CARMA: Novel Bayesian model for fine-mapping in meta-analysis studies

Zikun Yang\textsuperscript{1}, Chen Wang\textsuperscript{1,2}, Atlas Khan\textsuperscript{2}, Badri Vardarajan\textsuperscript{3}, Richard Mayeux\textsuperscript{3}, Krzysztof Kiryluk\textsuperscript{2}, Iuliana Ionita-Laza\textsuperscript{1,\#}

\textsuperscript{1} Department of Biostatistics, Columbia University, New York
\textsuperscript{2} Division of Nephrology, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York
\textsuperscript{3} Department of Neurology, College of Physicians and Surgeons, Columbia University, New York
\textsuperscript{\#} Correspondence: 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 methods, the proposed model has several appealing features, such as 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 values in meta-analysis settings. Using simulations, we compare performance with commonly used fine-mapping methods, including fastPAINTOR and SuSiE, and show that the proposed model has smaller credible sets with high power when using in-sample LD, and lower FDR/higher precision and higher coverage for credible sets when using external LD. 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, including \textit{NDUFS2}, \textit{INPP5D}, \textit{SORL1}, \textit{BIN1}, \textit{CD2AP}, \textit{CASS4}, \textit{PICALM}, \textit{MS4A6A} and others.

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.\textsuperscript{1} 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 \textit{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.\textsuperscript{2} The LD matrix is estimated using individuals of European descent in UK Biobank (UKBB).\textsuperscript{3} Two fine-mapping models, SuSiE\textsuperscript{4} and fastPAINTOR,\textsuperscript{5} 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.\textsuperscript{4–12} 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, i.e. some methods restrict the number of possible causal variants,\textsuperscript{6,7,10,12} 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.\textsuperscript{5,8,9} 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,\textsuperscript{13} and several existing methods address this issue formally by introducing prior probabilities on model space.\textsuperscript{8,9} 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;\textsuperscript{14} 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 in Meta-Analysis studies), 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\textsuperscript{15} 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\textsuperscript{11} 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 (\(X\)) and phenotype (\(y\)) data for \(n\) subjects. We assume a standard linear model \(y = X\beta + \epsilon, \epsilon \sim \text{MVN}(0, \sigma_y^2 I_n)\), where \(y\) is a \(n \times 1\) vector of quantitative phenotype values, \(X\) is a standardized \(n \times p\) genotype matrix, \(\beta\) is a \(p\)-dimensional vector of effect sizes of variants, and \(\epsilon\) 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 \(se(\hat{\beta}_i)\). Then the Z-score is defined as \(Z_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}\) with \(Z_i \sim N(0, 1)\) under the null hypothesis \(H_0: \beta_i = 0\).

Let \(\gamma' = \{0, 1\}^p\) denote an indicator vector, such that \(\gamma_i = 1\) if \(\hat{\beta}_i \neq 0\). Let \(\lambda \propto \beta\) (i.e., \(\lambda_i \neq 0\) if and only if \(\beta_i \neq 0\)). The sampling distribution of \(Z\) can be written as:

\[
Z|\lambda, \sigma_y^2, \Sigma \sim \text{MVN}(\Sigma \lambda, \sigma_y^2 \Sigma),
\]

where \(\Sigma = \frac{X^TX}{n}\). Given any model \(\gamma\), we make the following model assumptions:

\[
\lambda_\gamma|\sigma_y^2, \tau, \gamma \sim \text{MVN}(0, \frac{\sigma_y^2}{\tau} \Sigma_\gamma^{-1}),
\]

\[
\tau \sim \text{Gamma}(0.5, 0.5),
\]

\[
\sigma_y^2 \sim \frac{1}{\sigma_y^2},
\]

where \(\lambda_\gamma\) represents the vector of non-zero entries of \(\lambda\) (corresponding to \(\gamma_i = 1\)) and \(\Sigma_\gamma\) represents the submatrix of \(\Sigma\) corresponding to model \(\gamma\). We want to identify the true model (denoted by the indicator vector \(\gamma_T\)) that generated the summary statistics, through posterior inference within a Bayesian paradigm.

Note that the marginal prior distribution of \(\lambda_\gamma\) is a multivariate Cauchy distribution after integrating out the mixing parameter \(\tau\). We additionally assume a truncated Poisson prior on the size of the model, i.e. \(\sum_{i=1}^p \gamma_i \sim \text{Truncated Poisson}(\eta)\) 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\textsuperscript{16} 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.
Credible sets and credible models. In the authors define a credible set as a subset of correlated variants (with correlation within the set greater than some threshold \( r \)) that has probability \( \rho \) or greater of containing at least one causal variant, e.g. the \( i \)th and \( j \)th SNPs could form a credible set if \( \Pr (\gamma_i = 1 | Z) + \Pr (\gamma_j = 1 | Z) \geq \rho \) and \( \text{cor} (Z_i, Z_j) \geq r \). The credible set provides a small set of SNPs for follow-up studies.

However, as also mentioned in, a SNP should not be considered as non-causal if it is not included in any credible set. For example, there are circumstances where a group of highly correlated SNPs containing one causal SNP might not result in a sufficiently large PIP to form a credible set. Hence, we 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.

Simulations

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

Generating simulated datasets

Genotype simulation. We use the R package ‘sim1000G’ 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. 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). Let \( \mathbf{\theta} \) 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 \( N(0,0.01) \). 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 \( i \) being causal as \( \Pr (\gamma_i = 1 | \mathbf{\theta}, \mathbf{w}_i) = \frac{\exp \{ \mathbf{w}_i' \mathbf{\theta} \}}{1 + \exp \{ \mathbf{w}_i' \mathbf{\theta} \} } \), where \( \mathbf{w}_i \) is the vector of annotations. The causal SNPs are selected among those with highest prior probabilities. In simulations we assume three causal variants at each locus.

Phenotype generation. For a given locus, let \( T \) denote the index set of the true causal SNPs, i.e., if \( i \in T \), then \( \gamma_i = 1 \). For each \( i \in T \), the effect sizes of the causal SNPs are drawn independently from \( \beta_i \sim N(0, 0.5^2) \), and for all \( i \notin T \) we set \( \beta_i = 0 \). The phenotypic variance \( \sigma_y^2 \) is computed such that \( \phi = 0.0075 \), where \( \phi = \frac{\text{Var} (X\mathbf{\beta})}{\sigma_y^2 + \text{Var} (X\mathbf{\beta})} \). Then we sample \( \mathbf{y} \) such that \( \mathbf{y} = X\mathbf{\beta} + \mathbf{e}; \mathbf{e} \sim N(0, \sigma_y^2 I_{n \times n}) \).
Remarks on some implementation details:

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. SuSiE can only include one annotation in the form of prior probabilities. To make fair comparisons with CARMA, in simulations we use the same prior probabilities estimated by CARMA.

