where $\boldsymbol{\gamma}^{(l)} \in \{0,1\}^J$ is a vector of indicator variables in which exactly one of the J elements is one and the remaining are zero, $\boldsymbol{b}^{(l)} \in \mathbb{R}^R$ is a vector of regression coefficients, and $\boldsymbol{u} \otimes \boldsymbol{v} := \boldsymbol{u}\boldsymbol{v}^\intercal$ denotes the outer product of (column) vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. The coefficients $\boldsymbol{B}$ defined in this way are a sum of L "single effects" $\boldsymbol{B}^{(l)}$. In particular, matrix $\boldsymbol{B}^{(l)} \in \mathbb{R}^{J\times R}$ has at most one row containing non-zero values, and these non-zero values are determined by $\boldsymbol{b}^{(l)}$. We therefore refer to $\boldsymbol{B}^{(l)}$ as a "single effect matrix" because it encodes the effects of a single SNP. The final set of coefficients, $\boldsymbol{B}$, is a matrix with at most L rows containing non-zero values.

Similar to SuSiE, we introduce priors for the indicator variables $\boldsymbol{\gamma}^{(l)}$ and regression coefficients $\boldsymbol{b}^{(l)}$,

$$\boldsymbol{\gamma}^{(l)} \sim \text{Multinom}(1, \boldsymbol{\pi}) \tag{5}$$

$$\boldsymbol{b}^{(l)} \sim g_l, \tag{6}$$

in which $\text{Multinom}(m, \boldsymbol{\pi})$ denotes the multinomial distribution for m random multinomial trials with category probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$, such that $\pi_j \geqslant 0$, $\sum_{j=1}^{J} \pi_j = 1$. The π_j's are the prior inclusion probabilities. By default, we assume a uniform prior; that is, $\pi_j = 1/J$, for $j = 1, \ldots, J$. (All the

results in this paper use this default prior.) Our software implementation of mvSuSiE also support for other choices of $\boldsymbol{\pi}$; for example, $\boldsymbol{\pi}$ could be determined by external biological information about the SNPs [e.g., 77].

The prior distribution g_l for each single effect $\boldsymbol{b}^{(l)}$ should capture the variety of effect sharing patterns we expect from the multiple traits. To this end, we use a prior similar to the mixture of multivariate normals prior introduced in [89],

$$g_l(\boldsymbol{b}) = \sum_{k=1}^{K} \omega_k N_R(\boldsymbol{b}; \mathbf{0}, \sigma_{0l}^2 \boldsymbol{U}_k), \tag{7}$$

in which each $\boldsymbol{U}_k$ is a (possibly singular) covariance matrix, $\sigma_{0l}^2 \geqslant 0$ scales the prior for each single effect l, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$ is a vector of mixture weights, such that $\omega_k \geqslant 0$, $\sum_{k=1}^{K} \omega_k = 1$, and $N_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution on $\boldsymbol{x} \in \mathbb{R}^d$ with mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and $d \times d$ covariance $\boldsymbol{\Sigma}$. The covariance matrices $\mathcal{U} := \{\boldsymbol{U}_1, \ldots, \boldsymbol{U}_K\}$ and the mixture weights $\boldsymbol{\omega}$ must be chosen beforehand, whereas prior scaling parameters $\sigma_{01}^2, \ldots, \sigma_{0L}^2$ are treated as unknown, and are estimated from the data.

In summary, mvSuSiE is a multivariate regression model with a flexible mixture-of-normals prior on the "single effects," $\boldsymbol{b}^{(l)}$. The unknowns of primary interest are the single effect matrices $\boldsymbol{B}^{(l)}$. As we explain in more detail below, we compute a posterior distribution of the single effects, which is then used to compute key fine-mapping statistics, specifically the posterior inclusion probabilities (PIPs) and credible sets (CSs). The scaling factors σ_{0l}^2 are not of primary interest to the fine-mapping ("nuisance parameters"), and are estimated from the data to aid in better posterior estimation of the single effects. Other model parameters, such as the residual covariance matrix $\boldsymbol{V}$, are assumed to be known, or should have been estimated previously. Below we give guidance on choosing these parameters or estimating them from data.

Posterior computation approach

Here we outline our approach to estimating the posterior distribution for the unknowns of primary interest, the single effect matrices $\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}$, building on the ideas introduced in [106]. A more formal mathematical development of the mvSuSiE algorithms is given in the Supplementary Note.

In this section, we assume that the model parameters $\boldsymbol{V}, \boldsymbol{\pi}, \boldsymbol{\omega}$ and $\mathcal{U}$, as well as L, the maximum number of single effects, are known, or have been estimated in earlier steps in the analysis. We also assume in this section that the scaling factors $\sigma_1^2, \ldots, \sigma_L^2$ are known (in the Supplementary Note we describe how the scaling factors are estimated).

The posterior distribution of $\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}$, as in other Bayesian variable selection models, is intractable, and therefore we must resort to numerical approximations. A key consideration is that we would like these computations to scale well to large genetic data sets, which makes intensive Monte Carlo techniques such as Markov chain Monte Carlo [e.g., 23, 27, 31, 35, 41, 64, 102, 105] less attractive. Another key consideration is that we would like accurate estimates of posterior quantities which can be difficult to achieve when many variables (the SNPs) are highly correlated, or correlated in complicated ways, which is typically the case in genetic fine-mapping. These considerations, as well as others, motivated us to develop an alternative posterior computation approach for SuSiE based on variational approximation ideas [96]. The algorithm for performing the approximate posterior computations in SuSiE was called "Iterative Bayesian Stepwise Selection" (IBSS). In this paper, we have extended the SuSiE variational approach to the mvSuSiE model. Therefore, applying the ideas developed in [96] leads to an IBSS algorithm for fitting the mvSuSiE model (Algorithm 1 in the Supplementary Note).

Choice of L

The number of effects, L, is typically not known in advance. However, mvSuSiE is generally robust to misspecification of L so long as L is chosen to be larger than the (true) number of effects. That's because mvSuSiE prunes single effects when they are not needed by estimating the scaling factors σ_{0l}^2

in the prior (7) as zero or close to zero. This approach to estimating the number of single effects, L, by adapting the prior is closely related to "automatic relevance determination" [65, 86], and this same approach was used in SuSiE [96].

Extension of posterior computation approach to work with summary data

The strategy used in [106] to extend SuSiE to summary data is quite general, and we take this same here: first, in "mvSuSiE with sufficient statistics," we describe an algorithm that uses *sufficient statistics*, and produces the same result as running the mvSuSiE IBSS algorithm on the individual-level data; in "mvSuSiE with summary data: mvSuSiE-RSS", we consider summary data that approximate the sufficient statistics, and therefore yield results that do not exactly reproduce mvSuSiE with individual-level data; and since many genetic association studies provide z-scores, or other summary statistics that can be used to compute z-scores, we describe mvSuSiE-RSS with z-scores in greater detail ("Special case when $\boldsymbol{X}, \boldsymbol{Y}$ are standardized: mvSuSiE-RSS with z-scores").

mvSuSiE with sufficient statistics. The data enter the mvSuSiE model only through the likelihood, which from (2) is

$$\ell(\boldsymbol{B}; \boldsymbol{X}, \boldsymbol{Y}) = |2\pi \boldsymbol{V}|^{-N/2} \exp\left\{ -\frac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y}^\intercal \boldsymbol{Y} - 2\boldsymbol{B}^\intercal \boldsymbol{X}^\intercal \boldsymbol{Y} + \boldsymbol{B}^\intercal \boldsymbol{X}^\intercal \boldsymbol{X}\boldsymbol{B})] \right\}. \tag{8}$$

Here we treat $\boldsymbol{V}$ as a fixed quantity so we do not explicitly mention this dependency in the notation for the likelihood. It is clear from this expression that the data influence the likelihood only through the quantities $\boldsymbol{X}^\intercal \boldsymbol{Y}$ and $\boldsymbol{X}^\intercal \boldsymbol{X}$. Therefore, $\boldsymbol{X}^\intercal \boldsymbol{Y}$ and $\boldsymbol{X}^\intercal \boldsymbol{X}$ are *sufficient statistics* for $\boldsymbol{B}$. Thus, by rearranging the computations, we obtain a variant of the IBSS algorithm that fits the mvSuSiE model using only sufficient statistics. We call this algorithm "IBSS-ss", and it is outlined in Algorithm 2 in the Supplementary Note.

We use IBSS$(\boldsymbol{X}, \boldsymbol{Y})$ to denote the result of applying the IBSS algorithm (Algorithm 1 in Supplementary Note) to the individual-level data, and we use IBSS-ss$(\boldsymbol{X}^\intercal \boldsymbol{X}, \boldsymbol{X}^\intercal \boldsymbol{Y})$ to denote the result of applying the IBSS-ss algorithm (Algorithm 2 in Supplementary Note) to the sufficient statistics. These two algorithms will give the same result, IBSS$(\boldsymbol{X}, \boldsymbol{Y})$ = IBSS-ss$(\boldsymbol{X}^\intercal \boldsymbol{X}, \boldsymbol{X}^\intercal \boldsymbol{Y})$. However, the computational complexity of the two approaches is different. The computational complexity of one iteration of the IBSS algorithm is $O(L \times (NJR + KJR^3))$, whereas the complexity of a single iteration of the IBSS-ss algorithm is $O(L \times (J^2R + KJR^3))$. Therefore, when $N \gg J$, which is often the case in fine-mapping studies, IBSS-ss will usually be faster. We note, however, that computing the $J \times J$ matrix $\boldsymbol{X}^\intercal \boldsymbol{X}$ can be expensive, and potentially more expensive than running mvSuSiE itself. So IBSS-ss will be more computationally attractive than IBSS if N is much larger than J and if $\boldsymbol{X}^\intercal \boldsymbol{X}$ can be computed efficiently using a software such as PLINK [19] or LDStore [7].

mvSuSiE with summary data: mvSuSiE-RSS. We define mvSuSiE-RSS as the application the IBSS-ss algorithm to the sufficient statistics or approximations to these statistics (e.g., an LD estimate obtained from different genotype data than the genotype data used to obtain the other statistics). Conceptually, this approach combines the mixture prior (7) with an approximation to the likelihood (8). To formalize this, we write the likelihood as a function of the sufficient statistics,

$$\ell_{\mathrm{ss}}(\boldsymbol{B}; \boldsymbol{S}_{xx}, \boldsymbol{S}_{xy}, N) := |2\pi \boldsymbol{V}|^{-N/2} \exp(-\frac{N}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y}^\intercal \boldsymbol{Y}/N - 2\boldsymbol{B}^\intercal \boldsymbol{S}_{xy} + \boldsymbol{B}^\intercal \boldsymbol{S}_{xx}\boldsymbol{B})]), \tag{9}$$

so that

$$\ell_{\mathrm{ss}}(\boldsymbol{B}; \tfrac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{X}, \tfrac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{Y}, N) = \ell(\boldsymbol{B}; \boldsymbol{X}, \boldsymbol{Y}). \tag{10}$$

Replacing $\boldsymbol{S}_{xx} = \frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{X}$ with an estimate $\hat{\boldsymbol{S}}_{xx} \approx \boldsymbol{S}_{xx}$ is therefore the same as replacing the sufficient-statistics likelihood (9) with

$$\ell_{\mathrm{RSS}}(\boldsymbol{B}) := \ell_{\mathrm{ss}}(\boldsymbol{B}; \hat{\boldsymbol{S}}_{xx}, \tfrac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{Y}, N). \tag{11}$$

Note that when $\hat{\boldsymbol{S}}_{xx} = \boldsymbol{S}_{xx}$, the approximation is exact; that is, $\ell_{\text{RSS}}(\boldsymbol{B}) = \ell(\boldsymbol{B}; \boldsymbol{X}, \boldsymbol{Y})$.

In summary, applying mvSuSiE with $\boldsymbol{S}_{xx}$ is equivalent to using the individual-data likelihood (8), and applying mvSuSiE with $\hat{\boldsymbol{S}}_{xx}$ is equivalent to using the approximate likelihood (11).

Special case when $\boldsymbol{X}, \boldsymbol{Y}$ are standardized: mvSuSiE-RSS with z-scores. Now we consider the special case when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are standardized, which is common in genetic association studies. By "standardized", we mean that the columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$ have been scaled to have unit variance; $\sum_{i=1}^{N} x_{ij}^2 = N$, $j = 1, \ldots, J$, and $\sum_{i=1}^{N} y_{ir}^2 = N$, $r = 1, \ldots, R$. (See [106] for exact definitions of the z-scores and the LD matrix $\boldsymbol{R}$.) This is in addition to the assumption, mentioned earlier, that the columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are centered to have means of zero. See [83, 93] for a discussion on the choice to standardize.

With standardized $\boldsymbol{X}$ and $\boldsymbol{Y}$, the sufficient statistics $\boldsymbol{X}^\intercal\boldsymbol{Y}$ and $\boldsymbol{X}^\intercal\boldsymbol{X}$ can be recovered from the sample size, N, the (in-sample) LD matrix, $\boldsymbol{R}$, and the marginal association z-scores, $\hat{z}_{jr}$, which are obtained from simple linear regressions between the traits r and the SNPs j. (The z-scores should ideally be computed using the same samples for each trait so that the correlations among SNPs are same for all traits.) In particular, the sufficient statistics are recovered by the following two equations,

$$\boldsymbol{X}^\intercal\boldsymbol{X} = N \times \boldsymbol{R} \tag{12}$$

$$\boldsymbol{X}^\intercal\boldsymbol{Y} = \sqrt{N} \times \tilde{\boldsymbol{Z}} \tag{13}$$

in which $\tilde{\boldsymbol{Z}}$ denotes the $J \times R$ matrix of "adjusted z-scores",

$$\tilde{z}_{jr} := \frac{N}{N + \hat{z}_{jr}} \times \hat{z}_{jr}. \tag{14}$$

Note that, when the effects are small, $\tilde{z}_{jr} \approx \hat{z}_{jr}$.

Substituting equations (12–13) into the sufficient-statistics likelihood (9) gives

$$\ell_{\text{ss}}(\boldsymbol{B}; \boldsymbol{S}_{xx}, \boldsymbol{S}_{xy}, N) = \ell_{\text{ss}}(\boldsymbol{B}; \mathbf{R}, \tilde{\boldsymbol{Z}}/\sqrt{N}, N). \tag{15}$$

When the in-sample LD matrix $\boldsymbol{R}$ is not available, and is replaced with $\hat{\boldsymbol{R}} \approx \boldsymbol{R}$, the mvSuSiE-RSS likelihood (11) becomes

$$\ell_{\text{RSS}}(\boldsymbol{B}) = \ell_{\text{ss}}(\boldsymbol{B}; \hat{\boldsymbol{R}}, \tilde{\boldsymbol{Z}}/\sqrt{N}, N). \tag{16}$$

In summary, when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are standardized, applying mvSuSiE with $\boldsymbol{R}$ is equivalent to using the individual-data likelihood (8), and applying mvSuSiE with $\hat{\boldsymbol{R}}$ is equivalent to using the approximate likelihood (16).

mvSuSiE posterior statistics

Here we describe the posterior statistics used in an mvSuSiE fine-mapping analysis.

Basic posterior quantities. We start with two basic posterior statistics used to calculate other statistics. The first posterior quantity is the posterior probability that the lth single effect is nonzero for SNP j,

$$\alpha_j^{(l)} := \Pr(\gamma_j^{(l)} = 1 \mid \boldsymbol{X}, \boldsymbol{Y}). \tag{17}$$

The second posterior quantity, denoted $\mathit{clfsr}_{jr}^{((l)}$, is the *local false sign rate* [82, 89] for SNP j in trait r and single effect l *conditioned on* SNP j having a nonzero effect in single effect l,

$$\mathit{clfsr}_{jr}^{(l)} := 1 - \max\{\Pr(b_{jr}^{(l)} > 0 \mid \boldsymbol{X}, \boldsymbol{Y}, \gamma_j^{(l)} = 1), \Pr(b_{jr}^{(l)} < 0 \mid \boldsymbol{X}, \boldsymbol{Y}, \gamma_j^{(l)} = 1)\}. \tag{18}$$

Intuitively, the *clfsr* ("conditional *lfsr*") measures how confident we can be in the sign of SNP j in trait r and single effect l given that SNP j has a nonzero effect in single effect l. A small *lfsr* indicates high confidence in the sign of an effect. The *lfsr* is robust to modeling assumptions [82] which is helpful for reducing sensitivity to the choice of prior.

Cross-trait PIP. The posterior inclusion probability (PIP) is a standard quantity reported by most fine-mapping methods, so PIPs are convenient for comparing performance of different fine-mapping methods. PIPs are also useful for visualizing the fine-mapping signal within a candidate fine-mapping region. For mvSuSiE, we define the PIP for SNP j as the posterior probability that at least one of the regression coefficients for the jth SNP is not zero,

$$\begin{aligned}\mathrm{PIP}_j &:= \Pr(\boldsymbol{b}_j \neq \boldsymbol{0} \mid \boldsymbol{X}, \boldsymbol{Y}) \\ &= 1 - \Pr(\boldsymbol{b}_j = \boldsymbol{0} \mid \boldsymbol{X}, \boldsymbol{Y}) \\ &= 1 - \textstyle\prod_{l=1}^{L}(1 - \alpha_j^{(l)}).\end{aligned} \tag{19}$$

in which $\alpha_j^{(l)}$ is defined in (17).

min-lfsr. The PIP tells us whether or not a SNP has an effect on at least one trait, but it does not tell us *which traits* are affected by the SNP. To quantify this, we calculate a *minimum lfsr* (*min-lfsr*), which we define as the smallest *lfsr* among the L single effects,

$$\textit{min-lfsr}_{jr} := \min_{l \in \{1,\ldots,L\}} \textit{lfsr}_{jr}^{(l)}, \tag{20}$$

in which $\textit{lfsr}_{jr}^{(l)}$ is the (unconditional) *lfsr* for SNP j in outcome r and single effect l,

$$\textit{lfsr}_{jr}^{(l)} := 1 - \max\{\Pr(b_{jr}^{(l)} > 0 \mid \boldsymbol{X}, \boldsymbol{Y}), \Pr(b_{jr}^{(l)} < 0 \mid \boldsymbol{X}, \boldsymbol{Y})\} \tag{21}$$

$$= 1 - \alpha_j^{(l)}(1 - \textit{clfsr}_{jr}^{(l)}), \tag{22}$$

and we use the definition of $\textit{clfsr}_{jr}^{(l)}$ from (18). In other words, SNP j is considered "significant" in trait r if and only if it is significant in at least one of the L effects.

Credible sets. A cross-trait credible set $\mathrm{CS}(\boldsymbol{\alpha}^{(l)}; \rho)$ is defined as a set of SNPs that has probability at least ρ of containing an effect SNP [59]. The calculation of cross-trait CSs is described in [96].

A CS does not indicate *which traits* are affected by the SNPs. To assess significance of a CS for a specific trait r, we compute the *average lfsr*, defined as a weighted average of the conditional *lfsr*'s for all SNPs,

$$\textit{lfsr}_r^{(l)} := \sum_{j=1}^{J} \alpha_j^{(l)} \, \textit{clfsr}_{jr}^{(l)}. \tag{23}$$

If the *average lfsr* for trait r is small, this indicates high confidence in the sign of the effect, and so we say the effects of the SNPs in the CS are *significant for trait* r ("trait-specific CS").

Specifying $\boldsymbol{V}$ *and* g

In order to run mvSuSiE or mvSuSiE-RSS, we must first specify the $R \times R$ residual variance-covariance matrix, $\boldsymbol{V}$, and the prior on the regression coefficients, g. In the next two sections, we describe the steps that were taken to specify these model parameters in the simulations and the UK Biobank blood traits case study. Since we always applied mvSuSiE to summary data ("mvSuSiE-RSS"), and specifically z-scores, we describe estimation of $\boldsymbol{V}$ and g using the z-scores.

Estimating the residual variance matrix. When the traits are measured in the same samples, it is important to account for possible correlations among the measurements of the different traits; failure to do so can result in miscalibrated fine-mapping statistics (Supplementary Fig. 2). The residual covariance matrix $\boldsymbol{V}$ is used to account for correlations among the measurements; the special case of independent measurements can be modeled by setting $\boldsymbol{V}$ to a diagonal matrix.

To estimate $\boldsymbol{V}$, we adapted the approach described in [89], in which $\boldsymbol{V}$ was estimated from z-scores (e.g., z-scores obtained from marginal association tests). Specifically, we took the following steps. First, we pooled the z-scores from all the fine-mapping regions considered. Then we filtered out large (in magnitude) z-scores; specifically, we only considered SNPs in which the largest z-score magnitude for any trait was less than 2. This improved the estimate of $\boldsymbol{V}$ by removing SNPs that might affect one or more of the traits. Denoting the number of SNPs used in this calculation by J', and letting $\hat{\boldsymbol{z}}_j$ denote the vector of z-scores obtained from the R association tests for SNP j, we estimated $\boldsymbol{V}$ as

$$\hat{\boldsymbol{V}} = \frac{1}{J'} \sum_{j=1}^{J'} \hat{\boldsymbol{z}}_j \hat{\boldsymbol{z}}_j^{\mathsf{T}}. \tag{24}$$

To verify this estimator, in the simulations we compared (24) with the sample covariance of $\boldsymbol{Y}$ (Supplementary Fig. 2). Although genetic effects should also ideally be removed before estimating $\boldsymbol{V}$, in the simulations the genetic effects were all very small, and so should have little impact on this estimate.

The estimate for the UK Biobank blood cell trait data is given in Supplementary Table 3. Since the blood cell traits were standardized, we scaled the resulting estimate to obtain a correlation matrix.

