CARMA: Novel Bayesian model for fine-mapping with high-dimensional functional data

Zikun Yang[1], Chen Wang[1,2], Atlas Khan[2], Krzysztof Kiryluk[2], Iuliana Ionita-Laza[1,#]

February 1, 2022

[1] Department of Biostatistics, Columbia University, New York

[2] Division of Nephrology, Department of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York

[#]Correspondance: ii2135@columbia.edu

Abstract

We propose a novel Bayesian model for fine-mapping in order to identify putative causal variants at GWAS loci. Relative to existing fine-mapping models, the proposed model has several appealing features, including assuming a heavy-tail distribution on effect sizes, joint modeling of summary statistics and large number of functional annotations, and accounting for discrepancies between summary statistics and external linkage disequilibrium (LD) values in meta-analysis settings. We compare performance with commonly used fine-mapping methods, including fastPAINTOR, SuSiE and PolyFun+SuSiE in simulations and show that the proposed model has improved performance, including higher power to identify known causal variants, higher coverage and less false positives. We further illustrate our approach by applying it to a meta-analysis of Alzheimer's Disease GWAS data where we prioritize putatively causal variants and genes.

Introduction

Meta-analyses of GWAS studies have identified a large number of significant loci. Fine-mapping is the natural next step in order to identify putative causal genetic variants at these loci. Meta-analyses however pose several challenges that can invalidate the results from existing fine-mapping methods. For example, using linkage disequilibrium (LD) from external panels can create inconsistencies with GWAS summary statistics which can lead fine-mapping methods to prioritize non-causal variants. Similarly, uneven sample size coverage at different variants can lead to biased posterior inclusion probability (PIP) values. To illustrate this point we show the example of a GWAS locus *SCIMP* (SLP adaptor and CSK interacting membrane protein) for Alzheimer's disease (AD). We use summary statistics from a large meta-analysis GWAS of clinically diagnosed AD and AD-by-proxy with 71,880 cases and 383,378 controls of European ancestry from three consortia.[1] The LD matrix is estimated using individuals of European descent in UK Biobank.[2] Two fine-mapping models, SuSiE[3] and fastPAINTOR,[4] prioritize variants with low Z-scores due to discrepancies between summary statistics and LD values, whereas the results of the proposed model appear as expected (Figure 1). In this paper we introduce a new fine-mapping method that improves the power and reduces false positives in such situations.

There are many statistical fine-mapping methods in the literature.[3–11] Most of the existing methods can work with summary statistics and LD information from relevant reference panels, and make certain assumptions on the number of causal variants; some methods restrict the number of possible causal variants,[5,6,9,11] while others relax this assumption by introducing a prior distribution on model space, and implementing stochastic algorithms such as Markov Chain Monte Carlo (MCMC) to reduce the computational cost.[4,7,8] Multiplicity control is another important aspect of any Bayesian fine-mapping method in order to control the false discovery rate in the context of multiple testing,[12] and several existing methods address this issue formally by introducing prior probabilities on model space.[7,8] It is also worth mentioning that existing Bayesian fine-mapping methods

usually assume a Normal distribution as the prior distribution on effect sizes, whereas it has been shown that assuming a heavy-tail prior distribution can substantially increase association power;[13] in particular, a heavy-tail distribution is substantially more sensitive to large signals.

Here we propose a new Bayesian model, CARMA (CAusal Robust Mapping method with Annotations), that attempts to improve upon existing methods especially in complex meta-analyses settings as described above. Our proposed model has several technical innovations: (1) it replaces the usual Normal-Gamma prior family for the effect size distribution by a heavy-tail Cauchy distribution, which was previously shown to satisfy basic consistency requirements for variable selection[14] and may better reflect the empirical data; (2) it jointly models summary statistics and high-dimensional functional annotations - this is different from other recent models such as PolyFun[10] that first estimate prior causal probabilities based on functional annotations and summary statistics, and then apply fine-mapping methods such as SuSiE with these prior probabilities; (3) it introduces a novel Bayesian hypothesis testing approach to account for discrepancies between summary statistics and LD from external reference panels in order to avoid an increase in false positives. We illustrate the proposed method using simulations and applications to an AD GWAS meta-analysis.

Results

Overview of the proposed model

We assume that we have genotype ($\boldsymbol{X}$) and phenotype ($\boldsymbol{y}$) data for n subjects. For a given locus, we assume a standard linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_y^2 I_n)$, where $\boldsymbol{y}$ is a $n \times 1$ vector of quantitative phenotype values, $\boldsymbol{X}$ is a standardized $n \times p$ genotype matrix, $\boldsymbol{\beta}$ is a p-dimensional vector of effect sizes of variants, and $\boldsymbol{\epsilon}$ is a Gaussian noise vector. For each variant i with $i = 1, \ldots, p$, we obtain the estimated marginal effect $\hat{\beta}_i$ and standard error $\text{se}\left(\hat{\beta}_i\right)$. Then the Z-score is defined as $Z_i = \frac{\hat{\beta}_i}{\text{se}\left(\hat{\beta}_i\right)}$ with $Z_i \sim N(0,1)$ under the null hypothesis $H_0 :\ \beta_i = 0$. Let $\boldsymbol{\gamma}' = \{0,1\}^p$ denote an indicator vector, such that $\gamma_i = 1$ iff $\beta_i \neq 0$. Let $\boldsymbol{\lambda} \propto \boldsymbol{\beta}$. We want to identify the true model (denoted by the indicator vector $\boldsymbol{\gamma}_T$) that generated the summary statistics through posterior inference within a Bayesian paradigm.

Given any model $\boldsymbol{\gamma}$, we make the following model assumptions:

$$\begin{aligned}
\boldsymbol{Z}|\boldsymbol{\lambda}_{\boldsymbol{\gamma}}, \sigma_y^2, \boldsymbol{\Sigma} &\sim \text{MVN}(\boldsymbol{\Sigma}\boldsymbol{\lambda}_{\boldsymbol{\gamma}}, \sigma_y^2\boldsymbol{\Sigma}), \\
\boldsymbol{\lambda}_{\boldsymbol{\gamma}}|\sigma_y^2, \tau &\sim \text{MVN}(0, \frac{\sigma_y^2}{\tau}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}), \\
\tau &\sim \text{Gamma}(0.5, 0.5), \\
\sigma_y^2 &\sim \frac{1}{\sigma_y^2},
\end{aligned}$$

where $\boldsymbol{\Sigma} = \frac{\boldsymbol{X}'\boldsymbol{X}}{n}$. Note that the marginal prior distribution of $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ is a multivariate Cauchy distribution after integrating out the mixing parameter τ. We additionally assume a truncated Poisson prior on the size of the model, i.e. $\sum_{i=1}^p \gamma_i \sim$ Truncated Poisson (η) and $\sum_{i=1}^p \gamma_i \in \{0, 1, \ldots, p\}$, to control the total number of causal variants assumed by a given model, and hence provide multiplicity control or control of the false discovery rate. We implement a Shotgun stochastic search algorithm[15] for searching the posterior distribution over model space, which has advantages over the more commonly used MCMC-based methods in that it leads to a more complete exploration of the areas with high marginal likelihood in the posterior model space. This semi-exhaustive search feature alleviates the problem of unequal PIPs for perfectly correlated variants that can happen with other MCMC algorithms.

High-dimensional functional annotations and meta-analyses settings. The main advantages of the proposed model over existing fine-mapping models are when incorporating large number of functional annotations, and in the context of meta-analyses when mismatch between external LD and GWAS summary statistics can lead to false positive results when using existing methods.

CARMA seeks to distinguish highly correlated variants by leveraging functional annotations. Let $\boldsymbol{W}$ be a $p \times q$ matrix for p SNPs and q functional annotations. We maximize the joint likelihood of the summary statistics and functional annotations

$$L(\boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{\gamma}, \boldsymbol{W}) = f(\boldsymbol{Z}|\boldsymbol{\gamma}) \Pr(\boldsymbol{\gamma}|\boldsymbol{W}, \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is a q-dimensional vector of the corresponding coefficients, through an EM algorithm with the prior probability $\Pr(\boldsymbol{\gamma}|\boldsymbol{W}, \boldsymbol{\theta})$ being linked to functional annotations via a penalized Poisson regression model using Elastic Net[16] to account for correlations among the annotations while still providing sparse estimates.

To account for possible discrepancies between summary statistics and LD estimated from external reference panels, we propose a Bayesian hypothesis testing procedure for detecting inconsistencies among highly correlated variants. We also introduce the concept of credible models, as an alternative to credible sets, based on the top candidate models in terms of posterior probability. Compared to credible sets, credible models typically involve less variants while identifying a higher proportion of causal variants as shown in simulations below. More details are available in the Methods section.

Simulations

We first perform simulations to investigate the performance of the CARMA model and competitor methods, SuSiE, fastPAINTOR and SuSiE+PolyFun.

Generating simulated datasets

Genotype simulation. We use the R package 'sim1000G'[17] to simulate genotypes based on the 1000 Genomes Project data (phase 3, European population). To select regions representative of GWAS loci in terms of size and LD structure, we focus on 94 loci identified as risk regions in a recent GWAS on breast cancer.[18] Within each region, we filter out rare variants (MAF<0.01); the number of variants in each region ranges between $\sim 1,500 - 4,000$. We simulate genotype data for $n = 10,000$ individuals.

Causal SNP selection. We assume two scenarios (1) no functional annotation and (2) with functional annotations (919 DeepSEA chromatin features[19]). Let $\boldsymbol{\theta}_{\text{True}}$ denote a 920-dimensional coefficient vector including the intercept term; for each chromosome, we randomly select 200 chromatin features as being related to the causal status of variants, with the coefficients being sampled from a Normal distribution $N(0, 0.01)$. Hence the dimension of the relevant annotations is high and the impact of each individual annotation is weak. We compute the prior probability of SNP ith being causal as $\Pr(\gamma_i = 1|\boldsymbol{\theta}_{\text{True}}, \boldsymbol{w}_i) = \frac{\exp\{\boldsymbol{w}_i'\boldsymbol{\theta}_{\text{True}}\}}{1+\exp\{\boldsymbol{w}_i'\boldsymbol{\theta}_{\text{True}}\}}$, where $\boldsymbol{w}_i$ is the vector of the corresponding DeepSEA features. The total number of causal SNPs per locus is set at 3, selected among those with highest prior probabilities.

Phenotype generation. For a given locus, let T denote the index set of the true causal SNPs, i.e., if $i \in T$, then $\gamma_i = 1$. The effect sizes for the causal SNPs are related to the MAF as follows: $\beta_i = \left\{ \sqrt{\frac{\max_{i \in T}\{\sum_{h=1}^{n} x_{h,i}\}}{\sum_{h=1}^{n} x_{h,i}}}; i \in T \right\}$, where $\boldsymbol{x}_i$ is the vector of genotypes for the ith SNP. We set $\beta_i = 0$ if $i \notin T$. The phenotypic variance σ_y^2 is computed such that $\phi = 0.0075$, where $\phi = \frac{\text{Var}(\boldsymbol{X}\boldsymbol{\beta}_T)}{\sigma_y^2 + \text{Var}(\boldsymbol{X}\boldsymbol{\beta}_T)}$. Then we sample the phenotype $\boldsymbol{y}$ such that $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_T + \boldsymbol{\epsilon}$; $\boldsymbol{\epsilon} \sim \text{N}(0, \sigma_y^2 I_{n \times n})$.

Remarks on some implementation details. We compare performance with fastPAINTOR and SuSiE (+PolyFun).

1. fastPAINTOR cannot handle high-dimensional annotations therefore we select a subset of the ten most informative annotations as follows: we first compute the correlations between the summary statistics and each functional annotation, then we iteratively select top correlated annotations such that no annotation

has absolute correlation > 0.3 with previously selected annotations. Then, fastPAINTOR is run based on the first 10 selected annotations.

2. For SuSiE+PolyFun (when including functional annotations), we first compute prior causal probabilities via an L2-regularized extension of S-LDSC implemented in PolyFun. Then, the prior causal probabilities are provided as input to the SuSiE model.

3. For CARMA, the value of hyper-parameter η is set at the default setting $\sqrt{\frac{1}{p}}$ unless otherwise noted, and hence depends on the number of SNPs at each locus. Additional results using different values of η can be found in Figures S1 to S4 and tables S1 and S2.

4. We run fastPAINTOR and CARMA models at a chromosome level, while SuSiE (+PolyFun) is run one locus at a time (since SuSiE does not have the option to aggregate multiple loci).

Simulation Results

Performance as measured by AUROC and AUPR. To assess accuracy in predicting causal variants, we compute AUROC and AUPR values based on PIPs from the different models for the two scenarios, no functional annotation, or with functional annotations (Figure 2a). When no functional annotation, the proposed model performs similarly to SuSiE and better than fastPAINTOR. With functional information as prior, the accuracy of the three models tends to improve, especially for the proposed model and fastPAINTOR in terms of AUPR. The SuSiE+PolyFun model consists of two separate modeling procedures, and hence the summary statistics and the functional information are not modeled jointly. That might explain the relatively minor improvement of SuSiE+PolyFun over SuSiE alone. Overall, the proposed model performs substantially better relative to SuSiE and fastPAINTOR especially in terms of AUPR.

We also compute the AUROC and AUPR using the estimated prior probability of the CARMA model $\Pr(\boldsymbol{\gamma}|\boldsymbol{W}, \boldsymbol{\theta})$ as a predictor, for each locus separately. The mean AUROC/AUPR value across loci is 0.96/0.45, therefore the estimated prior probability provides useful information on the SNPs being causal, and leads to improved accuracy when including functional annotations (Figure S5).

Credible sets. We next examine properties of the credible sets from each model. For SuSiE, we use the credible sets reported by SuSiE. For CARMA and fastPAINTOR we compute credible sets as in.[3] More details are in the Methods section. Furthermore, we compute four statistics for the credible sets as in:[3]

1. **Power**: The overall proportion of simulated causal variants included in a credible set.

2. **Coverage**: The proportion of credible sets that contain a causal variant.

3. **Size**: The number of variants included in a credible set.

4. **Purity**: The average squared correlation of variants in a credible set.

The results are shown in Figure 2b for level $\rho = 0.99$ (results for $\rho = 0.95$ can be found in Figure S6). When functional annotations are included, the power of CARMA increases substantially relative to the scenario with no functional annotation, and relative to the competitor models, SuSiE and fastPAINTOR. Note that CARMA has somewhat lower coverage (but still high > 0.9) compared to SuSiE, but has considerably lower size for the credible sets and comparable values for purity. As with the AUROC/AUPR results above, the inclusion of functional annotations has a greater impact for the models that jointly integrate summary statistics and functional annotations like CARMA and fastPAINTOR, compared to SuSiE+PolyFun. These results suggest that to better leverage functional annotations, algorithms that maximize the marginal likelihood conditional on both summary statistics and annotations as proposed here are preferable. Finally, the substantially lower coverage of the credible sets from fastPAINTOR suggests the need of including dimensional penalization to control the false positives.

