nature genetics

Article

https://doi.org/10.1038/s41588-023-01597-3

Improving fine-mapping by modeling infinitesimal effects

Received: 24 October 2022

Accepted: 26 October 2023

Published online: 30 November 2023

Check for updates

Ran Cui[1,2,3,4] ✉, Roy A. Elzur[1,2,3], Masahiro Kanai[1,2,3,5,6,7], Jacob C. Ulirsch[1,2,3,8], Omer Weissbrod[9], Mark J. Daly[1,2,3,6], Benjamin M. Neale[1,2,3,4], Zhou Fan[10,11] ✉ & Hilary K. Finucane[1,2,3,11] ✉

Fine-mapping aims to identify causal genetic variants for phenotypes. Bayesian fine-mapping algorithms (for example, SuSiE, FINEMAP, ABF and COJO-ABF) are widely used, but assessing posterior probability calibration remains challenging in real data, where model misspecification probably exists, and true causal variants are unknown. We introduce replication failure rate (RFR), a metric to assess fine-mapping consistency by downsampling. SuSiE, FINEMAP and COJO-ABF show high RFR, indicating potential overconfidence in their output. Simulations reveal that nonsparse genetic architecture can lead to miscalibration, while imputation noise, nonuniform distribution of causal variants and quality control filters have minimal impact. Here we present SuSiE-inf and FINEMAP-inf, fine-mapping methods modeling infinitesimal effects alongside fewer larger causal effects. Our methods show improved calibration, RFR and functional enrichment, competitive recall and computational efficiency. Notably, using our methods' posterior effect sizes substantially increases polygenic risk score accuracy over SuSiE and FINEMAP. Our work improves causal variant identification for complex traits, a fundamental goal of human genetics.

Over the past two decades, genome-wide association studies (GWAS) have successfully identified thousands of loci that are associated with various diseases and traits[1]. However, refining these associations to identify causal variants remains challenging, due to extensive linkage disequilibrium (LD) among associated variants[2]. Many approaches can be taken to help nominate causal variants from associations, such as overlapping GWAS signals with coding or functional elements of the genome[3], with expression quantitative trait loci[4], and across populations having different ancestries and patterns of LD[5–7]. Complementary to and in conjunction with these approaches, Bayesian sparse regression and variable selection methods, which aim to identify causal variants and quantify their uncertainty based on a statistical model (for example, SuSiE[8], FINEMAP[9,10], ABF[11] and COJO[12]-ABF), are widely applied in practice[13–19].

The appeal of Bayesian approaches to fine-mapping is two-fold. First, these methods determine a posterior inclusion probability (PIP) for each variant, quantifying the probability that the variant is causal under the model, which can reflect uncertainty due to LD. For example, two variants in strong LD and harboring a strong association with the phenotype may each have PIP close to 50%, representing confidence

[1]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [6]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [7]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. [8]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA. [9]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [10]Department of Statistics and Data Science, Yale University, New Haven, CT, USA. [11]These authors jointly supervised this work: Zhou Fan, Hilary K. Finucane. ✉e-mail: rancui@broadinstitute.org; zhou.fan@yale.edu; finucane@broadinstitute.org

that there is a causal signal but uncertainty about which variant(s) is/are causal. Second, these methods incorporate assumptions about genetic architecture—namely the relative probabilities of different numbers of and configurations of causal single nucleotide polymorphisms (SNPs), as reflected by a Bayesian prior—to improve statistical power for identifying high-confidence variants.

Bayesian fine-mapping methods are correctly calibrated when the PIPs accurately reflect the true proportions of causal variants, for example, nine out of ten variants having PIP 90% are truly causal for the trait. Calibration (that is, whether or not the posterior probability of causality reflects the true proportion of causal variants) is ensured when the linear model for genetic effects and Bayesian prior for genetic architecture across loci are both correctly specified, and accurate calibration has also been demonstrated empirically in simulations to be robust under mild model misspecifications[20]. However, the actual calibration and false discovery rates of these methods in real data applications are not easily determined, as true causal variants and the sources of model misspecification may be unknown.

In this Article, we propose the replication failure rate (RFR) to assess the stability of fine-mapping methods by evaluating the consistency of PIPs in random subsamples of individuals from a larger well-powered cohort. We found the RFR to be higher than expected across traits for several Bayesian fine-mapping methods. Moreover, variants that failed to replicate at the higher sample size were less likely to be coding. Together these analyses suggest that SuSiE, FINEMAP and COJO-ABF may be miscalibrated in real data applications. In other words, they may return a disproportionately large number of false discoveries among high-PIP variants.

We performed large-scale simulations to assess the effects of several plausible sources of model misspecification on calibration. These simulations—which include, among other factors, varying levels of nonsparsity and stratification—suggest that a denser and more polygenic architecture of genetic effects may be a major contributor to PIP miscalibration. We thus propose incorporating a model of infinitesimal effects when performing Bayesian sparse fine-mapping, recasting the goal of fine-mapping as the identification of a sparse set of large-effect causal variants among many variants having smaller effects. We develop and implement fine-mapping tools SuSiE-inf and FINEMAP-inf that extend the computational ideas of SuSiE and FINEMAP to model additional infinitesimal genetic effects within each fine-mapped locus.

Applying SuSiE-inf and FINEMAP-inf to ten quantitative traits in the UK Biobank (UKBB) shows improved RFR. SuSiE-inf high-PIP variants are more functionally enriched than SuSiE high-PIP variants. Cross-ancestry phenotype prediction using SuSiE-inf/FINEMAP-inf shows substantial improvement over SuSiE/FINEMAP across seven traits and six diverse ancestries. These results suggest that explicit modeling of a polygenic genetic architecture, even within individual genome-wide significant loci, may substantially improve fine-mapping accuracy.

Results

Current methods are probably miscalibrated in real data

Real data benchmarking of fine-mapping methods is challenging due to the lack of ground truth. However, downsampling large cohorts allows assessment of the methods' stability. We chose ten well-powered quantitative phenotypes (Methods) in the UKBB and computed the RFR for SuSiE and FINEMAP as follows (for results related to ABF and COJO-ABF, see Supplementary Note 1). Our group previously performed fine-mapping[20] on a cohort of 366,194 unrelated "white British" ancestry individuals defined in the Neale Lab UKBB GWAS[21]. We downsampled this cohort to a random subsample of 100,000 and performed fine-mapping with the same pipeline (Methods). RFR is defined as the proportion of high-confidence (PIP >0.9) variants fine-mapped in the 100,000 subsample that failed to replicate (PIP <0.1) in the full 366,000 cohort. This RFR is an estimate of the conditional probability $\Pr(\mathrm{PIP}^{366,000} < 0.1 \mid \mathrm{PIP}^{100,000} > 0.9)$ for a randomly chosen variant. In a truly sparse causal model, assuming that the method is well powered at sample size $N = 366{,}000$ to detect true causal variants, which are identified with high confidence at 100,000, the RFR is an approximate lower bound for the false discovery rate $\Pr(\text{not causal} \mid \mathrm{PIP}^{100,000} > 0.9)$ (Supplementary Note 2).

Across all ten traits, we observed different levels of RFR for different phenotypes, and an aggregated RFR of 15% for SuSiE and 12% for FINEMAP (Fig. 1a,b; see Extended Data Fig. 1 for different PIP thresholds). These values far exceed the false discovery rate expected in a correctly specified sparse Bayesian model (SuSiE 1.8% and FINEMAP 2.0%), which we denote by expected proportion of noncausal variants (EPN) and estimate from the mean reported PIPs exceeding 0.9. In contrast, ideal simulations under correctly specified models show close agreement between RFR and EPN (Fig. 1a, Methods and Supplementary Note 3).

To gain insight into whether nonreplicating variants (PIP 0.9 at 100,000 and PIP <0.1 at 366,000) are causal, we examined the functional annotations, focusing on two distinct categories: coding and putative regulatory (Methods). We found a significant depletion of functionally important variants in the nonreplicating set compared to the replicated set ($P = 1.7 \times 10^{-7}$) (Fig. 1c and Methods). This suggests that many nonreplicating variants may be noncausal, and that SuSiE and FINEMAP may be miscalibrated when applied in real data (for our investigation into other potential causes for high RFR, see Supplementary Note 4). We found higher functional enrichment in the set of nonreplicating variants than the background, suggesting some PIPs at $N = 366{,}000$ may be too conservative. However, here we focus on investigating the more concerning under-conservative PIPs that can lead to elevated false discovery rate.

Unmodeled nonsparse effects can lead to miscalibration

Bayesian sparse variable selection approaches to fine-mapping, including SuSiE and FINEMAP, commonly rely on some of the following assumptions. (1) Within each genome-wide significant locus, one or a small number of variants have a true causal contribution to the phenotype. (2) All true causal variants within the locus are included in, or tagged by a sparse subset of, the analyzed genotypes. (3) The distribution of causal variant effect sizes is well approximated by a simple, oftentimes Gaussian, prior. (4) There is no uncorrected confounding, and the residual error is uncorrelated with the genotype. (5) There is no imputation noise or error in the genotypes. Violations of any of these assumptions can, in principle, cause miscalibration, although the severity of such miscalibration under the degrees of violation that are present in fine-mapping applications is unclear a priori.

We designed large-scale simulations to investigate how SuSiE and FINEMAP may be affected by these five sources of misspecification. Our simulations use UKBB genotypes ($N = 149{,}630$ individuals of white British ancestry) and BOLT linear mixed model (BOLT–LMM)[22] for GWAS, incorporating (1) varying amounts of unmodeled nonsparse causal effects (varying both the coverage of nonsparsity, that is, the proportion of variants with nonzero effects, and the amount of heritability the nonsparse component explains), (2) missing causal variants that are removed by quality control filtering before fine-mapping, (3) effect size distributions for the large and sparse causal variants that reflect estimates from fine-mapping of real traits, (4) varying amounts of uncorrected population stratification, and (5) imputation noise in the input genotypes (see Methods for detailed description of our simulations and other misspecifications we considered). In previous work[20] by our group, we found that quality control filters and imputation noise did not contribute to miscalibration in simulations; here we continued to include them while adding nonsparsity, effect size estimates from real data, and uncorrected population stratification as additional sources of miscalibration. Note that we simulated a single cohort, without the heterogeneity that often comes with meta-analysis,

https://doi.org/10.1038/s41588-023-01597-3

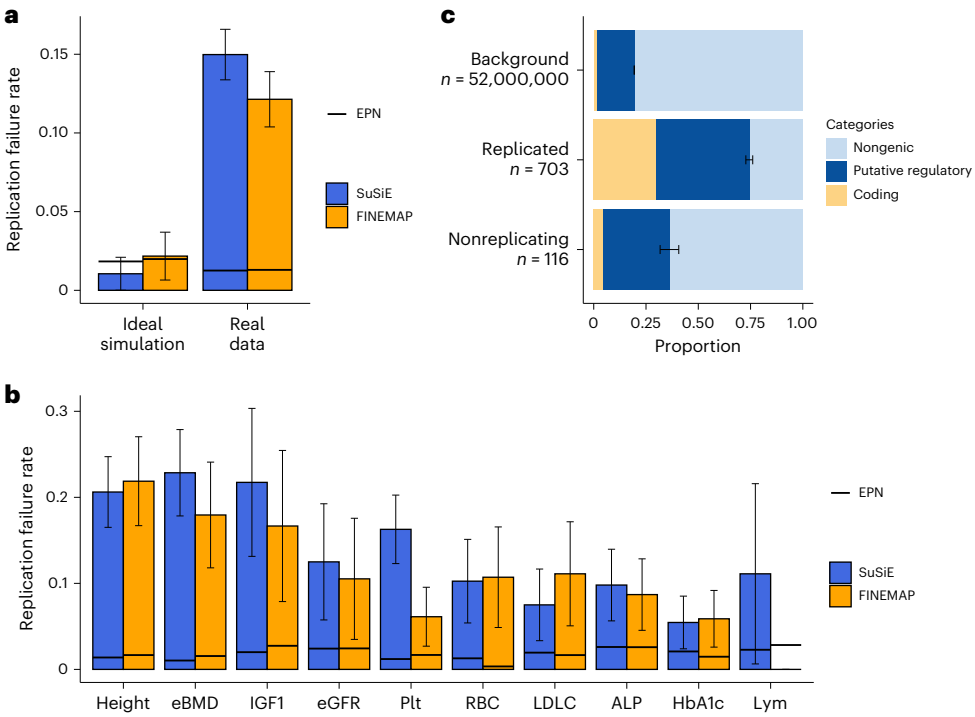


Fig. 1 | RFRs and functional enrichments. a, RFRs for SuSiE and FINEMAP aggregated across ten UKBB quantitative phenotypes, contrasted with RFRs in ideal simulations and with EPN. **b**, Trait-separated RFRs for SuSiE and FINEMAP. The ten UKBB traits are height, eBMD, Plt, HbA1c, RBC, ALP, IGF1, LDLC, Lym and eGFR. **c**, Functional annotations in three disjoint categories: coding, putative regulatory and nongenic (see Methods for detailed definitions). Variants are aggregated between SuSiE and FINEMAP with nonreplicating, the group of nonreplicating variants (PIP >0.9 at $N = 100{,}000$ and PIP <0.1 at $N = 366{,}000$); replicated, the group of replicated variants (PIP >0.9 at both $N = 100{,}000$ and $N = 366{,}000$); and background, the group of all variants included in the fine-mapping analysis, aggregated across ten traits. For method-separated plots and more sets of variants, see Supplementary Fig. 1. Error bars represent one standard deviation of the corresponding binomial distribution Binom(n,p), where n is the total number of variants (for **a** and **b**, n is the total number of high-PIP variants at sample size $N = 100{,}000$; for **c**, n is the total number of variants in each group) and p is the corresponding proportion (RFR in **a** and **b** and proportion of annotated variants in **c**). Bar plot data are presented as proportion ± standard deviation. Numerical results are available in Supplementary Tables 4 and 5.

where quality control (QC) and imputation are important contributors to miscalibration[23]. Moreover, we did not consider error in the probabilities outputted by standard imputation software or different types of genotyping error, which could contribute to miscalibration even in the absence of heterogeneity.

Our simulations show that missing causal variants due to QC, use of a realistic non-Gaussian effect size distribution estimated from real data, and imputation error did not induce miscalibration, consistent with and extending previous results[20].