Specifying the prior. The prior (7) can accommodate many different effect sharing patterns. However, to maximize the benefit of using this prior, it should capture the effect sharing patterns that are actually present in the data. Following [89], we considered three approaches to obtaining g (in the simulations, we assessed the advantages of each of these approaches; see Supplementary Fig. 1):

- The simple "random effects prior" that assumes the effects are independent across traits. This is a special case of (7) in which the mixture consists of a single mixture component ($K = 1, \omega_1 = 1$) with covariance matrix $\boldsymbol{U}_1 = \boldsymbol{I}_R$. Although simple, this prior is used—implicitly or explicitly—by methods that assume the effects are independent across traits.

- Prior with a mixture of "canonical" sharing patterns. (See "Canonical prior" below for details.) This prior is not as flexible as the "data-driven" prior described next, but has the advantage of being easy to implement because it does not involve any separate estimation steps.

- A "data-driven" prior in which the covariances and weights are estimated from the data. The basic idea behind this prior is to adapt the sharing patterns $\boldsymbol{U}_k$ and corresponding mixture weights ω_k to be consistent with the data. (See "Data-driven prior" below for details.) This requires more work to design but has a potentially greater payoff.

Canonical prior The generation of the canonical covariance matrices for the prior g is implemented by the `create_cov_canonical` function in the `mvsusieR` package. Following [89], this function generates the following matrices: the $R \times R$ identity matrix, $\boldsymbol{I}_R$, modeling the case when all effects are independent; the "equal effects" matrix, an $R \times R$ matrix of all ones, which models the case in which all effects are the same; rank-1 matrices modeling trait-specific effects of the form $\boldsymbol{e}_r \boldsymbol{e}_r^{\mathsf{T}}$, in which $\boldsymbol{e}_r$ is a vector of length R containing all zeros except for a 1 in the rth position; and matrices modeling uniformly heterogeneous effects, with ones on the diagonal and σ on the off-diagonal, where σ is 0.25, 0.5 or 0.75. In total, this resulted in $R + 5$ covariance matrices, where R is the number of traits.

Note that the canonical covariance matrices are all at the same scale (each matrix has entries spanning the range 0 to 1), and none of the matrices allow for negatively correlated effects (all of the entries in these matrices are non-negative).

To complete the canonical prior, we assigned uniform weights $\omega_k = 1/K$ to the $K = R + 5$ mixture components.

Data-driven prior. We also took an approach similar to [89] to generate the covariance matrices $\boldsymbol{U}_k$ and mixture weights ω_k in the "data-driven" prior.

First, we prepared a data set to learn the prior. For each candidate fine-mapping region, we identified the top z-score which was defined as the vector of association z-scores for the R traits containing the largest (in magnitude) z-score among all SNPs in the given fine-mapping region. Letting M denote the number of fine-mapping regions, we formed an $M \times R$ matrix containing the top z-scores. Here we denote this matrix by $\boldsymbol{Z}$.

Next, we generated *initial estimates* of covariance matrices using a variety of approaches:

- $R + 5$ canonical covariance matrices (see "Canonical prior" above).
- The empirical covariance matrix $\boldsymbol{Z}^\intercal \boldsymbol{Z}/M$.
- Three rank-1 matrices of the form $\boldsymbol{v}_r \boldsymbol{v}_r^\intercal$, $r = 1, 2, 3$, in which $\boldsymbol{v}_r$ is the rth right singular vector of the reduced singular value decomposition (SVD) of $\boldsymbol{Z}$, $\boldsymbol{Z} = \sum_{r=1}^{R} \sigma_r \boldsymbol{w}_r \boldsymbol{v}_r{}^\intercal$, in which σ_r is the rth singular value and $\boldsymbol{w}_r$ is the rth left singular vector.
- A rank-3 approximation of $\boldsymbol{Z}$ based its SVD, $\boldsymbol{U} \approx \sum_{r=1}^{3} \sigma_r^2 \boldsymbol{v}_r \boldsymbol{v}_r^\intercal / M$.
- A sparse, low-rank approximation of $\boldsymbol{Z}$ obtained using the R package `flashr` [97] (version 0.6-8), $\boldsymbol{U} \approx \boldsymbol{F}\boldsymbol{L}^\intercal \boldsymbol{L}\boldsymbol{F}^\intercal/M$, where $\boldsymbol{L}$ is the $M \times R'$ loadings matrix and $\boldsymbol{F}$ is the $R \times R'$ matrix of estimated factors, and $R' \leqslant R$ is the rank of the approximation. The rank, R', was automatically determined by `flashr`.
- R' rank-1 matrices of the form $\boldsymbol{f}_r \boldsymbol{f}_r^\intercal$, $r = 1, \ldots, R'$, in which $\boldsymbol{f}_r$ denotes the rth column of $\boldsymbol{F}$.

After completing these steps, we had initial estimates for $K = R + R' + 11$ covariance matrices.

Next, we ran Extreme Deconvolution (ED) [12] to refine the initial estimates of the $\boldsymbol{U}_k$ and simultaneously estimate the mixture weights ω_k. (We used the ED algorithm implemented in the `cov_ed` function from the `mashr` R package, version 0.2.59, which was adapted from [12].) The mixture weights initialized to uniform values, $\omega_k = 1/K$, $k = 1, \ldots, K$.

Finally, to avoid poor estimation of the *lfsr* that can happen when the prior covariances are singular (*i.e.*, not invertible), we added a small positive constant to the diagonals of all the covariance matrices $\boldsymbol{U}_k$ in the data-driven prior. This step ensured that these matrices were all invertible.

The data-driven prior obtained by running the above procedure on the UK Biobank data is shown in Supplementary Fig. 7. The data-driven priors obtained by running the above procedure separately in Scenarios a and b of the simulations are shown in Supplementary Figures 8 and 9, respectively.

UK Biobank data

The UK Biobank is a prospective cohort study with detailed phenotype and genotype data collected from approximately 500,000 participants recruited in the United Kingdom, with ages between 40 and 69 at time of recruitment [15, 84]. For fine-mapping, we focussed on a subset of 16 blood cell traits from the UK Biobank haematology data collection [80]. These blood cell traits were also the focus of a recent association analysis [6, 48] and fine-mapping studies [88, 92]. Several of the UK Biobank blood cell traits are based on the same measured quantities and are therefore highly correlated so we did not include all the blood cell traits in our analyses. For example, relative volume of erythrocytes, also known as "hematocrit" (HCT), is calculated from mean corpuscular volume (MCV) and red blood cell count (RBC#), so to avoid including highly correlated traits we did not include HCT. The blood cell trait used in our fine-mapping analyses are summarized in Supplementary Table 1.

The UK Biobank imputed genotypes feature a high density of available SNPs, so they are well-suited for fine-mapping [15, 84]. We used a subset of the 502,492 available UK Biobank genotypes (version 3), removing samples that met one or more of the following criteria for exclusion: mismatch between self-reported and genetic sex; pregnant; one or more data entries needed for the analysis or data preparation steps are missing; and, following [6, 92], a blood-related disease was reported in the hospital in-patient data (blood-related diseases included were eukemia, lymphoma, bone marrow transplant, chemotherapy, myelodysplastic syndrome, anemia, HIV, end-stage kidney disease, dialysis, cirrhosis,

multiple myeloma, lymphocytic leukemia, myeloid leukemia, polycythaemia vera, haemochromatosis). Additionally, we excluded outlying genotype samples based on heterozygosity and/or rate of missing genotypes as defined by UK Biobank (data field 22027), and we removed any individuals having at least one relative in the cohort based on UK Biobank kinship calculations (samples with a value other than zero in data field 22021). Finally, to limit confounding due to population structure, we included only genotype samples marked as "White British" (based on a principal components analysis of the genotypes [15] stored in data field 22009). After filtering genotype samples according to these criteria, 257,605 samples remained.

We applied quantile normalization to the 16 blood cell traits measured in the 257,605 samples, separately for each trait, to transform each trait to the standard normal distribution. Since ultimately we aimed to jointly model the 16 blood cell traits, we removed outlying phenotypes according to a simple multivariate normal distribution of the phenotypes. Specifically, after quantile normalization, we measured the Mahalanobis distance $\boldsymbol{y}_i^\intercal \hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{y}_i$ for each individual i, where $\boldsymbol{y}_i$ is the vector of 16 blood cell traits measured in individual i, and $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix estimated from the 257,605 UK Biobank samples. We discarded samples with Mahalanobis distance falling within the $[0.99, 1]$ quantile of the chi-square distribution with 16 degrees of freedom. This step removed 8,625 samples, for a final total of 248,980 UK Biobank samples.

Base-pair positions of the SNPs in the UK Biobank genotype data are reported using Genome Reference Consortium human genome assembly 37 (hg19).

Association analyses of UK Biobank blood cell traits

Using the UK Biobank genotype and phenotype data prepared as described above, we computed association statistics for each of the 16 blood cell traits and for all available biallelic SNPs on autosomal chromosomes meeting the following criteria: minor allele frequency of 0.1% or greater; information ("INFO") score of 0.6 or greater (the INFO score quantifies imputation quality). The same criteria were used in [98] to filter the SNPs.

Association statistics were computed using the `--glm` function in PLINK (version 2.00a2LM, 64-bit Intel, Feb 21, 2009) [19] with `hide-covar no-x-sex omit-ref --vif 100`. Following [18, 92], we included the following covariates in the association analyses: sex (data field 31), age at recruitment (21022), age × age, assessment centre (54), and genotype measurement batch (22000). To limit inflation of spurious associations due to population structure, we also included the top 10 genotype PCs as covariates following previous association analyses of UK Biobank data [e.g., 61]. (These PCs were previously computed by UK Biobank [15] and stored in data field 22009.) The covariates input file for PLINK was prepared by calling the `model.matrix` function in R and standardizing quantitative covariates (age, PCs) to have mean 0 and variance 1.

The summary data provided as input to SuSiE and mvSuSiE were the z-scores and p-values extracted from the `T_STAT` and `P` columns in the `plink2 --glm` outputs. The association statistics computed using PLINK have been made available in a Zenodo repository [**?**]. *Note: Zenodo or Data Dryad repository will be created for this and URL will be provided.*

Selection of regions for fine-mapping

To select regions for fine-mapping, we adapted the approach used in [92] to the multivariate setting. In brief, we began by identifying regions separately for each trait. For each significant association (PLINK p-value less than 5×10^{-8}), we defined the fine-mapping region as all SNPs within ± 250 kb of the significant association. Next, any regions overlapping by one or more SNPs were combined into a larger region. We repeated combining regions until no regions overlapped. This resulted in a set of fine-mapping regions for each of the 16 blood cell traits, similar to [92]. To form a single set of fine-mapping regions for all 16 traits, we then merged two regions from different traits whenever they overlapped. The end result of this procedure was a set 975 of disjoint fine-mapping regions satisfying these properties: all "significant SNPs" (with PLINK p-value less than 5×10^{-8}) belong to exactly one region; all SNPs within 250 kb of a significant SNP belong to exactly one region. This procedure generated fine-mapping regions that varied considerably in size: their lengths ranged from 411 kb to

8.73 Mb (average size: 961 kb; median size: 686 kb); and the number of SNPs ranged from 93 SNPs to 36,605 SNPs (average number of SNPs: 4,776; median number of SNPs: 3,514). A listing of all 975 regions is given in Supplementary Table 2. These same regions were used in both single-trait and multi-trait fine-mapping.

Note that we did not fine-map the extended MHC [44], defined as base-pair positions 25–36 Mb on chromosome 6. The MHC is particularly challenging to analyze and interpret, and therefore is typically analyzed separately [17, 60, 73].

Simulations using UK Biobank genotypes

We evaluated the fine-mapping methods on data sets generated using real genotypes $\boldsymbol{X}$ and simulated phenotypes $\boldsymbol{Y}$. For the genotypes, we used the UK Biobank imputed genotypes. We simulated $\boldsymbol{Y}$ from different mvSuSiE models.

The genotype data were curated following the data preparation steps described above, so $N =$ 248,980 in all our simulations. To clarify, these data preparation steps included removing outlying blood cell trait observations (see above). Even though this particular filtering step was not needed since we did not use the UK Biobank phenotype data in the simulations, for convenience we used the data prepared with this filtering step.

Simulation scenarios. We implemented three fine-mapping scenarios in the simulations.

The first scenario was by far the simplest: 2 traits in which the traits were independent and the effects of each causal SNP on the two traits were also independent. This simple scenario was intended mainly for comparisons with PAINTOR and flashfm so as to not unfairly disadvantage these methods: flashfm cannot handle a large number of traits; PAINTOR cannot handle a large number of causal SNPs, and assumes independent traits and independent effects (Table 1).

For comparing other fine-mapping methods (SuSiE, mvSuSiE, CAFEH), we simulated data sets under two more complex scenarios, which we refer to as "Scenario a" and "Scenario b".

In Scenario a, we simulated 20 independent traits in which the SNP effects were either specific to one trait or shared among traits in simple ways (equal effects among 2 traits, equal effects among half of the traits, or correlated equally among all 20 traits). In the results, we call Scenario a the "Trait-specific + Shared Effects" scenario.

Scenario b was intended to capture a combination of factors that one might more realistically encounter in fine-mapping studies. It is also more challenging because the traits are correlated and the effects are shared among the traits in complex ways. Specifically, we simulated using a residual covariance matrix $\boldsymbol{V}$ and sharing patterns $\boldsymbol{U}_k$ obtained from our analyses of the UK Biobank blood cell traits. In the results, we refer to Scenario b as the "Complex Shared Effects" scenario.

Simulation procedure. Let $\boldsymbol{X}$ denote the $N \times J$ genotype matrix for a given fine-mapping region where J is the number of SNPs in the region, and $N = 248{,}980$.

The procedure we used to simulate an $N \times R$ matrix $\boldsymbol{Y}$ was the following.

1. Center and scale the columns of $\boldsymbol{X}$ so that each column has a mean of 0 and a variance of 1.
2. Choose S, the number of causal SNPs. For Scenarios a and b, set S to 1, 2, 3, 4 or 5 with probabilities $0.3, 0.3, 0.2, 0.1, 0.1$, respectively. For the 2-trait simulations, set $S = 2$.
3. Sample the indices of the S causal SNPs uniformly at random from $\{1, \ldots, J\}$. Denote the set of causal SNPs by $\mathcal{C}$.
4. For each SNP $j \in \mathcal{C}$, simulate the R effects, $\boldsymbol{b}_j \in \mathbb{R}^R$, from the mixture of multivariate normals (7), in which $\sigma_{0l}^2 = 1$. In the 2-trait simulations, we set $K = 1, \omega_1 = 1, \boldsymbol{U}_1 = \boldsymbol{I}_2$, which is simply the multivariate normal with zero mean and identity covariance. In Scenario a, we simulated the effects of the causal SNPs $\boldsymbol{b}_j$ using a mixture of 19 covariance matrices (Supplementary Fig. 6). For Scenario b, we simulated the effects using the mixture of 15 covariance matrices (Supplementary Fig. 7).

5. For each SNP $j \notin \mathcal{C}$, set $\boldsymbol{b}_j = \mathbf{0}$.

6. Choose the residual variance σ^2. To set σ^2 to a realistic value, we set σ^2 so that the greatest proportion of variance in a trait explained by the SNPs was 0.05%, which roughly corresponds to the proportion of variance explained in the mvSuSiE fine-mapping analyses of the UK Biobank blood cell traits. In particular, we solved for σ^2 satisfying $\frac{\sigma_g^2}{\sigma^2+\sigma_g^2} = 0.0005$, where $\sigma_g^2 := \mathrm{Var}(\hat{\boldsymbol{y}}_r)$ is the variance in the rth trait explained by the SNPs, in which $\mathrm{Var}(\boldsymbol{\theta})$ denotes the sample variance, $\hat{\boldsymbol{y}}_r := \boldsymbol{x}_1 b_{1r} + \cdots + \boldsymbol{x}_J b_{Jr}$, and $r := \mathrm{argmax}_{r' \in \{1,\ldots,R\}} \mathrm{Var}(\hat{\boldsymbol{y}}_{r'})$.

7. Specify the $R \times R$ residual correlation matrix $\boldsymbol{C}$, then set $\boldsymbol{V} = \sigma^2 \boldsymbol{C}$. For Scenario a and the 2-trait scenario, $\boldsymbol{C} = \boldsymbol{I}_R$. For Scenario b, the 16×16 covariance matrix $\boldsymbol{C}$ was set to the correlation matrix estimated from the 16 blood cell traits after removing the linear effects of covariates (Supplementary Table 3).†

8. Simulate $\boldsymbol{Y}$ using model (1).

9. Center and scale the columns of $\boldsymbol{Y}$ so that each column has a mean of zero and a variance of 1.

10. Compute the summary statistics—effect estimates $\hat{\beta}_{jr}$, standard errors $\hat{s}_{jr}$, z-scores z_{jr} and the in-sample LD matrix $\boldsymbol{R}$—using PLINK [19] and LDstore [7]. For these summary statistics, we extracted the `BETA`, `SE`, `T_STAT` and `P` columns from the `plink2 --glm` output (see "Association analyses of UK Biobank blood cell traits" for more details on how PLINK was called). Note PLINK was applied to the raw genotypes without centering or scaling. We computed the $J \times J$ in-sample LD matrix $\hat{\boldsymbol{R}} = \boldsymbol{D}^{-1/2}\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{D}^{-1/2}$, where $\boldsymbol{D} := \mathrm{diag}(\boldsymbol{X}^\intercal\boldsymbol{X})$, using LDstore version 1.1.

This procedure produced an empirical distribution of z-scores roughly similar to the z-scores seen in association analyses of the blood cell traits; in our simulations, the largest z-score magnitude in each fine-mapping region had a median of 11.05, mean of 10.97 and a third quantile of 11.71, whereas the corresponding statistics for the UK Biobank blood cell traits were 8.01, 10.85 and 12.18.

For each of the three scenarios (Scenario a, Scenario b, and the 2-trait scenario), we simulated 600 data sets for 600 fine-mapping regions selected from the curated set of 975 regions for the UK Biobank blood cell traits (Supplementary Table 2). All selected regions had at least 1,000 SNPs and no more than 5,000 SNPs, and were at least 400 Kb in size and at most 1.6 Mb. In total, we simulated $600 \times 3 = 1{,}800$ fine-mapping data sets.

Details of the methods compared

In this section we describe how we ran the methods on the simulated data sets.

SuSiE, mvSuSiE and PAINTOR were run using the z-scores and in-sample LD. CAFEH and flashfm were run using the effect estimates, standard errors of these effect estimates, and in-sample LD. Some methods, including mvSuSiE, also accepted an additional input, the sample size (N), in which case we provided this as well. flashfm also required the "reference" allele frequencies, which in all our analyses were the minor allele frequencies.

PAINTOR. We ran PAINTOR [50] in only the 2-trait simulations. PAINTOR was designed to work with functional genomic annotation data, so to run PAINTOR we created a single "dummy" annotation in which all SNPs were assigned to this annotation (that is, all entries of the annotation matrix were set to 1). For all data sets, we asked PAINTOR to enumerate all possible configurations up to 2 causal SNPs. (In the 2-trait simulations, the true number of causal SNPs was always 2.) We did not use the "mcmc" option (`-mcmc`) because the outputted PIPs when using this option were all zero in our tests. (The same issue was reported in `https://github.com/gkichaev/PAINTOR_V3.0/issues/5`.)

†Note that, although this correlation matrix was estimated from the UK Biobank data, for the simulations, the $\boldsymbol{V}$ used in the mvSuSiE analyses of the simulated data was estimated using the simulated data and so the $\boldsymbol{V}$ used by mvSuSiE differed from the $\boldsymbol{V}$ used to simulate $\boldsymbol{Y}$.

All other PAINTOR options were kept at their default settings. Note that PAINTOR does not accept N (the sample size) as input. Also note that PAINTOR assumes that both traits and effects are independent across traits (Table 1).

flashfm. We ran flashfm [39] in only the 2-trait simulations. We ran flashfm by calling function `FLASHFMwithFINEMAP` from R package `flashfm` (version 0.0.0.9000). This function internally calls FINEMAP [8, 9] (we used FINEMAP 1.4.1) with settings `--sss --n-configs-top 1000 --n-causal-snps 10`, which allows configurations of up to 10 causal SNPs. We ran flashfm with 4 CPUs (`NCORES = 4`). All other flashfm settings were kept at their defaults. The inputs to `FLASHFMwithFINEMAP` were the effect estimates, the standard errors of these effect estimates, minor allele frequencies, vector of trait means, and sample size N. Since $\boldsymbol{Y}$ was centered and standardized in the simulations, the vector of trait means was simply a vector of zeros of length R.

As a side note, we learned that flashfm implicitly assumes that $\boldsymbol{X}$ is not standardized since flashfm uses the MAFs to obtain a covariance matrix from the correlation (LD) matrix $\boldsymbol{R}$. Therefore, when $\boldsymbol{X}$ is standardized, it may be preferrable to replace the MAFs with a constant value for all SNPs.

CAFEH. We ran CAFEH [3] on all data sets in Scenarios a and b. Specifically, we used the `fit_cafeh_summary` interface in CAFEH 1.0 installed with Python 3.7.4. The `fit_cafeh_summary` function accepts the following data inputs: effect estimates, standard errors of those estimates, LD matrix, and sample size N. When calling `fit_cafeh_summary`, all optional arguments were kept at the software defaults. CAFEH's default setting for the upper limit on the number of single effects (denoted as "K" in the CAFEH model), is 10, which is the same default used in SuSiE and mvSuSiE. Note that CAFEH assumes that both traits and effects are independent across traits (Table 1).

For assessing performance of CAFEH PIPs and trait-specific PIPs (in CAFEH, these are called "study PIPs"), we called `get_pip` and `get_study_pip`.

CAFEH outputs credible sets without any filter on the purity of the CSs. Therefore, to make the CAFEH credible sets comparable to SuSiE and mvSuSiE credible sets, we filtered out CSs with purity less than 0.5.

