SuSiE 2.0: improved methods and implementations for genetic fine-mapping and phenotype prediction

Alexander McCreight[1,2,†], Yanghyeon Cho[1,3,†], Ruixi Li[1], Daniel Nachun[4], Hao-Yu Gan[1], Peter Carbonetto[5], Matthew Stephens[5,6,*], William R.P. Denault[5,7,8,*], and Gao Wang[1,2,9,*]

[1]Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
[2]Department of Biostatistics, Columbia University, New York, NY, USA
[3]Department of Mathematics and Statistical Science, University of Idaho, Moscow, ID, USA
[4]Department of Genetics, Stanford University, Stanford, CA, USA
[5]Department of Human Genetics, The University of Chicago, IL, USA
[6]Department of Statistics, The University of Chicago, IL, USA
[7]Data Science Institute, University of Chicago, IL, USA
[8]Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
[9]Department of Neurology, Columbia University, New York, NY, USA

***Corresponding authors:**
wang.gao@columbia.edu, wdenault@uchicago.edu, mstephens@uchicago.edu

Abstract

Sum of Single Effects regression (SuSiE) has become widely adopted for genetic fine-mapping, yet its original implementation faces architectural limitations that hinder extensibility and performance. We present SuSiE 2.0, featuring a modular redesign for extensibility, up to 5x speed improvements for summary statistics applications, and several useful extensions including SuSiE-ash, a new method that improves calibration when strong signals coexist with moderate effects. Simulations and real data benchmarks demonstrate performance across diverse genetic architectures, highlighting improved calibration of SuSiE-ash for fine-mapping under complex polygenic backgrounds with 1.5–3x FDR reduction while maintaining power, and revealing SuSiE-based methods as effective yet underappreciated tools for TWAS prediction.

Keywords: fine-mapping, TWAS, Bayesian variable selection, eQTL, polygenic architecture, R package

Background

Sum of Single Effects regression (SuSiE) [1] has emerged as a powerful Bayesian variable selection tool, producing posterior inclusion probabilities (PIPs) and single-effect credible sets (CSs) that quantify uncertainty in selected variables. These make SuSiE particularly suited for genetic fine-mapping, where causal signals are sparse yet genetic variables are highly

correlated due to linkage disequilibrium, and capturing all potential causal effects is essential for downstream biological interpretation. The susieR package has received over 180,000 downloads on CRAN (as of 2025), and has inspired numerous methodological extensions [2–14], with integration into major analysis pipelines including COLOC [15] and large-scale genetic studies such as UK Biobank [16], GTEx [17], and FinnGen [18].

However, the original susieR implementation suffers from architectural limitations that hinder extensibility and performance, and many extensions exist only as standalone command-line tools [2,6,14,19], making them difficult to integrate into R-based pipelines or benchmark systematically. In applying SuSiE to expression QTL (eQTL) fine-mapping, we observed potential calibration issues under complex genetic architectures where strong regulatory signals coexist with moderate and weak effects. Existing extensions such as SuSiE-inf [2] model a pervasive infinitesimal effects background, but proved overly conservative in our applications.

In this brief report we present SuSiE 2.0, a modular redesign that maintains backward compatibility while enabling seamless integration of extensions and improved performance. Under this framework we developed SuSiE-ash, a new method that places an adaptive shrinkage prior on moderate to weak effects, which proves to improve calibration across diverse genetic architectures. We also incorporate several published extensions [2,7], and demonstrate that SuSiE-based methods serve as effective yet underappreciated tools for TWAS prediction.

Results and discussion

Figure 1A illustrates the SuSiE 2.0 architecture, which organizes the computational workflow into four stages: interface, constructor, workhorse, and refinement. User-facing functions accept individual-level data, sufficient statistics, or summary statistics, harmonized into a common internal representation before executing Iterative Bayesian Stepwise Selection (IBSS). This modular design uses S3 generic dispatch to separate data-type specific operations from core algorithm logic, eliminating code duplication while enabling seamless integration of methodological extensions, as many reduce to customizations in Bayes factor computation or residual variance estimation. New features include (1) new prior on residual variance [20] for improved coverage particularly in small samples [7], (2) new model SuSiE-ash using adaptive shrinkage [21,22] to model moderate to weak effects (**Methods** and **Supplementary Notes**), (3) SuSiE-inf for modeling infinitesimal backgrounds [2], (4) up to 5x speed improvements for summary statistics with regularized LD matrices (**Figure S1A**), and (5) enhanced model refinement algorithm (**Figure S1B**), flexible convergence criteria, and additional residual estimation methods for greater robustness. SuSiE 2.0 includes comprehensive unit tests covering 99% of code.

To assess performance, we developed *simxQTL*, an R package implementing diverse genetic architectures for benchmarking (**Methods**). We evaluated SuSiE, SuSiE-ash, and SuSiE-inf using power and false discovery rate (FDR) at 95% credible set (CS) coverage, ROC curves at variant level, and phenotype prediction accuracy as a proxy for TWAS model performance.

Under sparse settings with k = 1–5 causal variants (n=1,000, p=5,000), all methods appear reasonably calibrated (**Figure 1C**, **Figure S2**), with SuSiE achieved the highest power followed closely by SuSiE-ash, while SuSiE-inf was considerably more conservative (**Figure 1B**). At k = 5, SuSiE-inf maintained the lowest FDR whereas SuSiE-ash offered a more favorable power-FDR tradeoff. Prediction accuracy was nearly identical across methods, with SuSiE and SuSiE-ash slightly outperforming SuSiE-inf (**Figure 1D**). For variant-level evaluation, SuSiE-ash achieved the best ROC performance at low false positive rates (FPR), followed by SuSiE, with SuSiE-inf substantially lower.

Under an oligogenic setting more representative of eQTL architecture (3 strong, 5 moderate effect variants and 15 polygenic background effects; **Supplementary Notes S4**), SuSiE-ash maintained power nearly identical to SuSiE while achieving substantially lower FDR; SuSiE-inf remained the most conservative (**Figure 1F–G**, **Figure S3**). Prediction accuracy was comparable across methods, with SuSiE-ash showing a slight advantage (**Figure 1H**). For variant-level ROC performance at low FPR, SuSiE-ash and SuSiE performed similarly, with SuSiE-inf trailing behind (**Figure 1I**). Under settings with stronger infinitesimal backgrounds (**Figure S4–5**), SuSiE-inf, while still conservative in power, achieved the best FDR control and improved ROC performance approaching the other methods, and achieved the best prediction accuracy, though closely followed by SuSiE-ash.

SuSiE's elevated FDR under polygenic architectures can arise from synthetic associations, where non-causal variants accumulate spurious signals through LD with multiple true effect variants. SuSiE interprets these synthetic signals as distinct true effects (**Figure S7**). SuSiE-ash mitigates this by modeling the polygenic background with adaptive shrinkage, attributing this diffuse signal to residual variance rather than credible sets, improving its sensitivity to sparse effects.

We also implemented and evaluated other proposed extensions for improving credible set coverage, including attainable coverage (SparsePro [19]) and Bayesian Linear Programming [23]. Attainable coverage showed limited benefit in our benchmarks (**Figure S8**) but is included in SuSiE 2.0 as a convenient alternative for constructing credible sets at different coverage levels post-analysis when LD matrices are not readily available to implement the *purity* filter. Bayesian Linear Programming provided no improvement and is not included in SuSiE 2.0 (**Supplementary Notes S5**).

While SuSiE-ash was motivated by eQTL fine-mapping where moderate polygenic backgrounds are common, different applications may warrant different approaches. For exploratory genome-wide analysis prioritizing sensitivity, standard SuSiE provides the most signal and can identify candidates for follow-up with SuSiE-ash or SuSiE-inf. For targeted candidate regions, running all three methods helps ensure robustness. Other molecular QTLs and GWAS may exhibit distinct genetic architectures, and method choice ultimately depends on the application and tolerance for false discoveries.

Conclusions

We present SuSiE 2.0, a modular reimplementation of SuSiE that improves extensibility, performance, and calibration for genetic fine-mapping and TWAS prediction. SuSiE-ash addresses elevated FDR under complex genetic architectures by modeling moderate to weak effects through adaptive shrinkage, achieving improved calibration without sacrificing power. The four-stage architecture readily accommodates future extensions such as generalized linear models or integration into a generalized IBSS framework, and developers can build directly on the SuSiE 2.0 codebase to ensure compatibility with existing workflows. The software is available as an R package with comprehensive documentation and unit tests.

METHODS

Overview of *SuSiE-ash* Model

We model the phenotype as sparse effects targeted for fine-mapping and a background of unmappable moderate to weak effects, aiming to improve power and reduce false discovery by capturing variation unexplained by the sparse component:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}), \tag{1}$$

where vector $\boldsymbol{y}$ is mean-centered and $\boldsymbol{X}$ is the standardized $n \times p$ genotype matrix. The sparse component $\boldsymbol{\beta}$ is represented using the Sum of Single Effects model (*SuSiE*) [1],

$$\boldsymbol{\beta} = \sum_{\ell=1}^{L} b_\ell \, \boldsymbol{\gamma}_\ell, \tag{2}$$

where each $\boldsymbol{\gamma}_\ell = (\gamma_{\ell 1}, \ldots, \gamma_{\ell p})$ is a one-hot indicator putative causal variant in the ℓ-th *mappable* effect, and $b_\ell \sim N(0, \sigma_{0\ell}^2)$ (Eqs. S2–S5 in **Supplementary Notes S1**).

The remaining moderate genetic effects, scaled by σ^2, are modeled using an adaptive-shrinkage mixture-of-normals prior,

$$\theta_j \sim \sum_{k=1}^{K} \pi_k \, N(0, \sigma^2 \sigma_k^2), \tag{3}$$

with a fixed variance grid $\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2 = (\sigma_1^2, \ldots, \sigma_K^2)$ and mixture weights $\boldsymbol{\pi}$. Together, the sparse and unmappable components induce a marginal precision structure of the form

$$\Omega = \left(\sigma^2 \boldsymbol{I} + \tau^2 \boldsymbol{X}\boldsymbol{X}^\top\right)^{-1}, \tag{4}$$

where $\tau^2 = \text{var}(\theta_j) = \sigma^2 \sum_k \pi_k \sigma_k^2$ accounts for variations that *SuSiE* (essentially setting $\tau^2 = 0$) cannot fine-map (See **Supplementary Notes S1** for the full description).

Posterior inference proceeds via coordinate-ascent variational inference (VI). Under a mean-field approximation,

$$q(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{\ell=1}^{L} q(\boldsymbol{\beta}^{(\ell)}) \prod_{j=1}^{p} q(\theta_j), \tag{5}$$

the evidence lower bound (ELBO) (Eqs. S8–S9) decomposes into tractable subproblems corresponding to single-effect regression (SER) updates for $\boldsymbol{\beta}$ and normal-means (NM) updates for $\boldsymbol{\theta}$ (Eqs. S13–S14), as outlined below.

