Fine-mapping methods for complex traits: essential adaptations for samples of related individuals

Junjian Wang[1], Francesco Tiezzi[1,2], Yijian Huang[3], Julong Wei[4], Christian Maltecca[1], and Jicai Jiang[1*]

[1] Department of Animal Science, North Carolina State University, Raleigh, NC, 27695, USA

[2] Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Firenze, 50144, Italy

[3] Smithfield Premium Genetics, Rose Hill, NC, 28458, USA

[4] Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, 48201, USA

*Corresponding author.

E-mail address: jjiang26@ncsu.edu

Abstract

Fine-mapping causal variants from GWAS loci is challenging in populations with substantial relatedness, such as livestock, as standard methods often assume unrelatedness, leading to poor fine-mapping accuracy. Here, we introduce a comprehensive Bayesian framework to address this. Our approach features BFMAP-SSS for individual-level data, which uses a linear mixed model (LMM) and shotgun stochastic search with simulated annealing. For summary statistics, we develop FINEMAP-adj and SuSiE-adj, novel strategies that directly use standard FINEMAP and SuSiE for samples of related individuals by employing LMM-derived inputs (particularly a relatedness-adjusted linkage disequilibrium matrix). Furthermore, gene-level posterior inclusion probability (PIP_{gene}) is proposed to enhance detection power by aggregating variant signals. Extensive simulations based on pig genotypes show our methods substantially outperform existing tools (FINEMAP, SuSiE, FINEMAP-inf, SuSiE-inf, and GCTA-COJO) in samples of related individuals, achieving notable improvements in fine-mapping accuracy (e.g., up to several-fold increases in AUPRC). PIP_{gene} markedly improves candidate gene identification. Application to

Duroc pig traits demonstrates practical utility. This work provides robust, validated methods and associated software and scripts for accurate fine-mapping in populations with complex relatedness.

Key words: fine-mapping; complex trait; Bayesian; related individuals

Introduction

Genome-wide association studies (GWAS) have successfully discovered many genetic associations with complex traits and diseases (1). However, these studies primarily identify associated genomic regions rather than pinpoint the actual causal variants, largely due to linkage disequilibrium (LD), whereby non-causal variants are correlated with true causal variants (2, 3). Such associated regions often contain many SNPs, the majority of which have no direct effect on the trait (4), thereby complicating downstream functional validation. Fine-mapping aims to address this by statistically prioritizing the most likely causal variants within a GWAS-identified locus, assuming at least one such variant exists (5) and thus narrowing the field for functional studies.

Statistical fine-mapping methods typically account for LD structure to identify a small set of putative causal variants, effectively treating the problem as variable selection in a regression model (6, 7). Given m SNPs in a region, 2^m possible causal models exist, making exhaustive evaluation computationally prohibitive for densely genotyped regions. Early approaches like conditional stepwise selection (e.g., GCTA-COJO) (8) identify independent signals but can be suboptimal, particularly when multiple causal variants in LD are present or when a non-causal variant strongly tags multiple causal variants (5, 7). Methods like CAVIAR (7) and its extension CAVIARBF (9) improved upon this by jointly modeling multiple variants and exhaustively evaluating all possible models with up to a pre-specified small number of causal variants, but faced computational limitations with larger regions or higher numbers of true causal variants. To address

this challenge, newer approaches were developed. FINEMAP (10) employs shotgun stochastic search (SSS) (11) to efficiently explore the vast model space, while SuSiE uses Iterative Bayesian Stepwise Selection (IBSS) with the "Sum of Single Effects" model (12, 13) to search multiple effects. These methods have become widely adopted due to their improved computational efficiency. Recent extensions like FINEMAP-inf and SuSiE-inf (14) further refine these by attempting to model infinitesimal effects within the candidate region.

Despite these advancements, a major limitation of most existing fine-mapping methods, particularly those based on summary statistics, is their foundation in linear regression models that assume samples of unrelated individuals. The use of related individuals violates the underlying model assumption of unrelatedness, potentially distorting LD and subsequent causal variant identification. While it has become standard to use summary statistics derived from linear mixed models (LMMs) that account for relatedness and/or infinitesimal background at the GWAS stage (15, 16), the downstream fine-mapping tools often still do not correctly incorporate this relatedness structure into their own models. This critical issue of appropriately adapting fine-mapping for related individuals has particularly been overlooked in livestock genetics, where populations typically exhibit substantial and complex inter-individual relatedness due to breeding practices.

This study introduces a comprehensive Bayesian fine-mapping framework specifically designed to address these challenges in samples of related individuals. We aim to demonstrate that existing summary-statistics-based fine-mapping methods cannot be naively applied to samples of related individuals and develop comprehensive methodological solutions for related samples. We build upon our previously developed BFMAP approach, which uses an LMM for individual-level data and originally employed forward selection (BFMAP-Forward) (17). Here, we enhance the BFMAP framework by implementing a more robust model exploration strategy, BFMAP-Shotgun

Stochastic Search (BFMAP-SSS), to better handle scenarios with highly correlated causal variants. Recognizing the widespread use of summary statistics, we further develop novel adaptation strategies, FINEMAP-adj and SuSiE-adj. These approaches enable the direct use of standard FINEMAP and SuSiE software with summary statistics obtained from related individuals by inputting LMM-derived inputs and a relatedness-adjusted LD matrix. Finally, to improve detection power in genomic regions with challenging LD patterns, as is common in livestock, we propose and validate the use of gene-level posterior inclusion probability (PIP_{gene}) that aggregates variant-level evidence.

Through simulations based on real livestock genotypes and application to economically important traits in Duroc pigs, we demonstrate both the inadequacy of naively applying existing summary-statistics methods to related individuals and the substantial improvements offered by our comprehensive methodological solutions (BFMAP-SSS, FINEMAP-adj, SuSiE-adj, and PIPgene) in fine-mapping performance. This work provides both novel methodologies and practical adaptations essential for robust identification of causal variants and trait-associated genes in samples with complex genetic relatedness.

Materials and methods

Fine-mapping with unrelated individuals

For a sample of unrelated individuals, the genetic fine-mapping analysis can be represented using the following linear model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \text{ with } \mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2) \text{ and } \mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2), \quad (1)$$

where $\mathbf{y}$ is an n-vector of covariate-adjusted phenotypes for n individuals, $\mathbf{X}$ is an n-by-m matrix of standardized genotypes for n individuals and m SNPs in a genomic region of interest, $\mathbf{b}$ is an m-

vector of joint SNP effects, $\mathbf{e}$ is an n-vector of residual effects, and φ and σ_e^2 are hyperparameters related to variance components. Specifically, φ represents the ratio of per-SNP genetic variance to residual variance (σ_e^2), and we can set φ=0.01 to specify a moderate prior variance for SNP effects. Note that model (1) does not account for genetic relatedness between individuals, restricting its application to samples of unrelated individuals.

Model (1) can be expressed as follows:

$$\mathbf{y} \sim \mathcal{N}(0, \sigma_e^2(\mathbf{X}\mathbf{X}^{\mathrm{T}}\varphi + \mathbf{I})), \tag{2}$$

which is further transformed to $\mathbf{X}^{\mathrm{T}}\mathbf{y}/\sqrt{n\sigma_e^2} \sim \mathcal{N}(0, \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{X}\varphi/n + \mathbf{X}^{\mathrm{T}}\mathbf{X}/n)$. Denoting $\mathbf{z} = \mathbf{X}^{\mathrm{T}}\mathbf{y}/\sqrt{n\sigma_e^2}$ and $\mathbf{\Sigma} = \mathbf{X}^{\mathrm{T}}\mathbf{X}/n$ (we ignore the error introduced by the estimation of σ_e^2 and the use of n instead of $n-1$), we have

$$\mathbf{z} \sim \mathcal{N}(0, \mathbf{\Sigma}\mathbf{\Sigma}\varphi n + \mathbf{\Sigma}), \tag{3}$$

where $\mathbf{z}$ is an m-vector of z-scores from linear regression associations and $\mathbf{\Sigma}$ is an m-by-m matrix of Pearson's correlations between m SNPs. It is important to note that model (3) holds for any set of SNPs in the genomic region. For a set of SNPs (denoted as s), let $\mathbf{z}_s$ and $\mathbf{\Sigma}_s$ represent the z-score vector and correlation matrix, respectively. Given model (3), we can compute the Bayes factor of the model (denoted as M) for s over the reduced model including no SNPs:

$$\mathrm{BF}_M = |n\varphi\mathbf{\Sigma}_s + \mathbf{I}|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{z}_s^{\mathrm{T}}(\mathbf{I}n^{-1}\varphi^{-1} + \mathbf{\Sigma}_s)^{-1}\mathbf{z}_s\right). \tag{4}$$

Equation (4) demonstrates the computation of Bayes factors for multi-SNP models using GWAS summary statistics and reference panel SNP correlations. This approach forms the foundation for genetic fine-mapping tools like CAVIARBF (9) and FINEMAP (10). Compared to individual-level data calculations under model (2), this method offers enhanced computational

efficiency, particularly when leveraging reference panels like the 1000 Genomes Project (18) for SNP correlation estimation. However, like model (2), this approach is restricted to unrelated individuals.

Genetic fine-mapping is fundamentally a model selection problem. Although some fine-mapping methods, such as GCTA-COJO (8), use *P*-values rather than Bayes factors, they maintain consistency with Bayesian approaches. Specifically, when GCTA-COJO identifies a small conditional *P*-value for a SNP, this typically corresponds to a substantially larger log(*BF*) for the model including this SNP than for the comparable model without this SNP. Bayesian fine-mapping methods, such as CAVIARBF (9) and FINEMAP (10), evaluate sufficiently many multi-SNP models to compute posterior inclusion probability (PIP) for each variant in the candidate genomic region.

SuSiE (13) introduces a novel approach by decomposing the overall genetic effect into a sum of individual single effects, rather than explicitly evaluating different combinations of multiple SNPs. This computational innovation, combined with the implementation of variational Bayes for approximate posterior inference, enables SuSiE to identify multiple causal SNPs with high computational efficiency. Despite these methodological advances, SuSiE operates within the same linear framework as CAVIARBF and FINEMAP.

Fine-mapping with related individuals: individual-level data

We previously developed a Bayesian framework for genetic fine-mapping with related individuals, which we refer to as BFMAP (17). BFMAP uses individual-level genotype and phenotype data based on the following linear mixed model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{g} + \mathbf{e} \text{ with } \mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2),\ \mathbf{g} \sim \mathcal{N}(0, \mathbf{G}\eta\sigma_e^2), \text{ and } \mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2). \quad (5)$$

Model (5) extends model (1) by including an additional random-effect term, **g**, to account for genetic relatedness and whole-genome infinitesimal effects. **G** represents the genomic relationship matrix (GRM) (19), and η denotes the ratio of additive genetic variance to residual variance (σ_e^2). Other terms remain as defined in model (1).

For model implementation, we need to specify both η and σ_e^2. When the effects of candidate variants (**Xb**) account for only a small proportion of phenotypic variance (typically true when modeling variants from a small genomic region), we can set the value of η based on the SNP heritability (h^2) by $\eta = h^2/(1-h^2)$. In practice, we can use the SNP h^2 estimate from the null model without the **Xb** term to determine η. For the residual variance σ_e^2, we assign the non-informative Jeffrey's prior, $P(\sigma_e^2) \propto 1/\sigma_e^2$.

For any model M that specifies a set of SNPs to be included, let s denote this SNP set. Given $D = \{\mathbf{y}, \mathbf{X}, \mathbf{G}, \varphi, \eta\}$, we have the following equation for the marginal likelihood after integrating out σ_e^2:

$$P(D|M) \propto P(\mathbf{y}|\mathbf{X}, \mathbf{G}, \varphi, \eta, M) \propto (\mathbf{y}^{\mathrm{T}}(\varphi\, \mathbf{X}_s\mathbf{X}_s^{\mathrm{T}} + \eta\mathbf{G} + \mathbf{I})^{-1}\mathbf{y})^{-\frac{1}{2}n}|\varphi\, \mathbf{X}_s\mathbf{X}_s^{\mathrm{T}} + \eta\mathbf{G} + \mathbf{I}|^{-\frac{1}{2}}, \quad (6)$$

where $\mathbf{X}_s$ denotes the genotype matrix for SNP set s. The effective computation of $P(D|M)$ has been detailed in our previous study (17).

Computing $P(D|M)$ for every possible model is computationally infeasible. We have implemented two methods in BFMAP to facilitate model selection: one based on forward selection and the other using shotgun stochastic search (SSS) (11), which we refer to as BFMAP-Forward and BFMAP-SSS, respectively. The implementation of BFMAP-Forward has been detailed in our previous study (17). The forward selection procedure sequentially adds independent signals to the model, repositions them for refinement, and identifies variants related to each signal. For BFMAP-

SSS, we largely follow the implementation in FINEMAP (10). SSS efficiently explores the vast model space by evaluating neighboring models at each iteration. However, a naïve implementation may fail to discover causal variants even with a very long chain for certain LD structures. To address this limitation, we incorporate simulated annealing (20) to ensure more thorough search. We apply the following linear cooling scheme:

$$T_k = T_0 - k \cdot \Delta_T, \tag{7}$$

where T_0 is the initial temperature, T_k is the temperature at the k-th step, and Δ_T is the temperature decrement per step. For all analyses in this study, T_0 was set to 100, and Δ_T was fixed at 1. We performed 100 SSS iterations at each temperature step and set the final temperature to 1. An additional 100 SSS iterations were conducted at this final temperature to ensure thorough exploration of the model space.

Fine-mapping with related individuals: summary statistics

In this section, we present a summary-statistics-based fine-mapping method for related individuals, following the framework of BFMAP (17). Denoting $\mathbf{H} = \mathbf{G}\eta + \mathbf{I}$, we transform model (5) to:

$$\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{y} \sim \mathcal{N}\left(0, \sigma_e^2(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\varphi + \mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X})\right). \tag{8}$$

Defining a diagonal matrix $\mathbf{D} = \mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X})$ and an m-by-m matrix $\widetilde{\mathbf{\Sigma}} = \mathbf{D}^{-1/2}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\mathbf{D}^{-1/2}$, we obtain from (8):

$$\mathbf{z} \sim \mathcal{N}\left(0, \widetilde{\mathbf{\Sigma}}\mathbf{D}\widetilde{\mathbf{\Sigma}}\varphi + \widetilde{\mathbf{\Sigma}}\right), \tag{9}$$

where $\mathbf{z} = (\mathbf{D}\sigma_e^2)^{-1/2}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{y}$ (that is, an m-vector of z-scores from linear mixed model associations). For a set of SNPs (denoted as s) in a candidate genomic region, we can readily derive the Bayes factor of the model (denoted as M) for s over the reduced model including no SNPs:

$$\mathrm{BF}_M = \left|\mathbf{D}_s\widetilde{\mathbf{\Sigma}}_s\varphi + \mathbf{I}\right|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{z}_s^{\mathrm{T}}\left(\mathbf{D}_s^{-1}\varphi^{-1} + \widetilde{\mathbf{\Sigma}}_s\right)^{-1}\mathbf{z}_s\right), \tag{10}$$

where $\mathbf{z}_s$, $\mathbf{D}_s$, and $\widetilde{\mathbf{\Sigma}}_s$ represent $\mathbf{z}$, $\mathbf{D}$, and $\widetilde{\mathbf{\Sigma}}$, respectively, corresponding to SNP set s.

