Identification of cell-type-specific genetic regulation of gene expression for transcriptome-wide association studies

Duo Zhang*[1], Qiurui Ma*[2], Brandon Jew[3], Sriram Sankararaman[4]

[1]Shandong University; [2]Hong Kong University of Science and Technology; [3] Bioinformatics Interdepartmental Program, UCLA; [4]Department of Computer Science, Department of Human Genetics, UCLA

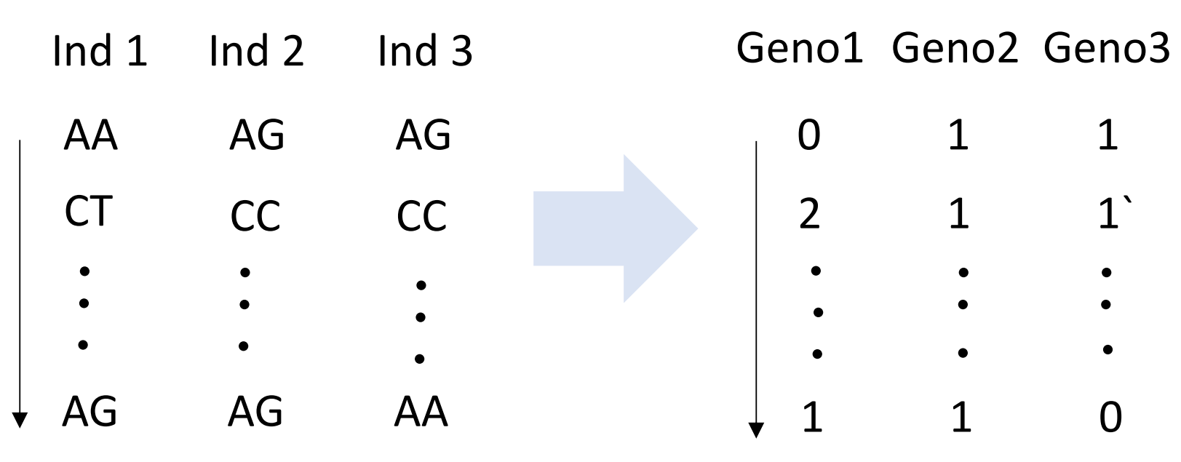
*** Equal contribution**

I. BACKGROUNDS

i. The Central Dogma

- Single nucleotide polymorphisms (**SNPs**) are sites of variation in our DNA
- Gene expression (**GE (Z)**) is the level of mRNA in one cell type. **Bulk level GE (G)** is the combined GE of all cell types in a tissue.

Encode DNA into SNP data



Ind 1	Ind 2	Ind 3
AA	AG	AG
CT	CC	CC
⋮	⋮	⋮
AG	AG	AA

Geno1	Geno2	Geno3
0	1	1
2	1	1`
⋮	⋮	⋮
1	1	0

ii. Current Studies

- Genome-wide association studies (**GWAS**)[2] linearly associate SNPs with phenotypes
- Transcriptome-wide association studies (TWAS)[3] linearly characterize the association of **GE** regulated by SNPs and phenotypes