Updating $\boldsymbol{\beta}$. Different from *SuSiE*, *SuSiE-ash* updates each single-effect component under a marginal likelihood that incorporates the precision matrix Ω (Eq. 4) following the same formulation used in *SuSiE-inf*. Conditioning on the current variance components σ^2 and τ^2, the ℓ-th effect is

updated using the leave-one-effect-out residual

$$\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell} = \boldsymbol{y} - \sum_{\ell' \neq \ell} \boldsymbol{X}\boldsymbol{\beta}^{(\ell')},$$

together with a marginal likelihood involving Ω, ensuring that the SER update is performed under the correct effective noise structure.

Given this likelihood, *SuSiE-ash* computes the posterior inclusion probabilities $\alpha_{\ell j}$, posterior means $m_{\ell j}$, and variances $\sigma^2_{\ell j}$ for each single-effect component. This Ω-adjusted update mitigates PIP miscalibration under non-sparse architectures and improves the recovery of strong causal signals (Eqs. S13–S14).

Updating $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. Conditioned on the current sparse component, we use a data-driven approach to initialize variance grid $\boldsymbol{\sigma}^2_{\boldsymbol{\theta}}$ (**Supplementary Notes S3**) and update each θ_j using normal-means posterior computations from *Mr.ASH* [22], yielding posterior mixture weights ϕ_{1jk}, shrinkage-adjusted means μ_{1jk}, and variances s^2_{1jk} (Eqs. S19–S21). Mixture proportions are updated as

$$\tilde{\pi}_k = \frac{1}{p} \sum_{j=1}^{p} \phi_{1jk}, \tag{6}$$

which yields the updated precision matrix Ω.

The complete *SuSiE-ash* procedure is summarized in Algorithm 1, with further details in **Supplementary Notes S1–S3**.

Algorithm 1 Iterative Bayesian Stepwise Selection (IBSS) algorithm for *SuSiE-ash*

Require: $\boldsymbol{y}, \boldsymbol{X}, L, \sigma^2, \boldsymbol{\sigma}_{0b}^2$

  while not converged **do**

  Step 1: Update sparse effects

    for $\ell = 1$ to L **do**

      $\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell} = \boldsymbol{y} - \boldsymbol{X} \sum_{\ell' \neq \ell} \bar{\boldsymbol{\beta}}^{\ell'}$

      $(\boldsymbol{\alpha}^\ell, \boldsymbol{m}_1^\ell, \boldsymbol{\sigma}_1^{2,\ell}) \leftarrow$ SER$(\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell}, \boldsymbol{X})$

      $\bar{\boldsymbol{\beta}}^\ell = \boldsymbol{\alpha}^\ell \odot \boldsymbol{m}_1^\ell$

    end for

    $\bar{\boldsymbol{\beta}} = \sum_\ell \bar{\boldsymbol{\beta}}^\ell$

  Step 2: Convergence check using $\boldsymbol{\alpha}$

    $\Delta_\alpha = \max_{i,j} |\alpha_{ij}^{(t)} - \alpha_{ij}^{(t-1)}|$

    if $\Delta_\alpha <$ `tol` **then**

      break

    end if

  Step 3: Update unmappable oligogenic / polygenic effects

    Update (σ^2, τ_0^2) using the MoM estimator for provisional variance components

    Construct the variance grid $\{\sigma_k^2\}_{k=1}^K$ based on $\hat{\tau}_0^2/\hat{\sigma}^2$

    Fit *Mr.ASH* with current σ^2 and variance grid until convergence:

    for $j = 1$ to p **do**

      $\bar{\boldsymbol{r}}_{\theta_j} = \boldsymbol{y} - \boldsymbol{X}\bar{\boldsymbol{\beta}} - \sum_{j' \neq j} \boldsymbol{x}_{j'} \bar{\mu}_{j'}$

      $(\boldsymbol{\mu}_{1j}, \boldsymbol{s}_{1j}^2, \boldsymbol{\phi}_{1j}) \leftarrow$ NM$^{post}(\bar{\boldsymbol{r}}_{\theta_j}, \boldsymbol{X})$

      $\bar{\mu}_j = \sum_k \phi_{1jk} \mu_{1jk}$

    end for

    $\tilde{\pi}_k = \frac{1}{p} \sum_j \phi_{1jk}$

  Step 4: Update variance components via *Mr.ASH*

    Extract updated σ^2 and mixture weights $\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_K)'$

    Update $\tau^2 = \sigma^2 \sum_{k=1}^K \tilde{\pi}_k \sigma_k^2$

    Update precision matrix $\Omega = (\tau^2 \boldsymbol{X}\boldsymbol{X}^T + \sigma^2 \boldsymbol{I})^{-1}$

  end while

  Output: Posterior summaries for $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, mixture weights $\boldsymbol{\pi}$, and variance components (σ^2, τ^2).

Overview of Simulation Study Design

We evaluated *SuSiE*, *SuSiE-ash*, and *SuSiE-inf* using genotype data from UK Biobank. Under sparse settings, we varied the number of causal variants (k=1–5) with fixed per-SNP heritability. Under complex genetic architectures mimicking realistic eQTL settings, we partitioned genetic effects into three components: sparse (3 variants with large effects), oligogenic (5–10 variants with moderate effects), and polygenic background (15 variants with small effects). We also evaluated settings with an infinitesimal background, where instead of polygenic background with limited variants, *all* remaining variants collectively contribute a small portion of heritability. Total heritability was fixed at $h^2 = 0.25$ across scenarios. To facilitate reproducible benchmarking, we implemented these simulation designs in `simxQTL`, an R package providing standardized genetic architectures for systematic evaluation of gene-mapping methods (https://github.com/StatFunGen/simxQTL). See **Supplementary Notes S4** for full simulation details.

Code Availability

SuSiE 2.0 is available in the susieR package (https://github.com/stephenslab/susieR). Simulation functions are provided in the simxQTL package (https://github.com/StatFunGen/simxQTL). Real-data analysis scripts and TWAS weight functions are available at https://github.com/StatFunGen/xqtl-protocol and https://github.com/StatFunGen/pecotmr, respectively. Scripts to reproduce all analyses are available at https://github.com/alexmccreight/susieR2.0-paper.

Acknowledgements

We thank Angela Helfrich and Mark Bronnimann from Amazon Web Services for providing cloud computing support for real-world data analysis. This work was supported in part by NIH grants R01HG002585 and R35GM153249 (to M.S., P.C.), NIH grants R01AG076901 (to G.W., R.L.), R01AG086467 (to A.M., Y.C.), and a grant from the Urbut Family Foundation (to G.W.). This project is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences, LLC program. This research was conducted using data from the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP). We thank the participants and investigators of these studies.

Author Contributions

GW conceived and designed the experiments. GW, WD and MS jointly supervised the research. YC, AM developed the *SuSiE-ash* model, and AM developed SuSiE 2.0 package with input from GW. AM implemented the numerical experiments. RL and AM performed data applications. PC, HG and DN contributed to improvements on the SuSiE 2.0 package. AM, YC and GW wrote the manuscript.

Competing Interests

The authors declare no competing interests.

References

[1] 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: Statistical Methodology. 2020 Dec 1;82(5):1273–300.

[2] Cui R, Elzur RA, Kanai M, Ulirsch JC, Weissbrod O, Daly MJ, et al. Improving fine-mapping by modeling infinitesimal effects. Nature Genetics. 2024 Jan 1;56(1):162–9.

[3] Gao B, Zhou X. MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. Nature Genetics. 2024 Jan 1;56(1):170–9.

[4] Lu Z, Wang X, Carr M, Kim A, Gazal S, Mohammadi P, et al. Improved multiancestry fine-mapping identifies cis-regulatory variants underlying molecular traits and disease risk. Nature Genetics. 2025 Aug 1;57(8):1881–9.

[5] Zhao S, Crouse W, Qian S, Luo K, Stephens M, He X. Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. Nature Genetics. 2024 Feb 1;56(2):336–47.

[6] Yuan K, Longchamps RJ, Pardiñas AF, Yu M, Chen TT, Lin SC, et al. Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. Nature Genetics. 2024 Sept 1;56(9):1841–50.

[7] Denault WRP, Carbonetto P, Li R, The Alzheimer's Disease Functional Genomics Consortium, Wang G, Stephens M. Accounting for uncertainty in residual variances improves calibration for fine-mapping with small sample sizes. bioRxiv. 2025 Jan 1;2025.05.16.654543.

[8] Zou Y, Carbonetto P, Xie D, Wang G, Stephens M. Fast and flexible joint fine-mapping of multiple traits via the Sum of Single Effects model. bioRxiv. 2025 Jan 1;2023.04.14.536893.

[9] Denault WRP, Sun H, Carbonetto P, Liu A, De Jager PL, Bennett D, et al. fSuSiE enables fine-mapping of QTLs from genome-scale molecular profiles. bioRxiv. 2025 Jan 1;2025.08.17.670732.

[10] Rossen J, Shi H, Strober BJ, Zhang MJ, Kanai M, McCaw ZR, et al. MultiSuSiE improves multi-ancestry fine-mapping in All of Us whole-genome sequencing data. medRxiv. 2024 Jan 1;2024.05.13.24307291.

[11] Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nature Genetics. 2020 Dec 1;52(12):1355–63.

[12] Zhang X, Jiang W, Zhao H. Integration of expression QTLs with fine mapping via SuSiE. PLOS Genetics. 2024 Jan 25;20(1):e1010929.

[13] Zou Y, Carbonetto P, Wang G, Stephens M. Fine-mapping from summary data with the “Sum of Single Effects” model. PLOS Genetics. 2022 July 19;18(7):e1010299.

[14] Strober BJ, Zhang MJ, Amariuta T, Rossen J, Price AL. Fine-mapping causal tissues and genes at disease-associated loci. Nature Genetics. 2025 Jan 1;57(1):42–52.

[15] Rasooly D, Peloso GM, Giambartolomei C. Bayesian Genetic Colocalization Test of Two Traits Using coloc. Current Protocols. 2022 Dec 1;2(12):e627.

[16] Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023 Oct 1;622(7982):329–38.

[17] Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature. 2017 Oct 1;550(7675):204–13.

[18] Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023 Jan 1;613(7944):508–18.

[19] Zhang W, Najafabadi H, Li Y. SparsePro: An efficient fine-mapping method integrating summary statistics and functional annotations. PLOS Genetics. 2023 Dec 28;19(12):e1011104.

[20] Servin B, Stephens M. Imputation-Based Analysis of Association Studies: Candidate Regions and Quantitative Traits. PLOS Genetics. 2007 July 27;3(7):e114.

[21] Stephens M. False discovery rates: a new deal. Biostatistics. 2017 Apr 1;18(2):275–94.

[22] Kim Y, Wang W, Carbonetto P, Stephens M. A flexible empirical Bayes approach to multiple linear regression and connections with penalized regression. Journal of Machine Learning Research. 2024;25(185):1–59.