SuSiE and FINEMAP were both miscalibrated in simulations with nonsparse effects. For example, when nonsparse causal effects explain 75% of the total SNP heritability, only about 80% of variants with PIP ≥0.9 are causal, far below the rate of approximately 97% that we would expect given the variants' mean PIP. Miscalibration increased and recall decreased as we increased the proportion of total SNP heritability (set at 0.5; for comparison, see Supplementary Table 1 for common SNP heritability in real traits) explained by nonsparse effects from 58% to 100% (Fig. 2a,b and Table 1) while fixing the coverage to 1%. This trend was consistently observed at different levels of coverage. See Methods and Extended Data Fig. 2 for results at 0.5% and 5% coverage. We emphasize that calibration was measured against the set of all causal variants, including the nonsparse causal effects.

To further support that unmodeled nonsparse causal effects—among all the misspecification we incorporated—formed the primary driver of the observed miscalibration, we decomposed the simulated genetic component $X\boldsymbol{\beta}$ of the phenotype into the sum of four subcomponents representing sparse causal effects, missing causal variants, uncorrected stratification and unmodeled nonsparse causal effects. Regressing each of these four subcomponents on the true and false positive variants (respectively defined as causal and noncausal variants with PIP ≥0.9), false positive variants were more correlated with the nonsparse causal effects than true positive variants (Fig. 2c and Methods).

Our simulated population stratification in the standard pipeline (Methods), where BOLT–LMM was used for association mapping, failed to induce miscalibration. Replacing BOLT–LMM with ordinary least squares (OLS) for association mapping allowed us to induce higher levels of uncorrected confounding (Supplementary Table 2) that did lead to miscalibration (Fig. 3), but are less true to the pipeline used in our real data applications. See Methods for interpretation and more discussion on these results.

In conclusion, the presence of nonsparse effects is a driver of miscalibration for SuSiE and FINEMAP. The stratification we simulated only induced miscalibration when using OLS for association mapping but not when using BOLT–LMM. None of the other sources of misspecification incorporated in our simulations caused miscalibration within our fine-mapping pipeline.

Modeling infinitesimal effects in addition to sparse effects

To address PIP miscalibration that may arise from nonsparse causal effects, we propose to explicitly incorporate a model of broad infinitesimal genetic effects when fine-mapping causal variants. Here, we describe two specific implementations of this idea that extend FINEMAP and SuSiE. We call the resulting methods FINEMAP-inf and SuSiE-inf.

FINEMAP-inf and SuSiE-inf are based on a random-effects linear model $y = X(\boldsymbol{\beta} + \boldsymbol{\alpha}) + \epsilon$ for observed phenotypes y across n samples, where X is an n by p genotype matrix for p variants, $\boldsymbol{\beta}$ is a vector of sparse genetic effects of interest, $\boldsymbol{\alpha}$ is an additional vector of dense

https://doi.org/10.1038/s41588-023-01597-3

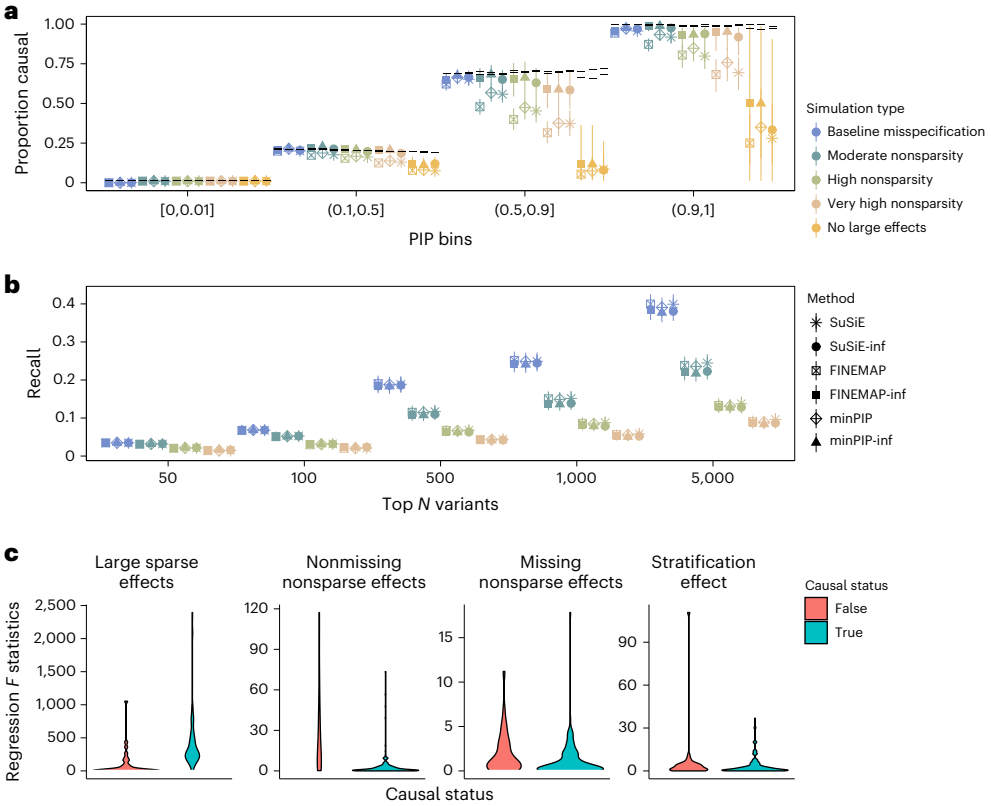


Fig. 2 | Simulations with nonsparse effects. a, Calibration for SuSiE, FINEMAP, minPIP and corresponding 'inf' methods under nonsparsity simulation settings detailed in Table 1 and Methods. minPIP and minPIP-inf are aggregating methods: minPIP-inf is equal to min(PIP) between SuSiE-inf and FINEMAP-inf, and minPIP is equal to min(PIP) between SuSiE and FINEMAP. The colored markers show true proportion of causal variants, and the short black lines show the expected proportion of causal variants in each PIP bin for each method. **b**, Recall for the same methods, defined as the percentage of simulated large effects among the top N variants when ranked by PIP. Error bars on calibration and recall plots correspond to 95% Wilson confidence interval. Note that 'no large effects' simulations are not shown on the recall plot because there are zero simulated large effects. **c**, Regressing subcomponents of 'high nonsparsity' phenotype on true versus false positives (variants with PIP >0.9 that are either causal or noncausal). Numerical results are available in Supplementary Tables 6–8.

infinitesimal effects and ϵ is the residual error. In the context of such a model, we define the primary goal of fine-mapping as inferring the nonzero coordinates of the sparse component $\boldsymbol{\beta}$. We will refer to these coordinates as the 'causal model' and the 'causal variants, although in this model, every variant may have an additional small causal effect on y through the infinitesimal component α.

We model coordinates of $\boldsymbol{\alpha}$ and of the residual error ϵ as independent and identically distributed with normal distributions $\mathcal{N}(0, \tau^2)$ and $\mathcal{N}(0, \sigma^2)$, respectively, where τ^2 is the effect size variance for the infinitesimal effect. For FINEMAP-inf, coordinates of the sparse effects $\boldsymbol{\beta}$ are also modeled as independent and identically distributed, with point-normal distribution $\pi_0 \mathcal{N}(0, s^2) + (1 - \pi_0)\,\delta_0$. We use a shotgun stochastic search procedure as in FINEMAP for performing approximate posterior inference of the sparse component $\boldsymbol{\beta}$, marginalizing its posterior distribution over both the infinitesimal effects $\boldsymbol{\alpha}$ and the residual errors ϵ. The shotgun stochastic search is divided into several epochs, and we propose a method-of-moments approach to update estimates of the variance components (σ^2, τ^2) between epochs.

For SuSiE-inf, we follow the approach of SuSiE and instead parametrize the sparse causal effects as a sum of single effects $\boldsymbol{\beta} = \sum_{l=1}^{L} \boldsymbol{\beta}^{(l)}$ for a pre-specified number of causal variants L. As in SuSiE, we perform posterior inference for $\boldsymbol{\beta}$ using a variational approximation for the joint posterior distribution of $\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(L)}$, again marginalizing over both $\boldsymbol{\alpha}$ and ϵ. The approximation is computed by iterative stepwise optimization of an evidence lower bound, where updated estimates of the variance components (σ^2, τ^2) are computed within each iteration using a method-of-moments approach.

The resulting models are similar to LMMs commonly used in contexts of association testing and phenotype prediction[22,24–26]. See Discussion for an explanation of why we do not apply existing methods for fitting LMMs.

Both methods take as input either the GWAS data (y, X) or sufficient summary statistics given by the un-standardized per-SNP z scores, $z = (1/\sqrt{n})X^{\mathrm{T}}y$, the in-sample LD correlation matrix $\mathrm{LD} = (1/n)X^{\mathrm{T}}X$, and the mean-squared phenotype $\langle y^2 \rangle = (1/n)\,y^{\mathrm{T}}y$. Both methods output estimates of (σ^2, τ^2) for each locus fine-mapped, together with a PIP and posterior mean effect size estimate for each SNP. Computational cost is reduced by expressing all operations in terms of the eigenvalues and eigenvectors of LD, which may be pre-computed separately for each fine-mapped locus (Fig. 4). Details of the methods and computations are provided in Supplementary Note 5. We have released open-source software implementing these methods ('Code availability').

SuSiE-inf and FINEMAP-inf show improved performance

In our simulations, we find that SuSiE-Inf and FINEMAP-Inf have improved calibration over SuSiE and FINEMAP, respectively, except for simulations using OLS that introduced uncorrected population stratification, which are less relevant to our findings in real data using BOLT–LMM (Figs. 2a and 3a and Methods). Recall of SuSiE-Inf and FINEMAP-Inf was very similar to, but slightly lower than, that of SuSiE and FINEMAP, respectively (Figs. 2b and 3b). With improved performance in simulations having nonsparse genetic effects, and similar performance in simulations with stratification using BOLT–LMM (Fig. 3a), we turned to real data benchmarking to assess whether SuSiE-inf and FINEMAP-inf improve performance in practice.

https://doi.org/10.1038/s41588-023-01597-3

Table 1 | Parameter settings for large-scale simulations

	Imputation noise	Sparse causal prior	20 PC effects multiplier	PCs corrected in GWAS	Nonsparse causal effects	Missing causal effects
Ideal	No	Uniform	0	0	None	None
Baseline misspecification	Yes	SuSiE Height posterior	1	19 out of 20	None	None
Moderate stratification with BOLT	Yes	SuSiE Height posterior	5	0 out of 20	None	None
Severe stratification with BOLT	Yes	SuSiE Height posterior	8	0 out of 20	None	None
Moderate stratification with OLS	Yes	SuSiE Height posterior	1	0 out of 20	None	None
Severe stratification with OLS	Yes	SuSiE Height posterior	2	0 out of 20	None	None
Moderate nonsparsity	Yes	SuSiE Height posterior	1	19 out of 20	58% of h^2, 1% coverage	Yes
High nonsparsity	Yes	SuSiE Height posterior	1	19 out of 20	75% of h^2, 1% coverage	Yes
Very high nonsparsity	Yes	SuSiE Height posterior	1	19 out of 20	83% of h^2, 1% coverage	Yes
No large effects	Yes	SuSiE Height posterior	1	19 out of 20	100% of h^2, 1% coverage	Yes
0.5% coverage, ratio 3:1	Yes	SuSiE Height posterior	1	19 out of 20	75% of h^2, 0.5% coverage	Yes
5% coverage, ratio 3:1	Yes	SuSiE Height posterior	1	19 out of 20	75% of h^2, 5% coverage	Yes
5% coverage, ratio 15:1	Yes	SuSiE Height posterior	1	19 out of 20	94% of h^2, 5% coverage	Yes

Different parameter settings for ten sets of simulations are mentioned in the main text. Note that PCs corrected in GWAS used in-sample (N=150,000) PCs as covariates for phenotypes generated with full sample (N=366,000) PCs. See Methods for details on how each misspecification is incorporated. h^2, heritability.

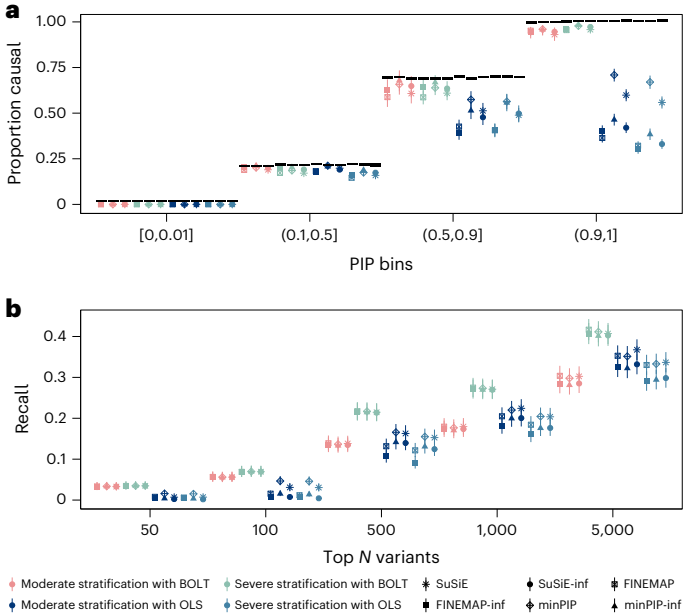


Fig. 3 | Simulations with uncorrected stratification. a, Calibration plot for SuSiE, FINEMAP, minPIP and corresponding "inf" methods in four stratification simulation settings (Table 1). The colored markers show true proportion of causal variants, and the short black lines show the expected proportion of causal variants in each PIP bin for each method. **b**, Recall for the same methods and simulations, defined as the percentage of simulated large effects among the top N (N = 50, 100, 500, 1,000 and 5,000) variants when ranked by PIP. Error bars correspond to 95% Wilson confidence interval. Numerical results are available in Supplementary Tables 6 and 7.