iii. Challenges & Goals

Methodological

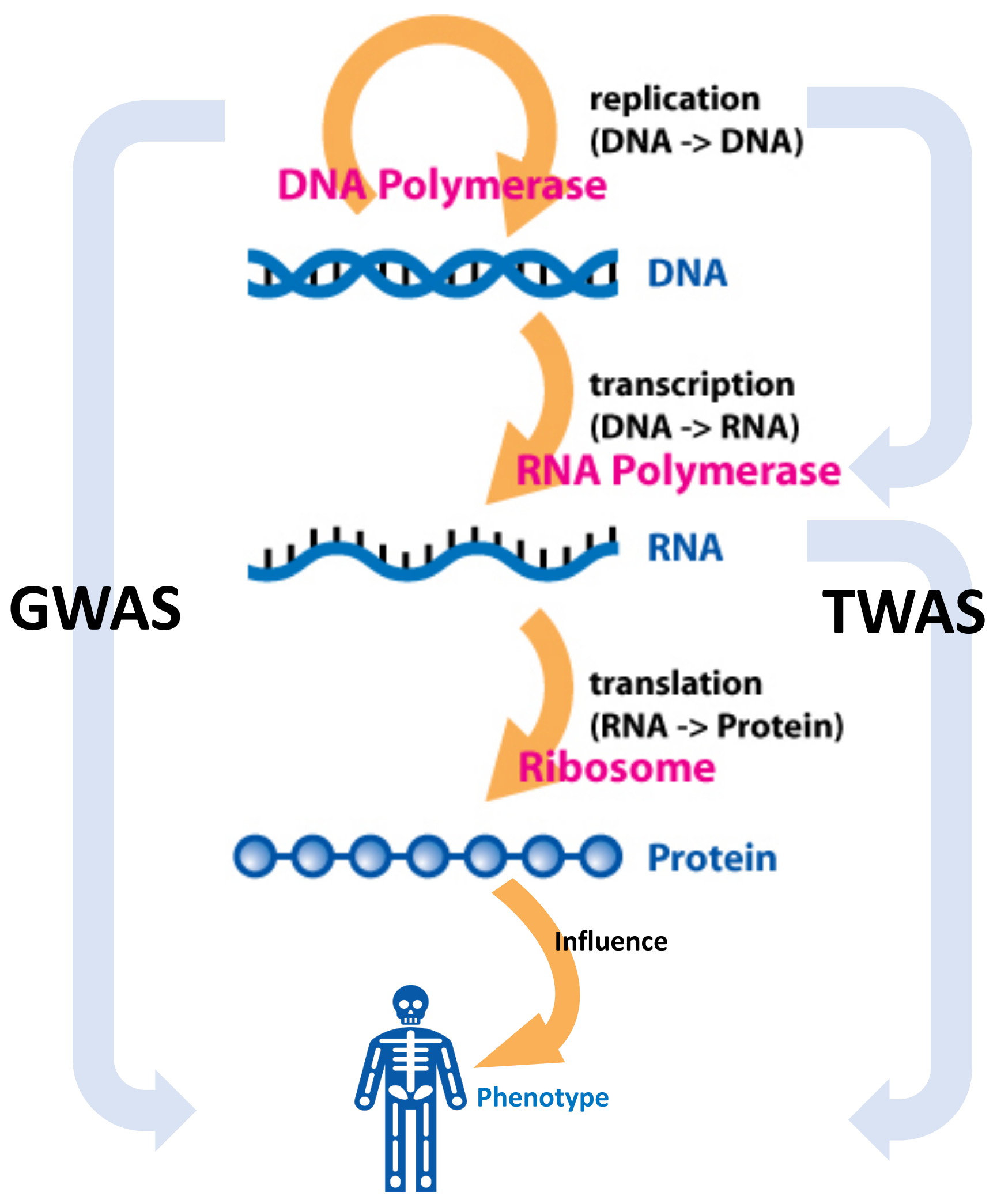
Unclear how SNPs affect phenotypes

- Missing cell type information
- Identified associations do not indicate causality

Our Goal

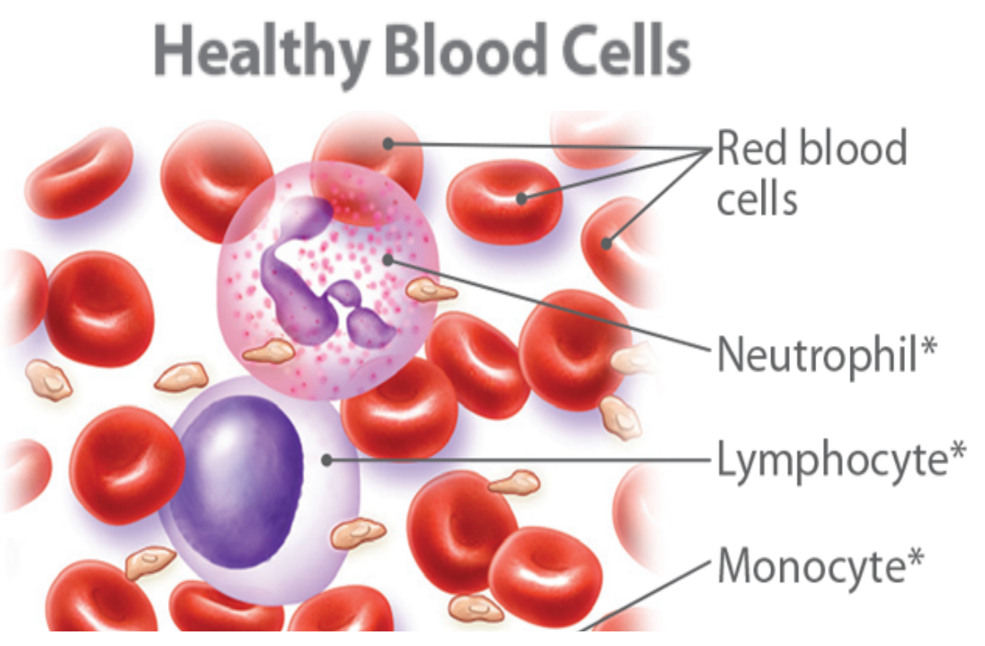
Deconvolute bulk level GE into cell-specific GE with SNPs and cell-type weights. Associate cell-type specific GE with phenotypes