Credible model. The CARMA model provides posterior probabilities of the visited candidate models for each locus. Let $\gamma_{(1)}$ denote the leading model that receives the largest posterior probability among the visited candidate models; then we define as credible models the set of candidate models such that the posterior odds between the leading model and the models in the set is smaller than a pre-determined threshold, such as 3.2 or 10 as recommended in.[20] We view credible models as complementary to credible sets. Therefore, for each locus, we compare the results based on the variants included by the credible models in CARMA and the variants included by all the credible sets for CARMA, SuSiE and fastPAINTOR. Compared to the credible sets of CARMA, SuSiE and fastPAINTOR, the CARMA credible models achieve considerable higher power with smaller sets of selected variants, which may facilitate follow-up investigations (Figure S7, Table S3).

PIPs for perfectly correlated SNPs. Methods based on stochastic algorithms such as MCMC often fail to evenly explore the area of the posterior model space which creates problems when there are perfectly correlated SNPs. Due to the semi-exhaustive search feature of the Shotgun algorithm, CARMA examines the neighborhood of the currently selected model, including candidate models that exchange perfectly correlated SNPs. This way, the resulting PIPs tend to be more consistent between highly correlated SNPs. To illustrate this point, we examine the standard deviation of the PIPs within groups of perfectly correlated SNPs based on the results of the three models when no functional annotation is included. Across the 94 loci, there are 30,131 groups of perfectly correlated SNPs with an average of 317 groups at each locus, and 4.23 SNPs per group. Since the values of the PIPs returned by each model can vary across different models, we standardized the PIPs within each group by dividing the PIPs by the maximum PIP in each group. CARMA has a standard deviation of 0.005 for PIPs for perfectly correlated SNPs. This is in contrast to fastPAINTOR which has a standard deviation of 0.509. Note that the SuSiE model has identical PIPs for perfectly correlated SNPs by definition due to its algorithm that runs univariate regression individually.

Effect of LD-summary statistics inconsistencies in meta-analyses. In complex meta-analyses studies, with LD estimated from external reference panels, it is not uncommon to have discrepancies between Z-scores and LD values, which leads to biased PIPs for existing fine-mapping models, and false prioritization of non-causal variants. Here we show the robustness of the CARMA model in such scenarios.

We artificially create such scenarios in simulations for all 94 loci. For each locus, we first identify all groups with at least three SNPs such that the minimum correlation within the group is greater than 0.9. We rank these groups in terms of the absolute value of the average Z-score within group. Then starting with the group with the largest value, we select a random SNP in the set and add a random value $\sim Unif(1,2)$ to its Z-score, with the sign randomly drawn with probability 0.5. We compute the corresponding Bayes factor for the test on whether the Z-score of the selected SNP has the same distribution as the rest of the group (see Methods). If the corresponding Bayes factor is smaller than 10^{-20} (note that this is a very stringent threshold), then we move to the next group, otherwise, we repeat the procedure for the current group until the Bayes factor is less than this threshold. Each locus includes 10 such outliers, one for each group of highly correlated SNPs, for a total of 940 outliers across the 94 loci. An example illustrating such a spike-in dataset along with the results of CARMA, SuSiE and fastPAINTOR on the original and the spike-in datasets are shown in Figure 3.

To assess the robustness of the three models in the presence of outliers, we apply the three fine-mapping models on these spike-in datasets and compare results to those based on the original datasets. The presence of outliers can substantially bias the term $\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}$ (e.g. $\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}$ can become negative or extremely large), which leads to invalid marginal likelihood for the CARMA model (Methods). We fix $\sigma_y^2 = 1$ for these scenarios because this way the term $\exp\left\{-\frac{\boldsymbol{Z}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}}{2\sigma_y^2}\right\}$ can be ignored (it is constant across all candidate models). We also set $\eta = 5$ to account for the change in likelihood when fixing σ_y^2 at 1 (Figures S1 to S4 and tables S1 and S2).

CARMA is largely unaffected in terms of AUROC and AUPR values when outliers are artificially introduced in the data (Figure 4a). For SuSiE and fastPAINTOR the prioritization of the causal SNPs based on PIPs is affected by the presence of outliers, i.e., outliers receive large PIPs and therefore are identified as the potentially causal SNPs instead of the true causal SNPs, as also exemplified in Figure 3. The coverage values for the credible sets for SuSiE and fastPAINTOR drop substantially, to 51% and 14% respectively, due to many outliers

receiving high PIPs leading to false credible sets, whereas the proposed model still maintains high coverage (94%). Similarly, in terms of power, CARMA is largely unaffected by the existence of outliers, whereas SuSiE and fastPAINTOR suffer substantial power loss (Figure 4b). Similar robustness for CARMA can be observed with respect to credible models (Figure S8 and Table S2).

Sensitivity of the outlier detection. In all 94 loci, we detect a total number of 666 outliers, 95% of which are the simulated outliers. We also run the outlier detection algorithm on the original data at the 94 loci. Only 6 out of the 94 loci have detected outliers, for a total number of 58, concordant with chance expectation.

Credible sets with only one SNP. A particular situation arises when a credible set contains only one SNP (i.e. its PIP exceeds ρ). Such scenarios require that a SNP have a large summary statistic and be relatively independent of all the other SNPs. An outlier therefore may lead to such a situation because the discordant LD/Z-score values may create the appearance of a SNP with large summary statistic but weakly correlated to surrounding SNPs. Figure 5 shows the total number of credible sets with one SNP for the three testing models across the 94 loci, and the corresponding causal status of the SNPs in these credible sets. As shown, for the original datasets with no spike-in outliers, the three models work well. However, when outliers are introduced in the data, the number of credible sets with one SNP for SuSiE and fastPAINTOR increases substantially, and the coverage levels of the credible sets are poor, showing that indeed outliers, when present, receive large PIP values from these existing models.

Remarks on limitations of SuSiE and fastPAINTOR. The results of the simulation studies show that fastPAINTOR suffers from poor coverage even without the existence of the simulated outliers. This is due to the lack of multiplicity control in the EM algorithm of fastPAINTOR, which can make the PIP values sensitive to the prior probabilities generated by the functional annotations. We also note that the EM algorithm in fastPAINTOR is sensitive to the choice of seed for the random number generator (see Supplemental material in section "Additional results on real data analyses"). The main limitation of SuSiE+PolyFun is that the summary statistics and the functional annotations are not modeled jointly. Therefore, the results of the SuSiE model after including functional annotations do not improve as much as those for CARMA and fastPAINTOR. Furthermore, the inference based on the SuSiE and fastPAINTOR models can be greatly influenced by discrepancies between summary statistics and external LD values as shown above.

Fine-mapping Alzheimer's disease GWAS loci

In this section we present fine-mapping results at 30 GWAS loci identified in a large meta-analysis of clinically diagnosed AD and AD-by-proxy with 71,880 cases and 383,378 controls of European ancestry.[1] The clinically diagnosed AD case-control data are from 3 consortia (Alzheimer's disease working group of the Psychiatric Genomics Consortium (PGC-ALZ), the International Genomics of Alzheimer's Project (IGAP), and the Alzheimer's Disease Sequencing Project (ADSP)), and the AD-by-proxy data are based on 376,113 individuals of European ancestry from UKBB. We use the leading SNP at each locus from the meta-analysis (phase 3) and for the purposes of fine-mapping define the locus as $\pm$ 500kb centered around the leading SNP. We do not include the HLA and APOE loci due to long-range LD in these regions and extreme values of the summary statistics, which cause numerous false findings when fine-mapping these regions. We use the LD matrix from the UKBB provided by PolyFun for all three models. For each model, we consider two scenarios: (1) no functional annotation, and (2) including functional annotations. SuSiE can only include one annotation, therefore we use prior causal probabilities made available by PolyFun based on a meta-analysis of several UKBB traits.[10] For the CARMA model, we include 924 functional annotations including DeepSEA,[19] CADD,[21] PO-EN,[22] and PolyFun.[10] For fastPAINTOR, we adopt the strategy used in simulations, namely we pre-select 10 functional annotations (among 924) that are most correlated to the summary statistics.

We present results in terms of PIPs of SNPs at the 30 loci, impact of including functional annotations, and comparisons of credible sets. As discussed already, one challenging aspect of meta-analyses is unequal sample sizes available at different genetic variants, which can affect the accuracy of the resulting PIPs. For the

current study, the sample sizes can vary from 9,703 to 444,006 depending on which datasets are included in the meta-analyses. In addition to analyzing the complete dataset (referred to as the heterogeneous dataset), we also consider a reduced dataset that focuses only on SNPs included in at least IGAP, PGC-ALZ, and the large AD-by-proxy study analyses (referred to as the homogeneous dataset); the sample sizes for the SNPs included in this homogeneous set vary between 418,339 and 444,006. The homogeneous dataset effectively contains a much smaller number of SNPs relative to the heterogeneous dataset, and serves as a comparison for the more realistic, heterogeneous dataset.

Fine-mapping results at 30 GWAS loci. We show fine-mapping results for individual loci using the three models in Figures S9-S38. The SNPs with largest PIP are reported in Tables S4 and S5. There is a rather large overlap between CARMA and SuSiE for the homogenous dataset with no functional information (21/30, Figure 6). Due to differences in how functional annotations are integrated into the different models, the concordance for the top SNPs among the three models decreases when we include functional annotations. Furthermore, for the heterogeneous dataset the overlap between CARMA and SuSiE diminishes with SuSiE showing many more SNPs with PIPs equal to 1 (Figures 6-8). For example, at loci *CR1*, *BIN1*, *CLNK*, *HS3ST1*, *CD2AP*, *ZCWPW1*, *MS4A6A*, *PICALM*, *ADAM10*, *SCIMP*, *ABI3*, and *ABCA7*, SuSiE reports multiple SNPs with PIPs equal or close to 1 due to biased marginal likelihood caused by discrepancies between Z-scores and LD values. One typical example is *CD2AP* (CD2 associated protein), where CARMA identifies a top SNP rs9381563 (eQTL for *CD2AP* in brain spinal cord and a few other tissues in GTEx) with PIP$>$ 0.6, while SuSiE seems to identify several SNPs with PIP=1 at this locus (Figure 7). Furthermore, CARMA identifies the same SNP across the different analyses (homogeneous vs. heterogeneous dataset, and with or without functional annotation, Figure S17). There is accumulating evidence that CD2AP is implicated in AD pathogenesis. In particular, CD2AP loss of function is linked to enhanced Aβ production, Tau-induced neurotoxicity, abnormal neurite structure modulation and reduced blood-brain barrier integrity.[23]

For 15/30 loci the SNP with the largest PIP is shared across the different analyses in CARMA (homogeneous/heterogeneous, with or without functional annotation). Given that the homogenous dataset contains only a subset of the complete set of SNPs, such SNPs with robust evidence may be interesting for follow-up investigations.

Credible sets. For the homogeneous dataset, results for CARMA and SuSiE are largely similar to each other, except for smaller credible sets for CARMA (Figure 8a). fastPAINTOR has smaller average $|Z|$ and larger variance for variants in a credible set, as well as higher number of credible sets. For the heterogeneous dataset, SuSiE and fastPAINTOR show marked decreases in average $|Z|$ for variants in a credible set (Figure 8b), suggesting that these two models may include in credible sets variants with smaller Z-scores but highly correlated (based on external LD) with SNPs with larger Z-scores (see also Figure 1). Also, the average number of credible sets with one SNP in the SuSiE model increases significantly as shown in Figure 8b, which suggests that SuSiE is affected by outliers in this heterogeneous dataset. These results together show that discrepancies between Z-scores and LD values in the heterogeneous dataset lead to biased PIP values for SuSiE and fastPAINTOR; the average size of the credible sets for both SuSiE and fastPAINTOR also decreases substantially in the heterogeneous dataset as a consequence. The credible sets generated by CARMA also have smaller sizes, for a different reason. The strong penalization $\eta = 1/\sqrt{p}$ assigned to limit false positives in the heterogeneous dataset also decreases the PIPs of SNPs with weak or medium Z-scores. Therefore, the credible sets tend to contain stronger signals as illustrated by the increased mean $|Z|$ (Figure 8b).

Credible models. The overall number of SNPs included in the credible models is smaller than for the credible set (Figure S39), consistent with the simulation results. When the PIPs are not large enough to formulate a credible set, the credible models can still provide a set of leading SNPs accounting for LD, which is also the reason why the average of the absolute values of the summary statistics $|Z|$ of the SNPs selected by the credible models is smaller than for the credible sets as shown in Figure S39. When functional annotations are included, the corresponding posterior probability of the top candidate model may increase, therefore the number of credible models and the number of SNPs included in the credible models may decrease (Table S8). Notice

that the credible models are also less affected by potential outliers in the heterogeneous datasets, similar to the credible sets of the CARMA model. More results can be found in Table S8.

Discussion

We proposed here a novel Bayesian fine-mapping method, CARMA, which is designed to prioritize potentially causal variants within GWAS risk loci by leveraging the LD structure and functional annotations available for variants at the locus under investigation. Different from existing methods, CARMA assumes a heavy-tail Cauchy prior distribution on effect sizes, and jointly maximizes the likelihood of summary statistics and functional annotations in a unified EM algorithm with multiplicity control. Importantly, we introduce a novel Bayesian hypothesis testing approach to account for mismatches between summary statistics and LD values from external reference panels in order to avoid prioritization of non-causal variants. Through extensive simulations, we demonstrate that CARMA has higher accuracy to predict causal variants relative to SuSiE and fastPAINTOR, especially when functional annotations are included and in complex meta-analyses settings with discrepancies between summary statistics and LD values estimated from reference panels.

We further illustrate the use of CARMA in a large fine-mapping analysis of 30 GWAS loci for Alzheimer's disease identified in.[1] The results show that CARMA has the ability to handle possible discrepancies between the LD matrix from the UKBB and the summary statistics in the meta-analysis, while the results from SuSiE and fastPAINTOR can be greatly affected in such scenarios. In particular, we highlight a higher confidence list of 15 out of 30 loci where CARMA consistently identifies the same top SNP across the different analyses (homogeneous or heterogeneous, with or without functional annotations).

Very recently, a new version of SuSiE has been released that is applicable to summary statistics and LD matrix extracted from reference panels.[24] In this context, the authors develop a likelihood ratio test for identifying a particular type of inconsistency, namely the "allele flip" scenario. Note that our proposed outlier detection method is more general and deals with broad types of inconsistencies, including those generated by different sample size coverage at different SNPs in the data. We also note that the diagnostic procedure in SuSiE is a separate step and not integrated into the main algorithm, and requires the inversion of the entire LD matrix for identifying one outlier, which is computationally intensive. In our applications to the motivating example in Figure 1, this new implementation has failed to solve the issue of biased PIPs (more details are in the Supplemental material).

CARMA has been implemented in a computationally efficient R package.