Note that the two CAFEH summary data interfaces—`fit_cafeh_summary` and `fit_cafeh_z`—produce the same or very similar results when $\boldsymbol{X}$ is standardized, so we expect that both CAFEH summary-data interfaces would perform similarly when $\boldsymbol{X}$ is standardized. Both functions internally call function `CAFEHSummary` with the same LD matrix, but provide different effect estimates and standard errors of the effect estimates. Let $\hat{\boldsymbol{\beta}}$ denote the vector of effect estimates (with one entry per SNP) and let $\hat{\boldsymbol{s}}$ denote the vector of standard errors (also with one entry per SNP). If `fit_cafeh_summary` calls `CAFEHSummary` with inputs $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{s}}$, and assuming $\boldsymbol{X}$ is standardized, then it can be shown that `fit_cafeh_z` calls `CAFEHSummary` with inputs $\sqrt{N}\hat{\boldsymbol{\beta}}, \sqrt{N}\hat{\boldsymbol{s}}$. Since `CAFEHSummary` is invariant to rescaling of $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{s}}$—that is, `CAFEHSummary` generates the same result with inputs $a\hat{\boldsymbol{\beta}}$ and $a\hat{\boldsymbol{s}}$ for any choice of scalar $a > 0$—it follows that `fit_cafeh_summary` and `fit_cafeh_z` also produce the same result when $\boldsymbol{X}$ is standardized. In practice, this invariance does not hold exactly since it requires that the prior on the effects also be appropriately rescaled, but empirically we have found that the CAFEH PIPs and posterior effect estimates are almost the same for different choices of $a > 0$.‡

SuSiE. We ran SuSiE in all simulations by calling function `susie_rss` from the `susieR` R package [96] (version 0.12.12). In each data set, we ran `susie_rss` once per trait. The `susie_rss` interface accepts different types of summary data; we provided z-scores, in-sample LD, and sample size N. For all simulations, we set L, the maximum number of non-zero effects, to 10. We also set $L = 10$ for the 2-trait simulations even though there were never more than 2 causal SNPs in these simulations. We estimated the residual variance (`estimate_residual_variance = TRUE`), which is the recommended

‡See `https://github.com/karltayeb/cafeh/blob/current_working_branch/notebooks/CAFEHS_scale_invariance.ipynb`.

setting when the LD is estimated from the "in-sample" data. We set the maximum number of IBSS iterations to 1,000 (`max_iter = 1000`). The remaining optional arguments were kept at their defaults.

Since SuSiE analyzes each trait separately, it does not directly provide a systematic way to quantify evidence for a SNP being a cross-trait causal SNP. To quantify performance in this task and compare with mvSuSiE, we quantified the evidence for a cross-trait causal SNP using an *ad hoc* metric, the "maximum PIP", defined as

$$\text{max-PIP}_j := \max_{r \in \{1,\ldots,R\}} \text{PIP}_{jr}, \tag{25}$$

where PIP_{jr} is the (trait-specific) PIP for SNP j obtained from the SuSiE analysis of trait r.

mvSuSiE. We ran mvSuSiE using the `mvsusie_rss` interface from the `mvsusieR` R package (version 0.0.3.0518, git commit id 9f28916). While `susie_rss` accepts a *vector* of z-scores, `mvsusie_rss` accepts a *matrix* of z-scores (specifically, a $J \times R$ matrix). In the simulations, we compared several mvSuSiE variants using different prior choices; for more details, see "Specifying the prior" above and Supplementary Fig. 1. (In the 2-trait simulations, we only used the canonical prior. This was a mixture of multivariate normals with $K = 7$ components.) We also compared mvSuSiE with different settings of the residual covariance $\boldsymbol{V}$; see "Estimating the residual variance matrix" above, and Supplementary Figures 2 and 5. In all cases, we ran `mvsusie_rss` with the following settings: `L = 10`, `max_iter = 1000`, `estimate_prior_variance = TRUE`, `estimate_prior_method = "EM"`, `precompute_covariances = TRUE` and `n_thread = 4`. We set $L = 10$ for the 2-trait simulations even though there were never more than 2 causal SNPs in these simulations. All other options were kept at the default settings.

Computing environment

All analyses of the simulated data sets were run on Linux machines (Scientific Linux 7.4) with 4 Intel Xeon E5-2680v4 ("Broadwell") processors, and with R 4.1.0 [72] linked to the OpenBLAS 0.3.13 optimized numerical libraries. At most 10 GB of memory was needed to perform a fine-mapping analysis of a single simulated data set using one of the methods. We used the Dynamic Statistical Comparisons system (`https://github.com/stephenslab/dsc`) to perform the simulations.

Fine-mapping of UK Biobank blood cell traits using SuSiE and mvSuSiE

We fit a mvSuSiE model to each fine-mapping data set—specifically the prepared z-scores matrix $\boldsymbol{Z}$ and LD matrix $\boldsymbol{R}$—by calling `mvsusie_rss` from the `mvsusieR` package with the following settings: `L = 10`, `N = 248980`, `precompute_covariances = TRUE`, `estimate_prior_variance = TRUE`, `estimate_prior_method = "EM"`, `max_iter = 1000` and `n_thread = 1`. We ran SuSiE on each data set $\{\boldsymbol{Z}, \boldsymbol{R}\}$ separately for each trait (*i.e.*, column of $\boldsymbol{Z}$) by calling `susie_rss` with the following settings: `n = 248980`, `L = 10`, `max_iter = 1000`, `estimate_prior_variance = TRUE`, `refine = TRUE`. Any CSs returned by `susie_rss` or `mvsusie_rss` with purity less than 0.5 were removed.

Data availability

The genotype and phenotype data used in our analyses are available from UK Biobank. Association test statistics for the UK Biobank blood cell traits, results of the simulations, results of the blood cell trait fine-mapping analyses, and other results needed to reproduce the figures in the paper are also available online (see URLs).

Code availability

Source code implementing our methods, including SuSiE and mvSuSiE, including the code implementing the simulations and analysis of the UK Biobank blood cell traits, is available online and is distributed under the BSD and MIT open source licenses (see URLs).

References

[1] Alper, S. (2017). Genetic diseases of PIEZO1 and PIEZO2 dysfunction. In P. A. Gottlieb (Ed.), *Piezo Channels*, Volume 79 of *Current Topics in Membranes*, pp. 97–134. Academic Press.

[2] Altshuler, D., M. J. Daly, and E. S. Lander (2008). Genetic mapping in human disease. *Science 322*(5903), 881–888.

[3] Arvanitis, M., K. Tayeb, B. J. Strober, and A. Battle (2022). Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. *American Journal of Human Genetics 109*(2), 223–239.

[4] Asano, K., S. Ueki, M. Tamari, Y. Imoto, S. Fujieda, and M. Taniguchi (2020). Adult-onset eosinophilic airway diseases. *Allergy 75*(12), 3087–3099.

[5] Asimit, J. L., D. B. Rainbow, M. D. Fortune, N. F. Grinberg, L. S. Wicker, and C. Wallace (2019). Stochastic search and joint fine-mapping increases accuracy and identifies previously unreported associations in immune-mediated diseases. *Nature Communications 10*, 3216.

[6] Astle, W. J., H. Elding, T. Jiang, D. Allen, D. Ruklisa, et al. (2016). The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell 167*(5), 1415–1429.

[7] Benner, C., A. S. Havulinna, M.-R. Järvelin, V. Salomaa, S. Ripatti, and M. Pirinen (2017). Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *American Journal of Human Genetics 101*(4), 539–551.

[8] Benner, C., A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen (2018). Refining fine-mapping: effect sizes and regional heritability. *bioRxiv*, doi:10.1101/318618.

[9] Benner, C., C. C. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics 32*(10), 1493–1501.

[10] Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: a review for statisticians. *Journal of the American Statistical Association 112*(518), 859–877.

[11] Bottolo, L., M. Banterle, S. Richardson, M. Ala-Korpela, M.-R. Järvelin, and A. Lewin (2021). A computationally efficient Bayesian seemingly unrelated regressions model for high-dimensional quantitative trait loci discovery. *Journal of the Royal Statistical Society, Series C 70*(4), 886–908.

[12] Bovy, J., D. W. Hogg, and S. T. Roweis (2011). Extreme Deconvolution: inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Annals of Applied Statistics 5*(2B), 1657–1677.

[13] Brady, G., H. J. Whiteman, L. C. Spender, and P. J. Farrell (2009). Downregulation of RUNX1 by RUNX3 requires the RUNX3 VWRPY sequence and is essential for Epstein-Barr virus-driven B-cell proliferation. *Journal of Virology 83*(13), 6909–6916.

[14] Buniello, A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, et al. (2018). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research 47*(D1), D1005–D1012.

[15] Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, and J. Marchini (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature 562*(7726), 203–209.

[16] Cahalan, S. M., V. Lukacs, S. S. Ranade, S. Chien, M. Bandell, and A. Patapoutian (2015). Piezo1 links mechanical forces to red blood cell volume. *eLife 4*, e07370.

[17] Caliskan, M., C. D. Brown, and J. C. Maranville (2021). A catalog of GWAS fine-mapping efforts in autoimmune disease. *American Journal of Human Genetics 108*(4), 549–563.

[18] Canela-Xandri, O., K. Rawlik, and A. Tenesa (2018). An atlas of genetic associations in UK Biobank. *Nature Genetics 50*(11), 1593–1599.

[19] Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience 4*(1), s13742–015–0047–8.

[20] Chen, W., B. R. Larrabee, I. G. Ovsyannikova, R. B. Kennedy, I. H. Haralambieva, G. A. Poland, and D. J. Schaid (2015). Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics 200*(3), 719–736.

[21] Chipman, H., E. I. George, and R. E. McCulloch (2001). The practical implementation of Bayesian model selection. In P. Lahiri (Ed.), *Model Selection*, Volume 38 of *IMS Lecture Notes*, pp. 65–116. Beachwood, OH: Institute of Mathematical Statistics.

[22] Chun, S., A. Casparino, N. A. Patsopoulos, D. C. Croteau-Chonka, B. A. Raby, P. L. De Jager, S. R. Sunyaev, and C. Cotsapas (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nature Genetics 49*(4), 600–605.

[23] Clyde, M. A., J. Ghosh, and M. L. Littman (2011). Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics 20*(1), 80–101.

[24] Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory* (2nd ed.). Wiley-Interscience.

[25] Crispino, J. D. and M. J. Weiss (2014). Erythro-megakaryocytic transcription factors associated with hereditary anemia. *Blood 123*(20), 3080–3088.

[26] Dawid, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika 68*(1), 265–274.

[27] Dellaportas, P., J. J. Forster, and I. Ntzoufras (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing 12*, 27–36.

[28] Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*(1), 1–22.

[29] Foley, C. N., J. R. Staley, P. G. Breen, B. B. Sun, P. D. W. Kirk, S. Burgess, and J. M. M. Howson (2021). A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nature Communications 12*, 764.

[30] Fortune, M. D., H. Guo, O. Burren, E. Schofield, N. M. Walker, M. Ban, S. J. Sawcer, J. Bowes, J. Worthington, A. Barton, S. Eyre, J. A. Todd, and C. Wallace (2015). Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nature Genetics 47*(7), 839–846.

[31] George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

[32] George, E. I. and R. E. McCulloch (1997). Approaches for Bayesian variable selection. *Statistica Sinica 7*(2), 339–373.

[33] Giambartolomei, C., D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, and V. Plagnol (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genetics 10*(5), 1–15.

[34] Giambartolomei, C., J. Zhenli Liu, W. Zhang, M. Hauberg, H. Shi, J. Boocock, J. Pickrell, A. E. Jaffe, T. C. Consortium, B. Pasaniuc, and P. Roussos (2018). A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics 34*(15), 2538–2545.

[35] Guan, Y. and M. Stephens (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Annals of Applied Statistics 5*(3), 1780–1815.

[36] Gupta, A. and D. Nagar (2000). *Matrix variate distributions*. Boca Raton, FL: Chapman & Hall.

[37] Han, B. and E. Eskin (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American Journal of Human Genetics 88*(5), 586–598.

[38] Helms, C., L. Cao, J. G. Krueger, E. M. Wijsman, F. Chamian, D. Gordon, M. Heffernan, J. A. W. Daw, J. Robarge, J. Ott, P.-Y. Kwok, A. Menter, and A. M. Bowcock (2003). A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. *Nature Genetics 35*(4), 349–356.

[39] Hernández, N., J. Soenksen, P. Newcombe, M. Sandhu, I. Barroso, C. Wallace, and J. Asimit (2021). The flashfm approach for fine-mapping multiple quantitative traits. *Nature Communications 12*, 6147.

[40] Heskes, T., O. Zoeter, and W. Wiegerinck (2004). Approximate expectation maximization. In S. Thrun, L. K. Saul, and B. Schölkopf (Eds.), *Advances in Neural Information Processing Systems 16*, pp. 353–360. MIT Press.

[41] Hoggart, C. J., J. C. Whittaker, M. De Iorio, and D. J. Balding (2008). Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genetics 4*(7), e1000130.

[42] Hormozdiari, F., E. Kostem, E. Y. Kang, B. Pasaniuc, and E. Eskin (2014). Identifying causal variants at loci with multiple signals of association. *Genetics 198*(2), 497–508.

[43] Hormozdiari, F., M. Van De Bunt, A. V. Segre, X. Li, J. W. J. Joo, M. Bilow, J. H. Sul, S. Sankararaman, B. Pasaniuc, and E. Eskin (2016). Colocalization of GWAS and eQTL signals detects target genes. *American Journal of Human Genetics 99*(6), 1245–1260.

[44] Horton, R., L. Wilming, V. Rand, R. C. Lovering, E. A. Bruford, V. K. Khodiyar, M. J. Lush, S. Povey, C. C. Talbot, M. W. Wright, H. M. Wain, J. Trowsdale, A. Ziegler, and S. Beck (2004). Gene map of the extended human MHC. *Nature Reviews Genetics 5*(12), 889–899.

[45] Hukku, A., M. Pividori, F. Luca, R. Pique-Regi, H. K. Im, and X. Wen (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *American Journal of Human Genetics 108*(1), 25–35.

[46] Hutchinson, A., J. Asimit, and C. Wallace (2020). Fine-mapping genetic associations. *Human Molecular Genetics 29*(R1), R81–R88.

[47] Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1999). An introduction to variational methods for graphical models. *Machine Learning 37*(2), 183–233.

[48] Kachuri, L., S. Jeon, A. T. DeWan, C. Metayer, X. Ma, J. S. Witte, C. W. Chiang, J. L. Wiemels, and A. J. de Smith (2021). Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *American Journal of Human Genetics 108*(10), 1823–1835.

[49] Kanai, M., R. Elzur, W. Zhou, M. Kanai, K.-H. H. Wu, et al. (2022). Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genomics 2*(12), 100210.

[50] Kichaev, G., M. Roytman, R. Johnson, E. Eskin, S. Lindstroem, P. Kraft, and B. Pasaniuc (2017). Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics 33*(2), 248–255.

[51] Kichaev, G., W.-Y. Yang, S. Lindstrom, F. Hormozdiari, E. Eskin, A. L. Price, P. Kraft, and B. Pasaniuc (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genetics 10*(10), e1004722.

[52] Kote-Jarai, Z., E. J. Saunders, D. A. Leongamornlert, M. Tymrakiewicz, T. Dadaev, et al. (2013). Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with tert expression. *Human Molecular Genetics 22*(12), 2520–2528.

[53] LaPierre, N., K. Taraszka, H. Huang, R. He, F. Hormozdiari, and E. Eskin (2021). Identifying causal variants by fine mapping across multiple studies. *PLoS genetics 17*(9), e1009733.

[54] Lee, Y., L. Francesca, R. Pique-Regi, and X. Wen (2018). Bayesian multi-SNP genetic association analysis: control of FDR and use of summary statistics. *bioRxiv*, doi:10.1101/316471.

[55] Lewin, A., H. Saadi, J. E. Peters, A. Moreno-Moral, J. C. Lee, K. G. Smith, E. Petretto, L. Bottolo, and S. Richardson (2016). MT-HESS: an efficient Bayesian approach for simultaneous association detection in OMICS datasets, with application to eQTL mapping in multiple tissues. *Bioinformatics 32*(4), 523–532.

[56] Li, Y. and M. Kellis (2016). Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic Acids Research 44*(18), e144.

[57] Ling, T. and J. D. Crispino (2020). GATA1 mutations in red cell disorders. *IUBMB Life 72*, 106–118.

[58] Ma, S., A. E. Dubin, Y. Zhang, S. A. R. Mousavi, Y. Wang, A. M. Coombs, M. Loud, I. Andolfo, and A. Patapoutian (2021). A role of PIEZO1 in iron metabolism in mice and humans. *Cell 184*(4), 969–982.e13.

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

[60] Matzaraki, V., V. Kumar, C. Wijmenga, and A. Zhernakova (2017). The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biology 18*, 76.

[61] Mbatchou, J., L. Barnard, J. Backman, A. Marcketta, J. A. Kosmicki, A. Ziyatdinov, C. Benner, C. O'Dushlaine, M. Barber, B. Boutkov, L. Habegger, M. Ferreira, A. Baras, J. Reid, G. Abecasis, E. Maxwell, and J. Marchini (2021). Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics 53*(7), 1097–1103.

[62] McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics 9*(5), 356–369.

[63] Mitchell, T. J. and J. J. Beauchamp (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association 83*(404), 1023–1032.

[64] Moser, G., S. H. Lee, B. J. Hayes, M. E. Goddard, N. R. Wray, and P. M. Visscher (2015). Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics 11*(4), e1004969.

[65] Neal, R. M. (1996). *Bayesian learning for neural networks*, Volume 118 of *Lecture Notes in Statistics*. New York, NY: Springer.

[66] Neal, R. M. and G. E. Hinton (1998). A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan (Ed.), *Learning in Graphical Models*, pp. 355–368. New York, NY: Springer.

[67] Newcombe, P. J., D. V. Conti, and S. Richardson (2016). JAM: a scalable Bayesian framework for joint analysis of marginal SNP effects. *Genetic Epidemiology 40*(3), 188–201.

[68] Nichols, K. E., J. D. Crispino, M. Poncz, J. G. White, S. H. Orkin, J. M. Maris, and M. J. Weiss (2000). Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nature Genetics 24*(3), 266–270.

[69] Page, G. P., T. Kanias, Y. J. Guo, M. C. Lanteri, X. Zhang, A. E. Mast, R. G. Cable, B. R. Spencer, J. E. Kiss, F. Fang, S. M. Endres-Dighe, D. Brambilla, M. Nouraie, V. R. Gordeuk, S. Kleinman, M. P. Busch, M. T. Gladwin, and the National Heart AND Lung, and Blood Institute (NHLBI) Recipient Epidemiology Donor Evaluation Study–III (REDS-III) program (2021). Multiple-ancestry genome-wide association study identifies 27 loci associated with measures of hemolysis following blood storage. *Journal of Clinical Investigation 131*(13), e146077.

[70] Pasaniuc, B. and A. L. Price (2017). Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics 18*(2), 117–127.

[71] Pearson, T. A. and T. A. Manolio (2008). How to interpret a genome-wide association study. *Journal of the American Medical Association 299*(11), 1335–1344.

[72] R Foundation for Statistical Computing (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

[73] Raychaudhuri, S., C. Sandor, E. a. Stahl, J. Freudenberg, H.-S. Lee, X. Jia, L. Alfredsson, L. Padyukov, L. Klareskog, J. Worthington, K. a. Siminovitch, S.-C. Bae, R. M. Plenge, P. K. Gregersen, and P. I. W. de Bakker (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genetics 44*(3), 291–296.

[74] Ruffieux, H., A. C. Davison, J. Hager, J. Inshaw, B. P. Fairfax, S. Richardson, and L. Bottolo (2020). A global-local approach for detecting hotspots in multiple-response regression. *Annals of Applied Statistics 14*(2), 905–928.

[75] Schaid, D. J., W. Chen, and N. B. Larson (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics 19*(8), 491–504.

[76] Scheitz, C. J. F. and T. Tumbar (2013). New insights into the role of Runx1 in epithelial stem cell biology and pathology. *Journal of Cellular Biochemistry 114*(5), 985–993.

[77] Schwartzentruber, J., S. Cooper, J. Z. Liu, I. Barrio-Hernandez, E. Bello, N. Kumasaka, A. M. H. Young, R. J. M. Franklin, T. Johnson, K. Estrada, D. J. Gaffney, P. Beltrao, and A. Bassett (2021). Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nature Genetics 53*(3), 392–402.

[78] Servin, B. and M. Stephens (2007). Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genetics 3*(7), e114.

[79] Sesia, M., E. Katsevich, S. Bates, E. Candès, and C. Sabatti (2020). Multi-resolution localization of causal variants across the genome. *Nature Communications 11*, 1093.

[80] Sheard, S. M., R. Nicholls, and J. Froggatt (2017). UK Biobank haematology data companion document. Technical report, UK Biobank `https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/haematology.pdf`.

[81] Stephens, M. (2013). A unified framework for association analysis with multiple related phenotypes. *PLoS ONE 8*(7), e65245.

[82] Stephens, M. (2017). False discovery rates: a new deal. *Biostatistics 18*(2), 275–294.

[83] Stephens, M. and D. J. Balding (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics 10*(10), 681–690.

[84] Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, and R. Collins (2015). UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine 12*(3), e1001779.

[85] Tam, V., N. Patel, M. Turcotte, Y. Bossé, G. Paré, and D. Meyre (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics 20*(8), 467–484.

[86] Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research 1*, 211–244.