Real data benchmarking shows improvements by several metrics. RFR was substantially decreased for SuSiE-inf (Fig. 5a). SuSiE-inf-specific high-PIP variants (variants that are assigned a high PIP by SuSiE-inf but not by SuSiE) are 58% more enriched in functionally important categories than SuSiE-specific high-PIP variants ($P = 6 \times 10^{-4}$); the analogous difference in functional enrichment for FINEMAP versus FINEMAP-inf was nonsignificant (38% more for FINEMAP-inf specific variants, $P = 0.07$; Fig. 5b,c). In addition to high-PIP variants identified with PIP >0.9, we also observed better functional enrichment for top N (N = 500, 1,000, 1,500 and 3,000) variants (Extended Data Fig. 3a), demonstrating better prioritization of variants by our methods. Similar improvements were observed when using OLS for GWAS instead of BOLT–LMM (Extended Data Fig. 3b,c), upon correcting for stratification using principal component (PC) analysis. Compared to SuSiE and FINEMAP, we obtained fewer high-PIP variants (16% reduction aggregated between SuSiE and FINEMAP); however, the reduction is smaller for high-confidence variants, characterized either by replicated variants (11% reduction) or variants achieving PIP >0.9 for both SuSiE-Inf and FINEMAP-Inf/both SuSiE and FINEMAP (11% reduction) (Extended Data Fig. 3d). We observed a more substantial reduction of 42% in the number of credible sets when using SuSiE-inf; however, the reduction for smaller credible sets (number of variants <10) was somewhat smaller (36% reduction). Credible sets generated by SuSiE-inf are smaller on average than those generated by SuSiE (Extended Data Fig. 3e). Together, these results demonstrate both that SuSiE-inf and FINEMAP-inf allow for more confident identification of likely causal variants than the current state of the art, and that there is room for further methodological improvement.

In simulation, estimates of the infinitesimal variance τ^2 were higher on average for simulation settings with higher true infinitesimal

https://doi.org/10.1038/s41588-023-01597-3

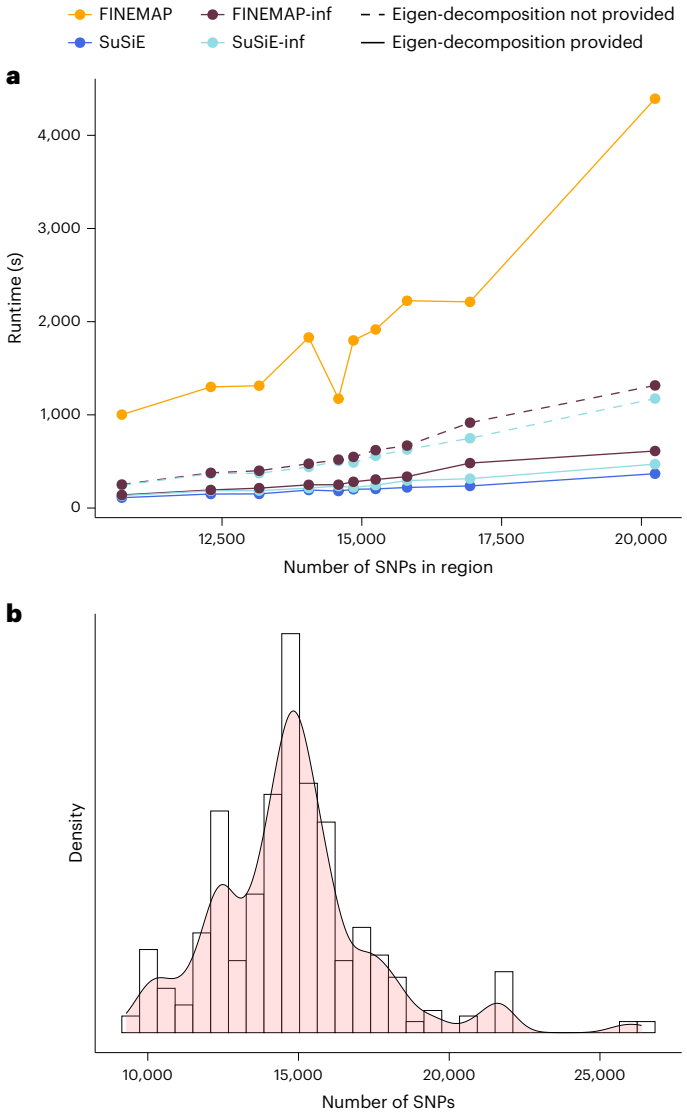


Fig. 4 | Runtime comparison. a, Runtime for SuSiE, FINEMAP, as well as SuSiE-inf and FINEMAP-inf, with/without provided eigen-decomposition of the LD matrix, is shown. We took all fine-mapped loci across the ten phenotypes we examined in UKBB and grouped them into ten quantiles based on the number of SNPs in the locus. Each point in the plot represents the average runtime for a method in a quantile. **b**, Distribution of fine-mapped locus sizes in terms of number of SNPs in the locus, aggregated across ten UKBB phenotypes and across two sample sizes: $N = 100{,}000$ and $N = 366{,}000$. Numerical results available in Supplementary Table 9.

variance (Extended Data Fig. 4a,b). Estimates of τ^2 were also higher in the presence of more residual stratification in the simulations, fixing all other simulation parameters (Extended Data Fig. 4c). In UKBB data, estimates of τ^2 varied across traits, with height showing the highest estimates and LDL showing the lowest estimates (Extended Data Fig. 5a,b). We also found that estimates of τ^2 increased, on average, as the number of credible sets in a locus increased (Extended Data Fig. 5c). Estimates of τ^2 varied across loci for a given trait, due either to differences in genetic architecture, residual stratification or estimation noise. We caution against the interpretation of τ^2 as a direct reflection of trait heritability or genetic architecture, without further investigations into these factors that may contribute to the τ^2 estimates.

To further validate our methods in real data, we performed cross-ancestry polygenic risk score (PRS) prediction[27,28], using posterior effect sizes estimated on 366,000 samples from the 'white British' cohort in UKBB to predict phenotypes in six held-out cohorts of different ancestries[29]: African ancestry ($N = 6{,}637$), admixed American ancestry ($N = 982$), Central/South Asian ancestry ($N = 8{,}876$), East Asian ancestry ($N = 2{,}709$), European ancestry test ($N = 54{,}337$) and Middle Eastern ancestry ($N = 1{,}599$). Prediction accuracy is measured by 'ΔR^2', which is the difference in R^2 from a model that includes both the covariates and genotype effects relative to a model that includes the covariates alone. Using posterior mean effect size estimates for the sparse component β in SuSiE-inf/FINEMAP-inf yields, on average, a near ten-fold increase in ΔR^2 over SuSiE/FINEMAP across all held-out cohorts and traits (Methods, Fig. 5c,d and Supplementary Table 3). Here we compute PRS using only the sparse component, to provide a validation metric for the fine-mapped SNPs. We leave an exploration of improving PRS accuracy by integrating estimated infinitesimal effect sizes to future work.

Our group has shown previously that combining SuSiE and FINEMAP can yield more reliable PIPs[30]. Here we recommend using the minimum PIP between SuSiE-inf and FINEMAP-inf (minPIP-inf) for each fine-mapped variant. Compared to minPIP (minimum PIP between SuSiE and FINEMAP), minPIP-inf retains more high confidence variants, showing better agreement between SuSiE-inf and FINEMAP-inf (Extended Data Fig. 6). We observed substantially improved RFR for minPIP-inf over minPIP (Extended Data Fig. 7a). Functional enrichment for the top N variants, simulation and PRS performance for minPIP-inf is comparable to either SuSiE-inf or FINEMAP-inf individually (Fig. 2a,b and Extended Data Fig. 7b–d). As examples of the improved effectiveness of the minPIP-inf method over minPIP, we examined two loci. At the *PCSK9* locus for low-density lipoprotein cholesterol (LDLC), in addition to the well-known causal variant rs11591147, SuSiE-inf and FINEMAP-inf consistently identified two intronic variants: rs499883 and rs7552841 with high confidence, replicating a previous finding using functionally informed priors[31], whereas SuSiE did not identify variant rs499883. At the *AK3* locus for platelet count (Plt), a known causal missense variant, rs7412, is in high LD with variant rs1065853. Only FINEMAP-inf captured rs7412, while SuSiE, SuSiE-inf and FINEMAP-inf captured another known causal variant, rs429358. MinPIP between SuSiE and FINEMAP missed both, whereas minPIP-inf captured one (Extended Data Figs. 8 and 9).

Discussion

We propose fine-mapping methods that control for infinitesimal causal effects while fine-mapping sparse causal effects. Using our methods, we observed substantial improvements in simulations with nonsparse genetic architecture. Our results when simulating uncorrected stratification were ambiguous: when using BOLT–LMM, stratification did not lead to miscalibration and our methods performed similarly to the previous methods; however, when using OLS, stratification led to substantial miscalibration that was similar between FINEMAP and FINEMAP-inf and worse for SuSiE-inf than SuSiE. In contrast, real data benchmarking demonstrated an unambiguous improvement in performance, for example, decreased RFR, improved functional enrichment of top variants, and large gains in polygenic risk prediction. Put together, the accuracy of identifying sparse causal variants is greatly improved when incorporating the infinitesimal model, although our results show that there is also room for further methodological improvement.

The models we propose here are similar to models that have been used previously to model genome-wide genetic architecture for risk prediction, heritability estimation and association mapping[22,24,26,32]. Fine-mapping differs from these other applications in that (1) fine-mapping requires inclusion of a denser set of variants with higher LD in each locus, so that the causal variants are likely to be included; (2) fine-mapping requires accurate inference of posterior inclusion probabilities; (3) fine-mapping is often performed at very large sample sizes; and (4) fine-mapping does not require joint modeling of genome-wide data, which would be computationally challenging given the density of variants and typical sample sizes.

https://doi.org/10.1038/s41588-023-01597-3

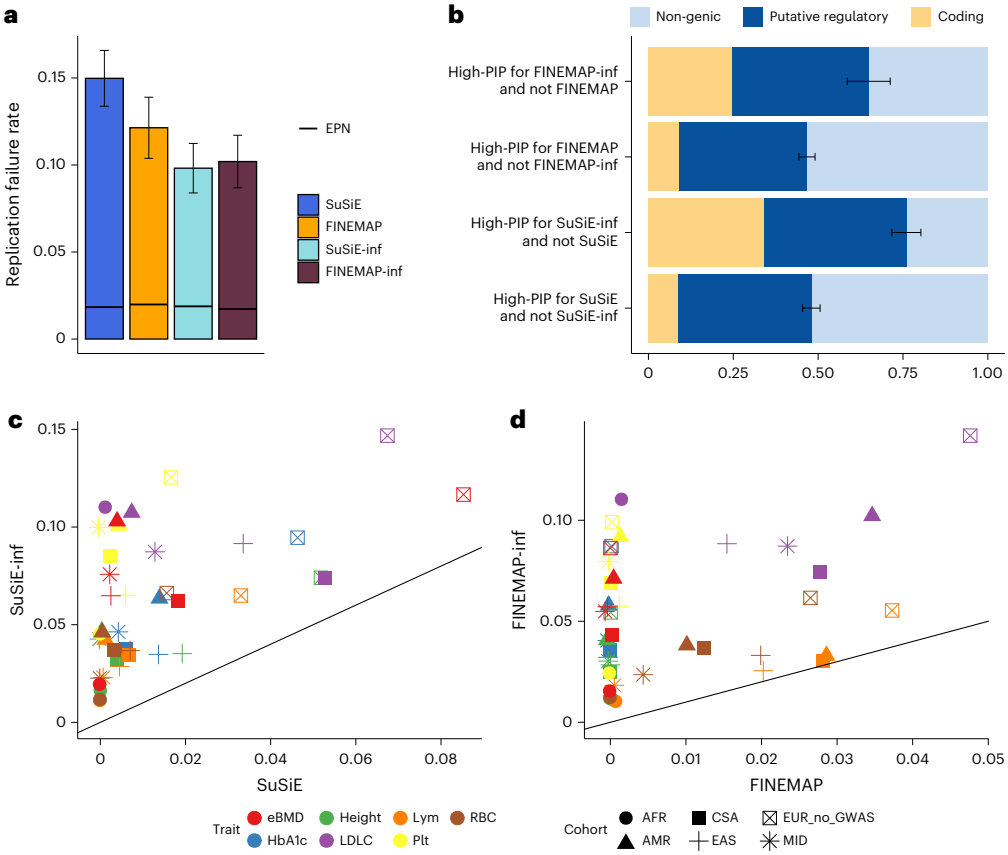


Fig. 5 | Real data performance improvements. a, RFRs are shown for SuSiE, FINEMAP, SuSiE-inf and FINEMAP-inf, aggregated across ten UKBB traits (Supplementary Table 4). **b**, Functional enrichment of the set differences between SuSiE and SuSiE-inf high-PIP variants (PIP >0.9), and FINEMAP and FINEMAP-inf high-PIP variants, is shown. Error bars represent one standard deviation of the corresponding binomial distribution Binom(n,p), where n is the total number of variants (for **a**, n is the total number of high-PIP variants at sample size $N = 100{,}000$; for **b**, n is the total number of variants in each set) and p is the corresponding proportion (RFR in **a** and proportion of annotated variants in **b**). Bar plot data is presented as proportion ± standard deviation. Numerical results are available in Supplementary Table 10. **c**,**d**, PRS accuracy, in terms of ΔR^2, when applying SuSiE posterior effect sizes versus SuSiE-inf sparse component of the posterior effect sizes as weights (**c**), as well as PRS accuracy when applying FINEMAP posterior effect sizes versus FINEMAP-inf sparse component of the posterior effect sizes as weights (**d**). PRSs were computed for six out-of-sample cohorts and seven traits (see Methods for cohort and trait definition). For PRS accuracy plots with standard errors, see Extended Data Fig. 10. Numerical results are available in Supplementary Tables 11 and 12. AFR, African ancestry; CSA, Central/South Asian ancestry; EUR, European ancestry; AMR, admixed American ancestry; EAS, East Asian ancestry; MID, Middle Eastern ancestry.

To emphasize the distinction between fine-mapping and risk prediction, if two variants are in perfect LD with large marginal effect sizes, a risk prediction method would perform equally well upon attributing this effect to either variant, whereas the desired outcome in fine-mapping is a more precise quantification of uncertainty for which variant(s) harbors the true effect. Because of these factors, we do not apply existing methods for fitting LMMs in other contexts. We instead extend algorithmic ideas in the fine-mapping literature to better estimate and quantify uncertainty for a sparse genetic component in the presence of strong LD, while estimating an infinitesimal variance component separately for each genome-wide significant locus. Our model incorporates infinitesimal effects for variants in LD with those of the sparse component, which we believe is important for obtaining improved calibration and fine-mapping accuracy. With careful translation, we anticipate that methodological innovations in risk prediction may continue to lead to advances in fine-mapping and vice versa.

We view our methods as complementary to a body of recent statistical developments that seek to accurately quantify and control false discoveries under weaker modeling assumptions, using constructions of knock-off variables and related conditional re-randomization ideas[33,34]. Such methods have been applied to GWAS and genetic fine-mapping applications[35–37]. Our perspective differs in the following ways: we choose not to assume a sparse causal model or test null hypotheses of exact conditional independence and instead aim to accurately identify large effects that drive observed GWAS associations in a model where every variant may be causal. To yield adequate statistical power for detecting causal variants at fine-mapping resolutions, we rely on a strong assumption about genetic architecture, as reflected by a Bayesian prior probability for each candidate model, rather than testing a null hypothesis for each variant that allows for an arbitrary genetic architecture excluding that variant. Thus, our methods remain largely dependent on relatively strong assumptions of the underlying genetic architecture, and we view the potential integration of these ideas into a more model-robust framework as an important direction for future research.

