Accounting for uncertainty in residual variances improves calibration for fine-mapping with small sample sizes

William R.P. Denault[1], Peter Carbonetto[1], Ruixi Li[2], The Alzheimer's Disease Functional Genomics Consortium, Gao Wang[2,3], and Matthew Stephens[1,4]

[1]Department of Human Genetics, University of Chicago, Chicago, IL, USA
[2]Center for Statistical Genetics, The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA
[3]Department of Neurology, Columbia University, New York, NY, USA
[4]Department of Statistics, University of Chicago, Chicago, IL, USA
[*]Correspondence: William R.P. Denault, wdenault@uchicago.edu; Matthew Stephens, mstephens@uchicago.edu

Abstract

The Sum of Single Effects (SuSiE) model is a widely adopted method for genetic fine-mapping. We show that, in small-sample studies, the original SuSiE fitting procedure produces substantially higher rates of false positive findings. We show that a simple modification to SuSiE improves performance in small-sample studies. This modification is particularly important for emerging molecular QTL applications in rare cell types and primary tissues where sample sizes are inherently limited.

Main

Wang et al [1] introduced the "Sum of Single Effects" (SuSiE) model as a way to capture uncertainty in variable selection in multiple linear regression. This work was motivated by settings like genetic fine-mapping which often involves highly correlated variables, and where it is difficult to confidently identify the *relevant* variables, i.e., those with a non-zero regression coefficient. They pointed out that, even in such settings, it may be possible, and useful, to narrow down the potentially relevant variables by identifying sets of (highly correlated) variables that, with high probability, contain at least one variable with a non-zero coefficient. They refer to these sets of variables as "Credible Sets" (CSs), and the ability of SuSiE to efficiently generate reliable CSs is a key feature of the method. Since its introduction, a number of studies have extended the method in various ways [2–7].

Here we highlight a limitation of the original SuSiE method: it can produce severely miscalibrated CSs when applied to data sets with small sample sizes ($n \leq 50$). We show that this miscalibration occurs even in the simplest case where exactly one variable is assumed to have a non-zero coefficient. While many fine-mapping studies are large, this miscalibration issue is particularly important in smaller studies such as pilot experiments, complex experimental designs (e.g., dynamic eQTLs from [8], $n = 19$, and more recently [9], $n = 53$), costly multimodal-omics experiments (e.g., multimodal QTL mapping in influenza-infected cells [10], $n = 35$), and studies of rare cell types (e.g., eQTL mapping in microglia from the subventricular zone [11], $n = 47$).

To address this limitation, we show that the miscalibration can be greatly improved by a simple modification of the procedures in [1] to account for uncertainty in the residual variance parameter. Specifically, whereas the original SuSiE method estimates the residual variance by maximum likelihood, here we instead integrate it out under a diffuse prior, following ideas from the classical Bayesian literature [12–15]. Breifly, the SuSiE model builds on the simpler "Single Effect Regression" (SER) model, which assumes that exactly one variable has a nonzero effect on the outcome. The original SER formulation from [1] does not account for uncertainty in the residual variance; here we augment it by placing a conjugate inverse-gamma prior on this parameter. This augmented SER is, in essence, the model considered in [13], so we refer to it as the "SS-SER" model (short for "Servin and Stephens").

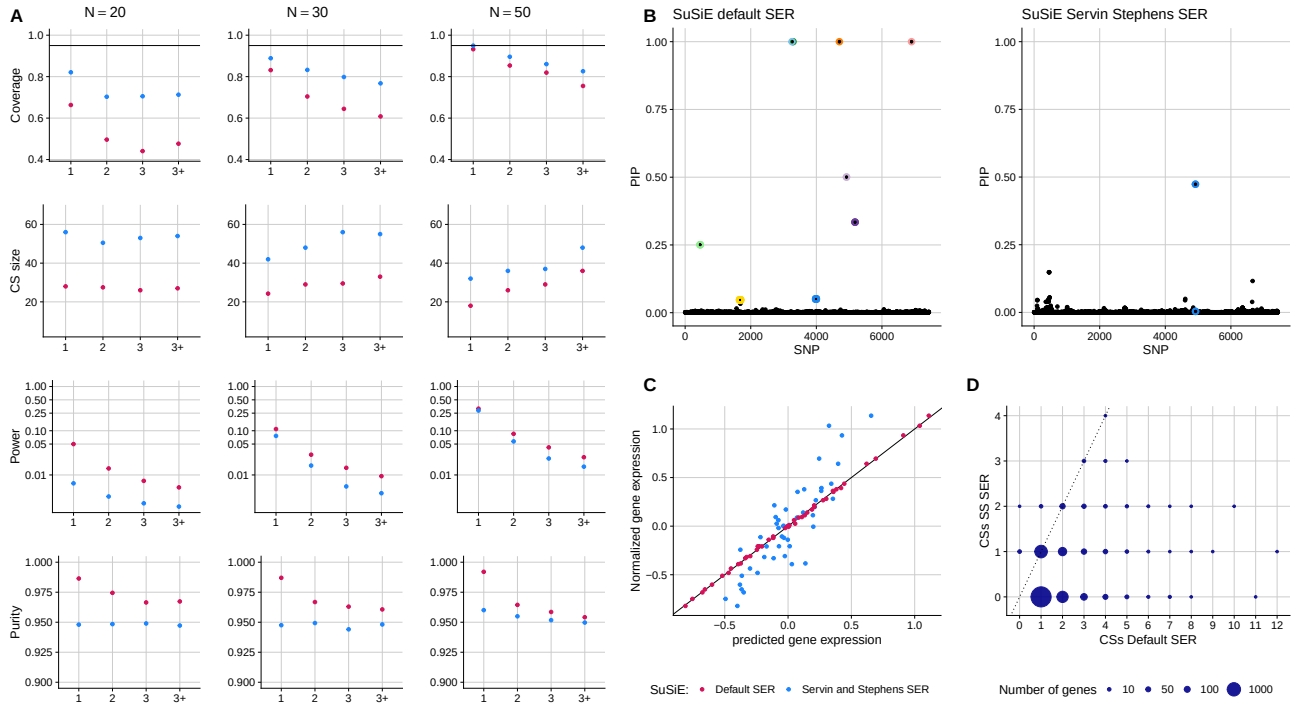

Fig. 1. Panel A: Comparison of 95% CSs from SuSiE with the default SER (red) and SuSiE with the SS-SER (blue) in a simulation setting where the genotype accounts for 25% of the trait variance. The top row displays the observed empirical coverage of the two methods for *N*= 20,30, and 50. The black line represents the target 95% coverage. Coverage was calculated as $1 - \frac{FP}{TP+FP}$ where FP and TP denote the number of false positives and true positives, respectively. The second row gives the mean CS size. The third row gives the power. Note the y-axis is on a logarithmic scale. The fourth row presents the median CS purity. **Panels B, C:** Illustration of severe miscalibration of SuSiE with the default SER in fine-mapping *ADAM17* expression using data from microglia in the subventricular zone from [11], *N* = 47. **Panel B** shows the estimated posterior inclusion probabilities (PIP) and credible sets (CS) for both SuSiE with the default SER and SS-SER. The x-axis represents the SNP index, while the y-axis gives the PIPs. Colors correspond to different CSs. SuSiE with the default SER identifies 9 CSs, whereas SuSiE with SS-SER reports 1 CS (this CS overlaps a CS from SuSiE with default SER). **Panel C** compares the predicted *ADAM17* gene expression under both models against normalized gene expression. The default SER exhibits explains $> 99\%$ of the variance, suggesting overfitting, while the SS-SER model provides a more realistic fit (68% variance explained). **Panel D:** summary of genome-wide fine-mapping using RNA-seq data from microglia in the subventricular zone [11]. It compares of the number of 95% credible sets (CSs) reported by SuSiE using the default SER (x-axis) versus SuSiE with the SS-SER (y-axis). For improved visualization, the point at (0,0) is removed; in 12,232 genes, both methods give no CSs.

This simple modification substantially improves CS calibration in simulations, and provides more realistic results when applying the SuSiE model to small-sample QTL association studies (Fig. 1).

We compared our new approach (SuSiE with the *SS-SER*) with the original approach (SuSiE with the *default SER*) on simulated data. For each simulation, we sampled genotypes by randomly selecting loci from the 1000 Genomes Project Phase 3 [16], including only unrelated individuals. Within these loci, we retained biallelic single nucleotide polymorphisms (SNPs) with minor allele frequencies $\geq 5\%$. From the remaining SNPs, we randomly selected 1–10 causal variants that collectively explained a specified proportion of the phenotypic variance ($h^2 = 25\%, 30\%, 50\%, 75\%$). We evaluated different sample sizes ($N = 20, 30, 50, 75, 100$). Both methods were run using an upper bound of $L = 10$ causal variants, and with the purity filtering step from [1], which removes CSs containing pairs of variables with correlation less than 0.5. See **Methods** for details.