Methods

Basic notations and assumptions

Assume that we have genotype ($\boldsymbol{X}$) and phenotype ($\boldsymbol{y}$) data for n subjects. For a given locus, we assume a standard linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_y^2 I_n)$$

where $\boldsymbol{y}$ is a standardized $n \times 1$ vector of quantitative phenotype values, $\boldsymbol{X}$ is a standardized $n \times p$ genotype matrix, $\boldsymbol{\beta}$ is a p-dimensional vector of effect sizes of SNPs, and $\boldsymbol{\epsilon}$ is a Gaussian noise vector. GWAS are usually performed in a univariate fashion, so that for each variant i with $i = 1, \ldots, p$, we obtain the estimated marginal effect and standard error:

$$\hat{\beta}_i = \frac{\boldsymbol{x}_i'\boldsymbol{y}}{\boldsymbol{x}_i'\boldsymbol{x}_i},\quad \text{se}\left(\hat{\beta}_i\right) = \sqrt{\frac{s_i^2}{\boldsymbol{x}_i'\boldsymbol{x}_i}},$$

where $s_i^2 = \frac{(\boldsymbol{y}-\boldsymbol{x}_i\hat{\beta}_i)'(\boldsymbol{y}-\boldsymbol{x}_i\hat{\beta}_i)}{n-2}$. Then the Z-score (Wald statistic) is defined as $Z_i = \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)}$ with $Z_i \sim N(0,1)$ under the null hypothesis $H_0: \beta_i = 0$. Asymptotically, it can be shown that

$$\boldsymbol{Z} \asymp \frac{\boldsymbol{X'y}}{O(\sqrt{n})} = \frac{\boldsymbol{X'X\beta} + \boldsymbol{X'\epsilon}}{O(\sqrt{n})}.$$

Details are in the Supplemental Material (section 'Summary statistics'). Note that the matrix $\boldsymbol{X'X}$ is the LD correlation matrix of SNPs up to a constant associated with sample size n, and can be approximated using the appropriate, population-matched reference panel. Assuming that $\boldsymbol{\Sigma} = \frac{\boldsymbol{X'X}}{n}$ and $\boldsymbol{\lambda} \propto \boldsymbol{\beta}$ (i.e., $\lambda_i \neq 0$ if and only if $\beta_i \neq 0$), the sampling distribution of $\boldsymbol{Z}$ can be represented as

$$\boldsymbol{Z}|\boldsymbol{\lambda}, \sigma_y^2, \boldsymbol{\Sigma} \sim \text{MVN}(\boldsymbol{\Sigma\lambda}, \sigma_y^2\boldsymbol{\Sigma}).$$

Indicator vector $\boldsymbol{\gamma}$ and true model. Let $\boldsymbol{\gamma}' = \{0,1\}^p$ denote an indicator vector, such that $\lambda_i \neq 0$ iff $\gamma_i = 1$. Also, let $S = \{i; \lambda_i \neq 0\}$ denote the index set of causal SNPs assumed by a particular model, such that if $i \in S$, then $\gamma_i = 1$ and $\lambda_i \neq 0$. Then, each indicator vector $\boldsymbol{\gamma}_S$ uniquely defines a model with dimension equal to $|S| = \sum_{i=1}^p \gamma_i$. For a particular $\boldsymbol{\gamma}_S$, we denote by $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}$ the corresponding coefficient vector, such that $\lambda_i \neq 0$ if $i \in S$.

Review of existing fine-mapping models

We briefly review several representative fine-mapping models, including JAM, fastPAINTOR and SuSiE.

JAM.[7] Building upon prior Bayesian fine-mapping methods such as CAVIARBF[11] and FINEMAP,[8] JAM assumes the popular g-prior[25] to model the effect sizes:

$$\begin{aligned}\boldsymbol{\lambda}_\gamma|\sigma_y^2 &\sim \text{MVN}\left(0, g\sigma_y^2\boldsymbol{\Sigma}_\gamma\right),\\ \sigma_y^2 &\sim \text{Inv-Gamma}\left(0.01, 0.01\right),\end{aligned}$$

where g is a pre-determined constant. JAM assigns a beta-binomial distribution as the prior distribution of the model size, i.e., the number of the assumed causal SNPs in a model:

$$\begin{aligned}\Pr\left(\{\boldsymbol{\gamma}; \mathbf{1}\cdot\boldsymbol{\gamma} = s\}\right) &\sim \text{Beta-binomial}(s+a, p-s+b),\\ \Pr\left(\boldsymbol{\gamma}\right) &= \frac{\Pr\left(\{\boldsymbol{\gamma}; \mathbf{1}\cdot\boldsymbol{\gamma} = s\}\right)}{\binom{p}{s}},\end{aligned}$$

where $\{\boldsymbol{\gamma}; \mathbf{1}\cdot\boldsymbol{\gamma} = s\}$ is the set of models having the same dimension as s, and all models of the same size are equally likely. $a = 1$ and $b = 9$ are used in applications, corresponding to a mean proportion of truly casual SNPs of 10%.

Relative to previous fine-mapping models, JAM does not restrict the number of causal SNPs in the region and provides a computationally efficient method based on Reversible Jump MCMC to explore a wide range of candidate models. Also, unlike previous fine-mapping methods which assume that σ_y^2 is a plug-in estimate, JAM marginalizes out both $\boldsymbol{\lambda}$ and σ_y^2 to improve the power. However, it has been shown in[14] that the g-prior leads to issues with model selection consistency. Another potential issue is that the correlation matrix $\boldsymbol{\Sigma}$ is not necessarily positive definite, and JAM cannot work properly in such cases.

fastPAINTOR[4]**.** fastPAINTOR uses as input Z-scores for individual variants, $Z_i = \frac{\hat{\beta}_i}{\text{se}(\hat{\beta}_i)}$, and the LD matrix $\boldsymbol{\Sigma}$, usually estimated from a reference panel. fastPAINTOR assumes the following sampling distribution of $\boldsymbol{Z}$:

$$\boldsymbol{Z}|\boldsymbol{\lambda}_\gamma, \boldsymbol{\Sigma} \sim \text{MVN}(\boldsymbol{\Sigma\lambda}_\gamma, \boldsymbol{\Sigma}),$$

for a given model $\boldsymbol{\gamma}$. The assumed prior distribution of $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ can be written as

$$\begin{aligned}\boldsymbol{\lambda}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \sigma^2_{\boldsymbol{\lambda}} &\sim \text{MVN}(0, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}),\\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} &= \sigma^2_{\boldsymbol{\lambda}}\text{Diag}(\boldsymbol{\gamma}) + \text{Diag}(\sigma^2_y).\end{aligned}$$

The distribution of $\boldsymbol{Z}$ after integrating out $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$ is

$$\boldsymbol{Z}|\boldsymbol{\gamma}, \boldsymbol{\Sigma} \sim \text{MVN}(0, \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma})P(\boldsymbol{\gamma}).$$

The prior distribution of γ_i can be Uniform, or a Bernoulli distribution, where the probability of being causal can be related to functional annotations as follows:

$$\gamma_i \sim \text{Bern}(\frac{\exp\{\boldsymbol{w}_i'\boldsymbol{\theta}\}}{1+\exp\{\boldsymbol{w}_i'\boldsymbol{\theta}\}}),$$

where $\boldsymbol{w}_i$ is the vector of functional annotations for SNP i. Posterior probabilities for each SNP are calculated using an importance sampling algorithm. Note that fastPAINTOR cannot handle large number of functional annotations, and a pre-selection of a small number of annotations may be necessary.[10]

SuSiE.[3] The SuSiE model works particularly well in situations where variables are highly correlated and the effects are sparse. It is conceptually different from the other fine-mapping models. The inference in the SuSiE model is based on multiple basic models, the so-called Single-Effect Regression (SER) models. Under the assumption of no prior information, the SER model is defined as:

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \text{MVN}(\boldsymbol{0}, \sigma^2_y I_n)\\ \boldsymbol{\beta} &= b\boldsymbol{\gamma}\\ b &\sim N(0, \sigma^2_0)\\ \boldsymbol{\gamma} &\sim \text{Mult}(1, \boldsymbol{\pi}),\end{aligned}$$

where $\text{Mult}(m, \boldsymbol{\pi})$ denotes the multinomial distribution on class counts that is obtained when m samples are drawn with class probability $\boldsymbol{\pi}$. For the SER model, there is only one non-zero effect, and hence the indicator vector $\boldsymbol{\gamma}$ has only one non-zero element. SuSiE assumes by default that $\boldsymbol{\pi} = (1/p, \ldots, 1/p)$, but one could define $\boldsymbol{\pi}$ based on prior probabilities derived from functional annotations. Calculating the posterior inclusion probabilitles (PIP) $\Pr\left(\gamma_i = 1|\boldsymbol{y}, \boldsymbol{X}, \sigma^2_y, \sigma^2_0\right)$ involves fitting p univariate regression of $\boldsymbol{y}$ on the columns of $\boldsymbol{x}_i$ of $\boldsymbol{X}$.

The SER model assumes only one causal SNP. For extensions to multiple causal variants within a locus, the final SuSiE model is built based on L SER models. The idea is to introduce multiple single-effect vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L$ and construct the overall effect vector $\boldsymbol{\beta}$ as the sum of these single effects. The model is as follows:

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim \text{MVN}(\boldsymbol{0}, \sigma^2_y I_n)\\ \boldsymbol{\beta} &= \sum_{l=1}^{L}\boldsymbol{\beta}_l\\ \boldsymbol{\beta}_l &= \boldsymbol{\gamma}_l b_l\\ b_l &\sim N(0, \sigma^2_0)\\ \boldsymbol{\gamma}_l &\sim \text{Mult}(1, \boldsymbol{\pi}).\end{aligned}$$

SuSiE performs an iterative Bayesian stepwise selection (IBSS) algorithm to fit this model; at each iteration it uses the SER model to estimate $\boldsymbol{\beta}_l$ given current estimate of $\boldsymbol{\beta}_{l'}$ for $l' \neq l$. Specifically, it fits the SER model for $\boldsymbol{\beta}_l$ using the residual $\bar{r} \leftarrow \boldsymbol{y} - \boldsymbol{X}\sum_{l'\neq l}\boldsymbol{\beta}_{l'}$. The result of the SuSiE model consists of L fitted $\hat{\boldsymbol{\beta}}_l$, and L corresponding PIP vectors $\boldsymbol{\alpha}_l = \left\{\Pr\left(\boldsymbol{\beta}_{l,1} \neq 0|\boldsymbol{X}, \boldsymbol{y}\right), \ldots, \Pr\left(\boldsymbol{\beta}_{l,p} \neq 0|\boldsymbol{X}, \boldsymbol{y}\right)\right\}'$. Then the final PIP is defined as

$$\Pr\left(\hat{\beta}_j \neq 0|\boldsymbol{X}, \boldsymbol{y}\right) \approx 1 - \prod_{l\in\{1,\ldots,L\}}(1-\alpha_{l,j}),$$

assuming that the $\beta_{l,j}$ are independent across $l = 1 \dots L$. SuSiE naturally produces credible sets; a level ρ credible set is defined as a subset of variables that has probability ρ or more to contain at least one causal variable.

SuSiE+PolyFun.[10] To fully utilize the potential of the fine-mapping methods that can only take as input univariate prior causal probability, Weissbrod et al.[10] proposed to estimate prior probabilities through the S-LDSC model,[26] which essentially estimates global, genome-wide functional genomic enrichments using only summary data. Specifically, PolyFun estimates the prior probability for each SNP in proportion to per-SNP heritability estimates: $\Pr(\beta_i \neq 0 | \boldsymbol{w}_i) \propto \hat{\text{var}}[\beta_i | \boldsymbol{w}_i] = \hat{\boldsymbol{\theta}}' \boldsymbol{w}_i$ where

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta} \in R^q}{\text{argmin}} \sum_i \left[\chi_i^2 - n \sum_{j=1}^{q} \theta_j l(i,j) - nb - 1 \right]^2 + \alpha ||\boldsymbol{\theta}||^2;$$

q is the number of functional annotations, $\chi_i^2 = \frac{(\boldsymbol{x}_i' \boldsymbol{y})^2}{n}$ is the χ^2 statistic of SNP i, $l(i,j) = \sum_{k=1}^{p} \text{cor}(\boldsymbol{x}_i, \boldsymbol{x}_k) w_{j,k}$ is the LD-score for SNP i weighted by functional annotation j, and b measures the contribution of confounding biases.

In applications, SuSiE+PolyFun was shown to lead to more causal variant discoveries relative to SuSiE without functional annotations. One advantage of PolyFun is that the contribution of a functional annotation to the heritability is estimated using genome-wide SNPs. However, one implicit assumption is that the effect of a functional annotation is consistent across loci which might not be true. Another possible concern is the overfitting potential due to the use of summary statistics for both the PolyFun and SuSiE components of the final model.

We summarize the main features and assumptions of commonly used fine-mapping methods in the literature in Table 2.

Proposed Bayesian fine-mapping model

We introduce here the details of our model; we refer to it as CARMA (CAusal Robust Mapping method with Annotations). Given any model $\boldsymbol{\gamma}$, we assume that the summary statistics $\boldsymbol{Z}$ follow a multivariate normal distribution:

$$\boldsymbol{Z} | \boldsymbol{\lambda}_{\boldsymbol{\gamma}}, \sigma_y^2, \boldsymbol{\Sigma} \sim \text{MVN}(\boldsymbol{\Sigma} \boldsymbol{\lambda}_{\boldsymbol{\gamma}}, \sigma_y^2 \boldsymbol{\Sigma}).$$

We want to identify the true model (denoted by the true indicator vector $\boldsymbol{\gamma}_T$) that generated the summary statistics through posterior inference within a Bayesian paradigm.

Heavy tail prior distribution on effect sizes. We assume a slab-spike prior for the prior distribution of coefficient λ_i, i.e.,

$$\begin{aligned} \lambda_i &\sim \text{Mixture of Normal distributions if } \gamma_i = 1, \\ \lambda_i &= 0 \text{ if } \gamma_i = 0. \end{aligned}$$

Specifically, the prior distribution of the assumed non-zero effect size $\boldsymbol{\lambda}_S$ that is associated with $\boldsymbol{\gamma}_S$ is

$$\begin{aligned} \boldsymbol{\lambda}_S | \sigma_y^2, \tau &\sim \text{MVN}(0, \frac{\sigma_y^2}{\tau} \boldsymbol{\Sigma}_{S,S}^{-1}), \\ \tau &\sim f(\tau), \ \tau \in (0, \infty), \end{aligned}$$

where $f(\tau)$ denotes the prior distribution on τ. For the prior distribution of σ_y^2, we assume a non-informative prior distribution, such that

$$\sigma_y^2 \sim \frac{1}{\sigma_y^2}.$$

We also include an alternative setting when σ_y^2 is fixed. By assigning a prior distribution on the mixing parameter τ, we intrinsically assign a mixture of normal distribution on the effect size $\boldsymbol{\lambda}_S$. The choice of $f(\tau)$ directly impacts the computation of posterior probabilities for candidate models. Here, we explore two specific choices of $f(\tau)$: Zellner-Siow's Cauchy prior[27] and the hyper-g prior,[14] as follows:

$$\begin{aligned} \text{Zellner-Siow's Cauchy:} \quad \tau &\sim \text{Gamma}(\tfrac{1}{2}, \tfrac{1}{2}), \\ \text{hyper-}g\text{:} \quad \tau &\sim \text{beta-prime}(\tfrac{1}{2}, 1). \end{aligned}$$

Under both priors, integrating out τ yields heavy-tail prior distributions on the effect sizes $\boldsymbol{\lambda}_S$; in fact, the marginal prior distribution of $\boldsymbol{\lambda}_S$ under Zellner-Siow's prior is a multivariate Cauchy distribution. Replacing the conventional Gaussian prior on SNP effect sizes with a heavy-tail distribution to increase association power has been used before, such as in BOLT-LMM.[13]

The advantages of assigning a mixture of normal prior distribution over normal prior in model selection are well documented in the Bayesian literature.[14,27,28] Specifically, in[14,28] the authors show that, under mild conditions, a variety of mixtures of normal distributions can achieve model selection consistency $\lim_n \Pr(M_T|\boldsymbol{Z}) = 1$, whereas the normal prior suffers two well-known model selection paradoxes, Bartlett's Paradox[29] and Information Paradox,[25] which prevents the posterior probability of the true model from converging to 1. As shown in the Supplement, given a Z score with a relatively large value, the logarithm of the Bayes factor between the model with the mixture of normal prior and the model with normal prior is asymptotically proportional to Z^2, which is overwhelmingly in favor of the model with a mixture of normal prior and would result in a stronger PIP to support the causality of the corresponding SNP.