While our work improves fine-mapping accuracy, further advances are needed. First, exploring the effects of stratification and different association mapping methods on fine-mapping should be a priority. Second, our methods improve on RFR over previous methods, but RFR is still elevated compared to ideal simulations, suggesting room for improved modeling. In addition to better modeling, independent replication in another biobank[30] and incorporation of functional evidence such as annotations and expression quantitative trait loci[20] can help boost accuracy of discovery. Further methodological advancements, which may come from more flexible models of genetic architecture or further study of uncorrected confounding, may also contribute to further improvements in cross-population polygenic risk prediction.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-023-01597-3.

References

1. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
2. Shendure, J., Findlay, G. M. & Snyder, M. W. Genomic medicine—progress, pitfalls, and promise. *Cell* **177**, 45–57 (2019).
3. Hukku, A. et al. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am. J. Hum. Genet.* **108**, 25–35 (2021).
4. Hormozdiari, F. et al. Colocalization of GWAS and eQTL signals detects target genes. *Am. J. Hum. Genet.* **99**, 1245–1260 (2016).
5. LaPierre, N. et al. Identifying causal variants by fine mapping across multiple studies. *PLoS Genet.* **17**, e1009733 (2021).
6. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
7. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
8. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. Stat. Methodol.* **82**, 1273–1300 (2020).
9. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
10. Benner, C., Havulinna, A. S., Salomaa, V., Ripatti, S. & Pirinen, M. Refining fine-mapping: effect sizes and regional heritability. Preprint at *bioRxiv* https://doi.org/10.1101/318618 (2018).
11. Wakefield, J. Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
12. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). S1-3.
13. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
14. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
15. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
16. Ulirsch, J. C. et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat. Genet.* **51**, 683–693 (2019).
17. Westra, H.-J. et al. Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. *Nat. Genet.* **50**, 1366–1374 (2018).
18. Zhang, Z. et al. Genetic analyses support the contribution of mRNA N^6-methyladenosine (m^6A) modification to human disease heritability. *Nat. Genet.* **52**, 939–949 (2020).
19. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
20. Ulirsch, Jacob C. Identification and interpretation of causal genetic variants underlying human phenotypes (accession no. 2022.29209644). Doctoral dissertation, Harvard University (2022).
21. Howrigan, D. P. et al. Nealelab/UK_Biobank_GWAS: v2. *Zenodo* https://doi.org/10.5281/zenodo.8011558 (2023).
22. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
23. Kanai, M. et al. Meta-analysis fine-mapping is often miscalibrated at single-variant resolution. *Cell Genomics* **2**, 100210 (2022).
24. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
25. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
26. Erbe, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129 (2012).
27. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
28. Márquez-Luna, C. & Loh, P.-R. South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
29. Pan UK Biobank. *Broad Institute* https://pan.ukbb.broadinstitute.org/ (2023).
30. Kanai, M. et al. Insights from complex trait fine-mapping across diverse populations. Preprint at *medRxiv* https://doi.org/10.1101/2021.09.03.21262975 (2021).
31. Weissbrod, O. et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).
32. Lloyd-Jones, L. R. et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).
33. Bates, S., Candès, E., Janson, L. & Wang, W. Metropolized knockoff sampling. *J. Am. Stat. Assoc.* **116**, 1413–1427 (2021).
34. Candès, E., Fan, Y., Janson, L. & Lv, J. Panning for gold: 'model-x' knockoffs for high dimensional controlled variable selection. *J. R. Stat. Soc. Ser. Stat. Methodol.* **80**, 551–577 (2018).
35. Sesia, M., Katsevich, E., Bates, S., Candès, E. & Sabatti, C. Multi-resolution localization of causal variants across the genome. *Nat. Commun.* **11**, 1093 (2020).
36. Sesia, M., Bates, S., Candès, E., Marchini, J. & Sabatti, C. False discovery rate control in genome-wide association studies with population structure. *Proc. Natl Acad. Sci. USA* **118**, e2105841118 (2021).
37. He, Z. et al. GhostKnockoff inference empowers identification of putative causal variants in genome-wide association studies. *Nat. Commun.* **13**, 7209 (2022).

Methods

Our research complies with all relevant ethical regulations. Our data source is the UKBB. UKBB has approval from the North West Multicentre Research Ethics Committee as a Research Tissue Bank approval. This approval means that researchers do not require separate ethical clearance and can operate under the Research Tissue Bank approval. Data for this work was obtained under application number 31063. Additional ethical approval was not required.

Statistics and reproducibility

The UKBB contains 500,000 participants with various ancestries. Our research analyzed the ancestry with the largest sample size: 366,194 unrelated 'white British' individuals. Quality control of this cohort was previously done by Neale Lab GWAS[21]. The individuals of British ancestry were determined by the PC analysis-based sample selection criteria[38,39], and were further filtered to self-reported 'white British', 'Irish' or 'white'. Our downsampling analysis consists of randomly selected 100,000 individuals from the 366,194 individuals. High RFRs across multiple quantitative traits are reproducible with other subsets of randomly selected 100,000 individuals from this cohort.

Selection of UKBB phenotypes and downsampling analysis

To select the ten phenotypes for which to perform downsampling analyses, we used results from ref. 30 and computed the combined number of high-PIP (PIP >0.9) variants fine-mapped at $N = 366{,}000$ samples using both SuSiE and FINEMAP. From the top 15 phenotypes (out of 94), with the highest number of high-PIP variants (Supplementary Table 13), we selected: height, estimated heel bone mineral density (eBMD), Plt, hemoglobin A1c (HbA1c), red blood cell count (RBC), alkaline phosphatase (ALP), insulin-like growth factor 1 (IGF1), LDLC, lymphocyte count (Lym) and estimated glomerular filtration rate based on serum creatinine (eGFR) to perform downsampling analyses.

We downsampled from $N = 366{,}000$ to a random subset of $N = 100{,}000$ twice (to increase the number of discoveries and therefore statistical power for RFR analyses) and performed GWAS and fine-mapping on both sets of the $N = 100{,}000$ individuals using the same pipeline used at $N = 366{,}000$ (for pipeline description, see below). The sample size $N = 100{,}000$ was chosen to resemble the UKBB interim release dataset of total $N = 150{,}000$ with $N = 107{,}000$ white British individuals.

Fine-mapping pipeline

GWAS and fine-mapping in this paper were performed following the pipeline described in our group's previous work[30]. Briefly, GWAS summary statistics were computed using BOLT–LMM (v2.3.2) with covariates including sex, age, age^2, age and sex interaction term, age^2 and sex interaction term, and the top 20 genotype PCs. Fine-mapping regions were defined using a 3 Mb window around each lead GWAS variant, with merging of overlapping regions. Fine-mapping was performed with in-sample LD computed using LDstore v2.0 (ref. 40).

Excessively large regions (consequence of merging) that could not be fine-mapped due to computational limitations were tiled with overlapping 3 Mb loci, with 1 Mb spacing between the start points of consecutive loci. For these tiled regions, we computed a PIP for each SNP based on the 3 Mb locus whose center was closest to the SNP. This tiling approach was previously described and applied[31].

Although BOLT–LMM is the GWAS method of choice in our group's previous work[30], we also used OLS regression for some of our simulations and real data applications.

We then performed fine-mapping using the following tools, namely multiple causal variant methods SuSiE[8] v894ba2f and FINEMAP[9,10] v1.3.1, single causal variant method ABF[11], and conditional association (COJO[12] v1.93.0beta) plus ABF fine-mapping method, which we denote COJO-ABF. Fine-mapping pipeline scripts are available in ref. 41.

Ideal simulations

To establish reference RFR and calibration for all tested methods, we performed ideal simulations without model misspecification using UKBB genotypes. For simulating RFR, we performed two sets of simulations each at sample size $N = 366{,}000$ and subsample size $N = 100{,}000$. We used UKBB imputed dosages as true genotypes, and only selected 'white British' individuals defined previously in the Neale lab GWAS[21]. We drew 1,000 causal variants per simulation uniformly randomly from a total of 6.6 million common (minor allele frequency (MAF) ≥1%) imputed variants genome wide. We standardized genotypes to mean 0 and variance 1 and drew per-standardized genotype causal effect sizes from the same normal distribution $N(0, 0.5/1{,}000)$ for all selected causal variants. We then added errors randomly drawn from a normal distribution $N(0, 0.5)$ to simulate phenotypes. For comparison of calibration with our simulations under model misspecifications, three additional sets of ideal simulations at a matching sample size $N = 150{,}000$ were performed. Phenotypes were generated similarly, with 700 uniformly sampled true causal variants having effect sizes drawn from $N(0, 0.5/700)$.

Functional enrichment

We analyzed functional annotations to gain insights into the potential causal status of nonreplicating variants (defined in the main text and in the next paragraph). We define three main disjoint functional categories: coding, putative regulatory and nongenic. These categories are derived from the seven main functional categories defined in ref. 30. The 'coding' category is the union of predicted loss of function and missense categories; the 'putative regulatory' category is the union of synonymous, 5′ untranslated region, 3′ untranslated region, promoter and *cis*-regulatory element categories. We compare the proportion of nongenic variants in the following groups of variants:

1. Nonreplicating, the group of variants with PIP ≥0.9 at $N = 100{,}000$ and PIP ≥0.1 at $N = 366{,}000$.
2. Replicated, the group of variants with PIP ≥0.9 at $N = 100{,}000$ and PIP ≥0.9 at $N = 366{,}000$.
3. Matched on PIP at 100,000, a group of replicated variants chosen to match the nonreplicating variants on PIP at $N = 100{,}000$. For each nonreplicating variant with PIP <1, we find a replicated variant whose PIP is the closest as its match, and the matched variant is removed for future matches. If the nonreplicating variant has PIP of 1, we match a random (if there are multiple) replicated variant with the PIP of 1. If there are more nonreplicating variants with a PIP of 1 than there are replicated variants with PIP of 1, we do not remove the matched replicated variant from future matches, resulting in repeated matches.
4. Matched on PIP at 366,000, a group of low-PIP variants (PIP ≤0.1 at $N = 366{,}000$) chosen to match the nonreplicating variants on PIP at $N = 366{,}000$. Matching is performed the same way as described above, except that there are no repeated matches.
5. Background, defined as the union of all variants included in fine-mapping from all ten phenotypes.

P values are reported when assessing the significance of the difference between proportions of nongenic variants in different groups of variants. Fisher's exact test was performed using the R (version 4.2.1) function fisher.test, and one-sided *P* values were reported from the output of this function.

Large-scale simulations with misspecification

We selected 149,630 UKBB individuals from a set of 366,194 unrelated 'white British' individuals defined previously in the Neale Lab GWAS[21] for our large-scale simulations. We performed simulations under models that are misspecified in the following ways: (1) genotype imputation noise, (2) nonuniform probabilities for the identities of causal variants, (3) nonsparsity of true causal effects, (4) uncorrected population

stratification, and (5) missing causal variants. We performed nine sets of simulations. All simulations included the same amount of (1) imputation noise, (2) nonuniform prior causal probabilities, and (5) missing causal variants. The first simulation, 'baseline misspecification' in Table 1, included a small amount of (4) uncorrected stratification. Another four simulations varied, in addition, (3) the level of nonsparsity of causal effects. Finally, four additional simulations varied in (4) the amount of simulated stratification and the methods for correcting this stratification (Population stratification below).

Simulated genotypes

To simulate genotypes for 149,630 individuals, we randomly drew true genotypes for all autosomes based on the imputed genotype probabilities in the 'bgen' files provided by UKBB. Briefly, probabilistic true genotypes (pGTs) for a given variant , denoted pGT_i, were computed via $\mathrm{pGT}_i = \lceil u_i - \mathrm{GP}(X_i = 0) \rceil + \lceil u_i - \mathrm{GP}(X_i = 0) - \mathrm{GP}(X_i = 1) \rceil$, where $\mathrm{GP}(X_i = k), k \in \{0,1,2\}$ represents the genotype probability of having k copies of alternative alleles and $u_i \sim \mathrm{Uniform}(0,1)$ represents a uniform random variable. Phenotypes were generated using the pGTs. In downstream GWAS and fine-mapping, we use imputed genotype dosages provided by UKBB, thus simulating imputation noise. We only included variants with minor allele count >10, INFO score >0.2 and Hardy–Weinberg equilibrium P value $>1 \times 10^{-10}$ in our simulations.

Causal variants simulation

To incorporate a more realistic nonuniform distribution over causal variants, we simulated sparse causal effects from the SuSiE posterior distribution for UKBB height, as computed in the larger 366,000 sample[30]. Specifically, in each locus, for each credible set $\mathrm{CS}_i, i \in I$, where I indexes all credible sets outputted by fine-mapping height with SuSiE, we chose a causal variant according to normalized posterior inclusion probabilities within the corresponding SuSiE single effect (denoted α_{ik} for $k \in \mathrm{CS}_i$). We then drew the chosen variant's raw effect size (to be scaled later) from a normal distribution with mean and standard deviation given by the SuSiE posterior mean and standard deviation conditional on inclusion in the model. In total, 1,434 sparse causal variants were chosen.

For the simulations that investigated nonsparsity of causal effects, we drew additional causal variants uniformly at random such that x% ($x \in \{0.5,1,5\}$) of all simulated variants have a nonzero effect. For each selected variant, we sampled its raw effect size (to be scaled later) from $\mathcal{N}(0,v)$ where $v = [2p(1-p)]^{\alpha}$, p represents the MAF, and $\alpha = -0.38$. The value α is estimated in ref. 42. For all simulation settings, simulated nonsparse effects had an overall effect size standard deviation approximately on the order of 1×10^{-4} units per normalized genotype.