It is important to clarify the treatment of σ_e^2 in equation (10). While our BFMAP framework (e.g., equation (6)) involves integrating out σ_e^2 using its Jeffrey's prior, the z-scores used as input for equation (10) inherently treat σ_e^2 as a fixed quantity. This is because z-scores are standardized statistics that already incorporate an estimate of σ_e^2. However, this difference in treatment is generally of little practical consequence. For reasonably large samples (n > 1,000), the estimate of σ_e^2 is highly precise. Therefore, using z-scores that embed a point estimate of σ_e^2 typically yields results very similar to those that would be obtained from a full Bayesian integration of σ_e^2.

Equation (10) has a similar form to equation (4). Specifically, while the genotype correlation matrix in equation (4) for unrelated individuals is defined as $\mathbf{\Sigma} = \mathbf{X}'\mathbf{X}/n$, this is replaced in equation (10) for related individuals by $\widetilde{\mathbf{\Sigma}} = \mathbf{D}^{-1/2}\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}\mathbf{D}^{-1/2}$. $\widetilde{\mathbf{\Sigma}}$ can be regarded as relatedness-adjusted genotype correlations, which may substantially differ from $\mathbf{\Sigma}$. Using $\widetilde{\mathbf{\Sigma}}$ instead of $\mathbf{\Sigma}$ is critical for fine-mapping with related individuals. Misuse of unadjusted correlations can lead to poor performance.

Applying the approximation $\mathbf{D} \approx \mathbf{I}\tilde{n}$ (where $\tilde{n}$ is the average of the diagonal elements of $\mathbf{D}$ and $\mathbf{I}$ is the identity matrix) to equation (10) transforms it into equation (11):

$$\mathrm{BF}_M = \left|\tilde{n}\varphi\widetilde{\mathbf{\Sigma}}_s + \mathbf{I}\right|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\mathbf{z}_s^{\mathrm{T}}\left(\mathbf{I}\tilde{n}^{-1}\varphi^{-1} + \widetilde{\mathbf{\Sigma}}_s\right)^{-1}\mathbf{z}_s\right), \tag{11}$$

which has a structural form the same as equation (4). This approximation mirrors that used by GRAMMAR-Gamma (21), $\mathbf{X}_j^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}_j \approx c\mathbf{X}_j^{\mathrm{T}}\mathbf{X}_j = cn$, where $\mathbf{X}_j$ represents the vector of

standardized genotypes for SNP j and c is a constant across all SNPs. The structural similarity between equations (11) and (4) enables direct use of existing summary-statistics-based fine-mapping tools (e.g., FINEMAP and SuSiE) for related individuals. The use involves two steps: 1) calculate $\mathbf{z}$, $\tilde{n}$, and $\tilde{\mathbf{\Sigma}}$ as described in this section, and 2) input these as z-scores, sample size, and LD matrix, respectively, into FINEMAP and SuSiE. We refer to these adapted methods as FINEMAP-adj and SuSiE-adj.

Posterior probability metrics

Given the Bayes factor BF_M or the marginal likelihood $P(D|M)$ of model M (which defines a specific set of candidate causal SNPs, s), its posterior model probability is calculated as:

$$P(M|D) = \frac{P(D|M)P(M)}{\sum_{M'\in\Gamma} P(D|M')P(M')} = \frac{\mathrm{BF}_M P(M)}{\sum_{M'\in\Gamma} \mathrm{BF}_{M'} P(M')}, \tag{12}$$

where Γ represents the model space and M' represents any individual model in the model space. We use a standard binomial prior, $P(M) = \pi^{|M|}(1-\pi)^{m-|M|}$, to induce sparsity in BFMAP, where $|M|$ is the model size (number of SNPs in M) and π is the prior probability that a SNP is causal. This prior implies an expected model size of $m\pi$. Following standard fine-mapping practice (9, 10), π is set to $1/m$, reflecting an expectation of one causal variant *a priori*.

The posterior inclusion probability (PIP) quantifies the probability that a specific SNP has a non-zero effect, given the observed data and priors (22). For SNP j, the PIP is computed as:

$$\mathrm{PIP}_j = \sum_{M\in\Gamma} \mathbb{I}(j \in M) P(M|D), \tag{13}$$

where $\mathbb{I}(j \in M)$ is an indicator function returning 1 if j is included in model M and 0 otherwise. This definition of PIP is used by methods that explore a broad model space, such as FINEMAP, SuSiE, FINEMAP-inf, SuSiE-inf, FINEMAP-adj, SuSiE-adj, and BFMAP-SSS.

222 Our BFMAP-Forward implementation, however, operates via a stepwise forward selection
223 procedure. Consequently, it calculates a posterior conditional inclusion probability (PCIP). A
224 variant's PCIP is the posterior probability of the variant having a non-zero effect, conditional on
225 the set of variants already incorporated into the model. While PCIPs are by definition conditional
226 (unlike marginal PIPs), they serve as the direct analogue to PIPs for BFMAP-Forward throughout
227 this study.

228 Finally, to aggregate evidence at the gene level, we propose gene-level PIP ($\mathrm{PIP}_{\mathrm{gene}}$),
229 defined as the posterior probability that a given gene harbors at least one non-zero-effect variant.
230 This metric is computed by summing the posterior probabilities of all models containing at least
231 one SNP within the gene:

232 $$\mathrm{PIP}_{\mathrm{gene}} = \sum_{M \in \Gamma} \mathbb{I}(\mathrm{gene} \cap M \neq \emptyset) P(M|D), \quad (14)$$

233 where $\mathbb{I}(\mathrm{gene} \cap M \neq \emptyset)$ indicates whether any SNP in the gene is included in model M. This
234 gene-level aggregation offers a complementary perspective involving a trade-off between
235 localization precision and power. For $\mathrm{PIP}_{\mathrm{gene}}$ calculation, gene boundaries were defined using
236 Ensembl annotations and extended by 3kb upstream and downstream to include potential
237 regulatory regions.

238 **Data simulation**

239 We evaluated fine-mapping methods using genotype data from 31,301 Duroc pigs provided by
240 Smithfield Premium Genetics (Roanoke Rapids, NC, USA). The animals were genotyped with the
241 PorcineSNP60 BeadChip (Illumina Inc., San Diego, CA, USA) and genotypes were imputed to
242 34,615,361 autosomal variants using SWIM 1.0 (23). This full set of imputed variants was then
243 filtered for minor allele frequency (MAF) ≥1%, Hardy-Weinberg equilibrium ($P \geq 1 \times 10^{-8}$), and

244 IMPUTE2 info score ≥0.8, resulting in ~11.7 million quality-controlled variants on autosomes.
245 From these quality-controlled variants, we randomly selected a 4-Mb region on chromosome 1
246 containing 6,504 variants for the simulation study.

247 Quantitative traits were simulated using model (5) with a heritability of 0.5 using MPH
248 (24). The GRM **G**, used to model the covariance of polygenic effects due to genetic relatedness,
249 was computed using 10,000 randomly selected, quality-controlled chip SNPs. We positioned 1, 2,
250 or 3 causal variants within the central 2-Mb region, collectively explaining either 1% or 4% of
251 phenotypic variation (PVE). Simulations used samples of 5,000 or 10,000 randomly selected
252 individuals, generating 100 replicates for each of the 12 scenarios (3 causal variant counts × 2 PVE
253 levels × 2 sample sizes).

254 For gene-based fine-mapping evaluation, we partitioned the 4-Mb candidate region into
255 100 equal blocks to serve as gene proxies. This block size approximates the average pig gene
256 length of ~33 kb derived from the Sscrofa11.1 genome annotation (Ensembl release 113) (25). To
257 generate the required summary statistics for fine-mapping from each simulated dataset, we
258 conducted association tests for variants in the candidate region using two approaches: 1) a linear
259 mixed model (LMM) implemented in GCTA-MLMA (v1.94.1) (26), and 2) a linear regression
260 model in PLINK 2.0 (--glm) (27) using the top 20 genotype principal components as covariates
261 for population structure control. GCTA-MLMA used the same GRM as in the trait simulation.
262 Furthermore, required LD matrices were generated from each simulation sample's genotype data:
263 both the standard SNP correlation matrix ($\mathbf{\Sigma}$) as in model (3) and the relatedness-adjusted
264 correlation matrix ($\tilde{\mathbf{\Sigma}}$) as in model (9). The standard matrix ($\mathbf{\Sigma}$) was used by GCTA-COJO,
265 FINEMAP, SuSiE, FINEMAP-inf, and SuSiE-inf, while the relatedness-adjusted matrix ($\tilde{\mathbf{\Sigma}}$) was
266 used by FINEMAP-adj and SuSiE-adj.

267 **Performance evaluation in simulations**

268 The performance of various fine-mapping approaches was benchmarked using the datasets
269 generated across the 12 distinct simulation scenarios (3 causal variant counts × 2 PVE levels × 2
270 sample sizes). It is important to note that while 100 replicates were generated for each scenario,
271 not all replicates resulted in a significant association signal within the candidate region when
272 analyzed by GCTA-MLMA. Specifically, only those simulation replicates where at least one
273 variant in the candidate region achieved a GCTA-MLMA *P*-value $< 1 \times 10^{-5}$ were retained for
274 subsequent fine-mapping evaluation. This yielded a minimum of 74 such replicates for any given
275 scenario, with detailed counts available in Supplementary Table S1.

276 The methods evaluated included our novel approaches (BFMAP-SSS, FINEMAP-adj,
277 SuSiE-adj, and PIP_{gene}), our previously described BFMAP-Forward method (17), and several well-
278 established reference methods: FINEMAP v1.4.2 (10), SuSiE (as implemented in susieR v0.12.35)
279 (12, 13), FINEMAP-inf v1.3, SuSiE-inf v1.4 (14), and GCTA-COJO (--cojo-slct in GCTA
280 v1.94.1) (8, 28). BFMAP-Forward and BFMAP-SSS were implemented in BFMAP v0.65, while
281 FINEMAP-adj and SuSiE-adj directly used FINEMAP and SuSiE, respectively, with adapted
282 inputs designed for related individuals. For consistency across all applicable methods, a causal
283 variant's prior PVE and the maximum number of causal variants were set to 0.01 and 5,
284 respectively, in all simulation datasets. Furthermore, the same GRM, constructed from the 10,000
285 quality-controlled chip SNPs as detailed in the trait simulation, was used by all our LMM-based
286 fine-mapping methods requiring such an input, specifically BFMAP (both BFMAP-Forward and
287 BFMAP-SSS) and our -adj approaches (FINEMAP-adj and SuSiE-adj). For a comprehensive
288 comparison, reference methods were applied twice, using association statistics as input from LMM
289 (GCTA-MLMA) and, separately, from linear regression (PLINK 2.0 --glm).

290 We applied consistent *P*-value-based pre-filtering for fine-mapping across methods.
291 Specifically, for GCTA-COJO (--cojo-slct), the significance threshold in its stepwise procedure (-
292 -cojo-p) was set to 1×10^{-3}. Similarly, for all other fine-mapping tools evaluated (FINEMAP,
293 SuSiE, FINEMAP-inf, SuSiE-inf, and our BFMAP and -adj approaches), only variants with a
294 GWAS *P*-value $< 1 \times 10^{-3}$ from the respective input summary statistics (GCTA-MLMA or PLINK
295 2.0 --glm) were included in the fine-mapping analysis. This initial filtering threshold was chosen
296 to reduce the computational burden of model searching for each tool while aiming to be sufficiently
297 inclusive of potentially causal variants. Unless otherwise specified, all other parameters for these
298 software tools were maintained at their default values.

299 To evaluate fine-mapping accuracy, variants from each method were designated as
300 putatively causal based on their respective PIPs (or PCIPs for BFMAP-Forward) using thresholds
301 ranging from 0.0 to 1.0. Performance was primarily quantified using recall (also known as
302 sensitivity or power) and precision (also known as positive predictive value). Recall is the
303 proportion of true causal variants correctly identified (calculated as TP / (TP + FN)), and precision
304 is the proportion of designated putatively causal variants that are indeed true causal (calculated as
305 TP / (TP + FP)). Here, TP, FP, and FN represent the number of true positives, false positives, and
306 false negatives, respectively, determined at a given PIP (or PCIP) threshold and averaged over all
307 simulation replicates retained for fine-mapping evaluation per scenario (Supplementary Table S1).
308 Precision and recall were also calculated for gene-level PIPs using the same approach, with genes
309 rather than variants as the unit of evaluation.

310 Given that GCTA-COJO (--cojo-slct) does not compute PIPs, a modified evaluation
311 strategy was necessary. Variants identified by GCTA-COJO were ranked by first prioritizing those
312 from its jma output, followed by variants in its cma output, with the latter ranked by their

313 conditional *P*-values. This *P*-value-based ranking was then used to assess GCTA-COJO's ability
314 to prioritize true causal variants relative to the PIP/PCIP-based rankings from other methods.
315 Specifically, we evaluated its recall by selecting the top *k*-ranked variants (where *k* ranged from 1
316 to 20) as putatively causal. For each *k*, recall was calculated as the proportion of all true causal
317 variants that were included in this top *k* selection, averaged across all qualifying simulation
318 replicates per scenario.

319 **Duroc pig trait analysis**

320 This study also analyzed several economically important traits from the same cohort of Duroc pigs
321 described in the 'Data simulation' section. Phenotype data, collected by Smithfield Premium
322 Genetics between 2015 and 2021, included growth traits for ~27,000 animals (off-test body weight
323 [WT], back fat thickness [BF], and loin muscle depth [MS]) and reproduction traits for 3,290 sows
324 (number of piglets born alive [NBA], born dead [NBD], and weaned [NW]). The three growth
325 traits were collected at the end of the performance test, as detailed in Bergamaschi et al. (29). All
326 phenotypes were pre-adjusted to remove systematic environmental effects, including farm and
327 physiological sources of variation. For association and fine-mapping analyses, we utilized the
328 ~11.7 million quality-controlled autosomal variants (derived as detailed in the 'Data simulation'
329 section).

330 GWAS was performed using SLEMM-GWA v0.89.5 (30), which provides accuracy
331 comparable to GCTA-MLMA but with orders of magnitude greater computational efficiency
332 (Supplementary Figure S1). Association peaks were identified by visual inspection of the
333 Manhattan plot for each trait, with a *P* -value threshold of 1×10^{-5}. For each retained peak,
334 candidate region boundaries were established by extending 2 Mb upstream and downstream from
335 the position of the minimum *P*-value variant. Variants with $P < 1 \times 10^{-3}$ within each defined

336 candidate region were then retained for subsequent fine-mapping analyses. The same 10,000
337 quality-controlled chip SNPs selected for the trait simulation were used to construct the GRM for
338 methods requiring it (e.g., SLEMM and our BFMAP and -adj approaches).