The Central Dogma



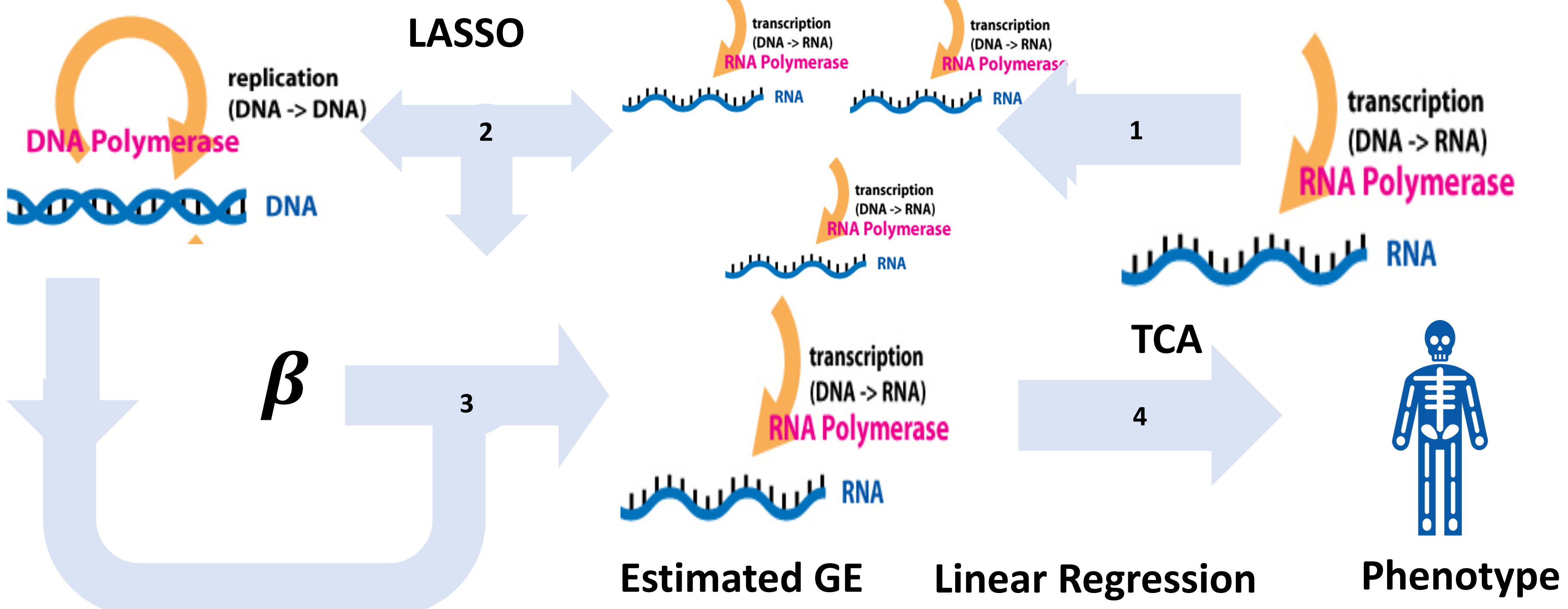
Practical

Cell-type-specific biological data is resource intensive and expensive to acquire.



II. METHODOLOGY

i. Pipeline Framework



1. TCA deconvolutes bulk level GE into cell-type-specific ones
2. Effect size of SNPs on cell-type-specific GE imputed by LASSO
3. Cell-type-specific gene expression imputed from effect size for external cohorts
4. Estimated cell-type-specific GE is regressed into phenotype

ii. TCA model

$$Z_h^i = \overbrace{\overbrace{\epsilon_z + \mu_h + \begin{bmatrix}1\\27\\0\end{bmatrix}^T \gamma_h}^{\text{TCA Model: mdl2}} + \begin{bmatrix}0\\2\\1\\\vdots\\0\\1\end{bmatrix}^T \beta_h}^{\text{Added SNPs effect: mdl1}} \qquad \epsilon_z \in N(0,\sigma_z)$$

Cell-Specific GE **Covariate:** **(Gender, Age, Smoking)** **SNPs**

$$G_i = c_i^2\delta + \sum_{h=1}^{k} w_{hi} z_{hi} + \epsilon_g \qquad \epsilon_g \in N(0,\sigma_g)$$