3. For CARMA, the value of hyper-parameter $\eta$ is set at the default setting $\sqrt{1/p}$ when using in-sample LD, and hence depends on the number of SNPs at each locus. In simulations using external LD, we use an adaptive procedure for selecting $\eta$ accounting for the inconsistency between Z-scores and external LD matrix (Methods).

4. We run fastPAINTOR and CARMA models at a chromosome level, while SuSiE 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 different models (Figure 2a). CARMA and SuSiE have similar performance, and generally outperform fastPAINTOR. We also compute for each locus the AUROC and AUPR using the estimated prior probability of the CARMA model $Pr(\gamma|W, \theta)$ as a predictor. The mean AUROC/AUPR value across loci is 0.91/0.46, which suggests that the estimated prior probabilities provide useful information on the SNPs being causal, and lead to improved accuracy when including functional annotations (Figure S1).

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 $^4$ (Methods). Furthermore, we compute four statistics for the credible sets as in $^4$

1. **Power**: The overall proportion of simulated causal variants included in any 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 S2). As above, CARMA and SuSiE have similar performance, except that the size of the credible sets for CARMA is smaller than that for SuSiE. In contrast, fastPAINTOR has lower power and coverage. When functional annotations are included, the AUPR for all methods increases. Notice that due to the heavy-tailed prior distribution on the effect sizes assumed by CARMA, the credible size of CARMA tends to be lower than for SuSiE even with the same prior. Finally, the substantially lower coverage of the credible sets from fastPAINTOR suggests the need of including dimensional penalization to control the false positives.

Credible models. CARMA 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 $^{21}$ We view credible models as complementary to credible sets. Therefore, for each locus, we compare the results based on the variants included in the credible models in CARMA and the variants
included in all the credible sets for SuSiE and fastPAINTOR. Compared to the credible sets of SuSiE and fastPAINTOR, CARMA credible models generally achieve higher power with smaller sets of selected variants, which may facilitate follow-up investigations (Figure 2c).

Positive predictions. We also compare methods in terms of power and false discovery rate (FDR) using positive predictions, i.e. PIP is greater than a given threshold. With no functional annotation, the performances of CARMA and SuSiE are comparable whereas fastPAINTOR has higher FDR (Figure 3a). These results also show the improved performance (higher power and lower FDR) of CARMA and SuSiE when introducing functional annotations, while fastPAINTOR still experiences higher FDR than the other two models.

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 0.01 for PIPs of perfectly correlated SNPs. This is in contrast to fastPAINTOR which has a standard deviation of 0.494. 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. Furthermore, more complications can arise due to sample size heterogeneity across different SNPs. Allele flipping (alleles are encoded differently in the study and reference panel) can also create problems for fine-mapping. Here we show the robustness of the CARMA model in such scenarios.

Specifically, we artificially create such scenarios by adding outliers to all 94 GWAS loci, and then perform fine-mapping on these spike-in datasets to assess the robustness of the three testing models. First, we identify all groups with at least three SNPs and such that the minimum correlation within 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 associated with the hypothesis 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 $10^{-4}$, then we move to the next group, otherwise, we repeat the procedure for the current group until the Bayes factor is less than the threshold. For each locus, we add 10 outliers, one for each group of highly correlated SNPs, for a total of 940 outliers across the 94 loci.

This scenario can lead to substantial bias in the term $Z^T\Sigma^{-1}Z$ (e.g. $Z^T\Sigma^{-1}Z$ can become negative or extremely large), which leads to invalid marginal likelihood for CARMA (Methods). We fix $\sigma^2_y = 1$ for these scenarios because this way the term $\exp\left\{\frac{-Z^T\Sigma^{-1}Z}{2\sigma^2_y}\right\}$ can be ignored (it is constant across all candidate models). We select $\eta$ through an adaptive procedure (Methods).

CARMA is less affected than SuSiE and fastPAINTOR when using the external LD matrix, especially in terms of AUPR and coverage values for the credible sets (Figure 4a). In particular, the coverage values for the credible sets for SuSiE and fastPAINTOR drop substantially, to 64% and 23% respectively, whereas CARMA still maintains relatively high coverage (> 80%, Figure 4b). In terms of credible models, CARMA also performs well and generally produces smaller models relative to the credible sets in SuSiE (Figure 4c). Notably, Figure 3b
shows that the FDR for CARMA using spike-in data is substantially lower than for the competing methods, whereas the FDR of SuSiE and fastPAINTOR can be very high when using external LD regardless of the PIP threshold used. This is due to the fact that there are many non-causal SNPs that are mistakenly assigned PIPs close to 1 by SuSiE and fastPAINTOR (see below).

Credible sets with only one SNP. A particular situation arises when a credible set contains only one SNP (i.e. its PIP exceeds \( \rho \)). Such scenarios require that a SNP have a large summary statistic and be relatively independent of all the other SNPs. Discordant LD/Z-score values when using external LD may create the appearance that a SNP with large summary statistic is weakly correlated to surrounding SNPs. Figure 5 shows the total number of credible sets with one SNP for the three models across the 94 loci, and the corresponding coverage of these credible sets. As shown, when using in-sample LD CARMA identifies more credible sets with one causal SNP relative to SuSiE and fastPAINTOR, which suggests that CARMA can better identify strong signals due to the heavy-tailed prior distribution on the effect sizes. When analyzing the spike-in 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 non-causal SNPs receive large PIP values from these existing models.