[23] Spector A, Janson L. Controlled Discovery and Localization of Signals via Bayesian Linear Programming. Journal of the American Statistical Association. 2025 Jan 2;120(549):460–71.

FIGURES

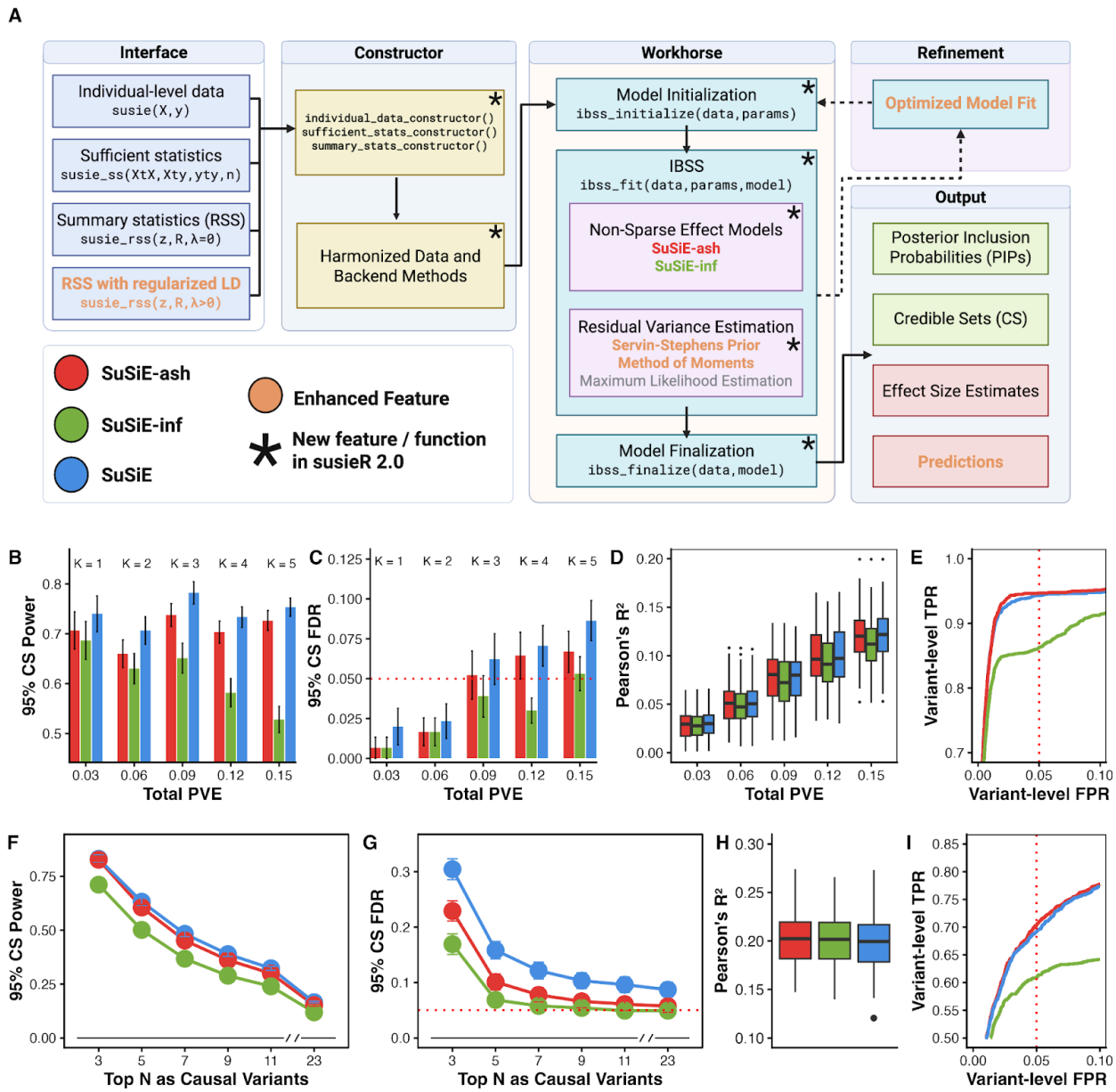

Figure 1. SuSiE 2.0 software architecture and performance across sparse and complex genetic architectures. (A) SuSiE 2.0's modular design unifies individual-level data, sufficient statistics, and summary statistics under a single algorithmic pipeline with two approaches for modeling moderate and weak effects: *SuSiE-ash* and *SuSiE-inf*. **(B-E)** Sparse genetic effects (K = 1, 2, 3, 4, 5 causal variants; n = 1,000, p = 5,000 variants, h^2_{snp} = 0.03; 150 replicates per K). **(B)** 95% credible set (CS) power across varying total proportion of variance explained (PVE). **(C)** 95% CS false discovery rate (FDR) with nominal 5% FDR threshold (dotted red line). **(D)**

5-fold cross-validated prediction accuracy (Pearson's R^2). **(E)** Variant-level ROC curves pooled across all values of K and replicates with 5% false positive rate (FPR) threshold (dotted red line). **(F-I)** Oligogenic effects on polygenic background with total 23 causal variants, mimicking a complex yet realistic cis-eQTL scenario (K = 3 strong effects (50% total PVE), 5 moderate effects (35% total PVE),15 polygenic background (15% total PVE); n = 1,000, p = 5,000, total PVE h^2_g = 0.25; 150 replicates). **(F)** 95% CS power when considering the N strongest simulated effects as causal variants. **(G)** 95% CS FDR across top N causal variant thresholds. **(H)** 5-fold cross-validated prediction accuracy (Pearson's R^2). **(I)** Variant-level ROC curve with 5% FPR (dotted red line) using top 8 variants by effect size magnitude as causal (to cover the simulated strong and moderate effects).

Supplementary Notes

S.1 *SuSiE-ash* Model and Assumptions

SuSiE-ash is a new Bayesian variable selection regression approach that improves fine-mapping by combining *SuSiE* [1] for sparse variable selection and *Mr.ASH* [2] for adapative shrinkage estimation of unmappable effects. The key idea is to iteratively update the strong, sparse effect $\boldsymbol{\beta}$ using *SuSiE* marginalizing over $\boldsymbol{\theta}$ and $\boldsymbol{\epsilon}$, followed by updating the unmappable effects (oligogenic and polygenic) $\boldsymbol{\theta}$ using *Mr.ASH* based on the residual after removing the updated sparse effects.

SuSiE-ash is based upon the following model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{S1}$$

where $\boldsymbol{y}$ is a centered $n \times 1$ vector of phenotype, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p]$ is a standardized $n \times p$ matrix of genotypes for p genetic variants in a genomic region of interest, with $\boldsymbol{x}_j$ being the j-th column of $\boldsymbol{X}$, the p-vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ represent strong spare effect and oligogenic/polygenic effects, respectively, which are independent of each other, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Here, we construct $\boldsymbol{\beta}$ by *SuSiE* to model the strong sparse component [1]. We assume that precisely L variants have a non-zero effect on the outcome:

$$\boldsymbol{\beta} \sim \sum_{\ell=1}^{L} \boldsymbol{\beta}^{(\ell)}, \tag{S2}$$

$$\boldsymbol{\beta}^{(\ell)} = b^{\ell} \boldsymbol{\gamma}^{\ell}, \tag{S3}$$

$$\boldsymbol{\gamma}^{\ell} \sim \text{Mult}(1, \boldsymbol{p}), \tag{S4}$$

$$b^{\ell} \sim N(0, \sigma_{0\ell}^2), \tag{S5}$$

where $\boldsymbol{\beta}^{(\ell)}$ denotes the ℓ-th single-effect vector, $\boldsymbol{\gamma}^{\ell} = (\gamma_1^{\ell}, \ldots, \gamma_p^{\ell})'$ is a p-vector indicating the location of the causal SNP in the ℓ-th single effect, with $\boldsymbol{p} = (p_1, \ldots, p_p)'$ representing the prior weight that sum to 1, and b^{ℓ} is a scalar representing the causal effect size in the ℓ-th single effect. Note that the single-effect regression (SER) model is a special case of the above-specified model when $L = 1$.

Then we construct $\boldsymbol{\theta}$ using an adaptive shrinkage prior for the scaled coefficient ([2], [3]) to model the remaining unmappable effects:

$$\theta_j \sim \sum_{k=1}^{K} \pi_k N(0, \sigma^2 \sigma_k^2), \tag{S6}$$

where $\boldsymbol{\pi} = (\pi_1, ...\pi_K)'$ represents the mixture proportions (non-negative and sum to one), and $\sigma_1^2, ..., \sigma_K^2$ are a non-negative, increasing, pre-specified grid of component variances such that $0 \leq \sigma_1^2, ..., \sigma_K^2 < \infty$ with σ_1^2 set to 0.

Remark 1 (Background variance modeling). Conceptually, *SuSiE-ash* separates sparse effect localization from background variance modeling. Following *SuSiE-inf*, the background component $\boldsymbol{\theta}$ is modeled at the variant level but enters the likelihood only through its marginal variance $\tau^2 = \text{Var}(\theta_j)$, inducing the precision structure $\Omega = (\sigma^2 I + \tau^2 X X^\top)^{-1}$. This formulation allows diffuse polygenic effects to be absorbed as structured background variation while preserving a clear inferential focus on sparse effects. At the same time, by modeling $\boldsymbol{\theta}$ using a flexible shrinkage prior, *SuSiE-ash* enables more accurate estimation of the residual variance σ^2 and background effects, yielding a principled compromise between the overly optimistic behavior of *SuSiE* and the overly conservative behavior of *SuSiE-inf*.

One may consider forming the likelihood using variant-specific background variances obtained from fitting $\boldsymbol{\theta}$ via adaptive shrinkage, yielding a precision matrix of the form $(\sigma^2 \boldsymbol{I} + \boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top)^{-1}$, where $\boldsymbol{D} = \text{diag}(\tau_1^2, \ldots, \tau_p^2)$ and τ_j^2 denotes the posterior variance of the background effect at variant j. However, such a formulation blurs the distinction between sparse and background effects and leads to unstable iterative updates: large τ_j^2 values may reflect either unmappable background signal or moderate effects better attributed to the sparse component, while near-zero τ_j^2 values lead to numerically unstable inversion of the precision matrix. Moreover, because $\boldsymbol{D}$ is repeatedly updated, such formulations preclude reuse of a fixed spectral decomposition of $\boldsymbol{X}\boldsymbol{X}^\top$, requiring repeated large-scale matrix inversions or decompositions and incurring prohibitive computational cost in high-dimensional settings. In contrast, a global background variance stabilizes the separation between sparse effect localization and background variation, with $\boldsymbol{\theta}$ estimated for downstream analyses rather than fine-mapping.

