Colocalization analysis further examines causal consistency through joint inference across phenotypes. When combined with TWAS-based integration at the gene level, this workflow ultimately forms a complete inferential chain from GWAS to fine-mapping, then to colocalization and TWAS, and finally to candidate gene identification:

GWAS→fine-mapping→colocalization→TWAS→candidate gene identification

This continuous analytical path indicates that the focus of statistical genetics has moved beyond the identification of association signals and is gradually shifting toward the systematic interpretation of molecular mechanisms (Hormozdiari et al., 2016; Okamoto et al., 2023).

6 Practical Guidelines for Method Selection: A Layer-Based Decision Framework

6.1 A unified view of trade-offs: computation, identifiability, and information sources

In fine-mapping and cross-omics analysis, method selection is not simply a matter of tool preference, but rather the result of a systematic trade-off among different inferential strategies. At its core, this trade-off usually involves three mutually constraining aspects: computational complexity, causal resolution or identifiability, and the structure of information sources, including LD, functional annotations, and cross-study data. Within a unified Bayesian framework, the differences among methods can essentially be understood as different ways of imposing constraints on the same causal configuration space by drawing on different types of information (Schaid et al., 2018; Hutchinson et al., 2020).

(1) FastPAINTOR: improving resolution through prior information
The core of fastPAINTOR lies in its use of functional annotations to construct informed priors, thereby increasing the identifiability of causal variants while keeping the likelihood structure, that is, the LD structure, unchanged (Kichaev et al., 2014, 2016). The strengths of this method are mainly reflected in two respects. On the one hand, it can operate efficiently at the genome-wide scale and is therefore suitable for large-scale analyses. On the other hand, when functional annotation information is sufficiently informative, it can substantially shrink the credible set and thus improve the ranking resolution of candidate causal variants. Previous studies have shown that integrating functional annotations not only reduces the size of the candidate set but also improves the prioritization of causal variants (Kichaev et al., 2016; Zou et al., 2021).

However, this very advantage also defines its limitation. The inferential results of fastPAINTOR are highly sensitive to the quality of annotations. Once functional annotations contain noise, missingness, or systematic bias, these defects may be propagated into the posterior inference and introduce systematic error. From a statistical perspective, fastPAINTOR is therefore essentially an annotation-informed prior model that improves causal resolution by relying on prior information.

(2) CAVIAR: robustness through likelihood modeling
Unlike fastPAINTOR, CAVIAR does not rely on functional annotations, but instead distinguishes correlated signals from causal signals by explicitly modeling LD structure at the likelihood level (Schaid et al., 2018). This characteristic makes it less dependent on external information and therefore generally more robust when functional annotations are sparse or of uncertain reliability. At the same time, CAVIAR allows multiple causal variants to exist within a region, giving it considerable flexibility in handling complex genetic architectures.

This robustness, however, comes at a cost. Because CAVIAR must evaluate high-dimensional causal configuration spaces, its computational complexity increases rapidly as the number of candidate SNPs and the number of potential causal variants rise. CAVIAR is therefore best understood as a likelihood-driven model that emphasizes robustness in causal inference.

(3) MsCAVIAR: enhancing identifiability through cross-study information
MsCAVIAR extends CAVIAR into cross-study or cross-population settings. By jointly modeling summary signals from multiple studies together with their respective LD structures, it further improves the identification of causal variants (Lapierre et al., 2020). Its most important advantage lies in its ability to exploit differences in LD patterns