Fine-mapping Alzheimer’s disease GWAS loci

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. The clinically diagnosed AD case-control data are from three 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). 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 ±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. For the CARMA model, we include 924 functional annotations including DeepSEA, CADD, PO-EN, and PolyFun. 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 smaller number of SNPs at the 30 loci relative to the heterogeneous dataset (70,698 vs. 127,373), 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 S3-S32. SNPs with largest PIP at each locus, for each model and each scenario, are reported in Supplemental Table S1. Loci where all three models identify the same top SNP are reported in Table S1, for different scenarios. There is a rather large overlap between CARMA and SuSiE for the homogenous dataset with no functional information, as expected given the similar performance of CARMA and SuSiE in simulations using in-sample LD (20/30, 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 (Figure 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 discrepancies between Z-scores and LD values. One example is CD2AP (CD2 associated protein), where CARMA identifies a top SNP rs9381563 (eQTL for CD2AP in Brain Dorsolateral Prefrontal Cortex in the The Religious Orders Study and Memory and Aging Project (ROSMAP), mQTL associated with methylation status for CD2AP in Brain Dorsolateral Prefrontal Cortex in ROSMAP), while SuSiE identifies 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 S11). 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.

For 16/30 loci the SNP with the largest PIP is shared across the different analyses in CARMA (homogeneous/heterogeneous, with or without functional annotation, Table 1). We have used several functional annotation tools including CADD, Eigen/EigenPC, and data on eQTL and sQTL from GTEx and ROSMAP to functionally annotate these SNPs (Tables S2-S3). We also report for each SNP the gene with the highest V2G score in the Open Targets Genetics Portal. There is a rather high concordance between the gene with highest score in Open Targets and the gene identified by mQTLs in Brain Dorsolateral Prefrontal Cortex in ROSMAP (10/16), suggesting that these SNPs and the putative effector genes could 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), concordant with simulations results. 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 (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. Namely, the heterogeneous dataset contains more variants with stronger SNPs, which eventually receive larger PIPs. Therefore, the credible sets tend to contain stronger signals as illustrated by the increased mean |Z| (Figure 8b), and therefore become smaller.

Credible models. 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 |Z| of the SNPs selected by the credible models is smaller than for the credible sets (Figure S33). Note that the credible models are also less affected by potential outliers in the heterogeneous datasets with similar results as in the homogeneous dataset. More results can be found in Table S4.

Discussion

We have 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 precision and lower FDR when predicting causal variants relative to SuSiE and fastPAINTOR, especially in complex meta-analyses settings with LD from external 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. The results show that CARMA has the ability to better 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 16 out of 30 loci where CARMA consistently identifies the same top SNP across the different analyses (homogeneous or heterogeneous, with or without functional annotations). We also link these top SNPs to putative effector genes using two independent approaches; many of the putative effector genes have strong literature support for a role in Alzheimer’s disease pathogenesis.

Very recently, a new version of SuSiE has been released that is applicable to summary statistics and LD matrix extracted from reference panels. In this context, the authors developed 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, i.e. 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 and other loci with biased PIPs in our analyses, this new implementation has failed to solve the problem (more details are in the Supplemental Material section).

CARMA has been implemented in a computationally efficient R package.

Methods

Basic notations and assumptions

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

\[
y = X\beta + \epsilon, \quad \epsilon \sim MVN(0, \sigma_y^2I_n)
\]

where \(y\) is a standardized \(n \times 1\) vector of quantitative phenotype values, \(X\) is a standardized \(n \times p\) genotype matrix, \(\beta\) is a \(p\)-dimensional vector of effect sizes of SNPs, and \(\epsilon\) 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:

\[
\hat{\beta}_i = \frac{x'_iy}{x'_ix_i}, \quad se(\hat{\beta}_i) = \sqrt{\frac{s_i^2}{x'_ix_i}},
\]

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

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

Details are in the Supplemental Material. Note that the matrix \(X'X\) 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 \(\Sigma = \frac{X'X}{n}\) and \(\lambda \propto \beta\) (i.e., \(\lambda_i \neq 0\) if and only if \(\beta_i \neq 0\)), the sampling distribution of \(Z\) can be represented as

\[
Z | \lambda, \sigma_y^2, \Sigma \sim MVN(\Sigma\lambda, \sigma_y^2\Sigma).
\]

8
Indicator vector $\gamma$ and true model. Let $\gamma' = \{0, 1\}^p$ 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 $\gamma_S$ uniquely defines a model with dimension equal to $|S| = \sum_{i=1}^p \gamma_i$. For a particular model $\gamma_S$, we denote the $p$-dimensional coefficient vector by $\lambda_S$ and denote the $|S|$-dimensional subvector of the non-zero entries of $\lambda_S$ by $\lambda_{\gamma_S}$, 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. Building upon prior Bayesian fine-mapping methods such as CAVIARBF and FINEMAP, JAM assumes the popular $g$-prior model to model the effect sizes for a given model $\gamma_S$:

$$
\lambda_{\gamma_S} \sigma_y^2 \sim MVN \left( 0, g \sigma_y^2 \Sigma_{\gamma_S} \right),
\sigma_y^2 \sim \text{Inv-Gamma}(0.01, 0.01),
$$

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:

$$
Pr(\{\gamma; 1 \cdot \gamma = s\}) \sim \text{Beta-binomial}(s + a, p - s + b),
Pr(\gamma) = \frac{Pr(\{\gamma; 1 \cdot \gamma = s\})}{\binom{p}{s}},
$$

where $\{\gamma; 1 \cdot \gamma = s\}$ 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 $\sigma_y^2$ is a plug-in estimate, JAM marginalizes out both $\lambda$ and $\sigma_y^2$ to improve the power. However, it has been shown in $^{15}$ that the $g$-prior leads to issues with model selection consistency. Another potential issue is that the correlation matrix $\Sigma$ is not necessarily positive definite, and JAM cannot work properly in such cases.

