One common scenario is horizontal pleiotropy, in which a single variant affects multiple traits through independent pathways. In such cases, colocalization may still detect a shared signal, even though no direct mediation relationship exists between the molecular phenotype and the complex trait (Rasooly et al., 2022). Evidence from epigenomic studies further suggests that many disease-associated loci involve complex regulatory architectures with multiple parallel pathways (Shikov et al., 2020; Boix et al., 2021; Khan et al., 2024).

LD structure and allelic heterogeneity can also complicate interpretation, particularly in regions with multiple signals where the true causal variant may be obscured by correlated variants (Wallace, 2021). For these reasons, colocalization is best viewed as a filtering step rather than a definitive test.

In analytical workflows, loci with strong colocalization support are often prioritized for further evaluation using complementary approaches. For example, Mendelian randomization can be applied to assess the direction and magnitude of potential effects, whereas loci with weak or inconsistent evidence may require re-examination at the level of fine-mapping or data harmonization before further interpretation.

4.3 Applications in plant systems

In plant systems, colocalization analysis has proven useful for disentangling tissue-specific and environment-dependent regulatory effects. Studies across multiple tissues and environmental conditions have shown that the effects of regulatory variants can vary substantially depending on developmental stage or external stimuli, leading to context-dependent contributions to complex traits.

For instance, analyses in crops and model plants have identified distinct regulatory patterns across organs such as leaves, roots, and fruits, as well as across developmental stages. Work in tomato and other species has further demonstrated that conserved developmental genes may exhibit pleiotropic effects, while still showing variation in regulatory behavior across environments or species (Hendelman et al., 2021).

In practice, it is often beneficial to construct eQTL maps in trait-relevant tissues and under representative environmental conditions, followed by stratified colocalization analyses across these contexts. The use of multi-parent populations, such as NAM or MAGIC, can improve resolution by reducing LD and helping to distinguish multiple signals. In polyploid crops, additional attention is required to differentiate homologous gene copies and to accurately quantify expression, in order to minimize ambiguity in gene assignment.

Candidate loci identified through colocalization can then be further evaluated using downstream approaches, including Mendelian randomization, near-isogenic lines, genome editing, and expression assays, ultimately contributing to a more complete characterization of the relationship between genetic variation and phenotypic traits.

5 Mendelian Randomization as A Framework for Causal Inference

Within integrative analysis pipelines, Mendelian randomization (MR) is typically applied at a later stage, once genetic associations and molecular evidence have been established. Its primary purpose is to evaluate the direction and magnitude of potential effects by using genetic variants as instruments and treating molecular traits-such as gene expression-as exposures.

Compared with earlier steps, which focus on mapping or aligning signals across datasets, MR aims to quantify relationships under a set of assumptions. As a result, the interpretation of MR findings depends critically on how instruments are selected and on the plausibility of the underlying assumptions.

5.1 Core assumptions of instrumental variables

MR analyses rely on three basic conditions. First, the selected genetic variants must be sufficiently associated with the exposure, ensuring that they carry informative signal. Second, these variants should be independent of confounding factors, an assumption that is partly justified by the approximate random allocation of alleles at conception. Third, the effect of the instruments on the outcome should operate primarily through the exposure of interest, rather than through alternative pathways.