Prior distribution on model space. A fine-mapping locus typically contains a large number of variants which can be a challenge for model selection. Without an appropriate prior distribution, all candidate models including the finite true model receive very small posterior probabilities, as also noted in.[5] A dimensional penalization on the model space, through the prior distribution, is required for model selection consistency in such high-dimensional regime as shown in.[30,31]

We introduce a prior distribution on model space to control the total number of causal SNPs that any candidate model assumes, which is analogous to controlling the false discovery rate. Let $|S| = \sum_{\gamma_i \in \boldsymbol{\gamma}_S} \gamma_i$ denote the total number of causal SNPs for a given $\boldsymbol{\gamma}_S$. We first place a discrete prior distribution on the random variable $|S|$ and let $\Pr(|S||\eta)$ denote the p.m.f with η as a hyperparameter. Given $|S|$, we assume that all those models have the same prior probability. Hence, the prior probability of $\boldsymbol{\gamma}_S$ is

$$\Pr(\boldsymbol{\gamma}_S|\eta) = \frac{\Pr(|S||\eta)}{\binom{p}{|S|}}.$$

We propose to use the truncated-Poisson prior ($|S| \in \{0, \ldots, p\}$), which has been introduced before and shown to enjoy model selection consistency in.[31] Let $F(\cdot|\eta)$ be the cumulative distribution function of a Poisson distribution with mean η, the truncated-Poisson prior is defined as follows:

$$\Pr(|S||\eta) = \frac{\eta^{|S|}\exp\{-\eta\}}{|S|!F(p|\eta)} \propto \frac{\eta^{|S|}\exp\{-\eta\}}{|S|!}.$$

Since the total number of SNPs p is usually a large number and η is chosen to be small in order to reflect the sparse scenario for the true causal SNPs, $F(p|\eta) \to 1$. Then given any specific model $\boldsymbol{\gamma}_S$, the prior probability of this model under Poisson distribution is

$$\Pr(\boldsymbol{\gamma}_S|\eta) \propto \frac{\eta^{|S|}\exp\{-\eta\}(p-|S|)!}{p!}.$$

The Poisson distribution on model space provides necessary dimension penalization for multiplicity control. The hyper-parameter η plays a critical role in this dimension penalization mechanism. Note that the computation of the PIPs is based on the unnormalized posterior probabilities for the candidate models, i.e.,

$\Pr(\boldsymbol{\gamma}_S|\boldsymbol{Z}) \propto f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)\Pr(\boldsymbol{\gamma}_S|\eta)$. Therefore, the choice of η depends on the value of the marginal likelihood, i.e, $f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)$. When σ_y^2 is integrated out to yield a strong marginal likelihood, we recommend to set $\eta \propto p^{-a}$, for $a > 0$ to provide a strong dimensional penalization for controlling FDR.[32] On the other hand, if there are potential outliers and σ_y^2 is set equal to 1 (assuming $\boldsymbol{Z}$ is standardized) yielding a moderate marginal likelihood, we recommend to set η to be an integer to encourage a more aggressive exploration of the posterior model space by the algorithm. One caveat is that when the marginal likelihood is severely biased by outliers, such as the heterogeneous dataset of the GWAS meta-analysis for Alzheimer's disease, then we set $\eta \propto p^{-a}$ to account for the artificially increased marginal likelihood. Accuracy measures such as AUROC and AUPR that are rank-based are not sensitive to the choice of η as long as the value of η is within same magnitude; however the magnitude of the PIPs and the number of credible sets is affected by this choice (see Supplement section 'Hyper-parameter η of the Poisson prior distribution' for more details).

Marginal likelihood and posterior probability. Given the Z-scores $\boldsymbol{Z}$ and LD correlation matrix $\boldsymbol{\Sigma}$, the marginal likelihood conditional on $\boldsymbol{\gamma}_S$ is

$$f(\boldsymbol{Z}|\boldsymbol{\gamma}_S) = \int_{\tau}\int_{\sigma_y^2}\int_{\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}} f(\boldsymbol{Z}|\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}, \sigma_y^2) f(\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}|\boldsymbol{\gamma}_S, \tau, \sigma_y^2) f(\tau) f(\sigma_y^2) \mathrm{d}\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S} \mathrm{d}\sigma_y^2 \mathrm{d}\tau,$$

where $f(\boldsymbol{Z}|\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}, \sigma_y^2)$ is the density function of MVN($\boldsymbol{\Sigma}\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}, \sigma_y^2\boldsymbol{\Sigma}$), and $f(\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}|\boldsymbol{\gamma}_S, \tau, \sigma_y^2)$ is the density function of MVN($0, \frac{\sigma_y^2}{\tau}\boldsymbol{\Sigma}_{\gamma}^{-1}$). The marginal likelihood after integrating out $\boldsymbol{\lambda}_{\boldsymbol{\gamma}_S}$ and σ_y^2 is

$$f(\boldsymbol{Z}|\boldsymbol{\gamma}_S) = \int_0^{\infty} \frac{\Gamma(\frac{p}{2})|\boldsymbol{\Sigma}|^{1/2}}{\left[\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} - \boldsymbol{Z}_S'\frac{\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_S}^{-1}}{1+\tau}\boldsymbol{Z}_S\right]^{\frac{p}{2}}} (\pi)^{-\frac{p}{2}} \left(\frac{1+\tau}{\tau}\right)^{-\frac{|S|}{2}} f(\tau)\mathrm{d}\tau,$$

and the ratio between the marginal likelihood of $\boldsymbol{\gamma}_S$ and the null model $\boldsymbol{\gamma}_0$ is

$$\frac{f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)}{f(\boldsymbol{Z}|\boldsymbol{\gamma}_0)} = \int_0^{\infty} \left[1 - \frac{\boldsymbol{Z}_S'\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_S}^{-1}\boldsymbol{Z}_S}{\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\,(1+\tau)}\right]^{-\frac{p}{2}} \left(\frac{1+\tau}{\tau}\right)^{-\frac{|S|}{2}} f(\tau)\mathrm{d}\tau.$$

The details can be found in the Supplement section 'Marginalization of likelihood with respect to Cauchy' . Let $\mathcal{M}$ denote the model set that contains all candidate models. Then the posterior probability of any non-null model $\boldsymbol{\gamma}_S$ and the posterior probability of γ_i being equal to 1 (PIP) can be computed as

$$\Pr(\boldsymbol{\gamma}_S|\boldsymbol{Z}) = \frac{PO_{\boldsymbol{\gamma}_S:\boldsymbol{\gamma}_0}}{\sum_{\boldsymbol{\gamma}_A\in\mathcal{M}} PO_{\boldsymbol{\gamma}_A:\boldsymbol{\gamma}_0}},$$
$$\Pr(\gamma_i = 1|\boldsymbol{Z}) = \sum_{\boldsymbol{\gamma}_S:i\in S} \Pr(\boldsymbol{\gamma}_S|\boldsymbol{Z}),$$

where the posterior odds ($PO_{\boldsymbol{\gamma}_S:\boldsymbol{\gamma}_0}$) is defined as the product of the Bayes factor $\left(\frac{f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)}{f(\boldsymbol{Z}|\boldsymbol{\gamma}_0)}\right)$ and the prior odds $\left(\frac{\Pr(\boldsymbol{\gamma}_S|\eta)}{\Pr(\boldsymbol{\gamma}_0|\eta)}\right)$:

$$\begin{aligned} PO_{\boldsymbol{\gamma}_S:\boldsymbol{\gamma}_0} &= \frac{f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)}{f(\boldsymbol{Z}|\boldsymbol{\gamma}_0)} \frac{\Pr(\boldsymbol{\gamma}_S|\eta)}{\Pr(\boldsymbol{\gamma}_0|\eta)} \\ &= \frac{\eta^{|S|}(p-|S|)!}{p!} \int_0^{\infty} \left[1 - \frac{\boldsymbol{Z}_S'\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_S}^{-1}\boldsymbol{Z}_S}{\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\,(1+\tau)}\right]^{-\frac{p}{2}} \left(\frac{1+\tau}{\tau}\right)^{-\frac{|S|}{2}} f(\tau)\mathrm{d}\tau. \end{aligned}$$

Shotgun stochastic search algorithm. Most fine mapping methods adopt MCMC-based algorithms to explore the posterior model space. However such algorithms can be ineffective in high dimension situations because of slow convergence and inefficient proposal distribution.[15] Furthermore, due to high correlations

among SNPs at a locus, MCMC algorithms may not visit perfectly correlated SNPs evenly, which leads to unequal PIPs for such SNPs. Therefore, we adopt here the Shotgun stochastic search (Shotgun) algorithm[15] for exploring the posterior distribution over model space. The main benefit of using the Shotgun algorithm is that it not only records the visiting candidate models while running, but also the neighborhood of the visiting candidate models, hence the Shotgun algorithm has a more complete exploration of the areas with high marginal likelihoods in the posterior model space. This semi-exhaustive search feature alleviates the problem of unequal PIPs for perfectly correlated SNPs. This feature will also play an important role when integrating functional annotations into the computation of the proposed model. We note that running the Shotgun algorithm in this setting is feasible due to the strong dimensional penalization introduced by the prior Poisson distribution that constrains the search in the posterior model space.

Note that for a specific model denoted by an index set S, the unnormalized posterior distribution of the indicator vector $\boldsymbol{\gamma}_S$ is proportional to a product of the marginal likelihood and the prior distribution:

$$\Pr(\boldsymbol{\gamma}_S|\boldsymbol{Z}) = \frac{f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)f(\boldsymbol{\gamma}_S)}{f(\boldsymbol{Z})} \propto f(\boldsymbol{Z}|\boldsymbol{\gamma}_S)f(\boldsymbol{\gamma}_S).$$

The Shotgun algorithm is an iterative procedure that exhaustively examines the neighborhood of the current model, defined as:

$$\begin{aligned}
&\Gamma_-(S) := \{A : A \subset S, |S| - |A| = 1\} \text{ (one less SNP than } S\text{)},\\
&\Gamma_+(S) := \{A : A \supset A, |A| - |S| = 1\} \text{ (one more SNP than } S\text{)},\\
&\Gamma_\Leftrightarrow(S) := \{A : |S| - |A \cap S| = 1, |A| = |S|\} \text{ (models that replaces one SNP in } S\text{)}.
\end{aligned}$$

Then, all the unnormalized posterior probabilities of the neighborhood models, i.e., $\{\Gamma_-(S) \cup \Gamma_+(S) \cup \Gamma_\Leftrightarrow(S)\}$, will be computed. To update the current model, the algorithm first randomly selects one candidate model from each neighborhood set according to the unnormalized posterior probabilities, then randomly selects the next current model from the three selected models according to the corresponding posterior probabilities. By doing so, the algorithm stochastically moves towards the high posterior area in the model space. The Shotgun algorithm is stopped when the sum of the absolute difference of the PIPs between iterations is smaller than a pre-determined threshold, and we keep the top B visited candidate models with the largest unnormalized posterior probabilities. We denote by Γ the set of models selected during the algorithm.

Remarks

The inversion of Σ. Due to high correlations among SNPs at a locus, the LD matrix is not full rank most of the time. We use the Moore–Penrose inverse of the LD matrix instead; note that the corresponding consistency theorem regarding Bayesian model selection of the mixture of g-prior and the Moore–Penrose inverse of the correlation matrix has been established by Maruyama and George in.[33]

Fixing σ_y^2. There are situations where the LD correlation matrix $\boldsymbol{\Sigma}$ estimated from the reference panel might not be positive definite, i.e., $\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} < 0$, which leads to an invalid density function $f(\boldsymbol{Z}|\tau, \boldsymbol{\gamma})$. To accommodate such a situation, one solution is to fix σ_y^2 at a pre-determined value, such as 1, which should provide similar performance to the scenario of integrating out σ^2 (see Supplement section on 'Fixing σ^2 at 1' for more details and Figures S1 and S2 and table S1).

Posterior probability of candidate models. Most of the fine-mapping methods that have been proposed focus on the goal of variable selection, accounting for LD. One advantage of our proposed model is that in addition to PIPs, we provide the posterior probability of candidate models which can be informative about the putative causal variants at the test locus.

Specifically, the CARMA model returns a vast library of visited candidate models that can be ranked according to the corresponding marginal likelihoods. Let $\boldsymbol{\gamma}_{(b)}$, $b = 1, \ldots, B$, denote the ranked candidate models, such as $\boldsymbol{\gamma}_{(1)}$ receives the largest marginal likelihood. We use $\boldsymbol{\gamma}_{(1)}$ as the reference model to select

all other candidate models that are not significantly different from $\boldsymbol{\gamma}_{(1)}$ through a Bayesian hypothesis testing procedure as follows. The posterior odds of the hypothesis test for comparing $\boldsymbol{\gamma}_{(1)}$ to all other candidate models is

$$\mathrm{PO}_{\boldsymbol{\gamma}_{(1)}:\boldsymbol{\gamma}_{(b)}} = \frac{\Pr\left(\boldsymbol{\gamma}_{(1)}|\boldsymbol{Z},\eta\right)}{\Pr\left(\boldsymbol{\gamma}_{(b)}|\boldsymbol{Z},\eta\right)} = \frac{\Pr\left(\boldsymbol{Z}|\boldsymbol{\gamma}_{(1)}\right)\Pr\left(\boldsymbol{\gamma}_{(1)}|\eta\right)}{\Pr\left(\boldsymbol{Z}|\boldsymbol{\gamma}_{(b)}\right)\Pr\left(\boldsymbol{\gamma}_{(b)}|\eta\right)},$$

for $b = 1, \ldots, B$. The posterior odds quantifies the strength of evidence in favor of the leading causal configuration, represented by $\boldsymbol{\gamma}_{(1)}$ relative to other candidate models. If all candidate models have the same prior, then the posterior odds is equal to the Bayes factor. We use commonly accepted thresholds for $\mathrm{BF}_{\boldsymbol{\gamma}_{(1)}:\boldsymbol{\gamma}_{(b)}}$ as the threshold for $PO_{\boldsymbol{\gamma}_{(1)}:\boldsymbol{\gamma}_{(b)}}$ to determine if a candidate model $\gamma_{(b)}$ is significantly different from the reference model $\gamma_{(1)}$ (see Table S9 and,[20] where the authors assume equal prior). We only keep those models with the corresponding Bayes factor smaller than the threshold (e.g. 3.2 or 10), and refer to these selected models as the credible models.