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

$$
Z | \lambda_S, \Sigma \sim MVN(\Sigma \lambda_S, \Sigma),
$$

for a given model $\gamma$. The assumed prior distribution of $\lambda_S$ can be written as

$$
\lambda_S | \gamma, \sigma^2_\lambda \sim MVN(0, \Sigma_{\gamma}),
\Sigma_{\gamma} = \sigma^2_\lambda \text{Diag}(\gamma) + \text{Diag}(\sigma^2_y).
$$

The distribution of $Z$ after integrating out $\lambda_S$ is

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

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

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

where $w_i$ 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. $^{11}$
SuSiE.⁴ 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:

\[ y = X\beta + \epsilon, \quad \epsilon \sim \text{MVN}(0, \sigma^2_y I_n) \]
\[ \beta = b\gamma \]
\[ b \sim N(0, \sigma^2_0) \]
\[ \gamma \sim \text{Mult}(1, \pi), \]

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

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 \( \beta_1, \ldots, \beta_L \) and construct the overall effect vector \( \beta \) as the sum of these single effects. The model is as follows:

\[ y = X\beta + \epsilon, \quad \epsilon \sim \text{MVN}(0, \sigma^2_y I_n) \]
\[ \beta = \sum_{l=1}^{L} \beta_l \]
\[ \beta_l = \gamma_l b_l \]
\[ b_l \sim N(0, \sigma^2_0) \]
\[ \gamma_l \sim \text{Mult}(1, \pi). \]

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

\[ \Pr (\beta_j \neq 0|X, y) \approx 1 - \prod_{l \in \{1, \ldots, L\}} (1 - \alpha_l_{j,l}), \]

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

SuSiE+PolyFun.¹¹ To fully utilize the potential of the fine-mapping methods that can only take as input univariate prior causal probability, Weissbrod et al.¹¹ proposed to estimate prior probabilities through the S-LDSC model,⁵⁰ 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: \( \Pr (\beta_i \neq 0|w_i) \propto \text{var} [\beta_i|w_i] = \theta' w_i \), where

\[ \hat{\theta} := \arg\min_{\theta \in \mathbb{R}^q} \sum_i \left[ \chi^2_i - n \sum_{j=1}^q \theta_j l(i, j) - nb - 1 \right]^2 + \alpha \| \theta \|^2; \]

\( q \) is the number of functional annotations, \( \chi^2_i = \frac{(x_i'y_i)^2}{n} \) is the \( \chi^2 \) statistic of SNP \( i \), \( l(i, j) = \sum_{k=1}^p \text{cor} (x_i, x_k) w_{j,k} \) 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. The advantage of PolyFun is that the contribution of a functional annotation to the heritability is estimated using genome-wide SNPs.

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

**Proposed Bayesian fine-mapping model: CARMA**

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

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

We want to identify the true model (denoted by the true indicator vector $\gamma_T$) 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 $\lambda_i$, i.e.,

$$\lambda_i \sim \text{Mixture of Normal distributions if } \gamma_i = 1,$$

$$\lambda_i = 0 \text{ if } \gamma_i = 0.$$

Specifically, given an index set $S$, the prior distribution of the assumed non-zero effect sizes $\lambda_{\gamma_S}$ that is associated with $\gamma_S$ is

$$\lambda_{\gamma_S} | \sigma_y^2, \tau, \gamma_S \sim \text{MVN}(0, \sigma_y^2 \tau^{-1} \Sigma^{-1}_{\gamma_S}),$$

$$\tau \sim f(\tau), \tau \in (0, \infty),$$

where $f(\tau)$ denotes the prior distribution on $\tau$. For the prior distribution of $\sigma_y^2$, 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 $\sigma_y^2$ is fixed. By assigning a prior distribution on the mixing parameter $\tau$, we intrinsically assign a mixture of normal distribution on the effect size $\lambda_{\gamma_S}$. 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\(^{31}\) and the hyper-g prior,\(^{15}\) as follows:

Zellner-Siow’s Cauchy: $\tau \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$,

hyper-g: $\tau \sim \text{beta-prime}(\frac{1}{2}, 1)$.

Under both priors, integrating out $\tau$ yields heavy-tail prior distributions on the effect sizes $\lambda_{\gamma_S}$; in fact, the marginal prior distribution of $\lambda_{\gamma_S}$ under Zellner-Siow’s prior is a multivariate Cauchy distribution (Supplemental Material). 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.\(^{14}\)

The advantages of assigning a mixture of normal prior distribution over normal prior in model selection are well documented in the Bayesian literature.\(^{15,31,32}\) Specifically, in\(^{15,32}\) the authors show that, under mild conditions, a variety of mixtures of normal distributions can achieve model selection consistency $\lim_n \Pr (M_T | Z) = 1$, whereas the normal prior suffers two well-known model selection paradoxes, Bartlett’s Paradox\(^{33}\) and Information Paradox,\(^{29}\) which prevents the posterior probability of the true model from converging to 1. As shown in the Supplemental Material, 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.6 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.34,35

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 \(|S| = \sum_{i=1}^{\gamma} \gamma_i\) denote the total number of causal SNPs for a given \(\gamma_S\). We first place a discrete prior distribution on the random variable \(|S|\) and let \(\Pr(|S|)\) denote the p.m.f with \(\eta\) as a hyperparameter. Given \(|S|\), we assume that all those models have the same prior probability. Hence, the prior probability of \(\gamma_S\) is

\[
\Pr(\gamma_S|\eta) = \frac{\Pr(|S|)\Pr(|S|)}{|S|!}. 
\]

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.5 Let \(F(-|\eta|)\) be the cumulative distribution function of a Poisson distribution with mean \(\eta\), the truncated Poisson prior is defined as follows:

\[
\Pr(|S|) = \frac{\eta^{|S|}\exp\{-\eta\}}{|S|!} \propto \frac{\eta^{|S|}\exp\{-\eta\}}{|S|!}.
\]

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

\[
\Pr(\gamma_S|\eta) \propto \frac{\eta^{|S|}\exp\{-\eta\}(p - |S|)!}{p!}.
\]

The Poisson distribution on model space provides necessary dimension penalization for multiplicity control. The hyper-parameter \(\eta\) plays a critical role in this dimension penalization mechanism. Note that the computation of the PIs is based on the unnormalized posterior probabilities for the candidate models, i.e., \(\Pr(\gamma_S|Z) \propto f(Z|\gamma_S)\Pr(\gamma_S|\eta)\). Therefore, the choice of \(\eta\) depends on the value of the marginal likelihood, i.e., \(f(Z|\gamma_S)\). In the ideal setting when in-sample LD is being used, we recommend \(\eta \propto p^{-0.5}\) to address the scenario where there are only a handful of causal SNPs in the testing locus harboring thousands of SNPs, i.e. the ultra-sparse scenario as \(p \rightarrow \infty\) and \(|T|\) constant. When using external LD, with heterogeneous datasets as in typical meta-analyses, where inconsistencies between Z-scores and external LD can lead to artificially large marginal likelihood, we propose to use an adaptive procedure to select \(\eta\) explained in the remark section below.