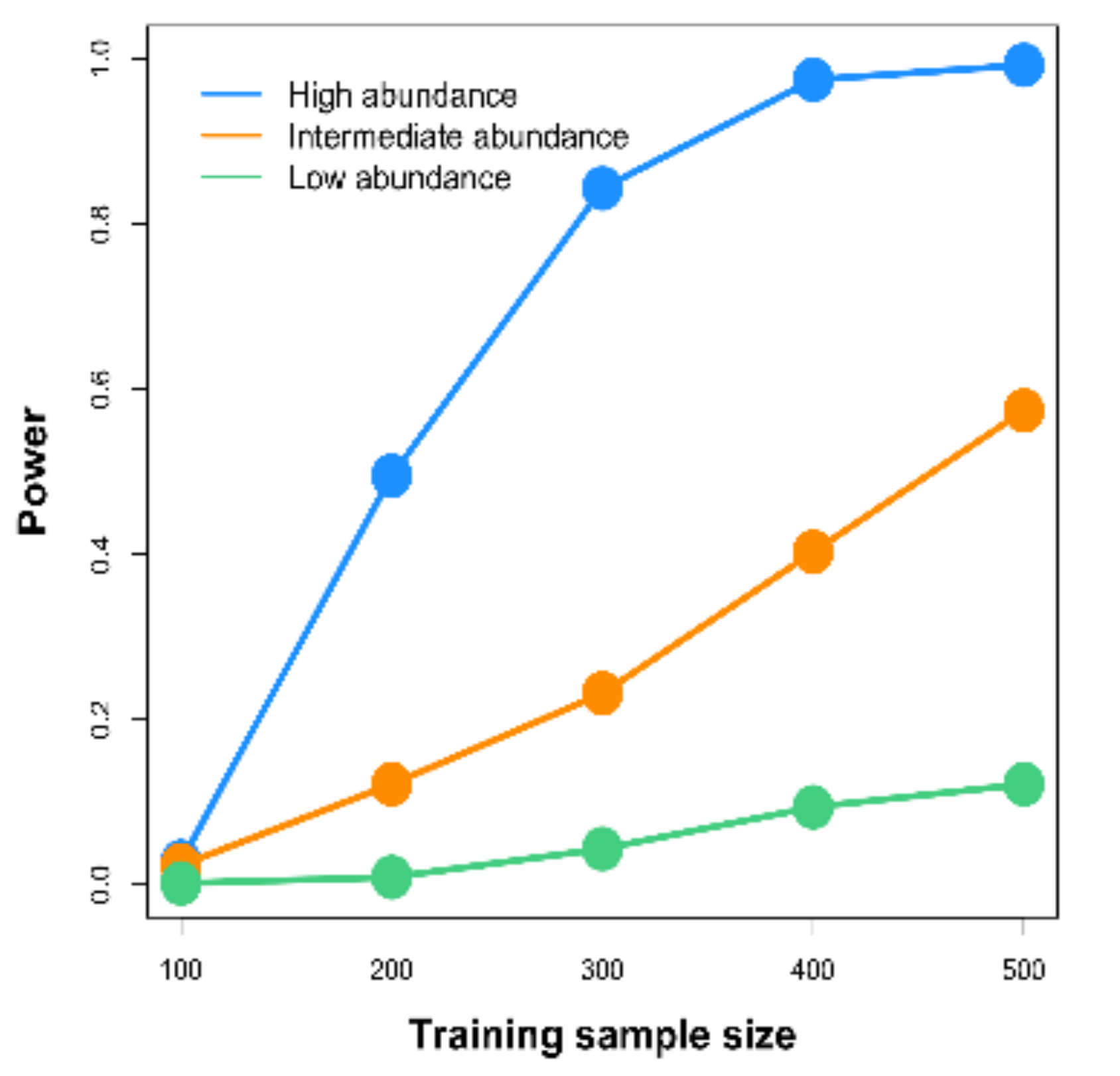
TCA assumes bulk level **GE** is a linear combination of GEs