Incorporating functional annotations. Leveraging functional annotation data on genetic variants at a locus of interest may improve the results of fine-mapping studies.[4,10] Most fine-mapping methods allow specification of only one functional annotation in the form of the prior probability for each SNP. However, it is difficult to choose one best annotation. fastPAINTOR cannot deal with high-dimensional functional annotation data and a pre-selection of a small number of functional annotations needs to be performed before running fastPAINTOR. SuSiE+PolyFun is essentially a two-step procedure: first, it estimates prior probabilities using the S-LDSC regression framework, and then those priors are provided as input to SuSiE.

By adding the functional annotations, the likelihood can be written as:

$$L(\boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{\gamma}, W) = f(\boldsymbol{Z}|\boldsymbol{\gamma})\Pr\left(\boldsymbol{\gamma}|W,\boldsymbol{\theta}\right).$$

Conventionally, the prior probability $\Pr\left(\boldsymbol{\gamma}|W,\boldsymbol{\theta}\right)$ is modeled using logistic regression. Since the indicator vector $\boldsymbol{\gamma}$ is unknown to us, an EM algorithm is typically used to replace the unknown indicator by its expectation conditional on summary statistics and functional annotations. Here we propose a new marginal likelihood of the summary statistics and the functional annotations, that uses the reconstruction of the posterior model space.

First, let Γ denote a truncated model space defined as a set of indicator vectors, such as $\Gamma' = \{\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_B\}$. We assume that the indicator vectors included in Γ represent the top B candidate models in terms of the posterior probability, i.e., $\min_{\boldsymbol{\gamma}\in\Gamma}\left\{\Pr\left(\boldsymbol{\gamma}|\boldsymbol{Z},W,\boldsymbol{\theta}\right)\right\} > \max_{\boldsymbol{\gamma}\notin\Gamma}\left\{\Pr\left(\boldsymbol{\gamma}|\boldsymbol{Z},W,\boldsymbol{\theta}\right)\right\}$. Notice that the semi-exhaustive search feature of the Shotgun algorithm together with the sparsity assumption on the number of causal SNPs at the test locus leads to a fairly complete reconstruction of the posterior model space (with plausible dimensions), therefore this property of Γ approximately holds.

Given the truncated model space Γ, let $\boldsymbol{G}' = \{G_1, \ldots, G_p\}$ denote the count vector associated with Γ, where $G_i \in \{0, 1, \ldots, B\}$ is the count of $\gamma_i = 1$ appearing in Γ. Given that B is relatively large, we assume that asymptotically $G_i \sim \mathrm{Poisson}(\exp\{\boldsymbol{w}_i'\boldsymbol{\theta}\})$. Then, the log-likelihood can be written as

$$\begin{aligned}
\ell(\boldsymbol{\theta}; \boldsymbol{Z}, \boldsymbol{G}, W) &= \log\left(f(\boldsymbol{Z}|\boldsymbol{G})f\left(\boldsymbol{G}|W,\boldsymbol{\theta}\right)\right) \\
&= \log\left(f(\boldsymbol{Z}|\boldsymbol{G})\right) + \sum_{i=1}^{p}\left[G_i\boldsymbol{w}_i'\boldsymbol{\theta} - \exp\left\{\boldsymbol{w}_i'\boldsymbol{\theta}\right\} - \log\left(G_i!\right)\right] \\
&\propto \sum_{i=1}^{p}\left[G_i\boldsymbol{w}_i'\boldsymbol{\theta} - \exp\left\{\boldsymbol{w}_i'\boldsymbol{\theta}\right\}\right]. \qquad (1)
\end{aligned}$$

The EM algorithm that maximizes the joint marginal likelihood of the summary statistics and the functional annotations is presented below.

EM algorithm. We wish to estimate $\boldsymbol{\theta}$ through maximizing the log-likelihood (Equation (1)), but we can not directly observe the count vector $\boldsymbol{G}$, therefore we also use the EM algorithm to find the MLE of $\boldsymbol{\theta}$ by treating $\boldsymbol{G}$ as missing variables. One particular difficulty of the proposed EM algorithm is that there is no analytical closed-form for $\mathbf{E}\left[G_i|\boldsymbol{Z}, \boldsymbol{w}_i, \boldsymbol{\theta}^{(s)}\right]$. Let $g_i^{(s)}$ denote the actual count of γ_i appearing in $\Gamma^{(s)}$ after running Shotgun algorithm at step (s) of the EM algorithm. Then we approximate $\mathbf{E}\left[G_i|\boldsymbol{Z}, \boldsymbol{w}_i, \boldsymbol{\theta}^{(s)}\right]$ by $g_i^{(s)}$, where the generating process of $g_i^{(s)}$ is still conditional on $\boldsymbol{Z}, \boldsymbol{w}_i$, and $\boldsymbol{\theta}^{(s)}$. The proposed EM algorithm is as follows:

Input: Summary statistics $\boldsymbol{Z}$, functional annotations W, hyperparameter η of the Poisson prior distribution, and B.
Initialization: Run Shotgun algorithm with the prior distribution Poisson(η) to generate $\Gamma^{(0)}$ and $\boldsymbol{g}^{(0)}$.

for $s = 0, 1, \ldots$ **do**

- **EM**

E-step Replace G_i by $\mathbf{E}\left[G_i|\boldsymbol{Z}, \boldsymbol{w}_i, \boldsymbol{\theta}^{(s)}\right]$, which is approximated by $g_i^{(s)}$, $i = 1, \ldots, p$.

M-step Maximize the penalized log-likelihood as,

$$\boldsymbol{\theta}^{(s+1)} := \underset{\boldsymbol{\theta} \in R^{q+1}}{\operatorname{argmax}} \sum_{i=1}^{p} \left[g_i^{(s)} \boldsymbol{w}_i' \boldsymbol{\theta} - \exp\left\{\boldsymbol{w}_i' \boldsymbol{\theta}\right\} \right] - \frac{(1-\alpha)}{2} ||\boldsymbol{\theta}||^2 - \alpha ||\boldsymbol{\theta}||.$$

Adjust the prior probability to introduce the multiplicity control (see details below),

$$\hat{\theta}_1^{(s+1)} = \log\left(\frac{\eta B^{(s)}}{\eta + p}\right). \tag{2}$$

Then, compute the prior probability of the $(s+1)$ step:

$$\hat{\Pr}\left(\gamma_i = 1 | \boldsymbol{w}_i, \boldsymbol{\theta}^{(s+1)}\right) = \frac{\exp\left\{\boldsymbol{w}_i' \boldsymbol{\theta}^{(s+1)}\right\}}{B^{(s)}}, \tag{3}$$

where $B^{(s)}$ is the minimum between B and the total number of models visited by the Shotgun algorithm in step (s).

- **Shotgun**

Initiate Shotgun algorithm with the estimated prior probability vector $\left\{\hat{\Pr}(\gamma_1), \ldots, \hat{\Pr}(\gamma_p)\right\}'$. After running Shotgun algorithm, acquire $\Gamma^{(s+1)}$ and $\boldsymbol{g}^{(s+1)}$, which depends on $\boldsymbol{Z}$, W, and $\boldsymbol{\theta}^{(s+1)}$.

end

Algorithm 1: EM algorithm with functional annotations.

Remarks

Feature selection on functional annotations. By introducing the elastic net penalty, the proposed model has the feature of performing variable selection on the potentially high-dimensional functional annotation data.

Multiplicity control. An important aspect of any Bayesian fine-mapping method is the need to introduce multiplicity control through a prior distribution on model space in order to control the false discovery rate in the context of multiple testing.[12] Notice that when the prior distribution is conditional on functional annotations, the prior distribution is no longer a Poisson distribution but a discrete distribution conditional on W and $\boldsymbol{\theta}$.

Therefore, the results of the Poisson regression might not provide adequate dimensional penalization without additional adjustments to ensure multiplicity control. Here we seek to bridge the dimensional penalization between the situation with or without functional annotations through η, where η is a generic dimensional penalization parameter in both situations.

It can be shown that when there is no prior information, the prior probability of an individual SNP being causal is

$$\Pr(\gamma_i = 1|\eta) = \frac{\eta}{\eta + p}, \text{for } \forall i \in \{1, \ldots, p\}$$

which is also the geometric mean, since the prior probability of any SNP being causal is identical. When we include functional annotations, the geometric mean of the estimated prior probability defined in Equation (3) is

$$\begin{aligned}
\left[\prod_{i=1}^{p} \hat{\Pr}\left(\gamma_i = 1|\boldsymbol{w}_i, \boldsymbol{\theta}^{(s+1)}\right)\right]^{\frac{1}{p}} &= \exp\left\{\frac{1}{p}\sum_{i=1}^{p}\log\left(\frac{\exp\left\{\boldsymbol{w}_i'\boldsymbol{\theta}^{(s+1)}\right\}}{B^{(s)}}\right)\right\} \\
&= \exp\left\{\frac{1}{p}\left[\sum_{i=1}^{p}\boldsymbol{w}_i'\boldsymbol{\theta}^{(s+1)}\right] - \log\left(B^{(s)}\right)\right\} \\
&= \exp\left\{\frac{1}{p}\left[p\theta_1^{(s+1)}\right] - \log\left(B^{(s)}\right)\right\} \\
&= \frac{\exp\left\{\theta_1^{(s+1)}\right\}}{B^{(s)}},
\end{aligned}$$

where $\theta_1^{(s+1)}$ is the intercept term (we used the fact that the design matrix for the functional annotations is standardized such as $\boldsymbol{w}_i$ has mean zero for $i = 2, \ldots, q$). To regulate the prior probability that is conditional on the annotations and introduce a similar control on multiplicity based on the Poisson distribution through η, we adjust the value of the intercept term by setting the two geometric means to be equal, such that

$$\hat{\theta}_1^{(s+1)} = \log\left(\frac{\eta B^{(s)}}{\eta + p}\right). \tag{4}$$

Therefore, when the functional annotations are included, the dimensional penalization of moving from the null model to any model with one variable is still asymptotically equal to

$$\frac{\hat{\Pr}(\gamma_S|\eta)}{\Pr(\gamma_0|\eta)} \approx \frac{\eta}{p},$$

where $|S| = 1$. Hence, the estimated prior probability of the Poisson regression also provides similar dimensional penalization to control FDR while ranking the causality of the testing SNPs based on the input functional annotations.

Choosing η. As previously mentioned, a rule of thumb to set up the value of η is $\eta \propto p^{-a}$, for $a > 0$ as discussed in.[30,32,34] We show more details on the choice of η in the Supplement section 'Hyper-parameter η of the Poisson prior distribution', especially when σ_y^2 is fixed at 1.

Identifying SNPs with Z-score/LD values mismatch

A challenge when doing fine-mapping with summary statistics and LD from reference panels is the possible mismatch between the LD values and the Z-score values. This can happen for several reasons. For example when using a reference panel to estimate the LD, or when summary statistics at neighboring SNPs are based on different studies with different sample sizes. Such scenarios can result in unrealistically large values for the marginal likelihood and high PIP values, especially given the large sample sizes for meta-analysis studies. We identify as outliers those SNPs with abnormal sampling distribution compared to all other highly correlated SNPs conditional on the summary statistics and the LD matrix in the region. We describe here a Bayesian hypothesis testing approach to identify such outliers.

The outlier detection procedure for a group of highly correlated SNPs. Let D denote the index set for a group of highly correlated SNPs, such that $\text{cor}(Z_i, Z_j) \geq \rho_{\text{outlier}}$, for $\forall i, j \in D$, where ρ_{outlier} is typically set at 0.9 or 0.99. Let $\tilde{\boldsymbol{Z}} = \{Z_1, \ldots, Z_{|D|}\}'$ denote the corresponding vector of the summary statistics. Given the high correlation for the SNPs in D, we can assume that the random variable vector $\tilde{\boldsymbol{Z}}$ follow a $\text{MVN}(\mathbf{1}_{|D|}\beta, \boldsymbol{\Sigma}_D)$, where $\mathbf{1}_{|D|}$ is the $|D|$-dimensional vector of 1, β is a scalar, and $\boldsymbol{\Sigma}_D$ is the corresponding LD matrix. Starting with the first SNP, suppose that we wish to test whether Z_1 is generated by a different distribution, i.e. $N(\beta, c)$, and the corresponding hypotheses are:

$$H_0 : c = 1;\ Z_1 \text{ is not an outlier vs.}$$
$$H_1 : c \neq 1;\ Z_1 \text{ is an outlier.}$$

Let $\tilde{\boldsymbol{Z}}_{-1} = \{Z_2, \ldots, Z_{|D|}\}'$ denote the vector of summary statistics for the SNPs in D, excluding SNP 1, and we assume that $\tilde{\boldsymbol{Z}}_{-1} = \{Z_2, \ldots, Z_{|D|}\}'$ are not outliers and follow a $\text{MVN}(\mathbf{1}_{|D|-1}\beta, \boldsymbol{\Sigma}_{D_{-1}})$ (it is possible that there are a few outliers in D_{-1}, however their influence will likely be small given the potentially large number of variants in the set).

Given an equal prior probability on the null and alternative hypothesis, the Bayes factor in favor of H_0 is

$$B_{0:1} = \frac{m^0\left(Z_1|\tilde{\boldsymbol{Z}}_{-1}, c = 1\right)}{\int m^1(Z_1|\tilde{\boldsymbol{Z}}_{-1}, c)g(c)\text{d}c},$$

where $g(c)$ is the prior distribution of the testing parameter c, and m^0 and m^1 are the posterior predictive densities at Z_1 assuming that the model generating Z_1 is the null or the alternative model, respectively (details are in section 'Outlier detection' in the Supplement). Following the robust Bayesian approach discussed in,[35] we can compute

$$\hat{B}_{0:1} = \inf_{g \in G} B_{0:1} = \frac{m^0\left(Z_1|\tilde{\boldsymbol{Z}}_{-1}, c = 1\right)}{\sup_{g \in G} \int m^1(Z_1|\tilde{\boldsymbol{Z}}_{-1}, c)g(c)\text{d}c},$$

where G represents the class of all possible prior densities. The minimization of the Bayes factor leads to the following equation,

$$\hat{B}_{0:1} = \begin{cases} \sqrt{e}\ a\ \exp\{-\frac{a}{2}\} & \text{for } a > 1, \\ 1 & \text{for } a \leq 1, \end{cases}$$

where $a = \frac{|Z_1 - E[Z_1|\tilde{\boldsymbol{Z}}_{-1}]|}{\sqrt{\text{Var}[Z_1|\tilde{\boldsymbol{Z}}_{-1}]}}$ is the distance between Z_1 and the conditional mean $E\left[Z_1|\tilde{\boldsymbol{Z}}_{-1}\right]$, in unit of the conditional s.d. $\text{Var}\left[Z_1|\tilde{\boldsymbol{Z}}_{-1}\right]$. We iteratively test all SNPs $d = 1, \ldots, |D|$, by computing the corresponding $\hat{B}_{0:1,d}$. Among SNPs with the corresponding Bayes factor smaller than a pre-determined threshold, we drop the one with the smallest Bayes factor. We repeat this procedure until all Bayes factors of the remaining SNPs are larger than the threshold. In this paper, for outlier detection we use a stringent threshold of 10^{-4} for the Bayes factor. More details on the computation of the Bayes factor and examples can be found in the section 'Outlier detection' in Supplement section.