Our results (Fig. 1A, Fig. 2) show that, for small-sample studies ($n \leq 50$), SuSiE with the default SER had consistently poor CS coverage, substantially below the nominal 95% level. By comparison,

SuSiE with SS-SER substantially improved coverage (although not attaining 95%). For example, with $N = 20$ and $h^2 = 0.25$, the default SER CS coverage dropped as low as 44.1%, but increased to 82.1% using the SS-SER (Fig. 1A). This improved coverage occurs because SS-SER produces larger CSs (Fig. 1A, Fig. 3) with lower purity (Fig. 1A, Fig. 4). Consequently, more CSs are discarded by the purity filter, resulting in a corresponding drop in power (Fig. 1A, Fig. 5). This loss of power and localization appear to be an inevitable consequence of improving calibration.

To gauge the impact of estimating the number of causal variants (as well as the purity filtering step), we conducted additional simpler simulations where a single variant affects the trait, assessing CS coverage when (i) allowing SuSiE to estimate the number of causal effects (setting $L = 10$ as above), and (ii) running SuSiE with the correct number of causal effects ($L = 1$). We ran (i) and (ii) both with and without purity filtering. Note that fitting SuSiE with $L = 1$ does not involve any variational approximation, so (ii) removes any potential error from the approximation. Results are summarized in Figures 6 and 7.

In these simpler simulations, the miscalibration issue appeared only with the smallest sample sizes ($n = 20, 30$) and smallest effect sizes ($h^2 = 0.25, 0.3$), and only in the cases with purity filtering. Allowing SuSiE to estimate the number of causal effects had only a minor impact on the calibration in all cases. These results may seem to suggest avoiding the purity filter altogether; however, in practice this produces many "noninformative CSs" that contain numerous variables. These noninformative CSs technically improve calibration bby including almost every variable, and thus likely including a relevant variable by chance. However, outputting many uninformative CSs to improve calibration in this way is typically undesirable. Therefore, in small-sample studies we recommend maintaining the purity filter, using the SS-SER, and bearing in mind that the resulting CSs may still be somewhat less reliable than in large studies.

To compare the methods on real data, we fine-mapped gene expression QTL (eQTL) from two RNA-seq studies. The first dataset from Aracena et al [10] examined the transcriptional response in primary macrophages (n = 35) before and after infection with influenza. The second dataset from Lopes et al [11]investigated eQTLs in microglia, a key cell type involved in brain aging. For each gene, we selected SNPs within ±200 kb of the gene body with MAF > 0.05, and fine-mapped these SNPs using both approaches (default SuSiE and SuSiE with SS-SER).

For the Aracena et al [10] study, across 13,680 fine-mapped genes, default SuSiE reported 1,020 unique CSs, whereas SuSiE with SS-SER reported 389 CSs. Among the genes where SuSiE with SS-SER identified a single CS (381 out of 389), this CS consistently overlapped with a CS reported by default SuSiE. For the Lopes et al [11] study, across 17,132 fine-mapped genes, default SuSiE identified 6,918 CSs, while SuSiE with SS-SER reported 1,556 CSs. Again, nearly all genes with a single CS from SuSiE with SS-SER (844 out of 858) had overlapping CSs from default SuSiE, which exhibited clear signs of miscalibration (Fig. 1B–D), with 146 genes having 4 or more CSs, including up to 12 CSs for *GAL3ST2*. By contrast, SuSiE with SS-SER reported only one gene with 4 CSs.

In conclusion, we have highlighted a critical limitation of the SuSiE method: it produces poorly calibrated CSs in studies with small sample sizes. We present a simple modification that uses the SS-SER to better account for uncertainty in the residual variance parameter, substantially improving CS calibration. Our method retains the computational advantages of SuSiE and can be applied to summary statistics [17], making it practical for a wide range of fine-mapping studies. While our approach improves credible set coverage, some miscalibration remains in certain settings, and further improving performance for small-sample studies remains an area for future work.

References

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

[2] Gao, B. and X. Zhou (2024). MESuSiE enables scalable and powerful multi-ancestry fine-mapping of causal variants in genome-wide association studies. *Nature Genetics 56*, 170–179.

[3] Rossen, J., H. Shi, B. J. Strober, M. J. Zhang, M. Kanai, Z. R. McCaw, L. Liang, O. Weissbrod,

and A. L. Price (2024). MultiSuSiE improves multi-ancestry fine-mapping in All of Us whole-genome sequencing data. *medRxiv*, doi:10.1101/2024.05.13.24307291.

[4] Yuan, K., R. J. Longchamps, A. F. Pardiñas, M. Yu, T.-T. Chen, S.-C. Lin, et al. (2024). Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. *Nature Genetics 56*(9), 1841–1850.

[5] Zou, Y., P. Carbonetto, D. Xie, G. Wang, and M. Stephens (2023). Fast and flexible joint fine-mapping of multiple traits via the Sum of Single Effects model. *bioRxiv*, doi:10.1101/2023.04.14.536893.

[6] Cui, R., R. A. Elzur, M. Kanai, J. C. Ulirsch, O. Weissbrod, M. J. Daly, B. M. Neale, Z. Fan, and H. K. Finucane (2024). Improving fine-mapping by modeling infinitesimal effects. *Nature Genetics 56*, 162–169.

[7] Weissbrod, O., F. Hormozdiari, C. Benner, R. Cui, J. Ulirsch, S. Gazal, A. P. Schoech, B. van de Geijn, Y. Reshef, C. Márquez-Luna, L. O'Connor, M. Pirinen, H. K. Finucane, and A. L. Price (2020, November). Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nature Genetics*, 1–9. Publisher: Nature Publishing Group.

[8] Strober, B., R. Elorbany, K. Rhodes, N. Krishnan, K. Tayeb, A. Battle, and Y. Gilad (2019). Dynamic genetic regulation of gene expression during cellular differentiation. *Science 364*(6447), 1287–1290.

[9] Popp, J. M., K. Rhodes, R. Jangi, M. Li, K. Barr, K. Tayeb, A. Battle, and Y. Gilad (2024). Cell type and dynamic state govern genetic regulation of gene expression in heterogeneous differentiating cultures. *Cell Genomics 4*(12), 100701.

[10] Aracena, K. A., Y.-L. Lin, K. Luo, A. Pacis, S. Gona, Z. Mu, et al. (2024). Epigenetic variation impacts individual differences in the transcriptional response to influenza infection. *Nature Genetics 56*(3), 408–419.

[11] Lopes, K. P., G. J. L. Snijders, J. Humphrey, A. Allan, M. A. M. Sneeboer, E. Navarro, B. M. Schilder, R. A. Vialle, M. Parks, R. Missall, W. van Zuiden, F. A. J. Gigase, R. Kübler, A. B. van Berlekom, E. M. Hicks, C. Böttcher, J. Priller, R. S. Kahn, L. D. de Witte, and T. Raj (2022, Jan). Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies. *Nature Genetics 54*, 4–17.

[12] Gönen, M., W. O. Johnson, Y. Lu, and P. H. Westfall (2005). The Bayesian two-sample t test. *American Statistician 59*(3), 252–257.

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

[14] Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*(481), 410–423.

[15] Rouder, J. N., P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review 16*, 225–237.

[16] The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature 526*(7571), 68–74.

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

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

Code availability. SuSiE with SS-SER is implemented into the original susieR package available at `https://github.com/stephenslab/susieR`. The scripts to replicate these simulations are available at `https://github.com/william-denault/susie_small_sample`.

Methods

Single effect regression accounting for uncertainty in residual variance estimates

The Sum of Single Effects (SuSiE) model is built on the simpler "Single Effect Regression" (SER) model, which assumes that *exactly one* of the variables in the regression has a non-zero effect on the outcome. The SER model from [1] is

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \epsilon \\ \boldsymbol{\beta} &= \boldsymbol{\gamma}\beta \\ \boldsymbol{\gamma} &= \text{Multinom}(1, \boldsymbol{\pi}) \\ \epsilon &\sim N(0, \sigma^2) \\ \beta &\sim N(0, \sigma_0^2\sigma^2), \end{aligned} \tag{1}$$

where $\boldsymbol{y}$ is the vector of observed outcomes, $\boldsymbol{X}$ is a matrix of explanatory variables, $\boldsymbol{\beta}$ is the vector containing the effects of these variables on the phenotype, and ϵ is Gaussian noise with variance σ^2. Note that $\boldsymbol{\gamma}$ is a vector in which exactly one entry is non-zero. The prior probability for which entry is non-zero is encoded by the vector $\boldsymbol{\pi}$ (a vector of non-negative entries that sum to one). The hyperparameter σ_0^2 determines the prior on the effect size relative to the residual variance. In [1], the residual variance σ^2 was estimated by maximizing the likelihood (or, in SuSiE, by maximizing the "evidence lower bound"). Given σ^2 and σ_0^2, it is straightforward to compute the posterior distribution on $\boldsymbol{\beta}$ analytically, essentially by summing over all the possible values for $\boldsymbol{\gamma}$ [1, 13].