S.2 Variational Inference Framework for *SuSiE-ash*

Note that both *SuSiE* and *Mr.ASH* adopt the variational approximation (VA) method [4] to approximate the posterior distribution under their respective models. By assuming a fully factorized variational approximation, they simplify the optimization of the evidence lower bound (ELBO) over joint prior variables, making it tractable. This tractability is achieved by employing the coordinate ascent algorithm [5], which converts the complex joint optimization problem into a series of simpler tasks. In *SuSiE*, this coordinate-ascent procedure is implemented through the Iterative Bayesian Stepwise Selection (IBSS) algorithm, in which each single-effect component is updated by fitting a single-effect regression (SER) model to partial residuals. Likewise, *Mr.ASH* performs analogous coordinate-wise updates for each coefficient under the normal-mean mixture model (NM).

In *SuSiE-ash*, we similarly assume that the approximation of the joint posterior q for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is factorized as:

$$q(\boldsymbol{\beta}, \boldsymbol{\theta}) = q_{\boldsymbol{\beta}}(\boldsymbol{\beta}) q_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \prod_{\ell=1}^{L} q_{\boldsymbol{\beta}^\ell}(\boldsymbol{\beta}^\ell) \prod_{j=1}^{p} q_{\theta_j}(\theta_j). \tag{S7}$$

Our proposed Algorithm is iteratively optimizing variational approximations $q_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ and $q_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ by

maximizing the following ELBOs, respectively, under *SuSiE-ash* in Eq. (S1):

$$F_{\boldsymbol{\beta}}(\boldsymbol{\beta};\sigma^2,\tau^2,\boldsymbol{\sigma}_{0b}^2) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log\det\Omega + E_{q_{\boldsymbol{\beta}}}\Big[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top\Omega(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\Big] + E_{q_{\boldsymbol{\beta}}}\Big[\log\frac{g_{\boldsymbol{\beta}}(\boldsymbol{\beta})}{q_{\boldsymbol{\beta}}(\boldsymbol{\beta})}\Big] \tag{S8}$$

and

$$F_{\boldsymbol{\theta}}(\boldsymbol{\theta};\sigma^2,\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2,\boldsymbol{\pi},q_{\boldsymbol{\beta}}) = -\frac{n}{2}\log(2\pi\sigma^2) + E_{q_{\boldsymbol{\theta}}}\Big[-\frac{1}{2\sigma^2}(\bar{\boldsymbol{r}}_{\boldsymbol{\theta}}-\boldsymbol{X}\boldsymbol{\theta})^\top(\bar{\boldsymbol{r}}_{\boldsymbol{\theta}}-\boldsymbol{X}\boldsymbol{\theta})\Big] + E_{q_{\boldsymbol{\theta}}}\Big[\log\frac{g_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{q_{\boldsymbol{\theta}}(\boldsymbol{\theta})}\Big], \tag{S9}$$

where $\boldsymbol{\sigma}_{0b}^2 = (\sigma_{01}^2,\ldots,\sigma_{0L}^2)'$ and $\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2 = (\sigma_1^2,\ldots,\sigma_K^2)'$ denote the vectors of single-effect and adaptive shrinkage variances, respectively. We define the precision matrix $\Omega = (\tau^2\boldsymbol{X}\boldsymbol{X}^\top + \sigma^2\boldsymbol{I})^{-1}$, with $\tau^2 = \mathrm{Var}(\theta_j) = \sigma^2\sum_{k=1}^K \pi_k\sigma_k^2$ being the total prior variance of $\boldsymbol{\theta}$. The functions $g_{\boldsymbol{\beta}}$ and $g_{\boldsymbol{\theta}}$ denote the prior distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, respectively. Finally, $\bar{\boldsymbol{r}}_{\boldsymbol{\theta}} = \boldsymbol{y} - \boldsymbol{X}\bar{\boldsymbol{\beta}}$, is the residual after removing the current posterior mean of $\boldsymbol{\beta}$, $\bar{\boldsymbol{\beta}} = E_{q_{\boldsymbol{\beta}}}[\boldsymbol{\beta}]$.

SuSiE-ash is fitted using a generalized iterative Bayesian Stepwise-selection (GIBSS) as outlined in Algorithm 1. Implementation within an iteration loop is outlined below.

Update the strong mappable effect $\boldsymbol{\beta}$. Given variance components σ^2 and τ^2, we update each $q_{\boldsymbol{\beta}^\ell}$, $\ell = 1,\ldots,L$ by fitting the following single-effect regression (SER) model:

$$\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell} := E_{q_{\boldsymbol{\beta}}}\Big[\boldsymbol{y} - \sum_{\ell'\neq\ell}\boldsymbol{X}\boldsymbol{\beta}^{\ell'}\Big] = \boldsymbol{X}\boldsymbol{\beta}^\ell + \tilde{\epsilon}_{\boldsymbol{\beta}}, \tag{S10}$$

where $\tilde{\epsilon}_{\boldsymbol{\beta}} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon} \sim N(0,\Omega^{-1})$. The posterior distribution for $\boldsymbol{\beta}^\ell = b^\ell\boldsymbol{\gamma}^\ell$ under the SER model (Eq. S10) is:

$$\boldsymbol{\gamma}^\ell \mid \boldsymbol{X},\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell},\sigma^2,\tau^2,\boldsymbol{\sigma}_{0b}^2 \sim \mathrm{Mult}(1,\boldsymbol{\alpha}^\ell), \tag{S11}$$

$$b^\ell \mid \boldsymbol{X},\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell},\sigma^2,\tau^2,\boldsymbol{\sigma}_{0b}^2,\gamma_j^\ell = 1 \sim N(m_{1j}^\ell,\sigma_{1j}^{2,\ell}), \tag{S12}$$

where $\boldsymbol{\alpha}^\ell = (\alpha_1^\ell,\ldots,\alpha_p^\ell)'$ is the vector of the posterior inclusion probabilities (PIPs). Note that the posterior distribution for the ℓ-th single effect can be obtained by maximizing the following simpler ELBO:

$$F_\ell(\boldsymbol{\beta}^\ell;\sigma^2,\tau^2,\boldsymbol{\sigma}_{0b}^2) = -\frac{1}{2}E_{q^\ell}\big[-2\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell}^\top\Omega\boldsymbol{X}\boldsymbol{\beta}^\ell + (\boldsymbol{X}\boldsymbol{\beta}^\ell)^\top\Omega\boldsymbol{X}\boldsymbol{\beta}^\ell\big] + E_{q^\ell}\big[\log\frac{g_{\beta^\ell}(\boldsymbol{\beta}^\ell)}{q_{\beta^\ell}(\boldsymbol{\beta}^\ell)}\big] + c_\ell \tag{S13}$$

$$= \sum_{j=1}^p \alpha_j^\ell(\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell}^\top\Omega\boldsymbol{x}_j)m_{1j}^\ell - \frac{1}{2}\sum_{j=1}^p \alpha_j^\ell\boldsymbol{x}_j^\top\Omega\boldsymbol{x}_j(m_{1j}^{2,\ell}+\sigma_{1j}^{2,\ell}) + \sum_{j=1}^p\frac{\alpha_j^\ell}{1}\Big[1+\log\frac{\sigma_{1j}^{2,\ell}}{\sigma_{0\ell}^2} - \frac{m_{1j}^{2,\ell}+\sigma_{1j}^{2,\ell}}{\sigma_{0\ell}^2}\Big] + \sum_{j=1}^p\alpha_j^\ell\log\frac{p_j}{\alpha_j^\ell} + c_\ell, \tag{S14}$$

where c_ℓ is the sum of terms that are not dependent on q^ℓ. Then, by taking partial derivatives of $F_\ell(\boldsymbol{\beta}^\ell; \sigma^2, \tau^2, \boldsymbol{\sigma}_{0b}^2)$ with respect to parameters $\boldsymbol{\alpha}^\ell$, $\boldsymbol{m}_1^\ell = (m_{11}^\ell, \ldots, m_{1p}^\ell)'$, and $\boldsymbol{\sigma}_1^{2,\ell} = (\sigma_{11}^{2,\ell}, \ldots, \sigma_{1p}^{2,\ell})'$, the explicit formulas for this update are expressed as:

$$\alpha_j^\ell = \frac{p_j \exp\left(\log \frac{\sigma_{1j}^\ell}{\sigma_{0\ell}} + \frac{m_{1j}^{2,\ell}}{2\sigma_{1j}^{2,\ell}}\right)}{\sum_{j'=1}^p p_{j'} \exp\left(\log \frac{\sigma_{1j'}^\ell}{\sigma_{0\ell}} + \frac{m_{1j'}^{2,\ell}}{2\sigma_{1j'}^{2,\ell}}\right)} \tag{S15}$$

and

$$m_{1j}^\ell = \bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell}^\top \Omega \boldsymbol{x}_j \times \sigma_{1j}^{2,\ell} \text{ and } \sigma_{1j}^{2,\ell} = \left[\boldsymbol{x}_j^\top \Omega \boldsymbol{x}_j + 1/\sigma_{0\ell}^2\right]^{-1}. \tag{S16}$$

For brevity, we introduce the following function that returns arguments of the posterior distribution of $\boldsymbol{\beta}^\ell$ in Algorithm1:

$$\text{SER}(\bar{\boldsymbol{r}}_{\boldsymbol{\beta},\ell}, \boldsymbol{X}) = (\boldsymbol{\alpha}^\ell, \boldsymbol{m}_1^\ell, \boldsymbol{\sigma}_1^{2,\ell}). \tag{S17}$$

Update the unmappable effects θ and the mixture proportion π in the shrinkage prior. To update $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$, one could theoretically work with the marginal distribution over $\boldsymbol{\beta}$. However, this approach is computationally demanding, particularly for large datasets. As a practical alternative, we leverage the existing *Mr.ASH* model implementation in the `R` package `mr.ash` in a modular fashion incorporated into GIBSS. This update involves fitting the *Mr.ASH* model to the residuals after removing the updated sparse effect, $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_p)'$, from $\boldsymbol{y}$, denoted as

$$\bar{\boldsymbol{r}}_{\boldsymbol{\theta}} = \boldsymbol{y} - \boldsymbol{X}\bar{\boldsymbol{\beta}} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}. \tag{S18}$$

Here, each $\bar{\beta}_j$ is computed as $\bar{\beta}_j = \sum_{\ell=1}^L \alpha_j^{(\ell)} m_{1j}^\ell$.