Implementing the outlier procedure within the Shotgun algorithm. When there are outliers in a dataset, we want to make sure that the Shotgun algorithm selects the 'current' model γ_S not because of the presence of outliers (which leads to large values for the marginal likelihood). Therefore, for each SNP included in the index set S, we identify all the highly correlated SNPs in the dataset, and perform the outlier detection procedure on the selected group as stated above. Algorithm 2 shows the outlier detection algorithm, implemented within the Shotgun algorithm.

At any step of the Shotgun algorithm, suppose that the current model is γ_S.
Input: The index set $S = \{s_1, \ldots, s_{|S|}\}$ for the current model γ_S, the threshold on the Bayes factor δ, and the threshold on the correlation ρ_{outlier}.

for $s = 1, \ldots, |S|$ **do**

- Given $s_s \in S$, identify the group of highly correlated SNPs indicated by the index set $D = \{i;\ \text{cor}\,(Z_{s_s}, Z_i) \geq \rho_{\text{outlier}} \text{ for } \forall i \in \{1, \ldots, p\}\}$.
- Define $\tilde{\boldsymbol{Z}} = \{Z_1, \ldots, Z_{|D|}\}'$ as the summary statistics vector of the set D .

repeat

for $d = 1, \ldots, |D|$ **do**

- Define the hypothesis test for Z_d, such that

$$\begin{array}{ll} H_0 : Z_d \sim N(\beta, 1); & Z_d \text{ is not an outlier} \\ H_1 : Z_d \sim N(\beta, c),\ c \neq 1; & Z_d \text{ is an outlier.} \end{array}$$

- Compute the corresponding Bayes factor $\hat{B}_d$ conditional on $\tilde{\boldsymbol{Z}}_{D_{-d}}$ and $\boldsymbol{\Sigma}_{D_{-d}}$.

end

if $\exists d \in \{1, \ldots, |D|\}, \hat{B}_d < \delta$ **then**

- Drop Z_d, where $\hat{B}_d = \min\left(\left\{\hat{B}_1, \ldots, \hat{B}_{|D|}\right\}\right)$, from the fine-mapping computation.
- Drop Z_d from $\tilde{\boldsymbol{Z}}$, i.e., $\tilde{\boldsymbol{Z}} = \{Z_1, \ldots, Z_{d-1}, Z_{d+1}, \ldots, Z_{|D|}\}'$.
- Drop dth index from the index set D.

until $\hat{B}_d \geq \delta$, for $\forall d \in \{1, \ldots, |D|\}$;

end

Algorithm 2: The outlier detection procedure implemented in Shotgun algorithm.

Multi-locus extension

As shown before,[4,10,36] a fine-mapping method may benefit from inference based on multiple loci. Suppose that there are a total of L GWAS loci, and we assume that each GWAS locus is independent of each other. The proposed model can be naturally extended to multiple loci. Since each locus $l = 1, \ldots, L$ has a different number of SNPs p_l, the hyperparameter η_l that provides multiplicity control should take a different value $\eta_l \propto p_l^{-a}$, $a > 0$. Thus, the baseline probability of SNPs being causal within each locus is still determined by the size of the corresponding locus. Note that the multiple locus model assumes that functional annotations are similarly associated with the probability to be causal, a rather strong assumption. In our applications we fit the proposed model at the chromosome level.

Credible sets

In[3] the authors define a credible set as follows.

Definition 1. *A level ρ credible set is defined to be a subset of correlated variables (with correlation within the set greater than some threshold r) that has probability ρ or greater of containing at least one effect variable (i.e. causal SNP).*

Constructing credible sets. Given a correlation threshold r, we can define multiple candidate credible sets as follows

$$S := \{i \in \{1, \ldots, p\} : \min\{cor(i, j) \geq r\}, \text{ for all } i, j \in S\}.$$

Given an index set S, let $s = \left\{s_1, \ldots, s_{|S|}\right\}$ denote the indices of the selected SNPs in the set S ranked in order of decreasing PIPs, such as $\Pr(\gamma_{s_1}|\boldsymbol{Z}) > \Pr(\gamma_{s_2}|\boldsymbol{Z}) > \ldots > \Pr\left(\gamma_{s_{|S|}}|\boldsymbol{Z}\right)$, and let P_l denote the cumulative sum of the l largest PIPs:

$$P_l = \sum_{j=1}^{l} \Pr\left(\gamma_{s_j}|\boldsymbol{Z}\right).$$

Then, the credible set is defined as $S_{l_0} = \{s_1, \ldots, s_{l_0}\}$, where $l_0 = \min\{l : P_l \geq \rho\}$. This procedure makes sure that the selected credible sets have the smallest size. Then, given a specific level ρ, we run this procedure on a predetermined sequence of candidate correlation values, such as $r \in \{0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$, and compute the corresponding number of credible sets for each candidate value. The largest correlation threshold r with level ρ credible sets (those sets with posterior probability greater than ρ) is selected along with the corresponding credible sets.

References

[1] Iris E Jansen, Jeanne E Savage, Kyoko Watanabe, Julien Bryois, Dylan M Williams, Stacy Steinberg, Julia Sealock, Ida K Karlsson, Sara Hägg, Lavinia Athanasiu, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. Nature genetics, 51(3):404–413, 2019.

[2] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. The uk biobank resource with deep phenotyping and genomic data. Nature, 562(7726):203–209, 2018.

[3] Gao Wang, Abhishek K Sarkar, Peter Carbonetto, and Matthew Stephens. 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(5):1273–1300, 2020.

[4] Gleb Kichaev, Megan Roytman, Ruth Johnson, Eleazar Eskin, Sara Lindstroem, Peter Kraft, and Bogdan Pasaniuc. Improved methods for multi-trait fine mapping of pleiotropic risk loci. Bioinformatics, 33(2):248–255, 2017.

[5] Yongtao Guan and Matthew Stephens. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. The Annals of Applied Statistics, pages 1780–1815, 2011.

[6] Julian B Maller, Gilean McVean, Jake Byrnes, Damjan Vukcevic, Kimmo Palin, Zhan Su, Joanna MM Howson, Adam Auton, Simon Myers, Andrew Morris, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nature genetics, 44(12):1294, 2012.

[7] Paul J Newcombe, David V Conti, and Sylvia Richardson. Jam: a scalable bayesian framework for joint analysis of marginal snp effects. Genetic epidemiology, 40(3):188–201, 2016.

[8] Christian Benner, Chris CA Spencer, Aki S Havulinna, Veikko Salomaa, Samuli Ripatti, and Matti Pirinen. Finemap: efficient variable selection using summary data from genome-wide association studies. Bioinformatics, 32(10):1493–1501, 2016.

[9] Gleb Kichaev, Wen-Yun Yang, Sara Lindstrom, Farhad Hormozdiari, Eleazar Eskin, Alkes L Price, Peter Kraft, and Bogdan Pasaniuc. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet, 10(10):e1004722, 2014.

[10] Omer Weissbrod, Farhad Hormozdiari, Christian Benner, Ran Cui, Jacob Ulirsch, Steven Gazal, Armin P Schoech, Bryce Van De Geijn, Yakir Reshef, Carla Márquez-Luna, et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. Nature Genetics, pages 1–9, 2020.

[11] Wenan Chen, Beth R Larrabee, Inna G Ovsyannikova, Richard B Kennedy, Iana H Haralambieva, Gregory A Poland, and Daniel J Schaid. Fine mapping causal variants with an approximate bayesian method using marginal test statistics. Genetics, 200(3):719–736, 2015.

[12] James G Scott and James O Berger. Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem. The Annals of Statistics, pages 2587–2619, 2010.

[13] Po-Ru Loh, George Tucker, Brendan K Bulik-Sullivan, Bjarni J Vilhjalmsson, Hilary K Finucane, Rany M Salem, Daniel I Chasman, Paul M Ridker, Benjamin M Neale, Bonnie Berger, et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. Nature genetics, 47(3):284–290, 2015.

[14] Feng Liang, Rui Paulo, German Molina, Merlise A Clyde, and Jim O Berger. Mixtures of g priors for bayesian variable selection. Journal of the American Statistical Association, 103(481):410–423, 2008.

[15] Chris Hans, Adrian Dobra, and Mike West. Shotgun stochastic search for "large p" regression. Journal of the American Statistical Association, 102(478):507–516, 2007.

[16] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2):301–320, 2005.

[17] Apostolos Dimitromanolakis, Jingxiong Xu, Agnieszka Krol, and Laurent Briollais. sim1000g: a user-friendly genetic variant simulator in r for unrelated individuals and family-based designs. BMC bioinformatics, 20(1):26, 2019.

[18] Laura Fachal, Hugues Aschard, Jonathan Beesley, Daniel R Barnes, Jamie Allen, Siddhartha Kar, Karen A Pooley, Joe Dennis, Kyriaki Michailidou, Constance Turman, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nature genetics, 52(1):56–73, 2020.

[19] Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning–based sequence model. Nature methods, 12(10):931–934, 2015.

[20] Robert E Kass and Adrian E Raftery. Bayes factors. Journal of the american statistical association, 90(430):773–795, 1995.

[21] Martin Kircher, Daniela M Witten, Preti Jain, Brian J O'roak, Gregory M Cooper, and Jay Shendure. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46(3):310–315, 2014.

[22] Zikun Yang, Chen Wang, Stephanie Erjavec, Lynn Petukhova, Angela Christiano, and Iuliana Ionita-Laza. A semisupervised model to predict regulatory effects of genetic variants at single nucleotide resolution using massively parallel reporter assays. Bioinformatics, 37(14):1953–1962, 2021.

[23] Qing-Qing Tao, Yu-Chao Chen, and Zhi-Ying Wu. The role of cd2ap in the pathogenesis of alzheimer's disease. Aging and disease, 10(4):901, 2019.

[24] Yuxin Zou, Peter Carbonetto, Gao Wang, and Matthew Stephens. Fine-mapping from summary data with the "sum of single effects" model. bioRxiv, 2021.

[25] Arnold Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. Bayesian inference and decision techniques, 1986.

[26] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature genetics, 47(11):1228, 2015.

[27] Arnold Zellner and Aloysius Siow. Posterior odds ratios for selected regression hypotheses. Trabajos de estadística y de investigación operativa, 31(1):585–603, 1980.

[28] Andrew J Womack, Luis León-Novelo, and George Casella. Inference from intrinsic bayes' procedures under model selection and uncertainty. Journal of the American Statistical Association, 109(507):1040–1053, 2014.

[29] Harold Jeffreys. The theory of probability. OUP Oxford, 1998.

[30] Ismaël Castillo, Aad van der Vaart, et al. Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. The Annals of Statistics, 40(4):2069–2101, 2012.

[31] Andrew J Womack, Claudio Fuentes, and Daniel Taylor-Rodriguez. Model space priors for objective sparse bayesian regression. arXiv preprint arXiv:1511.04745, 2015.

[32] Ismaël Castillo, Johannes Schmidt-Hieber, Aad Van der Vaart, et al. Bayesian linear regression with sparse priors. The Annals of Statistics, 43(5):1986–2018, 2015.

[33] Yuzo Maruyama, Edward I George, et al. Fully bayes factors with a generalized g-prior. The Annals of Statistics, 39(5):2740–2765, 2011.

[34] Ismael Castillo and Etienne Roquain. On spike and slab empirical bayes multiple testing. arXiv preprint arXiv:1808.09748, 2018.

[35] MJ Bayarri and James O Berger. Robust Bayesian bounds for outlier detection. De Gruyter, 1992.

[36] Wenan Chen, Shannon K McDonnell, Stephen N Thibodeau, Lori S Tillmans, and Daniel J Schaid. Incorporating functional annotations for fine-mapping causal variants in a bayesian framework using summary statistics. Genetics, 204(3):933–958, 2016.

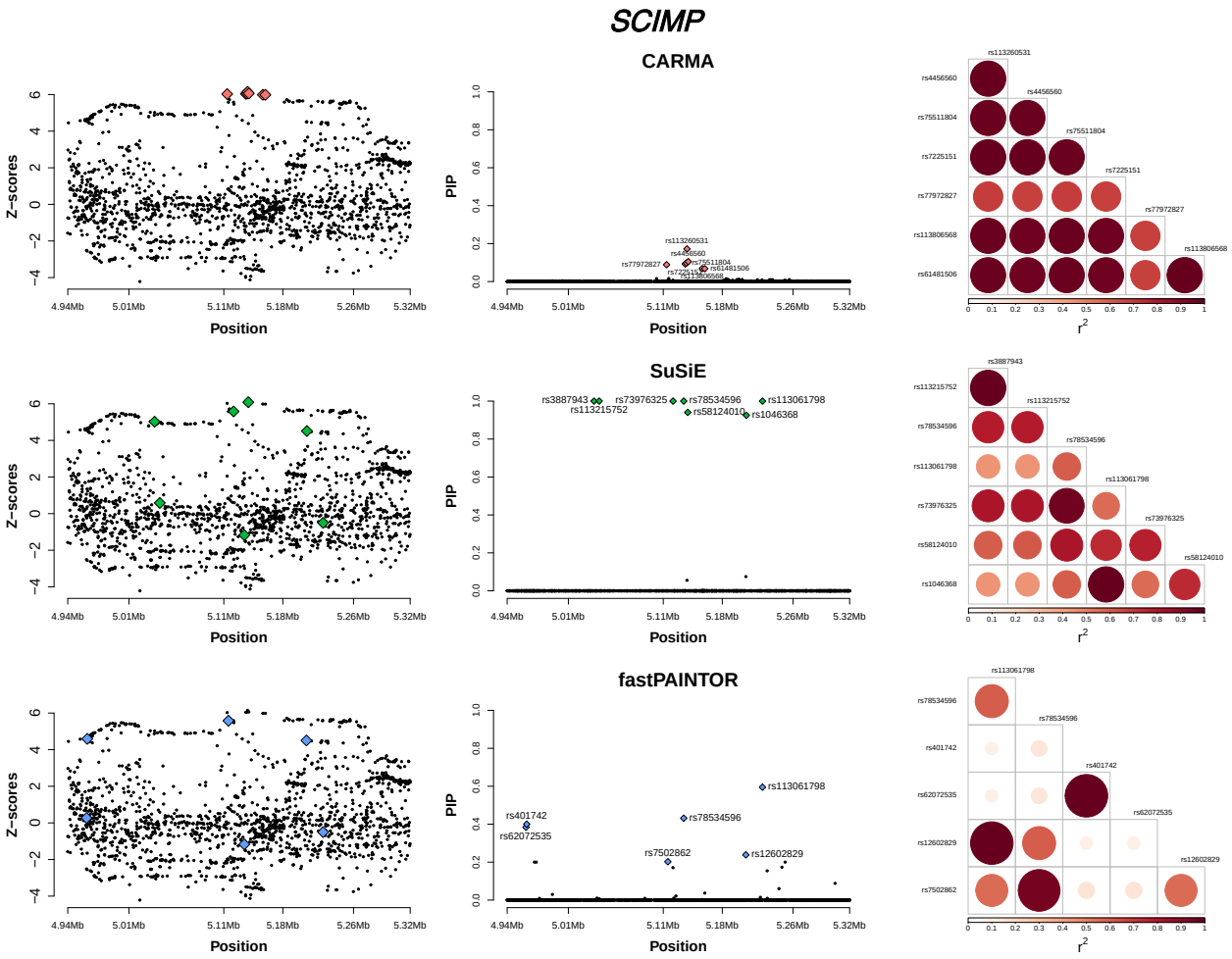

Figure 1: **Motivating example.** GWAS Z-scores along with PIPs from three models (CARMA, SuSiE and fastPAINTOR) for one Alzheimer's disease risk locus *SCIMP* (GRCh37/hg19). For each model, the heatmap depicts the LD (r^2 based on the UKBB data) between SNPs highlighted in the PIP panel on the left.

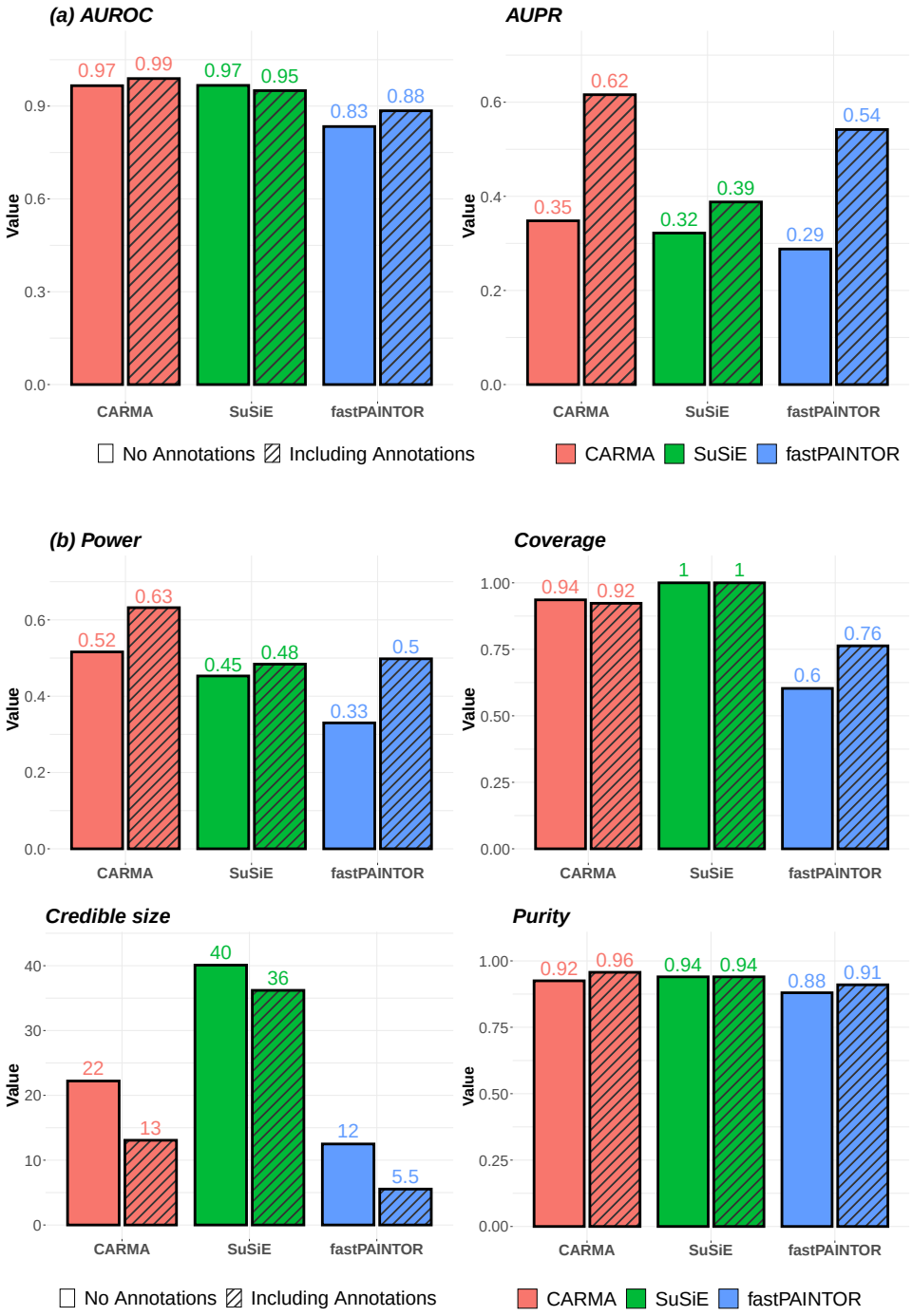

Figure 2: **Effect of functional annotation on AUROC, AUPR, and credible sets.** **(a)** Mean AUROC and AUPR values based on PIPs from the proposed model, SuSiE(+PolyFun), and fastPAINTOR for no functional annotation vs. including functional annotations. **(b)** Performance of credible sets ($\rho = 0.99$). Power: the overall proportion of simulated causal variables included in a credible set; Coverage: the proportion of credible sets that contain a causal variable; Size: the number of variables included in a credible set; Purity: the average squared correlation of variables in a credible set.

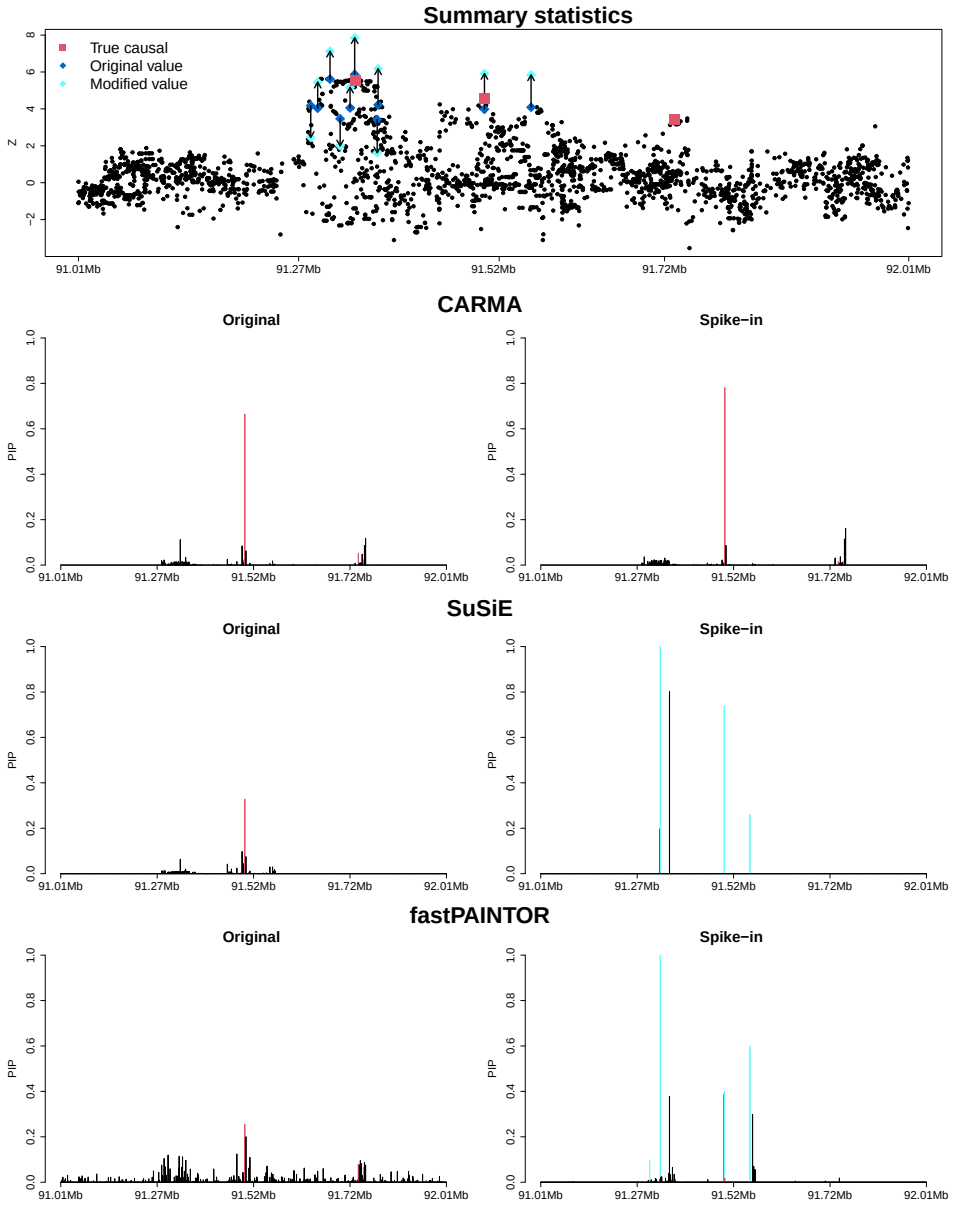

Figure 3: **Example illustrating the effect of summary statistics/LD inconsistencies.** The top panel shows the summary statistics with the original values and the modified values for the SNPs selected as outliers. The bottom six panels show the results of CARMA, SuSiE and fastPAINTOR on the original and spike-in datasets. The red bars and green bars highlight the PIPs of the true causal SNPs and the outliers, respectively.

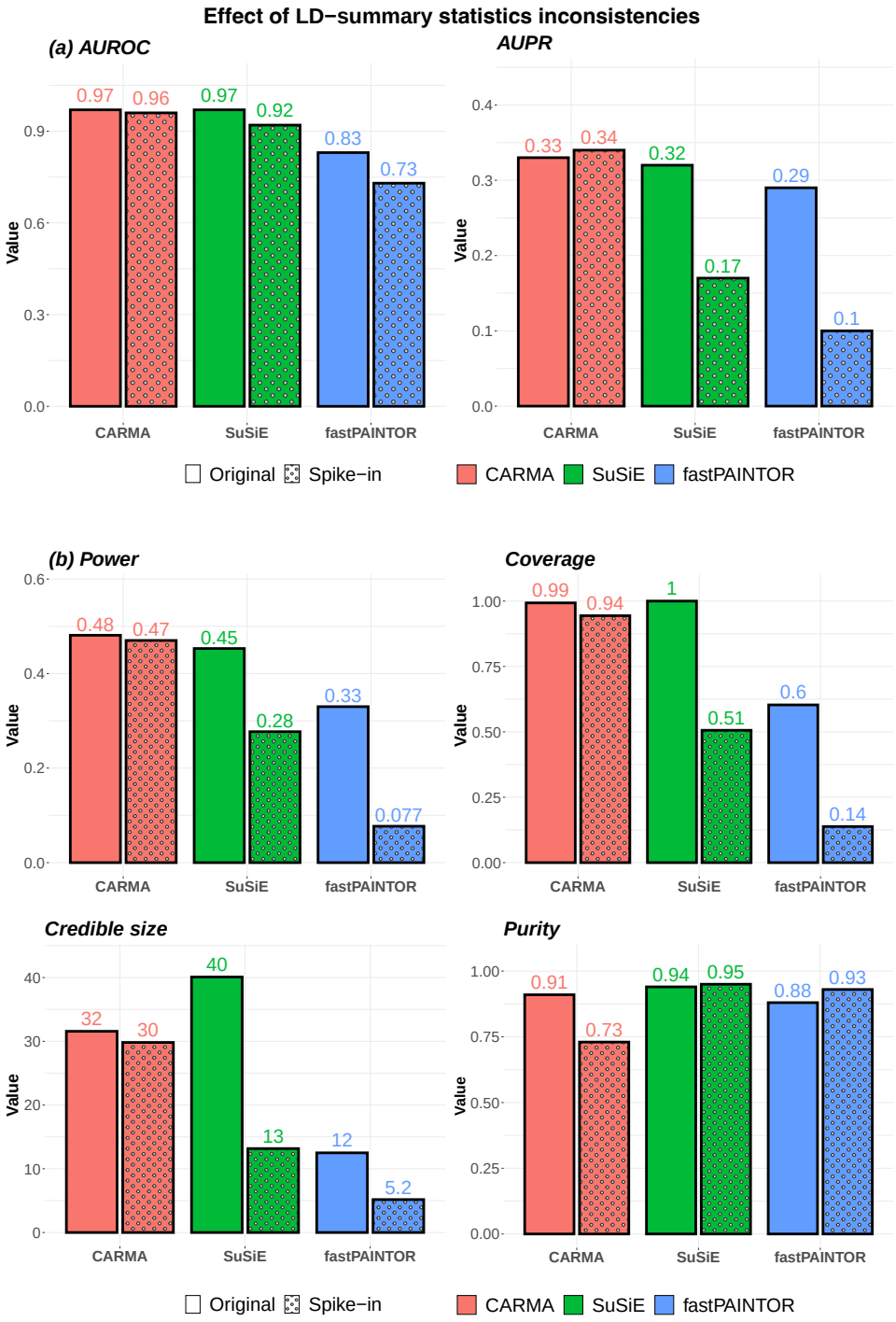

Figure 4: **Effect of summary statistics/LD inconsistencies in simulations.** **(a)** Mean AUROC and AUPR values in the original vs. outlier enriched datasets. **(b)** Performance of credible sets in the original vs. outlier enriched datasets ($\rho = 0.99$). Power: the overall proportion of simulated causal variables included in a credible set; Coverage: the proportion of credible sets that contain a causal variable; Size: the number of variables included in a credible set; Purity: the average squared correlation of variables in a credible set.

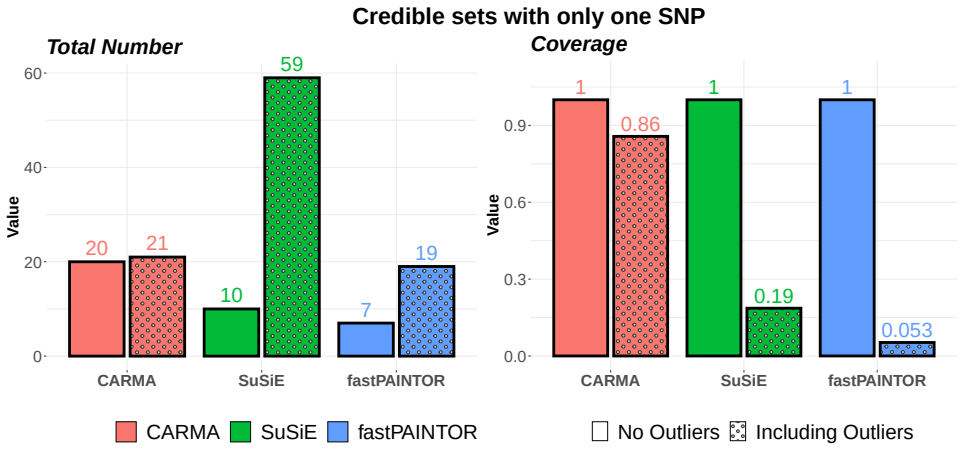

Figure 5: **Increased single SNP credible sets in the presence of summary statistics/LD inconsistencies.** Results on credible sets ($\rho = 0.99$) with only one SNP, i.e., the corresponding SNP receives a PIP larger than 0.99. For each model we report the total number of credible sets that contain only one SNP across all 94 loci, and the coverage of these sets, i.e. the proportion of these credible sets that contain a causal SNP.