III. RESULTS

i. Simulated data

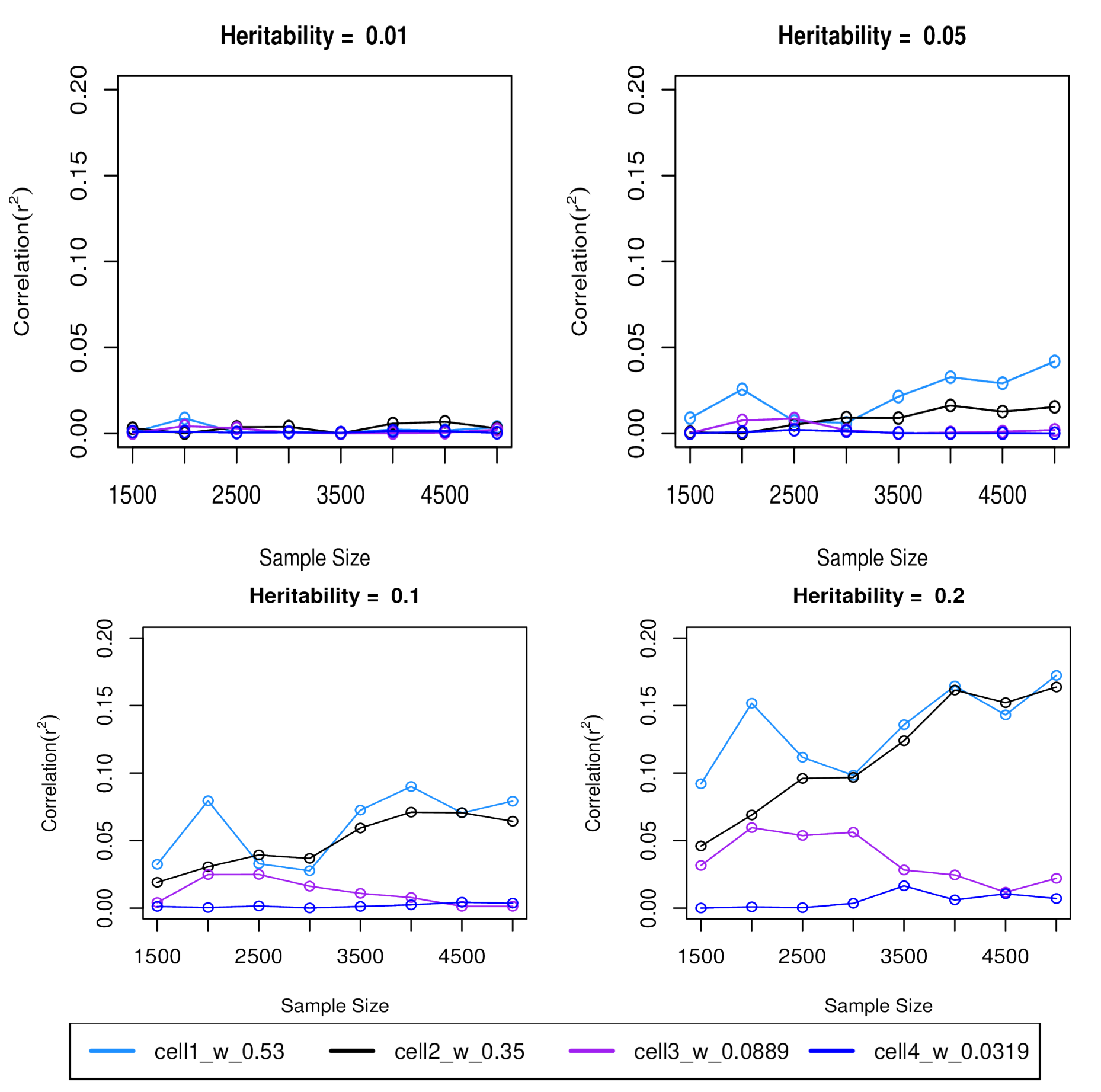
The method possesses sufficient power to detect cell-specific expression-phenotype associations

Power of cell-specific expression imputation



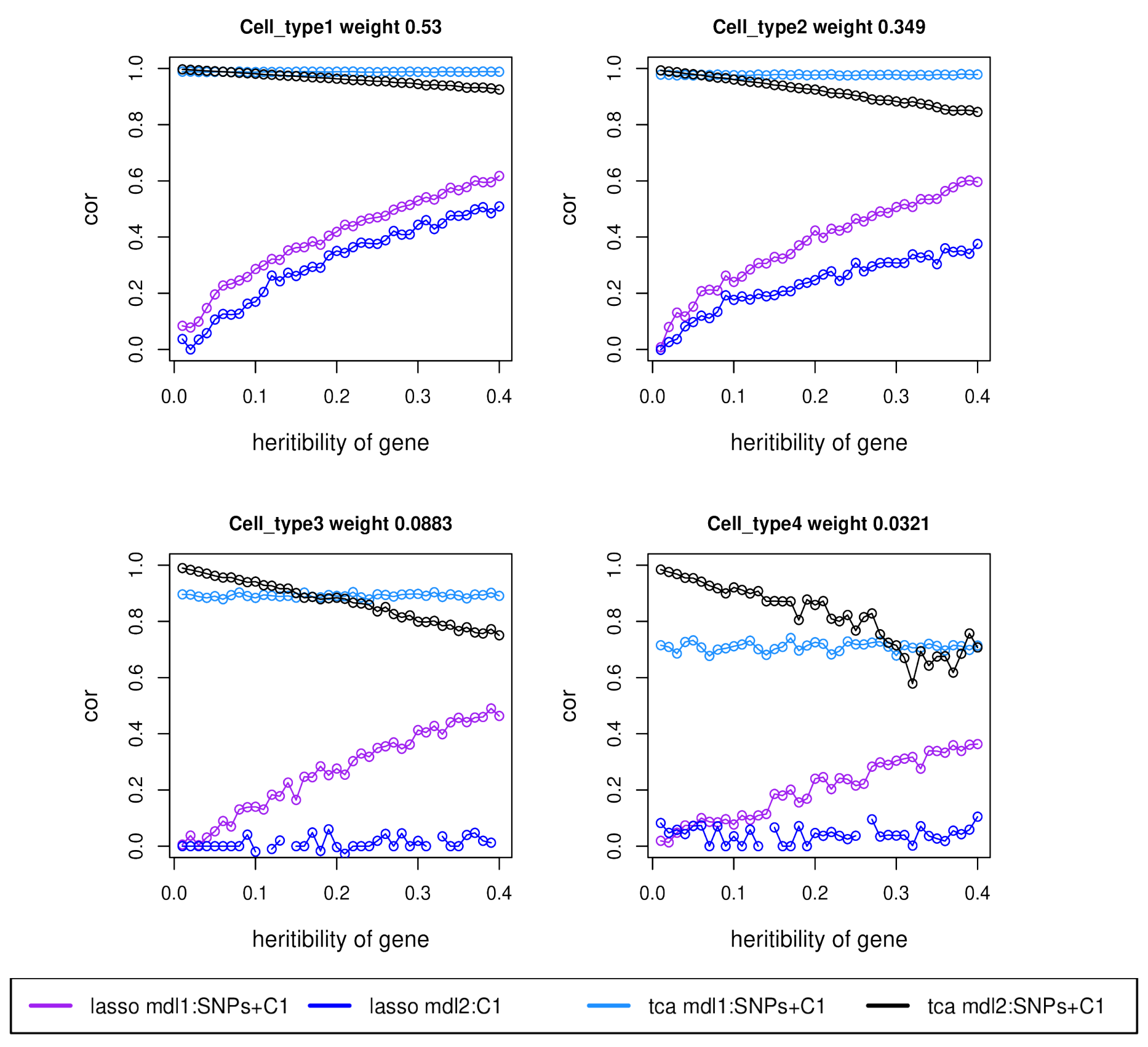
The variance explained by the model is lower than the theoretical upper bound