[87] Tokuhiro, S., R. Yamada, X. Chang, A. Suzuki, Y. Kochi, T. Sawada, M. Suzuki, M. Nagasaki, M. Ohtsuki, M. Ono, H. Furukawa, M. Nagashima, S. Yoshino, A. Mabuchi, A. Sekine, S. Saito, A. Takahashi, T. Tsunoda, Y. Nakamura, and K. Yamamoto (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics 35*(4), 341–348.

[88] Ulirsch, J. C., C. A. Lareau, E. L. Bao, L. S. Ludwig, M. H. Guo, C. Benner, A. T. Satpathy, V. K. Kartha, R. M. Salem, J. N. Hirschhorn, H. K. Finucane, M. J. Aryee, J. D. Buenrostro, and V. G. Sankaran (2019). Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature Genetics 51*(4), 683–693.

[89] Urbut, S. M., G. Wang, P. Carbonetto, and M. Stephens (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics 51*, 187–195.

[90] van der Harst, P., W. Zhang, I. Mateo Leach, A. Rendon, N. Verweij, et al. (2012). Seventy-five genetic loci influencing the human red blood cell. *Nature 492*(7429), 369–375.

[91] Visscher, P. M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang (2017). 10 years of GWAS discovery: biology, function, and translation. *American Journal of Human Genetics 101*(1), 5–22.

[92] Vuckovic, D., E. L. Bao, P. Akbari, C. A. Lareau, A. Mousas, et al. (2020). The polygenic and monogenic basis of blood traits and diseases. *Cell 182*(5), 1214–1231.e11.

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

[94] Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genetics 17*(9), e1009440.

[95] Wallace, C., A. J. Cutler, N. Pontikos, M. L. Pekalski, O. S. Burren, J. D. Cooper, A. R. García, R. C. Ferreira, H. Guo, N. M. Walker, D. J. Smyth, S. S. Rich, S. Onengut-Gumuscu, S. J. Sawcer, M. Ban, S. Richardson, J. A. Todd, and L. S. Wicker (2015). Dissection of a complex disease susceptibility region using a Bayesian stochastic search approach to fine mapping. *PLoS Genetics 11*(6), e1005272.

[96] Wang, G., A. Sarkar, P. Carbonetto, and M. Stephens (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society, Series B 82*(5), 1273–1300.

[97] Wang, W. and M. Stephens (2021). Empirical Bayes matrix factorization. *Journal of Machine Learning Research 22*(120), 1–40.

[98] Weissbrod, O., F. Hormozdiari, C. Benner, R. Cui, J. Ulirsch, S. Gazal, A. P. Schoech, B. van de Geijn, Y. Reshef, C. Márquez-Luna, L. O'Connor, M. Pirinen, H. K. Finucane, and A. L. Price (2020). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics 52*(12), 1355–1363.

[99] Wen, X., Y. Lee, F. Luca, and R. Pique-Regi (2016). Efficient integrative multi-SNP association analysis via deterministic approximation of posteriors. *American Journal of Human Genetics 98*(6), 1114–1129.

[100] Wen, X., R. Pique-Regi, and F. Luca (2017). Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. *PLoS Genetics 13*(3), 1–25.

[101] Yang, J., T. Ferreira, A. P. Morris, S. E. Medland, P. A. F. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics 44*(4), 369–375.

[102] Zanella, G. and G. Roberts (2019). Scalable importance tempering and Bayesian variable selection. *Journal of the Royal Statistical Society, Series B 81*(3), 489–517.

[103] Zhao, Z., M. Banterle, L. Bottolo, S. Richardson, A. Lewin, and M. Zucknick (2021). BayesSUR: An R package for high-dimensional multivariate Bayesian variable and covariance selection in linear regression. *Journal of Statistical Software 100*(11), 1–32.

[104] Zhao, Z., M. Banterle, A. Lewin, and M. Zucknick (2021). Structured Bayesian variable selection for multiple correlated response variables and high-dimensional predictors. *arXiv 2101.05899*.

[105] Zhou, X., P. Carbonetto, and M. Stephens (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics 9*(2), e1003264.

[106] Zou, Y., P. Carbonetto, G. Wang, and M. Stephens (2022). Fine-mapping from summary data with the "Sum of Single Effects" model. *PLOS Genetics 18*(7), e1010299.

Supplementary Note

More details about Table 1

Here are more details about the columns in Table 1 in which we compared different statistical methods for multivariate fine-mapping:

- Upper limit on number of causal SNPs: "User-specified" means that the method requires the user to specify an upper limit on the number of causal SNPs allowed. "No limit" means that the method does not constrain the number of causal SNPs.

- Data accepted, summary and sufficient. Methods marked "yes" under "sufficient" should reproduce the same result as if the methods were provided with the full data. By contrast (non-sufficient) summary data approximate the summary statistics, and therefore yield results that are not expected to exactly reproduce the results that would be obtained with the full data. (All methods implicitly allow full data because summary statistics can always be computed from full data.) See [106] or the Online Methods ("Extension of posterior computation approach to work with summary data") for definitions of "summary data" and "sufficient data".

- CSs. Does the method compute credible sets (CS)? Note that calculation of CSs is trivial when limiting to at most 1 causal SNP.

- Allows correlated traits. Methods marked as "no" assume measurement error is independent across traits. Assuming independent errors is appropriate if different, non-overlapping samples were used to measure the different traits.

- Models effect sharing. Methods with a "no" in the column (implicitly or explicitly) assume that the effects of a SNP on different traits are independent (conditioned on the SNP having a nonzero effect on one or more traits). A "yes" means that the method can model correlations among the effects on different traits.

- Sample runtimes. Sample runtimes were obtained by running the method on a simulated data set with $J = 5{,}000$ SNPs, $N = 250{,}000$ individuals, and $R = 2$ or $R = 20$ traits. When the method accepted either full data or summary data, the summary-data version of the method was used. Note the sample size, N, should only affect the running time for methods that only accept full data; it should not affect runtime of summary-data methods. When the method limits the number of causal SNPs, the upper limit was always set to 10 or the the largest acceptable value if this was less than 10. For PAINTOR, the upper limit was set to 2 because, in our tests, PAINTOR ran for a very long time when allowing 3 or more causal SNPs. See "Computing environment" for details about the computing environment used to obtain these runtimes.

- Software and version. The name of the software and the version number of the software that was used in our evaluations. For mvSuSiE, the git commit id was given instead of the version number.

Conventions for mathematical expressions

Here we summarize the notational conventions used in the main text and supplement. Matrices are written using bold, uppercase letters (e.g., $\boldsymbol{A}$), column vectors are written as bold, lowercase letters (e.g., $\boldsymbol{a}$), and scalars are written in plain font (e.g., a, A). For indexing, we usually use a capital letter to denote the total number of elements, and we use the corresponding lowercase symbol to denote the index; e.g., $j = 1, \ldots, J$. We use $\mathbb{R}$ to denote the real numbers and $\mathbb{R}^d$ for the set of real vectors of dimension d. We use Δ^d to denote the simplex on $\mathbb{R}^d$; that is, all $x \in \mathbb{R}^d$ such that $x_1 + \cdots + x_d = 1$, $x_i \geqslant 0$, $i = 1, \ldots, d$. We use $\mathbb{R}^{m \times n}$ to denote the set of all $m \times n$ matrices with real entries, $\mathbb{S}^n_{++}$ for the set of all $n \times n$ real, symmetric positive definite matrices, and $\mathbb{S}^n_{+}$ for the set of all $n \times n$ real, symmetric positive semi-definite matrices (this set may include matrices that are singular, or not invertible). We write the matrix transpose of $\boldsymbol{A}$ as $\boldsymbol{A}^\intercal$. For a square matrix $\boldsymbol{A}$, its inverse is $\boldsymbol{A}^{-1}$, its determinant is $\det \boldsymbol{A}$, also written as $|\boldsymbol{A}|$, the trace is $\mathrm{tr}(\boldsymbol{A})$, and $\boldsymbol{A}^\dagger$ denotes the Moore–Penrose inverse ("pseudoinverse") of $\boldsymbol{A}$. We use $\boldsymbol{I}_n$ for the $n \times n$ identity matrix, and we use $\boldsymbol{1}_n$ as a shorthand

for a column vector of ones of length n. We use $\boldsymbol{a}^\intercal$ to denote a row vector. We denote the outer product of (column) vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as $\boldsymbol{a} \otimes \boldsymbol{b} := \boldsymbol{a}\boldsymbol{b}^\intercal$, and we denote the elementwise product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as $\boldsymbol{A} \circ \boldsymbol{B}$. Finally, we typically denote ordered or unordered sets using calligraphic letters (e.g., $\mathcal{A}$).

The multivariate simple regression model

The mvSuSiE model (2) is based on a simple multivariate regression model with one variable which we refer to as the "multivariate simple regression model." This model is

$$\boldsymbol{Y} \sim MN_{N\times R}(\boldsymbol{x}\boldsymbol{b}^\intercal, \boldsymbol{I}_N, \boldsymbol{V}). \tag{26}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N\times R}$ is a matrix of R observed responses in N samples, $\boldsymbol{x}$ is a vector of N observations for a single explanatory variable, $\boldsymbol{b} \in \mathbb{R}^R$ is the (unknown) vector of regression coefficients for the R responses, $\boldsymbol{V} \in \mathbb{S}^R_{++}$ is an invertible covariance matrix, and $MN_{n\times m}(\boldsymbol{M}, \boldsymbol{U}, \boldsymbol{V})$ is the matrix normal distribution [26, 36] with mean $\boldsymbol{M} \in \mathbb{R}^{n\times m}$ and covariance matrices $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{S}^{m\times m}_+$. For now, this model does not include an intercept.

In the following, we give the expression for the likelihood for (26) and relate it to the more familiar multivariate normal distribution. The likelihood for the multivariate regression model (26) is

$$\begin{aligned}\ell(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) &:= p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{b}, \boldsymbol{V}) \\ &= |2\pi\boldsymbol{V}|^{-N/2} \exp\big\{ -\tfrac{1}{2}\mathrm{tr}\big[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)\big]\big\}.\end{aligned} \tag{27}$$

Given $\boldsymbol{V}$, the least-squares estimate of $\boldsymbol{b}$, denoted by $\hat{\boldsymbol{b}}$, and its variance-covariance matrix, $\hat{\boldsymbol{S}}$, are

$$\hat{\boldsymbol{b}} = \frac{\boldsymbol{Y}^\intercal\boldsymbol{x}}{\boldsymbol{x}^\intercal\boldsymbol{x}} \tag{28}$$

$$\hat{\boldsymbol{S}} = \frac{\boldsymbol{V}}{\boldsymbol{x}^\intercal\boldsymbol{x}}. \tag{29}$$

Note that $\hat{\boldsymbol{b}}$ is also the value of $\boldsymbol{b}$ maximizing the likelihood (27). Using these quantities, the likelihood (27) can be rewritten as

$$\ell(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) = |2\pi\boldsymbol{V}|^{-N/2} \exp\big\{ -\tfrac{1}{2}[\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{Y}^\intercal\boldsymbol{Y}) + (\boldsymbol{b} - \hat{\boldsymbol{b}})^\intercal\hat{\boldsymbol{S}}^{-1}(\boldsymbol{b} - \hat{\boldsymbol{b}}) - \hat{\boldsymbol{b}}^\intercal\hat{\boldsymbol{S}}^{-1}\hat{\boldsymbol{b}}]\big\}. \tag{30}$$

This expression is convenient because the terms involving $\boldsymbol{b}$ are multivariate normal up to a constant of proportionality, and in particular we have that

$$\ell(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) \propto N_R(\boldsymbol{b}; \hat{\boldsymbol{b}}, \hat{\boldsymbol{S}}), \tag{31}$$

where $N_n(\boldsymbol{\theta}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal density at $\boldsymbol{\theta} \in \mathbb{R}^n$ with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance $\boldsymbol{\Sigma} \in \mathbb{S}^n_+$.

Remark 1. Calculation of $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{S}}$ only requires the summary statistics $\boldsymbol{x}^\intercal\boldsymbol{x}$, $\boldsymbol{Y}^\intercal\boldsymbol{x}$. Also, since the likelihood (31) only involves $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{S}}$, the likelihood up to a constant of proportionality can also be computed with only these summary statistics.

The multivariate simple regression model with a normal prior

In the following proposition, we apply the above results for the multivariate simple regression model to a Bayesian multivariate simple regression model with a normal prior.

Proposition 1 (Bayesian multivariate simple regression with a normal prior). Consider the multivariate simple regression model (26) with a multivariate normal prior on the regression coefficients,

$$\boldsymbol{b} \mid \boldsymbol{S}_0 \sim N_R(\boldsymbol{0}, \boldsymbol{S}_0), \tag{32}$$

where $\boldsymbol{S}_0 \in \mathbb{S}^R_+$ is a (possibly singular) covariance matrix. The posterior of $\boldsymbol{b}$ is

$$\boldsymbol{b} \mid \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_0 \sim N_R(\boldsymbol{b}_1, \boldsymbol{S}_1), \tag{33}$$

where

$$\boldsymbol{b}_1 = \boldsymbol{S}_1 \hat{\boldsymbol{S}}^{-1} \hat{\boldsymbol{b}} \tag{34}$$

$$\boldsymbol{S}_1 = (\boldsymbol{S}_0^{-1} + \hat{\boldsymbol{S}}^{-1})^{-1}. \tag{35}$$

The Bayes Factor (BF) comparing this model against the null model ($\boldsymbol{b} = \boldsymbol{0}$) is

$$\begin{aligned} \mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_0) &= \frac{p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{V}, \boldsymbol{S}_0)}{p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{V}, \boldsymbol{b} = \boldsymbol{0})} \\ &= \frac{\int \ell(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V})\, p(\boldsymbol{b} \mid \boldsymbol{S}_0)\, d\boldsymbol{b}}{\ell(\boldsymbol{b} = \boldsymbol{0}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V})}, \end{aligned} \tag{36}$$

reduces to

$$\mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_0) = \frac{|\hat{\boldsymbol{S}}|^{1/2}}{|\boldsymbol{S}_0 + \hat{\boldsymbol{S}}|^{1/2}} \exp\{\tfrac{1}{2} \hat{\boldsymbol{b}}^\intercal \hat{\boldsymbol{S}}^{-1} \boldsymbol{S}_1 \hat{\boldsymbol{S}}^{-1} \hat{\boldsymbol{b}}\} \tag{37}$$

$$= \frac{|\hat{\boldsymbol{S}}|^{1/2}}{|\boldsymbol{S}_0 + \hat{\boldsymbol{S}}|^{1/2}} \exp\{\tfrac{1}{2} \boldsymbol{b}_1^\intercal \boldsymbol{S}_1^{-1} \boldsymbol{b}_1\}. \tag{38}$$

The same BF can also be equivalently expressed as a ratio of two multivariate normal densities,

$$\mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_0) = \frac{N_R(\hat{\boldsymbol{b}}; \boldsymbol{0}, \boldsymbol{S}_0 + \hat{\boldsymbol{S}})}{N_R(\hat{\boldsymbol{b}}; \boldsymbol{0}, \hat{\boldsymbol{S}})}. \tag{39}$$

Remark 2. Since the data $\boldsymbol{x}, \boldsymbol{Y}$ only enter the expressions for the posterior mean $\boldsymbol{b}_1$ and posterior covariance $\boldsymbol{S}_1$ through $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{S}}$, it follows that calculation of the posterior mean and covariance only requires summary statistics $\boldsymbol{x}^\intercal \boldsymbol{x}, \boldsymbol{Y}^\intercal \boldsymbol{x}$. Similarly, the BF (39) can be computed with only $\boldsymbol{x}^\intercal \boldsymbol{x}, \boldsymbol{Y}^\intercal \boldsymbol{x}$.

Updating a scaling factor in the prior. Suppose the prior covariance is parameterized as $\boldsymbol{S}_0 = \sigma_0^2 \boldsymbol{U}$, in which $\boldsymbol{U} \in \mathbb{S}_+^R$, $\sigma_0 \geqslant 0$. Here we assume $\boldsymbol{U}$ is a fixed parameter and we would like to estimate the scaling factor σ_0 by maximizing the likelihood,

$$\hat{\sigma}_0^2 := \underset{\sigma_0^2}{\operatorname{argmax}}\, p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{V}, \sigma_0^2 \boldsymbol{U}), \tag{40}$$

or, equivalently, by maximizing the Bayes Factor, which may be more convenient to compute,

$$\hat{\sigma}_0^2 = \underset{\sigma_0^2}{\operatorname{argmax}}\, \mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \sigma_0^2 \boldsymbol{U}). \tag{41}$$

When $\boldsymbol{U}$ is invertible, the maximum-likelihood estimate $\hat{\sigma}_0^2$ can be computed using a simple EM algorithm [28], in which the M-step update is

$$\sigma_0^2 = \mathrm{tr}(\boldsymbol{U}^{-1} \mathbb{E}[\boldsymbol{b}\boldsymbol{b}^\intercal])/R. \tag{42}$$

The E-step then consists of computing the posterior second moment,

$$\mathbb{E}[\boldsymbol{b}\boldsymbol{b}^\intercal] = \boldsymbol{b}_1 \boldsymbol{b}_1^\intercal + \boldsymbol{S}_1, \tag{43}$$

in which the posterior mean $\boldsymbol{b}_1$ and posterior covariance $\boldsymbol{S}_1$ are given by (34, 35). The maximum-likelihood estimate $\hat{\sigma}_0^2$ is then recovered by iterating the E-step (43) and M-step (42) until convergence.

The update (42) requires that $\boldsymbol{U}$ be invertible. To allow for singular matrices, a more general M-step update is

$$\sigma_0^2 = \mathrm{tr}(\boldsymbol{U}^\dagger \mathbb{E}[\boldsymbol{b}\boldsymbol{b}^\intercal])/R'. \tag{44}$$

in which $R' \leqslant R$ is the rank of $\boldsymbol{U}$. Note (42) and (44) are equivalent when $\boldsymbol{U}$ is invertible; that is, when $R' = R$.

The multivariate simple regression model with an intercept

Now we extend the multivariate simple regression model (26) to include an intercept. We show that including an intercept in the model is equivalent to "centering" $\boldsymbol{x}$ and the columns of $\boldsymbol{Y}$ so that they all have means of zero. More precisely, centering is equivalent to integrating out the intercept with respect to an (improper) uniform prior on the intercept. This is the multivariate generalization of the result for univariate regression given in [21]. This result is summarized in Proposition 2.

The multivariate simple regression model with an intercept is

$$\boldsymbol{Y} \sim MN_{N\times R}(\boldsymbol{1}_N\boldsymbol{\mu}^\intercal + \boldsymbol{x}\boldsymbol{b}^\intercal, \boldsymbol{I}_N, \boldsymbol{V}), \tag{45}$$

in which $\boldsymbol{\mu} \in \mathbb{R}^R$ is the (unknown) intercept. The likelihood of $\boldsymbol{\mu}, \boldsymbol{b}$ under this model is

$$\begin{aligned}\ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) &:= p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{V}) \\ &= |2\pi\boldsymbol{V}|^{-N/2} \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{1}\boldsymbol{\mu}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{1}\boldsymbol{\mu}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)]\}.\end{aligned} \tag{46}$$

Proposition 2 (Multivariate simple regression with an intercept). Consider the multivariate simple regression model with an intercept (45). The least-squares estimate of $\boldsymbol{\mu}$—which is also the value of $\boldsymbol{\mu}$ maximizing the likelihood (46)—and its covariance matrix are

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b} \tag{47}$$

$$\hat{\boldsymbol{S}}_\mu = \tfrac{1}{N}\boldsymbol{V}, \tag{48}$$

in which $\bar{x} := \frac{1}{N}\boldsymbol{x}^\intercal\boldsymbol{1}_N = \frac{1}{N}\sum_{i=1}^N x_i$ is the sample mean of $\boldsymbol{x}$, and $\bar{\boldsymbol{y}} := \frac{1}{N}\boldsymbol{Y}^\intercal\boldsymbol{1}_N$ is the vector containing the column means of $\boldsymbol{Y}$.

The profile likelihood for $\boldsymbol{b}$ is

$$\begin{aligned}\ell^*(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) &:= \max_{\boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) \\ &= \ell(\boldsymbol{b}; \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{Y}}, \boldsymbol{V}),\end{aligned} \tag{49}$$

in which $\tilde{\boldsymbol{x}} := \boldsymbol{x} - \bar{x}\boldsymbol{1}_N$ and $\tilde{\boldsymbol{Y}} := \boldsymbol{Y} - \boldsymbol{1}_N\bar{\boldsymbol{y}}^\intercal$ are the centered $\boldsymbol{x}$ and $\boldsymbol{Y}$. In other words, the profile likelihood for the multivariate simple regression with an intercept is the same as the likelihood for the multivariate simple regression *without* an intercept if we first center $\boldsymbol{x}$ and $\boldsymbol{Y}$. Centering $\boldsymbol{x}$ and $\boldsymbol{Y}$ is therefore equivalent to including an intercept in the multivariate regression and estimating the intercept by maximum-likelihood.