339 Results

340 Method overview

341 This study introduces a framework for fine-mapping complex traits in samples of related
342 individuals, addressing analyses with both individual-level data and summary statistics. Figure 1
343 provides a schematic representation of this framework, illustrating the relationships between
344 existing methods and the novel approaches developed herein.

345 Typical fine-mapping tools, such as FINEMAP and SuSiE, usually use summary statistics.
346 Their underlying models are derived from linear regression assuming samples of unrelated
347 individuals, which limits their direct applicability or accuracy when substantial relatedness is
348 present (Figures 1A and 1C). To overcome this, we developed BFMAP, a Bayesian fine-mapping
349 approach first introduced for individual-level livestock data using forward selection (BFMAP-
350 Forward) (17). BFMAP employs a linear mixed model (LMM) that explicitly accounts for whole-
351 genome infinitesimal effects and genetic relatedness through a genomic relationship matrix (GRM)
352 (Figure 1B). In this study, alongside the established BFMAP-Forward, we introduce BFMAP-SSS,
353 a novel BFMAP implementation coupling shotgun stochastic search (SSS) with simulated
354 annealing for enhanced model space exploration.

355 Building upon the BFMAP framework, we developed a strategy to leverage its advantages
356 within a summary statistics context. The BFMAP model, initially designed for individual-level
357 data, can be transformed into an equivalent model based on summary statistics (Figure 1D).

358 Through an approximation ($\mathbf{D} \approx \mathbf{I}\tilde{n}$), analogous to the one used in GRAMMAR-Gamma, this
359 transformed model yields a functional form identical to that underlying standard FINEMAP and
360 SuSiE. This key structural equivalence enables the direct adaptation of these existing tools for
361 accurate fine-mapping in related individuals. This adaptation involves a two-step process: first,
362 calculating relatedness-aware inputs (including standard LMM-derived z-scores ($\mathbf{z}$), an effective
363 sample size ($\tilde{n}$), and a relatedness-adjusted linkage disequilibrium (LD) matrix ($\tilde{\mathbf{\Sigma}}$), as detailed in
364 Materials and methods); and second, providing these as the conventional z-score, sample size, and
365 LD matrix inputs to FINEMAP or SuSiE. We term these adapted methods FINEMAP-adj and
366 SuSiE-adj.

367 It is noteworthy that recently proposed fine-mapping methods, FINEMAP-inf and SuSiE-
368 inf, are conceptually related to our LMM-based approaches (BFMAP, FINEMAP-adj, and SuSiE-
369 adj) in that both incorporate terms to account for infinitesimal effects (Figure 1E). However, a key
370 distinction is that FINEMAP-inf and SuSiE-inf model the infinitesimal effects of variants within
371 the candidate fine-mapping region, whereas our methods model whole-genome infinitesimal
372 effects via GRM.

373 Finally, recognizing the challenges of fine-mapping to single-variant resolution in
374 populations with extensive LD, such as livestock, we propose the use of gene-level posterior
375 inclusion probability (PIP_{gene}). This metric aggregates variant-level evidence to assess the overall
376 evidence for a gene harboring at least one non-zero-effect variant, offering a balance between
377 localization precision and detection power (Figure 1F). In the subsequent sections, the
378 performance of our novel contributions (BFMAP-SSS, FINEMAP-adj, and SuSiE-adj), alongside
379 our previously developed BFMAP-Forward approach (17), is evaluated in comparison with

380 existing approaches (FINEMAP, SuSiE, FINEMAP-inf, SuSiE-inf, and GCTA-COJO) for fine-
381 mapping in related individuals.

382 **Simulation benchmarks**

383 To rigorously evaluate the fine-mapping approaches, we conducted simulation studies based on
384 real Duroc pig genotypes across 12 distinct scenarios. These scenarios varied by the number of
385 true causal variants (1, 2, or 3), the total proportion of variance explained (PVE) by these causal
386 variants (1% or 4%), and the sample size (5,000 or 10,000 individuals). For each scenario, 100
387 replicates were generated, though only those yielding significant GWAS associations (GCTA-
388 MLMA $P < 1 \times 10^{-5}$) were retained for fine-mapping evaluation, ensuring robust downstream
389 analysis (see Materials and methods for details).

390 *Impact of relatedness adjustment on genotype correlations*

391 Summary-statistics-based fine-mapping methods critically rely on an accurate LD matrix
392 reflecting correlations between genotypes. In samples of related individuals, one might naively
393 compute a standard LD matrix directly from the observed genotypes. However, as we have
394 theoretically established in Material and methods (Section 'Fine-mapping with related individuals:
395 summary statistics'), a relatedness-adjusted LD matrix ($\tilde{\mathbf{\Sigma}}$) is required to properly account for the
396 genetic relatedness among individuals in fine-mapping. To empirically illustrate the impact of this
397 adjustment, we compared elements of the standard LD matrix ($\mathbf{\Sigma}$) with those of the relatedness-
398 adjusted LD matrix ($\tilde{\mathbf{\Sigma}}$) within our 4-Mb candidate region for simulations (Supplementary Figure
399 S2). This comparison revealed substantial differences between the two types of correlations.
400 Notably, the relatedness adjustment (transforming $\mathbf{\Sigma}$ to $\tilde{\mathbf{\Sigma}}$) often led to a shrinkage of correlation
401 magnitudes, particularly for variants that exhibited strong positive or negative correlations in the
402 standard $\mathbf{\Sigma}$ matrix. Consequently, applying fine-mapping methods with a standard, unadjusted LD

403 matrix in samples of related individuals, rather than the appropriate relatedness-adjusted version,
404 can severely compromise fine-mapping performance and the accuracy of causal variant
405 identification.

406 *Performance of reference methods with different input association statistics*

407 We first assessed the performance of established reference fine-mapping methods (FINEMAP,
408 SuSiE, FINEMAP-inf, SuSiE-inf, and GCTA-COJO) using two types of input association statistics:
409 those derived from a LMM (via GCTA-MLMA) accounting for relatedness and whole-genome
410 infinitesimal effects, and those from a standard linear regression model (via PLINK 2.0 --glm)
411 with top 20 genotype principal components for population structure control.

412 GCTA-COJO consistently demonstrated slightly superior performance when using LMM-
413 derived association statistics across all simulation scenarios (Supplementary Figure S3). For the
414 other reference methods (FINEMAP, SuSiE, and their -inf variants), the choice of input statistics
415 had a more nuanced impact. In scenarios with 1 or 2 causal variants, these methods generally
416 favored linear regression inputs in 4% PVE settings, whereas LMM-derived statistics were often
417 preferred in 1% PVE settings, as reflected by precision-recall curves (Figure 2) and area under the
418 precision-recall curve (AUPRC) values (Table 1). This distinction was less apparent in scenarios
419 with 3 causal variants. Notably, SuSiE-inf frequently encountered convergence issues when using
420 LMM-derived association statistics but performed robustly with linear regression inputs.

421 *Comparative performance of proposed and existing fine-mapping methods*

422 Overall, GCTA-COJO's performance in prioritizing causal variants was generally comparable to
423 or slightly poorer than established PIP-based methods across the simulation scenarios

424 (Supplementary Figure S3). Consequently, while its results are available, our primary comparative
425 focus shifts to the PIP/PCIP-based methods.

426 A consistent pattern emerged when comparing our developed methods (BFMAP-Forward,
427 BFMAP-SSS, FINEMAP-adj, and SuSiE-adj) against the established approaches (FINEMAP,
428 SuSiE, FINEMAP-inf, SuSiE-inf). Our methods designed for related individuals consistently and
429 substantially outperformed these existing tools across all 12 simulation scenarios, as illustrated by
430 the precision-recall curves (Figure 2) and the AUPRC values (Table 1). This performance
431 advantage for our methods was particularly pronounced in scenarios with larger sample sizes or
432 higher PVE. For instance, when AUPRCs were averaged across the six scenarios (3 causal variant
433 counts × 2 PVE levels) with 10,000 individuals, BFMAP-SSS (average AUPRC = 0.234) showed
434 approximately a 3.8-fold higher AUPRC than the best-performing reference method, SuSiE using
435 LMM-derived association statistics (average AUPRC = 0.049) (Table 1). Specifically, FINEMAP-
436 adj and SuSiE-adj, which incorporate relatedness-adjusted LD matrices and LMM-derived
437 association statistics, achieved remarkable improvements over their counterparts (FINEMAP and
438 SuSiE) that use standard LD matrices. When AUPRCs were averaged across the six scenarios (3
439 causal variant counts × 2 PVE levels) with 10,000 individuals, FINEMAP-adj yielded a 5.4-fold
440 higher AUPRC than standard FINEMAP, and SuSiE-adj a 3.1-fold higher AUPRC than standard
441 SuSiE. These substantial gains were achieved when standard FINEMAP and SuSiE were provided
442 with LMM-derived association statistics, their generally optimal input type under these conditions
443 with 10,000 individuals (Table 1).

444 Among our LMM-based approaches, the overall differences in fine-mapping performance
445 were relatively small (Table 1, Figure 2). BFMAP-SSS, employing shotgun stochastic search,
446 performed best overall. BFMAP-Forward, which uses a forward selection strategy, showed

447 performance comparable to or slightly less favorable than FINEMAP-adj and SuSiE-adj that
448 implement more exhaustive model space exploration.

449 *Fine-mapping challenges in livestock populations and impact of simulation parameters*

450 It is crucial to note the inherent difficulty of fine-mapping in livestock populations, characterized
451 by strong LD and extensive genetic relatedness among individuals. Even with methods explicitly
452 accounting for relatedness and whole-genome infinitesimal effects, performance can be
453 constrained. For instance, in the simulation scenario with 3 causal variants, 10,000 individuals,
454 and 1% PVE, the best-performing reference method (SuSiE, using its optimal input) yielded an
455 AUPRC of only 0.024 (Table 1). While our novel methods offered a substantial improvement (e.g.,
456 BFMAP-SSS achieved an AUPRC of 0.077 in the same scenario), these results underscore that
457 fine-mapping in such populations remains a formidable task.

458 The simulation parameters also highlighted expected trends. Performance for most
459 methods improved with fewer causal variants or larger PVE for the candidate region. A key
460 observation was the differential response to increased sample size: while our relatedness-aware
461 approaches (BFMAP-Forward, BFMAP-SSS, FINEMAP-adj, and SuSiE-adj) uniformly benefited
462 from larger sample sizes, the methods originally designed for unrelated individuals (FINEMAP,
463 SuSiE, and their -inf variants) showed limited improvement or, in some instances, even a decline
464 in performance as sample size increased when applied to these related samples (Table 1). This
465 underscores the critical importance of appropriately modeling genetic relatedness, especially in
466 larger cohorts of related individuals.

467 **Gene-level PIPs increase detection power in fine-mapping**

468 To address the challenges of fine-mapping at single-variant resolution, particularly in populations
469 with extensive LD like the Duroc pigs in this study, we evaluated the utility of our proposed gene-
470 level posterior inclusion probability (PIP_{gene}). The calculation of PIP_{gene} requires posterior model
471 probabilities for variant sets, which are not standard outputs for all fine-mapping tools. Therefore,
472 for this specific evaluation, we computed PIP_{gene} using the outputs from BFMAP-Forward,
473 BFMAP-SSS, and FINEMAP-adj, as these methods provide the necessary model-level
474 information. This selection of methods is sufficient to demonstrate the advantages of gene-level
475 aggregation over variant-level PIPs, an orthogonal comparison to the inter-method performance
476 benchmarks detailed previously.

477 The application of gene-level fine-mapping resulted in a substantial improvement in the
478 ability to correctly identify causal genes (i.e., our defined gene proxy blocks containing one or
479 more true causal variants) for all three methods evaluated (BFMAP-Forward, BFMAP-SSS, and
480 FINEMAP-adj). This enhancement is evident from the precision-recall curves, which consistently
481 show superior performance for gene-level PIPs compared to variant-level PIPs across all 12
482 simulation scenarios (Figure 3). The AUPRC values further underscore the benefits of gene-level
483 aggregation (Table 2). The AUPRC for identifying causal genes was markedly higher than that for
484 identifying individual causal variants across all methods and scenarios. Notably, the magnitude of
485 this performance gain was often greatest in the more challenging simulation scenarios. For instance,
486 in the scenario with 3 causal variants, 1% PVE, and 10,000 individuals (a setting where variant-
487 level fine-mapping proved particularly difficult), gene-level aggregation yielded dramatic
488 improvements. At the variant level, BFMAP-Forward, BFMAP-SSS, and FINEMAP-adj achieved
489 AUPRCs of 0.061, 0.077, and 0.076, respectively. In contrast, when evidence was aggregated to
490 the gene level using PIP_{gene}, these same methods achieved AUPRCs of 0.305, 0.420, and 0.406,

491 respectively (Table 2). This represents an approximate 4- to 5.5-fold increase in AUPRC,
492 highlighting the substantial power gained by shifting the inferential focus from individual variants
493 to gene units in complex fine-mapping contexts.

494 **Computational performance**

495 We evaluated the runtime of our LMM-based fine-mapping methods using simulated datasets with
496 2 causal variants (4% PVE) for 5,000 and 10,000 individuals, focusing on the same 4-Mb candidate
497 region (3,837 and 4,371 pre-filtered SNPs, respectively) as used in our main simulation
498 benchmarks. Analyses were conducted on an Intel Xeon Gold 6230 CPU. Key fine-mapping
499 parameters, such as the prior PVE per causal variant (0.01) and the maximum number of assumed
500 causal variants (5), were kept consistent with the main simulation benchmarks.

501 Wall-clock times for all methods, representing the analysis of a single simulation replicate,
502 are detailed in Supplementary Table S2. BFMAP-Forward was the fastest approach, completing
503 analyses in approximately 0.3 minutes (5,000 individuals) and 0.5 minutes (10,000 individuals)
504 with 20 CPU threads. FINEMAP-adj required approximately 1.5 minutes (5,000 individuals) and
505 2.3 minutes (10,000 individuals) using 20 threads; these times include the pre-calculation of the
506 relatedness-adjusted LD matrix. Both BFMAP-Forward and FINEMAP-adj showed limited
507 additional speedup when using 20 threads compared to single-thread execution. BFMAP-SSS,
508 while generally the most accurate in fine-mapping benchmarks, was the most computationally
509 intensive, with wall-clock times of approximately 52 minutes (5,000 individuals) and 170 minutes
510 (10,000 individuals) using 20 threads. SuSiE-adj, which utilizes the susieR package and operates
511 on a single thread, completed its analyses in approximately 1.3 minutes (5,000 individuals) and
512 2.3 minutes (10,000 individuals), comparable to single-threaded FINEMAP-adj.

513 **Application to Duroc pig traits**

514 We applied our novel fine-mapping methods to analyze six economically important traits in the
515 Duroc pig population previously described. To identify candidate genomic regions for fine-
516 mapping, we first performed GWAS for each trait using the mixed-model approach implemented
517 in SLEMM-GWA. All six traits produced high-quality Manhattan plots, five of which displayed
518 association peaks with an inclusive significance threshold of $P < 1 \times 10^{-5}$ (Supplementary Figure
519 S4). Using this threshold, we identified 46 candidate region-trait pairs: 43 for performance traits
520 and 3 for reproduction traits (Supplementary Table S3).

521 To identify putative causal variants and genes within these candidate regions, we initially
522 employed BFMAP-SSS due to its strong performance in simulations. This fine-mapping identified
523 30 variants with a PIP > 0.1 and 2 variants with a PIP > 0.5 across all region-trait pairs
524 (Supplementary Table S4). Recognizing the potential to enhance discovery by aggregating variant-
525 level evidence, we then calculated the gene-level PIP (PIP_{gene}) for all annotated genes within each
526 candidate region, with gene boundaries defined as detailed in Material and methods. This approach
527 identified 87 gene-trait pairs with $PIP_{gene} > 0.1$, of which 22 pairs had a $PIP_{gene} > 0.5$
528 (Supplementary Table S5). These results underscore the utility of this gene-based strategy for
529 prioritizing candidate genes.

530 The power of PIP_{gene} was particularly evident in regions where individual variant PIPs were
531 low. For example, a candidate region on chromosome 1 (chr1:51,300,453-55,300,453) associated
532 with back fat thickness (BF) (Figure 4A) contained no variants with a PIP greater than 0.03 (Figure
533 4B), making variant prioritization challenging. However, the PIP_{gene} analysis revealed that *MRAP2*
534 emerged with a strong signal ($PIP_{gene} = 0.669$, distinctly higher than other genes in the region;
535 Figure 4C). The *MRAP2* gene encodes melanocortin receptor accessory protein 2, which regulates

536 melanocortin receptor activity and energy homeostasis (31). Previous studies have identified loss-
537 of-function *MRAP2* variants as pathogenic for monogenic hyperphagic obesity in humans (32, 33),
538 and *Mrap2*-knockout mice develop severe early-onset obesity with increased fat mass and
539 adipocyte hypertrophy (33). These findings strongly support *MRAP2* as a biologically plausible
540 candidate for the BF association observed in our study.

541 A second illustrative example for BF is the candidate region on chromosome 6
542 (chr6:144,948,358-148,948,358; Figure 4D). In this locus, fine-mapping with BFMAP-SSS
543 identified three prioritized variants (each with PIP > 0.1) that are located within the *LEPR* gene.
544 This gene encodes the leptin receptor, which mediates leptin signaling in energy homeostasis and
545 body weight regulation (34). The *LEPR* gene showed a strong gene-level signal (PIP_{gene} = 0.861),
546 consistent with its well-established role in regulating energy homeostasis and body composition
547 (35–37). Mutations in this gene can cause severe early-onset obesity, hyperphagia, and
548 hypogonadotropic hypogonadism in humans (38, 39). While PIPs of all other variants in the
549 candidate region were < 0.05 (Figure 4E), our PIP_{gene} analysis highlighted an additional candidate
550 gene, *DNAJC6* (PIP_{gene} = 0.875; Figure 4F). This gene encodes auxilin-1 , a DNAJ/HSP40 family
551 protein that regulates resting metabolic rate (40) and has been shown to suppress adipogenesis and
552 energy metabolism in adipocytes (41). These findings demonstrate the ability of PIP_{gene} to enhance
553 detection power for relevant genes whose signals might be diffuse at the individual variant level.

554 Finally, we compared our summary-statistics-based methods (FINEMAP-adj and SuSiE-
555 adj) against individual-level BFMAP-SSS by re-analyzing the two illustrative BF candidate
556 regions. At the variant level, SuSiE-adj PIPs showed high concordance with BFMAP-SSS PIPs in
557 both regions (Supplementary Figures S5A and S5B). For FINEMAP-adj, variant PIPs on the
558 chromosome 1 region were highly concordant with BFMAP-SSS (Supplementary Figure S5A)

559 and stable across iteration lengths (Supplementary Figures S6A and S6C), including the default
560 iteration settings. The chromosome 6 region, however, proved more challenging for FINEMAP-
561 adj: while its concordance with BFMAP-SSS improved with increased iterations from the default
562 settings (plateauing around 10,000 iterations) (Supplementary Figures S6B and S6D), FINEMAP-
563 adj still showed substantial discrepancy from BFMAP-SSS, with a PIP correlation between the
564 two methods of 0.594 even with 50,000 iterations (Supplementary Figure S5B). Despite the
565 variant-level discrepancy for FINEMAP-adj on the chromosome 6 region, PIP_{gene} values derived
566 from the FINEMAP-adj runs with 50,000 iterations were highly correlated with those from
567 BFMAP-SSS in both regions (chr1: 0.999; chr6: 0.945; Supplementary Figure S5C and S5D).
568 Gene-level aggregation yielded more highly consistent inferences between FINEMAP-adj and
569 BFMAP-SSS, even when variant-level concordance is limited (Supplementary Figure S6).

570 # Discussion

571 Fine-mapping complex traits in populations with extensive relatedness, such as livestock, presents
572 significant statistical challenges. Standard fine-mapping methodologies, predominantly developed
573 for human genetics studies involving largely unrelated individuals, can falter or yield unreliable
574 results when applied naively to these populations. This study addresses this critical gap by
575 introducing a comprehensive framework. This includes BFMAP (a Bayesian approach for
576 individual-level data, encompassing our previously developed BFMAP-Forward (17) and the
577 novel BFMAP-SSS implementation), innovative summary-statistics adaptations (FINEMAP-adj
578 and SuSiE-adj), and a gene-level aggregation metric (PIP_{gene}), all designed to improve fine-
579 mapping under such complex conditions of relatedness.

A key contribution of this work is the clear demonstration, both theoretically and through extensive pig genotype-based simulations, that appropriately modeling relatedness is critical for achieving accurate fine-mapping. Our results unequivocally show that standard fine-mapping tools (GCTA-COJO, FINEMAP, SuSiE, and their -inf variants), even when provided with summary association statistics from mixed models, exhibit substantially poorer performance than our relatedness-aware methods (BFMAP-Forward, BFMAP-SSS, FINEMAP-adj, and SuSiE-adj) designed to correctly utilize relatedness information (Figure 2, Table 1). The performance gap, reflected in markedly poorer precision-recall profiles and lower AUPRC values for existing methods (GCTA-COJO, FINEMAP, SuSiE, and their -inf variants) (Figure 2, Table 1), underscores the potential for a high proportion of false positives in previous fine-mapping studies in livestock that applied human-centric fine-mapping methods not designed for samples of related individuals (e.g., 42–46). It is important to note that simply §using LMM-derived summary association statistics was insufficient to overcome the fundamental limitations of these methods when applied to samples of related individuals. This highlights the critical need for approaches that explicitly integrate relatedness, primarily through a relatedness-adjusted LD matrix, as implemented in our -adj methods, or via direct LMM modeling as in BFMAP.

Among our proposed relatedness-aware methods, BFMAP-SSS generally emerged as the most accurate, likely due to its sophisticated shotgun stochastic search with simulated annealing allowing for a more thorough exploration of the model space. FINEMAP-adj and SuSiE-adj also performed well, trailing BFMAP-SSS and slightly outperforming BFMAP-Forward (Table 1, Figure 2). This performance hierarchy aligns with theoretical expectations: the forward selection strategy of BFMAP-Forward is inherently less exhaustive than the search algorithms in BFMAP-SSS or the variable selection approaches underpinning SuSiE (and thus SuSiE-adj). The slight

603 performance edge of BFMAP-SSS over the -adj methods may stem from the GRAMMAR-
604 Gamma-like approximation employed to allow the FINEMAP and SuSiE programs to be directly
605 applied using appropriately transformed summary statistics (LMM-derived z-scores and effective
606 sample size $\tilde{n}$) and a relatedness-adjusted LD matrix ($\tilde{\mathbf{\Sigma}}$). While developing new programs for
607 FINEMAP-adj and SuSiE-adj from scratch to circumvent this approximation is feasible, the
608 current -adj approaches offer substantial practical advantages by leveraging well-established,
609 existing software, especially given their performance close to BFMAP-SSS.

610 The choice among our developed methods in practice will depend on a balance of factors.
611 BFMAP-SSS offers the highest accuracy but is the most computationally intensive. For studies
612 where computational resources are a constraint, FINEMAP-adj, SuSiE-adj, and BFMAP-Forward
613 provide much faster alternatives with only a modest trade-off in accuracy compared to BFMAP-
614 SSS (Figure 2, Supplementary Table S2). FINEMAP-adj, BFMAP-SSS, and BFMAP-Forward
615 also readily provide the posterior model probabilities necessary for calculating PIP_{gene}, a feature
616 not directly available from SuSiE-adj (which builds on SuSiE) without further methodological
617 development. Furthermore, users should be aware that FINEMAP-adj (which builds on FINEMAP)
618 can sometimes require extended iterations for convergence in complex regions, whereas SuSiE-
619 adj, benefiting from SuSiE's efficient variable selection, is generally robust in convergence. We
620 therefore recommend BFMAP-SSS when maximal accuracy is paramount and resources permit;
621 FINEMAP-adj when gene-level insights are desired with good speed; SuSiE-adj for fast and robust
622 variant-level fine-mapping without immediate need for PIP_{gene}; and BFMAP-Forward as a rapid
623 option for individual-level data.

624 A significant challenge in livestock genomics is the extensive and complex LD structure,
625 usually making fine-mapping at single-variant resolution exceedingly difficult. Our proposed

626 PIP_{gene} metric addresses this by aggregating variant-level evidence, thereby enhancing the power
627 to detect trait-associated genes, albeit with a trade-off in localization precision from variant to gene.
628 As demonstrated in our simulations, the improvement in detection power (AUPRC) using PIP_{gene}
629 was substantial, particularly in challenging scenarios with low PVE and multiple causal variants,
630 where gene-level AUPRCs were up to 5.5-fold higher than variant-level AUPRCs (Table 2, Figure
631 3). This utility was further underscored in our analysis of Duroc pig traits, where PIP_{gene}
632 successfully highlighted biologically relevant genes such as *MRAP2* and *DNAJC6* for back fat
633 thickness, even when underlying variant PIPs were diffuse or individually weak (Figure 4).
634 Furthermore, this approach led to the identification of other promising candidate genes, such as
635 *DYRK4* as a novel candidate for back fat thickness, whose high PIP_{gene} values warrant further
636 investigation (47). These findings collectively highlight the value of our fine-mapping framework
637 for enhancing candidate gene discovery in livestock.

638 Despite the advancements presented, this study has limitations. A primary limitation is our
639 simulation design. While based on real pig genotypes, evaluations focused on a single 4-Mb
640 candidate region where we simulated only 1-3 variants as having larger causal effects. This,
641 combined with an infinitesimal background modeled by 10,000 genome-wide chip SNPs,
642 represents one specific model of genetic architecture; real traits exhibit greater heterogeneity in
643 the number, effect size distribution, and genomic location of causal variants. Furthermore, the
644 GRM used to model this infinitesimal background was constructed directly from these 10,000
645 known infinitesimal-effect SNPs and was then supplied as input to GRM-dependent fine-mapping
646 methods. This perfect concordance between the SNPs defining the simulated infinitesimal
647 background and those in the analytical GRM is an idealization not reflective of real-world
648 scenarios, where GRMs are estimated from available marker data (often LD-pruned) and may not

649 fully capture true genetic covariance or perfectly align with the underlying polygenic architecture.
650 While our main simulations effectively highlighted performance differences using this consistent
651 GRM, we further investigated the impact of using a more "practical" GRM, constructed from
652 10,000 randomly selected quality-controlled genome-wide variants (from the ~11.7 million). Re-
653 analyzing all simulation scenarios with FINEMAP-adj and SuSiE-adj using this practical GRM
654 (for both LMM-GWAS summary statistic generation and as input to the -adj methods) yielded
655 fine-mapping performance highly concordant with that obtained using the "simulation-derived"
656 GRM (Supplementary Figure S7, Supplementary Table S6). This result suggests that while GRM
657 misspecification remains a theoretical concern, our -adj approaches demonstrate robustness to this
658 aspect of GRM construction, at least with reasonably dense marker-based GRMs. While our study
659 effectively demonstrates the advantages of relatedness-aware methods under the tested
660 architectures, further research exploring a wider array of simulated genetic complexities will be
661 valuable for comprehensively understanding their performance boundaries across all potential
662 real-world scenarios. Secondly, the current implementations, particularly those involving matrix
663 operations on the GRM or relatedness-adjusted LD, may face scalability challenges with datasets
664 involving hundreds of thousands of individuals, a scale becoming increasingly common. Further
665 algorithmic refinements will be needed to cater to such massive datasets. Thirdly, our current
666 PIP_{gene} formulation, which sums posterior model probabilities for models including any variant
667 within a gene region, can be systematically influenced by gene length and variant density. Under
668 the standard fine-mapping prior that assigns equal *a priori* causal probability to each variant,
669 longer genes with more variants inherently have greater opportunities to accumulate signal, leading
670 to higher PIP_{gene} values that may not solely reflect biological enrichment for causality. Future
671 refinements could explore incorporating more sophisticated priors that account for gene size or

variant density to further enhance the specificity of gene-level inference. Fourthly, our real data analysis relied on imputed sequence data for SNPs and indels; other structural variations were not considered, and differential imputation quality across variants could introduce unmodeled uncertainty.

Future research could focus on several exciting avenues. Incorporating functional annotation data into the prior model probabilities for variants or genes (17, 48–50) holds considerable promise for further enhancing the statistical power and accuracy of our methods. Additionally, while this study focused on fine-mapping GWAS signals for complex traits, the principles of accounting for relatedness (particularly through adjusted LD matrices and mixed models) are broadly applicable. Investigating the performance and adaptation of our framework for other types of genetic analyses in samples of related individuals, such as fine-mapping expression quantitative trait loci (eQTLs) (51–54), represents a valuable direction.

In conclusion, this study provides a robust suite of fine-mapping tools (BFMAP-Forward, BFMAP-SSS, FINEMAP-adj, and SuSiE-adj) and a novel gene-level aggregation strategy (PIP_{gene}) specifically designed for and validated in populations with complex relatedness, such as livestock. By properly accounting for genetic relationships and whole-genome infinitesimal effects, these methods offer substantial improvements in accuracy and power over existing approaches, paving the way for more reliable identification of causal variants and genes underlying complex traits in these important populations.

Acknowledgments

We appreciate Smithfield Premium Genetics (Rose Hill, NC, USA) for providing access to the data used in this study.

Author contributions

J.J. conceived and designed the study and developed the methodologies. J.Wang carried out the simulation studies and pig trait analyses. F.T., Y.H., and C.M. contributed to data preprocessing. J.Wang, J.J., and J.Wei performed data visualization. J.J. and J.Wang wrote the manuscript. All authors reviewed and approved the final manuscript.

Conflict of interest

Y.H. was employed by Smithfield Premium Genetics. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work is supported by the Agriculture and Food Research Initiative (AFRI) Foundational and Applied Science Program, project award no. 2023-67015-39260, and the Research Capacity Fund (HATCH), project award no. 7008128, from the U.S. Department of Agriculture's National Institute of Food and Agriculture.