In a similar manner to *SuSiE*, *Mr.ASH* employs a coordinate-ascent mean field variational inference: (1) the variational distribution of $\boldsymbol{\theta}$ is factorized as $q_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \prod_{j=1}^p q_{\boldsymbol{\theta}_j}(\theta_j)$; (2) the coordinate ascent update for each θ_j, $j = 1, \ldots, p$, is given by computing a posterior distribution under the following normal mean model:

$$\bar{\boldsymbol{r}}_{\theta_j} := E_{q_{\boldsymbol{\theta}_{-j}}}\left[\bar{\boldsymbol{r}}_{\boldsymbol{\theta}} - \sum_{j' \neq j} \boldsymbol{x}_{j'}\theta_{j'}\right] = \boldsymbol{x}_j\theta_j + \epsilon, \tag{S19}$$

where $(\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p)' \sim q_{\boldsymbol{\theta}_{-j}}$. Then, by [2], the posterior distribution for $\boldsymbol{\theta}_j$, denoted as q_{θ_j} under the normal mean model is given by:

$$q(\theta_j | \tilde{\theta}_j; \sigma^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2) = \sum_{k=1}^K \phi_{1jk} N(\mu_{1jk}, s_{1jk}^2), \tag{S20}$$

where

$$\mu_{1jk} = \frac{\sigma_k^2}{\sigma_k^2 + \gamma_j^2}\tilde{\theta}_j, \tag{S21}$$

$$s_{1jk}^2 = \frac{\sigma^2\sigma_k^2\gamma_j^2}{\sigma_k^2 + \gamma_j^2}, \tag{S22}$$

$$\phi_{1jk} = \frac{\pi_k L_{jk}}{\sum_{k=1}^K \pi_k L_{jk}}, \tag{S23}$$

with $\tilde{\theta}_j = (\boldsymbol{x}_j{}'\boldsymbol{x}_j)^{-1}\boldsymbol{x}_j'\bar{\boldsymbol{r}}_{\theta_j}$, $\gamma_j^2 = (\boldsymbol{x}_j'\boldsymbol{x}_j)^{-1}$, and $L_{jk} = N\big(\tilde{\theta}_j; 0, \sigma^2(\sigma_k^2 + \gamma_j^2)\big)$. For future reference, we define the function NM^{post}, which returns the estimated parameters for the posterior distribution of θ_j under the normal mean model:

$$\text{NM}^{\text{post}}\Big(\bar{\boldsymbol{r}}_{\theta_j}, \boldsymbol{X}; \sigma^2, \boldsymbol{\sigma}_{\boldsymbol{\theta}}^2\Big) = (\boldsymbol{\mu}_{1j}, \boldsymbol{s}_{1j}^2, \boldsymbol{\phi}_{1j}), \tag{S24}$$

where $\boldsymbol{\mu}_{1j} = (\mu_{1j1}, \ldots, \mu_{1jK})'$, $\boldsymbol{s}_{1j}^2 = (s_{1j1}^2, \ldots, s_{1jK}^2)'$, and $\boldsymbol{\phi}_{1j} = (\phi_{1j1}, \ldots, \phi_{1jK})'$. We can also obtain update the mixture proportion $\boldsymbol{\pi}$:

$$\tilde{\boldsymbol{\pi}} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_K)',$$

where $\tilde{\pi}_k = \frac{1}{p}\sum_{j=1}^p \phi_{1jk}$.

Remark 2: Note that in *SuSiE-ash* implementation, we used Bayes factor instead of likelihood in (S23) because the two are the same up to a constant term, and Bayes factor aligns with the *Mr.ASH* implementation, providing a numerically stable expression for evaluating the mixture components:

$$\phi_{1jk} = \frac{\pi_k \text{BF}_k}{\sum_{k'} \pi_k' \text{BF}_{k'}}, \tag{S25}$$

where

$$\text{BF}_k = \frac{N\big(\tilde{\theta}_j; 0, \sigma^2(\sigma_k^2 + \gamma_j^2)\big)}{N\big(\tilde{\theta}_j; 0, \sigma^2\gamma_j^2\big)} = \frac{1}{\sqrt{1 + \sigma_k^2/\gamma_j^2}} exp\Big(\frac{\tilde{\theta}_j \mu_{1jk}}{2\sigma^2\gamma_j^2}\Big). \tag{S26}$$

S.3 Implementation Details: Shrinkage Grid Construction and Variance Updates

We first calculate provisional variance components, denoted as $(\hat{\sigma}_0^2, \hat{\tau}_0^2)$, using a method-of-moments (MoM) estimator following *SuSiE-inf*. These values are obtained by plugging the posterior moments of $q_{\boldsymbol{\beta}}$, the variational posterior means and second moments of the single-effect coefficients, into the MoM equations. Specifically, $(\hat{\sigma}_0^2, \hat{\tau}_0^2)$ are computed by solving:

$$E_{q_\beta}\big(\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2\big) = \sigma^2 n + \tau^2 \,\text{tr}(\boldsymbol{X}^T\boldsymbol{X}),$$

$$E_{q_{\boldsymbol{\beta}}}\big[\|\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\|^2\big] = \sigma^2 \,\text{tr}(\boldsymbol{X}^T\boldsymbol{X}) + \tau^2 \,\text{tr}(\boldsymbol{X}^T\boldsymbol{X})^2.$$

Here, $\|\boldsymbol{u}\| = \sqrt{\boldsymbol{u}^T\boldsymbol{u}}$ denotes the Euclidean norm.

At iteration i, $\tau^{2(i)} = \sum_k \tilde{\pi}_k^{(i)}\sigma_k^2$ and the residual variance can be updated as:

$$\sigma^{2(i)} = \frac{\|\bar{\boldsymbol{r}}_\theta - \boldsymbol{X}\bar{\boldsymbol{\theta}}^{(i)}\|^2 + \sum_{j=1}^p \sum_{k=2}^K \phi_{1jk}^{(i)}(z_j + 1/\sigma_k^2)(\mu_{1jk}^{2(i)} + s_{1jk}^{2(i)}) - \sum_{j=1}^p z_j \bar{\theta}_j^{2(i)}}{n + p(1-\pi_1)},$$

where $z_j = \boldsymbol{x}_j'\boldsymbol{x}_j$ and $\bar{\boldsymbol{\theta}}^{(i)} = (\bar{\theta}_1^{(i)}, \ldots, \bar{\theta}_p^{(i)})'$ with $\bar{\theta}_j$ being $\sum_k \phi_{1jk}^{(i)}\mu_{1jk}^{(i)}$, $j = 1, \ldots, p$.

Data-driven variance grid for the Adaptive Shrinkage prior. To construct the variance grid $\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2 = (\sigma_1^2, \ldots, \sigma_K^2)$ for the adaptive shrinkage prior, we divide the range of plausible effect-size variances into three log-spaced regions centered around the moment-based estimate of the polygenic variance, $\hat{\tau}_0^2$. We then generate (i) a dense set of very small to near zero variances ($\hat{\tau}_0^2/100$ to $\hat{\tau}_0^2/10$) to capture very weak effects, (ii) a moderately dense region around $\hat{\tau}_0^2$ ($\hat{\tau}_0^2/10$ to $3\hat{\tau}_0^2$) to model average to moderate unmappable effects, and (iii) a coarse grid from $3\hat{\tau}_0^2$ up to the minimum sparse-effect variance estimated by *SuSiE*, $\min_\ell(\sigma_{11}^2, \ldots, \sigma_{L1}^2)$, to allow for additional larger effects not captured by *SuSiE*; notice that this is also the upper-bound of the grids. These grids are then combined with a point mass at zero to form the final mixture variance grid used in the *Mr.ASH* prior for each iteration.

S.4 Simulation Study Details

Sparse effects simulation design. We conducted extensive simulations to evaluate the performance of *SuSiE*, *SuSiE-ash*, and *SuSiE-inf* under a sparse genetic architecture, with slight modifications to the original *SuSiE* paper's simulation design. These simulations used genotype data from UK Biobank where we randomly sampled 150 LD blocks across chromosomes 1–22 and derived the corresponding genotype matrices to serve as the basis for our simulation. Each genotype matrix, $\mathbf{X} \in \mathbb{R}^{n\times p}$, has $n = 1,000$ and $p = 5,000$. Additionally, we required a minor allele frequency greater than 1%, and a missing rate below 5% (using mean imputation for missing data).

We generated gene expression levels, $\mathbf{y} \in \mathbb{R}^n$, based upon the following linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{S27}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ denotes the effect size vector and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ denotes the noise vector, such that $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_n)$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix. The effect size vector, $\boldsymbol{\beta}$, is constructed by first sampling a set of causal variant indices, $\mathcal{S}$, uniformly at random from $\{1, ..., p\}$. For each causal variant, $j \in \mathcal{S}$, we set $\beta_j = 1$, and for all non-causal variants, $j \notin \mathcal{S}$, we set $\beta_j = 0$. Finally, to achieve the desired heritability-level, h^2, we set the residual variance, σ^2, to solve the following equation:

$$\sigma^2 = \frac{(1-h^2)\text{var}(\mathbf{X}\boldsymbol{\beta})}{h^2} \tag{S28}$$

For each simulated effect, we fixed the per-snp heritability $(h^2_{\text{snp}}) = 0.03$. We then generated

simulated data sets varying the number of causal variants $k = \{1, 2, 3, 4, 5\}$ resulting in $h^2 = \{0.03, 0.06, 0.09, 0.12\}$, respectively. For each of the five scenarios, we generated 150 data sets as replicates.

Complex genetic architecture simulation design. We also evaluated *SuSiE*, *SuSiE-ash*, and *SuSiE-inf* under multiple complex genetic architectures to mimic a realistic eQTL architecture [6]. In these settings, gene expression is influenced both by a core set of variants that exert medium to large effects and by a polygenic background in which i) a number of variants (≈15) each contribute a small amount of phenotypic variants, or ii) all remaining variants each contribute a non-zero amount of variation. We used the same preprocessed genotype matrices from the UK Biobank data described in the sparse simulation design section.

We generated gene expression levels, $\mathbf{y} \in \mathbb{R}^n$, using a linear model that partitions genetic effects into three distinct components: sparse, oligogenic, and polygenic. The sparse component, $\mathcal{S}$, includes $k = 3$ variants which exhibit relatively large effects. These effects are drawn from a normal distribution, $\beta_j \sim \mathcal{N}(0, \sigma_{\mathcal{S}}^2)$, and then scaled to achieve the target heritability $h_{\mathcal{S}}^2$.

The oligogenic component, $\mathcal{O}$, has a small number of variants (5–10) that contribute moderate effects. These effects are modeled using a two-component Gaussian mixture distribution to allow for added variability in their magnitudes:

$$\beta_j \mid z_j \sim \mathcal{N}(0, \sigma_{z_j}^2), \tag{S29}$$

$$z_j \sim \text{Categorical}(\pi_1, \pi_2) \tag{S30}$$

for $j \in \mathcal{O}$, where $z_j \in \{1, 2\}$ denotes the mixture component assignment, σ_1^2 and σ_2^2 are the component-specific variances with $\sigma_1^2 < \sigma_2^2$ and $\pi_1 + \pi_2 = 1$. The oligogenic effects are then scaled to achieve the target heritability $h_{\mathcal{O}}^2$.