Next, consider Bayesian calculations for $\boldsymbol{\mu}$ with a multivariate normal prior, $\boldsymbol{\mu} \mid \boldsymbol{S}_{0\mu} \sim N_R(0, \boldsymbol{S}_{0\mu})$, in which $\boldsymbol{S}_{0\mu} \in \mathbb{S}_+^R$ is a (possibly singular) covariance matrix. The posterior for $\boldsymbol{\mu}$ conditioned on $\boldsymbol{b}$ is

$$\boldsymbol{\mu} \mid \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}, \boldsymbol{b} \sim N_R(\boldsymbol{\mu}_1, \boldsymbol{S}_{1\mu}), \tag{50}$$

where

$$\boldsymbol{\mu}_1 = \boldsymbol{S}_{1\mu}\hat{\boldsymbol{S}}_\mu^{-1}\hat{\boldsymbol{\mu}} \tag{51}$$

$$\boldsymbol{S}_{1\mu} = (\boldsymbol{S}_{0\mu}^{-1} + \hat{\boldsymbol{S}}_\mu^{-1})^{-1}. \tag{52}$$

The marginal likelihood obtained by averaging over the intercept is

$$\begin{aligned}\ell^\star(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}) &:= \textstyle\int \ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V})\, p(\boldsymbol{\mu} \mid \boldsymbol{S}_{0\mu})\, d\boldsymbol{\mu} \\ &= |2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{0\mu}^{-1}\boldsymbol{S}_{1\mu}|^{1/2} \exp\{\tfrac{1}{2}\boldsymbol{\mu}_1^\intercal\boldsymbol{S}_{1\mu}^{-1}\boldsymbol{\mu}_1 - \tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)]\}.\end{aligned} \tag{53}$$

In the special case of an (improper) uniform prior on $\boldsymbol{\mu}$, defined as $\boldsymbol{\mu} \sim N_R(0, \boldsymbol{S}_{0\mu})$ with $\boldsymbol{S}_{0\mu}^{-1} \to 0$, the posterior mean reduces to the least-squares estimate $\boldsymbol{\mu}_1 = \hat{\boldsymbol{\mu}}$, with covariance matrix $\boldsymbol{S}_{1\mu} = \hat{\boldsymbol{S}}_\mu$, and the marginal likelihood (53) simplifies to

$$\begin{aligned}\ell^\star(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}) &= |2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{0\mu}^{-1}\hat{\boldsymbol{S}}_\mu|^{1/2} \exp\left\{\tfrac{1}{2}\hat{\boldsymbol{\mu}}^\intercal\hat{\boldsymbol{S}}_\mu^{-1}\hat{\boldsymbol{\mu}} - \tfrac{1}{2}\mathrm{tr}\big[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)\big]\right\} \\ &= |\boldsymbol{S}_{0\mu}^{-1}\hat{\boldsymbol{S}}_\mu|^{1/2} \times \ell(\boldsymbol{b}; \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{Y}}, \boldsymbol{V}),\end{aligned} \tag{54}$$

in which $\tilde{\boldsymbol{x}} := \boldsymbol{x} - \bar{x}\mathbf{1}$ and $\tilde{\boldsymbol{Y}} := \boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^\intercal$ are the centered $\boldsymbol{x}$ and $\boldsymbol{Y}$. In other words, the marginal likelihood for multivariate simple regression with an intercept (45), when we use an improper uniform prior for the intercept, is the same (up to a constant of proportionality) as the likelihood for multivariate simple regression *without* an intercept (26) after first centering $\boldsymbol{x}$ and $\boldsymbol{Y}$.

See below for a proof of this result.

Remark 3. To account for an intercept when computing posterior quantities and BFs for the multivariate simple regression model (26), $\boldsymbol{x}$ and $\boldsymbol{Y}$ should be centered before computing the summary statistics; that is, the summary statistics should be $\tilde{\boldsymbol{x}}^\intercal\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{Y}}^\intercal\tilde{\boldsymbol{x}}$. See [106] for how to center summary statistics if they are not centered.

The Bayesian multivariate simple regression model with a mixture prior

Here we extend the Bayesian multivariate simple regression model with a normal prior to a model with a mixture-of-normals prior,

$$\boldsymbol{b} \mid \mathcal{S}_0, \boldsymbol{\omega} \sim \sum_{k=1}^{K} \omega_k N_R(\mathbf{0}, \boldsymbol{S}_{0k}), \tag{55}$$

in which $\mathcal{S}_0 := \{\boldsymbol{S}_{01}, \ldots, \boldsymbol{S}_{0K}\}$, each $\boldsymbol{S}_{0k} \in \mathbb{S}_+^R$ is a (possibly singular) covariance matrix, and $\boldsymbol{\omega} := (\omega_1, \ldots, \omega_K) \in \Delta^K$ are the mixture weights. The normal prior (32) is as a special case of (55) when $K = 1$.

To facilitate derivation of the posterior computations, we introduce the following data augmentation that recovers (55) after integrating over a latent random variable $\xi \in \{1, \ldots, K\}$,

$$\begin{aligned} &p(\xi = k \mid \boldsymbol{\omega}) = \omega_k \\ &\boldsymbol{b} \mid \mathcal{S}_0, \xi = k \sim N_R(\mathbf{0}, \boldsymbol{S}_{0k}). \end{aligned} \tag{56}$$

This augmented model allows us to reuse the posterior computations from the simpler models; in particular, posterior computations conditioned on ξ reduce to computations for the Bayesian multivariate regression model with a normal prior, which we state formally in the following proposition.

Proposition 3. Given $\mathcal{S}_0$ and $\boldsymbol{\omega}$, the Bayes factor comparing this model against the null model ($\boldsymbol{b} = \mathbf{0}$) is

$$\begin{aligned} \mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) &= \frac{p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})}{p(\boldsymbol{Y} \mid \boldsymbol{x}, \boldsymbol{V}, \boldsymbol{b} = \mathbf{0})} \\ &= \sum_{k=1}^{K} \omega_k \mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0k}), \end{aligned} \tag{57}$$

where the expressions for the individual BFs in the sum are given in Proposition 1. The posterior distribution of $\boldsymbol{b}$ is a mixture of normals,

$$\boldsymbol{b} \mid \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega} \sim \sum_{k=1}^{K} \omega_{1k} N_R(\boldsymbol{b}_{1k}, \boldsymbol{S}_{1k}), \tag{58}$$

in which $\boldsymbol{b}_{1k}$ and $\boldsymbol{S}_{1k}$ are the posterior mean and covariance of $\boldsymbol{b}$ conditioned on $\xi = k$, given by (34) and (35), respectively, after substituting $\boldsymbol{S}_0$ with $\boldsymbol{S}_{0k}$,

$$\boldsymbol{b}_{1k} := \boldsymbol{b}_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) = \boldsymbol{S}_{1k}\hat{\boldsymbol{S}}^{-1}\hat{\boldsymbol{b}} \tag{59}$$

$$\boldsymbol{S}_{1k} := \boldsymbol{S}_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) = (\boldsymbol{S}_{0k}^{-1} + \hat{\boldsymbol{S}}^{-1})^{-1}, \tag{60}$$

and the posterior mixture assignment probabilities ("responsibilities") are

$$\begin{aligned}\omega_{1k} &:= \omega_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \\ &= p(\xi = k \mid \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \\ &= \frac{\omega_k \mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0k})}{\sum_{k'=1}^{K} \omega_{k'} \mathrm{BF}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0k'})}. \end{aligned} \tag{61}$$

The posterior mean and covariance of $\boldsymbol{b}$ are

$$\boldsymbol{b}_1^{\mathsf{mix}} := \boldsymbol{b}_1^{\mathsf{mix}}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) = \textstyle\sum_{k=1}^{K} \omega_{1k} \boldsymbol{b}_{1k} \tag{62}$$

$$\boldsymbol{S}_1^{\mathsf{mix}} := \boldsymbol{S}_1^{\mathsf{mix}}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) = \textstyle\sum_{k=1}^{K} \omega_{1k}(\boldsymbol{b}_{1k}\boldsymbol{b}_{1k}^{\mathsf{T}} + \boldsymbol{S}_{1k}) - \boldsymbol{b}_1^{\mathsf{mix}}(\boldsymbol{b}_1^{\mathsf{mix}})^{\mathsf{T}}. \tag{63}$$

From the above remarks, the posterior quantities and Bayes factors for this model can be computed using the summary statistics $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}$, $\boldsymbol{Y}^{\mathsf{T}}\boldsymbol{x}$ instead of using the full data $\boldsymbol{x}, \boldsymbol{Y}$. To formalize these computations with summary statistics, we introduce notation for Bayes factors and posteriors in terms of summary statistics:

$$\boldsymbol{b}_{1k}^{\mathsf{ss}}(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) := \boldsymbol{b}_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{64}$$

$$\boldsymbol{S}_{1k}^{\mathsf{ss}}(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) := \boldsymbol{S}_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{65}$$

$$\omega_{1k}^{\mathsf{ss}}(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) := \omega_{1k}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{66}$$

$$\mathrm{BF}^{\mathsf{mix\text{-}ss}}(\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{x}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) := \mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}). \tag{67}$$

Updating a scaling factor in the prior. Similar to above, here we consider a special case of the mixture-of-normals prior in which the prior covariances are parameterized as $\boldsymbol{S}_{0k} = \sigma_0^2 \boldsymbol{U}_k$, $\boldsymbol{U}_k \in \mathbb{S}_+^R$, $k = 1, \ldots, K$, $\sigma_0 \geqslant 0$. We assume the $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_K$ are fixed parameters and we would like to estimate the scaling factor σ_0 by maximizing the likelihood,

$$\hat{\sigma}_0^2 := \underset{\sigma_0^2}{\operatorname{argmax}}\, p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}), \tag{68}$$

This is the same as maximizing the BF since the denominator in the BF does not depend on σ_0, and the BF may be more convenient to compute:

$$\hat{\sigma}_0^2 = \underset{\sigma_0^2}{\operatorname{argmax}}\, \mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}). \tag{69}$$

Again taking a simple EM approach to computing the maximum-likelihood estimate, the M-step update allowing for singular matrices is

$$\sigma_0^2 = \sum_{k=1}^{K} \frac{\omega_{1k}}{R_k} \times \mathrm{tr}(\boldsymbol{U}_k^{\dagger} \mathbb{E}[\boldsymbol{b}\boldsymbol{b}^{\mathsf{T}} \mid \xi = k]), \tag{70}$$

in which the E-step involves computing the posterior probabilities ω_{1k} and the posterior second moments,

$$\mathbb{E}[\boldsymbol{b}\boldsymbol{b}^{\mathsf{T}} \mid \xi = k] = \boldsymbol{b}_{1k}\boldsymbol{b}_{1k}^{\mathsf{T}} + \boldsymbol{S}_{1k}. \tag{71}$$

Here, $R_k \leqslant R$ denotes the rank of $\boldsymbol{U}_k$.

The multivariate single effect regression model

The single effect regression (SER) model is a multiple regression model in which exactly one of the regression coefficients is non-zero [96]. Here we define the multivariate single effect regression (MSER) model, which extends the SER model to the multivariate setting, and forms the basis for mvSuSiE.

The MSER model is

$$\begin{aligned} \boldsymbol{Y} &\sim MN_{N\times R}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{I}_N, \boldsymbol{V}) \\ \boldsymbol{B} &= \boldsymbol{\gamma} \otimes \boldsymbol{b}, \\ \boldsymbol{\gamma} &\sim \text{Multinom}(1, \boldsymbol{\pi}), \\ \boldsymbol{b} &\sim \sum_{k=1}^{K} \omega_k N_R(\boldsymbol{0}, \boldsymbol{S}_{0k}), \end{aligned} \tag{72}$$

in which $\boldsymbol{Y} \in \mathbb{R}^{N\times R}$ is a matrix storing N observations about R regression outcomes, $\boldsymbol{X} \in \mathbb{R}^{N\times J}$ is a matrix storing N observations about J regression variables, $\boldsymbol{B} \in \mathbb{R}^{J\times R}$ is a matrix of regression coefficients for the J variables and R outcomes, $\boldsymbol{b} \in \mathbb{R}^R$ is a vector of "single effect" regression coefficients, $\boldsymbol{V} \in \mathbb{S}^R_{++}$ is an invertible covariance matrix, $\boldsymbol{S}_{01}, \ldots, \boldsymbol{S}_{0K} \in \mathbb{S}^R_{+}$ are (possibly singular) prior covariance matrices, $\boldsymbol{\gamma} \in \{0,1\}^J$ is a random variable indicating which of the J variables explains the multivariate outcome (exactly one element is one, and the remaining are zero), $\boldsymbol{\pi} \in \Delta^J$ specifies the probabilities in the multinomial prior on $\boldsymbol{\gamma}$, $\boldsymbol{\omega} \in \Delta^K$ is a vector specifying the weights in the mixture-of-normals prior on $\boldsymbol{b}$, and $\text{Multinom}(m, \boldsymbol{\pi})$ denotes the multinomial distribution for m random trials with probabilities $\boldsymbol{\pi}$. Because $\boldsymbol{\gamma}$ is drawn from the multinomial distribution with a single random trial, $\boldsymbol{\gamma}$ has exactly one non-zero element; that is, $\gamma_k = 1$ for some k, and $\gamma_{k'} = 0$ for all other $k' \neq k$. From this property, it follows that the matrix $\boldsymbol{B}$ has at most one row containing non-zero elements, and this row is given by $\boldsymbol{b}$. The MSER model recovers the SER model when $R = 1$, $K = 1$, $\omega_1 = 1$ and $\boldsymbol{S}_{01} = 1$.

The posterior distribution of $\boldsymbol{B}, \boldsymbol{\gamma}$ is summarized by the following proposition.

Proposition 4. Let $\mathcal{S}_0 = \{\boldsymbol{S}_{01}, \ldots, \boldsymbol{S}_{0K}\}$ denote the K prior covariance matrices. Under the MSER model (72), the posterior distribution of $\boldsymbol{B} = \boldsymbol{\gamma} \otimes \boldsymbol{b}$ given parameter settings $\Theta := \{\boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}, \boldsymbol{\pi}\}$ is given by

$$\begin{aligned} \boldsymbol{\gamma} \mid \boldsymbol{X}, \boldsymbol{Y}, \Theta &\sim \text{Multinom}(1, \boldsymbol{\alpha}) \\ \boldsymbol{b} \mid \boldsymbol{X}, \boldsymbol{Y}, \Theta, \gamma_j = 1 &\sim \sum_{k=1}^{K} \omega_{1jk} N_R(\boldsymbol{b}_{1jk}, \boldsymbol{S}_{1jk}), \end{aligned} \tag{73}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ is the vector of posterior inclusion probabilities (PIPs), which can be written using the Bayes factors (57) for the Bayesian multivariate simple regression model with a mixture of normals prior,

$$\begin{aligned} \alpha_j &:= \alpha_j(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \\ &= \Pr(\gamma_j = 1 \mid \boldsymbol{X}, \boldsymbol{Y}, \Theta) \\ &= \frac{\pi_j \text{BF}^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})}{\sum_{j'=1}^{J} \pi_{j'} \text{BF}^{\mathsf{mix}}(\boldsymbol{x}_{j'}, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})}, \end{aligned} \tag{74}$$

and where the means $\boldsymbol{b}_{1jk}$, variances $\boldsymbol{S}_{1jk}$ and posterior mixture weights ω_{1jk} are given by

$$\boldsymbol{b}_{1jk} = \boldsymbol{b}_{1k}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{75}$$

$$\boldsymbol{S}_{1jk} = \boldsymbol{S}_{1k}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{76}$$

$$\omega_{1jk} = \omega_{1k}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}), \tag{77}$$

using the definitions of ω_{1k}, $\boldsymbol{b}_{1k}$ and $\boldsymbol{S}_{1k}$ in eqs. 59–61. The posterior mean and covariance of $\boldsymbol{b}$ conditioned on $\boldsymbol{\gamma}$ are

$$\mathbb{E}[\boldsymbol{b} \mid \gamma_j = 1] = \boldsymbol{b}_1^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{78}$$

$$\text{Cov}[\boldsymbol{b} \mid \gamma_j = 1] = \boldsymbol{S}_1^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}), \tag{79}$$

using definitions (62,63), and therefore the posterior mean of $\boldsymbol{B}$ is

$$\mathbb{E}[\boldsymbol{B}] = \begin{bmatrix} \alpha_1 \boldsymbol{b}_1^{\mathsf{mix}}(\boldsymbol{x}_1, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})^\mathsf{T} \\ \vdots \\ \alpha_J \boldsymbol{b}_1^{\mathsf{mix}}(\boldsymbol{x}_J, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})^\mathsf{T} \end{bmatrix}. \tag{80}$$

For describing the model fitting algorithms (below), we define a function, MSER, that returns the posterior distribution of $\boldsymbol{b}, \boldsymbol{\gamma}$ under the MSER model given the data $\boldsymbol{X}, \boldsymbol{Y}$ and the model parameters Θ:

$$\mathrm{MSER}(\boldsymbol{X}, \boldsymbol{Y}; \Theta) := (\boldsymbol{\alpha}, \boldsymbol{\Omega}_1, \mathcal{B}_1, \mathcal{S}_1), \tag{81}$$

in which $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ is the vector of PIPs (74), $\boldsymbol{\Omega}_1$ is the $J \times R$ matrix of posterior mixture weights ω_{1jk} (77), $\mathcal{B}_1$ is the set of posterior means $\boldsymbol{b}_{1jk}$ (75) for all j, k, and $\mathcal{S}_1$ is the set of posterior covariances $\boldsymbol{S}_{1jk}$ (76) for all j, k.

The multivariate single effect regression model with summary statistics. From the above remarks, $\boldsymbol{X}^\intercal \boldsymbol{X}$ and $\boldsymbol{X}^\intercal \boldsymbol{Y}$ are sufficient to compute the posterior distribution of $\boldsymbol{b}, \boldsymbol{\gamma}$. (More precisely, only the diagonal elements of $\boldsymbol{X}^\intercal \boldsymbol{X}$ are necessary to compute the posterior distribution of $\boldsymbol{b}, \gamma$.) To formalize these computations with summary statistics, we also define Bayes factors and posterior quantities in terms of summary statistics:

$$\boldsymbol{b}_{1jk}^{\mathsf{ss}} := \boldsymbol{b}_{1k}^{\mathsf{ss}}(\boldsymbol{x}_j^\intercal \boldsymbol{x}_j, \boldsymbol{Y}^\intercal \boldsymbol{x}_j, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{82}$$

$$\boldsymbol{S}_{1jk}^{\mathsf{ss}} := \boldsymbol{S}_{1k}^{\mathsf{ss}}(\boldsymbol{x}_j^\intercal \boldsymbol{x}_j, \boldsymbol{Y}^\intercal \boldsymbol{x}_j, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}) \tag{83}$$

$$\omega_{1jk}^{\mathsf{ss}} := \omega_{1k}^{\mathsf{ss}}(\boldsymbol{x}_j^\intercal \boldsymbol{x}_j, \boldsymbol{Y}^\intercal \boldsymbol{x}_j, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}), \tag{84}$$

and the PIPs,

$$\alpha_j^{\mathsf{ss}} = \frac{\mathrm{BF}^{\mathsf{mix\text{-}ss}}(\boldsymbol{x}_j^\intercal \boldsymbol{x}_j, \boldsymbol{Y}^\intercal \boldsymbol{x}_j, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})}{\sum_{j'=1}^{J} \pi_{j'} \mathrm{BF}^{\mathsf{mix\text{-}ss}}(\boldsymbol{x}_{j'}^\intercal \boldsymbol{x}_{j'}, \boldsymbol{Y}^\intercal \boldsymbol{x}_{j'}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega})}. \tag{85}$$

To describe the model fitting algorithms that work with summary statistics, we define a function, MSER-ss, that returns the posterior distribution of $\boldsymbol{b}, \boldsymbol{\gamma}$ under the MSER model given the summary statistics $\boldsymbol{X}^\intercal \boldsymbol{X}, \boldsymbol{X}^\intercal \boldsymbol{Y}$ and the model parameters Θ:

$$\text{MSER-ss}(\boldsymbol{X}^\intercal \boldsymbol{X}, \boldsymbol{X}^\intercal \boldsymbol{Y}; \Theta) := (\boldsymbol{\alpha}, \boldsymbol{\Omega}_1, \mathcal{B}_1, \mathcal{S}_1), \tag{86}$$

in which $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ is the vector of PIPs α_j^{ss} (85), $\boldsymbol{\Omega}_1$ is the $J \times R$ matrix of posterior mixture weights $\omega_{1jk}^{\mathsf{ss}}$ (84), $\mathcal{B}_1$ is the set of posterior means $\boldsymbol{b}_{1jk}^{\mathsf{ss}}$ (82) for all j, k, and $\mathcal{S}_1$ is the set of posterior covariances $\boldsymbol{S}_{1jk}^{\mathsf{ss}}$ (83) for all j, k.

Updating a scaling factor in the prior. Here we extend the MSER model to allow for estimating a scaling parameter, $\sigma_0^2 \geqslant 0$, in which $\boldsymbol{S}_{0k} = \sigma_0^2 \boldsymbol{U}_k, k = 1, \ldots, K$. As above, we estimate σ_0^2 by maximum likelihood,

$$\hat{\sigma}_0^2 := \underset{\sigma_0^2}{\operatorname{argmax}}\, p(\boldsymbol{Y} \mid \boldsymbol{X}, \Theta) \tag{87}$$

which is equivalent to maximizing a weighted sum of the Bayes factors,

$$\hat{\sigma}_0^2 = \underset{\sigma_0^2}{\operatorname{argmax}} \sum_{j=1}^{J} \pi_j \mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \mathcal{S}_0, \boldsymbol{\omega}). \tag{88}$$

As before, the maximum-likelihood estimate can be computed using a simple EM algorithm. The M-step update in the EM algorithm is given by

$$\sigma_0^2 = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{\alpha_j \omega_{1jk}}{R_k} \times \mathrm{tr}(\boldsymbol{U}_k^\dagger \mathbb{E}[\boldsymbol{b}\boldsymbol{b}^\intercal \mid \gamma_j = 1, \xi = k]), \tag{89}$$

in which $R_k \leqslant R$ denotes the rank of $\boldsymbol{U}_k$. The E-step in the EM algorithm consists of computing the posterior mixture weights ω_{1jk}, the posterior inclusion probabilities α_j, and the posterior second moments

$$\mathbb{E}[\boldsymbol{b}\boldsymbol{b}^\intercal \mid \gamma_j = 1, \xi = k] = \boldsymbol{b}_{1jk} \boldsymbol{b}_{1jk}^\intercal + \boldsymbol{S}_{1jk} \tag{90}$$

at the current setting of σ_0^2.

