Fine-mapping from summary data with the "Sum of Single Effects" model

Yuxin Zou[1], Peter Carbonetto[2,3], Gao Wang[4*], Matthew Stephens[1,2*]

1 Department of Statistics, University of Chicago, Chicago, IL, USA
2 Department of Human Genetics, University of Chicago, Chicago, IL, USA
3 Research Computing Center, University of Chicago, Chicago, IL, USA
4 Department of Neurology and the Gertrude. H. Sergievsky Center, Columbia University, New York, NY, USA

* wang.gao@columbia.edu
* mstephens@uchicago.edu

Abstract

In recent work, Wang *et al* introduced the "Sum of Single Effects" (*SuSiE*) model, and showed that it provides a simple and efficient approach to fine-mapping genetic variants from individual-level data. Here we present new methods for fitting the *SuSiE* model to summary data, for example to single-SNP z-scores from an association study and linkage disequilibrium (LD) values estimated from a suitable reference panel. To develop these new methods, we first describe a simple, generic strategy for extending any individual-level data method to deal with summary data. The key idea is to replace the usual regression likelihood with an analogous likelihood based on summary data. We show that existing fine-mapping methods such as FINEMAP and CAVIAR also (implicitly) use this strategy, but in different ways, and so this provides a common framework for understanding different methods for fine-mapping. We investigate other common practical issues in fine-mapping with summary data, including problems caused by inconsistencies between the z-scores and LD estimates, and we develop diagnostics to identify these inconsistencies. We also present a new refinement procedure that improves model fits in some data sets, and hence improves overall reliability of the *SuSiE* fine-mapping results. Detailed evaluations of fine-mapping methods in a range of simulated data sets show that *SuSiE* applied to summary data is competitive, in both speed and accuracy, with the best available fine-mapping methods for summary data.

Author summary

The goal of fine-mapping is to identify the genetic variants that causally affect some trait of interest. Fine-mapping is challenging because the genetic variants can be highly correlated, due to a phenomenon called linkage disequilibrium (LD). The most successful current approaches to fine-mapping frame the problem as a *variable selection problem*, and here we focus on one such approach based on the "Sum of Single Effects" (*SuSiE*) model. The main contribution of this paper is to extend *SuSiE* to work with summary data, which is often accessible when the full genotype and phenotype data are not. In the process of extending *SuSiE*, we also developed a new mathematical framework that helps to explain existing fine-mapping methods for summary data, why they work well (or not), and under what circumstances. In simulations, we show that

SuSiE applied to summary data is competitive with the best available fine-mapping methods for summary data. We also show how different factors such as accuracy of the LD estimates can affect the quality of the fine-mapping.

Introduction

Fine-mapping is the process of narrowing down genetic association signals to a small number of potential causal variants [1–4], and it plays an important part in the effort to understand the genetic causes of diseases [5,6]. However, fine-mapping is a difficult problem due to the strong and complex correlation patterns ("linkage disequilibrium", or LD) that exist among nearby genetic variants. Many different methods and algorithms have been developed to tackle the fine-mapping problem [2,7–19]. In recent work, Wang *et al* [17] introduced a new approach to fine-mapping, *SuSiE* (short for "SUm of SIngle Effects"), which has several advantages over existing approaches: it is more computationally scalable; and it provides a new, simple way to calculate "credible sets" of putative causal variants [2,20]. However, the algorithms in [17] also have an important limitation: they require individual-level genotype and phenotype data. In contrast, many other fine-mapping methods require access only to summary data, such as z-scores from single-SNP association analyses and an estimate of LD patterns from a suitable reference panel [7,8,11–13,15,16,21]. Requiring only summary data is useful because individual-level data are often difficult to obtain, both for practical reasons, such as the need to obtain many data sets collected by many different researchers, and for reasons to do with consent and privacy. By comparison, summary data are much easier to obtain, and many publications share such summary data [22].

In this paper, we introduce new variants of *SuSiE* for performing fine-mapping from summary data; we call these variants *SuSiE-RSS* (RSS stands for "regression with summary statistics" [23].) Our work exploits the facts that (i) the multiple regression likelihood can be written in terms of a particular type of summary data, known as *sufficient statistics* (explained below), and (ii) these sufficient statistics can be approximated from the types of summary data that are commonly available (e.g., z-scores from single-SNP association tests and LD estimates from suitable reference panel). In the special case where the sufficient statistics themselves are available the second approximation step is unnecessary and *SuSiE-RSS* yields the same results as *SuSiE* applied to the original individual-level data; otherwise it yields an approximation. By extending *SuSiE* to deal with widely-available summary statistics, *SuSiE-RSS* greatly expands the potential applicability of the *SuSiE* fine-mapping approach.

Although our main goal here is to extend *SuSiE* to deal with summary data, the approach we use, and the connections it exploits, are quite general, and could be applied to extend other individual-level data methods to deal with summary data. This general approach has two nice features. First it deals simply, and automatically, with non-invertible LD matrices, which arise frequently in fine-mapping. We argue, both through theory and example, that it provides a simpler and more effective solution to this issue than some existing approaches. Second, it shows how individual-level results can be obtained as a special case of summary-data analysis, by using the sufficient statistics as summary data.

By highlighting the close connection between the likelihoods for individual-level and summary data, our work generalizes results of [11], who showed a strong connection between Bayes Factors, based on specific priors, from individual-level data and summary data. Our results highlight that this connection is fundamentally due to a close connection between the likelihoods, and so will apply whatever prior is used (and indeed, to non-Bayesian approaches that do not use a prior). By focussing on likelihoods, our analysis also helps clarify the differences and connections between

existing fine-mapping methods such as FINEMAP v1.1 [12], FINEMAP v1.2 [21] and CAVIAR [7], which can differ in both the prior and likelihood used.

Finally, we introduce several other methodological innovations for fine-mapping. Some of these innovations are not specific to *SuSiE* and could be used to adapt other statistical methods for individual-level data to deal with summary data. We describe methods for identifying "allele flips"—alleles that are (erroneously) encoded differently in the study and reference data—and other inconsistencies in the summary data. (See also [24] for related ideas.) We illustrate how a single allele flip can lead to inaccurate fine-mapping results, emphasizing the importance of careful quality control when performing fine-mapping using summary data. We also introduce a new refinement procedure for *SuSiE* that sometimes improves estimates from the original fitting procedure.

Description of the method

We begin with some background and notation. Let $\boldsymbol{y} \in \mathbb{R}^N$ denote the phenotypes of N individuals in a genetic association study, and let $\boldsymbol{X} \in \mathbb{R}^{N \times J}$ denote their corresponding genotypes at J genetic variants (SNPs). To simplify the presentation, we assume the $\boldsymbol{y}$ are quantitative and approximately normally distributed, and that both $\boldsymbol{y}$ and the columns of $\boldsymbol{X}$ are centered to have mean zero, which avoids the need for an intercept term in (1) below [25]. We elaborate on treatment of binary and case-control phenotypes in the Discussion below.

Fine-mapping from individual-level data is usually performed by fitting the multiple linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}, \tag{1}$$

where $\boldsymbol{b} = (b_1, \ldots, b_J)^\intercal$ is a vector of multiple regression coefficients, $\boldsymbol{e}$ is an N-vector of error terms distributed as $\boldsymbol{e} \sim \mathcal{N}_N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_N)$, with (typically unknown) residual variance $\sigma^2 > 0$, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $\mathcal{N}_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the r-variate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$.

In this multiple regression framework, the question of which SNPs are affecting $\boldsymbol{y}$ becomes a problem of "variable selection"; that is, identifying which elements of $\boldsymbol{b}$ are not zero. While many methods exist for variable selection in multiple regression, fine-mapping has some special features—typically very high correlations among some columns of $\boldsymbol{X}$ and very sparse $\boldsymbol{b}$—that make Bayesian methods with sparse priors a preferred approach (e.g., [7–9]). These methods specify a sparse prior for $\boldsymbol{b}$, and perform inference by approximating the posterior distribution $p(\boldsymbol{b} \mid \boldsymbol{X}, \boldsymbol{y})$. In particular, the evidence for SNP j having a non-zero effect is often summarized by the "posterior inclusion probability" (PIP),

$$\mathrm{PIP}_j := \Pr(b_j \neq 0 \mid \boldsymbol{X}, \boldsymbol{y}). \tag{2}$$

The Sum of Single Effects (*SuSiE*) model

The key idea behind *SuSiE* [17] is to write $\boldsymbol{b}$ as a sum,

$$\boldsymbol{b} = \sum_{l=1}^{L} \boldsymbol{b}_l, \tag{3}$$

in which each vector $\boldsymbol{b}_l = (b_{l1}, \ldots, b_{lJ})^\intercal$ is a "single effect" vector; that is, a vector with exactly one non-zero element. The representation (3) allows that $\boldsymbol{b}$ has at most L non-zero elements, where L is a user-specified upper bound on the number of effects. (If

single-effect vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ have non-zero element at the same SNP j, $\boldsymbol{b}$ will have fewer than L non-zeros.)

The special case $L = 1$ corresponds to the assumption that a region has exactly one causal SNP (which we use as shorthand for a SNP with a non-zero effect); in [17] this special case is called the "single effect regression" (SER) model. The SER is particularly convenient because posterior computations are analytically tractable [9]; consequently, despite its clear limitations, the SER has been widely used [2, 26–28].

For $L > 1$, Wang *et al* [17] introduced a simple model fitting algorithm, which they called Iterative Bayesian Stepwise Selection (IBSS). In brief, IBSS iterates through the single-effect vectors $l = 1, \ldots, L$, at each iteration fitting $\boldsymbol{b}_l$ while keeping the other single-effect vectors fixed. By construction, each step thus involves fitting an SER, which, as noted above, is straightforward. Wang *et al* [17] show that IBSS can be understood as computing an approximate posterior distribution $p(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L \mid \boldsymbol{X}, \boldsymbol{y}, \sigma^2)$, and that the algorithm iteratively optimizes an objective function known as the "evidence lower bound" (ELBO).

Summary data for fine-mapping

Motivated by the difficulties in accessing the individual-level data $\boldsymbol{X}, \boldsymbol{y}$ from most studies, researchers have developed fine-mapping approaches that work with more widely available "summary data". Here we develop methods that use various combinations of the summary data.

(i) Vectors $\hat{\boldsymbol{b}} = (\hat{b}_1, \ldots, \hat{b}_J)^\intercal$ and $\hat{\boldsymbol{s}} = (\hat{s}_1, \ldots, \hat{s}_J)^\intercal$ containing estimates of marginal association for each SNP j, and corresponding standard errors, from a simple linear regression:

$$\hat{b}_j := \boldsymbol{x}_j^\intercal \boldsymbol{y} / (\boldsymbol{x}_j^\intercal \boldsymbol{x}_j), \tag{4}$$

$$\hat{s}_j := [(\boldsymbol{y} - \boldsymbol{x}_j \hat{b}_j)^\intercal (\boldsymbol{y} - \boldsymbol{x}_j \hat{b}_j) / (N \boldsymbol{x}_j^\intercal \boldsymbol{x}_j)]^{1/2}. \tag{5}$$

An alternative to $\hat{\boldsymbol{b}}, \hat{\boldsymbol{s}}$ is the vector $\hat{\boldsymbol{z}} = (\hat{z}_1, \ldots, \hat{z}_J)^\intercal$ of z-scores:

$$\hat{z}_j := \hat{b}_j / \hat{s}_j. \tag{6}$$

It is important that all $\hat{\boldsymbol{b}}, \hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{z}}$ be computed from the same N samples. Many studies provide $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{s}}$ (see [22] for examples), and many more provide the z-scores, or data that can be used to compute the z-scores (e.g., $\hat{z}_j$, can be recovered from the p-value and the sign of $\hat{b}_j$ [29]).