The polygenic component, $\mathcal{P}$, models the contribution of remaining causal variants. Effects for this component are drawn from a normal distribution with small variance, $\beta_j \sim \mathcal{N}(0, \sigma_{\mathcal{P}}^2)$ for $j \in \mathcal{P}$, and scaled to achieve the target heritability $h_{\mathcal{P}}^2$. Variants not assigned to any component have effects set to zero.

Collectively, these three mutually exclusive sets form a partition of the complete set of variants, i.e., $\mathcal{S} \cup \mathcal{O} \cup \mathcal{P} = \{1, ..., p\}$. These components comprise the effect vector $\boldsymbol{\beta} \in \mathbb{R}^p$ to create the following linear model:

$$\mathbf{y} = \mathbf{X}\left(\boldsymbol{\beta}_{\mathcal{S}} + \boldsymbol{\beta}_{\mathcal{O}} + \boldsymbol{\beta}_{\mathcal{P}}\right) + \boldsymbol{\epsilon} \tag{S31}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is the standardized genotype matrix and $\boldsymbol{\epsilon} \in \mathbb{R}^n$ is a noise vector with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

The total heritability, h_{total}^2 is partitioned among these components such that

$$h_{\text{total}}^2 = h_{\mathcal{S}}^2 + h_{\mathcal{O}}^2 + h_{\mathcal{P}}^2 \tag{S32}$$

with each h^2 value corresponding to the variance explained by the respective genetic component.

We scaled all effects to achieve their desired heritability proportions. Finally, the residual variance, σ^2, is chosen to ensure that the total variance of **y** reflects the desired total heritability.

In our benchmark, we evaluated three scenarios with varying genetic architectures, all with total heritability $h^2_{\text{total}} = 0.25$ and 150 replicates. The first scenario, oligogenic effects on a polygenic background (**Figure 1F–1I**), included 3 sparse effects (50% of h^2_{total}), 5 oligogenic effects (35% of h^2_{total}), and 15 polygenic effects (15% of h^2_{total}). The second scenario, oligogenic effects on a *moderate* infinitesimal background (**Figure S2 A–D**), maintained the same heritability proportions (50%, 35%, 15%) and variant counts for sparse and oligogenic components, but assigned the remaining heritability to *all* remaining variants. The third scenario, oligogenic effects on an *extensive* infinitesimal background (**Figure S2 E–H**), shifted the architecture toward a stronger infinitesimal contribution: 3 sparse effects (50% of h^2_{total}), 10 oligogenic effects (15% of h^2_{total}), and all remaining variants (35% of h^2_{total}).

S.5 Evaluation of Alternative Credible Set Coverage Methods

We evaluated two additional proposed approaches for improving credible set coverage: attainable coverage [7] and Bayesian Linear Programming (BLiP) [8]. Attainable coverage showed limited benefit in our benchmarks (**Figure S8**), and is included in *SuSiE 2.0* as a convenient alternative for constructing credible sets at different coverage levels post-analysis when LD matrices are not readily available to implement the purity filter. We note that the default entropy threshold parameter (ethresh = 20) performed poorly in both our sparse and complex benchmarks. We suspect this default was likely tuned for sparser marker sets, whereas our simulations use dense markers (p = 5,000). We therefore changed the default in *SuSiE 2.0* to max(100, 0.10p), where p is the number of variables. BLiP provided no improvement over the default *SuSiE* implementation and is not incorporated in *SuSiE 2.0*.

References for Supplementary Notes

[1] G. Wang, A. Sarkar, P. Carbonetto, and M. Stephens. “A simple new approach to variable selection in regression, with application to genetic fine mapping”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.5 (2020), pp. 1273–1300.

[2] Y. Kim, W. Wang, P. Carbonetto, and M. Stephens. “A flexible empirical Bayes approach to multiple linear regression and connections with penalized regression”. In: *Journal of Machine Learning Research* 25.185 (2024), pp. 1–59.

[3] M. Stephens. “False discovery rates: a new deal”. In: *Biostatistics* 18.2 (2017), pp. 275–294.

[4] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. “Variational inference: A review for statisticians”. In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

[5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[6] L. R. Lloyd-Jones, A. Holloway, A. McRae, et al. “The Genetic Architecture of Gene Expression in Peripheral Blood”. In: *American Journal of Human Genetics* 100.2 (Feb. 2017), pp. 228–237.

[7] W. Zhang, H. Najafabadi, and Y. Li. “SparsePro: An efficient fine-mapping method integrating summary statistics and functional annotations”. In: *PLOS Genetics* 19.12 (Dec. 2023). Publisher: Public Library of Science, e1011104.

[8] A. Spector and L. Janson. “Controlled Discovery and Localization of Signals via Bayesian Linear Programming”. In: *Journal of the American Statistical Association* 120.549 (Jan. 2025). Publisher: ASA Website, pp. 460–471. issn: 0162-1459. doi: 10.1080/01621459.2024.2347667.

SUPPLEMENTARY FIGURES

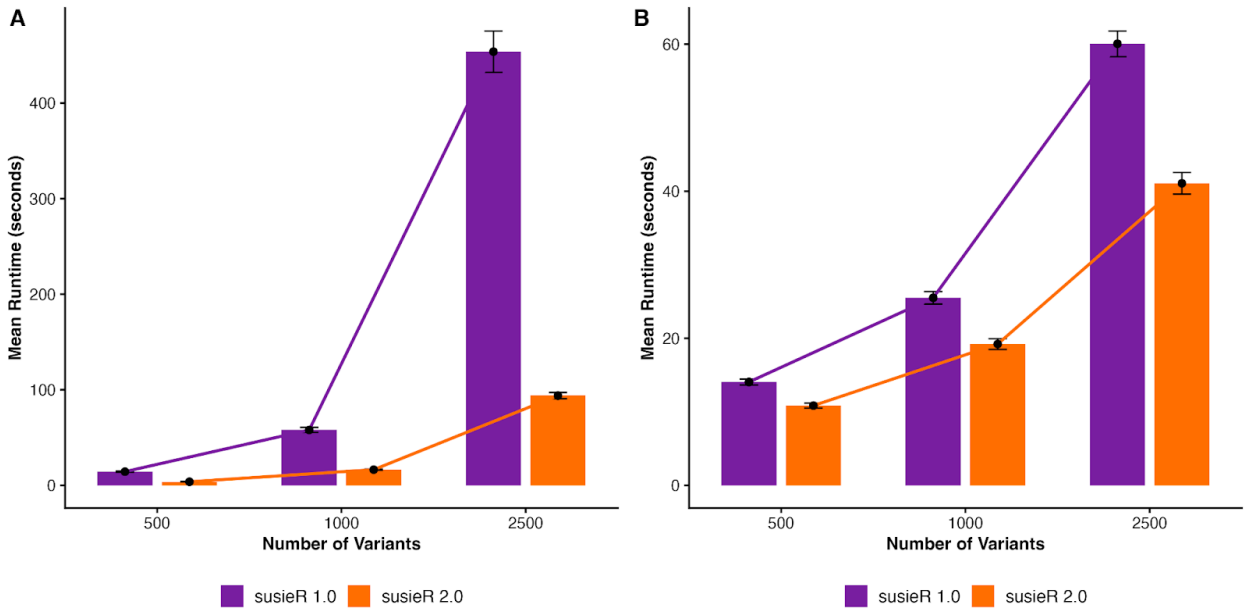

Figure S1. Runtime comparison of *SuSiE-RSS* with regularized LD and refinement between susieR 1.0 vs susieR 2.0. (A-B) Sparse genetic architecture simulation (K = 5 causal variants; n = 1,000, p = 500, 1,000, 2,500 variants, h^2_{snp} = 0.03; 150 replicates per p). **(A)** Performance of RSS with regularized LD. **(B)** Performance of refining model fit.

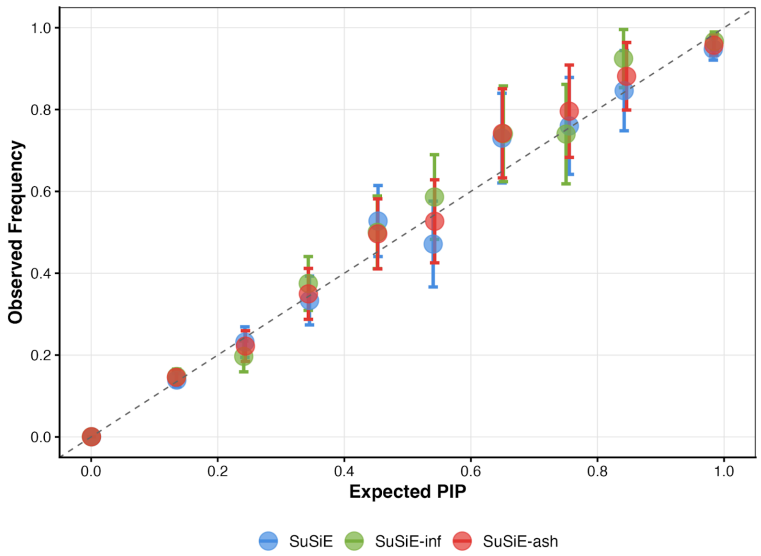

Figure S2. Sparse effects PIP calibration. Expected PIPs vs observed frequencies of true causal variants (diagonal line represents perfect calibration).