Algorithm 1 Iterative Bayesian Stepwise Selection (IBSS) for mvSuSiE

Require: Data $\boldsymbol{X} \in \mathbb{R}^{N \times J}$, $\boldsymbol{Y} \in \mathbb{R}^{N \times R}$.
Require: Maximum number of non-zero effects, $L \in \{1, \ldots, J\}$.
Require: Initial estimates of the posterior mean single effects, $\bar{\boldsymbol{B}}^{(l)} \in \mathbb{R}^{J \times R}$, $l = 1, \ldots, L$.
Require: Initial estimates of the prior scaling factors, $\sigma_{01}^2, \ldots, \sigma_{0L}^2 \geqslant 0$.
Require: Residual covariance matrix $\boldsymbol{V}$ (must be invertible); prior inclusion probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_J)$; prior mixture weights $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)$; and prior covariance matrices $\mathcal{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_K)$ (these do not need to be invertible).
Require: A function $\text{MSER}(\boldsymbol{X}, \boldsymbol{Y}; \Theta) \rightarrow (\boldsymbol{\alpha}, \boldsymbol{\Omega}_1, \mathcal{B}_1, \mathcal{S}_1)$ that returns the posterior distribution of $\boldsymbol{b}, \boldsymbol{\gamma}$ under the MSER model with data $\boldsymbol{X}, \boldsymbol{Y}$ and parameters Θ; see (81).

1: **repeat**
2: $\quad \bar{\boldsymbol{R}} \leftarrow \boldsymbol{Y} - \boldsymbol{X} \sum_{l=1}^{L} \bar{\boldsymbol{B}}^{(l)}$ ▷ Compute expected residuals.
3: $\quad$ **for** l in $1, \ldots, L$ **do**
4: $\quad\quad \bar{\boldsymbol{R}}_l \leftarrow \bar{\boldsymbol{R}} + \boldsymbol{X}\bar{\boldsymbol{B}}^{(l)}$ ▷ Disregard lth single effect in residuals.
5: $\quad\quad$ Update σ_{0l}^2 ▷ Optional; see (89).
6: $\quad\quad \Theta \leftarrow \{\boldsymbol{V}, \sigma_{0l}^2 \mathcal{U}, \boldsymbol{\omega}, \boldsymbol{\pi}\}$ ▷ Set MSER parameters.
7: $\quad\quad (\boldsymbol{\alpha}, \boldsymbol{\Omega}_1, \mathcal{B}_1, \mathcal{S}_1) \leftarrow \text{MSER}(\boldsymbol{X}, \bar{\boldsymbol{R}}_l; \Theta)$ ▷ Fit MSER to residuals.
8: $\quad\quad \boldsymbol{\alpha}_l \leftarrow \boldsymbol{\alpha}$ ▷ Store PIPs for lth single effect.
9: $\quad\quad$ Initialize $\boldsymbol{\mu}_l$ to a $J \times R$ matrix of zeros ▷ Compute conditional posterior means (78).
10: $\quad\quad$ **for** j in $1, \ldots, J$ **do**
11: $\quad\quad\quad$ Store $\sum_{k=1}^{K} \omega_{1jk} \boldsymbol{b}_{1jk}$ in row j of $\boldsymbol{\mu}_l$
12: $\quad\quad \bar{\boldsymbol{B}}^{(l)} \leftarrow (\boldsymbol{\alpha}_l \mathbf{1}_R^\intercal) \circ \boldsymbol{\mu}_l$ ▷ "$\circ$" denotes elementwise multiplication.
13: $\quad\quad \bar{\boldsymbol{R}} \leftarrow \bar{\boldsymbol{R}}_l - \boldsymbol{X}\bar{\boldsymbol{B}}^{(l)}$ ▷ Update expected residuals.
14: **until** convergence criterion is satisfied
15: **return** $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_L, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_L$

The mvSuSiE IBSS algorithm

The Iterative Bayesian Stepwise Selection (IBSS) algorithm for fitting mvSuSiE is derived by extending the ideas of [96] to the multivariate setting. Similar to IBSS for SuSiE, the IBSS algorithm for mvSuSiE is a coordinate ascent algorithm for optimizing a variational approximation [10, 47] to the posterior distribution for $\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}$ under the mvSuSiE model. The basic mvSuSiE IBSS algorithm is given in Algorithm 1.

Lines 6–11 compute the posterior mean regression coefficients $\boldsymbol{b}^{(l)}$ for the lth single effect conditioned on each of the variables $j = 1, \ldots, J$ being the "single effect variable." These posteriors are then stored in a $J \times R$ matrix (lines 10–11). Once these are computed and stored, the unconditional posterior means $\bar{\boldsymbol{B}}^{(l)}$ are simply the conditional posterior means, $\boldsymbol{\mu}_l$, weighted by the PIPs $\boldsymbol{\alpha}_l$.

The optional step of estimating the scaling factors σ_{0l}^2 (line 5) is mainly for pruning unneeded single effects (see "Choice of L"). This parameter estimation step can be viewed as an EM update in which the E-step is approximate [40, 66].

Derivation of the mvSuSiE IBSS algorithm. The IBSS algorithm fits an approximate posterior distribution for $\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}$ by minimizing a Kullback-Leibler (KL) divergence from the approximate posterior to the exact posterior subject to constraints on the approximate posterior. More precisely, denoting the exact posterior by p_{post} and the approximate posterior by q, we seek a q that minimizes the KL-divergence $D_{\mathsf{KL}}(q \,\|\, p_{\mathsf{post}})$ [24] subject to constraints on the q. Since the KL-divergence itself is hard to compute, we instead maximize the "evidence lower bound" (ELBO), which is equivalent to minimizing the KL-divergence, but easier to compute. The ELBO for the mvSuSiE model (2–7) is

$$F(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) = \mathbb{E}_q[\log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}, \boldsymbol{V})] + \mathbb{E}_q\left[\log\left\{\frac{p(\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)})}{q(\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)})}\right\}\right], \tag{91}$$

in which the prior $p(\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}) = \prod_{l=1}^{L} p_l(\boldsymbol{B}^{(l)})$ is defined according to (5–7).

Similar to [96], we constrain the approximate posterior so that it factorizes by the single effects:

$$q(\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}) = \prod_{l=1}^{L} q_l(\boldsymbol{B}^{(l)}). \tag{92}$$

Under this constraint, $\boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}$ are independent *a posteriori* annd each factor $q_l(\boldsymbol{B}^{(l)})$ is also a marginal posterior, and the right-hand part of the ELBO immediately simplifies:

$$F(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) = \mathbb{E}_q[\log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{B}^{(1)}, \ldots, \boldsymbol{B}^{(L)}, \boldsymbol{V})] + \sum_{l=1}^{L} \mathbb{E}_q\left[\log \frac{p_l(\boldsymbol{B}^{(l)})}{q_l(\boldsymbol{B}^{(l)})}\right], \tag{93}$$

In the next section, we show that this simple conditional independence assumption also results in tractable computations, and in particular the computations that make up Algorithm 1. Note that no other approximations or constraints are needed beyond the conditional independence assumption (92).

Fitting the MSER to the residuals in Algorithm 1 maximizes the ELBO for a single effect. The main result of this section is summarized by the following proposition.

PROPOSITION 5. Let $F(q, g, \boldsymbol{V}; \boldsymbol{Y})$ be the ELBO (91) under the mvSuSiE model (2–7), and suppose that q is constrained to factorize over the single effects $l = 1, \ldots, L$ as in (92). The setting of $q_l(\boldsymbol{B}^{(l)})$ that maximizes the ELBO (91) while $g, \boldsymbol{V}$ and the other factors $q_{l'}(\boldsymbol{B}^{(l')})$, for all $l' \neq l$, are fixed has the following closed-form solution:

$$\begin{aligned} \hat{q}_l &\coloneqq \operatorname{argmax}_{q_l} F(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) \\ &= \mathrm{MSER}(\boldsymbol{X}, \bar{\boldsymbol{R}}_l; \Theta), \end{aligned} \tag{94}$$

in which $\mathrm{MSER}(\boldsymbol{X}, \boldsymbol{Y}; \Theta)$ defined in (81) gives the posterior distribution of $\boldsymbol{B}$ under the multivariate single effect regression (MSER) model given data $\boldsymbol{X}, \boldsymbol{Y}$ and model parameters Θ, and we define

$$\begin{aligned} \bar{\boldsymbol{R}}_l &\coloneqq \mathbb{E}_q[\boldsymbol{Y} - \boldsymbol{X} \textstyle\sum_{l' \neq l} \boldsymbol{B}^{(l')}] \\ &= \boldsymbol{Y} - \boldsymbol{X} \textstyle\sum_{l' \neq l} \bar{\boldsymbol{B}}^{(l')} \end{aligned} \tag{95}$$

as the $N \times R$ matrix of expected residuals that ignore the lth single effect, and $\bar{\boldsymbol{B}}^{(l)} \coloneqq \mathbb{E}_q[\boldsymbol{B}^{(l)}]$.

COROLLARY 1. A corollary of Proposition 5 is that the optimal $q(\boldsymbol{B})$ factorizing as (92) is a product of factors $q_l(\boldsymbol{B}^{(l)})$ in which each factor is a posterior distribution under an MSER model (72).

This proposition and corollary are multivariate generalizations of the results for the univariate regression setting given in [96] (Propositions 1 and A1 in that paper). The proof of Proposition 5 is sketched out in the next two sections.

Special case when L = 1. The mvSuSiE model with $L = 1$ is an MSER model (72). For this special case, the ELBO (91) is

$$\begin{aligned} F_{\mathsf{MSER}}(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) &= \mathbb{E}_q[\log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{B}^{(1)}, \boldsymbol{V})] - D_{\mathsf{KL}}(q \,\|\, p) \\ &= -\tfrac{N}{2} \log |2\pi \boldsymbol{V}| - \tfrac{1}{2} \mathrm{tr}(\boldsymbol{V}^{-1} \mathrm{ERSS}) - D_{\mathsf{KL}}(q(\boldsymbol{B}^{(1)}) \,\|\, p(\boldsymbol{B}^{(1)})), \end{aligned} \tag{96}$$

in which "ERSS" is the multivariate expected residual sum of squares,

$$\mathrm{ERSS} \coloneqq \mathbb{E}_q[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(1)})^\intercal (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(1)})]. \tag{97}$$

In these expressions, we have dropped the "l" subscripts whenever they appear since there is only one single effect l. (We have kept the "(l)" superscript for $\boldsymbol{B}$ to avoid confusing $\boldsymbol{B}^{(1)}$ with $\boldsymbol{B}$ defined in eq. 3.)

The variational distribution maximizing the ELBO (96), $\hat{q} \coloneqq \operatorname{argmax}_q F_{\mathsf{MSER}}(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y})$, is the true posterior, $\hat{q}(\boldsymbol{B}^{(1)}) = p_{\mathsf{post}}(\boldsymbol{B}^{(1)}) \coloneqq p(\boldsymbol{B}^{(1)} \mid \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{V}, g)$ [10]; that is, q is the posterior distribution under an MSER model. At $\hat{q}$, the ELBO (96) is equal to the marginal log-likelihood; that is, $F_{\mathsf{MSER}}(\hat{q}, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) = \log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{V}, g)$.

Coordinate ascent update for lth single effect. To allow for posterior computations that are tractable, we restricted q to the set of posterior distributions that factorize over the single effects $l = 1, \ldots, L$ (92). With this approximation, we can now divide and conquer; we consider the problem of finding a $q_l(\boldsymbol{B}^{(l)})$ that maximizes the ELBO (91) while the remaining factors are fixed.

Expanding terms involving q_l, the ELBO (91) is

$$F(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) = -\tfrac{N}{2} \log |2\pi \boldsymbol{V}| - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\mathrm{ERSS}) - D_{\mathsf{KL}}(q_l(\boldsymbol{B}^{(l)}) \,\|\, g_l(\boldsymbol{B}^{(l)})) + \mathrm{const}, \tag{98}$$

where the "const" is a placeholder for terms in the ELBO not involving q_l, and the exprected residual sum of squares (ERSS) in this expression is

$$\mathrm{ERSS} := \mathbb{E}_q[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\intercal(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})]. \tag{99}$$

As a reminder, $\boldsymbol{B} = \sum_{l=1}^{L} \boldsymbol{B}^{(l)}$. Expanding the ERSS further, and making use of the property from (92) that the covariances are zero between all $\boldsymbol{B}^{(l)}$ and $\boldsymbol{B}^{(l')}$ whenever $l \neq l'$, the ELBO can be rewritten as

$$F(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}) = -\tfrac{N}{2} \log |2\pi \boldsymbol{V}| - \tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}\mathrm{ERSS}(l)] - D_{\mathsf{KL}}(q_l(\boldsymbol{B}^{(l)}) \,\|\, g_l(\boldsymbol{B}^{(l)})) + \mathrm{const}, \tag{100}$$

in which

$$\mathrm{ERSS}(l) := \mathbb{E}_q[(\bar{\boldsymbol{R}}_l - \boldsymbol{X}\boldsymbol{B}^{(l)})^\intercal(\bar{\boldsymbol{R}}_l - \boldsymbol{X}\boldsymbol{B}^{(l)})], \tag{101}$$

and $\bar{\boldsymbol{R}}_l$ is defined in (95). Assuming $l = 1$ without loss of generality, (100) is of the exact same form as (96) after ignoring terms that do not involve q_l, and after replacing $\boldsymbol{Y}$ with $\bar{\boldsymbol{R}}_l$.

In summary, the ELBO for mvSuSiE with $L > 1$ can be rearranged to exactly match the expression for the mvSuSiE ELBO with $L = 1$ if we ignore terms not involving q_l.

Computing the ELBO

While computing the ELBO (93) is not strictly needed to implement the IBSS algorithm, in practice it is useful for monitoring progress of the IBSS algorithm and for comparing different mvSuSiE model fits. Here we explain how we compute the ELBO.

Starting from (99), the "ERSS" part of the ELBO is

$$\begin{aligned}\mathrm{ERSS} &:= \mathbb{E}_q[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})^\intercal(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})] \\ &= (\boldsymbol{Y} - \boldsymbol{X}\bar{\boldsymbol{B}})^\intercal(\boldsymbol{Y} - \boldsymbol{X}\bar{\boldsymbol{B}}) - \textstyle\sum_{l=1}^{L}(\bar{\boldsymbol{B}}^{(l)})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}\bar{\boldsymbol{B}}^{(l)} + \mathbb{E}_q[(\boldsymbol{B}^{(l)})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{B}^{(l)}].\end{aligned} \tag{102}$$

From the definition of the mvSuSiE model, for a given single effect l only one row of $\boldsymbol{B}^{(l)}$ contains nonzero values, so $\mathbb{E}_q[b_{jr}b_{kr'}] = b_{jr}b_{kr'} = 0$ for any $j \neq k$ and $r, r' \in \{1, \ldots, R\}$, where b_{jr} is an entry of $\boldsymbol{B}^{(l)}$. Therefore, the ERSS simplifies to

$$\mathrm{ERSS} = (\boldsymbol{Y} - \boldsymbol{X}\bar{\boldsymbol{B}})^\intercal(\boldsymbol{Y} - \boldsymbol{X}\bar{\boldsymbol{B}}) - \sum_{l=1}^{L}(\bar{\boldsymbol{B}}^{(l)})^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}\bar{\boldsymbol{B}}^{(l)} + \sum_{l=1}^{L}(\bar{\boldsymbol{B}}^{(l)})^\intercal\boldsymbol{D}\bar{\boldsymbol{B}}^{(l)} + \sum_{l=1}^{L}\sum_{j=1}^{J} d_{jj}\boldsymbol{C}_j^{(l)}, \tag{103}$$

where $\boldsymbol{D}$ is a $J \times J$ diagonal matrix with diagonal entries $d_{jj} := (\boldsymbol{X}^\intercal\boldsymbol{X})_{jj}$ and $\boldsymbol{C}_j^{(l)}$ is the $R \times R$ covariance matrix for the jth row of $\boldsymbol{B}^{(l)}$ with respect to the approximate posterior q_l, which can be asily computed from the results of Proposition 4.

The remaining terms in the ELBO (93) are KL-divergences,

$$D_{\mathsf{KL}}(q_l \,\|\, p_l) = -\mathbb{E}_q\left[\log \frac{p_l(\boldsymbol{B}^{(l)})}{q_l(\boldsymbol{B}^{(l)})}\right]. \tag{104}$$

It is most convenient to compute this when $q_l(\boldsymbol{B}^{(l)})$ is updated, as we show next.

Algorithm 2 IBSS for mvSuSiE with sufficient statistics (IBSS-ss)

Require: Data $\boldsymbol{X}^\intercal\boldsymbol{Y} \in \mathbb{R}^{J\times R}$, $\boldsymbol{X}^\intercal\boldsymbol{X} \in \mathbb{R}^{J\times J}$.
Require: Maximum number of non-zero effects, $L \in \{1,\ldots,J\}$.
Require: Initial estimates of the posterior mean single effects, $\bar{\boldsymbol{B}}^{(l)} \in \mathbb{R}^{J\times R}$, $l = 1,\ldots,L$.
Require: Initial estimates of the prior scaling factors, $\sigma_{01}^2,\ldots,\sigma_{0L}^2 \geqslant 0$.
Require: Residual covariance matrix $\boldsymbol{V}$ (must be invertible); prior inclusion probabilities $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_J)$; prior mixture weights $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_K)$; and prior covariance matrices $\mathcal{U} = (\boldsymbol{U}_1,\ldots,\boldsymbol{U}_K)$ (these do not need to be invertible).
Require: A function MSER-ss$(\boldsymbol{X}^\intercal\boldsymbol{X}, \boldsymbol{X}^\intercal\boldsymbol{Y}; \Theta) \to (\boldsymbol{\alpha}, \boldsymbol{\omega}_1, \mathcal{B}_1, \mathcal{S}_1)$ that returns the posterior distribution of $\boldsymbol{b}, \boldsymbol{\gamma}$ under the MSER model with data $\boldsymbol{X}^\intercal\boldsymbol{X}, \boldsymbol{X}^\intercal\boldsymbol{Y}$ and parameters Θ.

repeat
  $\bar{\boldsymbol{P}} \leftarrow \boldsymbol{X}^\intercal\boldsymbol{Y} - \boldsymbol{X}^\intercal\boldsymbol{X}\sum_{l=1}^{L}\bar{\boldsymbol{B}}^{(l)}$ ▷ Compute expected residuals.
  for l in $1,\ldots,L$ **do**
    $\bar{\boldsymbol{P}}_l \leftarrow \bar{\boldsymbol{P}} + \boldsymbol{X}^\intercal\boldsymbol{X}\bar{\boldsymbol{B}}^{(l)}$ ▷ Disregard lth single effect in residuals.
    Update σ_{0l}^2 ▷ Optional; see (89).
    $\Theta \leftarrow \{\boldsymbol{V}, \sigma_{0l}^2\mathcal{U}, \boldsymbol{\omega}, \boldsymbol{\pi}\}$ ▷ Set MSER parameters.
    $(\boldsymbol{\alpha}, \boldsymbol{\Omega}_1, \mathcal{B}_1, \mathcal{S}_1) \leftarrow$ MSER-ss$(\boldsymbol{X}^\intercal\boldsymbol{X}, \bar{\boldsymbol{P}}_l; \Theta)$ ▷ Fit MSER to residuals.
    $\boldsymbol{\alpha}_l \leftarrow \boldsymbol{\alpha}$ ▷ Store PIPs for lth single effect.
    Initialize $\boldsymbol{\mu}_l$ to a $J \times R$ matrix of zeros ▷ Compute conditional posterior means (78).
    for j in $1,\ldots,J$ **do**
      Store $\sum_{k=1}^{K}\omega_{1jk}\boldsymbol{b}_{1jk}$ in row j of $\boldsymbol{\mu}_l$
    $\bar{\boldsymbol{B}}^{(l)} \leftarrow (\boldsymbol{\alpha}_l\boldsymbol{1}_R^\intercal) \circ \boldsymbol{\mu}_l$ ▷ "$\circ$" denotes elementwise multiplication.
    $\bar{\boldsymbol{P}} \leftarrow \bar{\boldsymbol{P}}_l - \boldsymbol{X}^\intercal\boldsymbol{X}\bar{\boldsymbol{B}}^{(l)}$ ▷ Update expected residuals.
until convergence criterion is satisfied
return $\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_L, \boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_L$

Computing the KL-divergence when L = 1. From (96), the KL-divergence for the mvSuSiE model with $L = 1$ is

$$D_{\mathsf{KL}}(q \,\|\, p) = -\tfrac{N}{2}\log|2\pi\boldsymbol{V}| - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\mathrm{ERSS}) - F_{\mathsf{MSER}}(q, g, \boldsymbol{V}; \boldsymbol{X}, \boldsymbol{Y}), \tag{105}$$

where the ERSS is given in (97). Recall, the optimal variational distribution is equal to the true posterior, $\hat{q}(\boldsymbol{B}^{(1)}) = p(\boldsymbol{B}^{(1)} \mid \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{V}, g)$. And when the optimal variational distribution is attained, the ELBO is equal to the marginal log-likelihood. Therefore, we have

$$\begin{aligned} D_{\mathsf{KL}}(\hat{q} \,\|\, p) &= -\tfrac{N}{2}\log|2\pi\boldsymbol{V}| - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\mathrm{ERSS}) - \log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{V}, g) \\ &= -\tfrac{N}{2}\log|2\pi\boldsymbol{V}| - \tfrac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\mathrm{ERSS}) - \log\textstyle\sum_{j=1}^{J}\mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \sigma_0^2\mathcal{U}, \boldsymbol{\omega}) \\ &\quad - \log p(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{B} = \boldsymbol{0}, \boldsymbol{V}) \\ &= -\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\mathrm{ERSS} - \boldsymbol{Y}^\intercal\boldsymbol{Y})] - \log\textstyle\sum_{j=1}^{J}\mathrm{BF}^{\mathsf{mix}}(\boldsymbol{x}_j, \boldsymbol{Y}, \boldsymbol{V}, \sigma_0^2\mathcal{U}, \boldsymbol{\omega}), \end{aligned} \tag{106}$$

in which the Bayes factor $\mathrm{BF}^{\mathsf{mix}}$ was derived in Proposition 4. Finally, to arrive at the desired KL-divergence $D_{\mathsf{KL}}(\hat{q}_l \,\|\, p_l)$, we substitute $\bar{\boldsymbol{R}}_l$ for $\boldsymbol{Y}$ in (106). A similar approach to computing the ELBO was taken in [96].

