

Accordingly, interpreting SNP-based heritability derived from methods such as GREML as direct evidence of “low trait heritability” is not statistically justified. Such interpretations overlook the dependence of the estimate on marker coverage, linkage disequilibrium structure, and parametric modeling assumptions.

A more appropriate perspective is that SNP-based heritability reflects the joint explanatory capacity of multiple factors within a given analytical framework. First, the extent to which genomic markers cover true genetic variation determines the upper bound of observable genetic signal. Second, the structure of linkage disequilibrium governs whether causal variants can be effectively proxied by measured markers. Third, assumptions regarding allele frequency distributions and effect sizes further influence both the bias and variance of the estimate (Speed et al., 2016; Yang et al., 2017; Génin, 2019; Wainschtein et al., 2022).

6.2 Interpretation checklist: a standardized workflow for GREML-based SNP heritability

After obtaining an SNP-based heritability estimate within the GREML framework, a single numerical value alone does not provide sufficient explanatory power. Its statistical significance and biological interpretation both depend on the data-generating process, model specification, and stability of the estimation procedure. Therefore, a sound interpretation should not be limited to reporting the estimate itself, but should be grounded in a systematic evaluation of the entire analytical process. In other words, interpreting SNP heritability is a form of “conditional inference,” whose validity depends on the consistency among data quality, model assumptions, and methodological suitability.

The foundation of result interpretation lies in the reliability of the data and the appropriateness of phenotypic modeling. The extent to which SNP markers cover genome-wide variation directly constrains the range of genetic variance that can be identified. In particular, when only common-variant genotyping array data are used, low-frequency variants, rare variants, and structural variants are not sufficiently captured. Their corresponding genetic contributions are therefore inevitably missed, leading to a systematic underestimation of SNP heritability, a phenomenon that has been clearly supported by large-scale sequencing studies (Wainschtein et al., 2019; 2022). Statistical processing of phenotypes is equally important. Phenotypes that have not been appropriately transformed or adjusted for systematic environmental factors often make effective variance decomposition difficult. In multi-environment or repeated-measure settings, if environmental heterogeneity is not explicitly modeled, part of the environmental effect may be incorrectly absorbed into the residual term, thereby weakening the ability to identify genetic variance (Evans et al., 2017; Yang et al., 2017).

The treatment of population structure and relatedness constitutes another important source of estimation bias. Systematic differences introduced by population stratification, together with correlation structures arising from cryptic relatedness, may distort heritability estimates if not adequately controlled, and the direction of such bias is not necessarily fixed. In individual-level analyses, correcting for principal components or constructing an appropriate mixed-model structure to absorb population-structure effects is a basic requirement for maintaining valid estimation. Meanwhile, the identification of close relatives and the thresholds used for their exclusion should also be subjected to sensitivity analysis, so as to avoid unstable inference caused by differences in sample structure. When individual-level data are unavailable, summary-statistics-based approaches, such as LDSC or SumHer, can serve as alternative strategies for robustly modeling population stratification and provide important references for interpreting GREML results (Ge et al., 2016; Speed et al., 2016; Speed and Balding, 2018; Speed et al., 2022).

The dependence of heritability estimation on the construction of the genomic relationship matrix (GRM) means that its interpretation must be situated within specific modeling assumptions. Because linkage disequilibrium (LD) patterns among SNPs are complex, failure to appropriately account for LD heterogeneity, or weak LD between genotyped markers and causal variants, may lead to systematic biases in different directions (Speed et al., 2012; 2016). In practice, a single standard GRM is often insufficient to fully characterize genetic architecture. Introducing LD correction or using stratified GRM models to partition SNPs by allele-frequency intervals or functional annotation categories can, to some extent, reduce model-specification bias and improve the resolution with which sources of genetic variance are interpreted. In addition, cross-checking results with frameworks that