and a shared underlying variant. Posterior probabilities are assigned to each scenario, with PPH4 commonly used as a summary measure of support for the shared-variant hypothesis (Zuber et al., 2022).

As genomic regions often contain multiple independent signals, the assumption of a single causal variant is frequently violated. To address this, more recent methods incorporate multi-signal modeling and integrate colocalization with fine-mapping, such as coloc combined with SuSiE or FINEMAP, as well as approaches like eCAVIAR and fastENLOC (Foley et al., 2021; Wallace, 2021). These developments improve performance in regions with complex signal structures and allow colocalization to be more closely aligned with locus-level inference.

In practice, reliable colocalization requires careful data harmonization. This includes aligning allele orientation, standardizing effect sizes and standard errors, and using LD reference panels that match the ancestry of the GWAS dataset. Because multiple independent signals may be present within a region, it is often advisable to perform conditional analysis or fine-mapping prior to colocalization, or to apply models that explicitly account for multiple signals. In multi-tissue settings, analyses can be conducted separately for each tissue and then integrated with functional annotations to form a more complete interpretation (Wallace, 2021; Zuber et al., 2022).

Although thresholds such as PPH4 > 0.8 are often used in practice, their interpretation depends on model assumptions, prior choices, and data quality. Sensitivity analyses and conditional results should therefore be considered when evaluating the robustness of findings (Figure 2) (Rasooly et al., 2022).

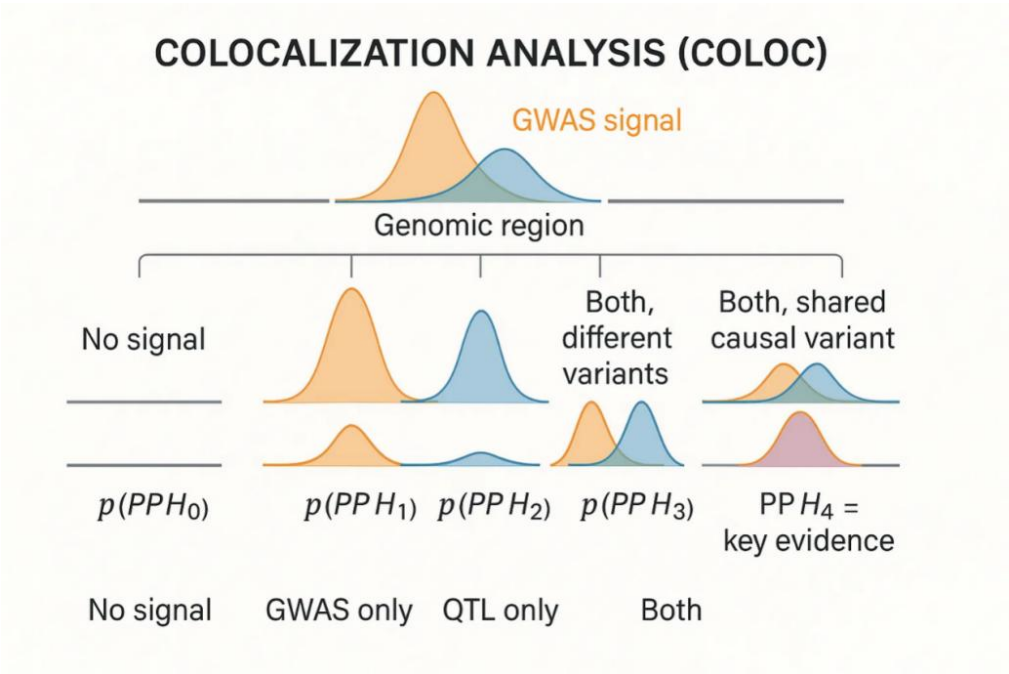

Figure 2 Statistical decision framework of colocalization analysis and its role in cross-dataset integration
Image caption: Colocalization analysis evaluates whether association signals from GWAS and molecular QTL (e.g., eQTL) within the same genomic region are consistent with a shared underlying genetic variant. Under a Bayesian framework, methods such as COLOC compare five mutually exclusive scenarios, including no signal, GWAS-only, QTL-only, independent signals, and shared signals, thereby quantifying alternative explanations for the observed patterns. The figure illustrates these scenarios and the corresponding signal configurations, with PPH4 representing the posterior support for the shared-variant hypothesis. It should be noted that PPH4 reflects statistical evidence for signal concordance rather than direct inference of causal mechanisms or mediation

4.2 Interpreting colocalization results

While colocalization is a useful tool for prioritizing candidate loci, its results should not be interpreted as direct evidence of causal relationships. Even when the posterior probability for a shared signal is high, this only indicates that two associations are consistent with the same underlying variant; it does not establish how that variant influences the trait.