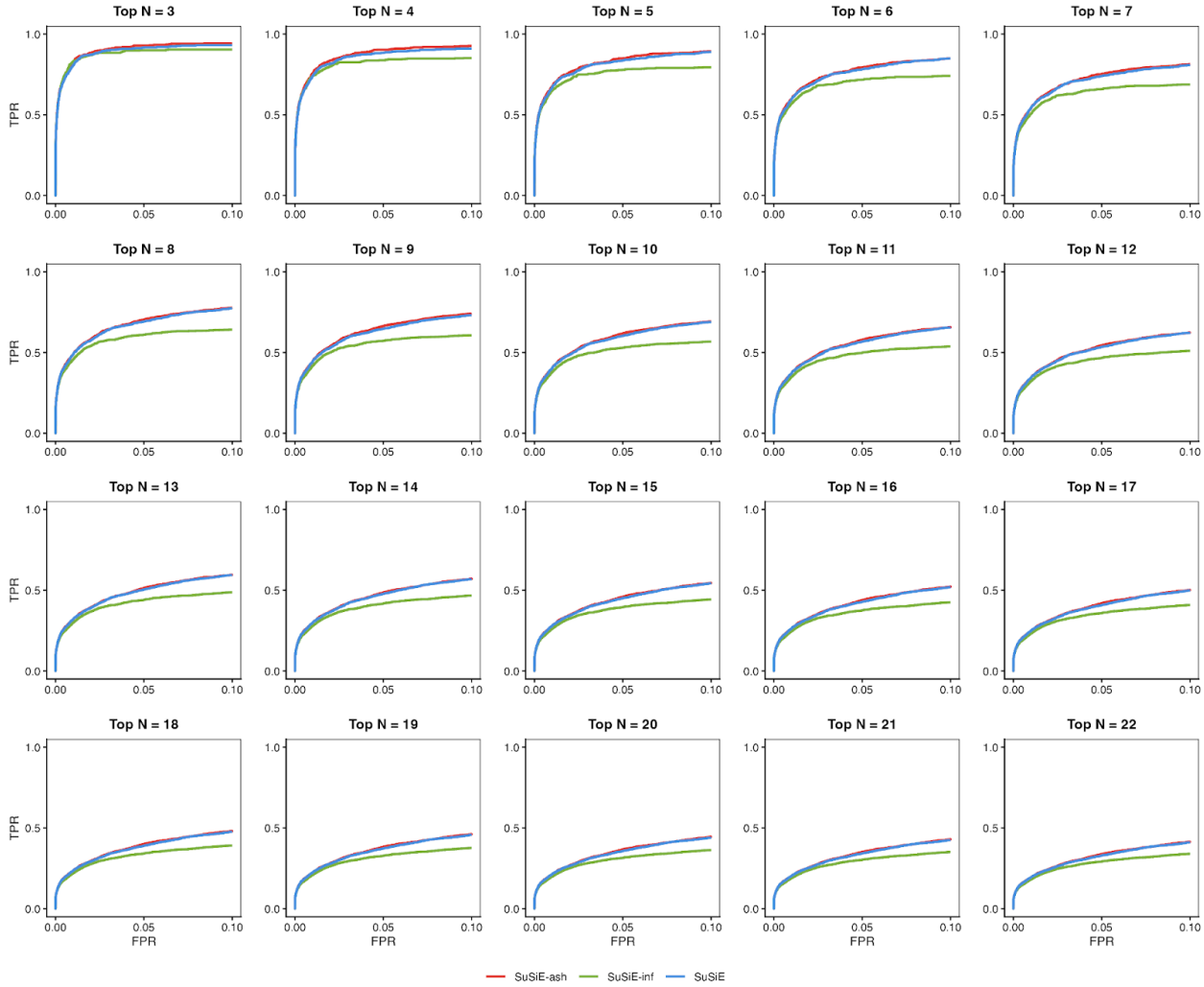

Figure S3. ROC curves for oligogenic effects on a polygenic background using top N = 3 to 22 as causal, matching the “top N as causal” approach to assess FDR and power of 95% CS in Figure 1F and 1G.

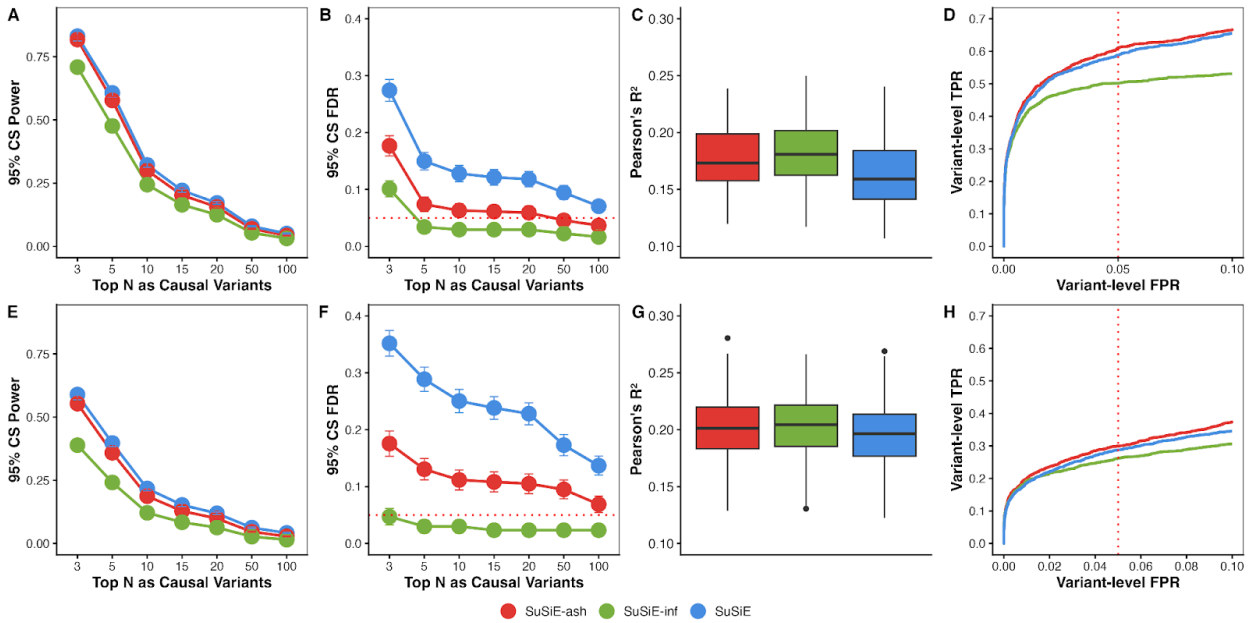

Figure S4. SuSiE 2.0 performance across complex genetic architectures with infinitesimal backgrounds. (A-D) Oligogenic effects on a moderate infinitesimal background (K = 3 sparse effects (50% total PVE), 5 oligogenic effects (35% total PVE), remaining variants (15% total PVE); n = 1,000, p = 5,000, total PVE h^2_g = 0.25; 150 replicates) **(A)** 95% CS power across top N variant thresholds. **(B)** 95% CS FDR across top N causal variant thresholds. **(C)** 5-fold cross-validated prediction accuracy (Pearson's R^2). **(D)** Variant-level ROC curve with 5% FPR (dotted red line) using top 8 variants as causal (to cover simulated strong and moderate effects). **(E-H)** Oligogenic effects on an extensive infinitesimal background (K = 3 sparse effects (50% total PVE), 10 oligogenic effects (15% total PVE), remaining variants (35% total PVE); n = 1,000, p = 5,000, total PVE h^2_g = 0.25; 150 replicates). **(E)** 95% CS power across top N variant thresholds. **(F)** 95% CS FDR across top N causal variant thresholds. **(G)** 5-fold cross-validated prediction accuracy (Pearson's R^2). **(H)** Variant-level ROC curve with 5% FPR (dotted red line) using top 13 variants as causal (to cover simulated strong and moderate effects).

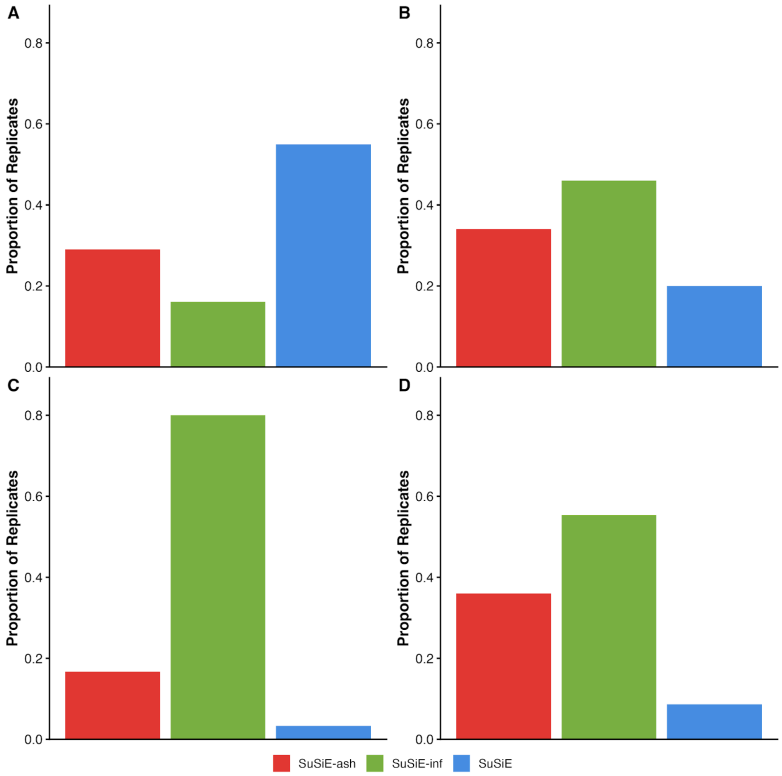

Figure S5. Proportion of replicates where each method achieves the best cross-validated prediction accuracy (Pearson's R^2) across sparse and complex genetic architectures. (A) Sparse effects. **(B)** Oligogenic effects on polygenic background. **(C)** Oligogenic effects on moderate infinitesimal background. **(D)** Oligogenic effects on extensive infinitesimal background.