We simulated three settings of nonsparsity coverage: 0.5%, 1% and 5%, where coverage is the percentage of all simulated variants with nonzero effects on the phenotype. For the simulations with 1% coverage, we varied the heritability explained by the nonsparse causal variants and it was set to be 58%, 75%, 83% and 100%, corresponding to heritability ratios between sparse and nonsparse causal effects of 1 to 1.4, 1 to 3, 1 to 5 and 0 to 1. To achieve these heritability proportions, we scaled all the simulated sparse and nonsparse causal effect sizes by corresponding constants. We observed that, for all simulation settings, simulated large effects had an overall effect size standard deviation approximately on the order of 1×10^{-2} units per normalized genotype. For the simulations with 0.5% coverage and one set of simulations with 5% coverage, we fixed the heritability ratio to 1 to 3. We performed an additional set of simulations with nonsparsity coverage of 5% and heritability ratio between sparse and nonsparse causal effects of 1 to 15. The purpose of this setting is to match the simulated per-SNP heritability with the 1% coverage 1-to-3 ratio simulations. For a summary of the different settings, see Table 1. We set the total SNP heritability to be 0.5. Note that the 0.5 heritability accounts for all simulated causal SNPs and not just the common SNPs. We have computed stratified LD score regression (s-LDSC) estimates of common SNP heritability for all the simulations and all ten UKBB phenotypes, and the results are available in Supplementary Tables 1 and 2.

Interestingly, changing the coverage of nonsparsity from 0.5% to 1% then to 5% while fixing the proportion of heritability explained by nonsparse effects showed a nonmonotonic behavior in the level of miscalibration. This is probably due to multiple factors influencing calibration: per-SNP heritability of nonsparse effects and LD between nonsparse and sparse causal variants. We observed increased miscalibration when per-SNP heritability is fixed and coverage changes from 1% to 5%. Similarly, when coverage is fixed at 1% and per-SNP heritability increases by 50% calibration also worsens (Fig. 2 and Extended Data Fig. 2).

Simulated population stratification

To simulate population stratification, we first regressed UKBB height on the top 20 PCs of the genotyped variants for $N = 360{,}415$ individuals. We then added the sum of the PC scores multiplied by their respective regression coefficients to the simulated phenotype, scaling this sum by a factor to vary the amount of simulated stratification. We assessed the amount of stratification by running s-LDSC[43] on the resulting GWAS summary statistics (without using any in-sample PCs as covariates) and examining the fitted intercept (Supplementary Table 2). As expected, we see higher s-LDSC intercept as we increase the PC scaling factor.

For the stratification simulations referenced in the main text and Table 1, we scaled PC effects by a factor of 5 (respectively 8) for moderate (respectively severe) stratification with BOLT, yielding a phenotype with 16.4% (respectively 42.9%) of its variance explained by stratification. For stratification with OLS, we scaled PC effects by 1 and 2 for moderate and severe stratification, yielding phenotypes with 0.6% and 2.6% of their variance due to stratification, respectively. s-LDSC intercepts of the stratification simulations are available in Supplementary Table 2.

Simulated phenotype

Phenotypes were generated as $y = X\boldsymbol{\beta} + C\boldsymbol{\zeta} + \epsilon$, where X is the above true genotype (pGT) matrix, $\boldsymbol{\beta}$ is a vector of the (sparse and nonsparse) causal effects, C is a matrix with top 20 PCs with corresponding effects $\boldsymbol{\zeta}$, and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ where σ^2 was chosen to yield total phenotypic variance equal to 1.

Incorporating missing causal variants

After generating phenotypes and before performing GWAS and fine-mapping, we applied variant-level quality control criteria as previously defined in the Neale Lab GWAS[21], which retained 13,364,303 variants after filtering for INFO >0.8, MAF >0.001 and Hardy–Weinberg equilibrium P value $>1 \times 10^{-10}$, with exception for the VEP-annotated coding variants where we allowed MAF $>1 \times 10^{-6}$. Notably, this quality control step resulted in the exclusion of approximately 71% of the simulated 'nonsparse' causal variants.

GWAS and fine-mapping for simulations

We performed GWAS on $N = 149{,}630$ individuals using BOLT–LMM[22] v2.3.2, with corresponding imputed variant dosages from UKBB. We used the top 19 PCs computed in-sample as covariates in the GWAS, except in the population stratification simulations, which included no covariates. For some of the population stratification simulations, we performed GWAS with OLS regression, rather than BOLT–LMM. We performed OLS using the linear regression rows method in Hail[44] v0.2.93. For fine-mapping we used the pipeline previously described in the subsection 'Fine-mapping pipeline'.

Interpreting population stratification simulation results

When scaling PC effects by a factor of 5 and computing GWAS summary statistics using BOLT–LMM, we observed an s-LDSC intercept of 1.07,

which is comparable to s-LDSC intercepts estimated in real complex traits (Supplementary Table 1), and we did not observe notable miscalibration in the downstream fine-mapping results. When we simulated a higher level of uncorrected stratification, scaling PC effects by a factor of 8 (s-LDSC intercept of 1.16, see 'Severe stratification with BOLT' in Table 1), PIPs obtained in downstream fine-mapping remained well calibrated (Fig. 3).

We hypothesize that the use of BOLT–LMM in our standard fine-mapping pipeline helped to correct for the simulated stratification effects, even though the in-sample PCs were not explicitly provided as covariates. This also probably explains the prima facie surprising recall results in Fig. 3, where the severe stratification simulations with BOLT have higher recall than the moderate stratification simulations with BOLT. In the severe simulations, stratification accounts for 42.9% of the phenotypic variance, whereas in the moderate simulations stratification accounts for only 16.4% of phenotypic variance. Because BOLT–LMM probably corrects for much of this simulated stratification, it effectively reduces the residual noise in the associations by much more for the severe simulations than for the moderate ones, allowing fine-mapping to nominate more causal variants.

To investigate stratification effects without using an LMM procedure, we performed two additional sets of simulations where GWAS summary statistics were computed using OLS instead. In these simulations, scaling PC effects by factors of 1 and 2 yielded average s-LDSC intercepts of 1.055 and 1.295, respectively (Supplementary Table 2), and induced notable miscalibration across all methods. This miscalibration was more severe for SuSiE-inf and FINEMAP-inf than for SuSiE and FINEMAP (Fig. 3).

It is unclear to us which of these simulation settings may be closer to reflecting the possible effects of uncorrected stratification in real fine-mapping applications, given that common methods of computing GWAS summary statistics do use LMM procedures and, in addition, explicitly control for in-sample PCs as covariates. Our real data results, including functional enrichment and PRS analyses, in UKBB show evidence that SuSiE-inf and FINEMAP-inf outperform existing methods in real data. We leave to future work a fuller investigation of the possible effects of uncorrected stratification on downstream fine-mapping, and a potential extension of these methods to address uncorrected stratification.

Regression of phenotype components on high-PIP variants

To identify which of several simulated model misspecifications were responsible for observed miscalibration, we decomposed the simulated genetic component $X\beta$ of the phenotype into the sum of four subcomponents representing sparse causal effects, nonsparse causal effects, nonsparse causal effects due to quality control and the effects of stratification. That is,

$$X\boldsymbol{\beta} = X\boldsymbol{\beta}_{\text{sparse}} + X\boldsymbol{\beta}_{\text{nonsparse}} + X\boldsymbol{\beta}_{\text{missing.nonsparse}} + XW\boldsymbol{\zeta}$$

where W is an $n \times 20$ matrix of UKBB PC loadings computed at a sample size of 360,415 and $\boldsymbol{\zeta}$ is a 20×1 vector of regression coefficients for the top 20 PCs in UKBB height at 360,415. For each simulation, we regressed each of the four genetic effect subcomponents on each of the PIP >0.9 variants independently, with 19 in-sample (n = 149,630) PCs as covariates in the regression (that is, the same covariates we used in GWAS in our simulations). For example, for the sparse genetic effect component, we compute the regression coefficient b and its associated F statistic for the following equation:

$$X\boldsymbol{\beta}_{\text{sparse}} = X_i b + C\mathbf{A}$$

where i is the index of a high-PIP (PIP >0.9) variant and C is a matrix that consists of 19 in-sample PCs, included here as covariates. $\mathbf{A}$ is a 19-by-1 vector of regression coefficients. We then compare the F statistics of truly causal and noncausal variants.

PRS cohort assignment

We used six ancestry groups derived by the Pan UKBB project:[29] European ancestry (N = 420,531, training 366,194 and testing 54,337), Central/South Asian ancestry (N = 8,876), African ancestry (N = 6,636), East Asian ancestry (N = 2,709), Middle Eastern ancestry (N = 1,599) and admixed American ancestry (N = 980). The 1,000 Genomes Project and Human Genome Diversity Panel were used as reference panels to assign continental ancestry.

PRS weights

We chose seven phenotypes: HbA1c, height, LDLC, Lym, Plt, RBC and eBMD for PRS predictions. We fine-mapped these seven phenotypes on the training cohort: European ancestry (N = 366,194 unrelated 'white British' individuals). SuSiE, FINEMAP, SuSiE-inf and FINEMAP-inf posterior effect sizes were obtained. Due to differences in computational efficiency, not all variants that are eligible for fine-mapping were able to be fine-mapped by all methods. To ensure fair comparison between SuSiE and SuSiE-inf (resp. FINEMAP and FINEMAP-inf), we include only variants that were fine-mapped by both SuSiE and SuSiE-inf (resp. FINEMAP and FINEMAP-inf) in the PRS analyses (Supplementary Table 3). PLINK2.0 (ref. 45) was then used to compute PRS for the six held-out cohorts using these posterior effect sizes. For SuSiE-inf and FINEMAP-inf, the assigned weight to each variant is the estimated posterior (mean) effect size from the sparse component β and does not include the estimated posterior effect size from the infinitesimal component α.

In case of minPIP or minPIP-inf, the assigned weight to each variant is the estimated posterior effect size of the variant outputted by methods whose PIP was taken as the minPIP. For example, if SNP1's PIP is 0.1 for SuSiE and 0.5 for FINEMAP, and the estimated posterior (mean) effect size for SNP1 is 0.01 for SuSiE and 0.005 for FINEMAP, then the minPIP for SNP1 is 0.1 and the weight we will use for SNP1 in PRS analysis for minPIP is 0.01.

PRS accuracy metric

We use ΔR^2 as our accuracy metric for PRS predictions[46]. To obtain ΔR^2, we fit two models:

- Model 0: a linear model using only covariates as predictor.
- Model 1: a linear model using true phenotype as target and both the PRS generated from multiplying the fine-mapped posterior effect size estimates with the genotypes and the covariates (sex, age, age^2, age and sex interaction term, age^2, and sex interaction term) as predictors.

We applied the function lm in R to obtain the adjusted R^2. The difference of adjusted R^2 of model 1 compared to adjusted R^2 of model 0 is defined as ΔR^2.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The main fine-mapping results at N = 100,000 sample size produced by this study are publicly available at https://doi.org/10.5281/zenodo.7055906. The fine-mapping results at N = 366,000 previously produced by our group are available at https://www.finucanelab.org/data. The UKBB individual-level data are accessible on request through the UKBB Access Management System (https://www.ukbiobank.ac.uk/). The UKBB analysis in this study was conducted via application number 31063.

Code availability

Software implementing SuSiE-inf and FINEMAP-inf are publicly available at https://github.com/FinucaneLab/fine-mapping-inf (https://doi.org/10.5281/zenodo.8427832). All scripts for figure

generation as well as simulation scripts are available at https://github.com/cuiran/improve-fine-mapping (https://doi.org/10.5281/zenodo.10037442).

References

38. Nealelab UK_Biobank_GWAS. *GitHub* https://github.com/Nealelab/UK_Biobank_GWAS/blob/master/ukb31063_eur_selection.R (2018).
39. FinucaneLab/fine-mapping-inf: fine-mapping-inf-published. *Zenodo* https://doi.org/10.5281/zenodo.10037442 (2023).
40. Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
41. mkanai/finemapping-pipeline. *Zenodo* https://zenodo.org/records/6908588 (2022).
42. Schoech, A. P. et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat. Commun.* **10**, 790 (2019).
43. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
44. Hail | Index. *Hail* https://hail.is/ (2023).
45. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
46. Weissbrod, O. et al. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nat. Genet.* **54**, 450–458 (2022).

Acknowledgements

This research has been conducted using the UKBB Resource under application number 31063. We acknowledge all the participants of UKBB. R.C., R.A.E. and H.K.F. are funded by DP5—5DP5OD024582-04, SFARI 704413 and 1U01HG011719-01. This work is also supported by the Novo Nordisk Foundation (NNF21SA0072102). M.K. is supported by a Nakajima Foundation Fellowship and the Masason Foundation. Z.F. is funded by NSF DMS-2142476. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank all the members of Finucane lab and Analytic and Translational Genetics Unit (ATGU) at Massachusetts General Hospital for their helpful feedback.

Author contributions

R.C. and H.K.F. conceived the idea of the RFR metric. Z.F. and H.K.F. conceived the idea of SuSiE-inf and FINEMAP-inf. Z.F. derived the methods and wrote the software. R.A.E. conducted the large-scale simulations and related analyses. M.K. provided GWAS and fine-mapping pipelines. R.C. conducted the analyses on UKBB and wrote the paper. J.C.U., O.W., B.M.N. and M.J.D. contributed to the analyses and helped interpret the results. Z.F. and H.K.F. jointly supervised this research and critically revised the paper.

Competing interests

J.C.U. is an employee of Illumina. O.W. is an employee and holds equity in Eleven Therapeutics. B.M.N. is a member of the scientific advisory board at Deep Genomics and Neumora, consultant of the scientific advisory board for Camp4 Therapeutics and consultant for Merck. The other authors declare no competing interests.

Additional information

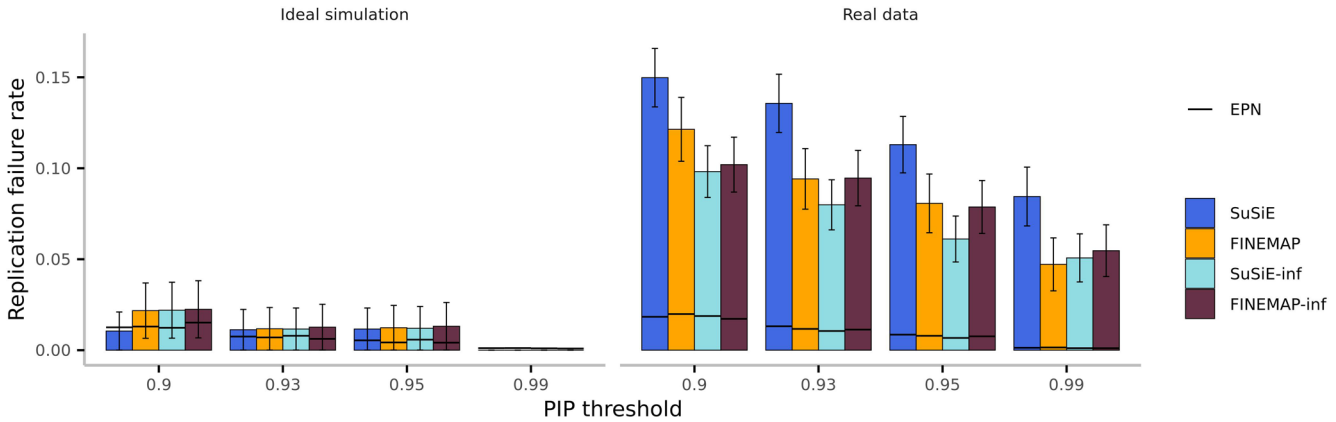
Extended data Available for this paper at https://doi.org/10.1038/s41588-023-01597-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-023-01597-3.