Marginal likelihood and posterior probability. Given the Z-scores \(Z\) and LD correlation matrix \(\Sigma\), the marginal likelihood conditional on \(\gamma_S\) is

\[
f(Z|\gamma_S) = \int_{\tau} \int_{\sigma_y^2} \int_{\lambda_S} f(Z|\lambda_S, \sigma_y^2) f(\lambda_S|\gamma_S, \tau, \sigma_y^2) f(\tau) f(\sigma_y^2) d\lambda_S d\sigma_y^2 d\tau, 
\]

where \(f(Z|\lambda_S, \sigma_y^2)\) is the density function of MVN(\(\Sigma\lambda_S, \sigma_y^2\Sigma\)), and \(f(\lambda_S|\gamma_S, \tau, \sigma_y^2)\) is the prior density function of the effect size, which is the product of \(p - |S|\) point mass distribution at 0 and the density function of MVN(0, \(\frac{\sigma_y^2}{\tau} \Sigma^{-1} \gamma_S^{-1}\)). The marginal likelihood after integrating out \(\lambda_S\) and \(\sigma_y^2\) is

\[
f(Z|\gamma_S) = \int_0^\infty \frac{\Gamma(\frac{|S|}{2}) |\Sigma|^{1/2}}{\left[ Z|\Sigma^{-1}Z - Z_S|_{S+1} \right]^{\frac{|S|}{2}} \left( \frac{1 + \tau}{\tau} \right)^{-\frac{|S|}{2}}} f(\tau) d\tau,
\]
and the ratio between the marginal likelihood of $\gamma_S$ and the null model $\gamma_0$ is

$$\frac{f(Z|\gamma_S)}{f(Z|\gamma_0)} = \int_0^\infty \left[ 1 - \frac{Z'_i \Sigma^{-1}_{\gamma_S} Z_S}{Z'_i \Sigma^{-1} Z (1+\tau)} \right]^{-\frac{p}{2}} \left( \frac{1+\tau}{\tau} \right)^{-\frac{|S|}{2}} f(\tau) d\tau.$$ 

The details can be found in the Supplemental Material. Let $M$ denote the model set that contains all candidate models. Then the posterior probability of any non-null model $\gamma_S$ and the posterior probability of $\gamma_i$ being equal to 1 (PIP) can be computed as

$$\Pr(\gamma_S|Z) = \frac{PO_{\gamma_S;\gamma_0}}{\sum_{\gamma_A \in M} PO_{\gamma_A;\gamma_0}},$$

$$\Pr(\gamma_i = 1|Z) = \sum_{\gamma_S:i \in S} \Pr(\gamma_S|Z),$$

where the posterior odds ($PO_{\gamma_S;\gamma_0}$) is defined as the product of the Bayes factor ($\frac{f(Z|\gamma_S)}{f(Z|\gamma_0)}$) and the prior odds ($\frac{\Pr(\gamma_S|\eta)}{\Pr(\gamma_0|\eta)}$):

$$PO_{\gamma_S;\gamma_0} = \frac{f(Z|\gamma_S) \Pr(\gamma_S|\eta)}{f(Z|\gamma_0) \Pr(\gamma_0|\eta)} = \frac{\eta^{|S|}(p-|S|)!}{p!} \int_0^\infty \left[ 1 - \frac{Z'_i \Sigma^{-1}_{\gamma_S} Z_S}{Z'_i \Sigma^{-1} Z (1+\tau)} \right]^{-\frac{p}{2}} \left( \frac{1+\tau}{\tau} \right)^{-\frac{|S|}{2}} f(\tau) d\tau.$$

**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. 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 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 $\gamma_S$ is proportional to a product of the marginal likelihood and the prior distribution:

$$\Pr(\gamma_S|Z) = \frac{f(Z|\gamma_S)f(\gamma_S)}{f(Z)} \propto f(Z|\gamma_S)f(\gamma_S).$$

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

$$\Gamma_-(S) := \{ A : A \subset S, |S| - |A| = 1 \} \text{ (one less SNP than } S),$$

$$\Gamma_+(S) := \{ A : A \supset A, |A| - |S| = 1 \} \text{ (one more SNP than } S),$$

$$\Gamma_{\Rightarrow}(S) := \{ A : |S| - |A \cap S| = 1, |A| = |S| \} \text{ (models that replaces one SNP in } S).$$

Then, all the unnormalized posterior probabilities of the neighborhood models, i.e., $\{\Gamma_-(S) \cup \Gamma_+(S) \cup \Gamma_{\Rightarrow}(S)\}$, 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 \( \Gamma \) the set of models selected during the algorithm.

**Remarks**

**The inversion of \( \Sigma \).** 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 \( \eta \)-prior and the Moore–Penrose inverse of the correlation matrix has been established by Maruyama and George in.\(^{36}\)

**Fixing \( \sigma^2_y \).** There are situations where the LD correlation matrix \( \Sigma \) estimated from the reference panel might not be positive definite, i.e., \( Z' \Sigma^{-1} Z < 0 \), which leads to an invalid density function \( f(Z|\tau, \gamma) \). To accommodate such a situation, one solution is to fix \( \sigma^2_y \) at a pre-determined value, such as 1, which should provide similar performance to the scenario of integrating out \( \sigma^2 \).

**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 \( \gamma(b) \), \( b = 1, \ldots, B \), denote the ranked candidate models, such as \( \gamma(1) \) receives the largest marginal likelihood. We use \( \gamma(1) \) as the reference model to select all other candidate models that are not significantly different from \( \gamma(1) \) through a Bayesian hypothesis testing procedure as follows. The posterior odds of the hypothesis test for comparing \( \gamma(1) \) to all other candidate models is

\[
PO_{\gamma(1):\gamma(b)} = \frac{Pr(\gamma(1)|Z, \eta)}{Pr(\gamma(b)|Z, \eta)} = \frac{Pr(Z|\gamma(1))Pr(\gamma(1)|\eta)}{Pr(Z|\gamma(b))Pr(\gamma(b)|\eta)},
\]

for \( b = 1, \ldots, B \). The posterior odds quantifies the strength of evidence in favor of the leading causal configuration, represented by \( \gamma(1) \) 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 \( BF_{\gamma(1):\gamma(b)} \) as the threshold for \( PO_{\gamma(1):\gamma(b)} \) to determine if a candidate model \( \gamma(b) \) is significantly different from the reference model \( \gamma(1) \) (see Table S5 and,\(^{21}\) 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.