To account for uncertainty in small-sample studies, we augment the SER model by placing a (conjugate) inverse-gamma prior on the residual variance:

$$\sigma^2 \sim \text{IG}\left(\tfrac{\kappa}{2}, \tfrac{\lambda}{2}\right). \tag{2}$$

This augmented SER is, in essence, the model considered in [13], and so we call it the "SS-SER" ("SS" being short for "Servin and Stephens"). Given the hyperparameters κ, λ and σ_0^2, it is straightforward to compute the posterior distribution on $\boldsymbol{\beta}$ analytically. Servin and Stephens [13] showed that this posterior distribution has a sensible limit as $\kappa \to 0, \lambda \to 0$, which we adopt for our implementation.

Extension to the SuSiE model: SuSiE with the SS-SER model The SER model is limited because it allows for only a single non-zero effect. Therefore, Wang et al [1] addressed this by allowing for $L \geq 1$ effects through summing L SER models, yielding the "sum of single effects" (SuSiE) model. They developed a fitting procedure, "iterative Bayesian stepwise selection" (IBSS), which fits the SuSiE model by iteratively fitting a sequence of SER models. They showed that this IBSS algorithm, when used with their SER, corresponds to optimizing a variational approximation [18] to the posterior distribution.

Here we use the same IBSS algorithm as in [1], but replace their SER (which fixes σ at a point estimate), with the SS-SER (which integrates out σ); see supplementary materials for details. Unlike the original IBSS algorithm, this modified IBSS is not guaranteed to exactly optimize a variational approximation to the posterior distribution; however, our empirical results show that in small-sample studies it produces better CS calibration while maintaining the computational simplicity of the original IBSS algorithm.

Univariate regression model. For simplicity, we present all derivations assuming κ and λ are strictly positive, ensuring a well-defined prior. However, the quantities can be computed with $\kappa = 0, \lambda = 0$.

Consider the following univariate regression model where σ^2 is treated as unknown:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}\beta + \epsilon \\ \epsilon &\sim N(0, \sigma^2) \\ \beta &\sim N(0, \sigma_0^2\sigma^2) \\ \sigma^2 &\sim \mathrm{IG}(\kappa/2, \lambda/2) \end{aligned} \tag{3}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are in $\in \mathbb{R}^n$, β is an unknown parameter, ϵ is Gaussian noise, and σ_0^2 is a scaling factor on the effect size. $\mathrm{IG}(\kappa, \lambda)$ denotes the inverse gamma distribution with shape κ and scale λ.

The posterior distribution of β conditional on the other inputs and σ^2 is

$$\beta \mid \boldsymbol{x}, \boldsymbol{y}, \sigma^2, \sigma_0^2 \sim N(\mu_1, \sigma^2 s_1^2), \tag{4}$$

where

$$\begin{aligned} s_1^2 &= \frac{r_0^2}{\boldsymbol{x}^T\boldsymbol{x}}, \\ \mu_1 &= r_0^2\hat{b}, \\ r_0^2 &:= \frac{\sigma_0^2}{\sigma_0^2 + 1/(\boldsymbol{x}^T\boldsymbol{x})}. \end{aligned} \tag{5}$$

The posterior distribution of σ^2 is

$$\sigma^2 \mid \boldsymbol{x}, \boldsymbol{y}, \sigma_0^2 \sim \mathrm{IG}\left(\frac{\kappa_1}{2}, \frac{\lambda_1}{2}\right), \tag{6}$$

where

$$\begin{aligned} \kappa_1 &= \kappa + n, \\ \lambda_1 &= \lambda + \mathrm{RSS}(\sigma_0^2), \\ \mathrm{RSS}(\sigma_0^2) &= \boldsymbol{y}^T\boldsymbol{y}\left(1 - r_0^2\sigma_{xy}^2\right), \\ \sigma_{xy} &= \frac{\boldsymbol{x}^T\boldsymbol{y}}{\sqrt{(\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{y}^T\boldsymbol{y})}}. \end{aligned} \tag{7}$$

Thus, the posterior distribution of β is

$$\beta \mid \boldsymbol{x}, \boldsymbol{y}, \sigma_0^2 \sim t_{\kappa_1}\left(\mu_1, \frac{\lambda_1}{\kappa_1}s_1^2\right). \tag{8}$$

The Bayes factor (BF) for this model is given by

$$\mathrm{BF} = \sqrt{1 - r_0^2} \times \left(\frac{\lambda_1}{\lambda + \boldsymbol{y}^T\boldsymbol{y}}\right)^{-\kappa_1/2}. \tag{9}$$

Further details of these derivations are provided in the Supporting derivations.

Bayes factor and posterior distribution for SS-SER model. Under the SS-SER model (1–2), the posterior distribution of $\boldsymbol{\beta}_j$ is a mixture of a point mass at zero and a t-distribution:

$$\boldsymbol{\beta}_j|\boldsymbol{X}, \boldsymbol{y}, \sigma_0^2 = (1 - \pi_{1j})\delta_0 + \pi_{1j}t_{\kappa_{1j}}\left(\mu_{1j}, \frac{\lambda_{1j}}{\kappa_{1j}}s_{1j}^2\right) \tag{10}$$

where $s_{1j}^2, \mu_{1j}, \kappa_{1j}$, and λ_{1j} follow the same definitions as in the univariate case, substituting $\boldsymbol{x}$ with $\boldsymbol{X}_j$ in equation (3). Here, $y = \boldsymbol{X}_j\beta_j + \epsilon$. The mixture weights are given by:

$$\pi_{1j} = \frac{\pi_j \mathrm{BF}_j}{\sum_{j'=1}^{p} \pi_{j'}\mathrm{BF}_{j'}}, \tag{11}$$

with where BF_j is the Bayes factor for the univariate association between column j of $\boldsymbol{X}$ and $\boldsymbol{y}$ (see eq. 9).

Simulations

In all simulations, we randomly sample a genomic region from the 1000 Genomes Phase 3 dataset [16] and select N unrelated individuals within this locus. We retain only biallelic SNPs with a minor allele frequency (MAF) $\geq$5% among the selected individuals.

Multiple causal SNPs simulations. For each simulation, we vary the number of individuals as $N = 20, 30, 50, 75$, and 100 and set the heritability, defined as the variance explained by the genotype, to $h^2 = 25\%, 30\%, 50\%$, and 75%.

For a given sample size N and heritability h^2, we randomly select N individuals as described above. We then randomly determine the number of causal SNPs, L, where L is drawn uniformly from $\{1, \ldots, 10\}$. The phenotype for individual i is generated as: $y_i = \mu_i + \epsilon_i$, where $\mu_i = \sum_{l=1}^{L} x_{il}$. Where x_{il} denotes the genotype at the lth causal SNP for individual i, and ϵ_i is a Gaussian noise term. To ensure that the in-sample variance explained by the genotype matches the specified h^2, we rescale the error term ϵ_i to satisfy the following identity:

$$h^2 = \frac{\sum_{i=1}^{n}(\mu_i - \bar{\mu})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{12}$$

where $\bar{\mu}$ and $\bar{y}$ are the empirical means of μ and y, respectively. In these simulations, we set the upper bound for the number of effects to $L = 10$ when running SuSiE, regardless of the selected single-effect regression (SER) variant, and allow the method to estimate L.

Single causal SNP simulations. Simulations for the single causal SNP scenario follow the same setup as in the multiple causal SNP scenario, except that we fix the number of causal SNPs at $L = 1$. For each combination of N and h^2, we evaluate the performance of SuSiE and its variants using the following configurations:

1. SuSiE with default SER, setting the upper bound on L to 1, with and without purity filtering.
2. SuSiE with default SER, setting the upper bound on L to 10, with and without purity filtering.
3. SuSiE with SS-SER, setting the upper bound on L to 1, with and without purity filtering.
4. SuSiE with SS-SER, setting the upper bound on L to 10, with and without purity filtering.