(ii) An estimate, $\hat{\boldsymbol{R}}$, of the in-sample LD matrix, $\boldsymbol{R}$. To be precise, $\boldsymbol{R}$ is the $J \times J$ SNP-by-SNP sample correlation matrix,

$$\boldsymbol{R} := \boldsymbol{D}_{xx}^{-1/2} \boldsymbol{X}^\intercal \boldsymbol{X} \boldsymbol{D}_{xx}^{-1/2} \tag{7}$$

where $\boldsymbol{D}_{xx} := \mathrm{diag}(\boldsymbol{X}^\intercal \boldsymbol{X})$ is a diagonal matrix that ensures the diagonal entries of $\boldsymbol{R}$ are all 1. Usually the estimate $\hat{\boldsymbol{R}}$ is taken to be an "out-of-sample" LD matrix—that is, the sample correlation matrix of the same J SNPs in a suitable reference panel, chosen to be genetically similar to the study population, possibly with additional shrinkage or banding steps to improve accuracy [14].

(iii) Optionally, the sample size N and the sample variance of $\boldsymbol{y}$. (Since $\boldsymbol{y}$ is centered, the sample variance of $\boldsymbol{y}$ is simply $v_y := \boldsymbol{y}^\intercal \boldsymbol{y} / N$). Knowing these quantities is obviously equivalent to knowing $\boldsymbol{y}^\intercal \boldsymbol{y}$ and N, so for brevity we will use the latter. These quantities are not required, but they can be helpful as we will see later.

Table 1. Summary of *SuSiE* and *SuSiE-RSS*, the different data they accept, and the corresponding likelihoods. In the "likelihood" column, $\tilde{\boldsymbol{z}} := \boldsymbol{D}_z^{1/2}\hat{\boldsymbol{z}}$ is the vector of adjusted z-scores; see eq. 17). In this summary we assume $\boldsymbol{X}, \boldsymbol{y}$ are standardized, which is common practice in genetic association studies. Note that when *SuSiE-RSS* is applied to sufficient statistics and σ^2 is estimated (second row), the likelihood is identical to the likelihood for *SuSiE* applied the individual-level data (first row). See https://stephenslab.github.io/susieR/articles/susie_rss.html for an illustration of how these methods are called in the R package, `susieR`.

method	data type	data	σ^2	likelihood	algorithm
SuSiE	individual	$\boldsymbol{X}, \boldsymbol{y}$	fit	$\ell(\boldsymbol{b}, \sigma^2) = \ell(\boldsymbol{b}, \sigma^2; \boldsymbol{X}, \boldsymbol{y})$	IBSS
SuSiE-RSS	sufficient	$\boldsymbol{R}, \hat{\boldsymbol{z}}, N$	fit	$\ell(\boldsymbol{b}, \sigma^2) = \ell_{ss}(\boldsymbol{b}, \sigma^2; \boldsymbol{R}, \tilde{\boldsymbol{z}}/\sqrt{N}, 1, N)$	IBSS-ss
SuSiE-RSS	summary	$\hat{\boldsymbol{R}}, \hat{\boldsymbol{z}}, N$	1	$\ell(\boldsymbol{b}) = \ell_{ss}(\boldsymbol{b}, 1; \hat{\boldsymbol{R}}, \tilde{\boldsymbol{z}}/\sqrt{N}, 1, N)$	IBSS-ss

We caution that if the summary statistics come from a meta-analysis, the summary statistics should be computed carefully to avoid the pitfalls highlighted in [24]. Importantly, SNPs that are not analyzed in all the individual studies in the meta-analysis should not be used in the fine-mapping.

SuSiE with summary data

A key question—and the question central to this paper—is, how do we use summary data to estimate the coefficients $\boldsymbol{b}$ in a multiple linear regression (1)? And, more specifically, how do we use them to estimate the single-effect vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L$ in *SuSiE* (3)? Here, we tackle these questions in two steps. First, we consider a special type of summary data, called "sufficient statistics," which contain the same information about the model parameters as the individual-level data $\boldsymbol{X}, \boldsymbol{y}$. Given such sufficient statistics, we develop an algorithm that *exactly reproduces* the results that would be obtained by running *SuSiE* on the original data $\boldsymbol{X}, \boldsymbol{y}$. Second, we consider the case where we have access to summary data that are not sufficient statistics; these summary data can be used to approximate the sufficient statistics, and hence approximate the results for individual-level data.

The IBSS-ss algorithm

The IBSS algorithm of [17] fits the *SuSiE* model to individual-level data $\boldsymbol{X}, \boldsymbol{y}$. The data enter the *SuSiE* model only through the likelihood, which from (1) is

$$\ell(\boldsymbol{b}, \sigma^2; \boldsymbol{X}, \boldsymbol{y}) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}^\intercal\boldsymbol{y} - 2\boldsymbol{b}^\intercal\boldsymbol{X}^\intercal\boldsymbol{y} + \boldsymbol{b}^\intercal\boldsymbol{X}^\intercal\boldsymbol{X}\boldsymbol{b})\right\}. \tag{8}$$

This likelihood depends on the data only through $\boldsymbol{X}^\intercal\boldsymbol{X}, \boldsymbol{X}^\intercal\boldsymbol{y}, \boldsymbol{y}^\intercal\boldsymbol{y}$ and N, and so these quantities are *sufficient statistics.* (Furthermore, these sufficient statistics can be computed from other combinations of summary data, which are therefore also sufficient statistics; we discuss on this point further below.) Careful inspection of the IBSS algorithm in [17] confirms that it depends on the data only through these sufficient statistics. Thus, by rearranging the computations we obtain a variant of IBSS, called "IBSS-ss", that can fit the *SuSiE* model from sufficient statistics; see S1 Text.

We use IBSS$(\boldsymbol{X}, \boldsymbol{y})$ to denote the result of applying the IBSS algorithm to the individual-level data, and IBSS-ss$(\boldsymbol{X}^\intercal\boldsymbol{X}, \boldsymbol{X}^\intercal\boldsymbol{y}, \boldsymbol{y}^\intercal\boldsymbol{y}, N)$ to denote the results of

applying the IBSS-ss algorithm to the sufficient statistics. These two algorithms will give the same result,

$$\text{IBSS-ss}(\boldsymbol{X}^\intercal \boldsymbol{X}, \boldsymbol{X}^\intercal \boldsymbol{y}, \boldsymbol{y}^\intercal \boldsymbol{y}, N) = \text{IBSS}(\boldsymbol{X}, \boldsymbol{y}), \tag{9}$$

however the computational complexity of the two approaches is different. First, computing the sufficient statistics requires computing the $J \times J$ matrix $\boldsymbol{X}^\intercal \boldsymbol{X}$, which is a non-trivial computation, requiring $O(NJ^2)$ operations. However, once this matrix has been computed, IBSS-ss requires $O(J^2)$ operations per iteration, whereas IBSS requires $O(NJ)$ operations per iteration. (The number of iterations should be the same.) Therefore, when $N \gg J$, which is often the case in fine-mapping studies, IBSS-ss will usually be faster. In practice, choosing between these workflows also depend on whether or not one prefers to precompute $\boldsymbol{X}^\intercal \boldsymbol{X}$, which can be done conveniently in programs such as PLINK [30] or LDstore [31]).

SuSiE with summary data: *SuSiE-RSS*

In practice, sufficient statistics may not be available; in particular, when individual-level data are unavailable, the matrix $\boldsymbol{X}^\intercal \boldsymbol{X}$ is also usually unavailable. A natural approach to deal with this issue is to approximate the sufficient statistics, then to proceed as if the sufficient statistics were available by inputting the approximate sufficient statistics to the IBSS-ss algorithm. We call this approach *SuSiE-RSS*.

For example, let $\hat{\boldsymbol{V}}_{xx}$ denote an approximation to the sample covariance $\boldsymbol{V}_{xx} = \frac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{X}$, and assume the other sufficient statistics $\boldsymbol{X}^\intercal \boldsymbol{y}, \boldsymbol{y}^\intercal \boldsymbol{y}, N$ are available exactly. (These are easily obtained from commonly available summary data, and $\hat{\boldsymbol{R}}$; see S1 Text.) Then *SuSiE-RSS* is the result of running the IBSS-ss algorithm on the sufficient statistics but with $N\hat{\boldsymbol{V}}_{xx}$ replacing $\boldsymbol{X}^\intercal \boldsymbol{X}$; that is, *SuSiE-RSS* is $\text{IBSS-ss}(N\hat{\boldsymbol{V}}_{xx}, \boldsymbol{X}^\intercal \boldsymbol{y}, \boldsymbol{y}^\intercal \boldsymbol{y}, N)$.

In practice, we found that estimating σ^2 sometimes produced very inaccurate estimates, presumably due to inaccuracies in $\hat{\boldsymbol{V}}_{xx}$ as an approximation to $\boldsymbol{V}_{xx}$. (This problem did not arise when $\hat{\boldsymbol{V}}_{xx} = \boldsymbol{V}_{xx}$.) Therefore, when running the IBSS-ss algorithm on approximate summary statistics, we recommend to fix the residual variance, $\sigma^2 = \boldsymbol{y}^\intercal \boldsymbol{y}/N$, rather than estimate it.

Interpretation in terms of an approximation to the likelihood

We defined *SuSiE-RSS* as the application the IBSS-ss algorithm to the sufficient statistics, or approximations to these statistics. Conceptually, this approach combines the *SuSiE* prior with an approximation to the likelihood (8).

To formalize this, we write the likelihood (8) explicitly as a function of the sufficient statistics, defining

$$\ell_{\text{ss}}(\boldsymbol{b}, \sigma^2; \boldsymbol{V}_{xx}, \boldsymbol{v}_{xy}, v_{yy}, N) := (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{N}{2\sigma^2}(v_{yy} - 2\boldsymbol{b}^\intercal \boldsymbol{v}_{xy} + \boldsymbol{b}^\intercal \boldsymbol{V}_{xx}\boldsymbol{b})\right\}, \tag{10}$$

so that

$$\ell(\boldsymbol{b}, \sigma^2; \boldsymbol{X}, \boldsymbol{y}) = \ell_{\text{ss}}(\boldsymbol{b}, \sigma^2; \boldsymbol{V}_{xx}, \tfrac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{y}, \tfrac{1}{N}\boldsymbol{y}^\intercal \boldsymbol{y}, N). \tag{11}$$

Replacing $\boldsymbol{V}_{xx}$ with an estimate $\hat{\boldsymbol{V}}_{xx}$ is therefore equivalent to replacing the likelihood (11) with

$$\ell_{\text{RSS}}(\boldsymbol{b}, \sigma^2) := \ell_{\text{ss}}(\boldsymbol{b}, \sigma^2; \hat{\boldsymbol{V}}_{xx}, \tfrac{1}{N}\boldsymbol{X}^\intercal \boldsymbol{y}, \tfrac{1}{N}\boldsymbol{y}^\intercal \boldsymbol{y}, N). \tag{12}$$

Note that when $\hat{\boldsymbol{V}}_{xx} = \boldsymbol{V}_{xx}$, the approximation is exact; that is, $\ell_{\text{RSS}}(\boldsymbol{b}, \sigma^2) = \ell(\boldsymbol{b}, \sigma^2; \boldsymbol{X}, \boldsymbol{y})$. Thus applying *SuSiE-RSS* with $\boldsymbol{V}_{xx}$ is equivalent to using

the individual-data likelihood (8), and applying it with $\hat{\boldsymbol{V}}_{xx}$ is equivalent to using the approximate likelihood (12). Finally, fixing $\sigma^2 = \frac{1}{N}\boldsymbol{y}^\intercal\boldsymbol{y}$ is equivalent to using the following likelihood:

$$\begin{aligned}\ell_{\mathrm{RSS}}(\boldsymbol{b}) &:= \ell_{\mathrm{RSS}}(\boldsymbol{b}, \tfrac{1}{N}\boldsymbol{y}^\intercal\boldsymbol{y}) \\ &= \ell_{\mathrm{ss}}(\boldsymbol{b}, \tfrac{1}{N}\boldsymbol{y}^\intercal\boldsymbol{y}; \hat{\boldsymbol{V}}_{xx}, \tfrac{1}{N}\boldsymbol{X}^\intercal\boldsymbol{y}, \tfrac{1}{N}\boldsymbol{y}^\intercal\boldsymbol{y}, N). \end{aligned} \tag{13}$$

General strategy for applying regression methods to summary data

The strategy used here to extend *SuSiE* to summary data is quite general, and could be used to extend essentially *any* likelihood-based multiple regression method for individual-level data $\boldsymbol{X}, \boldsymbol{y}$ to summary data. Operationally, this strategy involves two steps: (i) implement an algorithm that accepts as input sufficient statistics, and outputs the same result as the individual-level data; (ii) apply this algorithm to approximations of the sufficient statistics computed from (non-sufficient) summary data (optionally, fixing the residual variance to $\sigma^2 = \boldsymbol{y}^\intercal\boldsymbol{y}/N$). Conceptually, this strategy involves replacing the exact likelihood (18) with an approximate likelihood, either (12) or (13).

Special case when $\boldsymbol{X}, \boldsymbol{y}$ are standardized

In genetic association studies, is common practice to standardize both $\boldsymbol{y}$ and the columns of $\boldsymbol{X}$ to have unit variance—that is, $\boldsymbol{y}^\intercal\boldsymbol{y} = N$ and $\boldsymbol{x}_j^\intercal\boldsymbol{x}_j = N$ for all $j = 1, \ldots, J$—before fitting the model (1). (See [32, 33] for a discussion on the choice to standardize.) Standardizing $\boldsymbol{X}, \boldsymbol{y}$ is commonly done in genetic association analysis and fine-mapping, and results in some simplifications that facilitates connections with existing methods, so we consider this special case in detail.

When $\boldsymbol{X}, \boldsymbol{y}$ are standardized, the sufficent statistics are easily computed from the in-sample LD matrix $\boldsymbol{R}$, the single-SNP z-scores $\hat{\boldsymbol{z}}$, and the sample size N:

$$\boldsymbol{X}^\intercal\boldsymbol{X} = N\boldsymbol{R} \tag{14}$$

$$\boldsymbol{X}^\intercal\boldsymbol{y} = \sqrt{N}\tilde{\boldsymbol{z}} \tag{15}$$

$$\boldsymbol{y}^\intercal\boldsymbol{y} = N, \tag{16}$$

where we define

$$\tilde{\boldsymbol{z}} := \boldsymbol{D}_z^{1/2}\hat{\boldsymbol{z}}, \tag{17}$$

and we define $\boldsymbol{D}_z$ to be the diagonal matrix in which the jth diagonal element is $N/(N + \hat{z}_j^2)$ [21]. Note the elements of $\boldsymbol{D}_z$ have the interpretation as being one minus the estimated PVE ("Proportion of phenotypic Variance Explained"), so we refer to $\tilde{\boldsymbol{z}}$ as the vector of the "PVE-adjusted z-scores." If all the effects are small, the estimated PVEs will be close to zero, the diagonal of $\boldsymbol{D}_z$ will be close to one, and so $\tilde{\boldsymbol{z}} \approx \hat{\boldsymbol{z}}$.

Substituting equations (14)–(16) into (11) gives

$$\ell(\boldsymbol{b}, \sigma^2; \boldsymbol{X}, \boldsymbol{y}) = \ell_{ss}(\boldsymbol{b}, \sigma^2; \boldsymbol{R}, \tilde{\boldsymbol{z}}/\sqrt{N}, 1, N). \tag{18}$$

When the in-sample LD matrix $\boldsymbol{R}$ is not available, and is replaced with $\hat{\boldsymbol{R}} \approx \boldsymbol{R}$, the *SuSiE-RSS* likelihood (13) becomes

$$\ell_{\mathrm{RSS}}(\boldsymbol{b}) = \ell_{ss}(\boldsymbol{b}, 1; \hat{\boldsymbol{R}}, \tilde{\boldsymbol{z}}/\sqrt{N}, 1, N). \tag{19}$$

These expressions are summarized in Table 1.

Connections with previous work

The approach we take here is most closely connected with the approach used in FINEMAP (versions 1.2 and later) [21]. In essence, FINEMAP 1.2 uses the same likelihoods (18, 19) as we use here, but the derivations in [21] do not clearly distinguish the case where the in-sample LD matrix is available from the case where it is not. In addition, the derivations in [21] focus on Bayes Factors computed with particular priors, rather than focussing on the likelihood. Our derivations here emphasize that, when the in-sample LD matrix is available, results from "summary data" can be identical to those that would have been obtained from individual-level data. And our focus on the likelihoods draws attention the generality of this strategy; it is not specific to a particular prior, nor is it specific to the use of Bayesian methods.

Several other previous fine-mapping methods (e.g., [7, 8, 12, 16]) are based on the following model:

$$\hat{\boldsymbol{z}} \mid \boldsymbol{z}, \hat{\boldsymbol{R}} \sim \mathcal{N}_J(\hat{\boldsymbol{R}}\boldsymbol{z}, \hat{\boldsymbol{R}}), \tag{20}$$

where $\boldsymbol{z} = (z_1, \ldots, z_J)^\intercal$ is an unobserved vector of scaled effects, sometimes called the noncentrality parameters (NCPs),

$$\boldsymbol{z} := \frac{\boldsymbol{b}\sqrt{N}}{\sigma}. \tag{21}$$

(Earlier versions of *SuSiE-RSS* [34] were also based on this model.) To connect our method with this approach, note that, when $\hat{\boldsymbol{R}}$ is invertible, the likelihood (19) is equivalent to the likelihood for $\boldsymbol{b}$ in the following model:

$$\tilde{\boldsymbol{z}} \mid \boldsymbol{b}, \hat{\boldsymbol{R}} \sim \mathcal{N}_J(\sqrt{N}\hat{\boldsymbol{R}}\boldsymbol{b}, \hat{\boldsymbol{R}}). \tag{22}$$

(See S1 Text for additional notes.) This model is also used in Zhu and Stephens [23], where the derivation was based on the PVE-adjusted standard errors, which gives the same PVE-adjusted z-scores. This model (22) is essentially the same model as (20) but with the observed z-scores, $\hat{\boldsymbol{z}}$, replaced with the PVE-adjusted z-scores, $\tilde{\boldsymbol{z}}$. In other words, when $\hat{\boldsymbol{R}}$ is invertible, these previous approaches are the same as our approach except that they use the z-scores, $\hat{\boldsymbol{z}}$, instead of the PVE-adjusted z-scores, $\tilde{\boldsymbol{z}}$. Thus, where our approach uses the identity $\boldsymbol{X}^\intercal\boldsymbol{y} = \sqrt{N}\tilde{\boldsymbol{z}}$ (eq. 15), these previous approaches are implicitly making the approximation $\boldsymbol{X}^\intercal\boldsymbol{y} \approx \sqrt{N}\hat{\boldsymbol{z}}$. If all effect sizes are small (*i.e.*, PVE ≈ 0), then $\tilde{\boldsymbol{z}} \approx \hat{\boldsymbol{z}}$, and the approximation will be nearly exact; on the other hand, if the PVE is not close to zero for some SNPs, then the use of the PVE-adjusted z-scores is preferred [21]. Note that the PVE-adjusted z-scores require knowledge of N; in rare cases where N is unknown, replacing $\tilde{\boldsymbol{z}}$ with $\hat{\boldsymbol{z}}$ may be an acceptable approximation.

Approaches to dealing with a non-invertible LD matrix

One complication that can arise in working directly with models (20) or (22) is that $\hat{\boldsymbol{R}}$ is frequently non-invertible. For example, if $\hat{\boldsymbol{R}}$ is the sample correlation matrix from a reference panel, $\hat{\boldsymbol{R}}$ will be non-invertible (*i.e.*, singular) whenever the number of individuals in the panel is less than J, or whenever any two SNPs are in complete LD in the panel. In such cases, these models do not have a density (with respect to the Lebesgue measure). Methods using (20) have therefore required workarounds to deal with this issue. One approach is to modify ("regularize") $\hat{\boldsymbol{R}}$ to be invertible by adding a small, positive constant to the diagonal [7]. In another approach, the data are transformed into a lower-dimensional space [35, 36], which is equivalent to replacing $\hat{\boldsymbol{R}}^{-1}$ with its pseudoinverse (see S1 Text). Our approach circumvents these issues because the likelihood (19) is defined whether or not $\hat{\boldsymbol{R}}$ is invertible. (Indeed, it is

defined even if $\hat{\boldsymbol{R}}$ is not positive semi-definite, but its use in that case may be problematic as the likelihood can be unbounded; see also [37].) Our approach has several advantages over the data transformation approach: it is simpler; it does not involve inversion or factorization of a (possibly very large) $J \times J$ matrix; and it preserves the property that results under the SER model do not depend on LD (see Results and S1 Text). Also note that our approach can be combined with modifications to $\hat{\boldsymbol{R}}$, such as adding a small constant to the diagonal. The benefits of regularizing $\hat{\boldsymbol{R}}$ are investigated in the experiments below.

New refinement procedure for more accurate CSs

As noted in [17], the IBSS algorithm can sometimes converge to a poor solution (local optimum of the ELBO). Although this is rare, it can produce misleading results when it does occur; in particular it can produce false positive CSs (*i.e.*, CSs containing only null SNPs that have zero effect). To address this issue, we developed a simple refinement procedure for escaping local optima. The procedure is heuristic, and not guaranteed to eliminate all convergence issues, but in practice it often helps in those rare cases where the original IBSS has problems. The refinement procedure applies equally to both individual-level data and summary data.

In brief, the refinement procedure involves two steps: first, fit a *SuSiE* model by running the IBSS algorithm to convergence; second, for each CS identified from the fitted *SuSiE* model, rerun IBSS to convergence after first removing all SNPs in the CS (which forces the algorithm to seek alternative explanations for observed associations), then try to improve this fit by running IBSS to convergence again, with all SNPs. If these refinement steps improve the objective function, the new solution is accepted; otherwise, the original solution is kept. This process is repeated until the refinement steps no longer make any improvements to the objective. By construction, this refinement procedure always produces a solution whose objective is at least as good as the original IBSS solution. For full details, see S1 Text.

Because the refinement procedure reruns IBSS for each CS discovered in the initial round of model fitting, the computation increases with the number of CSs found. In data sets with many CSs the refinement procedure may be quite time-consuming.

Other improvements to fine-mapping with summary data

Here we introduce additional methods to improve accuracy of fine-mapping with summary data. These methods can be applied to other fine-mapping methods as well as *SuSiE*.

Regularization to improve consistency of the estimated LD matrix

Accurate fine-mapping requires $\hat{\boldsymbol{R}}$ to be an accurate estimate of $\boldsymbol{R}$. When $\hat{\boldsymbol{R}}$ is computed from a reference panel, the reference panel should not be too small [31], and should be of similar ancestry to the study sample. Even when a suitable panel is used, there will inevitably be differences between $\hat{\boldsymbol{R}}$ and $\boldsymbol{R}$. A common way to improve estimation of covariance matrices is to use regularization [38], replacing $\hat{\boldsymbol{R}}$ with $\hat{\boldsymbol{R}}_\lambda$,

$$\hat{\boldsymbol{R}}_\lambda := (1-\lambda)\hat{\boldsymbol{R}}_0 + \lambda \boldsymbol{I}, \tag{23}$$

where $\hat{\boldsymbol{R}}_0$ is the sample correlation matrix computed from the reference panel, and $\lambda \in [0,1]$ controls the amount of regularization. This strategy has previously been used in fine-mapping from summary data (e.g., [8,37,39]), but in previous work λ was usually

fixed at some arbitrarily small value, or chosen using cross-validation. Here, we estimate λ by maximizing the likelihood under the null ($\boldsymbol{z} = \mathbf{0}$),

$$\hat{\lambda} := \underset{\lambda \in [0,1]}{\operatorname{argmax}}\, \mathcal{N}_J(\tilde{\boldsymbol{z}}; \mathbf{0}, (1-\lambda)\hat{\boldsymbol{R}}_0 + \lambda \boldsymbol{I}). \tag{24}$$

The estimated $\hat{\lambda}$ reflects the consistency between the (PVE-adjusted) z-scores and the LD matrix $\hat{\boldsymbol{R}}_0$; if the two are consistent with one another $\hat{\lambda}$ will be close to zero.

Detecting and removing large inconsistencies in summary data

Regularizing $\hat{\boldsymbol{R}}$ can help address subtle inconsistencies between $\hat{\boldsymbol{R}}$ and $\boldsymbol{R}$. However, regularization cannot deal with large inconsistencies in the summary data, which, in our experience, occur often. One common source of such inconsistencies is an "allele flip" in which the alleles of a SNP are encoded one way in the study sample (used to compute $\hat{\boldsymbol{z}}$) and in a different way in the reference panel (used to compute $\hat{\boldsymbol{R}}$). Large inconsistencies can also arise from using z-scores that were obtained using different samples at different SNPs (and should be avoided by performing genotype imputation [23]). Anecdotally, we have found large inconsistencies like these often cause *SuSiE* to converge very slowly and produce misleading results, such as an unexpectedly large number of CSs, or two CSs containing SNPs that are in strong LD with each other. We have therefore developed diagnostics to help users detect such anomalous data. (We note that similar ideas were proposed in the recent paper [40].)

Under model (22), the conditional distribution of $\tilde{z}_j$ given the other PVE-adjusted z-scores is

$$\tilde{z}_j \mid \hat{\boldsymbol{R}}, \boldsymbol{b}, \tilde{\boldsymbol{z}}_{-j} \sim \mathcal{N}((\sqrt{N} b_j - \boldsymbol{\Omega}_{j,-j}\tilde{\boldsymbol{z}}_{-j})/\Omega_{jj}, 1/\Omega_{jj}), \tag{25}$$

where $\boldsymbol{\Omega} := \hat{\boldsymbol{R}}^{-1}$, $\tilde{\boldsymbol{z}}_{-j}$ denotes the vector $\tilde{\boldsymbol{z}}$ excluding $\tilde{z}_j$, and $\boldsymbol{\Omega}_{j,-j}$ denotes the jth row of $\boldsymbol{\Omega}$ excluding Ω_{jj}. This conditional distribution depends on the unknown b_j. However, provided that the effect of SNP j is small (*i.e.*, $b_j \approx 0$), or that SNP j is in strong LD with other SNPs, which implies $1/\Omega_{jj} \approx 0$, we can approximate this by

$$\tilde{z}_j \mid \hat{\boldsymbol{R}}, \tilde{\boldsymbol{z}}_{-j} \sim \mathcal{N}(-\boldsymbol{\Omega}_{j,-j}\tilde{\boldsymbol{z}}_{-j}/\Omega_{jj}, 1/\Omega_{jj}). \tag{26}$$

An initial quality control check can be performed by plotting the observed $\tilde{z}_j$ against its conditional expectation in (26), with large deviations potentially indicating anomalous z-scores. Since computing these conditional expectations involves the inverse of $\hat{\boldsymbol{R}}$, this matrix must be invertible. When $\hat{\boldsymbol{R}}$ is not invertible, we replace $\hat{\boldsymbol{R}}$ with the regularized (and invertible) matrix $\hat{\boldsymbol{R}}_\lambda$ following the steps described above.

Note that while we have written (25) and (26) in terms of the PVE-adjusted z-scores, $\tilde{\boldsymbol{z}}$, it is valid to use the same expressions for the unadjusted z-scores, $\hat{\boldsymbol{z}}$, so long as the effect sizes are small. The distribution (26) has been previously used to impute z-scores [41], and it is also used in DENTIST [40] (although DENTIST uses z-scores instead of the PVE-adjusted z-scores).

A more quantitative measure of the discordance of $\tilde{z}_j$ with its expectation under the model can be obtained by computing standardized differences between the observed and expected values,

$$t_j := \sqrt{\Omega_{jj}}(\tilde{z}_j + \boldsymbol{\Omega}_{j,-j}\tilde{\boldsymbol{z}}_{-j}/\Omega_{jj}). \tag{27}$$

SNPs j with largest (in magnitude) t_j are most likely to violate the model assumptions, and are therefore the top candidates for followup. When any such candidates are detected, the user should check the data pre-processing steps and fix any errors that cause inconsistencies in summary data. If there is no way to fix the errors, removing the anomalous SNPs is a possible workaround. Sometimes removing a single SNP is enough

to resolve the discrepancies—for example, a single allele flip can result in inconsistent z-scores among many other SNPs in LD with the allele-flip SNP. We have also developed a likelihood-ratio statistic based on (26) specifically for identifying allele flips; see S1 Text for a derivation of this likelihood ratio and an empirical assessment of its ability to identify allele-flip SNPs in simulations. After one or more SNPs are removed, one should consider re-running these diagnostics on the filtered summary data to search for additional inconsistencies that may have been missed in the first round. Alternatively, DENTIST provides a more automated approach to filtering out inconsistent SNPs [40].

We caution that computing these diagnostics requires inverting or factorizing a $J \times J$ matrix, and may therefore involve large computational expense—potentially a greater expense than the fine-mapping itself—when J, the number of SNPs, is large.

Software availability

The *SuSiE* and *SuSiE-RSS* methods are implemented in the R package `susieR`. It is available for download at https://github.com/stephenslab/susieR, and on CRAN at https://cran.r-project.org/package=susieR.

Verification and comparison

Fine-mapping with inconsistent summary data and a non-invertible LD matrix: an illustration

A technical issue that arises when developing fine-mapping methods for summary data (z-scores and LD matrix) is that the LD matrix is often not invertible. Several approaches to dealing with this have been suggested, including modifying the LD matrix to be invertible, transforming the data into a lower-dimensional space, or replacing the inverse with what is known as the "pseudoinverse" (see "Approaches to dealing with a non-invertible LD matrix" above). In *SuSiE-RSS*, we avoid this issue by directly approximating the likelihood, so *SuSiE-RSS* does not require the LD matrix to be invertible. We summarize the theoretical relationships between these methods in S1 Text. Here illustrate the practical advantage of the *SuSiE-RSS* approach in a toy example.

Consider a very simple situation with two SNPs, in strong LD, with observed z-scores $\hat{\boldsymbol{z}} = (6, 7)$. Both SNPs are significant, but the second SNP is more significant. Under the assumption that exactly one of these SNPs has an effect—which allows for exact posterior computations—the second SNP is the better candidate, and should have a higher PIP. Further, the PIPs should be unaffected by LD between the SNPs (see S1 Text). However, the transformation and pseudoinverse approaches—which are used by msCAVIAR [42] and in previous fine-mapping analyses [35, 36], and are also used in DENTIST to detect inconsistencies in summary data [40]—do not guarantee that either of these properties are satisfied. For example, suppose the two SNPs are in complete LD in the reference panel, so $\hat{\boldsymbol{R}}$ is a 2×2 (non-invertible) matrix with all entries equal to 1. Note that $\hat{\boldsymbol{R}}$ is inconsistent with the observed $\hat{\boldsymbol{z}}$ because complete LD betweeen SNPs implies their z-scores should be identical. (This could happen if the LD in the reference panel used to compute $\hat{\boldsymbol{R}}$ is slightly different from the LD in the association study.) The tranformation approach effectively adjusts the observed data $\hat{\boldsymbol{z}}$ to be consistent with the LD matrix before drawing inferences; here it would adjust $\hat{\boldsymbol{z}}$ to $\hat{\boldsymbol{z}} = (6.5, 6.5)$, removing the observed difference between the SNPs and forcing them to be equally significant, which seems undesirable. The pseudoinverse approach turns out to be equivalent to the tranformation approach (see S1 Text), and so behaves the same way. In contrast, our approach avoids this behaviour and correctly maintains the second SNP

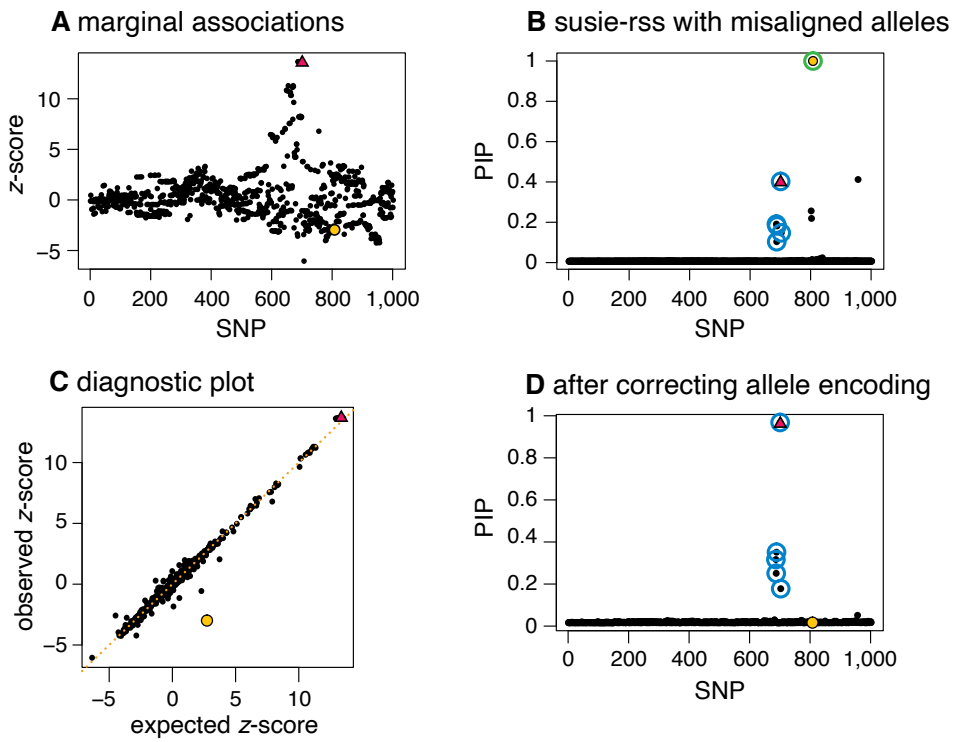

Fig 1. Example illustrating importance of identifying and correcting allele flips in fine-mapping. In this simulated example, one SNP (red triangle) affects the phenotype, and one SNP (yellow circle) has a different allele encoding in the study sample (the data used to compute the z-scores) and the reference panel (the data used to compute the LD matrix). Panel A shows the z-scores for all 1,002 SNPs. Panel B summarizes the results of running *SuSiE-RSS* on the summary data; *SuSiE-RSS* identifies a true positive CS (blue circles) containing the true causal SNP and a false positive CS (green circlesq) that incorrectly contains the mismatched SNP. The mismatched SNP is also incorrectly estimated to have an effect on the phenotype with very high probability (PIP = 1.00). The diagnostic plot (Panel C) compares the observed z-scores against the expected z-scores. In this plot, the mismatched SNP (yellow circle) shows the largest difference between observed and expected z-scores, and therefore appears furthest away from the diagonal. After fixing the allele encoding and recomputing the summary data, *SuSiE-RSS* identifies a single true positive CS (blue circles) containing the true-causal SNP (red triangle), and the formerly mismatched SNP is (correctly) not included in a CS (Panel D). This example is implemented as a vignette in the `susieR` package.

as the better candidate; applying *SuSiE-RSS* to this toy example yields PIPs of 0.001 for the first SNP and 0.998 for the second SNP, and a single CS containing the second SNP only. To reproduce this result, see the examples accompanying the `susie_rss` function in the `susieR` R package.

Effect of allele flips on accuracy of fine-mapping: an illustration

When fine-mapping is performed using z-scores from a study sample and an LD matrix from a different reference sample, it is crucial that the same alleles encodings are used in each sample. In our experience, "allele flips," in which different allele encoding are

used in the two samples, are a common source of fine-mapping problems. Here we use a simple simulation to illustrate this problem, and the steps we have implemented to diagnose and correct the problem.

We simulated a fine-mapping data set with 1,002 SNPs, in which one out of the 1,002 SNPs was causal, and we deliberately used different allele encodings in the study sample and reference panel for one of the non-causal SNPs; see S1 Text for more details on this simulation. The causal SNP is among the SNPs with the highest z-scores (the red triangle in Fig 1, Panel A), and *SuSiE-RSS* correctly includes this causal SNP in a CS (Panel B). However, *SuSiE-RSS* also wrongly includes the allele-flip SNP in a second CS (Panel B). This happens because the LD between the allele-flip SNP and other SNPs is incorrectly estimated. Fig 1, Panel C shows a diagnostic plot comparing each z-score against its expected value under model (22). The allele-flip SNP stands out as a likely outlier (yellow circle), and the likelihood ratio calculations identify this SNP as a likely allele flip: $\mathrm{LR} = 8.2 \times 10^3$ for the allele-flip SNP, whereas all the other 262 SNPs with z-scores greater than 2 in magnitude have likelihood ratios less than 1. (See S1 Text for a more systematic assessment of the use of these likelihood ratio for identifying allele-flip SNPs.) After correcting the allele encoding to be the same in the study and reference samples, *SuSiE-RSS* infers a single CS containing the causal SNP, and the allele-flip SNP is no longer included in a CS; see Fig 1, Panel D.

Note that these diagnostic tools are not sufficiently accurate to automatically identify and correct allele flips or other errors. Nonetheless they can be helpful to assist users with identifying errors in computational pipelines.

Simulations using UK Biobank genotypes

To systematically compare our new methods with existing methods for fine-mapping, we simulated fine-mapping data sets using the UK Biobank imputed genotypes [43]. The UK Biobank imputed genotypes are well suited to illustrate fine-mapping with summary data due to the large sample size, and the high density of available genetic variants after imputation. We randomly selected 200 regions on autosomal chromosomes for fine-mapping, such that each region contained roughly 1,000 SNPs (average size: 390 kb). Due to the high density of SNPs, these data sets often contain very strong correlations among SNPs; on average, a data set contained 30 SNPs with correlation exceeding 0.9 with at least one other SNP, and 14 SNPs with correlations exceeding 0.99 with at least one other SNP.

For each of the 200 regions, we simulated a quantitative trait under the multiple regression model (1) with $\boldsymbol{X}$ comprising genotypes of 50,000 randomly selected UK Biobank samples, and with 1, 2 or 3 causal variants explaining a total of 0.5% of variation in the trait (total PVE of 0.5%). In total, we simulated $200 \times 3 = 600$ data sets. We computed summary data from these real genotypes and the synthetic phenotypes. To compare how choice of LD matrix affects fine-mapping, we used three different LD matrices: the in-sample LD matrix computed from the 50,000 individuals ($\boldsymbol{R}$), and two out-of-sample LD matrices computed from randomly-sampled reference panels of 500 or 1,000 individuals, denoted $\hat{\boldsymbol{R}}_{500}$ and $\hat{\boldsymbol{R}}_{1000}$, respectively. The samples randomly chosen for each reference panel had no overlap with the study sample but were drawn from the same population, which mimics a situation where the reference sample is well matched to the study sample.

Refining *SuSiE* model fits improves fine-mapping performance

We begin by demonstrating the benefits of our new refinement procedure for improving *SuSiE* model fits. Fig 2 shows an example drawn from our simulations where the regular IBSS algorithm converges to a poor solution and our refinement procedure

Table 2. Runtimes on simulated data sets with in-sample LD matrix. Average runtimes are taken over 600 simulations. All runtimes are in seconds. All runtimes include time taken to read the data and write the results to files.

method	min.	average	max.
SuSiE-RSS, estimated σ, no refinement	0.65	1.33	18.89
SuSiE-RSS, estimated σ, with refinement	1.62	5.50	72.57
SuSiE-RSS, fixed σ, no refinement	0.40	1.40	18.61
SuSiE-RSS, fixed σ, with refinement	1.44	4.81	62.34
SuSiE-RSS, fixed σ, with refinement, $L = \text{true}$	0.37	1.52	4.95
DAP-G	0.66	5.70	371.76
FINEMAP	1.67	16.11	39.27
FINEMAP, $L = \text{true}$	1.00	12.92	42.93
CAVIAR, $L = \text{true}$	3.54	1,516.91	4,831.95

improves the solution. The example has two causal SNPs in moderate LD with one another, which have opposite effects that partially cancel out each others' marginal associations (Panel A). This example is challenging because the SNP with the strongest marginal association (SMA) is not in high LD with either causal SNP; it is in moderate LD with the first causal SNP, and low LD with the second causal SNP. Although [17] showed that the IBSS algorithm can sometimes deal well with such situations, that does not happen in this case; the IBSS algorithm yields three CSs, two of which are false positives that do not contain a causal SNP (Panel B). Applying our refinement procedure solves the problem; it yields a solution with higher objective function (ELBO), and with two CSs, each containing one of the causal SNPs (Panel C).

Although this sort of problem was not common in our simulations, it occurred often enough that the refinement procedure yielded a noticeable improvement in performance across many simulations (Fig 2, Panel D). (In this plot, power and false discovery rate (FDR) are 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.) In our remaining experiments we therefore always ran *SuSiE-RSS* with refinement.

Impact of LD accuracy on fine-mapping

We performed simulations to compare *SuSiE-RSS* with several other fine-mapping methods for summary data: FINEMAP [12, 21], DAP-G [14, 16] and CAVIAR [7]. These methods differ in the underlying modeling assumptions, the priors used, and in the approach taken to compute posterior quantities. For these simulations, *SuSiE-RSS*, FINEMAP and DAP-G were all very fast, usually taking no more than a few seconds per data set (see Table 2 in S1 Text); by contrast, CAVIAR was much slower because it exhaustively evaluated all causal SNP configurations. (Other Bayesian fine-mapping methods for summary data include PAINTOR [8], JAM [15] and CAVIARBF [11]. FINEMAP has been shown [12] to be faster and at least as accurate as PAINTOR and CAVIARBF. JAM is comparable in accuracy to FINEMAP [15] and is most beneficial when jointly fine-mapping multiple genomic regions, which we did not consider here.)

We compared methods based on both their posterior inclusion probabilities (PIPs) [44] and credible sets (CSs) [2, 17]. These quantities have different advantages. PIPs have the advantage that they are returned by most methods, and can be used to assess familiar quantities such as power and false discovery rates. CSs have the

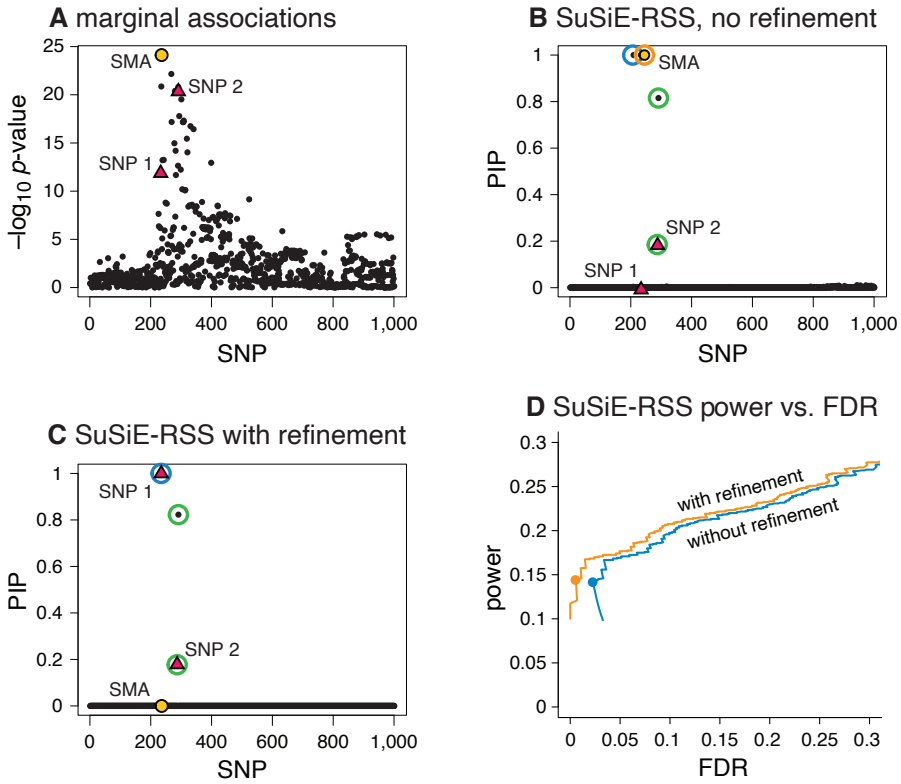

Fig 2. Refining *SuSiE* model fits improves fine-mapping accuracy. Panels A, B and C show a single example, drawn from our simulations, that illustrates how refining a *SuSiE-RSS* model fit improves fine-mapping accuracy. In this example, there are 1,001 candidate SNPs, and two SNPs (red triangles "SNP 1" and "SNP 2") explain variation in the simulated phenotype. The strongest marginal association (yellow circle, "SMA") is not a causal SNP. Without refinement, the IBSS-ss algorithm (applied to sufficient statistics, with estimated σ^2) returns a *SuSiE-RSS* fit identifying three 95% CSs (blue, green and orange circles); two of the CSs (blue, orange) are false positives containing no true effect SNP, one of these CSs contains the SMA (orange), and no CS includes SNP 1. After running the refinement procedure, the fit is much improved, as measured by the "evidence lower bound" (ELBO); it increases the ELBO by 19.06 (−70837.09 vs. −70818.03). The new *SuSiE-RSS* fit (Panel C) identifies two 95% CSs (blue and green circles), each containing a true causal SNP, and neither contains the SMA. Panel D summarizes the improvement in fine-mapping across all simulations; it shows power and false discovery rate (FDR) for *SuSiE-RSS* with and without using the refinement procedure as the PIP threshold for reporting causal SNPs is varied from 0 to 1. (This plot is the same as a precision-recall curve after flipping the x-axis because precision $= \frac{\mathrm{TP}}{\mathrm{TP+FP}} = 1 - \mathrm{FDR}$ and recall = power.) Circles are drawn at a PIP threshold of 0.95.

advantage that, when the data support multiple causal signals, the multiple causal signals is explicitly reflected in the number of CSs reported. Uncertainty in which SNP are causal is reflected in the size of a CS.

First, we assessed the performance of summary-data methods using the in-sample LD matrix. With an in-sample LD matrix, *SuSiE-RSS* applied to sufficient statistics (with estimated σ^2) will produce the same results as *SuSiE* on the individual-level data, so we did not include *SuSiE* in this comparison. The results show that *SuSiE-RSS*,

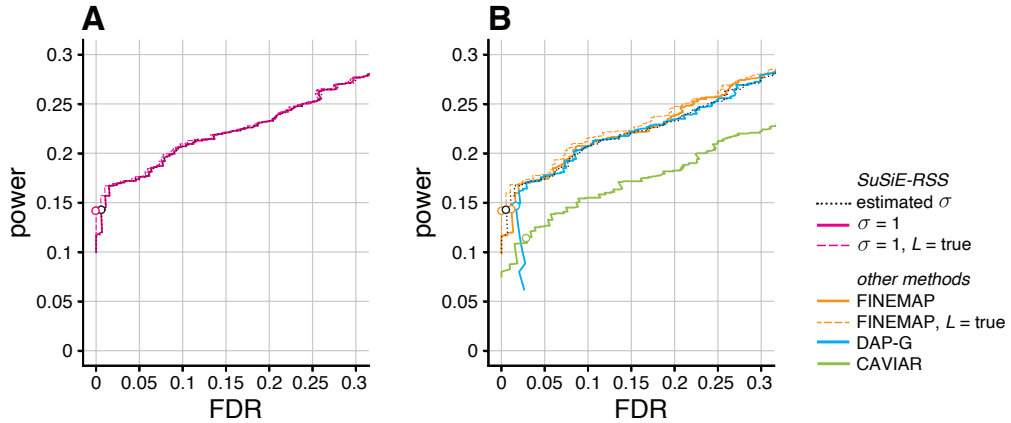

Fig 3. Discovery of causal SNPs using posterior inclusion probabilities—in-sample LD. Each curve shows power vs. FDR in identifying causal SNPs when the method (*SuSiE-RSS*, FINEMAP, DAP-G or CAVIAR) was provided with the in-sample LD matrix. FDR and power are calculated from 600 simulations as the PIP threshold is varied from 0 to 1. Open circles are drawn at a PIP threshold of 0.95. Two variants of FINEMAP and three variants of *SuSiE-RSS* are also compared: L, the maximum number of estimated causal SNPs, is the true number of causal SNPs, or larger than the true number; and, for *SuSiE-RSS* only, the residual variance σ^2 is estimated ("sufficient data") or fixed to 1 ("summary data"). Note that power and FDR are virtually identical for all three variants of *SuSiE-RSS* so the three curves almost fully overlap in Panel A.

FINEMAP and DAP-G have very similar performance, as measured by both PIPs (Fig 3) and CSs (the "in-sample LD" columns in Fig 4). The main difference between the methods is that DAP-G produced some "high confidence" (high PIP) false positives, which hinders its ability to produce very low FDR values. Further, all four methods produced CSs whose coverage was close to the target level of 95% (Panel A in Fig 4). Both *SuSiE-RSS* and FINEMAP require the user to specify an upper bound on the number of causal SNPs L. Setting this upper bound to the true value ("L = true" in the figures) only slightly improved their performance, demonstrating that, with an in-sample LD matrix, these methods are robust to overstating this bound. We also compared the sufficient-data (estimated σ^2) and summary-data (fixed σ^2) variants of *SuSiE-RSS*. The performance of the two variants was very similar, likely owing to the fact that the PVE was close to zero in all simulations, and so $\sigma = 1$ was not far from the truth. CAVIAR performed notably less well than the other methods for the PIP computations. (The CSs computed by CAVIAR are defined differently from CSs computed by other methods, so we did not include CAVIAR in the CS comparisons.)

Next, we compared the summary data methods using different out-of-sample LD matrices, again using *SuSiE-RSS* with in-sample LD (and estimated σ^2) as a benchmark. For every method, we computed out-of-sample LD matrices using two different panel sizes ($n = 500, 1000$) and three different values for the regularization parameter, λ (no regularization, $\lambda = 0$; weak regularization, $\lambda = 0.001$; and the regularization level λ estimated from the data as described in Methods). As might be expected, the performance of *SuSiE-RSS*, FINEMAP and DAP-G all degraded with out-of-sample LD compared with in-sample LD; see Fig 5 and Fig 4. Notably, the CSs no longer met the 95% target coverage (Panel A in Fig 4). In all cases, performance was notably worse with the smaller reference panel, which highlights the importance of using

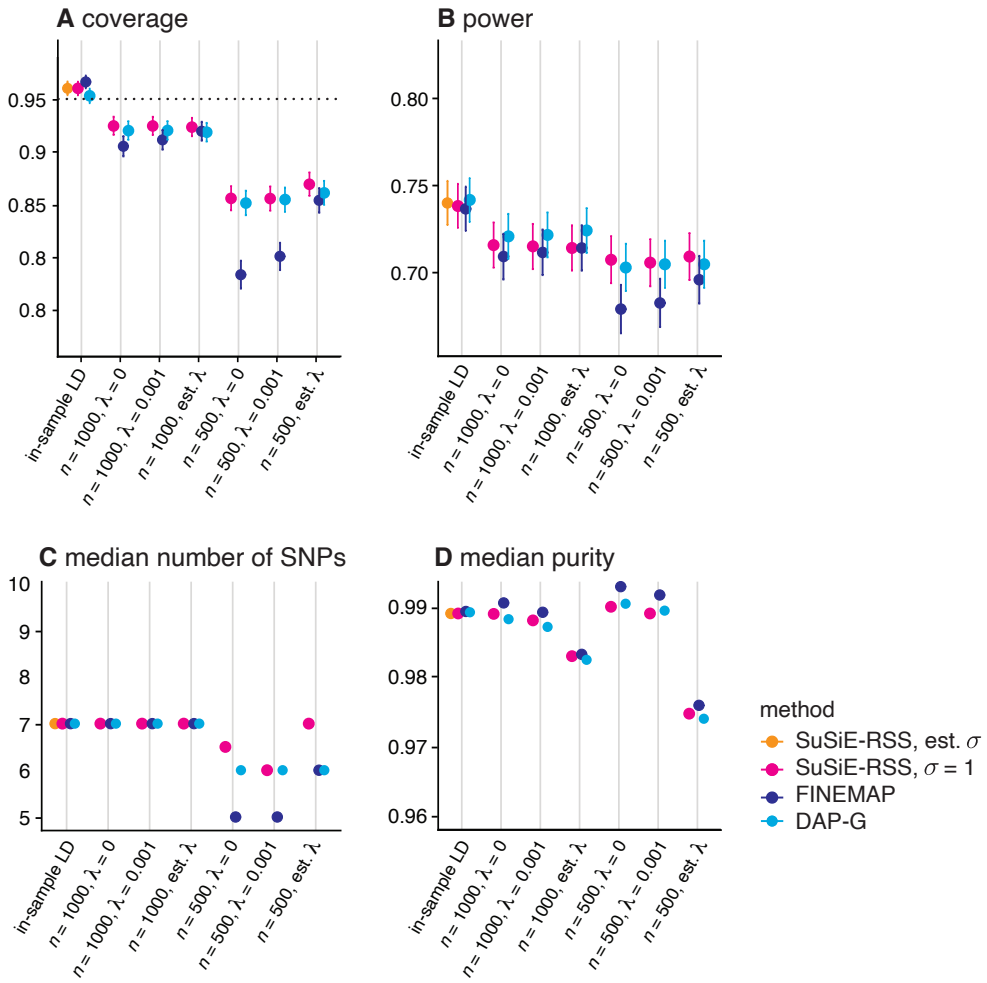

Fig 4. Assessment of 95% credible sets from *SuSiE-RSS*, FINEMAP and DAP-G with different LD estimates. Two variants of *SuSiE-RSS* were also compared: when the residual variance σ^2 was estimated ("sufficient data"), or fixed to 1 ("summary data"). We evalulate CSs using: (A) coverage, the proportion of CSs that contain a true causal SNP; (B) power, the proportion of true causal SNPs included in a CS; (C) median number of SNPs in each CS; and (D) median purity, where "purity" is defined as the smallest absolute correlation among all pairs of SNPs within a CS. Following [17], we discarded any CSs with purity less than 0.5. These statistics are taken as the mean (A, B) or median (C, D) over all simulations; error bars in A and B show two times the standard error. The target coverage of 95% is shown as a dotted horizontal line in Panel A.

a sufficiently large reference panel [31]. Regarding regularization, *SuSiE-RSS* and DAP-G performed similarly at all levels of regularization, and so do not appear to require regularization; in contrast, FINEMAP required regularization with an estimated regularization level λ to compete with *SuSiE-RSS* and DAP-G. Since estimating the regularization level is somewhat computationally burdensome, *SuSiE-RSS* and DAP-G have an advantage in this situation. All three methods benefited more from increasing the size of the reference panel than from regularization, again emphasizing the importance of sufficiently large reference panels. Interestingly, CAVIAR's performance was relatively insensitive to choice of LD matrix; however the other methods clearly

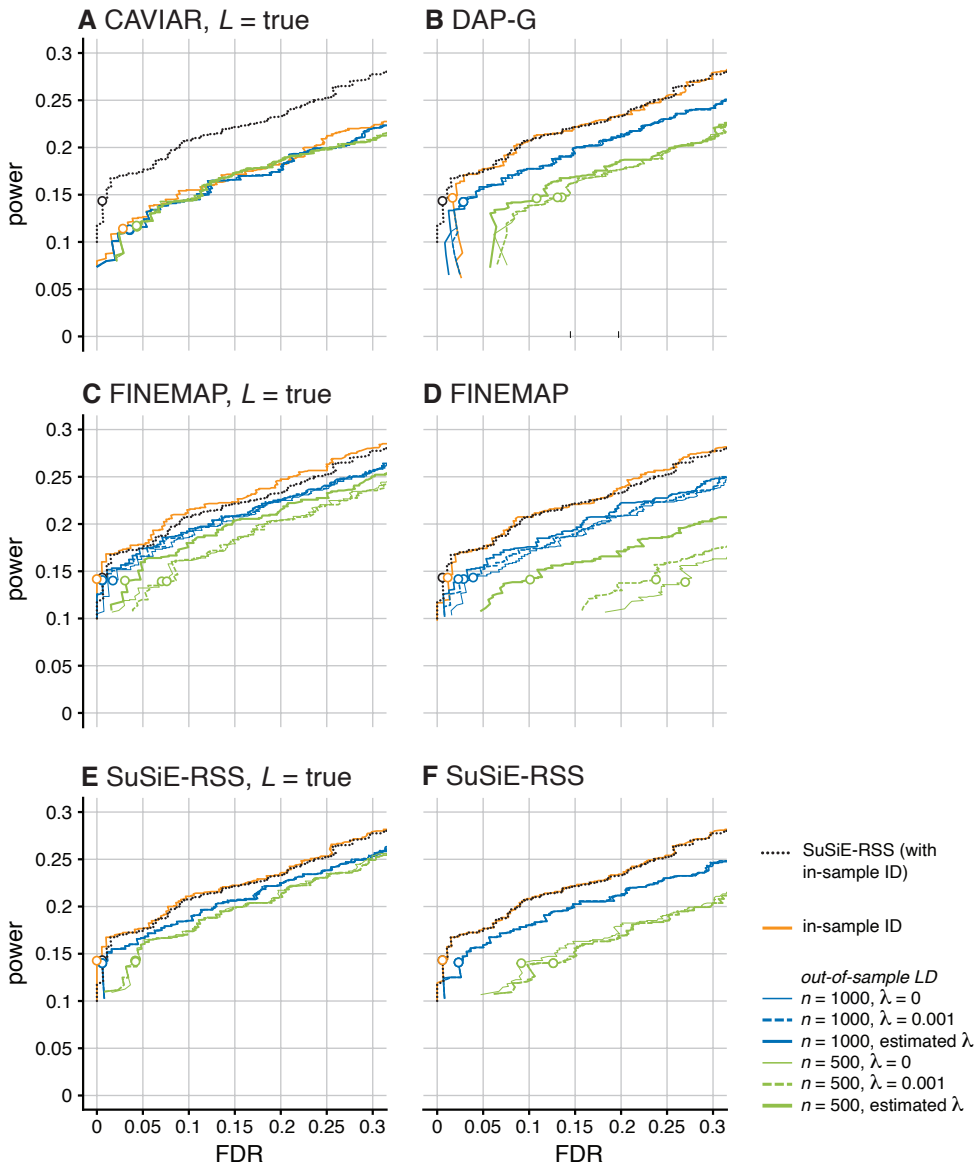

Fig 5. Discovery of causal SNPs using posterior inclusion probabilities—out-of-sample LD. Plots compare power vs. FDR of fine-mapping methods with different LD matrices, across 600 simulations, as the PIP threshold is varied from 0 to 1. Open circles indicate results at PIP threshold of 0.95. Each plot compares performance of one method (CAVIAR, DAP-G, FINEMAP or *SuSiE-RSS*) when provided with different LD estimates: in-sample ($\hat{\boldsymbol{R}} = \boldsymbol{R}$), or out-of-sample LD from a reference panel with either 1,000 samples ($\hat{\boldsymbol{R}} = \hat{\boldsymbol{R}}_{1000}$) or 500 samples ($\hat{\boldsymbol{R}} = \hat{\boldsymbol{R}}_{500}$). For out-of-sample LD, different levels of the regularization paremeter λ are also compared: $\lambda = 0$; $\lambda = 0.001$; and estimated λ (see Methods). Panels C–F show results for two variants of FINEMAP and *SuSiE-RSS*: in Panels C and E, the maximum number of causal SNPs, L, is set to the true value ($L =$ true); in Panels D and F, L is set larger than the true value ($L = 4$ for FINEMAP; $L = 10$ for *SuSiE-RSS*). In each panel, the dotted black line shows the results from *SuSiE-RSS* with in-sample LD and estimated σ^2, which provides a baseline for comparison (note that all the other *SuSiE-RSS* results were generated by fixing σ^2 to 1, which is the recommended setting for out-of-sample LD). Some power vs. FDR curves may not be visible in the plots because they overlap almost completely with another curve, such as some of the *SuSiE-RSS* results at different LD regularization levels.

outperformed CAVIAR with the larger ($n = 1,000$) reference panel.

The fine-mapping results with out-of-sample LD matrix also expose another interesting result: if FINEMAP and *SuSiE-RSS* are provided with the true number of causal SNPs (L = true), their results improve (Fig 5, Panels C vs. D, Panels E vs. F). This improvement is particularly noticeable for the small reference panel (light green lines). We interpret this result as indicating a tendency of these methods to react to misspecification of the LD matrix by sometimes including additional (false positive) signals. Specifying the true L reduces their tendency to do this because it limits the number of signals that can be included. This suggests that restricting the number of causal SNPs, L, may make fine-mapping results more robust to misspecification of the LD matrix (even for methods that are robust to overstating L when the LD matrix is accurate). Alternatively, priors or penalties that favor smaller L may also help. Indeed, when none of the methods are provided with information about the true number of causal SNPs, DAP-G slightly outperforms FINEMAP and *SuSiE-RSS*, possibly reflecting a tendency for DAP-G to favour models with smaller numbers of SNPs (either due to the differences in prior or differences in approximate posterior inference). Further study of this issue may lead to methods that are more robust to misspecified LD.

Fine-mapping causal SNPs with larger effects

So far, we have evaluated the performance of fine-mapping methods in simulations when the simulated effects of the causal SNPs were small (total PVE of 0.5%). This was intended to mimic the typical situation encountered in genome-wide association studies [45, 46]. Here we scrutinize the performance of fine-mapping methods when the effects of the causal SNPs are much larger, which might be more representative of the situation in expression quantitative trait loci (eQTL) studies [47–49]. FINEMAP and *SuSiE*—and therefore *SuSiE-RSS* with sufficient statistics—are expected to perform well in this setting [17, 21], but, as mentioned above, some summary-data methods make the (implicit) assumption that the effects are small (see "Connections with previous work"), and this assumption may affect performance in settings where this assumption is violated.

To assess the ability of the fine-mapping methods to identify causal SNPs with larger effects, we performed an additional set of simulations, again using the UK Biobank genotypes, except that here we simulated the 1–3 causal variants so that they explained, in total, a much larger proportion of variance in the trait (PVE of 10% and 30%). To evaluate these methods at roughly the same level of difficulty (*i.e.,* power), we simulated these fine-mapping data sets with much smaller sample sizes (n = 2,500 and 800, respectively).

The results of these high-PVE simulations are summarized in Fig 6. As expected, *SuSiE-RSS* with in-sample LD matrix performed consistently better than the other methods, which use an out-of-sample LD matrix, and therefore provides a baseline against which other methods can be compared. Overall, increasing PVE tended to increase variation in performance among the methods. In all PVE settings, *SuSiE-RSS* with out-of-sample LD was among the top performers, and it most clearly outperformed other methods in the highest PVE setting (30% PVE), where all of FINEMAP, DAP-G, and CAVIAR showed a notable decrease in performance. For DAP-G and CAVIAR, this decrease in performance was expected due to their implicit modeling assumption that the effect sizes are small. For FINEMAP, this drop in performance was unexpected, since FINEMAP uses the PVE-adjusted z-scores to account for larger effects. Although this situation is unusual in fine-mapping studies—it is unusual for a handful of SNPs to explain such a large proportion of variance in the trait—we examined these FINEMAP results more closely to understand why this was happening. (Also, see https://stephenslab.github.io/finemap/large_effect.html for a step-by-step

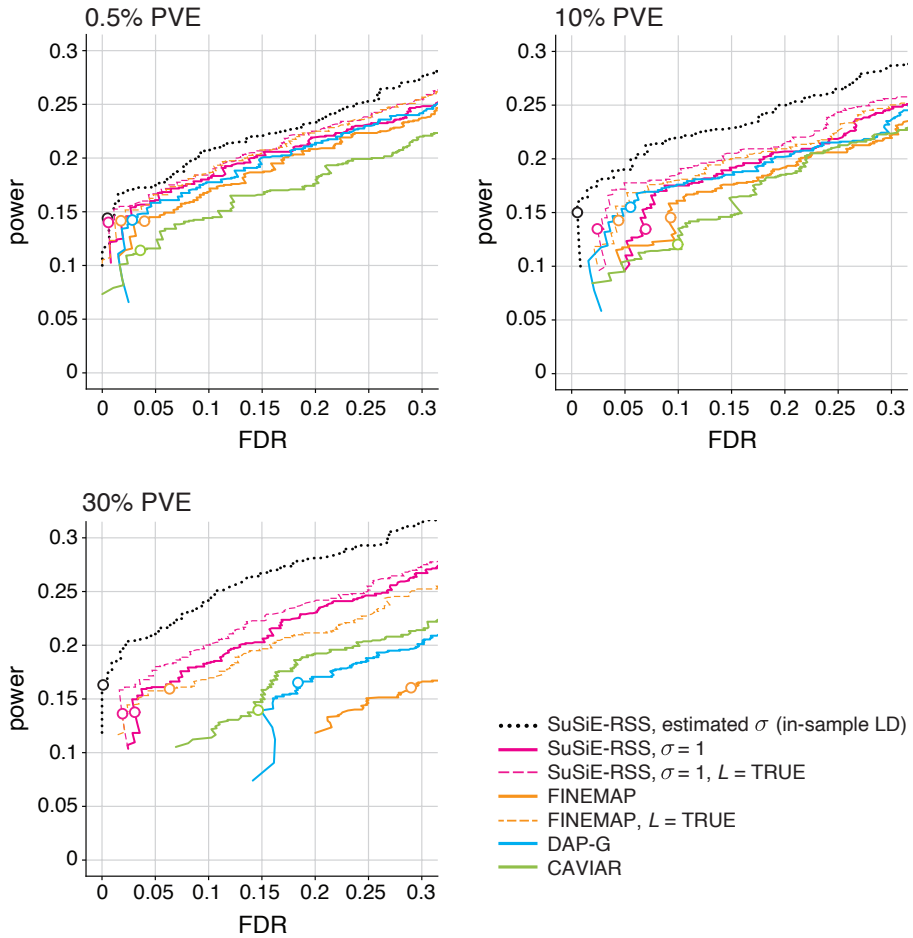

Fig 6. Discovery of causal SNPs using posterior inclusion probabilities—out-of-sample LD and larger effects. Each curve shows power vs. FDR for identifying causal SNPs with different effect sizes (total PVE of 0.5%, 10% and 30%). Each panel summarizes results from 600 simulations; FDR and power are calculated from the 600 simulations as the PIP threshold is varied from 0 to 1. Open circles depict power and FDR at a PIP threshold of 0.95. In addition to comparing different methods (*SuSiE-RSS*. FINEMAP, DAP-G, CAVIAR), two variants of FINEMAP and *SuSiE-RSS* are also compared: when L, the maximum number of estimated causal SNPs, is set to the true number of causal SNPs; and when L is larger than the true number. *SuSiE-RSS* with estimated residual variance σ^2 and in-sample LD (dotted black line) is shown as a "best case" method against which other methods can be compared. All other methods are given an out-of-sample LD matrix computed from a reference panel with 1,000 samples, with no regularization ($\lambda = 0$). The simulation results for 0.5% PVE (top-left panel) are the same as the results shown in previous plots (Fig 3, Fig 5), but presented differently to faciliate comparison with the results of the higher-PVE simulations. Note that the higher-PVE simulations used smaller sample sizes so that power is roughly equal in each panel.

example illustrating this result.) First, we confirmed that this performance drop only occurred with an out-of-sample LD matrix; with an in-sample LD matrix, FINEMAP's performance was very similar to *SuSiE-RSS*'s with an in-sample LD matrix (results not shown). A partial explanation for the much worse performance with out-of-sample LD

was that FINEMAP often overestimated the number of causal SNPs; in 17% of the simulations, FINEMAP assigned highest probability to configurations with more causal SNPs than the true number. By contrast, *SuSiE-RSS* overestimated the number of causal SNPs (*i.e.*, the number of CSs) in only 1% of the simulations. Fortunately, in settings where causal SNPs might have larger effects, FINEMAP's performance can be greatly improved by telling it the true number of causal SNPs ("L = TRUE"), which is consistent with our earlier finding that restricting L in *SuSiE-RSS* and FINEMAP can improve fine-mapping with an out-of-sample LD matrix.

Discussion

We have presented extensions of the *SuSiE* fine-mapping method to acccommodate summary data, with a focus on marginal z-scores and an out-of-sample LD matrix computed from a reference panel. Our approach provides a general template for how to extend any full-data regression method to analyse summary data: develop a full-data algorithm that works with sufficient statistics, then apply this algorithm directly to summary data. Although it is simple, as far as we are aware this template is novel, and it avoids the need for any special treatment of non-invertible LD matrices.

In simulations, we found that our new method, *SuSiE-RSS*, is competitive in both accuracy and computational cost with the best existing methods for fine-mapping from summary data, DAP-G and FINEMAP. Whatever method is used, our results underscore the importance of accurately computing out-of-sample LD from an appropriate and large reference panel (see also [31]). Indeed, for the best performing methods, performance depended more on choice of LD matrix than on choice of method. We also emphasize the importance of computing z-scores at different SNPs from the exact same samples, using genotype imputation if necessary [50]. It is also important to ensure that alleles are consistently encoded in study and reference samples.

Although our derivations and simulations focused on z-scores computed from quantitative traits with a simple linear regression, in practice it is common to apply summary data fine-mapping methods to z-scores computed in other ways, e.g., using logistic regression on a binary or case-control trait, or using linear mixed models to deal with population stratification and relatedness. The multivariate normal assumption on z-scores, which underlies all the methods considered here, should also apply to these settings, although as far as we are aware theoretical derivation of the precise form (20) is lacking in these settings (although see [12, 51, 52]). Since the model (20) is already only an approximation, one might expect that the additional effect of such issues might be small, particularly compared with the effect of allele flips or small reference panels. Nonetheless, since our simulations show that model misspecification can hurt performance of existing methods, further research to improve robustness of fine-mapping methods to model misspecification would be welcome.

Acknowledgements

This work was supported in part by NIH National Human Genome Research Institute grant R01HG002585 and by a grant from the Gordon and Betty Moore Foundation to M. Stephens. G. Wang was supported by NIH National Institute on Aging grant U01AG072572 and funding from the Thompson Family Foundation (TAME-AD). We thank Kaiqian Zhang for her contributions to the development and testing of the `susieR` package. We thank the University of Chicago Research Computing Center and the Center for Research Informatics for providing high-performance computing resources used to run the numerical experiments. This research has been conducted using data

from UK Biobank, a major biomedical database (under Application Number 27386).

References

1. Hutchinson A, Asimit J, Wallace C. Fine-mapping genetic associations. Human Molecular Genetics. 2020;29(R1):R81–R88.

2. Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nature Genetics. 2012;44(12):1294–1301.

3. Kote-Jarai Z, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Dadaev T, Jugurnauth-Little S, et al. Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. Human Molecular Genetics. 2013;22(12):2520–2528.

4. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nature Reviews Genetics. 2018;19(8):491–504.

5. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Human Molecular Genetics. 2015;24(R1):R111–R119.

6. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. American Journal of Human Genetics. 2017;101(1):5–22.

7. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. Genetics. 2014;198(2):497–508.

8. Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genetics. 2014;10(10):e1004722.

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

10. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics. 2012;44(4):369–375.

11. Chen W, Larrabee BR, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Poland GA, et al. Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. Genetics. 2015;200(3):719–736.

12. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493–1501.

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

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

15. Newcombe PJ, Conti DV, Richardson S. JAM: a scalable Bayesian framework for joint analysis of marginal SNP effects. Genetic Epidemiology. 2016;40(3):188–201.

16. Lee Y, Luca F, Pique-Regi R, Wen X. Bayesian multi-SNP genetic association analysis: control of FDR and use of summary statistics. bioRxiv. 2018;doi:10.1101/316471.

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

18. Sesia M, Katsevich E, Bates S, Candès E, Sabatti C. Multi-resolution localization of causal variants across the genome. Nature Communications. 2020;11:1093.

19. Wallace C, Cutler AJ, Pontikos N, Pekalski ML, Burren OS, Cooper JD, et al. Dissection of a complex disease susceptibility region using a Bayesian stochastic search approach to fine mapping. PLoS Genetics. 2015;11(6):e1005272.

20. Hutchinson A, Watson H, Wallace C. Improving the coverage of credible sets in Bayesian genetic fine-mapping. PLoS Computational Biology. 2020;16(4):e1007829.

21. Benner C, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. Refining fine-mapping: effect sizes and regional heritability. bioRxiv. 2018;doi:10.1101/318618.

22. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nature Reviews Genetics. 2017;18(2):117–127.

23. Zhu X, Stephens M. Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. Annals of Applied Statistics. 2017;11(3):1561–1592.

24. Kanai M, Elzur R, Zhou W, analysis Initiative GBM, Daly MJ, Finucane HK. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. medRxiv. 2022;doi:10.1101/2022.03.16.22272457.

25. Chipman H, George EI, McCulloch RE. The practical implementation of Bayesian model selection. In: Model Selection. vol. 38 of IMS Lecture Notes. Institute of Mathematical Statistics; 2001. p. 65–116.

26. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genetics. 2008;4(10):e1000214.

27. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. American Journal of Human Genetics. 2014;94(4):559–573.

28. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genetics. 2014;10(5):e1004383.

29. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–2191.

30. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

31. Benner C, Havulinna AS, Järvelin MR, Salomaa V, Ripatti S, Pirinen M. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. American Journal of Human Genetics. 2017;101(4):539–551.

32. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. Nature Reviews Genetics. 2009;10(10):681–690. doi:10.1038/nrg2615.

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

34. Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the "Sum of Single Effects" model. bioRxiv. 2021;doi:10.1101/2021.11.03.467167v1.

35. Lozano JA, Hormozdiari F, Joo JWJ, Han B, Eskin E. The multivariate normal distribution framework for analyzing association studies. bioRxiv. 2017;doi:10.1101/208199.

36. Park Y, Sarkar AK, He L, Davila-Velderrain J, De Jager PL, Kellis M. A Bayesian approach to mediation analysis predicts 206 causal target genes in Alzheimer's disease. bioRxiv. 2017;doi:10.1101/219428.

37. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genetic epidemiology. 2017;41(6):469–480.

38. Ledoit O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis. 2004;88(2):365–411.

39. Pasaniuc B, Zaitlen N, Shi H, Bhatia G, Gusev A, Pickrell J, et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. Bioinformatics. 2014;30(20):2906–2914.

40. Chen W, Wu Y, Zheng Z, Qi T, Visscher PM, Zhu Z, et al. Improved analyses of GWAS summary statistics by reducing data heterogeneity and errors. Nature Communications. 2021;12:7117.

41. Lee D, Bigdeli TB, Riley BP, Fanous AH, Bacanu SA. DIST: direct imputation of summary statistics for unmeasured SNPs. Bioinformatics. 2013;29(22):2925–2927.

42. LaPierre N, Taraszka K, Huang H, He R, Hormozdiari F, Eskin E. Identifying causal variants by fine mapping across multiple studies. bioRxiv. 2020;doi:10.1101/2020.01.15.908517.

43. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–209.

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

45. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nature Genetics. 2008;40(6):695–701.

46. Ioannidis JPA, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. American Journal of Epidemiology. 2006;164(7):609–614.

47. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, et al. Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Genome Biology. 2021;22:49.

48. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648–660.

49. Wen X, Luca F, Pique-Regi R. Cross-population joint analysis of eQTLs: fine mapping and functional annotation. PLoS Genetics. 2015;11(4):e1005176.

50. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics. 2016;48(5):481–487.

51. Han B, Kang HM, Eskin E. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. PLoS Genetics. 2009;5(4):1–13. doi:10.1371/journal.pgen.1000456.

52. Pirinen M, Donnelly P, Spencer CCA. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. Annals of Applied Statistics. 2013;7(1):369–390.

Supporting information

S1 Fig. Likelihood ratio for detecting allele flips. These plots summarize the likehood ratios LR_j for SNPs j in simulated fine-mapping data sets, separately for allele-flip SNPs with an effect (top row, right-hand side) and without an effect on the trait (top row, left-hand side), and for SNPs without a flipped allele that affect the trait (middle row, right-hand side) and do not affect the trait (middle row, left-hand side). The two histograms in the bottom row show likelihood ratios after restricting to SNPs with z-scores greater than 2 in magnitude. The bar heights in the histograms in the middle and bottom rows are drawn on the logarithmic scale to better visualize the smaller numbers of SNPs with likelihood ratios greater than 1 (*i.e.*, $\log \mathrm{LR}_j > 0$).

S1 Text. Detailed methods. More description of the methods, including: the single effect regression (SER) model with summary statistics; the IBSS-ss algorithm; computing the sufficient statistics; approaches to dealing with a non-invertible LD matrix; estimation of λ in the regularized LD matrix; likelihood ratio for detecting allele flips; *SuSiE* refinement procedure; detailed calculations for toy example; and more details on the simulations.