Correspondence and requests for materials should be addressed to Ran Cui, Zhou Fan or Hilary K. Finucane.

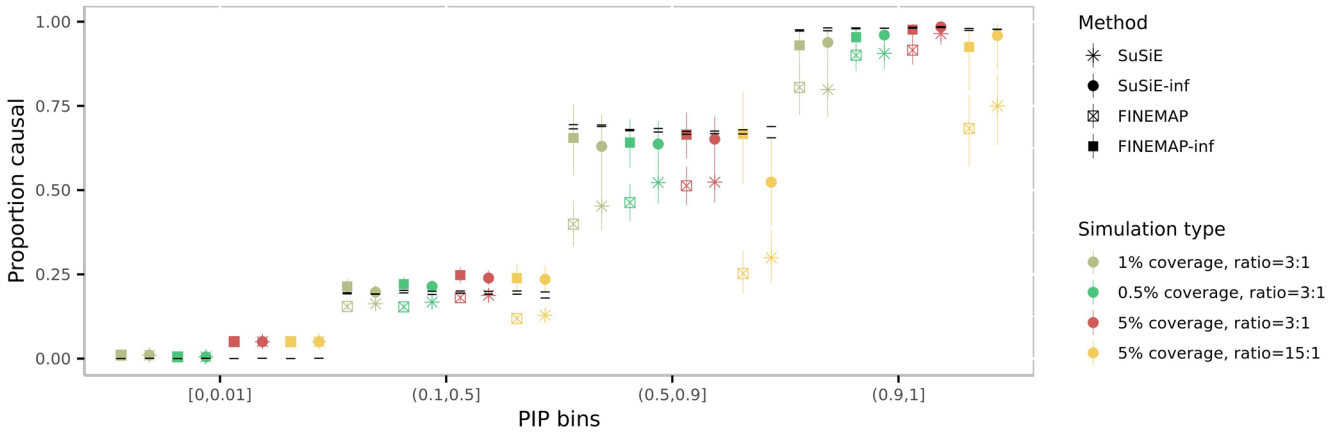
Peer review information *Nature Genetics* thanks the anonymou(s) reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-023-01597-3



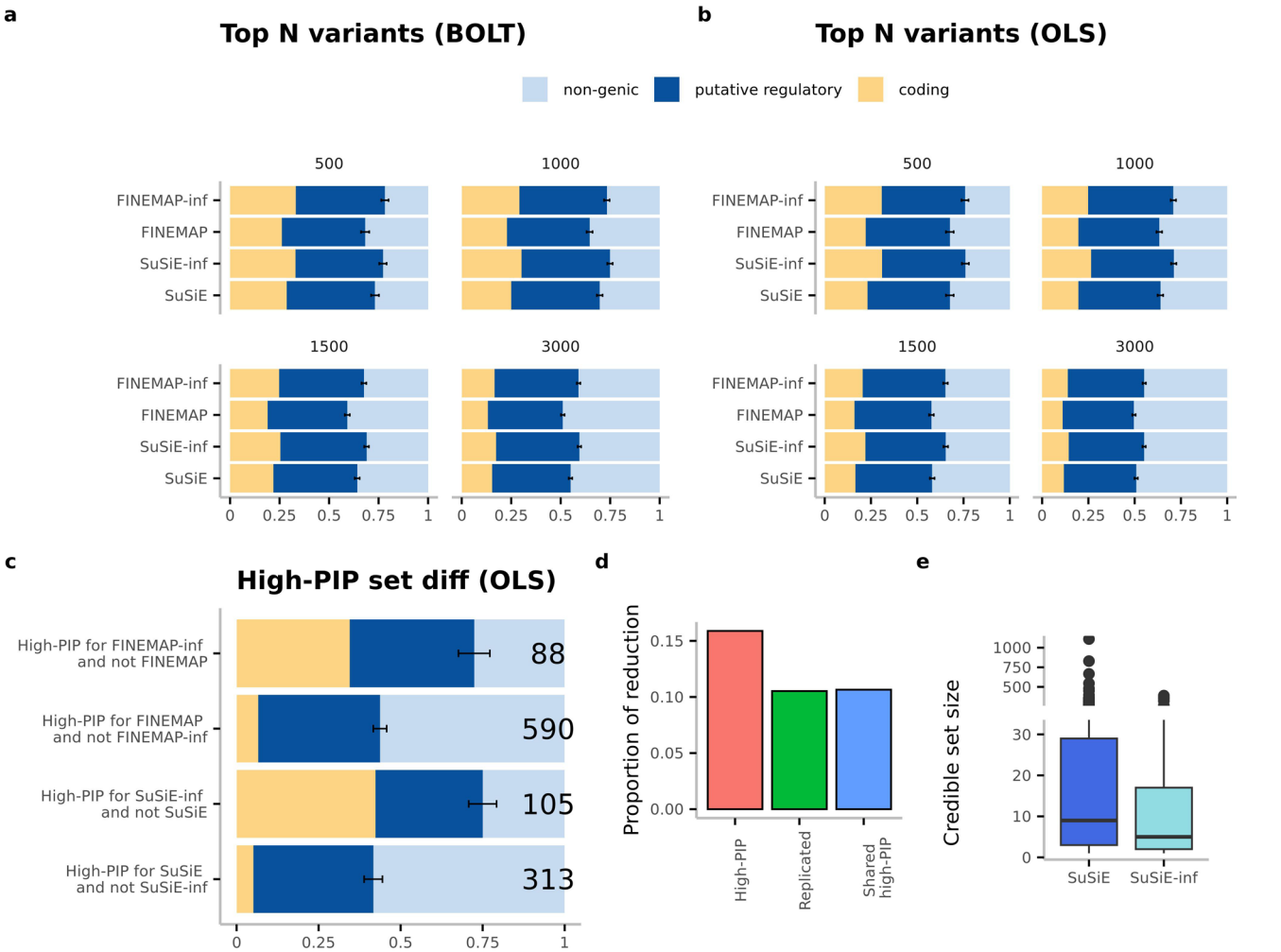
Extended Data Fig. 1 | Replication failure rates at different PIP thresholds. Replication failure rates at four different PIP thresholds: 0.9 (default), 0.93, 0.95, 0.99, for SuSiE, FINEMAP, SuSiE-inf, and FINEMAP-inf aggregated across 10 UKBB quantitative phenotypes, contrasted with RFRs in ideal simulations and with EPN. Error bars represent one SD of the corresponding binomial distribution Binom(n, p), where n is the total number of high-PIP variants at sample size N = 100 K, and p is the RFR. Bar plot data is presented as RFR + /- SD. Numerical results are available in Supplementary Table 14.

https://doi.org/10.1038/s41588-023-01597-3



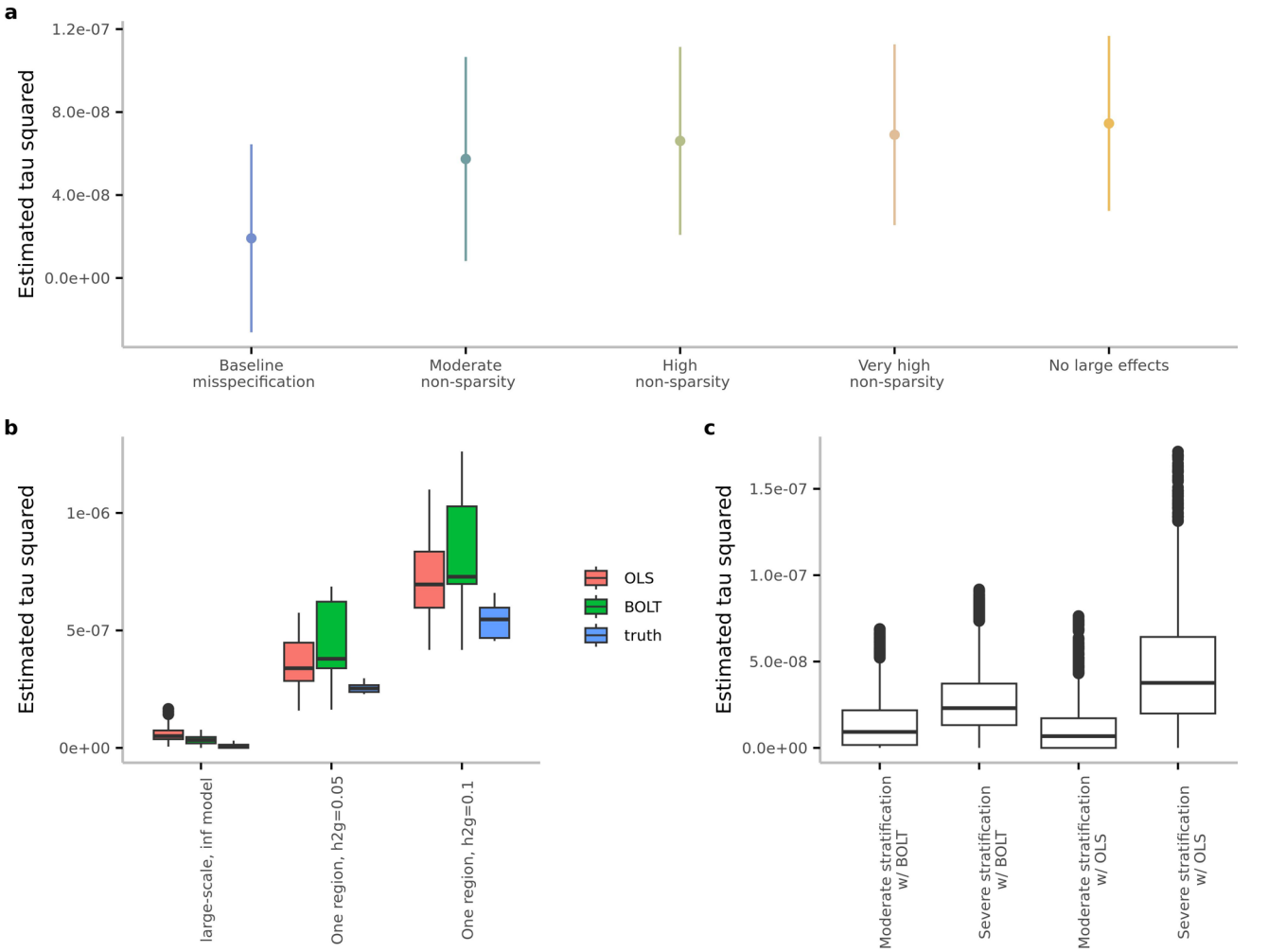
Extended Data Fig. 2 | Calibration in different non-sparsity coverage settings. Calibration for SuSiE, SuSiE-inf, FINEMAP and FINEMAP-inf in simulations with different non-sparsity coverage settings: 0.5%, 1%, and 5% (see Table 1 for more parameter settings in these simulations). Heritability ratio between small and large effects is fixed at 3:1 for three simulation scenarios, while the fourth scenario we set the coverage at 5% and heritability ratio at 15:1 to match the per-SNP heritability in simulations where 1% SNPs are causal and heritability ratio is 3:1. Error bars correspond to 95% Wilson confidence interval. Numerical results available in Supplementary Table 15.

https://doi.org/10.1038/s41588-023-01597-3



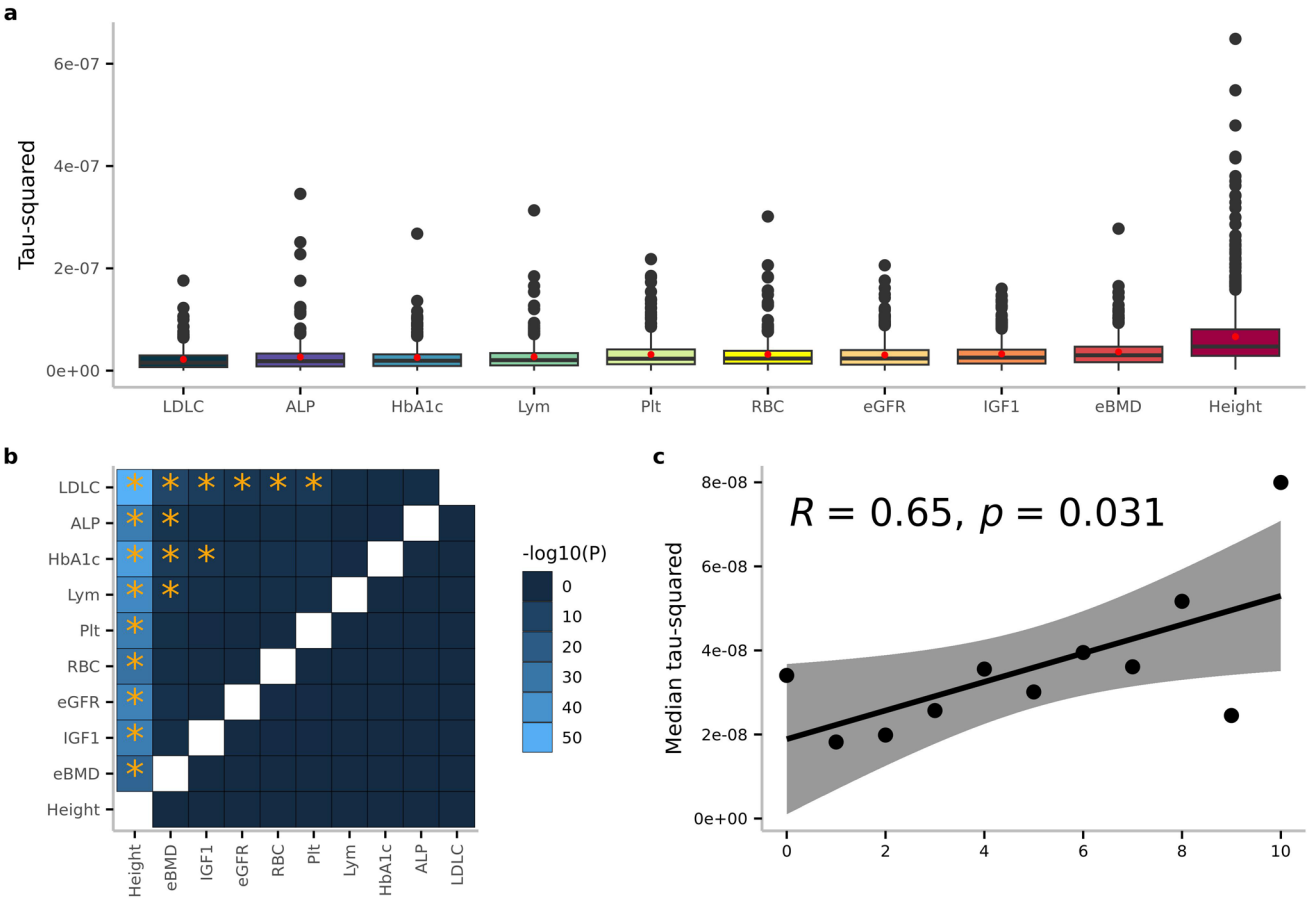
Extended Data Fig. 3 | Additional evidence of performance improvements in real data. a-b. Functional enrichment of top N (N = 500, 1000, 1500, and 3000) highest PIP variants from SuSiE, SuSiE-inf, FINEMAP, and FINEMAP-inf. GWAS summary statistics computed using BOLT-LMM and OLS. **c.** Functional enrichment of the set differences between SuSiE and SuSiE-inf high-PIP (PIP > 0.9) variants and FINEMAP and FINEMAP-inf high-PIP variants. Error bars represent one SD of the corresponding binomial distribution Binom(n,p), where n is the total number of variants in each set and p is the corresponding proportion of annotated variants). Bar plot data is presented as proportion +/- SD. **d.** The proportion of reduction for the number of variants in three categories when using the SuSiE-inf and FINEMAP-inf compared to using SuSiE and FINEMAP. The three categories are: High-PIP (PIP > 0.9 for either method, reduced from 1876 to 1578), Replicated (PIP > 0.9 at both sample sizes N = 100 K and N = 366 K, reduced from 665 to 595), and Shared high-PIP (PIP > 0.9 for both method, reduced from 723 to 646). **e.** Credible set sizes in all regions fine-mapped by SuSiE and SuSiE-inf. Box plot lower and upper hinges correspond to 1st and 3rd quantiles, whiskers extend no further than 1.5*IQR from the hinges, outliers are plotted as individual points, solid line in the boxes show medians. Numerical results available in Supplementary Table 16-20.