Additional figures

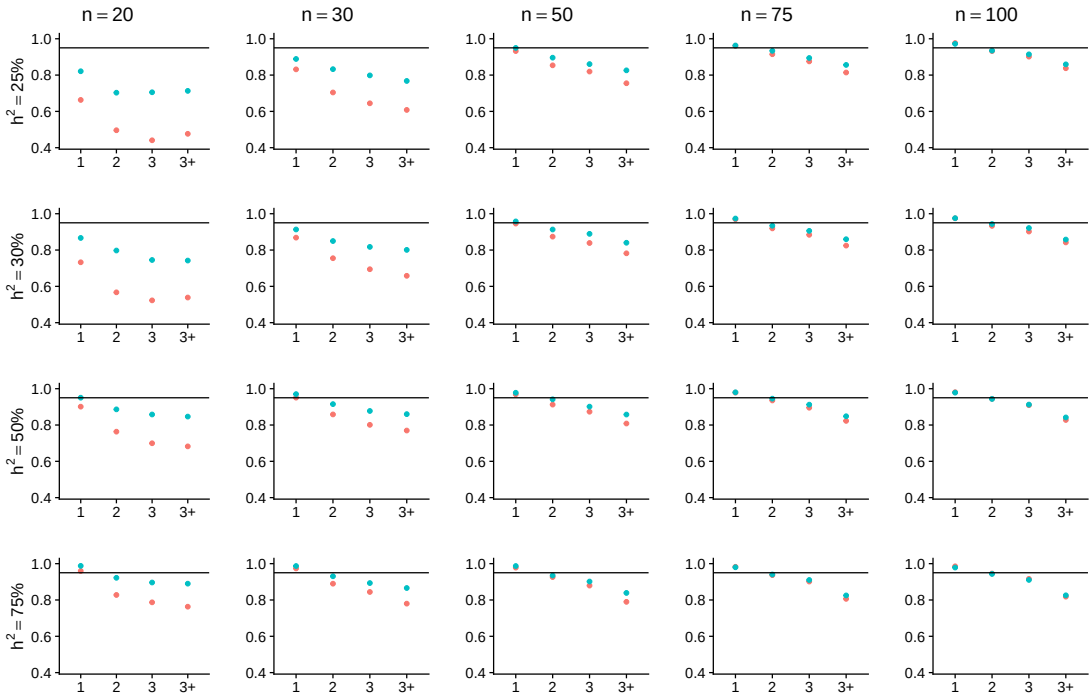

Fig. 2. Comparison of coverage (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue). The black line represents the target 95% coverage. Coverage was calculated as $1 - \frac{FP}{TP+FP}$ where FP and TP denote, respectively, the number of false positives and true positives.

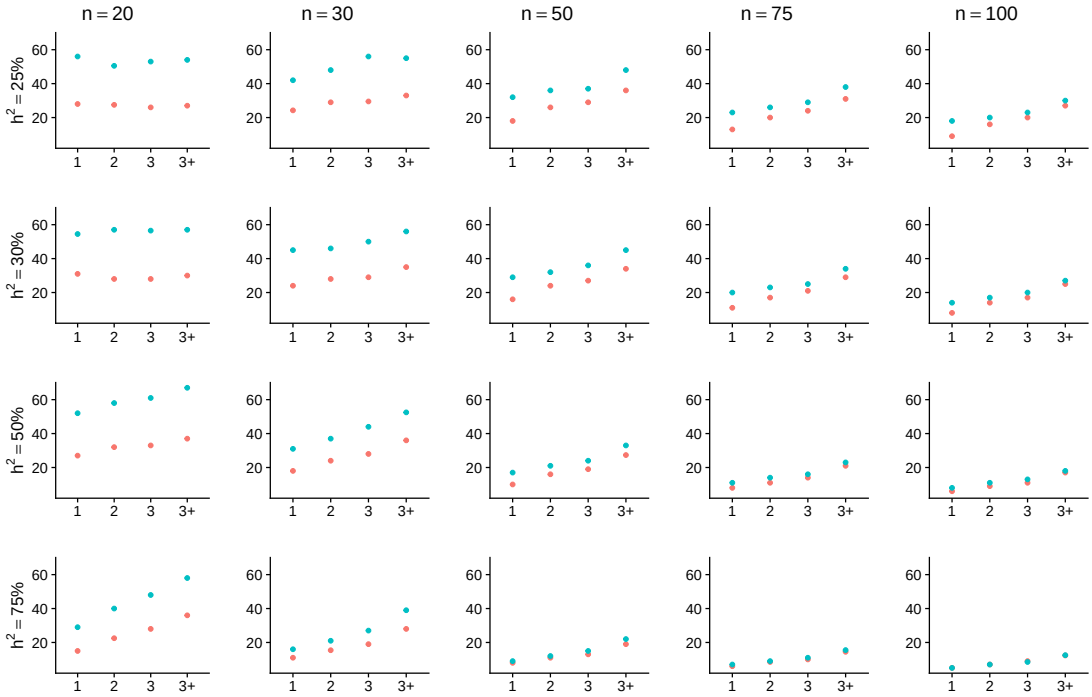

Fig. 3. Comparison of mean CS size (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue).

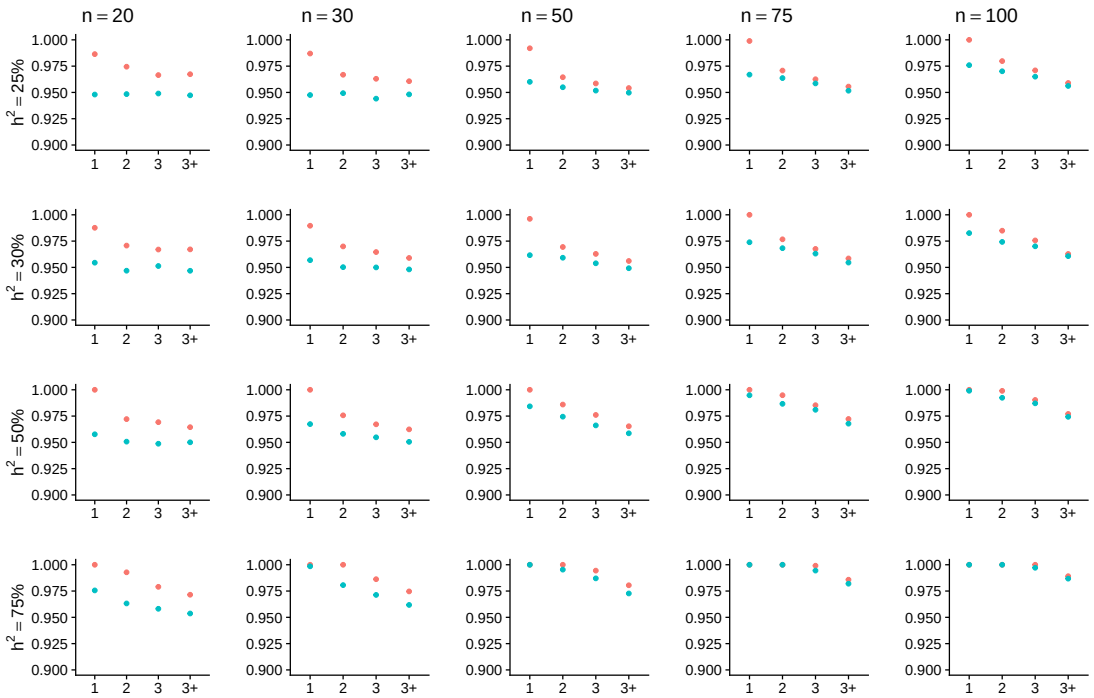

Fig. 4. Comparison of median CS purity (y-axis) for SuSiE with the default SER (red) and SuSiE with the SS-SER (blue).

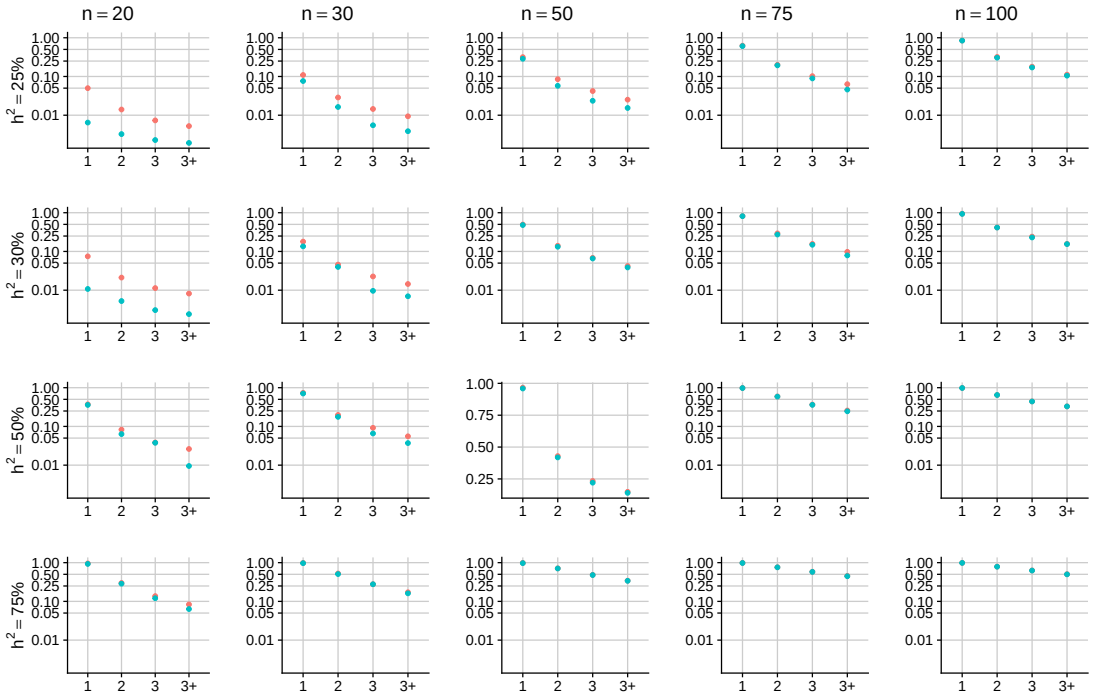

Fig. 5. Comparison of power (y-axis) for SuSiE using the default SER (red) and the SS-SER (blue). Note the y-axis is on a logarithmic scale.