Data availability

BFMAP can be downloaded at https://github.com/jiang18/bfmap. Tutorials for FINEMAP-adj and SuSiE-adj are available at https://github.com/JJWang259/FineMapping-RelatedIndividuals. The

713 Duroc pig data used in this study are property of Smithfield Premium Genetics (Rose Hill, NC,
714 USA). Restrictions apply to the availability of these data, which were used under license for the
715 current study and are not publicly available. Requests to access these datasets may be granted
716 under a negotiated research agreement and should be directed to Kent Gray, General Manager
717 (kgray@smithfield.com).

718 # References

719 1. Uffelmann,E., Huang,Q.Q., Munung,N.S., de Vries,J., Okada,Y., Martin,A.R., Martin,H.C.,
720 Lappalainen,T. and Posthuma,D. (2021) Genome-wide association studies. *Nat Rev Methods*
721 *Primers*, **1**, 59.

722 2. Stram,D.O. (2004) Tag SNP selection for association studies. *Genetic Epidemiology*, **27**, 365–374.

723 3. Boyle,E.A., Li,Y.I. and Pritchard,J.K. (2017) An expanded view of complex traits: from polygenic to
724 omnigenic. *Cell*, **169**, 1177–1186.

725 4. Lappalainen,T., Li,Y.I., Ramachandran,S. and Gusev,A. (2024) Genetic and molecular architecture of
726 complex traits. *Cell*, **187**, 1059–1075.

727 5. Schaid,D.J., Chen,W. and Larson,N.B. (2018) From genome-wide associations to candidate causal
728 variants by statistical fine-mapping. *Nat Rev Genet*, **19**, 491–504.

729 6. Spain,S.L. and Barrett,J.C. (2015) Strategies for fine-mapping complex traits. *Human Molecular*
730 *Genetics*, **24**, R111–R119.

731 7. Hormozdiari,F., Kostem,E., Kang,E.Y., Pasaniuc,B. and Eskin,E. (2014) Identifying Causal Variants at
732 Loci with Multiple Signals of Association. *Genetics*, **198**, 497–508.

733 8. Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication
734 And Meta-analysis (DIAGRAM) Consortium, Yang,J., Ferreira,T., Morris,A.P., Medland,S.E.,
735 Madden,P.A.F., Heath,A.C., Martin,N.G., Montgomery,G.W., *et al.* (2012) Conditional and joint
736 multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing
737 complex traits. *Nat Genet*, **44**, 369–375.

738 9. Chen,W., Larrabee,B.R., Ovsyannikova,I.G., Kennedy,R.B., Haralambieva,I.H., Poland,G.A. and
739 Schaid,D.J. (2015) Fine Mapping Causal Variants with an Approximate Bayesian Method Using
740 Marginal Test Statistics. *Genetics*, **200**, 719–736.

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

11. Hans,C., Dobra,A. and West,M. (2007) Shotgun Stochastic Search for “Large p” Regression. *Journal of the American Statistical Association*, **102**, 507–516.

12. Wang,G., Sarkar,A., Carbonetto,P. and Stephens,M. (2020) A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **82**, 1273–1300.

13. Zou,Y., Carbonetto,P., Wang,G. and Stephens,M. (2022) Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet*, **18**, e1010299.

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

15. Sul,J.H., Martin,L.S. and Eskin,E. (2018) Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genet*, **14**, e1007309.

16. Yu,J., Pressoir,G., Briggs,W.H., Vroh Bi,I., Yamasaki,M., Doebley,J.F., McMullen,M.D., Gaut,B.S., Nielsen,D.M., Holland,J.B., *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*, **38**, 203–208.

17. Jiang,J., Cole,J.B., Freebern,E., Da,Y., VanRaden,P.M. and Ma,L. (2019) Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun Biol*, **2**, 212.

18. Auton,A., Abecasis,G.R., Altshuler,D.M., Durbin,R.M., Abecasis,G.R., Bentley,D.R., Chakravarti,A., Clark,A.G., Donnelly,P., Eichler,E.E., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.

19. VanRaden,P.M. (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, **91**, 4414–4423.

20. Kirkpatrick,S., Gelatt,C.D. and Vecchi,M.P. (1983) Optimization by simulated annealing. *Science*, **220**, 671–680.

21. Svishcheva,G.R., Axenovich,T.I., Belonogova,N.M., van Duijn,C.M. and Aulchenko,Y.S. (2012) Rapid variance components-based method for whole-genome association analysis. *Nat Genet*, **44**, 1166–1170.

22. Stephens,M. and Balding,D.J. (2009) Bayesian statistical methods for genetic association studies. *Nat Rev Genet*, **10**, 681–690.

23. Ding,R., Savegnago,R., Liu,J., Long,N., Tan,C., Cai,G., Zhuang,Z., Wu,J., Yang,M., Qiu,Y., *et al.* (2023) The SWine IMputation (SWIM) haplotype reference panel enables nucleotide resolution genetic mapping in pigs. *Commun Biol*, **6**, 1–10.

24. Jiang,J. (2024) MPH: fast REML for large-scale genome partitioning of quantitative genetic variation. *Bioinformatics*, **40**, btae298.

779 25. Warr,A., Affara,N., Aken,B., Beiki,H., Bickhart,D.M., Billis,K., Chow,W., Eory,L., Finlayson,H.A.,
780 Flicek,P., *et al.* (2020) An improved pig reference genome sequence to enable pig genetics and
781 genomics research. *GigaScience*, **9**, giaa051.

782 26. Yang,J., Zaitlen,N.A., Goddard,M.E., Visscher,P.M. and Price,A.L. (2014) Advantages and pitfalls in the
783 application of mixed-model association methods. *Nat Genet*, **46**, 100–106.

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

786 28. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: A Tool for Genome-wide Complex
787 Trait Analysis. *The American Journal of Human Genetics*, **88**, 76–82.

788 29. Bergamaschi,M., Tiezzi,F., Howard,J., Huang,Y.J., Gray,K.A., Schillebeeckx,C., McNulty,N.P. and
789 Maltecca,C. (2020) Gut microbiome composition differences among breeds impact feed
790 efficiency in swine. *Microbiome*, **8**, 110.

791 30. Cheng,J., Maltecca,C., VanRaden,P.M., O'Connell,J.R., Ma,L. and Jiang,J. (2023) SLEMM: million-scale
792 genomic predictions with window-based SNP weighting. *Bioinformatics*, **39**, btad127.

793 31. Bernard,A., Naharros,I.O., Yue,X., Mifsud,F., Blake,A., Bourgain-Guglielmetti,F., Ciprin,J., Zhang,S.,
794 McDaid,E., Kim,K., *et al.* (2023) MRAP2 regulates energy homeostasis by promoting primary cilia
795 localization of MC4R. *JCI Insight*, **8**.

796 32. Baron,M., Maillet,J., Huyvaert,M., Dechaume,A., Boutry,R., Loiselle,H., Durand,E., Toussaint,B.,
797 Vaillant,E., Philippe,J., *et al.* (2019) Loss-of-function mutations in MRAP2 are pathogenic in
798 hyperphagic obesity with hyperglycemia and hypertension. *Nat Med*, **25**, 1733–1738.

799 33. Asai,M., Ramachandrappa,S., Joachim,M., Shen,Y., Zhang,R., Nuthalapati,N., Ramanathan,V.,
800 Strochlic,D.E., Ferket,P., Linhart,K., *et al.* (2013) Loss of Function of the Melanocortin 2 Receptor
801 Accessory Protein 2 Is Associated with Mammalian Obesity. *Science*, **341**, 275–278.

802 34. Villanueva,E.C. and Myers,M. (2008) Leptin receptor signaling and the regulation of mammalian
803 physiology. *International journal of obesity*, **32**, S8–S12.

804 35. Zhang,Y., Proenca,R., Maffei,M., Barone,M., Leopold,L. and Friedman,J.M. (1994) Positional cloning
805 of the mouse obese gene and its human homologue. *Nature*, **372**, 425–432.

806 36. Allison,M.B. and Myers Jr,M.G. (2014) Connecting leptin signaling to biological function. *The Journal*
807 *of endocrinology*, **223**, T25.

808 37. Friedman,J.M. (2019) Leptin and the endocrine control of energy balance. *Nature metabolism*, **1**,
809 754–764.

810 38. Chaves,C., Kay,T. and Anselmo,J. (2022) Early onset obesity due to a mutation in the human leptin
811 receptor gene. *Endocrinol Diabetes Metab Case Rep*, **2022**, 21–0124.

812 39. Clément,K., Vaisse,C., Lahlou,N., Cabrol,S., Pelloux,V., Cassuto,D., Gourmelen,M., Dina,C.,
813 Chambaz,J., Lacorte,J.-M., *et al.* (1998) A mutation in the human leptin receptor gene causes
814 obesity and pituitary dysfunction. *Nature*, **392**, 398–401.

815 40. Son,D. and Lee,M. (2023) Gene regulation of RMR-related *DNAJC6* on adipogenesis and
816 mitochondria function in 3T3-L1 preadipocytes. *Biochemical and Biophysical Research*
817 *Communications*, **672**, 1–9.

818 41. Kim,J. and Lee,M. (2022) RMR-related DNAJC6 expression suppresses adipogenesis in 3T3-L1 cells.
819 *Cells*, **11**, 1331.

820 42. Li,Y., Li,B., Yang,M., Han,H., Chen,T., Wei,Q., Miao,Z., Yin,L., Wang,R., Shen,J., *et al.* (2020) Genome-
821 Wide Association Study and Fine Mapping Reveals Candidate Genes for Birth Weight of
822 Yorkshire and Landrace Pigs. *Front. Genet.*, **11**.

823 43. Liu,K., Hou,L., Yin,Y., Wang,B., Liu,C., Zhou,W., Niu,P., Li,Q., Huang,R. and Li,P. (2023) Genome-wide
824 association study reveals new QTL and functional candidate genes for the number of ribs and
825 carcass length in pigs. *Animal Genetics*, **54**, 435–445.

826 44. Wu,Z., Gong,H., Zhou,Z., Jiang,T., Lin,Z., Li,J., Xiao,S., Yang,B. and Huang,L. (2022) Mapping short
827 tandem repeats for liver gene expression traits helps prioritize potential causal variants for
828 complex traits in pigs. *J Animal Sci Biotechnol*, **13**, 8.

829 45. Qiu,Z., Cai,W., Liu,Q., Liu,K., Liu,C., Yang,H., Huang,R., Li,P. and Zhao,Q. (2024) Unravelling novel and
830 pleiotropic genes for cannon bone circumference and bone mineral density in Yorkshire pigs.
831 *Journal of Animal Science*, **102**, skae036.

832 46. Gualdrón Duarte,J.L., Yuan,C., Gori,A.-S., Moreira,G.C.M., Takeda,H., Coppieters,W., Charlier,C.,
833 Georges,M. and Druet,T. (2023) Sequenced-based GWAS for linear classification traits in Belgian
834 Blue beef cattle reveals new coding variants in genes regulating body size in mammals. *Genet Sel*
835 *Evol*, **55**, 83.

836 47. Papadopoulos,C., Arato,K., Lilienthal,E., Zerweck,J., Schutkowski,M., Chatain,N., Müller-Newen,G.,
837 Becker,W. and de la Luna,S. (2011) Splice Variants of the Dual Specificity Tyrosine
838 Phosphorylation-regulated Kinase 4 (DYRK4) Differ in Their Subcellular Localization and Catalytic
839 Activity. *J Biol Chem*, **286**, 5494–5505.

840 48. Yang,Z., Wang,C., Liu,L., Khan,A., Lee,A., Vardarajan,B., Mayeux,R., Kiryluk,K. and Ionita-Laza,I. (2023)
841 CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses.
842 *Nat Genet*, **55**, 1057–1065.

843 49. Weissbrod,O., Hormozdiari,F., Benner,C., Cui,R., Ulirsch,J., Gazal,S., Schoech,A.P., van de Geijn,B.,
844 Reshef,Y., Márquez-Luna,C., *et al.* (2020) Functionally informed fine-mapping and polygenic
845 localization of complex trait heritability. *Nat Genet*, **52**, 1355–1363.

846 50. Kichaev,G., Yang,W.-Y., Lindstrom,S., Hormozdiari,F., Eskin,E., Price,A.L., Kraft,P. and Pasaniuc,B.
847 (2014) Integrating Functional Data to Prioritize Causal Variants in Statistical Fine-Mapping
848 Studies. *PLOS Genetics*, **10**, e1004722.

849 51. Wang,Q.S., Kelley,D.R., Ulirsch,J., Kanai,M., Sadhuka,S., Cui,R., Albors,C., Cheng,N., Okada,Y.,
850 Aguet,F., *et al.* (2021) Leveraging supervised learning for functionally informed fine-mapping of
851 cis-eQTLs identifies an additional 20,913 putative causal eQTLs. *Nat Commun*, **12**, 3394.

852 52. D'Antonio,M., Nguyen,J.P., Arthur,T.D., Matsui,H., D'Antonio-Chronowska,A. and Frazer,K.A. (2023)
853 Fine mapping spatiotemporal mechanisms of genetic variants underlying cardiac traits and
854 disease. *Nat Commun*, **14**, 1132.

855 53. Liu,L., Chandrashekar,P., Zeng,B., Sanderford,M.D., Kumar,S. and Gibson,G. (2021) TreeMap: a
856 structured approach to fine mapping of eQTL variants. *Bioinformatics*, **37**, 1125–1134.

857 54. Zhang,X., Jiang,W. and Zhao,H. (2024) Integration of expression QTLs with fine mapping via SuSiE.
858 *PLOS Genetics*, **20**, e1010929.

859

860

861 **Figure legends**

862 **Figure 1. Overview of fine-mapping methods for unrelated and related individuals.**

863 Overview of fine-mapping approaches for samples of unrelated (A, C, and E) and related (B, D,
864 and F) individuals. (A) Linear regression provides the foundation for fine-mapping in samples of
865 unrelated individuals. (B) BFMAP employs linear mixed models with genomic relationship matrix
866 for fine-mapping in samples of related individuals. (C) Standard summary-statistics methods
867 (FINEMAP and SuSiE) are derived from the linear regression framework. (D) Adapted methods
868 (FINEMAP-adj and SuSiE-adj) use relatedness-adjusted correlations ($\tilde{\mathbf{\Sigma}}$) and effective sample size
869 ($\tilde{n}$) derived from the BFMAP framework. Through the approximation $\mathbf{D} \approx \mathbf{I}\tilde{n}$, these methods
870 achieve compatibility with standard FINEMAP and SuSiE implementations. (E) FINEMAP-inf
871 and SuSiE-inf extend their respective base methods by modeling infinitesimal effects within
872 candidate regions. (F) All methods produce variant-level posterior inclusion probability (PIP; or
873 PCIP for BFMAP-Forward) as output. Gene-level PIP, available for methods that provide posterior
874 model probabilities (BFMAP-Forward, BFMAP-SSS, FINEMAP-adj, and FINEMAP),
875 aggregates variant-level evidence for prioritizing causal genes.

876 **Figure 2. Precision-recall curves for identifying causal variants using PIPs across simulation**
877 **scenarios.**

878 GWAS summary data were derived from GCTA-MLMA (LMM) or PLINK 2.0 --glm (LM).
879 Results are shown for different sample sizes (5,000 and 10,000 individuals), PVE levels (1% and
880 4%), and numbers of causal variants (1-3), with A-L representing the 12 simulation scenarios.

881 **Figure 3. Precision-recall curves for identifying causal genes across simulation scenarios.**

882 Results show the ability of three methods (BFMAP-Forward, BFMAP-SSS, and FINEMAP-adj)

883 to identify causal genes (defined as gene proxy blocks containing true causal variants) under
884 different sample sizes (5,000 and 10,000 individuals), PVE levels (1% and 4%), and numbers of
885 causal variants (1-3). A-L represent the 12 simulation scenarios.

886 **Figure 4. Fine-mapping results for two candidate regions.**

887 BFMAP-SSS fine-mapping analysis of two genomic regions associated with back fat thickness:
888 chr1:51,300,453–55,300,453 and chr6:144,948,358–148,948,358. (A) and (D) show locus zoom
889 plots displaying local association signals within each interval. (B) and (E) show the PIPs of
890 individual variants, and (C) and (F) show the PIPs of genes.

891 **Figure S1. Comparison of GCTA-MLMA and SLEMM-GWA for GWAS of back fat**
892 **thickness in 26,731 Duroc pigs.**

893 Both methods used identical GRMs constructed from 10,000 randomly selected chip SNPs and
894 analyzed 1,000,000 randomly selected imputed variants. (A) Chi-square association test statistic
895 comparison. (B) Computational time comparison. Analyses were conducted on a Linux server with
896 2×Intel Xeon Gold 6258R CPUs and 512 GB memory using 20 threads.

897 **Figure S2. Comparison of standard and relatedness-adjusted genotype correlation**
898 **matrices.**

899 (A) Scatter plot comparing pairwise genotype correlation coefficients between the standard matrix
900 ($\mathbf{\Sigma}$) and relatedness-adjusted matrix ($\tilde{\mathbf{\Sigma}}$) for all SNP pairs. The red dashed line represents y = x. (B)
901 Density distributions of genotype correlation coefficients for both matrix types. Both matrices
902 were computed using 5,000 randomly selected individuals and 2,456 pre-filtered variants on the
903 chr1 region from one simulation replicate. The standard matrix ($\mathbf{\Sigma}$) represents conventional SNP
904 correlations, while the adjusted matrix ($\tilde{\mathbf{\Sigma}}$) accounts for genetic relatedness among individuals as

905 described in model (9) with an input heritability parameter of 0.494.

906 **Figure S3. Comparison of fine-mapping methods for identifying causal variants across**
907 **simulation scenarios.**

908 The y-axis shows the proportion of true causal variants correctly identified when selecting the top
909 *k*-ranked variants (x-axis, k = 1 to 20) across 12 simulation scenarios. Results are shown for
910 different sample sizes (5,000 and 10,000 individuals), PVE levels (1% and 4%), and numbers of
911 causal variants (1-3). Panels A-L represent the 12 simulation scenarios.

912 **Figure S4. Manhattan plots for GWAS of six quantitative traits in Duroc pigs.**

913 Manhattan plots showing GWAS results for three growth traits (off-test body weight [WT], back
914 fat thickness [BF], and loin muscle depth [MS]) analyzed in ~27,000 pigs and three reproduction
915 traits (number of piglets born alive [NBA], born dead [NBD], and weaned [NW]) analyzed in
916 3,290 sows. Association analyses were performed using SLEMM-GWA with ~11.7 million
917 quality-controlled autosomal variants. GRMs were constructed from 10,000 randomly selected
918 chip SNPs. Horizontal lines indicate genome-wide significance thresholds ($P < 5 \times 10^{-8}$, red) and
919 suggestive significance thresholds ($P < 1 \times 10^{-5}$, blue).

920 **Figure S5. Comparison of PIPs between BFMAP-SSS and adapted summary-statistics-**
921 **based fine-mapping methods across two genomic regions.**

922 Scatter plots comparing variant-level (A, B) and gene-level (C, D) PIPs between BFMAP-SSS and
923 adapted summary-statistics methods for back fat thickness. (A) and (C) show results from
924 chr1:51,300,453-55,300,453 including 4,350 variants. (B) and (D) show results from
925 chr6:144,948,358-148,948,358 including 5,195 variants. The dashed line represents y = x.

926 **Figure S6. Impact of FINEMAP-adj iteration lengths on PIPs across two genomic regions.**

927 Root mean square error (RMSE) (A, B) and correlation (C, D) of PIPs calculated between
928 BFMAP-SSS and FINEMAP-adj at variant level (red) and gene level (blue) across different
929 iteration settings. (A) and (C) show results from chr1:51,300,453-55,300,453, and (B) and (D)
930 from chr6:144,948,358-148,948,358. The x-axis represents FINEMAP-adj iteration numbers.
931 Asterisks (*) indicate actual iteration numbers when applying default convergence settings.

932 **Figure S7. Impact of GRM construction on fine-mapping accuracy.**

933 Precision-recall curves comparing FINEMAP-adj and SuSiE-adj when using different GRMs for
934 fine-mapping. Solid lines represent analyses using the same 10,000 chip SNPs used to construct
935 the GRM during the trait simulation (Simulation-10k-SNP GRM), while dashed lines represent
936 analyses using a different set of 10,000 randomly selected variants from the imputed sequence data
937 (Random-10k-variant GRM). Results are shown under different sample sizes (5,000 and 10,000
938 individuals), PVE levels (1% and 4%), and numbers of causal variants (1-3). Panels A-L
939 correspond to the 12 simulation scenarios.

940 **Tables**

941 Table 1. AUPRC for causal variant identification by fine-mapping methods across simulation
942 scenarios.

Sample size	Method	Number of causal variants						Average
		1		2		3		
		PVE=0.01	PVE=0.04	PVE=0.01	PVE=0.04	PVE=0.01	PVE=0.04	
5,000	BFMAP-SSS	0.128	0.419	0.052	0.223	0.028	0.093	0.157
	BFMAP-Forward	0.124	0.379	0.032	0.143	0.011	0.061	0.125
	FINEMAP-adj	0.126	0.381	0.051	0.173	0.030	0.080	0.140
	SuSiE-adj	0.117	0.388	0.036	0.159	0.026	0.066	0.132
	FINEMAP (LMM)	0.067	0.064	0.024	0.020	0.013	0.013	0.034
	FINEMAP (LM)	0.033	0.169	0.010	0.042	0.006	0.020	0.047
	FINEMAP-inf (LMM)	0.033	0.021	0.016	0.012	0.010	0.011	0.017
	FINEMAP-inf (LM)	0.026	0.139	0.011	0.044	0.004	0.023	0.041
	SuSiE (LMM)	0.120	0.192	0.032	0.032	0.019	0.022	0.070
	SuSiE (LM)	0.037	0.187	0.009	0.039	0.005	0.019	0.049
	SuSiE-inf (LMM)	0.021	0.027	0.009	0.016	0.005	0.009	0.015
	SuSiE-inf (LM)	0.019	0.201	0.008	0.040	0.003	0.018	0.048
10,000	BFMAP-SSS	0.188	0.468	0.096	0.378	0.077	0.199	0.234
	BFMAP-Forward	0.140	0.469	0.053	0.250	0.061	0.157	0.188
	FINEMAP-adj	0.189	0.463	0.089	0.303	0.076	0.188	0.218
	SuSiE-adj	0.207	0.478	0.086	0.225	0.066	0.149	0.202
	FINEMAP (LMM)	0.046	0.056	0.023	0.037	0.015	0.027	0.034
	FINEMAP (LM)	0.027	0.065	0.009	0.039	0.017	0.035	0.032
	FINEMAP-inf (LMM)	0.024	0.028	0.011	0.017	0.006	0.018	0.017
	FINEMAP-inf (LM)	0.023	0.070	0.006	0.043	0.014	0.017	0.029
	SuSiE (LMM)	0.109	0.051	0.040	0.038	0.024	0.032	0.049
	SuSiE (LM)	0.028	0.119	0.009	0.022	0.016	0.034	0.038
	SuSiE-inf (LMM)	0.018	0.016	0.012	0.012	0.009	0.008	0.013
	SuSiE-inf (LM)	0.011	0.059	0.004	0.021	0.007	0.022	0.021

943 Note: AUPRC values were calculated by varying PIP thresholds (or PCIP for BFMAP-Forward) from 0 to 1 and
944 computing the area under the resulting precision-recall curves. GWAS summary data were derived from GCTA-
945 MLMA (LMM) or PLINK 2.0 --glm (LM). Results are shown for different sample sizes (5,000 and 10,000
946 individuals), PVE levels (1% and 4%), and numbers of causal variants (1-3).

947

948

949 Table 2. AUPRC for causal gene identification by fine-mapping methods across simulation
950 scenarios.

Sample size	Method	Number of causal variants					
		1		2		3	
		PVE=0.01	PVE=0.04	PVE=0.01	PVE=0.04	PVE=0.01	PVE=0.04
5,000	BFMAP-SSS	0.584	0.836	0.404	0.686	0.267	0.507
	BFMAP-Forward	0.535	0.780	0.277	0.548	0.149	0.371
	FINEMAP-adj	0.574	0.811	0.403	0.622	0.265	0.479
10,000	BFMAP-SSS	0.688	0.856	0.467	0.782	0.420	0.662
	BFMAP-Forward	0.618	0.828	0.341	0.585	0.305	0.527
	FINEMAP-adj	0.693	0.852	0.455	0.721	0.406	0.635

951 Note: AUPRC values were calculated by varying PIP thresholds (or PCIP for BFMAP-Forward) from 0 to 1 and
952 computing the area under the resulting precision-recall curves. Results show the ability of three methods (BFMAP-
953 Forward, BFMAP-SSS, and FINEMAP-adj) to identify causal genes (defined as gene proxy blocks containing true
954 causal variants) under different sample sizes (5,000 and 10,000 individuals), PVE levels (1% and 4%), and numbers
955 of causal variants (1-3).

956

957

Figure 1

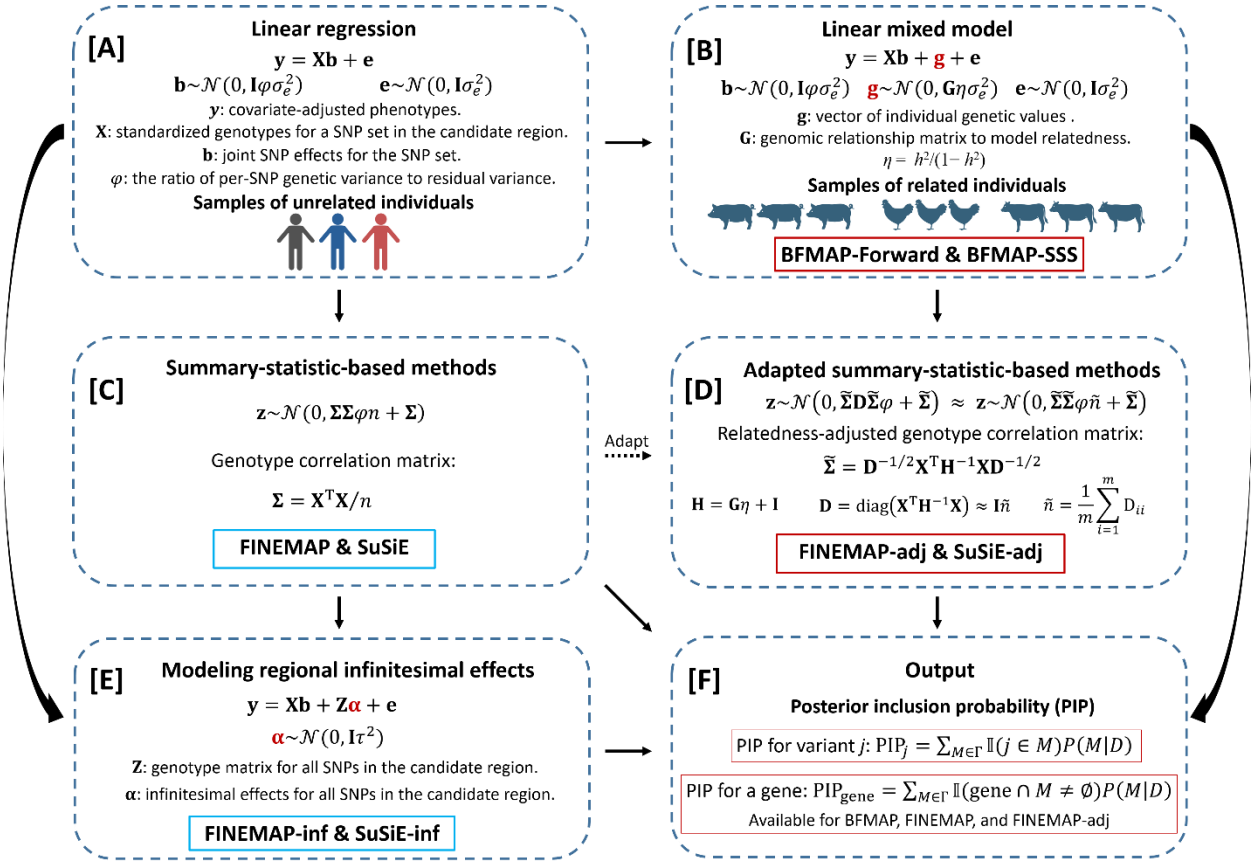

[A]
Linear regression
$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$
$\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2)$ $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$
$\mathbf{y}$: covariate-adjusted phenotypes.
$\mathbf{X}$: standardized genotypes for a SNP set in the candidate region.
$\mathbf{b}$: joint SNP effects for the SNP set.
φ: the ratio of per-SNP genetic variance to residual variance.
Samples of unrelated individuals
[B]
Linear mixed model
$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{g} + \mathbf{e}$
$\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2)$ $\mathbf{g} \sim \mathcal{N}(0, \mathbf{G}\eta\sigma_e^2)$ $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$
$\mathbf{g}$: vector of individual genetic values .
$\mathbf{G}$: genomic relationship matrix to model relatedness.
$\eta = h^2/(1-h^2)$
Samples of related individuals
BFMAP-Forward & BFMAP-SSS
[C]
Summary-statistic-based methods
$\mathbf{z} \sim \mathcal{N}(0, \mathbf{\Sigma}\mathbf{\Sigma}\varphi n + \mathbf{\Sigma})$
Genotype correlation matrix:
$\mathbf{\Sigma} = \mathbf{X}^{\mathrm{T}}\mathbf{X}/n$
FINEMAP & SuSiE
Adapt
[D]
Adapted summary-statistic-based methods
$\mathbf{z} \sim \mathcal{N}(0, \tilde{\mathbf{\Sigma}}\mathbf{D}\tilde{\mathbf{\Sigma}}\varphi + \tilde{\mathbf{\Sigma}}) \approx \mathbf{z} \sim \mathcal{N}(0, \tilde{\mathbf{\Sigma}}\tilde{\mathbf{\Sigma}}\varphi\tilde{n} + \tilde{\mathbf{\Sigma}})$
Relatedness-adjusted genotype correlation matrix:
$\tilde{\mathbf{\Sigma}} = \mathbf{D}^{-1/2}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\mathbf{D}^{-1/2}$
$\mathbf{H} = \mathbf{G}\eta + \mathbf{I}$ $\mathbf{D} = \mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}) \approx \mathbf{I}\tilde{n}$ $\tilde{n} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{D}_{ii}$
FINEMAP-adj & SuSiE-adj
[E]
Modeling regional infinitesimal effects
$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$
$\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{I}\tau^2)$
$\mathbf{Z}$: genotype matrix for all SNPs in the candidate region.
$\boldsymbol{\alpha}$: infinitesimal effects for all SNPs in the candidate region.
FINEMAP-inf & SuSiE-inf
[F]
Output
Posterior inclusion probability (PIP)
PIP for variant j: $\mathrm{PIP}_j = \sum_{M \in \Gamma} \mathbb{I}(j \in M) P(M|D)$
PIP for a gene: $\mathrm{PIP}_{\mathrm{gene}} = \sum_{M \in \Gamma} \mathbb{I}(\mathrm{gene} \cap M \neq \emptyset) P(M|D)$
Available for BFMAP, FINEMAP, and FINEMAP-adj

Figure 2

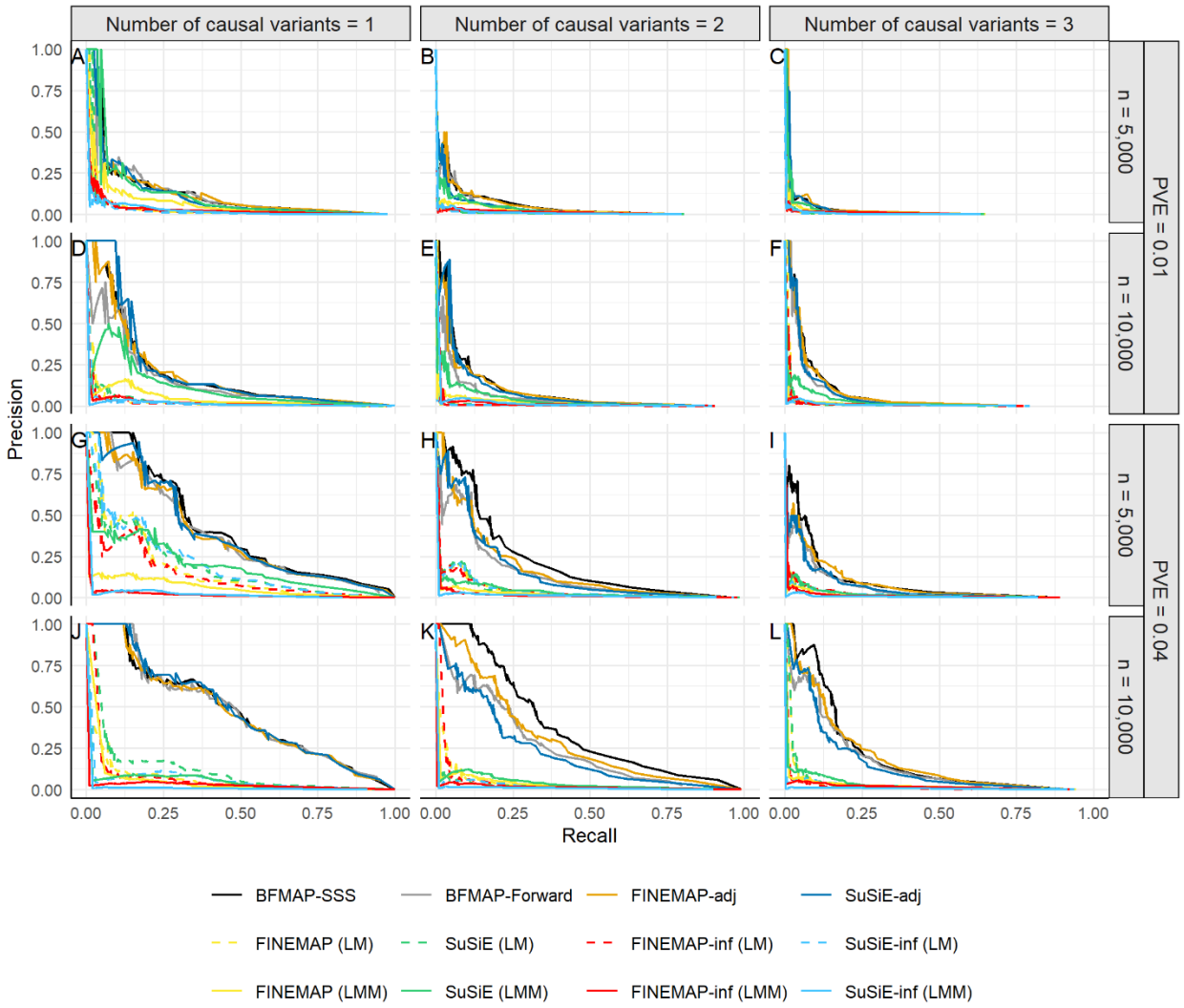

Number of causal variants = 1
Number of causal variants = 2
Number of causal variants = 3
PVE = 0.01
PVE = 0.04
n = 5,000
n = 10,000
A
B
C
D
E
F
G
H
I
J
K
L
Precision
Recall
0.00
0.25
0.50
0.75
1.00
BFMAP-SSS
BFMAP-Forward
FINEMAP-adj
SuSiE-adj
FINEMAP (LM)
SuSiE (LM)
FINEMAP-inf (LM)
SuSiE-inf (LM)
FINEMAP (LMM)
SuSiE (LMM)
FINEMAP-inf (LMM)
SuSiE-inf (LMM)

Figure 3

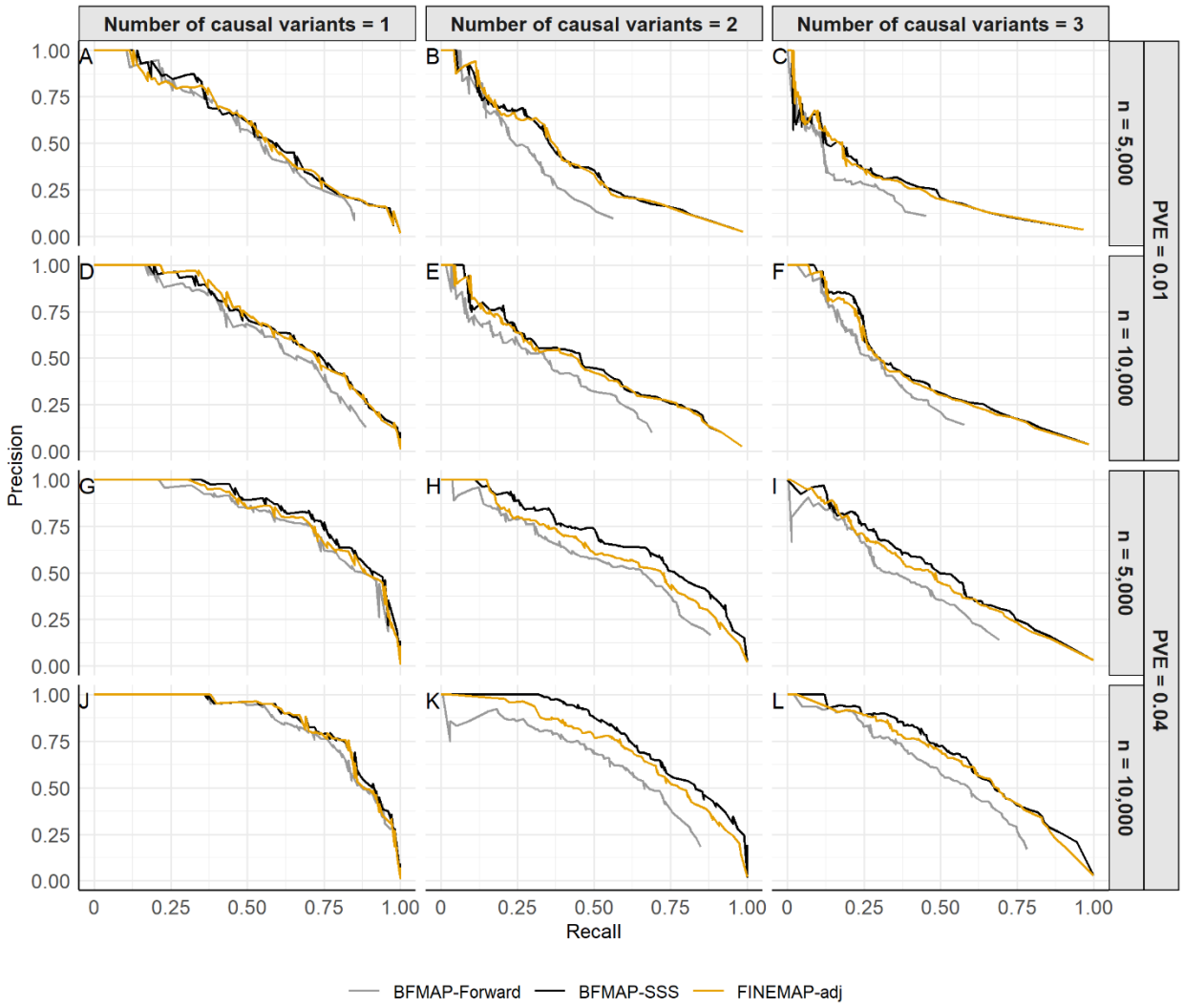
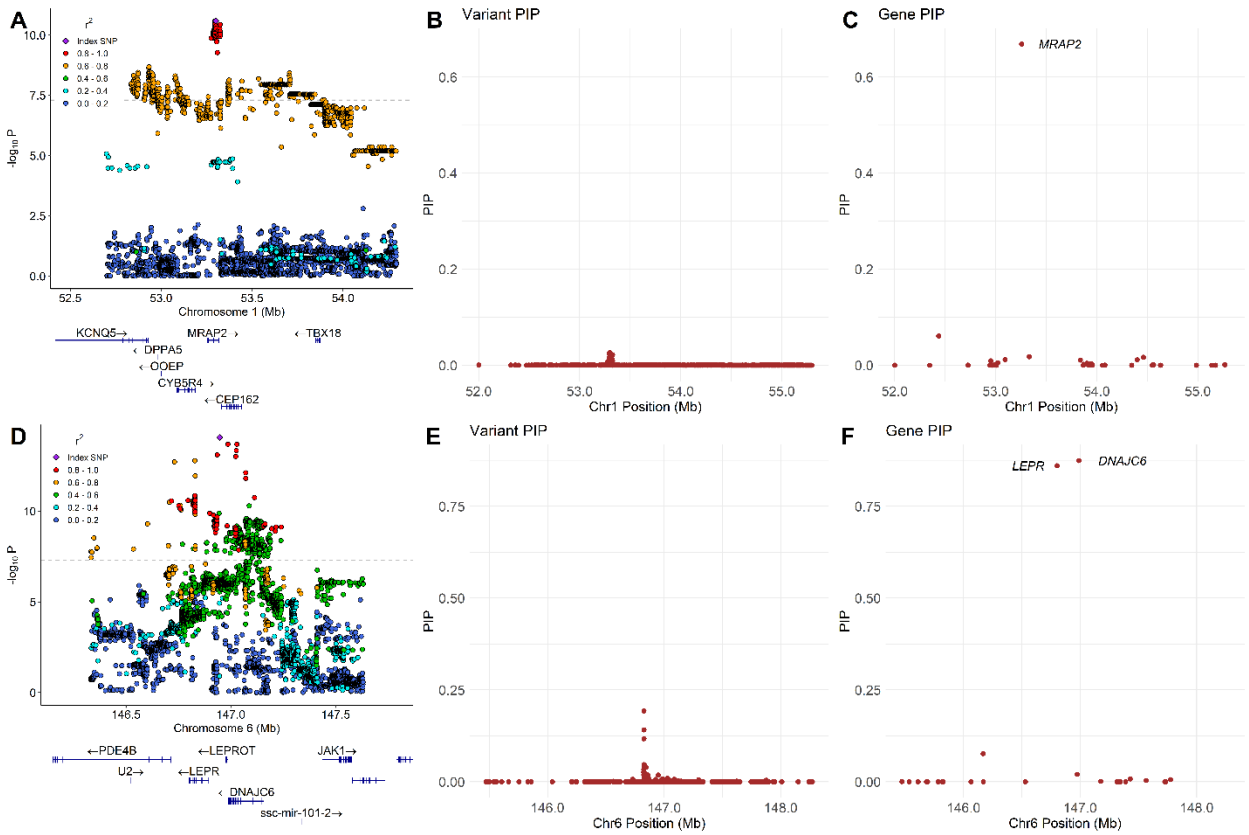

Number of causal variants = 1
Number of causal variants = 2
Number of causal variants = 3
PVE = 0.01
PVE = 0.04
n = 5,000
n = 10,000
n = 5,000
n = 10,000
A
B
C
D
E
F
G
H
I
J
K
L
Precision
Recall
1.00
0.75
0.50
0.25
0.00
0
0.25
0.50
0.75
1.00
BFMAP-Forward
BFMAP-SSS
FINEMAP-adj

Figure 4


A
B
Variant PIP
C
Gene PIP
MRAP2
D
E
Variant PIP
F
Gene PIP
LEPR
DNAJC6
Chromosome 1 (Mb)
Chr1 Position (Mb)
Chromosome 6 (Mb)
Chr6 Position (Mb)
PIP
KCNQ5
MRAP2
TBX18
DPPA5
OOEP
CYB5R4
CEP162
PDE4B
LEPROT
JAK1
U2
LEPR
DNAJC6
ssc-mir-101-2

[A] Linear regression

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2) \qquad \mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$$

$\boldsymbol{y}$: covariate-adjusted phenotypes.

$\mathbf{X}$: standardized genotypes for a SNP set in the candidate region.

$\mathbf{b}$: joint SNP effects for the SNP set.

φ: the ratio of per-SNP genetic variance to residual variance.

Samples of unrelated individuals

[B] Linear mixed model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{g} + \mathbf{e}$$

$$\mathbf{b} \sim \mathcal{N}(0, \mathbf{I}\varphi\sigma_e^2) \quad \mathbf{g} \sim \mathcal{N}(0, \mathbf{G}\eta\sigma_e^2) \quad \mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma_e^2)$$

$\mathbf{g}$: vector of individual genetic values .

$\mathbf{G}$: genomic relationship matrix to model relatedness.