The IBSS algorithm for mvSuSiE with sufficient statistics

Algorithm 2 outlines the IBSS algorithm for mvSuSiE with sufficient statistics in which the computations are rearranged so that they only require the sufficient statistics $\boldsymbol{X}^\intercal\boldsymbol{Y}$ and $\boldsymbol{X}^\intercal\boldsymbol{X}$.

Proofs and extended derivations

Proof of Proposition 2. First we give the proof from the maximum-likelihood estimation perspective in which we treat $\boldsymbol{\mu}.$ as a free parameter to be optimized.

The log-likelihood for $\boldsymbol{\mu}$ and $\boldsymbol{b}$ is

$$\log \ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) = -\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \mathbf{1}\boldsymbol{\mu}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \mathbf{1}\boldsymbol{\mu}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)] + \text{const},$$

in which the "const" includes additional terms that do not involve $\boldsymbol{\mu}$ or $\boldsymbol{b}$. Differentiating with respect to $\boldsymbol{\mu}$, setting the partial derivatives to zero, and solving for $\boldsymbol{\mu}$ yields

$$\boldsymbol{V}^{-1}\boldsymbol{\mu} = \boldsymbol{V}^{-1}(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b}),$$

and so

$$\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b}.$$

The profile likelihood for $\boldsymbol{b}$ is therefore

$$\begin{aligned}
\ell^*(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) &:= \max_{\boldsymbol{\mu}} \ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}) \\
&= |2\pi\boldsymbol{V}|^{-n/2} \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \mathbf{1}\hat{\boldsymbol{\mu}}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \mathbf{1}\hat{\boldsymbol{\mu}}^\intercal - \boldsymbol{x}\boldsymbol{b}^\intercal)]\} \\
&= |2\pi\boldsymbol{V}|^{-n/2} \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}((\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^\intercal) - (\boldsymbol{x} - \bar{x}\mathbf{1})\boldsymbol{b}^\intercal)^\intercal((\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^\intercal) - (\boldsymbol{x} - \bar{x}\mathbf{1})\boldsymbol{b}^\intercal)]\} \\
&= \ell(\boldsymbol{b}; \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{Y}}, \boldsymbol{V}).
\end{aligned}$$

The conditional posterior for $\boldsymbol{\mu}$ given $\boldsymbol{b}$ is

$$p(\boldsymbol{\mu} \mid \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}, \boldsymbol{b}) \propto \exp\{-\tfrac{1}{2}\boldsymbol{\mu}_1\boldsymbol{S}_{1\mu}^{-1}\boldsymbol{\mu}_1\},$$

which is the multivariate normal density with mean

$$\boldsymbol{\mu}_1 := N(\boldsymbol{S}_{0\mu}^{-1} + N\boldsymbol{V}^{-1})^{-1}\boldsymbol{V}^{-1}(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b}) = \boldsymbol{S}_{1\mu}\hat{\boldsymbol{S}}^{-1}\hat{\boldsymbol{\mu}}$$

and covariance

$$\boldsymbol{S}_{1\mu} := (\boldsymbol{S}_{0\mu}^{-1} + N\boldsymbol{V}^{-1})^{-1} = (\boldsymbol{S}_{0\mu}^{-1} + \hat{\boldsymbol{S}}^{-1})^{-1}.$$

The marginal likelihood for $\boldsymbol{b}$ obtained by integrating over $\boldsymbol{\mu}$ is

$$\begin{aligned}
\ell^\star(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0a}) &= \textstyle\int \ell(\boldsymbol{\mu}, \boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V})\, p(\boldsymbol{\mu} \mid \boldsymbol{S}_{0\mu})\, d\boldsymbol{\mu} \\
&= \textstyle\int |2\pi\boldsymbol{S}_{0\mu}|^{-1/2}|2\pi\boldsymbol{V}|^{-N/2} \\
&\qquad \times \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal - \mathbf{1}\boldsymbol{\mu}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal - \mathbf{1}\boldsymbol{\mu}^\intercal) + \boldsymbol{S}_{0\mu}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\intercal]\}\, d\boldsymbol{\mu}.
\end{aligned}$$

Expanding terms, we get

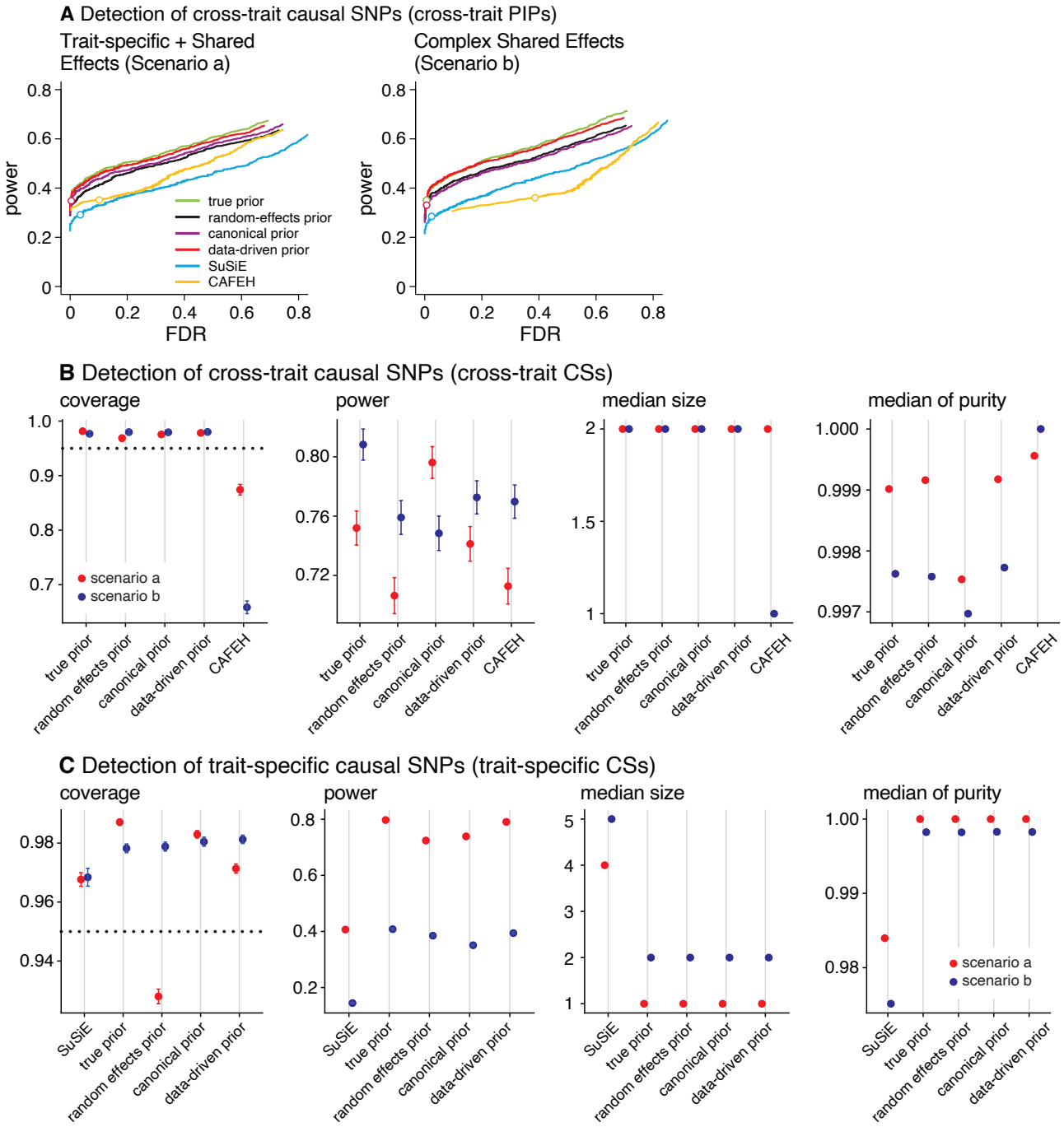
$$\begin{aligned}
&\ell^\star(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}) \\
&\quad = |2\pi\boldsymbol{S}_{0\mu}|^{-1/2}|2\pi\boldsymbol{V}|^{-N/2} \textstyle\int \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{S}_{1\mu}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\intercal - 2\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal\mathbf{1}\boldsymbol{\mu}^\intercal \\
&\qquad\qquad\qquad\qquad + \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)]\}\, d\boldsymbol{\mu} \\
&\quad = |2\pi\boldsymbol{S}_{0\mu}|^{-1/2}|2\pi\boldsymbol{V}|^{-N/2} \textstyle\int \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{S}_{1\mu}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_1)(\boldsymbol{\mu} - \boldsymbol{\mu}_1)^\intercal \\
&\qquad\qquad\qquad\qquad - N\boldsymbol{V}^{-1}(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})^\intercal N\boldsymbol{V}^{-1}\boldsymbol{S}_{1\mu} \\
&\qquad\qquad\qquad\qquad + \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)]\}\, d\boldsymbol{\mu} \\
&\quad = |\boldsymbol{S}_{0\mu}|^{-1/2}|2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{1\mu}|^{1/2} \exp\{-\tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal) \\
&\qquad\qquad\qquad\qquad - N\boldsymbol{V}^{-1}(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})^\intercal N\boldsymbol{V}^{-1}\boldsymbol{S}_{1\mu}]\} \\
&\quad = |2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{0\mu}^{-1}\boldsymbol{S}_{1\mu}|^{1/2} \exp\{\tfrac{1}{2}\boldsymbol{\mu}_1^\intercal\boldsymbol{S}_{1\mu}^{-1}\boldsymbol{\mu}_1 - \tfrac{1}{2}\mathrm{tr}[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)]\}.
\end{aligned}$$

When $\boldsymbol{S}_{0\mu}^{-1} \to \mathbf{0}$, the marginal likelihood simplifies to

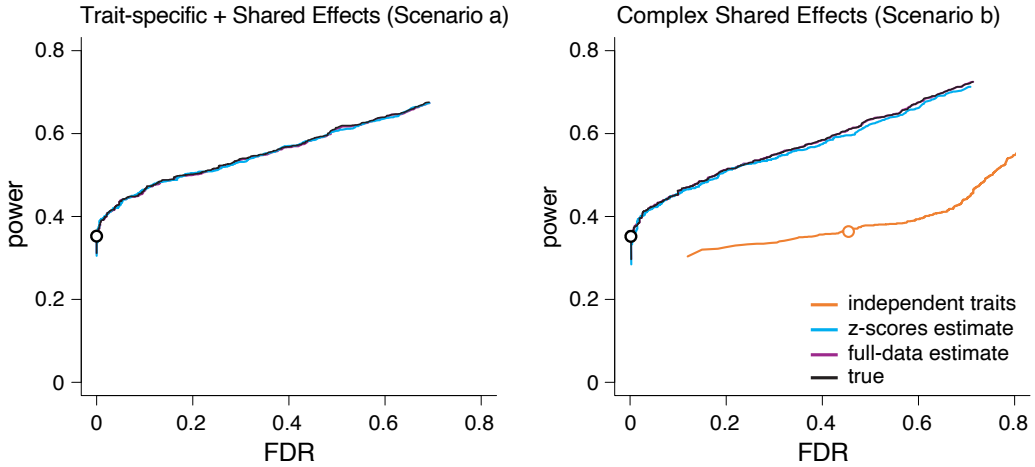
$$\begin{aligned}
&\ell^\star(\boldsymbol{b}; \boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{V}, \boldsymbol{S}_{0\mu}) \\
&\quad = |2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{0\mu}^{-1}\hat{\boldsymbol{S}}_\mu|^{1/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\left[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \boldsymbol{x}\boldsymbol{b}^\intercal) - N\boldsymbol{V}^{-1}(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})(\bar{\boldsymbol{y}} - \bar{x}\boldsymbol{b})^\intercal\right]\} \\
&\quad = |2\pi\boldsymbol{V}|^{-N/2}|\boldsymbol{S}_{0\mu}^{-1}\hat{\boldsymbol{S}}_\mu|^{1/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\left[\boldsymbol{V}^{-1}(\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^\intercal - (\boldsymbol{x} - \bar{x}\mathbf{1})\boldsymbol{b}^\intercal)^\intercal(\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^\intercal - (\boldsymbol{x} - \bar{x}\mathbf{1})\boldsymbol{b}^\intercal)\right]\} \\
&\quad = |\boldsymbol{S}_{0\mu}^{-1}\hat{\boldsymbol{S}}_\mu|^{1/2} \times \ell(\boldsymbol{b}; \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{Y}}, \boldsymbol{V}).
\end{aligned}$$

The marginal likelihood for $\boldsymbol{b}$ in the model with an intercept (and $\boldsymbol{x}, \boldsymbol{Y}$ are not centered) is therefore proportional to the likelihood for the model without an intercept and when $\boldsymbol{x}, \boldsymbol{Y}$ are centered.

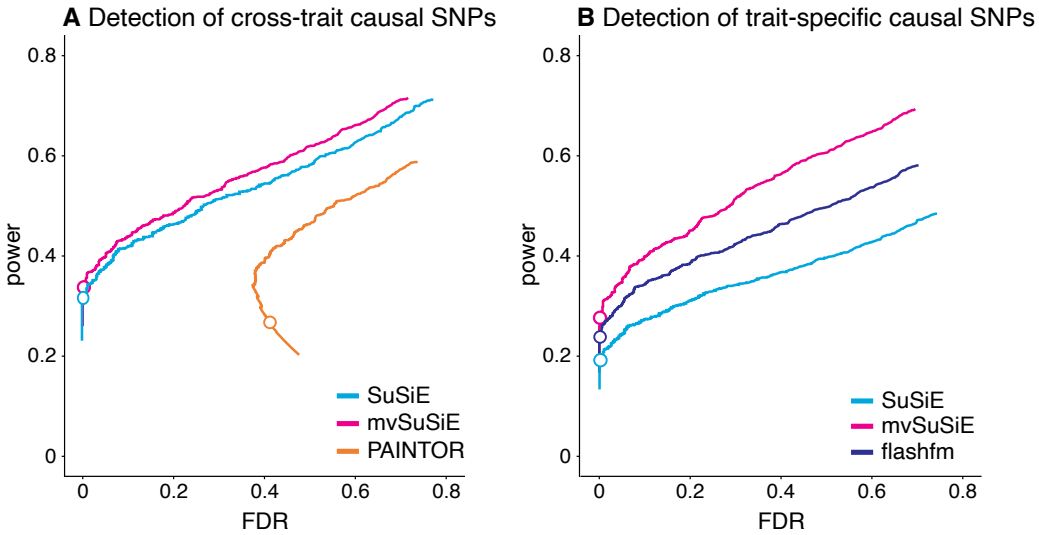
Supplementary Figures



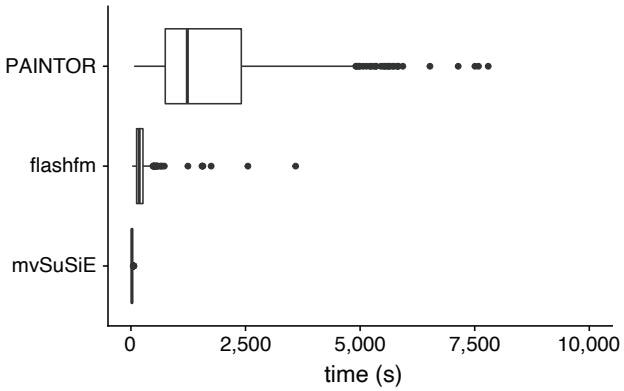
Supplementary Figure 1. Discovery of causal SNPs using PIPs and CSs—comparison of mvSuSiE variants with different priors. This figure is a companion to Fig. 2, giving more detail about the performance of mvSuSiE with different priors. In addition to SuSiE and CAFEH, four variants of mvSuSiE are compared: mvSuSiE with random effects prior; canonical prior; data-driven prior; and mvSuSiE with the true prior, that is the prior used to simulate the true effects. Panel A shows power vs. FDR in identifying causal SNPs using cross-trait measures (PIP or max-PIP). FDR and power are calculated from 600 simulations in each scenario as the measured is varied from 0 to 1. Open circles are drawn at a threshold of 0.95. Panels B and C evaluate CSs using the following metrics: *coverage*, the proportion of CSs containing a true causal SNP; *power*, the proportion of true causal SNPs included in at least one CS; *median size*, the median number of SNPs included in a CS; and *median purity*. Target coverage (95%) is shown as a dotted horizontal line. For C, mvSuSiE detects a trait-specific CS if the average *lfsr* in the CS is less than 0.05. Error bars show 2 times the empirical standard error from the results in all simulations. Note that SuSiE analyzes each trait independently and therefore is not included in Part B. CAFEH does not provide trait-specific CSs and therefore is not included in Part C.



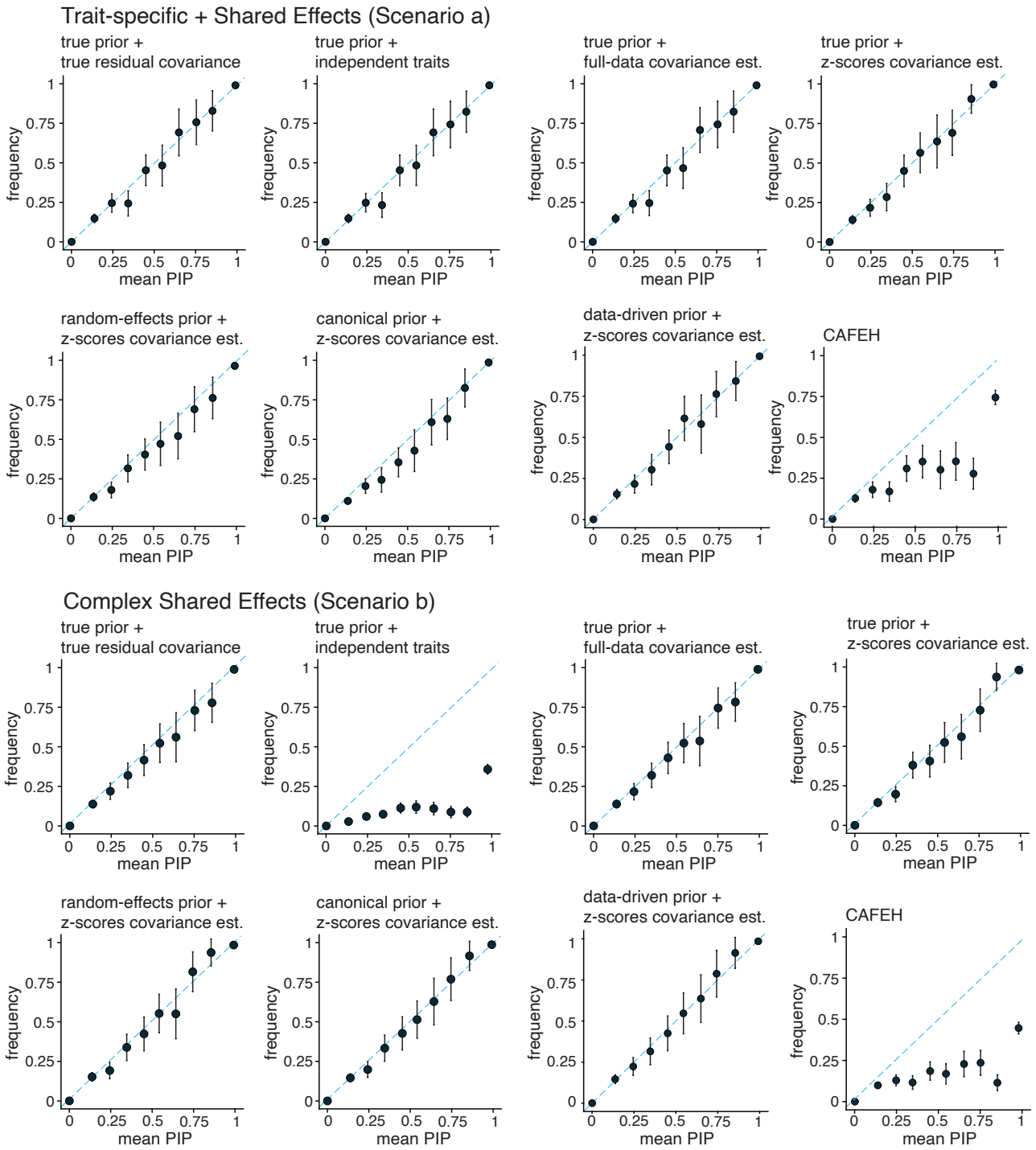
Supplementary Figure 2. Discovery of causal SNPs using cross-trait PIPs—comparison of mvSuSiE variants with different residual covariances. The following choices of $\boldsymbol{V}$ are compared: identity matrix ("independent traits"); $\boldsymbol{V}$ estimated from the individual-level data ("full-data estimate"); $\boldsymbol{V}$ estimated from the *z*-scores ("*z*-scores estimate"); the $\boldsymbol{V}$ used to simulate the data ("true"). Note that in Scenario a, the true residual covariance matrix was the identity matrix, so the results are exactly the same for both "true" and "independent traits." FDR and power were calculated from 600 simulations in each scenario as the cross-trait PIP was varied from 0 to 1. Open circles are drawn at a PIP threshold of 0.95. For these comparisons only, the prior was set to the mixture distribution used to simulate the effects. In Scenario a, in which the traits were simulated with $\boldsymbol{V} = \boldsymbol{I}_R$, there was almost no difference in assuming or not assuming $\boldsymbol{V} = \boldsymbol{I}_R$ in the mvSuSiE analysis; all the power-FDR curves closely overlap in the plot. In the right-hand plot (Scenario b), $\boldsymbol{Y}$ was simulated with $\boldsymbol{V} \neq \boldsymbol{I}_R$, and in that case there was a substantial reduction in performance when $\boldsymbol{V} = \boldsymbol{I}_R$ was assumed in the mvSuSiE analysis ("independent traits").



