

appropriate LD reference panels. When these conditions are not met, the scope of interpretation becomes more limited.

During candidate prioritization, locally acting signals are often easier to interpret and are therefore commonly prioritized. In contrast, distal regulatory signals typically require additional supporting evidence, such as network-based analyses or multi-step approaches, to support interpretation.

Functional annotation also plays an important role. When candidate variants align with known regulatory features-such as open chromatin regions or transcription factor binding sites-the biological plausibility of the findings is strengthened. Conversely, when evidence is limited or multiple signals coexist within a region, additional refinement or data integration may be required (Hormozdiari et al., 2016).

Under favorable conditions-where molecular QTL signals align with GWAS results and suitable instruments can be identified-further analyses can be carried out, with primary estimates reported alongside sensitivity analyses using complementary methods (Porcu et al., 2019; Zuber et al., 2022). When evidence is inconsistent, however, it is often preferable to revisit earlier steps rather than proceed with interpretation.

The final output of such analyses typically combines statistical results with diagnostic measures and functional annotations, allowing candidate genes to be ranked according to the strength of evidence and prioritized for experimental validation (Figure 3) (Votava and Parks, 2021; Lessard et al., 2024).

Image caption: This figure illustrates an integrated analytical pathway for causal inference in statistical genetics. The workflow begins with GWAS signals, followed by fine-mapping to refine candidate variants. Molecular QTL data are incorporated to link genetic variation with intermediate phenotypes. TWAS translates variant-level signals into gene-level associations, and colocalization evaluates whether signals across datasets are consistent with a shared underlying variant. Mendelian randomization is then applied to assess the direction and magnitude of potential effects. Diagnostic procedures and replication across populations or environments are used to evaluate robustness. The framework represents a progressive refinement of evidence rather than a strictly linear sequence

### **6.3 Examples of application**

This integrative approach has been applied across a range of biological systems. In human studies, such as those focusing on lipid-related traits, significant GWAS loci can be examined in relevant tissues (e.g., liver) to identify candidate regions. Colocalization analyses can then be used to prioritize loci showing consistent signals, followed by expression-based analyses and MR to evaluate potential relationships. When results are stable and supported by multiple lines of evidence, they can be further interpreted in the context of known biological pathways or potential therapeutic targets (Porcu et al., 2019; Wainberg et al., 2019; Votava and Parks, 2021; Lessard et al., 2024; Zhang et al., 2024).

In plant systems, similar strategies can be applied to complex traits such as disease resistance. Analyses across tissues and environmental conditions can help identify context-dependent regulatory signals. These can then be integrated with genetic and functional data to prioritize candidate genes. Given the complexity of plant genomes, including extended LD and structural variation, the use of multi-parent populations and pangenome references can improve resolution and interpretation (Zhang et al., 2024).

Together, these examples illustrate that integration across data types is not a matter of simply combining results, but of progressively refining evidence. Through this process, genetic associations can be translated into more specific and testable hypotheses about the mechanisms underlying complex traits.

## **7 Discussion**

### **7.1 Linking molecular associations with causal evaluation**

A central challenge in the study of complex traits lies in connecting genetic associations to mechanistic interpretation. Molecular QTL analyses and TWAS contribute by constraining association signals within a functional context, enabling signals dispersed across the genome to be interpreted at the level of genes or