https://doi.org/10.1038/s41588-023-01597-3



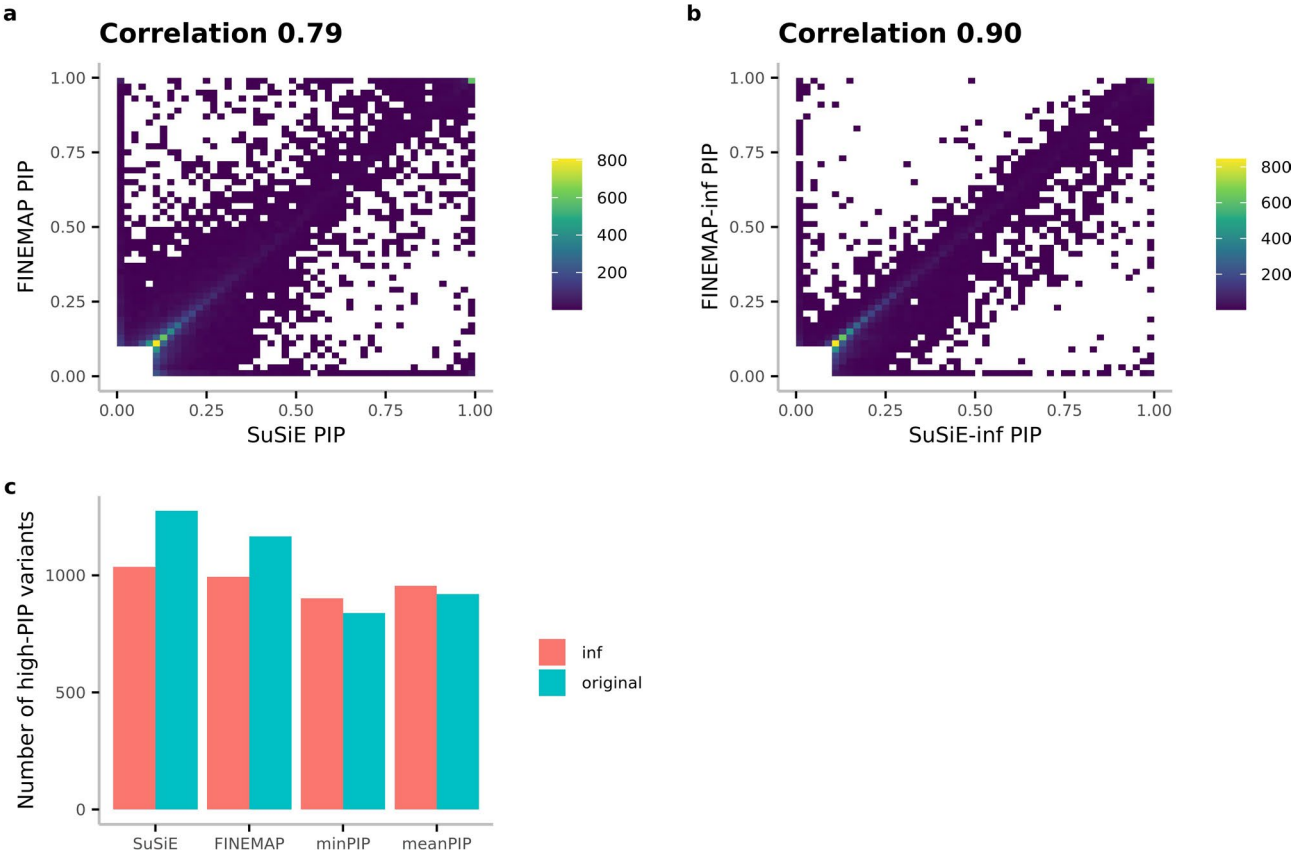
Extended Data Fig. 4 | Estimated infinitesimal variance (tau squared) in simulations. a. The mean estimated tau squared aggregated across all regions fine-mapped in each non-sparse simulation settings +/- SD, where SD is the in-sample standard deviation of estimated tau squared in the corresponding simulation setting. See Table 1 for simulation parameters. **b**. Estimated tau squared when OLS/BOLT-LMM is used to perform GWAS, and true tau squared in three sets of simulation settings are plotted. "Large-scale, inf model" represents the set of large-scale simulations (described in Methods) with 100% causal coverage setting and no missing causal variants are introduced. "One region, h2g = 0.05" represents the set of simulations using imputed genotypes in one region on Chromosome 1, with 100% causal coverage, no missing causal variants, and no exclusion of variants in the fine-mapping pipeline. The total SNP heritability is set to be 0.05. "One region, h2g = 0.1" is similar except with total SNP heritability set to be 0.1. **c**. Estimated tau squared in four stratification simulation settings with no non-sparse effects, see Table 1 for simulation parameters. Box plot lower and upper hinges correspond to 1st and 3rd quantiles, whiskers extend no further than 1.5*IQR from the hinges, outliers are plotted as individual points, solid line in the boxes show medians. Numerical results available in Supplementary Table 21-23.

https://doi.org/10.1038/s41588-023-01597-3



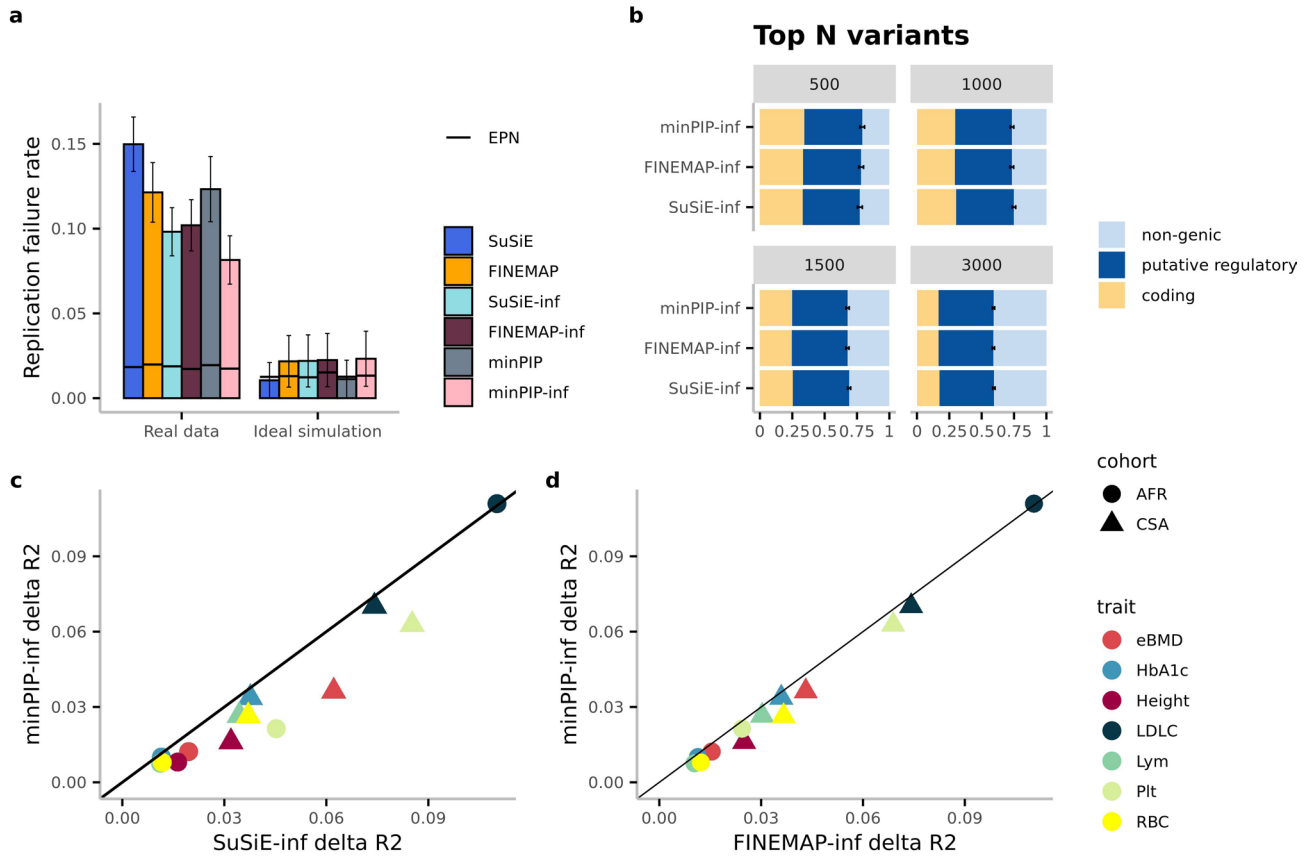
Extended Data Fig. 5 | Estimated infinitesimal variance (tau squared) in UK Biobank. a. Estimated tau squared in all fine-mapped regions for 10 UK Biobank phenotypes at sample size N = 366 K. Box plot lower and upper hinges correspond to 1st and 3rd quantiles, whiskers extend no further than 1.5*IQR from the hinges, outliers are plotted as individual points, solid line in the boxes show medians and the red dot denotes the mean. **b.** Comparing mean tau-squared estimates between traits. Heatmap shows the results of pair-wise Welch two-sample T-test with alternative hypothesis: mean of estimated tau squared in all regions for trait 1 (x-axis) is greater than that of trait 2 (y-axis). The test is one-sided. Multiple-testing adjusted p-value significance cutoff is set to be 0.05/90 = 5.5e-4, correcting for the total number of trait pairs tested. Stars indicate p-value has passed the significant threshold. **c.** Correlation between number of credible sets and the estimated infinitesimal variance (tau squared). Regions with the same number of credible sets are aggregated, and the median estimated tau squared are obtained from these regions. Scatter plot shows these medians. The best fitted line is plotted using ggscatter. R is the Pearson correlation, and p is the two-sided correlation p-value. The 95% confidence interval is shown on the plot as the gray shaded area. Numerical results available in Supplementary Table 24-26.