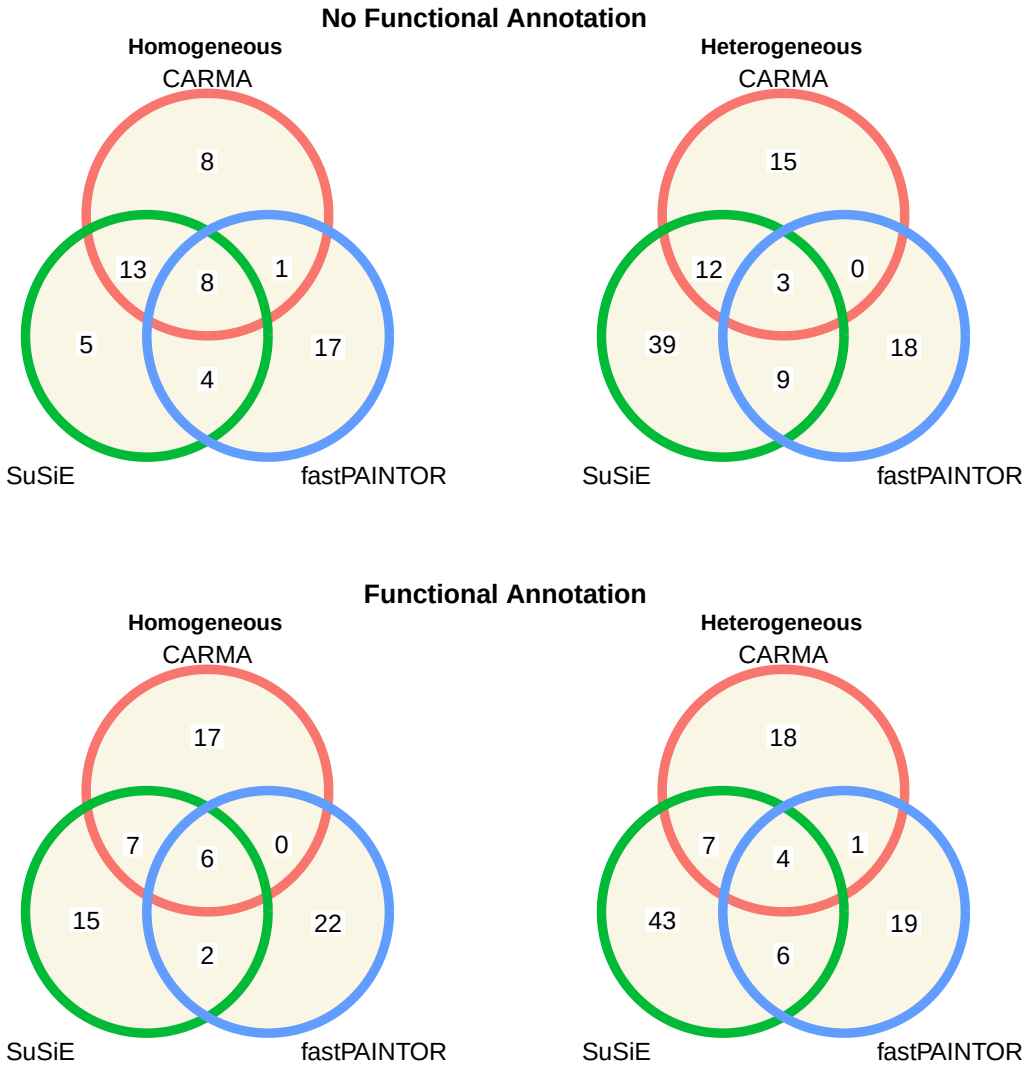

Figure 6: **Overlap of top PIP SNP at 30 Alzheimer's disease risk loci.** Venn diagrams for top SNPs at 30 Alzheimer's disease risk loci (highest PIP SNP at each locus) selected by CARMA, SuSiE, and fastPAINTOR with (a) no functional annotations, and (b) with functional annotations in the homogeneous and the heterogeneous datasets.

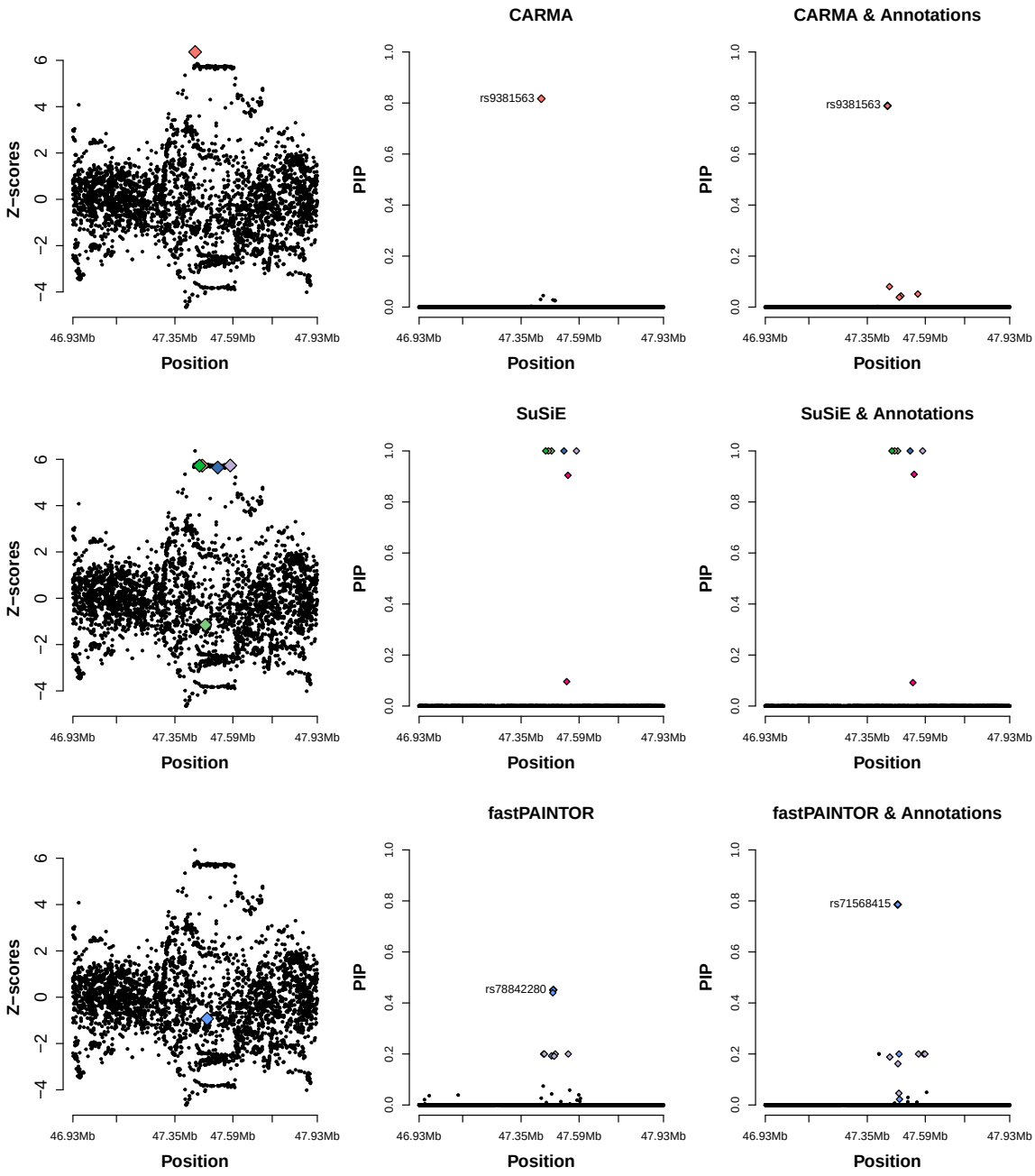

Figure 7: **PIPs of the three models for** *CD2AP*. The SNP with largest PIP (if unique) from each model is highlighted in each panel. The credible sets of each model are also highlighted by different colors. Results are based on the complete/heterogeneous dataset.

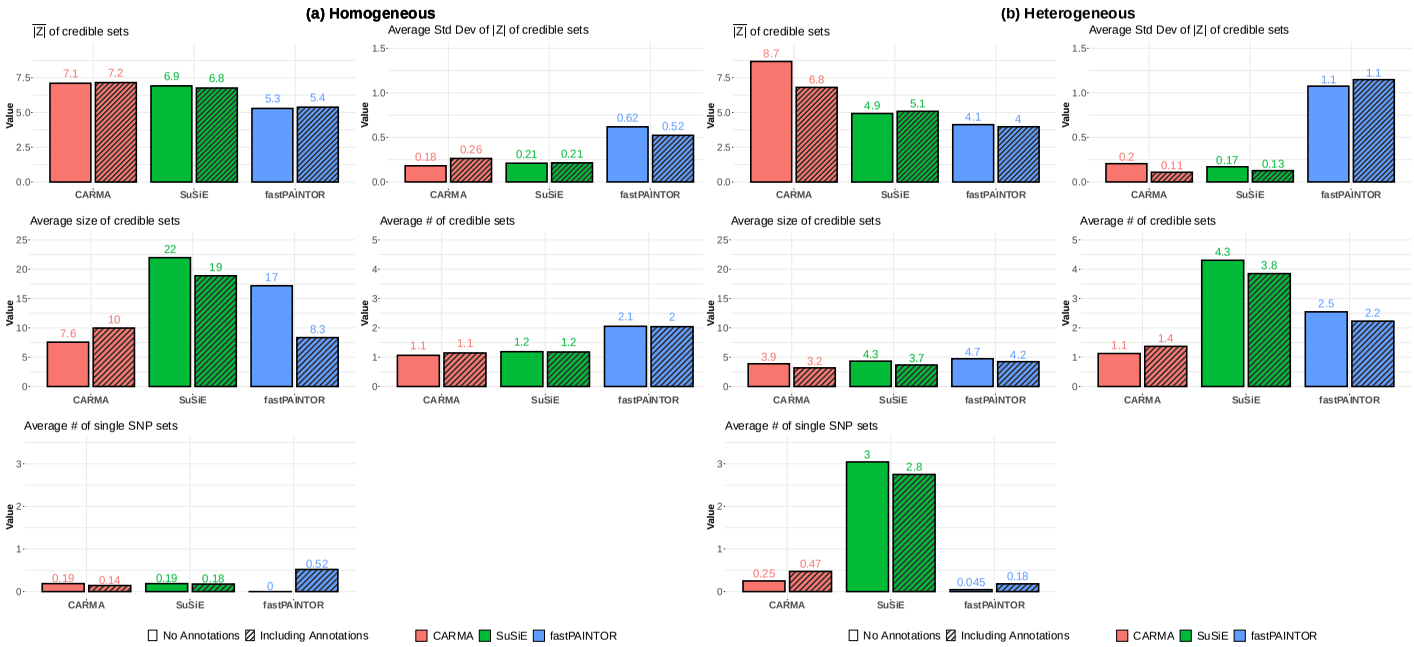

Figure 8: **Performance of the credible sets for the three models across 30 Alzheimer's disease risk loci.** Panel **(a)** depicts results for the homogeneous dataset, and panel **(b)** for the heterogeneous dataset. Top left: average absolute value of the summary statistics within each credible set; top right: standard deviation of the absolute value of summary statistics within each credible set; bottom left: size of the credible sets; bottom right: average number of credible sets per locus. The homogeneous dataset focused only on SNPs included in the IGAP, PGC-ALZ, and the large AD-by-proxy study analyses; the heterogeneous dataset focused on all SNPs available. The sample sizes associated with the heterogeneous dataset can vary from 9,703 to 444,006, whereas the sample sizes for the SNPs included in the homogeneous set vary between 418,339 and 436,498.

					CARMA PIP			
Locus	SNP	Z	N	EAF	$Homo_{\text{No Annot}}$	$Homo_{\text{Annot}}$	$Het_{\text{No Annot}}$	Het_{Annot}
ADAMTS4	rs4575098	6.37	434326	0.23	0.71	0.71	0.69	0.71
BIN1	rs4663105	14.01	425134	0.41	1.00	1.00	1.00	1.00
INPPD5	rs10933431	-6.15	436498	0.24	0.62	0.65	0.48	0.48
CD2AP	rs9381563	6.36	429204	0.36	0.61	0.61	0.82	0.79
ZCWPW1	rs1859788	-7.95	441735	0.32	0.42	0.41	0.42	0.42
CLU-PTK2B	rs4236673	-9.05	443360	0.38	0.35	0.34	0.31	0.37
ECHDC3	rs11257238	5.72	430330	0.36	0.29	0.30	0.28	0.42
MS4A6A	rs2081545	-8.01	441064	0.38	0.85	0.84	0.44	0.44
PICALM	rs677869	-8.75	432250	0.32	0.22	0.25	0.30	0.47
SORL1	rs11218343	-6.84	436249	0.04	0.99	0.99	1.00	1.00
SLC24A4	rs12590654	-6.42	426902	0.34	0.39	0.36	0.40	0.50
APH1B	rs117618017	5.52	444006	0.12	0.71	0.68	0.21	1.00
ALPK2	rs76726049	5.52	433227	0.01	0.99	0.99	0.28	0.40
CD33	rs3865444	-5.84	436498	0.30	0.46	0.45	0.35	0.49
CASS4	rs6014724	-6.21	433858	0.10	0.30	0.30	0.25	0.41

Table 1: **Loci where the same top SNP (with largest PIP) is identified by CARMA across all scenarios.** EAF: effect allele frequency, Z: meta-analysis Z-score (the same for the homogeneous and the heterogeneous datasets); N: number of samples in the homogeneous datasets; PIPs from the CARMA model across the four scenarios are reported.

Method	Prior on effect size	Prior on model space	Max#causal	Functional annot	Computation	Ref.
FINEMAP	Normal-Gamma	discrete Uniform	Fixed	No	Shotgun	[8]
CAVIARBF	Normal-Gamma	discrete probability	Fixed	Yes	Exhaustive	[36]
JAM	g-prior	Beta-Binomial	Unlimited	No	Exhaustive and MCMC	[7]
fastPAINTOR	Normal-Gamma	discrete probability	Unlimited	Yes	Exhaustive and MCMC	[4]
SuSiE (+PolyFun)	Normal-Gamma	Multinomial (discrete probability)	Unlimited	Yes	Variational Bayes	[3,10]
CARMA (proposed model)	Mixture of g-prior	Poisson	Unlimited	Yes	Shotgun	

Table 2: **Summary of commonly used Bayesian fine-mapping methods**.