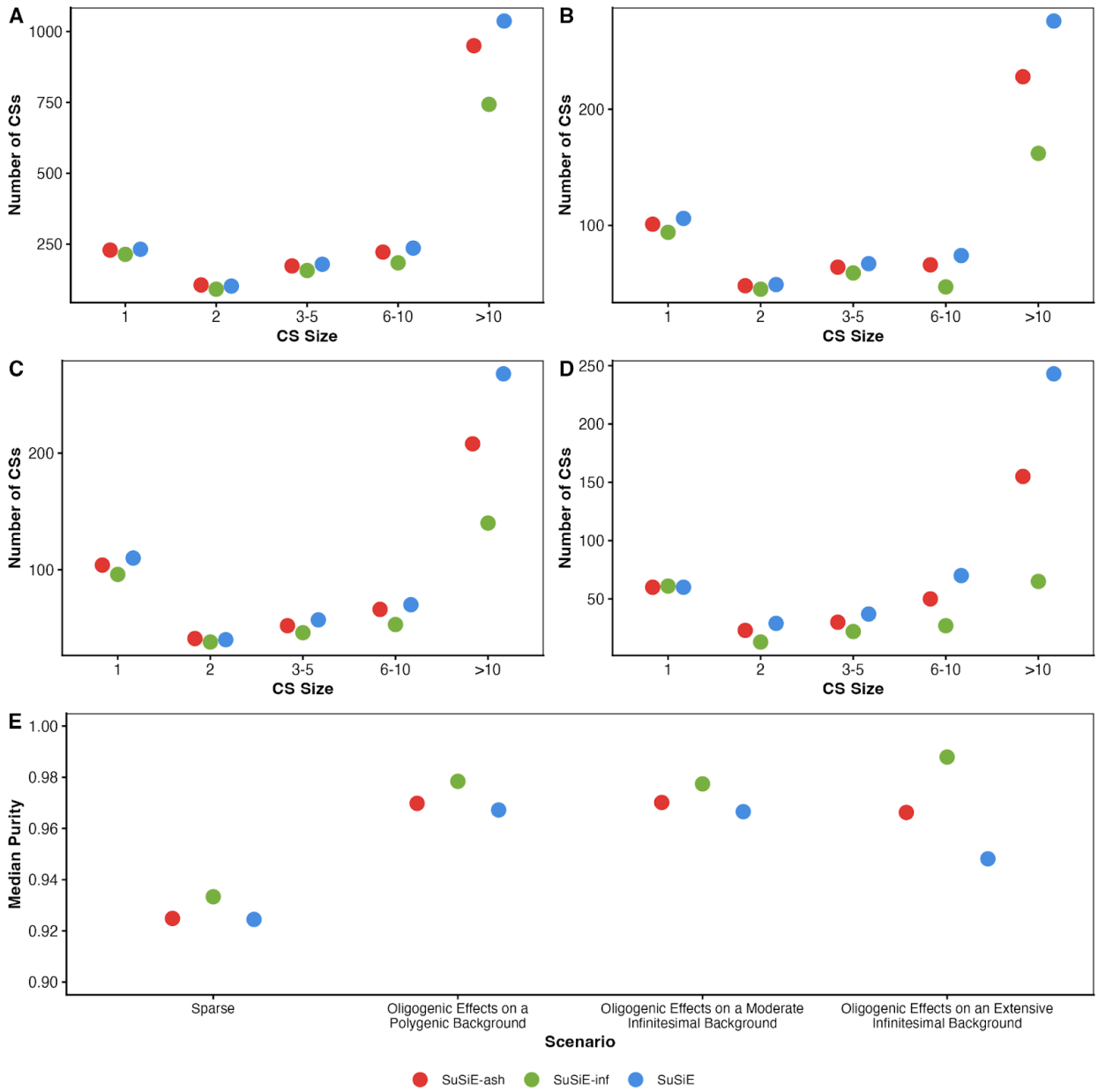

Figure S6. Credible set size and median purity across sparse and complex genetic architectures. (A) Sparse effect CS size. **(B)** Oligogenic effects on a polygenic background CS size. **(C)** Oligogenic effects on a moderate infinitesimal background CS size. **(D)** Oligogenic effects on an extensive infinitesimal background CS size. **(E)** Median purity across different genetic effect architectures.

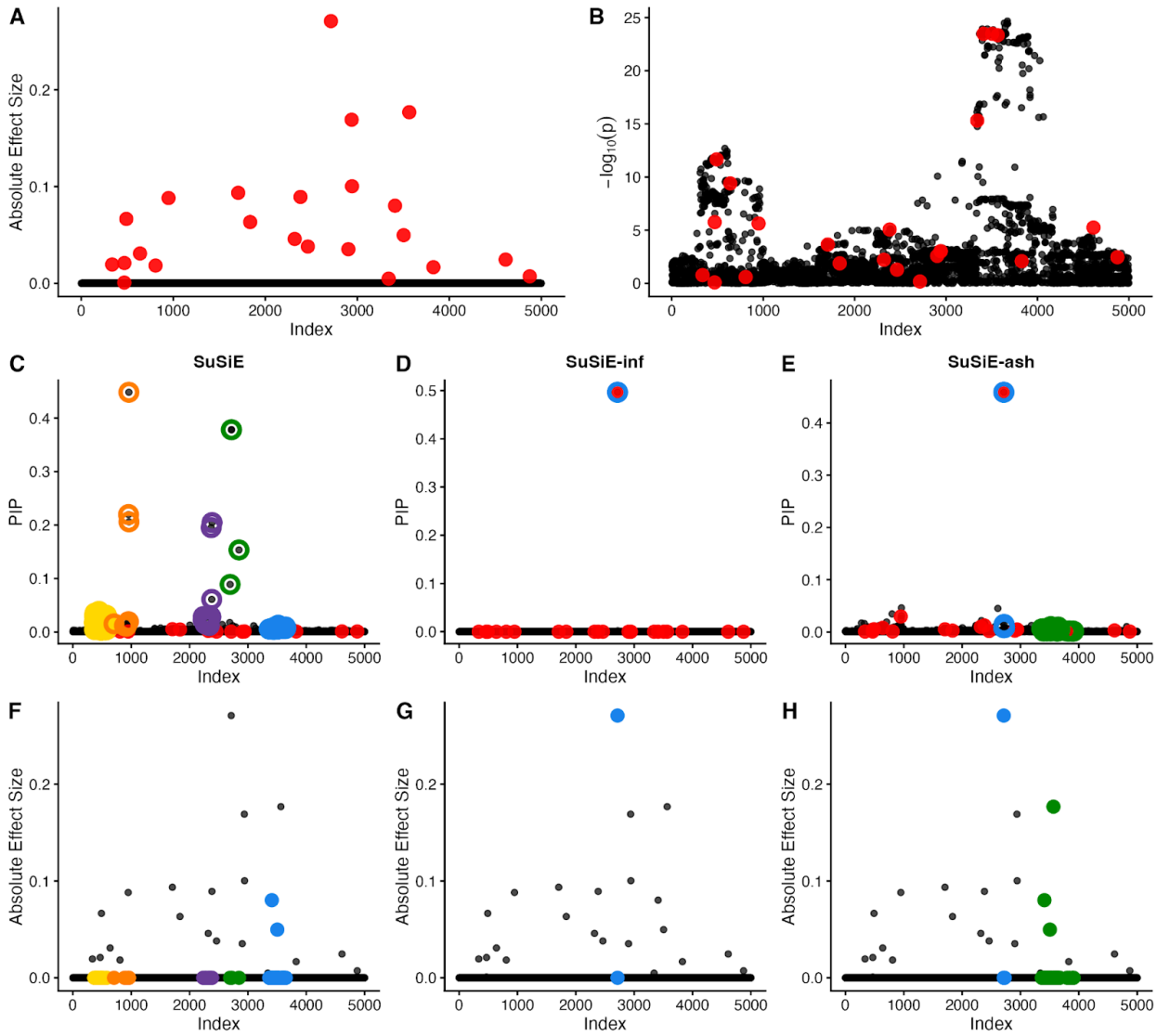

Figure S7. Fine-mapping performance in the presence of correlated polygenic signals. (A-H) Oligogenic effects on polygenic background with total 23 causal variants, mimicking a complex yet realistic cis-eQTL scenario (K = 3 strong effects (50% total PVE), 5 moderate effects (35% total PVE),15 polygenic background (15% total PVE); n = 1,000, p = 5,000, total PVE h^2_g = 0.25. **(A)** Absolute simulated effect sizes with causal variants highlighted in red. **(B)** Marginal association strength (-$\log_{10}$ p-values). **(C-E)** Posterior inclusion probabilities (PIPs) for SuSiE, SuSiE-inf, and SuSiE-ash, respectively; colored rings indicate credible set membership, causal variants highlighted in red. **(F-H)** Credible sets overlaid on absolute effect sizes for SuSiE, SuSiE-inf, and SuSiE-ash, respectively; colored points indicated variants captured in credible sets.

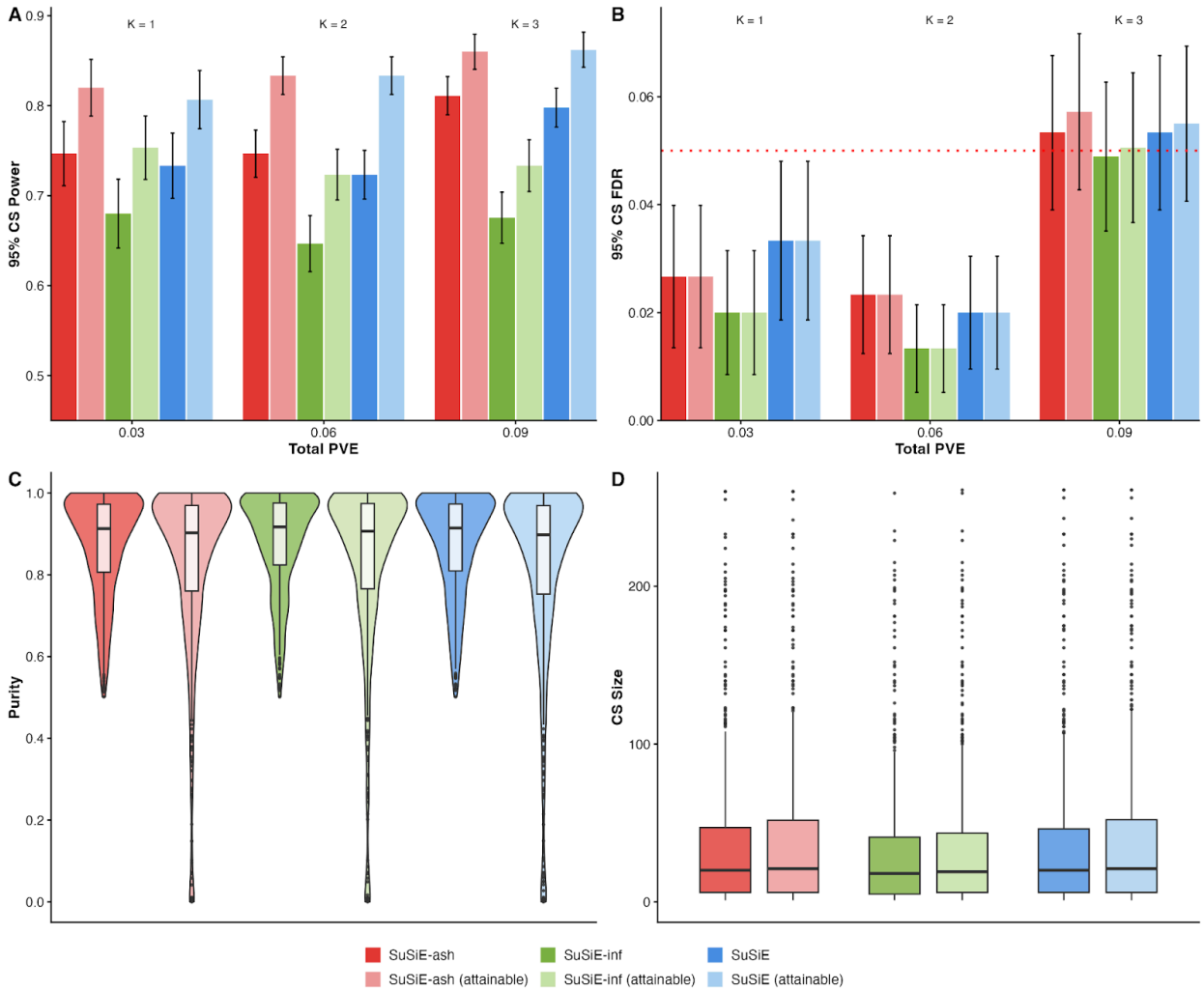

Figure S8. Performance comparison of purity-based and attainable coverage credible sets in sparse fine-mapping scenarios. (A-D) Sparse genetic effects (K = 1, 2, 3 causal variants; n = 1,000, p = 5,000 variants, h^2_{snp} = 0.03; 150 replicates per K). **(A)** 95% CS power across varying total PVE. **(B)** 95% CS FDR with nominal 5% FDR threshold (dotted red line). **(C)** Purity (minimum absolute correlation among variants within each CS) distribution pooled across K, showing many attainable coverage CS are very low in purity. **(D)** Credible set size distribution pooled across K.