**Adaptive procedure to select \( \eta \).** With external LD, to help limit false positives, we adaptively select \( \eta \). We assume \( \eta = p^{-a} \) and the value of \( a \) is selected in a data driven way. We want to utilize the dimensional penalization tool to limit the Shotgun algorithm from exploring the posterior space and focus more on the signals with leading Z-scores, i.e. those more likely to correspond to causal variants. At a given locus, let \( \gamma(0.99) \) denote an indicator vector of a candidate model assuming one causal SNP with the absolute value of Z-score of the assumed causal SNP being equal to \( |Z|_{(0.99)} \). Then, the posterior odds between the null model and the candidate model denoted by \( \gamma(0.99) \) can be written as:

\[
\log \left( \frac{f(Z|\gamma(0.99)) f(\gamma(0.99)|\eta = p^{-a})}{f(Z|\gamma_{null}) f(\gamma_{null}|\eta = p^{-a})} \right) = \log \left( \frac{f(Z|\gamma(0.99))}{f(Z|\gamma_{null})} \right) - (a + 1) \log (p).
\]
Note that this posterior odds also affects the probability that the Shotgun algorithm will visit model $\gamma_{(0.99)}$. We want to find the value of $\hat{a}$, such that by setting $\eta = p^{-\hat{a}}$, the posterior odds of the candidate model $\gamma_{(0.99)}$ and the null model is equal to 1, such that:

$$\hat{a} := \log \left( \frac{f(Z|\gamma_{(0.99)})}{f(Z|\gamma_{null})} \right) = (\hat{a} + 1)\log (p) \text{ for } \hat{a} \geq 0,$$

where the value of the logarithm of the Bayes factor can be computed as shown above. Therefore, the posterior odds between any other candidate model of one SNP with smaller Z-scores ($|Z| < |Z|_{(0.99)}$) and the null model is significantly decreased, which consequentially reduces the possibility of the Shotgun algorithm visiting candidate models with medium or low signals.

**Incorporating functional annotations.** Leveraging functional annotation data on genetic variants at a locus of interest may improve the results of fine-mapping studies.$^5,^{11}$ 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:

$$L(\theta; Z, \gamma, W) = f(Z|\gamma)\Pr(\gamma|W, \theta).$$

Conventionally, the prior probability $\Pr(\gamma|W, \theta)$ is modeled using logistic regression. Since the indicator vector $\gamma$ 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 $\Gamma$ denote a truncated model space defined as a set of indicator vectors, such as $\Gamma = \{\gamma_1, \ldots, \gamma_B\}$. We assume that the indicator vectors included in $\Gamma$ represent the top $B$ candidate models in terms of the posterior probability, i.e., $\min_{\gamma \in \Gamma} \Pr(\gamma|Z, W, \theta) > \max_{\gamma \in \Gamma} \Pr(\gamma|Z, W, \theta)$. 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 $\Gamma$ approximately holds.

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

$$\ell(\theta; Z, G, W) = \log (f(Z|G)f(G|W, \theta))$$

$$= \log (f(Z|G)) + \sum_{i=1}^{p} [G_iw_i^T\theta - \exp\{w_i^T\theta\} - \log (G_i!)]$$

$$\propto \sum_{i=1}^{p} [G_iw_i^T\theta - \exp\{w_i^T\theta\}] . \tag{1}$$

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

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

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

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

- **M-step** Maximize the penalized log-likelihood as,

$$\theta^{(s+1)} := \arg \max_{\theta \in \mathbb{R}^{q+1}} \sum_{i=1}^{p} \left[ g_i^{(s)} w_i' \theta - \exp \{ w_i' \theta \} \right] - \frac{(1 - \alpha)}{2}\|\theta\|^2 - \alpha\|\theta\|.$$ 

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

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

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

$$\hat{\Pr} \left( \gamma_i = 1 | w_i, \theta^{(s+1)} \right) = \frac{\exp \left\{ w_i' \theta^{(s+1)} \right\}}{B^{(s)}},$$ (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 $g^{(s+1)}$, which depends on $Z, W$, and $\theta^{(s+1)}$.

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. 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 $\theta$. 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 $\eta$, where $\eta$ 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

$$\hat{\Pr}(\gamma_i = 1|\mathbf{w}_i, \theta^{(s+1)}) = \exp\left\{\frac{1}{p} \sum_{i=1}^{p} \log\left(\frac{\exp\left\{\mathbf{w}_i^T \theta^{(s+1)}\right\}}{B^{(s)}}\right)\right\}$$

$$= \exp\left\{\frac{1}{p} \left[ \sum_{i=1}^{p} \mathbf{w}_i^T \theta^{(s+1)} \right] - \log\left(B^{(s)}\right)\right\}$$

$$= \exp\left\{\frac{1}{p} \left[ p\theta_1^{(s+1)} \right] - \log\left(B^{(s)}\right)\right\}$$

$$= \exp\left\{\theta_1^{(s+1)} \right\} \frac{\eta B^{(s)}}{\eta + p},$$

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 $\mathbf{w}_i$ 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 $\eta$, we adjust the value of the intercept term by setting the two geometric means to be equal, such that

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

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

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

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 $\eta$. As previously mentioned, a rule of thumb to set up the value of $\eta$ is $\eta \propto p^{-a}$, for $a \geq 0$ as discussed in.\textsuperscript{34,37,38} We use the adaptive procedure of choosing $\eta$ when using external LD. We show more details on the choice of $\eta$ in the Supplemental Material.

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 $\rho_{\text{outlier}}$ is typically set at 0.9 or 0.99. Let $\tilde{Z} = \{Z_1, \ldots, Z_{|D|}\}'$ 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 $\tilde{Z}$ follow a $\text{MVN}(1_{|D|}\beta, \Sigma_D)$, where $1_{|D|}$ is the $|D|$-dimensional vector of 1, $\beta$ is a scalar, and $\Sigma_D$ 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 $\tilde{Z}_{-1} = \{Z_2, \ldots, Z_{|D|}\}'$ denote the vector of summary statistics for the SNPs in $D$, excluding SNP 1, and we assume that $\tilde{Z}_{-1} = \{Z_2, \ldots, Z_{|D|}\}'$ are not outliers and follow a $\text{MVN}(1_{|D|-1}\beta, \Sigma_{D_{-1}})$ (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

$$B_{0:1} = \frac{m^0(\tilde{Z}_{-1}|c = 1)}{\int m^1(\tilde{Z}_{-1}|c)g(c)dc},$$

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 Supplemental Material). Following the robust Bayesian approach discussed in, we can compute

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

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

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

where $a = \frac{|Z_1 - E[Z_1|\tilde{Z}_{-1}]|}{\sqrt{\text{Var}[Z_1|\tilde{Z}_{-1}]}}$ is the distance between $Z_1$ and the conditional mean $E[Z_1|\tilde{Z}_{-1}]$, in unit of the conditional s.d. $\text{Var}[Z_1|\tilde{Z}_{-1}]$. 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 Supplemental Material.