$\eta = h^2/(1-h^2)$

Samples of related individuals

BFMAP-Forward & BFMAP-SSS

[C] Summary-statistic-based methods

$$\mathbf{z} \sim \mathcal{N}(0, \boldsymbol{\Sigma}\boldsymbol{\Sigma}\varphi n + \boldsymbol{\Sigma})$$

Genotype correlation matrix:

$$\boldsymbol{\Sigma} = \mathbf{X}^{\mathrm{T}}\mathbf{X}/n$$

FINEMAP & SuSiE

[D] Adapted summary-statistic-based methods

$$\mathbf{z} \sim \mathcal{N}\left(0, \widetilde{\boldsymbol{\Sigma}}\mathbf{D}\widetilde{\boldsymbol{\Sigma}}\varphi + \widetilde{\boldsymbol{\Sigma}}\right) \approx \mathbf{z} \sim \mathcal{N}\left(0, \widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Sigma}}\varphi\tilde{n} + \widetilde{\boldsymbol{\Sigma}}\right)$$

Relatedness-adjusted genotype correlation matrix:

$$\widetilde{\boldsymbol{\Sigma}} = \mathbf{D}^{-1/2}\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\mathbf{D}^{-1/2}$$

$$\mathbf{H} = \mathbf{G}\eta + \mathbf{I} \qquad \mathbf{D} = \mathrm{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{H}^{-1}\mathbf{X}\right) \approx \mathbf{I}\tilde{n} \qquad \tilde{n} = \frac{1}{m}\sum_{i=1}^{m} \mathrm{D}_{ii}$$

FINEMAP-adj & SuSiE-adj

[E] Modeling regional infinitesimal effects

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$$

$$\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{I}\tau^2)$$

$\mathbf{Z}$: genotype matrix for all SNPs in the candidate region.

$\boldsymbol{\alpha}$: infinitesimal effects for all SNPs in the candidate region.

FINEMAP-inf & SuSiE-inf

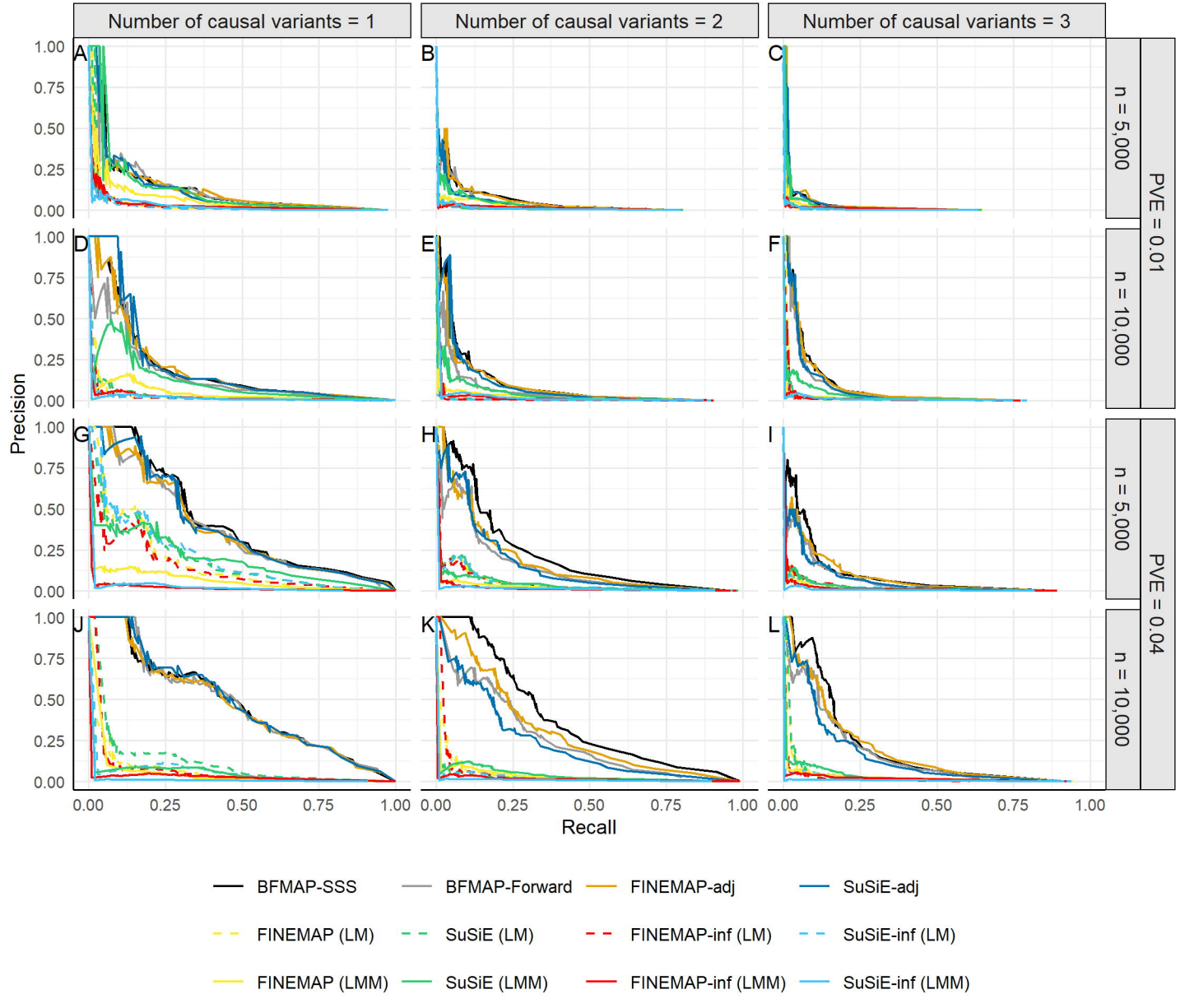
[F] Output

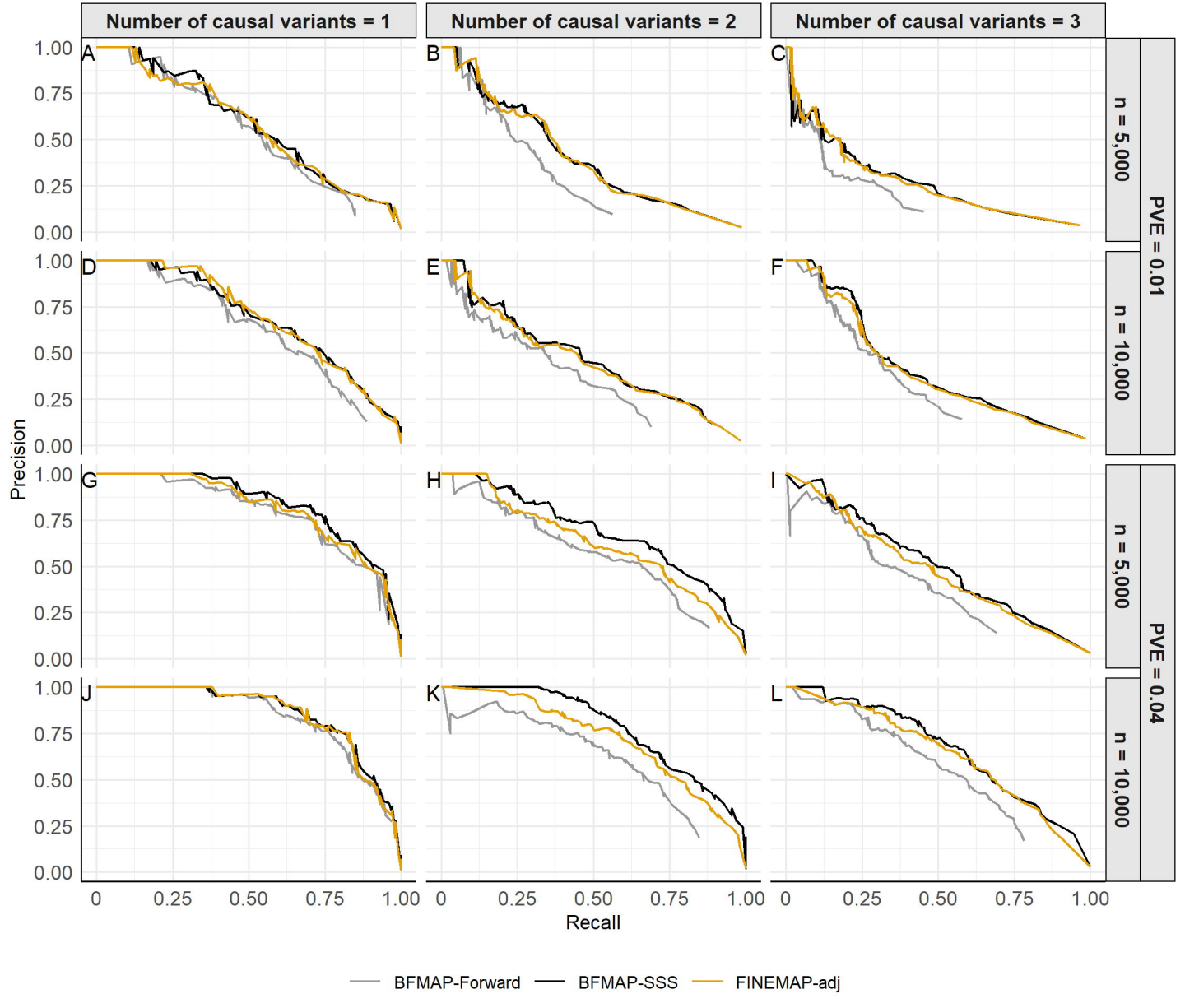
Posterior inclusion probability (PIP)

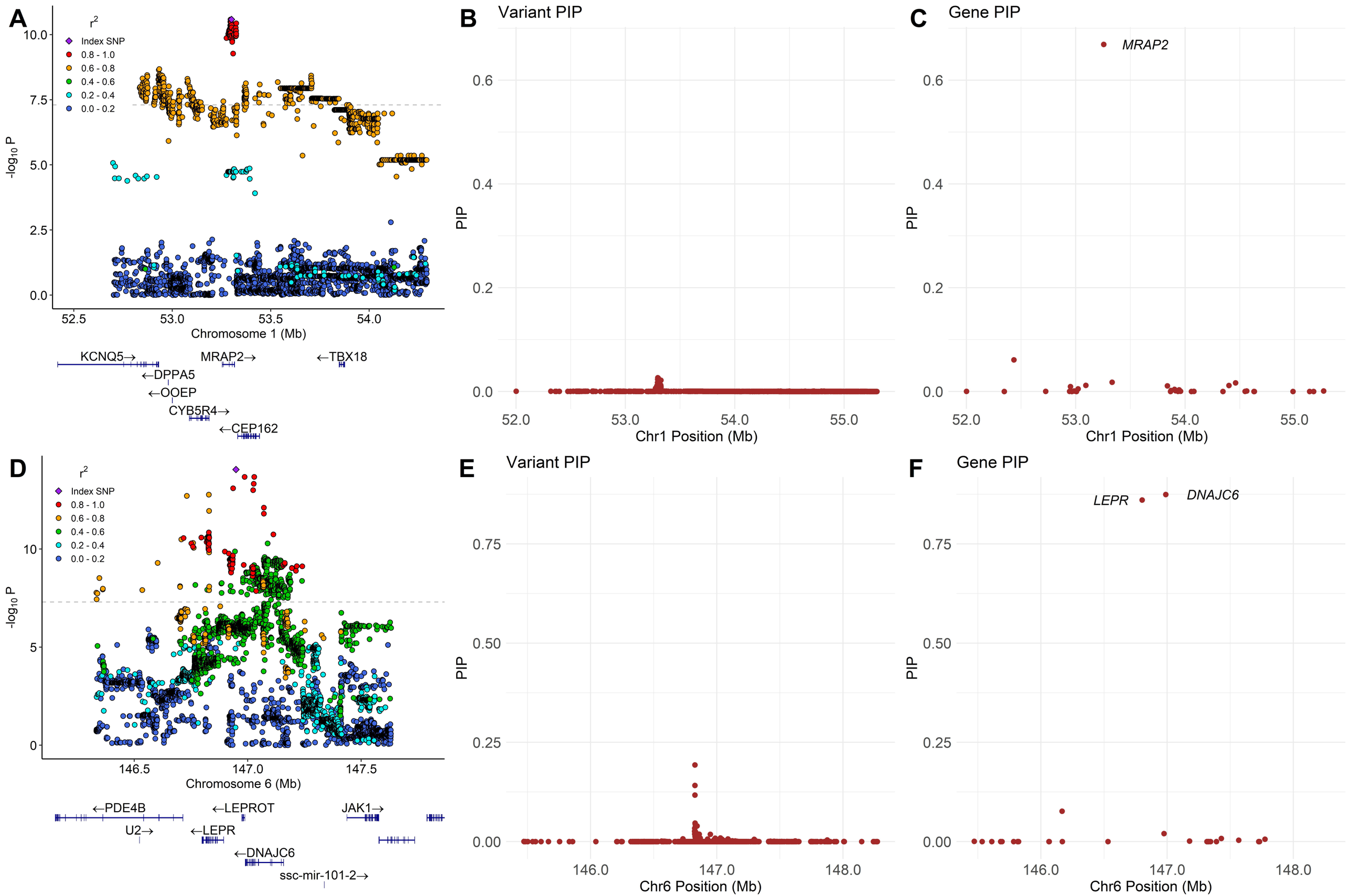
PIP for variant j: $\mathrm{PIP}_j = \sum_{M \in \Gamma} \mathbb{I}(j \in M) P(M|D)$

PIP for a gene: $\mathrm{PIP}_{\mathrm{gene}} = \sum_{M \in \Gamma} \mathbb{I}(\mathrm{gene} \cap M \neq \emptyset) P(M|D)$

Available for BFMAP, FINEMAP, and FINEMAP-adj


Number of causal variants = 1
Number of causal variants = 2
Number of causal variants = 3
PVE = 0.01
PVE = 0.04
n = 5,000
n = 10,000
n = 5,000
n = 10,000
A
B
C
D
E
F
G
H
I
J
K
L
Precision
Recall
0.00
0.25
0.50
0.75
1.00
BFMAP-SSS
BFMAP-Forward
FINEMAP-adj
SuSiE-adj
FINEMAP (LM)
SuSiE (LM)
FINEMAP-inf (LM)
SuSiE-inf (LM)
FINEMAP (LMM)
SuSiE (LMM)
FINEMAP-inf (LMM)
SuSiE-inf (LMM)


Number of causal variants = 1
Number of causal variants = 2
Number of causal variants = 3
PVE = 0.01
PVE = 0.04
n = 5,000
n = 10,000
n = 5,000
n = 10,000
A
B
C
D
E
F
G
H
I
J
K
L
Precision
Recall
0.00
0.25
0.50
0.75
1.00
0
BFMAP-Forward
BFMAP-SSS
FINEMAP-adj


A
r2
Index SNP
0.8 - 1.0
0.6 - 0.8
0.4 - 0.6
0.2 - 0.4
0.0 - 0.2
-log10 P
Chromosome 1 (Mb)
KCNQ5→
MRAP2→
←TBX18
←DPPA5
←OOEP
CYB5R4→
←CEP162
B
Variant PIP
PIP
Chr1 Position (Mb)
C
Gene PIP
MRAP2
D
Chromosome 6 (Mb)
←PDE4B
←LEPROT
JAK1→
U2→
←LEPR
←DNAJC6
ssc-mir-101-2→
E
Variant PIP
Chr6 Position (Mb)
F
Gene PIP
LEPR
DNAJC6