Prediction of cis–regulated cell–specfic expression from bulk data

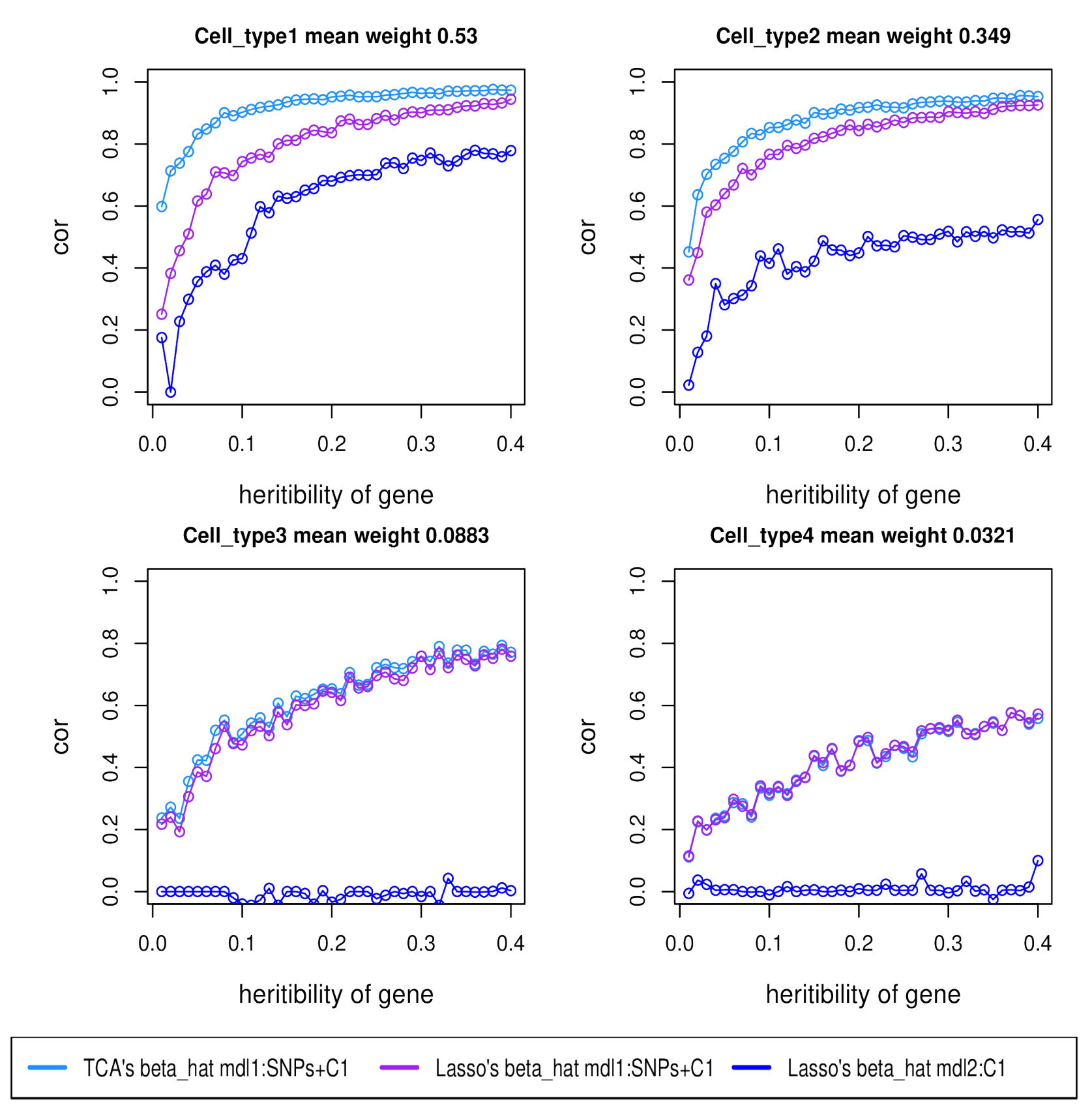


Our modified TCA performs better than the original TCA