**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 $\gamma_S$ 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 $\gamma_S$.

Input: The index set $S = \{s_1, \ldots, s_{|S|}\}$ for the current model $\gamma_S$, the threshold on the Bayes factor $\delta$, and the threshold on the correlation $\rho_{\text{outlier}}$.

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 $\tilde{Z} = \{Z_1, \ldots, Z_{|D|}\}'$ 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{align*}
      H_0 : Z_d &\sim N(\beta, 1); \\
      H_1 : Z_d &\sim N(\beta, c), \ c \neq 1;
      \end{align*}
      where $Z_d$ is not an outlier $Z_d$ is an outlier.
    - Compute the corresponding Bayes factor $\hat{B}_d$ conditional on $\tilde{Z}_{D-d}$ and $\Sigma_{D-d}$.
  if $\exists d \in \{1, \ldots, |D|\}, \hat{B}_d < \delta$ then
    - Drop $Z_d$, where $\hat{B}_d = \min \left(\hat{B}_1, \ldots, \hat{B}_{|D|}\right)$, from the fine-mapping computation.
    - Drop $Z_d$ from $\tilde{Z}$, i.e., $\tilde{Z} = \{Z_1, \ldots, Z_{d-1}, Z_{d+1}, \ldots, Z_{|D|}\}'$.
    - Drop $d$th 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,\textsuperscript{5,11,40} 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 $\eta_l$ 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 the authors define a credible set as follows.

Definition 1. A level $\rho$ credible set is defined to be a subset of correlated variables (with correlation within the set greater than some threshold $r$) that has probability $\rho$ 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

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

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

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

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 $\rho$, 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 $\rho$ credible sets (those sets with posterior probability greater than $\rho$) is selected along with the corresponding credible sets.

Data Availability. The manuscript used summary statistics from a meta-analysis Alzheimer’s Disease GWAS available at: https://ctg.cnmc.nl/software/summary_statistics.

