assess whether signals from different sources are compatible with a shared genetic origin, thereby reducing the likelihood of advancing non-causal genes into downstream analyses. Under these conditions, effect estimates obtained from MR become more interpretable in a biological context (Porcu et al., 2019).

With the increasing availability of diverse datasets, this integrative process has extended to multi-tissue and single-cell contexts, allowing regulatory effects to be examined at finer spatial and contextual resolution. For example, cell type–specific regulatory mechanisms have been shown to play a central role in certain disease processes, highlighting the value of incorporating such information into integrative analyses (Gleason et al., 2021).

At the implementation level, the choice of instruments remains a critical factor. Prioritizing cis-regulatory variants, combined with ancestry-matched LD reference panels and tissue-specific models, can improve stability. The use of multiple estimation methods alongside systematic diagnostic procedures further strengthens the robustness of findings (Hemani et al., 2018; Hu et al., 2022). In settings involving multiple molecular layers or tissues, multivariable models can help disentangle overlapping signals and avoid attributing shared regulation to a single pathway (Zuber et al., 2022).

7.2 The impact of pleiotropy and heterogeneity

Despite these advances, integrative analyses remain sensitive to pleiotropy and heterogeneity. Horizontal pleiotropy presents a major challenge, as a single genetic variant may influence multiple traits through independent pathways, complicating interpretation based on a single assumed mechanism. In such cases, effect estimates may deviate from the underlying biological process (Hemani et al., 2018).

Differences across datasets-such as mismatches in tissue relevance, population structure, or environmental context-can further contribute to variability in results. Even when colocalization indicates concordance between signals, this may reflect shared genetic architecture rather than a specific mechanistic pathway (Zuber et al., 2022). Consequently, colocalization should be interpreted as a filtering step rather than as evidence of causality.

Addressing these challenges requires both careful instrument selection and the use of complementary analytical strategies. Restricting analyses to locally acting variants and accounting for LD structure can reduce confounding. At the same time, comparing results across multiple methods helps identify inconsistencies that may indicate violations of assumptions. Diagnostic and sensitivity analyses play a key role in detecting influential or invalid instruments and in assessing the robustness of conclusions (Hu et al., 2022).

Recent methodological developments have sought to model pleiotropy and heterogeneity explicitly, particularly in multi-tissue and multi-context settings. While these approaches show promise, their performance remains dependent on data quality and appropriate model specification (Gleason et al., 2021; Lu et al., 2024).

7.3 Extensions to plant systems

In plant systems, integrative analyses face additional layers of complexity. Environmental effects often play a dominant role, such that the same genetic variant may exhibit different effects across developmental stages, tissues, or stress conditions. In addition, genomic features such as extended LD, structural variation, and polyploidy complicate the interpretation of association signals.

These characteristics necessitate adjustments to analytical strategies. For example, regulatory maps should ideally be constructed under trait-relevant tissues and environmental conditions, with stratified analyses used to capture context-dependent effects. Instrument selection must also account for gene copy number and homology, to avoid ambiguity in signal assignment (Porcu et al., 2019).

Population design is another important consideration. Multi-parent populations can improve resolution by reducing LD and enabling better separation of multiple signals. Replication across environmental conditions helps identify stable regulatory relationships, while multivariable approaches can be used to separate baseline and