Bioscience Evidence 2026, Vol.16, No.3, 186-201
http://bioscipublisher.com/index.php/be

Research Report Open Access

The Causal Inference Layer in Complex Trait Genetics: A Unified Statistical Framework from Fine-Mapping to Cross-Trait Integration

Xuanjun Fang
Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025, Hainan, China
Corresponding email: xuanjunfang@hitar.org

Bioscience Evidence, 2026, Vol.16, No.3 doi: 10.5376/be.2026.16.0015
Received: 22 May, 2026
Accepted: 22 Jun., 2026
Published: 30 Jun., 2026

Preferred citation for this article:
Fang X.J., 2026, The causal inference layer in complex trait genetics: a unified statistical framework from fine-mapping to cross-trait integration, Bioscience Evidence, 16(3): 186-201 (doi: 10.5376/be.2026.16.0015)

Abstract Genome-wide association studies (GWAS) have identified a large number of loci associated with complex traits and diseases. However, most of these signals arise from linkage disequilibrium (LD) rather than directly reflecting causal variants, thereby limiting their mechanistic interpretability. Probabilistic fine-mapping addresses this limitation by introducing posterior inclusion probabilities (PIPs) and credible sets, shifting the inferential target from a single significant locus to a distribution of causal probabilities and enabling a systematic characterization of genetic uncertainty. In recent years, with the growing availability of functional annotation data, multi-ancestry studies, and multi-omics resources, fine-mapping methods have continued to expand in both model architecture and application scope. Nevertheless, a unified theoretical perspective across these methods remains lacking. In this study, we develop a unified statistical framework centered on causal configurations by systematically integrating fine-mapping and colocalization analyses within a Bayesian inferential framework. Under this framework, fastPAINTOR constructs annotation-informed priors using functional annotations, whereas CAVIAR and its extension MsCAVIAR strengthen likelihood-based constraints through LD structure and cross-study information. Colocalization analysis further extends the inferential target from a single-trait setting to a multi-trait space, enabling probabilistic modeling of cross-trait causal consistency. Accordingly, research on complex traits can be organized into a continuous inferential pipeline from GWAS to fine-mapping, and further to colocalization and transcriptome-wide association studies (TWAS), thereby progressively translating statistical associations into biological mechanistic interpretation. On this basis, we further propose a method-selection strategy based on inferential hierarchy, clarifying the trade-offs among computational complexity, causal resolution, and information sources across different methods, and summarizing a practical workflow of "hierarchical inference." This framework is applicable not only to studies of human complex diseases, but also to applied contexts such as crop genetic improvement, where it can be used to assess causal consistency across environments and populations. By unifying fine-mapping and colocalization within the same causal inference layer, this study provides statistical genetics with a consistent conceptual language and analytical paradigm, thereby facilitating the systematic transition of complex trait research from association discovery to mechanistic interpretation and causal inference.

Keywords Complex traits; Fine-mapping; Causal inference; Colocalization analysis; Credible sets; Multi-omics integration

1 Introduction

Genome-wide association studies (GWAS) have substantially advanced the dissection of the genetic architecture of complex traits and diseases over the past decade, leading to the identification of numerous genomic regions associated with phenotypic variation in both human populations and crop systems (Lapierre et al., 2020). However, the evidence provided by GWAS is primarily statistical association rather than direct causal interpretation. Because linkage disequilibrium (LD) is widespread across the genome, significantly associated loci often function only as marker SNPs correlated with the true functional variants, rather than representing the actual causal variants with biological effects. At the same time, multiple independent causal signals may coexist within a single associated region, making the assumption that the most significant SNP directly corresponds to the causal variant statistically untenable in many cases (Hutchinson et al., 2020). Accordingly, how to distinguish causal signals from correlated signals within associated regions has become an important issue linking population genetic analysis with functional genomic interpretation.

From the perspective of the development of statistical genetics, genomic analysis of complex traits has not been built upon a single statistic, but has instead progressed around a series of different inferential targets. GWAS