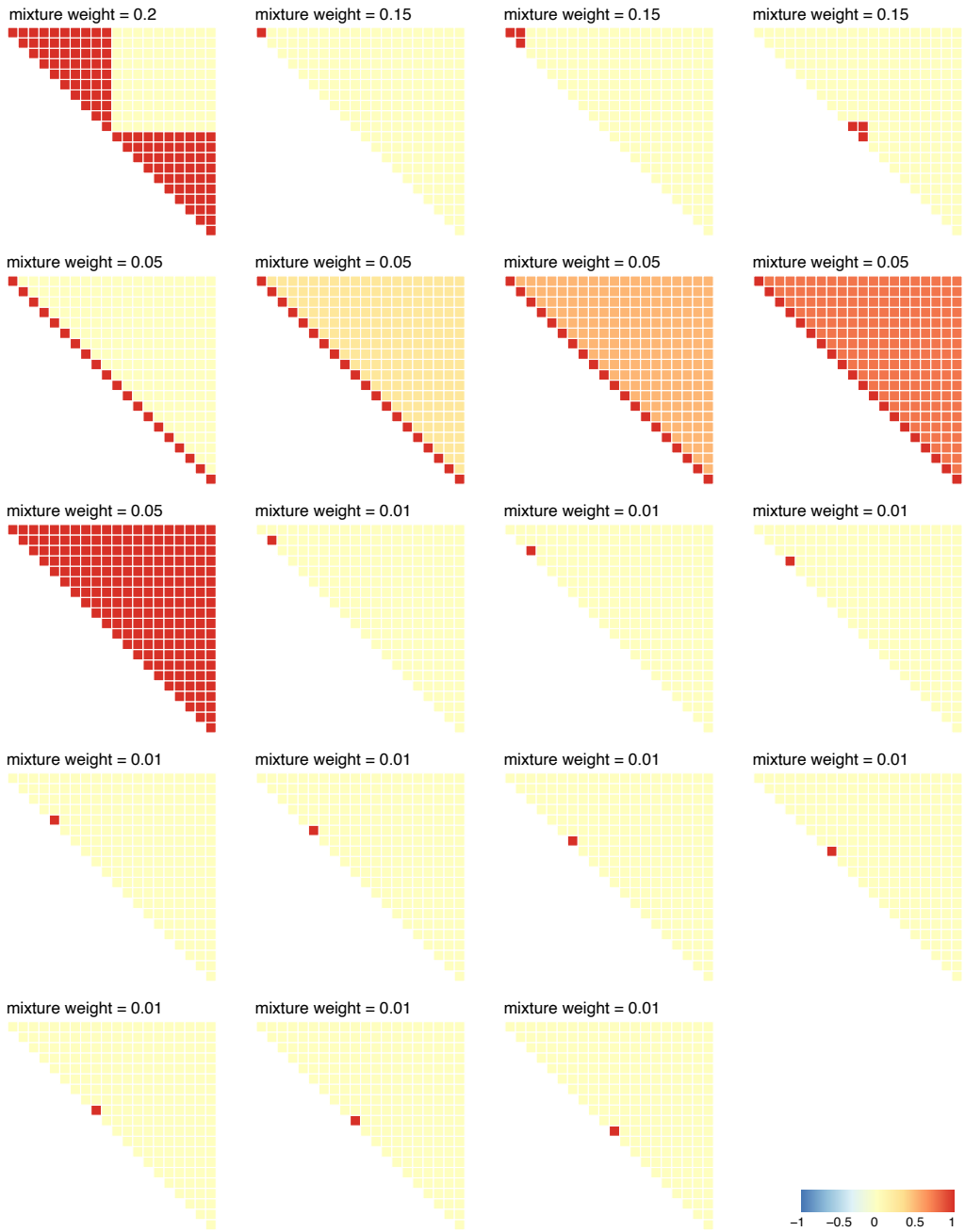
Supplementary Figure 3. Discovery of causal SNPs in 2-trait simulations. Panel A shows power vs. FDR in identifying cross-trait causal SNPs using PIPs (or max-PIP for SuSiE). In A, FDR and power are calculated from the 600 simulations in the 2-trait scenario as the threshold is varied from 0 to 1. Open circles are drawn at a threshold of 0.95. Note that flashfm does not provide a cross-trait measure so it is not included in A. Panel B shows power vs. FDR in identifying trait-specific causal SNPs. In B, FDR and power are calculated from the 600 simulations as the *marginal posterior probability (MPP)* threshold (flashfm), *PIP* (SuSiE), or *min-lfsr* (mvSuSiE) is varied from 0 to 1. Open circles are drawn at a threshold of 0.05. Note that PAINTOR does not provide a trait-specific measure so was not included in B.



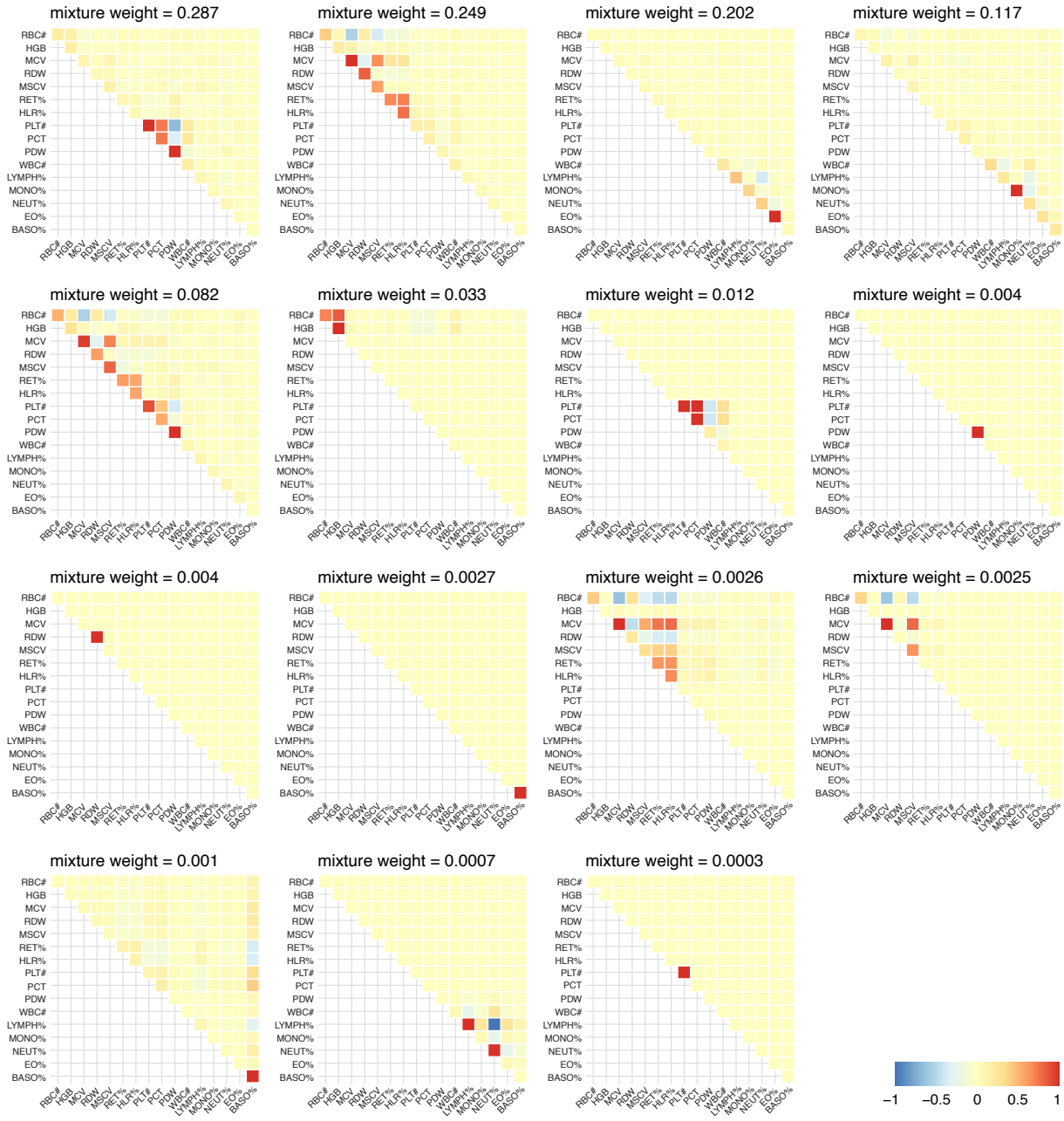
Supplementary Figure 4. Analysis runtimes in 2-trait simulations. The box plots summarize the analysis runtimes for PAINTOR, flashfm and mvSuSiE in the 2-trait simulations. The flashfm runtimes include FINEMAP computations. Note that a single run of PAINTOR takinng longer than 8 h (30,000 s) is not shown in the plot.



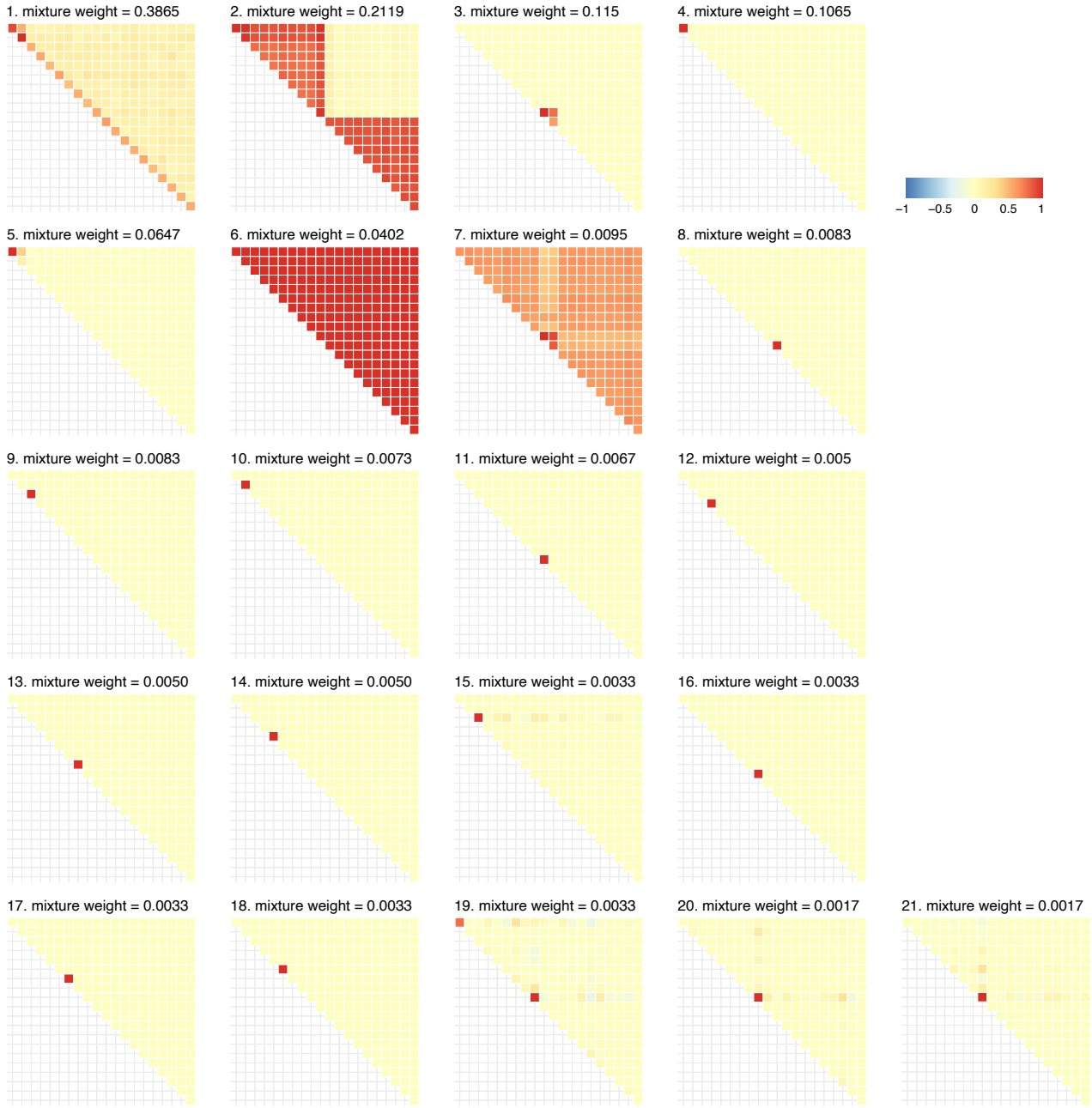
Supplementary Figure 5. Assessment of mvSuSiE and CAFEH PIP calibration. In each scenario, SNPs from all the simulations were grouped into bins according to their reported PIP (10 equally spaced bins from 0 to 1). The plots show the average PIP for each bin (X-axis) against the proportion of SNPs in that bin (Y-axis). Error bars depict 2 times the empirical standard error. A well-calibrated method should produce points near the diagonal. See Supplementary Figures 1 and 2 for details on the mvSuSiE variants compared.



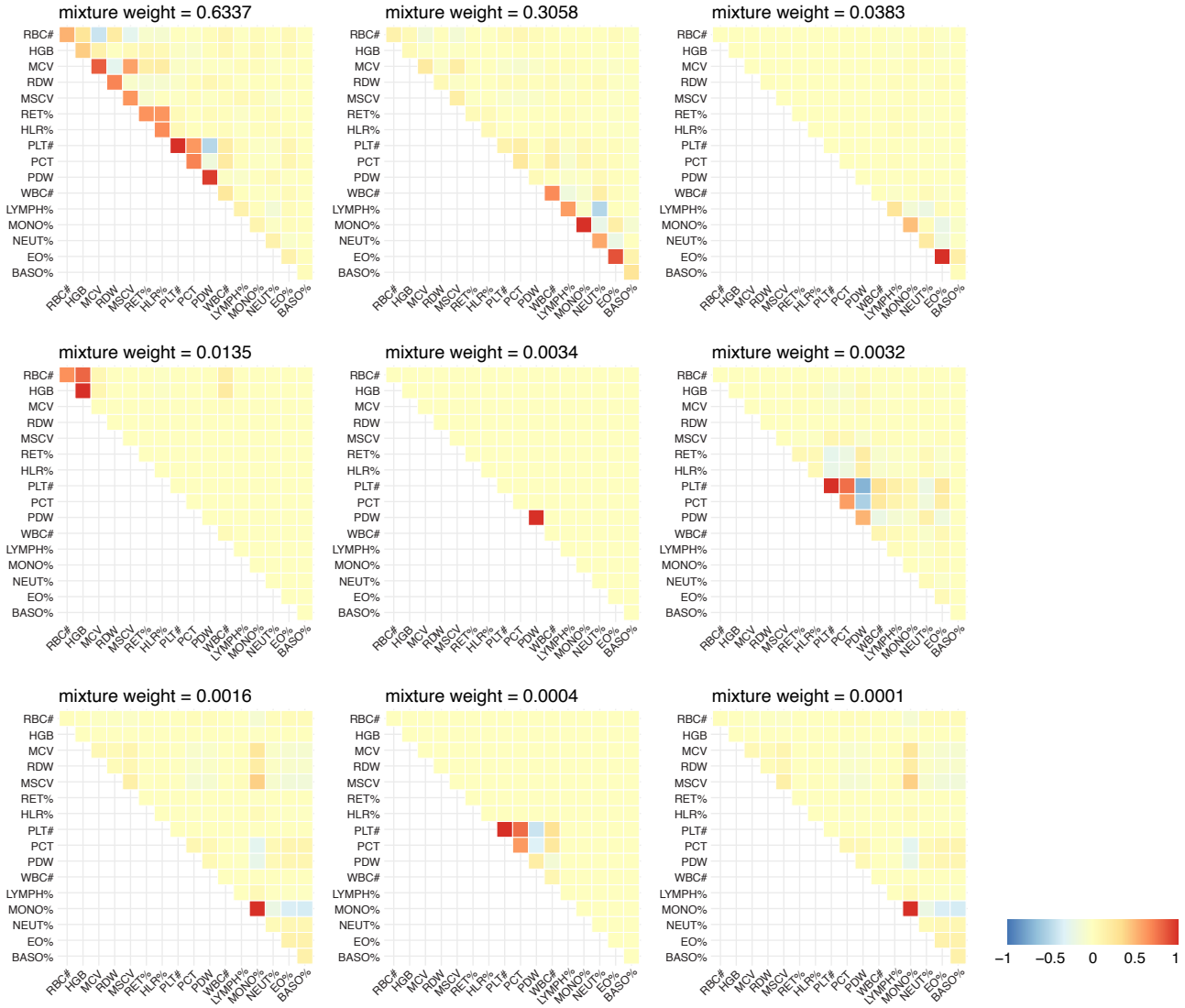
Supplementary Figure 6. Covariance matrices used to simulate the effects of the causal SNPs in Scenario a. Each plot shows a 20 $\times$ 20 covariance matrix, $\boldsymbol{U}_k$, and its corresponding mixture weight, ω_k, in the mixture-of-multivariate normals distribution used to simulate the effects of the causal SNPs. Note all covariance matrices contain elements spanning the range 0 to 1, and none of these matrices contain negative elements.



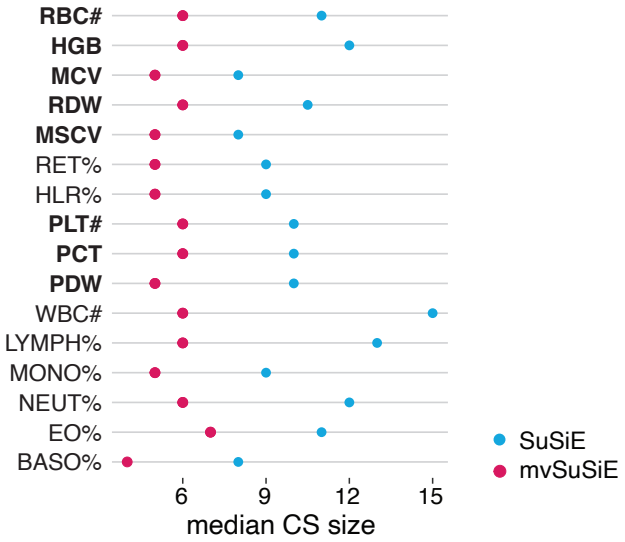
Supplementary Figure 7. Prior on multivariate SNP effects estimated from the UK Biobank blood cell traits. Each plot shows a 16 × 16 scaled covariance matrix $\boldsymbol{U}_k$ and its corresponding estimated mixture weight ω_k. These covariance matrices and mixture weights specify the mixture-of-multivariate normals prior used in the mvSuSiE analyses of the UK Biobank blood cell traits. For visualization purposes only, each plot shows the scaled covariance matrix $\boldsymbol{U}_k/s_k^2$, where s_k^2 is the absolute value of the largest (in magnitude) entry of $\boldsymbol{U}_k$, so that all of the plotted values lie between -1 and 1.



Supplementary Figure 8. Data-driven prior used in Scenario a. Each plot shows a 20 $\times$ 20 scaled covariance matrix $\boldsymbol{U}_k$ and its corresponding estimated mixture weight ω_k. These covariance matrices and mixture weights describe the "data-driven" mixture-of-multivariate normals prior used in the mvSuSiE analyses of the Scenario a simulated data sets. These covariance matrices capture many of the main effect sharing patterns used to simulate the data (compare to Supplementary Fig. 6) including tissue-specific effects (e.g., mixture components 8–18), independent effects (component 1), effects shared in subgroups (e.g., component 2), and effects shared equally across all tissues (component 6). For visualization purposes only, each plot shows the scaled covariance matrix $\boldsymbol{U}_k/s_k^2$, where s_k^2 is the absolute value of the largest (in magnitude) entry of $\boldsymbol{U}_k$, so that all of the plotted values lie between -1 and 1.



Supplementary Figure 9. Data-driven prior used in Scenario b. Each plot shows a 16 × 16 scaled covariance $\boldsymbol{U}_k$ and its corresponding estimated mixture weight ω_k. These covariance matrices and mixture weights describe the "data-driven" mixture-of-multivariate normals prior used in the mvSuSiE analyses of the Scenario b simulated data sets. Compare the covariances and mixture weights shown here to those in Supplementary Fig. 7, which were the covariances and weights used to simulate the data. For visualization purposes only, each plot shows the scaled covariance matrix $\boldsymbol{U}_k/s_k^2$, where s_k^2 is the absolute value of the largest (in magnitude) entry of $\boldsymbol{U}_k$, so that all of the plotted values lie between -1 and 1.



Supplementary Figure 10. Sizes of trait-specific CSs in SuSiE and mvSuSiE fine-mapping of UK Biobank blood cell traits. The plot compares the median sizes of the trait-specific CSs identified by SuSiE and mvSuSiE after removing CSs with purity less than 0.5.