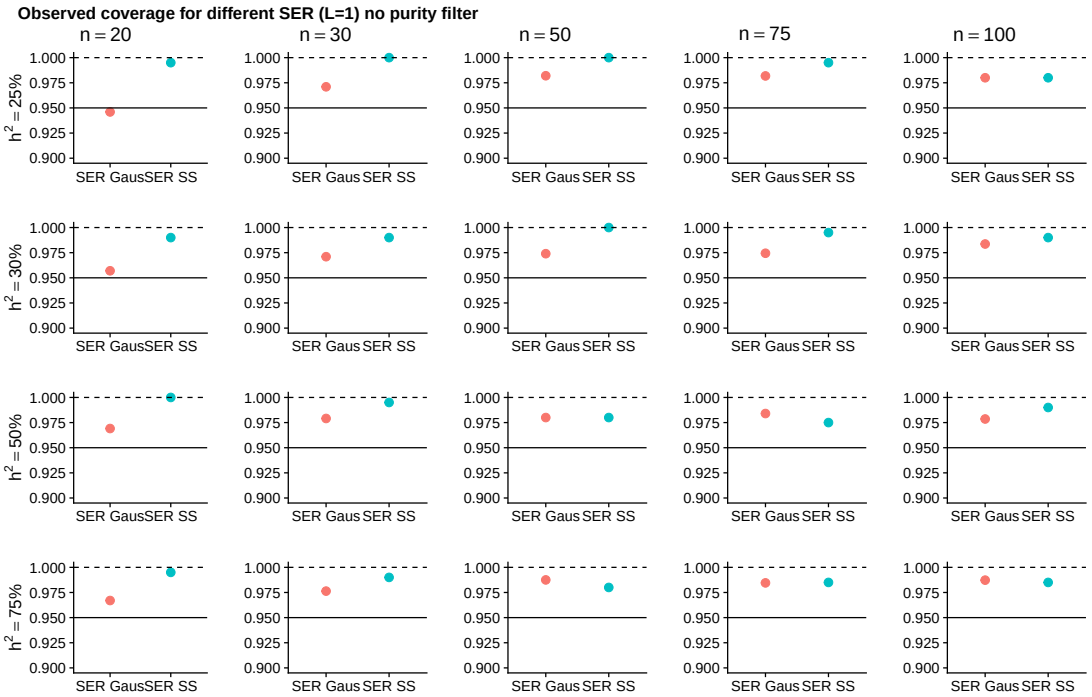

Fig. 6. Comparison of coverage between SuSiE using the default SER (red) and SuSiE using the SS-SER (blue) under a single causal SNP scenario without purity filtering. The solid line denotes the target 95% coverage level, while the dashed line indicates 100% coverage. Simulations were conducted with the $L = 1$, a configuration under which SuSiE computes exact posteriors for both SERs.

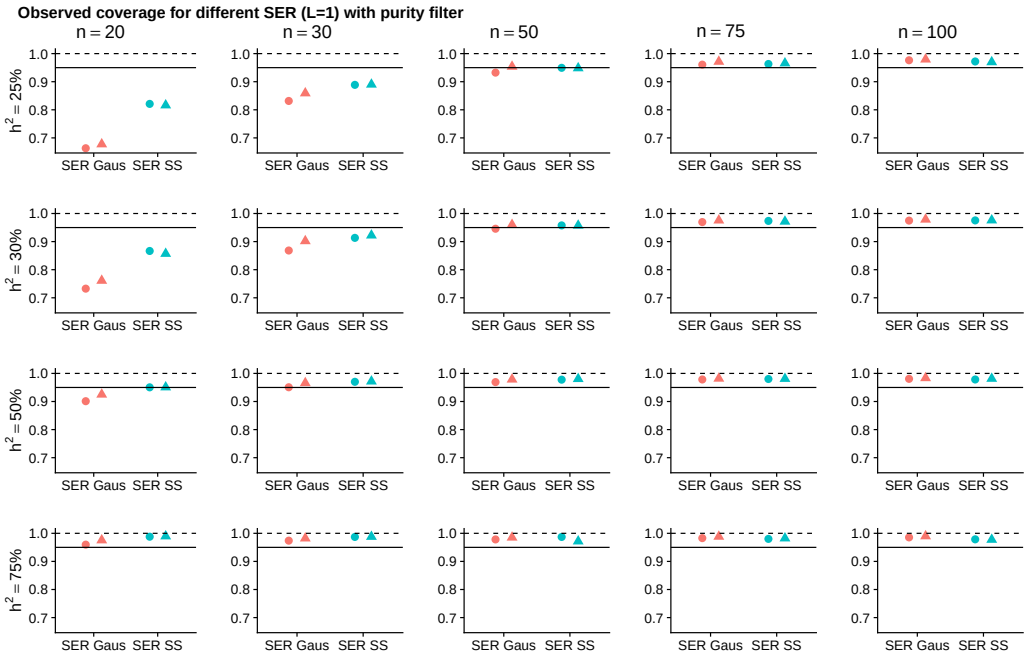

Fig. 7. Comparison of coverage between SuSiE using the default SER (red) and SuSiE using the SS-SER (blue) under a single causal SNP scenario with purity filtering. The solid line denotes the target 95% coverage level, while the dashed line indicates 100% coverage. Round markers represent simulations where SuSiE dynamically estimates the number of causal effects (with an upper bound of $L = 10$), whereas triangle markers correspond to simulations with the L parameter fixed to 1.

Supplementary Notes

Here, we provide detailed derivations of SuSiE with SS-SER model with sample code implementation appended to the end of this document.

Bayes factors

Consider the simple regression model to account for uncertainty in σ^2

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{x}b + \boldsymbol{e} \\ \boldsymbol{e} &\sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \\ b &\sim N(0, \sigma_0^2) \\ \sigma^2 &\sim \mathrm{IG}(\tfrac{\alpha}{2}, \tfrac{\beta}{2}) \\ \sigma_0^2 &= \sigma^2 \tilde{\sigma}_0^2, \end{aligned} \tag{13}$$

where $\mathrm{IG}(a, b)$ denotes the inverse gamma distribution with shape a and scale b, which has probability density function

$$\mathrm{IG}(x; a, b) = \Psi(a, b)\, x^{-(a+1)} e^{-b/x}, \tag{14}$$

and we note $\Psi(a, b) := \frac{b^a}{\Gamma(a)}$ the IG normalizing factor. We note $\boldsymbol{x} := (x_1, \ldots, x_n)^T \in \mathbf{R}^n$, $\boldsymbol{y} := (y_1, \ldots, y_n)^T \in \mathbf{R}^n$ and $\boldsymbol{e} := (e_1, \ldots, e_n)^T \in \mathbf{R}^n$, $b \in \mathbf{R}$ is the regression coefficient, $\sigma^2 > 0$ is the residual variance, $\sigma_0^2 > 0$ is the prior variance of the regression coefficient b, $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, $N(\mu, \sigma^2)$ denotes the univariate normal distribution with mean $\mu \in \mathbf{R}$ and variance $\sigma^2 > 0$, and $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal distribution with mean $\boldsymbol{\mu} \in \mathbf{R}^d$, and $d \times d$ covariance matrix $\boldsymbol{\Sigma}$. Here we assume that both $\boldsymbol{x}$ and $\boldsymbol{y}$ are centered to have mean zero, which removes the need for an intercept in the model [19]. In practice, we scale the prior variance of b by the sample variance of $\boldsymbol{y}$ to ensure that the prior is invariant to the scale of $\boldsymbol{y}$, but we can consider this to be "baked in" to the parameter $\tilde{\sigma}_0^2$, which is assumed to be known and is considered a fixed quantity in most of these derivations.

Given σ^2 and $\tilde{\sigma}_0^2$, the posterior computations for this model are simple: they can be conveniently written in terms of the usual least-squares estimate of b, its variance, and the z-score, which are

$$\hat{b} := \frac{\boldsymbol{x}^T \boldsymbol{y}}{\boldsymbol{x}^T \boldsymbol{x}} \tag{15}$$

$$s^2 := \frac{\sigma^2}{\boldsymbol{x}^T \boldsymbol{x}} \tag{16}$$

$$z := \hat{b}/s. \tag{17}$$

Note that scaling the prior variance of b by σ^2 is important to obtain analytical expressions for marginal likelihoods and Bayes factors.

A few side notes:

1. Placing an inverse gamma prior on σ^2 is equivalent to a gamma prior on the "precision" $\sigma^2 = 1/\sigma^2$ with shape a and inverse scale b.

2. The mean of $\mathrm{IG}(a, b)$ is $\frac{b}{a-1}$ (for $a > 1$), its variance is $\frac{b^2}{(a-1)^2(a-2)}$ (for $a > 2$), and the mode is $\frac{b}{a+1}$.

3. The "noninformative" prior for σ^2 is recovered with $\alpha \to 0, \beta \to 0$, and this was the prior used in [14]. For simplicity, we assume α and β are strictly positive so that the prior is well-defined. However, the derived quantities can be computed with $\alpha = 0, \beta = 0$.

From [13], the posterior distribution of $b \mid \sigma^2$ is

$$b \mid \boldsymbol{x}, \boldsymbol{y}, \sigma^2, \tilde{\sigma}_0^2 \sim N(\mu_1, \sigma^2 s_1^2), \tag{18}$$

where

$$s_1^2 = \frac{r_0^2}{\boldsymbol{x}^T\boldsymbol{x}} \tag{19}$$

$$\mu_1 = r_0^2 \hat{b} \tag{20}$$

$$r_0^2 := \frac{\tilde{\sigma}_0^2}{\tilde{\sigma}_0^2 + 1/(\boldsymbol{x}^T\boldsymbol{x})}. \tag{21}$$

Writing the posterior mean and variance in this way is helpful for intuition: as n becomes large, the "prior" term r_0^2 approaches 1 (because $1/(\boldsymbol{x}^T\boldsymbol{x})$ gets small), and therefore $\mu_1, \sigma^2 s_1^2$ approach the least-squares estimate $\hat{b}$ and its variance s^2. Also note that when $\tilde{\sigma_0}^2 = 0$, $r_0^2 = 0$.

The posterior distribution of $\sigma^2 \mid b$ is

$$\sigma^2 \mid \boldsymbol{x}, \boldsymbol{y}, b, \tilde{\sigma}_0^2 \sim \mathrm{IG}\big(\tfrac{\alpha+n}{2}, \tfrac{\beta+\mathrm{RSS}}{2}\big), \tag{22}$$

where $\mathrm{RSS} = \|\boldsymbol{y} - \boldsymbol{x}b\|_2^2 = (\boldsymbol{y} - \boldsymbol{x}b)^T(\boldsymbol{y} - \boldsymbol{x}b)$ is the residual sum of squares. It is interesting to note that the posterior mode of σ^2 when $\alpha = 0, \beta = 0$ is simply RSS/n, the mean of the residual squared errors.

The posterior distribution of σ^2 (averaging over b) is

$$\sigma^2 \mid \boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2 \sim \mathrm{IG}\big(\tfrac{\alpha_1}{2}, \tfrac{\beta_1}{2}\big), \tag{23}$$

where

$$\alpha_1 := \alpha + n \tag{24}$$

$$\beta_1 := \beta + \mathrm{RSS}(\tilde{\sigma}_0^2) \tag{25}$$

$$\mathrm{RSS}(\tilde{\sigma}_0^2) := \boldsymbol{y}^T\boldsymbol{y}(1 - r_0^2\sigma_{xy}^2) \tag{26}$$

$$\sigma_{xy} := \frac{\boldsymbol{x}^T\boldsymbol{y}}{\sqrt{(\boldsymbol{x}^T\boldsymbol{x})(\boldsymbol{y}^T\boldsymbol{y})}}. \tag{27}$$

Here, $\mathrm{RSS}(\tilde{\sigma}_0^2)$ can be viewed as a Bayesian analog of the RSS [22]. This expression agrees with [13] (specifically, see eq. 4 in Protocol S1 of [13]). It is interesting to note that when n is large and $a = 0, b = 0$, the posterior mode of σ^2 approaches $\boldsymbol{y}^T\boldsymbol{y}(1 - \sigma_{xy}^2)/n$, which is the sample variance of $\boldsymbol{y}$ scaled by a correction term for the correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$. Also, when $\alpha = 0, \beta = 0, \tilde{\sigma}_0^2 = 0$, the posterior mean of σ^2 is $\frac{\boldsymbol{y}^T\boldsymbol{y}}{n-1}$.

Therefore, the posterior distribution of b is

$$b \mid \boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2 \sim t_{\alpha_1}\big(\mu_1, \tfrac{\beta_1}{\alpha_1} s_1^2\big). \tag{28}$$

From this result, one can easily obtain the posterior mean and variance of b. Note that when $\alpha = 0, \beta = 0$ and n is large, the posterior variance of b approaches

$$\mathrm{Var}(b) \approx \frac{1}{n} \times \frac{\boldsymbol{y}^T\boldsymbol{y}}{\boldsymbol{x}^T\boldsymbol{x}} \times (1 - \sigma_{xy}^2). \tag{29}$$

The Bayes factor for this model is

$$\begin{aligned} \mathrm{BF} &= \frac{p(\boldsymbol{y} \mid \boldsymbol{x}, \tilde{\sigma}_0^2)}{p(\boldsymbol{y} \mid \boldsymbol{x}, b = 0)} \\ &= \frac{\int p(\boldsymbol{y} \mid \boldsymbol{x}, \sigma^2, \tilde{\sigma}_0^2)\, p(\sigma^2)\, d\sigma^2}{\int p(\boldsymbol{y} \mid \boldsymbol{x}, \sigma^2, b = 0)\, p(\sigma^2)\, d\sigma^2}. \end{aligned} \tag{30}$$

The dependencies on α, β are implicit in the expressions above and were not included to simplify notation. Let's denote the marginal likelihood (integrated over σ^2) as

$$I(\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2) := \int p(\boldsymbol{y} \mid \boldsymbol{x}, \sigma^2, \tilde{\sigma}_0^2)\, p(\sigma^2)\, d\sigma^2 \tag{31}$$

Then the Bayes factor for b in model 13 is

$$\text{BF} = \frac{I(\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2)}{I(\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2 = 0)}. \tag{32}$$

This integral works out to

$$I(\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2) = (2\pi)^{-n/2} \times \sqrt{1 - r_0^2} \times \frac{\Psi(\frac{\alpha}{2}, \frac{\beta}{2})}{\Psi(\frac{\alpha_1}{2}, \frac{\beta_1}{2})}, \tag{33}$$

using the α_1 and β_1 defined in (24) and (25). When $\tilde{\sigma}_0^2 = 0$, this simplifies to

$$I(\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2 = 0) = (2\pi)^{-n/2} \times \frac{\Psi(\frac{\alpha}{2}, \frac{\beta}{2})}{\Psi(\frac{\alpha_1}{2}, \frac{\beta + \boldsymbol{y}^T\boldsymbol{y}}{2})}. \tag{34}$$

Therefore, the Bayes factor for this model is

$$\text{BF} = \sqrt{1 - r_0^2} \times \left(\frac{\beta_1}{\beta + \boldsymbol{y}^T\boldsymbol{y}}\right)^{-\alpha_1/2}. \tag{35}$$

This expression agrees with the Bayes factor given in [13] (specifically, see eq. 12 in Protocol S1 of [13]). This also appears to agree with eq. 16 in [22].When $\beta = 0$, the Bayes factor simplifies to the somewhat intuitive expression

$$\text{BF} = \sqrt{1 - r_0^2} \times (1 - r_0^2\sigma_{xy}^2)^{-\alpha_1/2}. \tag{36}$$

Single effect regression

For convenience, we restate the single effect regression (SER) model from [1]. The SER extends (13) to a multiple linear regression model in which exactly one of the p variables affects the outcome, $\boldsymbol{y}$:

$$\begin{aligned}
\boldsymbol{y} &= \mathbf{X}\boldsymbol{b} + \boldsymbol{e} \\
\boldsymbol{e} &\sim N_n(0, \sigma^2\mathbf{I}_n) \\
\boldsymbol{b} &= b\boldsymbol{\gamma} \\
\boldsymbol{\gamma} &\sim \text{Multinom}(1, \boldsymbol{\pi}) \\
b &\sim N(0, \sigma_0^2) \\
\sigma^2 &\sim \text{IG}(\tfrac{\alpha}{2}, \tfrac{\beta}{2}) \\
\sigma_0^2 &= \sigma^2\tilde{\sigma}_0^2,
\end{aligned} \tag{37}$$

where $\mathbf{X} \in \mathbf{R}^{n\times p}$, $\boldsymbol{y} \in \mathbf{R}^n$, $\boldsymbol{\gamma} \in \{0,1\}^p$, and Mulitnom$(n, \boldsymbol{\pi})$ denotes the multinomial distribution with multinomial probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)$ and n trials. We assume $\boldsymbol{y}$ and the columns of $\mathbf{X}$ are centered to have mean zero, which implicitly accounts for an intercept. We will use $\boldsymbol{x}_j$ to denote a column of $\mathbf{X}$. Since our ultimate aim here is to extend this model to account for uncertainty, here we reparameterize the SER model slightly so that the prior variance is scaled by the residual variance, $\sigma_0^2 = \sigma^2\tilde{\sigma}_0^2$. This has no effect on the derivations except for the part where we update σ^2.