https://doi.org/10.1038/s41588-023-01597-3



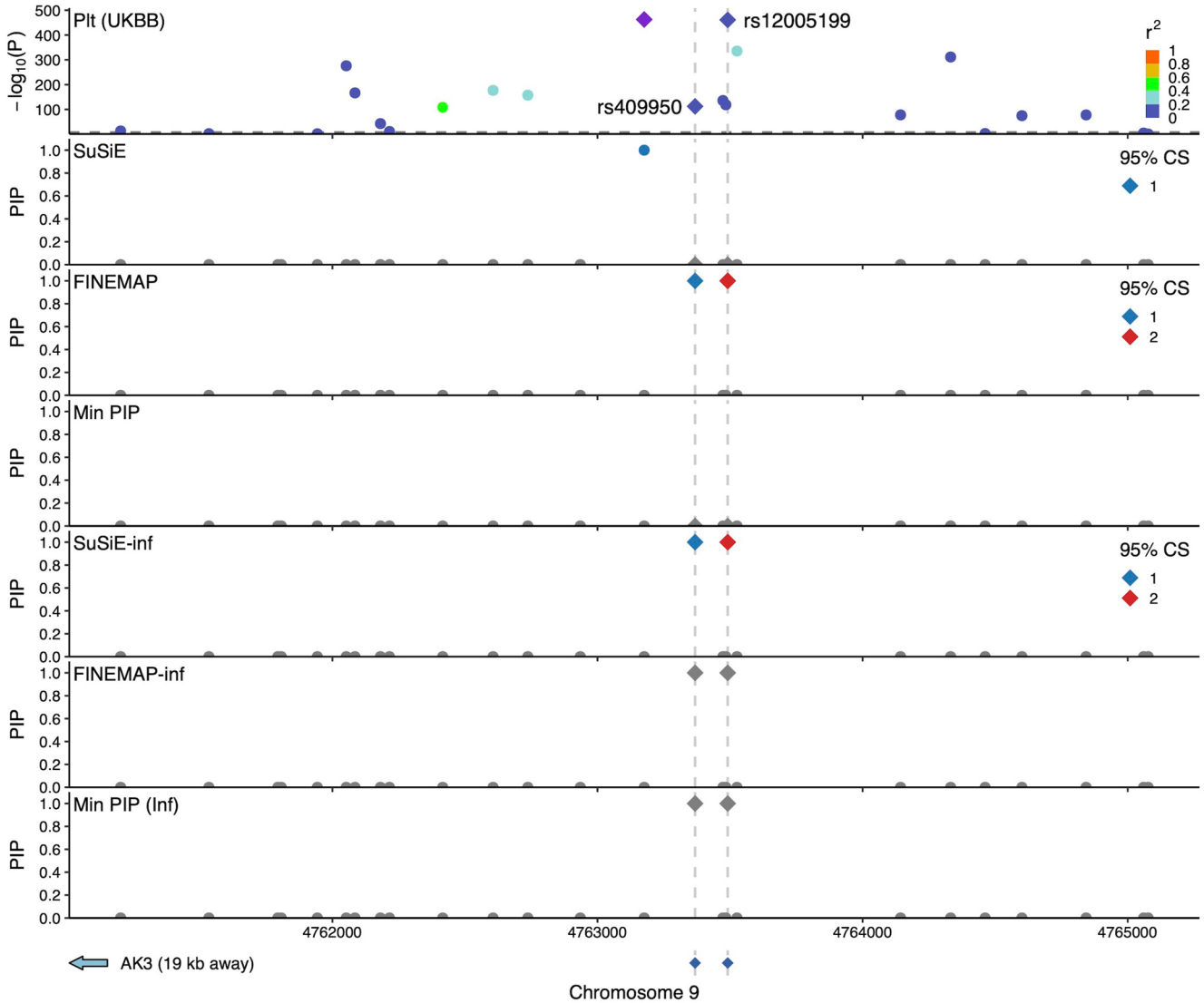
Extended Data Fig. 6 | Agreement between SuSiE and FINEMAP PIPs; SuSiE-inf and FINEMAP-inf PIPs. a-b. Density plots of PIPs from fine-mapping 10 UK Biobank traits at sample size N = 366 K. X-axis shows PIPs from running SuSiE (or SuSiE-inf), y-axis shows PIPs from running FINEMAP (or FINEMAP-inf). Only variants with PIP > = 0.1 for either method are shown on the plots. **c.** Number of high-PIP variants identified by SuSiE, SuSiE-inf, FINEMAP, FINEMAP-inf, minPIP, minPIP-inf, meanPIP and meanPIP-inf, where meanPIP(-inf) is defined as taking the average PIP between SuSiE and FINEMAP (resp. SuSiE-inf and FINEMAP-inf). Data aggregated across 10 UKBB traits fine-mapped at N = 366 K. Numerical results available in Supplementary Table 27.

https://doi.org/10.1038/s41588-023-01597-3



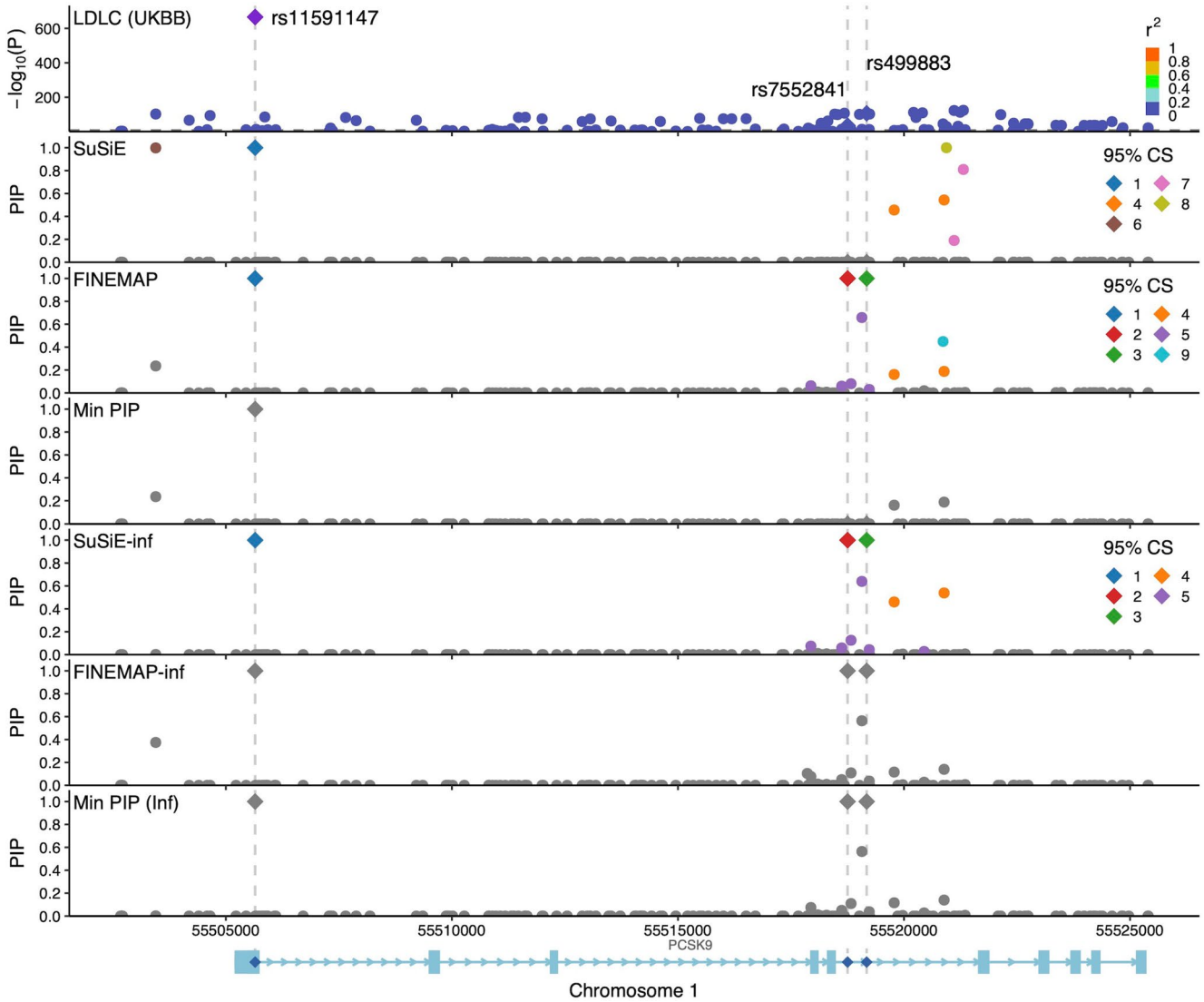
Extended Data Fig. 7 | minPIP-inf performance. a. Replication failure rates of minPIP and minPIP-inf in real data and in ideal simulations. Numerical results available in Supplementary Table 4. Error bars represent one SD of the corresponding binomial distribution Binom(n, p), where n is the total number of high-PIP variants at sample size N = 100 K, and p is the RFR. Bar plot data is presented as RFR + /- SD. **b.** Functional enrichment of top N (N = 500, 1000, 1500, and 3000) highest PIP variants from SuSiE-inf, FINEMAP-inf and minPIP-inf. Error bars represent one SD of the corresponding binomial distribution Binom(n,p), where n is the total number of variants in each set and p is the corresponding proportion of annotated variants). Numerical results available in Supplementary Table 16. **c-d.** PRS accuracy, in terms of delta R^2, when applying SuSiE-inf sparse component of the posterior effect sizes vs. minPIP-inf sparse component of the posterior effects sizes as weights; similarly, for FINEMAP-inf and minPIP-inf. PRS were computed for 2 out-of-sample cohorts and 7 traits. For descriptions of PRS weights, see Methods. Numerical results available in Supplementary Table 11-12.

https://doi.org/10.1038/s41588-023-01597-3



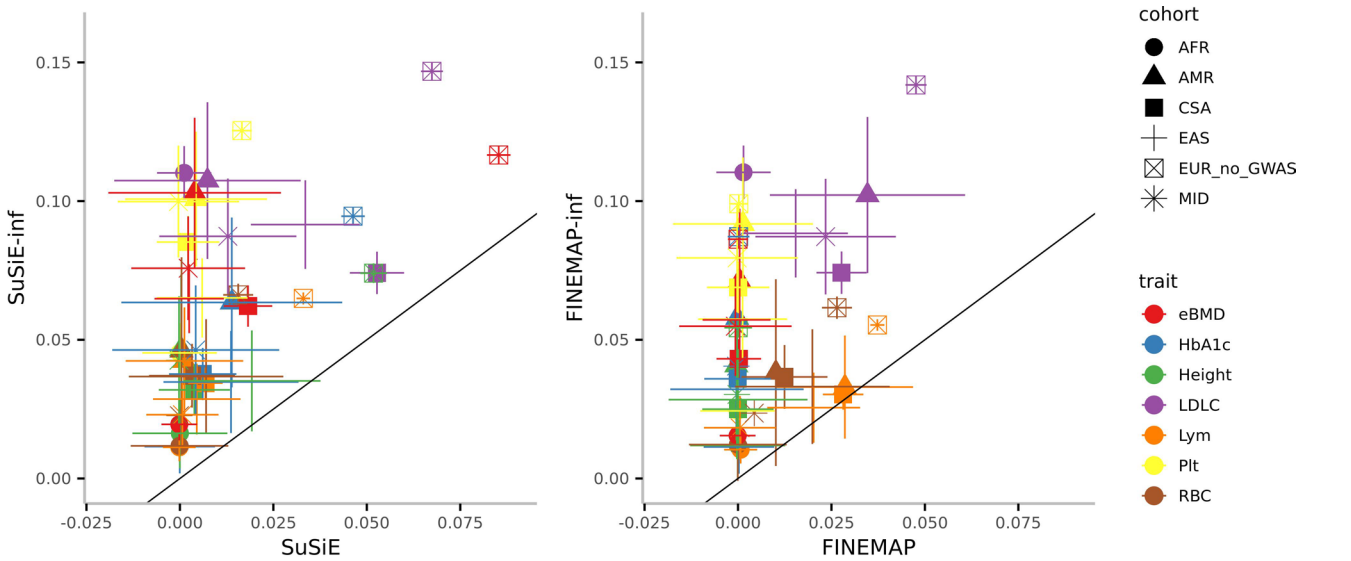
Extended Data Fig. 8 | AK3 locus for Plt. 4kbp window near the AK3 gene is shown on the plot. GWAS -log10 p-values from BOLT-LMM for the trait platelet count (Plt) are plotted on the top panel, PIPs from 4 fine-mapping methods and 2 aggregated methods are plotted on the subsequent panels. Variant rs12005199 and rs409950 are highlighted with dashed lines.

https://doi.org/10.1038/s41588-023-01597-3



Extended Data Fig. 9 | PCSK9 locus for LDLC. 23kbp window at the PCSK9 gene location is shown on the plot. GWAS -log10 p-values from BOLT-LMM for trait low density lipoprotein cholesterol (LDLC) are plotted on the top panel, PIPs from 4 fine-mapping methods and 2 aggregated methods are plotted on the subsequent panels. The well-known putative causal variant rs11591147 is highlighted with dash line, as well as two intronic variants: rs499883 and rs7552841.

https://doi.org/10.1038/s41588-023-01597-3



Extended Data Fig. 10 | PRS comparison with standard errors. a-b. Points represents the same values as in Fig. 5, standard errors represent 0.95 level confidence intervals of delta R^2. First, standard errors for the R^2 of Model 0 and Model 1 (defined in Methods) are computed separately using R function CI.rsq, then combined into the SE (standard error) of delta R^2 by taking the square root of the sum of squared. Data is presented as delta R^2 +/- SE for both axes.

nature portfolio

Corresponding author(s): Ran Cui

Last updated by author(s): Oct 10, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	SuSiE-inf and FINEMAP-inf software packages are available at https://github.com/FinucaneLab/fine-mapping-inf Access to the UK Biobank resource is available via application (http://www.ukbiobank.ac.uk/) SuSiE v0.9.1.0.894ba2f is available at https://github.com/stephenslab/susieR FINEMAP v1.3.1 is available at http://www.christianbenner.com LDStore v2.0 is available at http://www.christianbenner.com BOLT-LMM v2.3.2 is available at https://data.broadinstitute.org/alkesgroup/BOLT-LMM/ PLINK2 is available at https://www.cog-genomics.org/plink/2.0/ GCTA COJO v1.93.0beta is available at https://cnsgenomics.com/software/gcta/ fine-mapping pipeline scripts are available at https://doi.org/10.5281/zenodo.6908588

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data used for this study is approved under UK Biobank application number 31063. The main fine-mapping results at N=100K sample size produced by this study are publicly available at https://doi.org/10.5281/zenodo.7055906. The fine-mapping results at N=366K previously produced by our group is available at https://www.finucanelab.org/data. The UKBB individual-level data is accessible on request through the UK Biobank Access Management System (https://www.ukbiobank.ac.uk/).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Sex and gender-based analyses were not performed.
Population characteristics	The UK Biobank (UKBB) is a population-based cohort that recruited approximately 500,000 individuals in the United Kingdom between 2006 and 2010.
Recruitment	The UK Biobank recruited middle-aged (40–69 years old) volunteers across the United Kingdom.
Ethics oversight	UK Biobank

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The UK Biobank contains 500,000 participants with various ancestries. Our research analyzed 366,194 unrelated “white British” individuals. We chose this sample size to maximize the statistical power for single-cohort fine-mapping. We randomly downsampled this cohort to N=100K for Replication Failure Rate (RFR) part of the analyses. This sample size is chosen to resemble the sample size of the UK Biobank interim release dataset. We used as many samples as available in the UK Biobank after QC for the out-of-cohort PRS analyses: N=8876 Central/South Asian ancestry, N=6636 African ancestry, N=2709 East Asian ancestry, N=1599 Middle Eastern ancestry and N=980 Admixed American ancestry.
Data exclusions	We excluded "non-white-British" samples due to statistical complications when fine-mapping multiple ancestries. A combination of PCA based and self-reporting based quality controls are applied to exclude samples.
Replication	No replication dataset was analyzed as fine-mapping requires access to individual-level genotypic data from hundreds of thousands of individuals, which is generally not publicly available other than the UK Biobank.
Randomization	We randomly selected a subsample of N=100K out of the N=366K unrelated white-British samples in UK Biobank for the Replication Failure Rate (RFR) part of the analyses. No other randomization has been performed for this study. We included GWAS covariates age, age^2, inferred_sex, age * inferred_sex, age^2 * inferred_sex, and PCs 1-20 for association analyses at sample sizes N=366K and N=100K.
Blinding	We did not apply any blinding because no intervention was required in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging