credible sets or examine joint posterior probabilities to assess whether different signals are likely to share the same causal variant. In this sense, fine-mapping is not merely a tool for improving localization accuracy within a genomic region; it also serves to transmit uncertainty into subsequent analyses, enabling functional annotation integration, colocalization analysis, and cross-omics research to proceed on a unified probabilistic scale.

In addition, because PIP is fundamentally defined in probability space and does not depend on the form of a specific statistical measure, such as a p-value or effect size, it is more readily comparable across different study designs, populations, and omics data types. This property allows it to serve as a common language for expressing causal information across data domains and provides a methodological basis for cross-study integration (Shrestha et al., 2024). This is particularly valuable in multi-ancestry studies, where differences in LD structure across populations can be regarded as a naturally occurring additional constraint. Different LD patterns alter the correlations between non-causal variants and true causal loci, thereby helping to further shrink credible sets and improve causal resolution (Spain and Barrett, 2015). This feature also gives fine-mapping greater practical value in multi-ancestry research and naturally extends to the development of subsequent methods such as MsCAVIAR.

3 fastPAINTOR: An Annotation-Informed Model for Causal Probability

3.1 Model framework: joint inference of statistical evidence and functional annotation

Within the unified Bayesian framework of fine-mapping, the key distinction among methods lies not in whether probabilistic inference is performed, but in how the prior distribution over causal configurations is specified. From this perspective, fastPAINTOR can be understood as a prototypical *annotation-informed causal model*, whose central objective is to establish a probabilistic mapping between GWAS statistical evidence and functional annotations.

In general form, the fine-mapping problem can be expressed as estimating the posterior distribution over the causal configuration vector γ:

$$p(\gamma|D,R,A) \propto p(D|\gamma,R)\, p(\gamma|A)$$

From the perspective of statistical genetics, a basic fine-mapping model characterizes the distribution of causal probabilities under LD constraints, whereas fastPAINTOR further describes an annotation-informed causal probability distribution conditional on functional annotations, thereby giving its inferential target stronger biological interpretability. To estimate the posterior distribution within this framework, fastPAINTOR adopts approximate Bayesian methods, such as variational inference or importance sampling, in order to avoid the exponential enumeration of the full causal configuration space or the high computational cost associated with traditional Markov chain Monte Carlo (MCMC) approaches (Talukdar et al., 2023). Its optimization objective can be expressed as the evidence lower bound (ELBO):

$$\Pr(\gamma_j=1|a_j)=\text{logit}^{-1}(\alpha_0+\alpha^{\top}a_j)$$

where a_j denotes the annotation vector for SNP j (e.g., eQTL status, chromatin accessibility, transcription factor binding sites). This formulation effectively transforms biological functional potential into statistical prior probability, enabling a principled integration of functional and association evidence (Kichaev et al., 2014).

From the perspective of statistical genetics, a basic fine-mapping model characterizes the distribution of causal probabilities under LD constraints, whereas fastPAINTOR further describes an annotation-informed causal probability distribution conditional on functional annotations, thereby giving its inferential target stronger biological interpretability. To estimate the posterior distribution within this framework, fastPAINTOR adopts approximate Bayesian methods, such as variational inference or importance sampling, in order to avoid the exponential enumeration of the full causal configuration space or the high computational cost associated with traditional Markov chain Monte Carlo (MCMC) approaches (Talukdar et al., 2023). Its optimization objective can be expressed as the evidence lower bound (ELBO):

$$\mathcal{L}(q)=\mathrm{E}_{q(\gamma)}[\log p(D|\gamma,R)+\log p(\gamma|A)]-\mathrm{E}_{q(\gamma)}[\log q(\gamma)]$$