

full set of SNPs with subsets stratified by minor allele frequency (MAF) or linkage disequilibrium (LD), as well as LD-pruned variant sets-it is possible to assess the sensitivity of heritability estimates to marker selection, thereby determining whether the results depend on specific data-processing strategies.

Furthermore, regional sensitivity analysis involves removing specific high-LD regions (e.g., the MHC region) and re-estimating heritability to examine whether genetic variance is disproportionately concentrated in localized genomic segments. If the estimates change substantially after removal, this suggests that the region plays a dominant role in the genetic architecture of the trait. In addition, consistency comparisons across different methods (such as between GREML and LDSC or SumHer) constitute a critical step. Systematic discrepancies between methods are more likely to reflect differences in model assumptions or LD characterization, rather than being attributable to random estimation error.

S1.5 Interpretation framework

When interpreting SNP heritability, it is necessary to recognize that different methods correspond to different statistical targets (estimands), and therefore their estimates are generally not directly comparable. Specifically, GREML-based approaches rely on the genetic covariance structure among individuals as defined by the genetic relationship matrix (GRM), and essentially estimate the proportion of additive genetic variance captured by this matrix. In contrast, LDSC operates within an LD-weighted regression framework, providing an overall, aggregate-level interpretation of GWAS summary statistics. SumHer further incorporates explicit modeling of minor allele frequency (MAF) and LD, making its estimand dependent on the chosen weighting scheme and assumptions about the genetic architecture.

In this sense, the results h_{SNP}^2 obtained from different methods correspond to distinct statistical definitions. Only under ideal conditions-where the SNP set, LD structure, and model assumptions are completely aligned-can these estimates be expected to converge. Otherwise, the observed differences should be understood as reflecting differences in statistical targets, rather than inconsistencies in underlying biological mechanisms.

S1.6 Summary

In summary, this study establishes a systematic technical framework for SNP heritability analysis. Its core lies in providing a robust foundation for subsequent variance decomposition through standardized data preprocessing and GRM construction. At the same time, by integrating individual-level data methods with summary statistics approaches, it enables multi-path estimation and cross-validation. Furthermore, through comprehensive model diagnostics and sensitivity analyses, the framework ensures the reliability of the results and clearly delineates the boundaries of interpretation. This framework is not only applicable to large-scale human genetic datasets such as the UK Biobank, but also demonstrates strong scalability and can be extended to related research areas, including crop genetics and genomic selection.



Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.