Correlation of estimated gene expression and ground truth



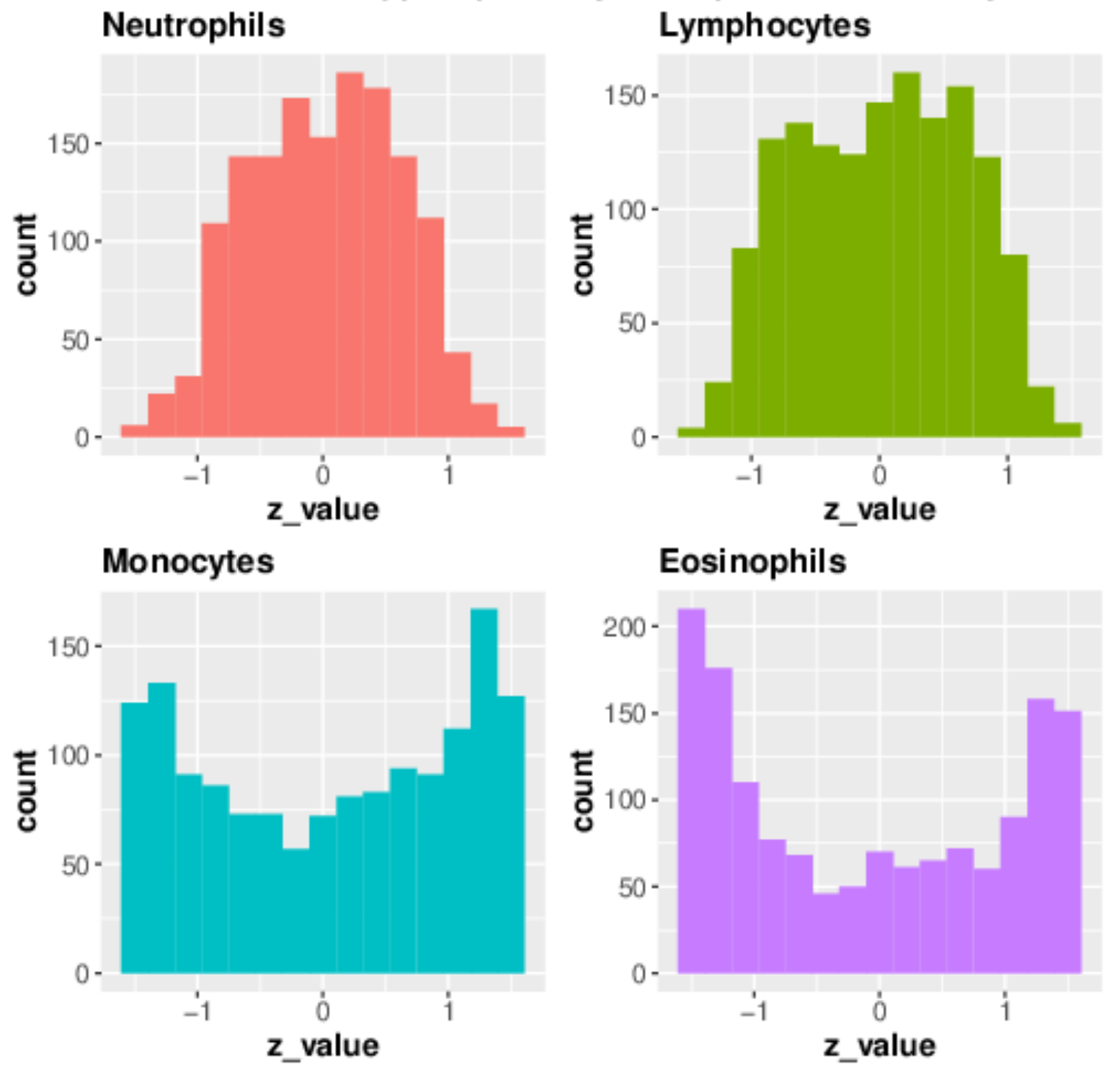
Correlation of estimated SNPs effect size and the ground truth



