


Research Report

Open Access

A Unified Framework for Causal Inference in Statistical Genetics: Integrating GWAS, Molecular QTL, Colocalization, and Mendelian RandomizationXuanjun Fang 

Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025, Hainan, China

 Corresponding email: xuanjunfang@hitar.orgPlant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0011](https://doi.org/10.5376/pgt.2026.17.0011)

Received: 30 Mar., 2026

Accepted: 26 Apr., 2026

Published: 20 May, 2026

Copyright © 2026 Fang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**Preferred citation for this article:**Fang X.J., 2026, A unified framework for causal inference in statistical genetics: integrating GWAS, molecular QTL, colocalization, and Mendelian randomization, Plant Gene and Trait, 17(3): 156-172 (doi: [10.5376/pgt.2026.17.0011](https://doi.org/10.5376/pgt.2026.17.0011))

Abstract Genome-wide association studies (GWAS) have identified thousands of loci associated with complex traits, yet translating these statistical signals into biological mechanisms remains a major challenge. A key difficulty lies in distinguishing between association, shared genetic architecture, and causal relationships across multiple layers of molecular regulation. In this study, we present a unified analytical framework for causal inference in statistical genetics that integrates GWAS, molecular quantitative trait loci (QTL), transcriptome-wide association studies (TWAS), colocalization analysis, and Mendelian randomization (MR). Within this framework, different methods address distinct inferential targets: GWAS identifies variant–trait associations; molecular QTL and TWAS link genetic variation to intermediate phenotypes; colocalization evaluates the consistency of signals across datasets; and MR estimates the direction and magnitude of potential effects under explicit assumptions. We emphasize that these components should not be interpreted in isolation but as part of a sequential process of evidence refinement. In particular, colocalization is necessary for prioritizing candidate mechanisms but does not establish causality, while MR provides effect estimates that remain sensitive to instrument validity, pleiotropy, and data heterogeneity. We further discuss practical considerations for implementation, including instrument selection, diagnostic evaluation, and cross-population validation, as well as challenges arising from pleiotropy, tissue specificity, and environmental interactions. Finally, we extend this framework to plant systems and emerging multi-omics contexts, highlighting the role of single-cell and epigenomic data in refining causal interpretation. By clarifying the roles and limitations of individual methods within an integrated framework, this study provides a structured approach for moving from genetic associations toward biologically interpretable and experimentally testable hypotheses.

Keywords Statistical genetics; Causal inference; Genome-wide association study (GWAS); Molecular QTL (eQTL, sQTL, pQTL); Transcriptome-wide association study (TWAS); Colocalization; Mendelian randomization; Multi-omics integration; Pleiotropy; Complex traits

1 Introduction

Genome-wide association studies (GWAS) have generated an unprecedented scale of statistical associations in complex trait genetics. However, these signals fundamentally represent association estimands—statistical relationships between genetic variants and traits—rather than direct evidence of biological causality. This distinction underlies a central gap in the field: many GWAS loci reside in non-coding regions and are influenced by linkage disequilibrium (LD) and multilayer regulatory architectures, such that a single association peak often implicates multiple candidate genes and mechanisms. Even when fine-mapping reduces signals to smaller credible sets, the resulting inference remains a causal probability estimand, rather than a direct estimate of causal direction or effect (Liu et al., 2019; Wainberg et al., 2019; Xie et al., 2021; Mostafavi et al., 2023).

From a unified statistical genetics perspective, the analysis of complex traits can be conceptualized as a multi-layer inferential chain defined by distinct statistical targets (estimands): GWAS characterizes association evidence, fine-mapping quantifies posterior probabilities of causality, and polygenic risk scores (PRS) translate these signals into individual-level predictive functionals. Yet a critical gap remains in this chain: how to move from causal probability to causal pathways and causal effect estimands. This gap defines the role of functional integration and causal inference methods.