References


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 in the middle.
Figure 2: Effect of functional annotation on AUROC, AUPR, credible sets, and credible models (in-sample LD). (a) Mean AUROC and AUPR values based on PIPs from CARMA, SuSiE, and fastPAINTOR for no functional annotation vs. including functional annotations. (b) Performance of credible sets ($\rho = 0.99$). Power: the proportion of simulated causal variants included in any credible set per locus; Coverage: the proportion of credible sets that contain a causal variant per locus; Size: the average number of variants included in a credible set; Purity: the average squared correlation of variants in a credible set. (c) Power is the percentage of the simulated causal variants identified by the credible models or the credible sets. Size is the average number of variants included in the credible models (CARMA) or all the credible sets (SuSiE, fastPAINTOR) for a locus. The credible models are computed based on a threshold of 10 for the posterior odds.
Figure 3: **Power and FDR using positive predictions.** (a) Comparison between no annotation and including annotations with in-sample LD. (b) Comparison between original and spike-in data. FDR vs. power for different methods using positive predictions as the PIP threshold varies from 0.1 to 1. These quantities are calculated as $\text{FDR} := \frac{FP}{TP+FP}$ and $\text{power} := \frac{TP}{TP+FN}$, where FP, TP, FN, TN denote the number of false positives, true positives, false negatives and true negatives respectively given a certain PIP threshold. Open circles denote the results at PIP threshold of 0.5, and solid circles denote the results at PIP threshold of 0.95.
Figure 4: Effect of external LD on AUROC, AUPR, credible sets, and credible models. (a) Mean AUROC and AUPR values based on PIPs from the proposed model, SuSiE, and fastPAINTOR for no functional annotation vs. including functional annotations. (b) Performance of credible sets ($\rho = 0.99$). Power: the proportion of simulated causal variants included in a credible set per locus; Coverage: the proportion of credible sets that contain a causal variant per locus; Size: the average number of variants included in a credible set; Purity: the average squared correlation of variants in a credible set. (c) Power is the percentage of the simulated causal variants identified by the credible models or the credible sets. Size is the average number of variants in the credible models or the credible sets for a locus. The credible models are computed based on a threshold of 10 for the posterior odds.
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 three different scenarios, 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 \textit{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 heterogeneous (complete) dataset.
Figure 8: Performance with respect to credible sets for the three models across 30 Alzheimer’s disease risk loci. (a) homogeneous dataset, and (b) heterogeneous dataset. For each scenario we report: average absolute value of the summary statistics within each credible set per locus; average standard deviation of the absolute value of summary statistics within each credible set per locus; average size of the credible sets; average number of credible sets per locus. Also reported is the average number of single SNP credible sets per locus for each model. 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.
<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Z</th>
<th>N</th>
<th>EAF</th>
<th>Hom</th>
<th>Hom_Annot</th>
<th>Het</th>
<th>Het_Annot</th>
<th>Gene</th>
<th>V2G</th>
<th>mQTL</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS4</td>
<td>rs4575998</td>
<td>6.37</td>
<td>434326</td>
<td>0.23</td>
<td>0.70</td>
<td>0.70</td>
<td>0.71</td>
<td>0.73</td>
<td>NDUFS2</td>
<td>0.23</td>
<td>NDUFS2</td>
<td>7.7e-28</td>
</tr>
<tr>
<td>BIN1</td>
<td>rs4663105</td>
<td>14.01</td>
<td>425134</td>
<td>0.41</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>BIN1</td>
<td>0.26</td>
<td>BIN1</td>
<td>1.0e-04</td>
</tr>
<tr>
<td>INPPD5</td>
<td>rs10933431</td>
<td>-6.15</td>
<td>436498</td>
<td>0.24</td>
<td>0.38</td>
<td>0.39</td>
<td>0.50</td>
<td>0.47</td>
<td>INPPD5</td>
<td>0.32</td>
<td>INPPD5</td>
<td>1.5e-05</td>
</tr>
<tr>
<td>CD2AP</td>
<td>rs9381563</td>
<td>6.36</td>
<td>429204</td>
<td>0.36</td>
<td>0.75</td>
<td>0.73</td>
<td>0.76</td>
<td>0.81</td>
<td>CD2AP</td>
<td>0.20</td>
<td>CD2AP</td>
<td>3.9e-14</td>
</tr>
<tr>
<td>ZCWPW1</td>
<td>rs1859788</td>
<td>-7.95</td>
<td>441735</td>
<td>0.32</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.41</td>
<td>PILRA</td>
<td>0.35</td>
<td>PILRA</td>
<td>9.7e-44</td>
</tr>
<tr>
<td>ECHDC3</td>
<td>rs11257238</td>
<td>5.72</td>
<td>430330</td>
<td>0.36</td>
<td>0.28</td>
<td>0.29</td>
<td>0.35</td>
<td>0.34</td>
<td>USP6NL</td>
<td>0.12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MS4A6A</td>
<td>rs2081545</td>
<td>-8.01</td>
<td>441064</td>
<td>0.38</td>
<td>0.56</td>
<td>0.55</td>
<td>0.43</td>
<td>0.43</td>
<td>MS4A6A</td>
<td>0.25</td>
<td>MS4A6A</td>
<td>3.6e-08</td>
</tr>
<tr>
<td>PICALM</td>
<td>rs677869</td>
<td>-8.75</td>
<td>432250</td>
<td>0.32</td>
<td>0.26</td>
<td>0.26</td>
<td>0.30</td>
<td>0.30</td>
<td>PICALM</td>
<td>0.31</td>
<td>PICALM</td>
<td>1.0e-03</td>
</tr>
<tr>
<td>SORL1</td>
<td>rs11218343</td>
<td>-6.84</td>
<td>436249</td>
<td>0.04</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>SORL1</td>
<td>0.17</td>
<td>SORL1</td>
<td>5.4e-03</td>
</tr>
<tr>
<td>APH1B</td>
<td>rs117618017</td>
<td>5.52</td>
<td>444006</td>
<td>0.12</td>
<td>0.58</td>
<td>0.57</td>
<td>0.88</td>
<td>1.00</td>
<td>APH1B</td>
<td>0.42</td>
<td>APH1B</td>
<td>2.2e-03</td>
</tr>
<tr>
<td>SCIMP</td>
<td>rs113260531</td>
<td>6.15</td>
<td>435724</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.18</td>
<td>0.17</td>
<td>RABEP1</td>
<td>0.24</td>
<td>RABEP1</td>
<td>2.9e-04</td>
</tr>
<tr>
<td>ABI3</td>
<td>rs28394864</td>
<td>5.64</td>
<td>430812</td>
<td>0.45</td>
<td>0.38</td>
<td>0.37</td>
<td>0.28</td>
<td>0.45</td>
<td>GNGT2</td>
<td>0.24</td>
<td>ZNF652</td>
<td>1.6e-03</td>
</tr>
<tr>
<td>SUZ12P1</td>
<td>rs112659010</td>
<td>-3.56</td>
<td>421059</td>
<td>0.08</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>RNF121</td>
<td>1.4e-02</td>
</tr>
<tr>
<td>ALPK2</td>
<td>rs76726049</td>
<td>5.52</td>
<td>433227</td>
<td>0.01</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>MALT1</td>
<td>0.17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CD33</td>
<td>rs3865444</td>
<td>-5.84</td>
<td>436498</td>
<td>0.30</td>
<td>0.44</td>
<td>0.45</td>
<td>0.46</td>
<td>0.44</td>
<td>CD33</td>
<td>0.38</td>
<td>SIGLECL1</td>
<td>1.0e-13</td>
</tr>
<tr>
<td>CASS4</td>
<td>rs6014724</td>
<td>-6.21</td>
<td>433858</td>
<td>0.10</td>
<td>0.28</td>
<td>0.28</td>
<td>0.30</td>
<td>0.29</td>
<td>CASS4</td>
<td>0.27</td>
<td>CASS4</td>
<td>2.9e-07</td>
</tr>
</tbody>
</table>

Table 1: Loci where the same top SNP (with largest PIP) is identified by CARMA across all scenarios. Z: meta-analysis Z-score (the same for the homogeneous and the heterogenous datasets); N: number of samples in the homogeneous datasets; EAF: effect allele frequency; PIPs from the CARMA model across the four scenarios are reported (homogeneous/heterogeneous with or without functional annotations; Gene: putative causal gene. V2G: variant-to-gene score from Open Targets; meQTL data from ROSMAP. Brain Dorsolateral Prefrontal Cortex. CpG sites (with the smallest p-value for each fine-mapped SNP) are assigned to genes based on shortest distance to gene footprint.
<table>
<thead>
<tr>
<th>Method</th>
<th>Prior on effect size</th>
<th>Prior on model space</th>
<th>Max#causal</th>
<th>Functional annot</th>
<th>Computation</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINEMAP</td>
<td>Normal-Gamma</td>
<td>discrete Uniform</td>
<td>Fixed</td>
<td>No</td>
<td>Shotgun</td>
<td>9</td>
</tr>
<tr>
<td>CAVIARBF</td>
<td>Normal-Gamma</td>
<td>discrete probability</td>
<td>Fixed</td>
<td>Yes</td>
<td>Exhaustive</td>
<td>40</td>
</tr>
<tr>
<td>JAM</td>
<td>$g$-prior</td>
<td>Beta-Binomial</td>
<td>Unlimited</td>
<td>No</td>
<td>Exhaustive and MCMC</td>
<td>8</td>
</tr>
<tr>
<td>fastPAINTOR</td>
<td>Normal-Gamma</td>
<td>discrete probability</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Exhaustive and MCMC</td>
<td>5</td>
</tr>
<tr>
<td>SuSiE (+PolyFun)</td>
<td>Normal-Gamma</td>
<td>Multinomial (discrete probability)</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Variational Bayes</td>
<td>4, 11</td>
</tr>
<tr>
<td>CARMA (proposed model)</td>
<td>Mixture of $g$-prior</td>
<td>Poisson</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Shotgun</td>
<td></td>
</tr>
</tbody>
</table>

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