From our derivations above, it follows that the posterior distribution of b_j, σ^2 under the SER model has the following form [the dependence on α, β and $\boldsymbol{\pi}$ is implicit]:

$$\begin{aligned}
p(b_j, \sigma^2|\mathbf{X}, \boldsymbol{y}, \tilde{\sigma}_0^2) &= p_{\mathsf{post}}(b_j \mid \sigma^2, \gamma_j = 1)\, p_{\mathsf{post}}(\sigma^2 \mid \gamma_j = 1)\, p_{\mathsf{post}}(\gamma_j = 1) \\
&= (1 - \pi_{1j)}\delta_0 + \pi_{1j}N(b_j; \mu_{1j}, \sigma^2 s_{1j}^2)\,\text{IG}(\sigma^2; \alpha_{1j}, \beta_{1j}),
\end{aligned} \tag{38}$$

where $s_{1j}^2 := s_1^2(\boldsymbol{x}_j, \boldsymbol{y})$, $\mu_{1j} := \mu_1(\boldsymbol{x}_j, \boldsymbol{y})$, $\beta_{1j} := \beta_1(\boldsymbol{x}_j, \boldsymbol{y})$, in which $s_1^2(\boldsymbol{x}, \boldsymbol{y})$, $\mu_1(\boldsymbol{x}, \boldsymbol{y})$, $\beta_1(\boldsymbol{x}, \boldsymbol{y})$ are given in equations (19, 20, 25). Here, we are reusing the expressions we have derived previously, with

one change made to the notation to make explicit the dependence of these expressions on $\boldsymbol{x}, \boldsymbol{y}$. As in the previous section, s_1^2 does not actually depend on $\boldsymbol{y}$, but we write it this way for consistency. The mixture weights are

$$\pi_{1j} = \frac{\pi_j \mathrm{BF}_j}{\sum_{j'=1}^{p} \pi_{j'} \mathrm{BF}_{j'}}, \tag{39}$$

where $\mathrm{BF}_j := \mathrm{BF}(\boldsymbol{x}_j, \boldsymbol{y})$, and $\mathrm{BF}(\boldsymbol{x}, \boldsymbol{y})$ is given by the previous result (35), again with the notation changed slightly to make the dependence on $\boldsymbol{x}, \boldsymbol{y}$ explicit. And $\alpha_{1j} = \alpha + n$ for all $j = 1, \ldots, p$.

Following the previous notation we refer to SER as the function that outputs the quantities to compute the posterior quantities under the SER model with unknown σ^2.

$$\mathrm{SER}(\boldsymbol{X}, \boldsymbol{y}, \sigma^2, \sigma_0^2) = (\boldsymbol{\pi}_1, \boldsymbol{\mu}_1, \boldsymbol{s}_1) \tag{40}$$

Updating $\tilde{\sigma}_0^2$ The only parameter that may need to be updated is the (unscaled) prior variance, $\tilde{\sigma}_0^2$. The EM update for this parameter is $(\tilde{\sigma}_0^2)^{\mathsf{new}} = (\sigma_0^2)^{\mathsf{new}}/\sigma^2$, in which $(\sigma_0^2)^{\mathsf{new}}$ is given by the posterior second moment of b^2. For the SER with unknown σ^2, this works out to

$$(\sigma_0^2)^{\mathsf{new}} = \sum_{j=1}^{p} \pi_{1j} \big(\mu_{1j}^2 + \tfrac{\beta_{1j}}{\alpha_{1j}-2} s_{1j}^2\big). \tag{41}$$

Posterior for σ^2 Following the fact that

$$\begin{aligned} p(b_j, \sigma^2 | \mathbf{X}, \boldsymbol{y}, \tilde{\sigma}_0^2) &= p_{\mathsf{post}}(b_j \mid \sigma^2, \gamma_j = 1)\, p_{\mathsf{post}}(\sigma^2 \mid \gamma_j = 1)\, p_{\mathsf{post}}(\gamma_j = 1) \\ &= (1 - \pi_{1j})\delta_0 + \pi_{1j} N(b_j; \mu_{1j}, \sigma^2 s_{1j}^2)\, \mathrm{IG}(\sigma^2; \alpha_{1j}, \beta_{1j}), \end{aligned} \tag{42}$$

Assuming independence between on $p_{\mathsf{post}}(b_j \mid \sigma^2, \gamma_j = 1)$ leads to the following quantities

$$p(\sigma^2 | \mathbf{X}, \boldsymbol{y}, \tilde{\sigma}_0^2) = \int \prod_j p_{\mathsf{post}}(b_j \mid \sigma^2, \gamma_j = 1)\, p_{\mathsf{post}}(\sigma^2 \mid \gamma_j = 1) p_{\mathsf{post}}(\gamma_j = 1) db_1 \ldots db_p \tag{43}$$

$$= \sum_j \pi_{1j} \mathrm{IG}(\sigma^2; \alpha_{1j}, \beta_{1j}). \tag{44}$$

The "Sum of Single Effects" (SuSiE) model with unknown σ^2

We now extend the SER model with unknown σ^2 to the "Sum of Single Effects" (SuSiE) model with unknown σ^2. Instead of a single variable affecting the outcome $\boldsymbol{y}$, the SuSiE model allows at most L variables to affect $\boldsymbol{y}$:

$$\begin{aligned} \boldsymbol{y} &= \mathbf{X}\boldsymbol{b} + \boldsymbol{e} \\ \boldsymbol{e} &\sim N_n(0, \sigma^2 \mathbf{I}_n) \\ \boldsymbol{b} &= \sum_{l=1}^{L} \boldsymbol{b}^{(l)} \\ \boldsymbol{b}^{(l)} &= b^{(l)} \boldsymbol{\gamma}^{(l)} \\ \boldsymbol{\gamma}^{(l)} &\sim \mathrm{Multinom}(1, \boldsymbol{\pi}) \\ b^{(l)} &\sim N(0, \sigma_{0l}^2) \\ \sigma^2 &\sim \mathrm{IG}\big(\tfrac{\alpha}{2}, \tfrac{\beta}{2}\big) \\ \sigma_{0l}^2 &= \sigma^2 \tilde{\sigma}_{0l}^2. \end{aligned} \tag{45}$$

Algorithm 1 General Iterative Bayesian Stepwise Selection (GIBBS) for Servin and Stephens prior

Require Data $\boldsymbol{X}$, $\boldsymbol{y}$
Require Number of effects, L; initial estimates of hyperparameters σ^2, $\tilde{\sigma}_0^2$
Require A function $\mathrm{SER}(\boldsymbol{X}, \boldsymbol{y}, \sigma^2, \tilde{\sigma}_0^2) \to (\boldsymbol{\pi}_1, \boldsymbol{\mu}_1, \boldsymbol{s}_1)$ that computes the posterior distribution for $\boldsymbol{b}$ under the SER with unknown σ^2 model; see section
Require Initial setting of $\hat{b}$, an estimate of the posterior mean of b_l, for $l = 1, \ldots, L$

repeat
$\hat{\boldsymbol{r}} \leftarrow \boldsymbol{y} - \boldsymbol{X} \sum_{l=1}^{L} \hat{\boldsymbol{b}}_l$ ▷ Compute expected residuals
for $l = 1, \ldots, L$ **do**
  $\hat{\boldsymbol{r}}_l \leftarrow \hat{\boldsymbol{r}} + \boldsymbol{X}\hat{\boldsymbol{b}}_l$ ▷ Disregard lth single effect in residuals
  $\tilde{\sigma}_{0l}^2 \leftarrow \arg\max_{\tilde{\sigma}_0^2} \mathrm{SER}(\hat{\boldsymbol{r}}_l; \sigma^2, \tilde{\sigma}_0^2)$ ▷ Update σ_{0l}^2 (optional; see 41)
  $(\boldsymbol{\pi}_l, \boldsymbol{\mu}_l, \boldsymbol{s}_l^2) \leftarrow \mathrm{SER}(X, \hat{\boldsymbol{r}}_l; \sigma^2, \sigma_{0l}^2)$ ▷ Fit SER to residuals
  $\hat{\boldsymbol{b}}_l \leftarrow \boldsymbol{\pi}_l \circ \boldsymbol{\mu}_l$ ▷ Elementwise multiplication
  $\boldsymbol{b}_l^2 \leftarrow \boldsymbol{\pi}_l \circ (\boldsymbol{s}_l^2 + \boldsymbol{\mu}_l^2)$ ▷ Compute posterior second moments
  $\hat{\boldsymbol{r}} \leftarrow \hat{\boldsymbol{r}}_l - \boldsymbol{X}\hat{\boldsymbol{b}}_l$ ▷ Update expected residuals
end for
until Stopping criterion satisfied or maxiter iterations completed

