Methylation QTL (meQTL) describe the influence of genetic variation on DNA methylation patterns. When combined with data on chromatin accessibility and histone modifications, they contribute to a broader view of regulatory architecture. In multi-tissue settings, meQTL signals often only partially overlap with eQTL or GWAS loci, suggesting that different regulatory layers may act through distinct pathways.

In plant systems, particular attention should be given to differences among methylation contexts (CG, CHG, and CHH) and their regulatory mechanisms. Incorporating tissue- or environment-specific analyses can improve the interpretability of these signals.

3 TWAS: From GWAS to Gene-Level Associations

Within integrative analyses, transcriptome-wide association studies (TWAS) provide a framework for translating GWAS signals from the variant level to the gene level. By incorporating gene expression as an intermediate phenotype, TWAS enables the evaluation of whether genetically regulated expression is associated with complex traits, thereby offering a structured link between genetic variation and downstream phenotypes.

It is important to note, however, that TWAS does not directly establish causality. Rather, it reorganizes SNP-level association signals into gene-level statistics under a specified expression prediction model and LD structure. In this sense, TWAS is best viewed as a structured projection of GWAS signals, rather than an independent causal inference approach.

3.1 Basic principles

The central idea of TWAS is to use reference datasets that contain both genotype and expression data to train predictive models of gene expression, and then apply these models to GWAS data to assess gene–trait associations (Li and Ritchie, 2021; Evans et al., 2024). For a given gene g, the predicted expression can be written as:

$$\widehat{E}_g = \sum_{j \in L_g} w_{gj} G_j$$

where L_g typically denotes SNPs within a cis region, w_{gj} represents weights estimated from reference data (e.g., using Elastic Net, BLUP, or BSLMM), and G_j denotes SNP dosage. Association is then tested between $\widehat{E}_g$ and the phenotype Y, with the goal of evaluating whether genetically driven variation in expression is related to the trait (Mai et al., 2023).

When only GWAS summary statistics are available, TWAS can be implemented using LD-based transformations. Let Z denote the vector of SNP-level GWAS Z-scores and R the corresponding LD matrix. The gene-level statistic can be approximated as:

$$Z_g \approx \frac{w_g^T Z}{\sqrt{w_g^T R w_g}}$$

The significance of this statistic depends on the effective information carried by the weights under the LD structure. Two practical considerations follow from this formulation. First, the LD reference panel should closely match the target GWAS population. Second, the choice of tissue or cell type used to train the expression model plays a critical role in determining both interpretability and effect direction (Li and Ritchie, 2021).

3.2 Common methods

Most TWAS methods follow a two-step strategy involving model training and downstream association testing, but differ in how expression weights are constructed and how multi-tissue information is incorporated. PrediXcan and its summary-based extension S-PrediXcan use elastic net regression to estimate cis-regulatory weights in reference data, which are then applied to individual-level or summary-level GWAS data. MultiXcan and UTMOST extend this framework by integrating information across tissues to distinguish shared from tissue-specific effects.

FUSION adopts a more flexible approach by integrating multiple prediction models-including BLUP, LASSO, Elastic Net, and BSLMM-within a unified framework, allowing direct computation of gene-level statistics from