

A NOVEL SNP HERITABILITY MODEL FOR HERITABILITY ANALYSES AND GENOMIC PREDICTION

ANUBHAV KAPHLE 

(Received 5 May 2023; first published online 23 June 2023)

2020 *Mathematics subject classification*: primary 62P10; secondary 92D10.

Keywords and phrases: heritability model, heritability analysis, phenotype prediction.

The SNP heritability model refers to a statistical model parametrising the variance of SNP effect sizes [4]. Most SNP heritability analyses to date assume a 1-parameter model that assigns the same variance to each SNP, which is unrealistic and has led to sub-optimal modelling of the genetic architecture and inaccurate estimates. Recent works have incorporated additional parameters to incorporate the properties of SNPs such as the minor allele fraction (MAF), local LD patterns and functional knowledge into the model to better capture heritability. This has led to the development of models such as the LDAK model [2], the Baseline LD model [1] and recently the BLD-LDAK model [3]. The BLD-LDAK heritability model is a complex (66 df) model containing highly correlated sets of predictors. My thesis is focused on exploring and testing existing and new predictors of SNP heritability and combining these to build a parsimonious heritability model and compare its performance with the BLD-LDAK model. I start by evaluating the BLD-LDAK heritability model using data from the UK Biobank project over 14 traits and provide updated results of SNP heritability and functional enrichment analyses. In the next step, I collect a comprehensive set of functional annotations that might be predictive of heritability from public genomic databases such as ENCODE, RoadMap Epigenome project, UCSC Genome Browser and RefSeq genes, and subject them to systematic variable selection to construct a new 10-parameter BIC10 heritability model. I then perform heritability analyses of traits recorded in the UK Biobank project to compare models. I also compare heritability models based on phenotype prediction accuracy across a range of diverse traits from the UK Biobank and assess the portability of heritability models across human ancestries. I show that the BIC10 and BLD-LDAK heritability models have equivalent performance, although the BIC10 model has 56 fewer parameters. The

Thesis submitted to the University of Melbourne in July 2022; degree approved on 22 December 2022; supervisors David Balding, Doug Speed and Damjan Vikcevic.

© The Author(s), 2023. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

fewer degrees of freedom provide better interpretability and computational advantages for heritability analysis without loss of accuracy.

The published thesis is available at <http://hdl.handle.net/11343/325143>.

References

- [1] S. Gazal, H. K. Finucane, N. A. Furlotte, P.-R. Loh, P. F. Palamara, X. Liu, A. Schoech, B. Bulik-Sullivan, B. M. Neale, A. Gusev and A. L. Price, 'Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection', *Nat. Genetics* **49**(10) (2017), 1421–1427.
- [2] D. Speed, N. Cai, M. R. Johnson, S. Nejentsev and D. J. Balding, 'Reevaluation of SNP heritability in complex human traits', *Nat. Genetics* **49**(7) (2017), 986–992.
- [3] D. Speed, J. Holmes and D. J. Balding, 'Evaluating and improving heritability models using summary statistics', *Nat. Genetics* **52**(4) (2020), 458–462.
- [4] D. Speed, A. Kaphle and D. J. Balding, 'SNP-based heritability and selection analyses: improved models and new results', *BioEssays* **44**(5) (2022), Article no. 2100170.

ANUBHAV KAPHLE, School of Mathematics and Statistics,
The University of Melbourne, Parkville, Victoria 3010, Australia

Current address:

Australian e-health Research Center,
Commonwealth Scientific and Industrial Research Organisation (CSIRO),
Melbourne, Victoria 3052, Australia
e-mail: anubhavkaphle@gmail.com