In the seminal SuSiE work [1], the authors assumed that the residual variance σ^2 is either known or could be estimated via maximum marginal likelihood. They fit the SuSiE model using a variational approximation (VA) of the posterior, where the variational posterior assumes that the L single-effect coefficients are independent under the posterior:

$$q(\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(L)}) = \prod_{l=1}^{L} q_l(\boldsymbol{b}^{(l)}). \tag{46}$$

Wang and colleagues [1] further demonstrated that under this conditional independence assumption, each q_l must be a valid posterior distribution for the corresponding SER model. When extending SuSiE to the case where σ^2 is unknown, the same conditional independence assumption leads to the following variational form:

$$q(\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(L)}, \sigma^2) = \prod_{l=1}^{L} q_l(\boldsymbol{b}^{(l)} \mid \sigma^2) \times q_\sigma(\sigma^2) \tag{47}$$

A key challenge in this variational approach is that $q_\sigma(\sigma^2)$ turns out to be a mixture of $p \times L$ inverse Gamma distributions, making it computationally intractable for efficient evaluation. However, given a known form for $q_\sigma(\sigma^2)$, obtaining $q_l(\boldsymbol{b}^{(l)} \mid \sigma^2)$ is straightforward. Despite this, employing such a VA necessitates numerical shortcuts and additional approximations to maintain computational efficiency. To circumvent these computational difficulties, we adopt an alternative heuristic approach called Generalized Iterative Bayesian Stepwise Selection (GIBSS) for fitting model 45, as described below.

GIBSS for SuSiE with unknown σ^2 A key feature of the (Gaussian SER) SuSiE model introduced by [1] is that, given estimates of $b_1, \ldots, b_{L-1}$, estimating b_L reduces to fitting a single- SER model with an offset:

$$o = \sum_{l=1}^{L-1} X b_l$$

As noted by Wang and colleagues [1], this observation suggests an iterative approach to fitting the SuSiE model by repeatedly fitting SER models—an algorithm they refer to as Iterative Bayesian

Stepwise Selection (IBSS). We build on this idea, we SuSiE with the Servin and Stephens prior using an iterative algorithm in which each iteration fits the SER with the Servin and Stephens prior. We refer to the resulting algorithm as Generalized IBSS (GIBSS), see Algorithm 1.

In short, GIBBS computes the approximate posterior distribution for each single effect b_l, which allows computing 1) the posterior mean of each estimated effect and their corresponding credible set (CS). The CSs are computing following [1]. As opposed to the original IBSS, it is not proven that GIBSS optimizes a proper variational approximation of the posterior distribution of b under the Servin and Stephens prior. Nonetheless, in our numerical experiment, GIBSS showed particularly good performances.

In summary, GIBSS computes an approximate posterior distribution for each single-effect coefficient b_l , enabling the derivation of both (1) the posterior mean of each estimated effect and (2) the corresponding credible sets (CSs), which are constructed following [1]. Unlike the original IBSS, GIBSS is not proven to optimize a formal variational approximation of the posterior distribution of b under the Servin and Stephens prior. Nevertheless, our numerical experiments demonstrate strong empirical performance of GIBSS.

GIBSS stopping criterion As described in Algorithm 1, GIBSS terminates when either the maximum number of iterations is reached (set to 100 by default) or a predefined stopping criterion is satisfied. In standard VA methods, it is common practice to monitor the evidence lower bound (ELBO) [18] and halt the algorithm when the increment in ELBO between consecutive iterations falls below a certain threshold (e.g., 10^{-6}).

Since GIBSS does not explicitly optimize a VA objective, we do not have direct access to an ELBO. Instead, we assess the variation in the parameters $(\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_L)$ across iterations and stop when the change between consecutive iterations is sufficiently small, i.e.,

$$\left\| (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_L)^t - (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_L)^{t+1} \right\|_2 < \epsilon, \tag{48}$$

where ϵ is a small threshold, set to 10^{-6} by default. While this stopping criterion is heuristic, we note that some fine-mapping methods that explicitly optimize a variational approximation employ a similar approach in practice (see the implementation of SuSiE-inf [6]).

Some notes on the t distribution

Let $x \sim t_\nu$ be random variable drawn from the t distribution with ν degrees of freedom. The probability density function is

$$t_\nu(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}. \tag{49}$$

Now let $y = \mu + \sigma x$. We denote its distribution as $y \sim t_\nu(\mu, \sigma^2)$, and its probability density function is

$$t_\nu(y; \mu, \sigma^2) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma^2}} \left(1 + \frac{1}{\nu}\left(\frac{y-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}. \tag{50}$$

The mean of y is μ (for $\nu > 1$) and its variance is $\frac{\nu}{\nu-2}\sigma^2$ (for $\nu > 2$).

A useful identity: if $\sigma^2 \sim \mathrm{IG}(\frac{\alpha}{2}, \frac{\beta}{2})$ and $x \mid \sigma^2 \sim N(\mu, s^2\sigma^2)$, then $x \sim t_\alpha(\mu, \frac{\beta}{\alpha}s^2)$.

Posterior distribution of σ^2, and b when σ^2 is unknown

Let's denote

$$\mu_j = \mu_{SS}(\boldsymbol{x}_j, \boldsymbol{y}, \tilde{\sigma}_0^2) \tag{51}$$

$$s_j^2 = \sigma_{SS}^2(\boldsymbol{x}_j, \boldsymbol{y}, \tilde{\sigma}_0^2) \tag{52}$$

Then the posterior first and second moment of b_j under the SER SS model 37 is given by

$$\mathbb{E}[b|\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2] = \pi_{1j}\mu_j \tag{53}$$
$$\mathbb{E}[b^2|\boldsymbol{x}, \boldsymbol{y}, \tilde{\sigma}_0^2] = \pi_{1j}(s_j^2 + \mu_j^2) \tag{54}$$

R script: Below the script to compute the posterior mean and variance for the SS model $\alpha = \beta = 0$

Sample code

Here the R code to compute the key quantities under the Servin and Stephens prior.

```
posterior_mean_SS<- function(x,y, s0_t=1){
  omega <- (( 1/s0_t^2)+crossprod(x))^-1
  b_bar<- omega%*%(crossprod(x,y))
  return( b_bar)
}

posterior_mean_SS_suff <- function(xtx,xty, s0_t=1){
  omega <- (( 1/s0_t^2)+xtx)^-1
  b_bar<- omega%*%(xty)
  return( b_bar)
}

posterior_var_SS <- function (x,y, s0_t=1){

  omega <- (( 1/s0_t^2)+crossprod(x))^-1
  b_bar<- omega%*%(crossprod(x,y))

  post_var_up <- 0.5*(crossprod(y)  -  b_bar *(omega ^(-1))*b_bar)
  post_var_down <- 0.5*(length(y)*(1/omega ))
  post_var <- (post_var_up/post_var_down)* length(y)/(length(y)-2)
  return(  post_var)
}


posterior_var_SS_suff<- function (xtx,xty, n,s0_t=1){
  omega <- (( 1/s0_t^2)+xtx)^-1
  b_bar<- omega%*%(xty)
  post_var_up <- 0.5*(yty  -  b_bar *(omega ^(-1))*b_bar)
  post_var_down <- 0.5*(n*(1/omega ))
  post_var <- (post_var_up/post_var_down)* n/(n-2)
  return(  post_var)
}



  compute_log_ssbf <- function (x, y, s0) {
    x   <- x - mean(x)
    y   <- y - mean(y)
    n   <- length(x)
    xx  <- sum(x*x)
    xy  <- sum(x*y)
    yy  <- sum(y*y)
    r0  <- s0/(s0 + 1/xx)
```

```
    sxy <- xy/sqrt(xx*yy)
    return((log(1 - r0) - n*log(1 - r0*sxy^2))/2)
  }

# assuming centered x and y
compute_log_ssbf_suff <- function ( xtx,  xty, yty , s0,n) {
    r0  <- s0/(s0 + 1/xtx)
    sxty <- xty/sqrt(xtx*yty)
    return((log(1 - r0) - n*log(1 - r0*sxty^2))/2)
}
```

References for Supplementary Notes

[19] George, E. I. and R. E. McCulloch (1997, April). Approaches for Bayesian Variable Selection. *Statistica Sinica 7*(2), 339–373. Publisher: Institute of Statistical Science, Academia Sinica.

[14] Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*(481), 410–423.

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

[22] Stephens, M. (2013, July). A Unified Framework for Association Analysis with Multiple Related Phenotypes. *PLOS ONE 8*(7), e65245. Publisher: Public Library of Science.

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

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

[6] Cui, R., R. A. Elzur, M. Kanai, J. C. Ulirsch, O. Weissbrod, M. J. Daly, B. M. Neale, Z. Fan, and H. K. Finucane (2024). Improving fine-mapping by modeling infinitesimal effects. *Nature Genetics 56*, 162–169.