ii. Real data

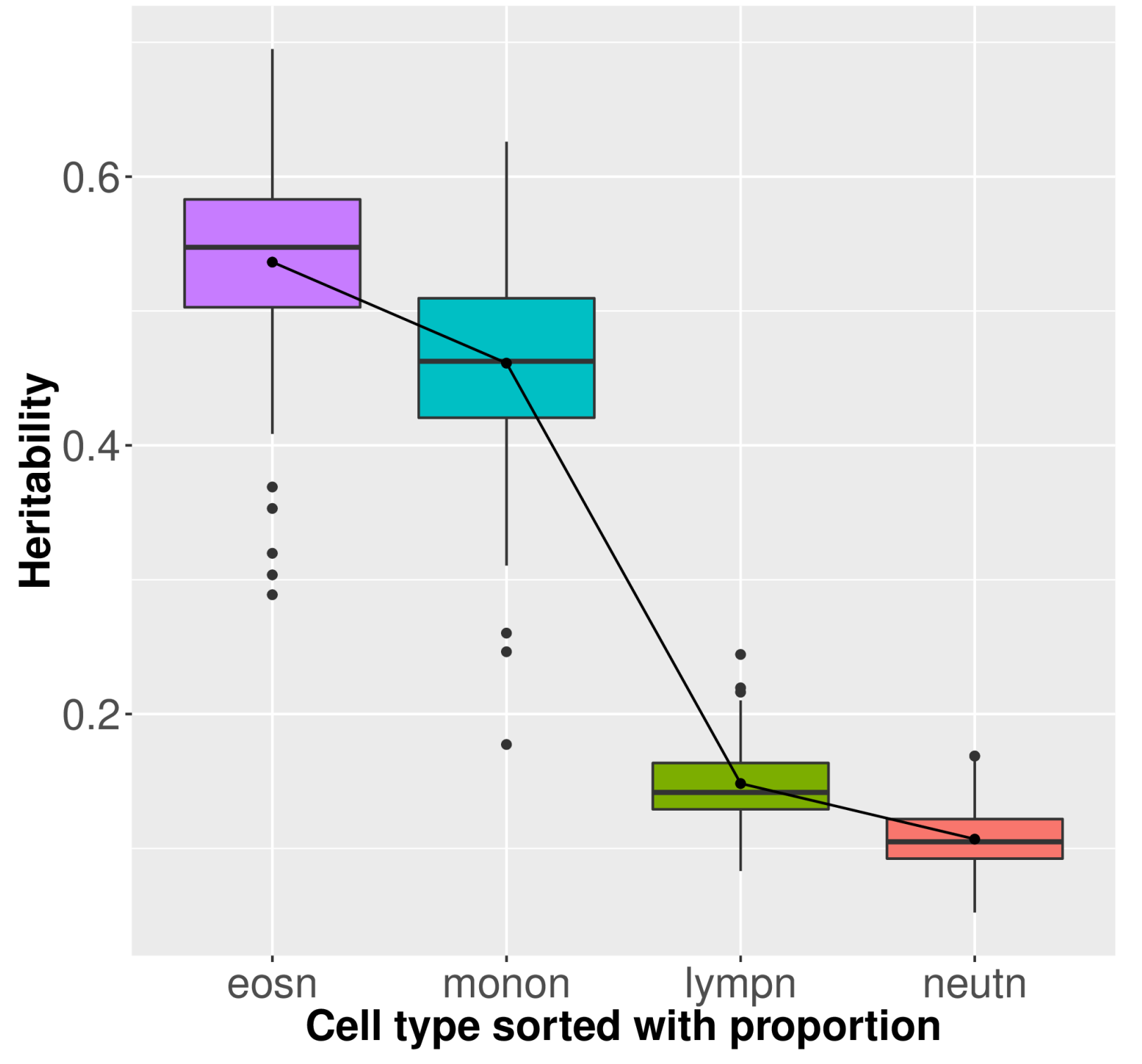
The model's performance is inconsistent among cell types

TCA estimated cell–type–specific gene expression for one gene



The model overfits on less abundance cell types

Cell–type–specific heritability



IV. CONCLUSION

Cell-specific expression-phenotype associations in large datasets (UK BioBank)) could be learnt with its SNPs and readily available, abundant datasets with bulk level gene expressions.

References

[1] Rahmani, E., Schweiger, R., Rhead, B., Criswell, L. A., Barcellos, L. F., Eskin, E., ... & Halperin, E. (2019). Cell-type-specific resolution epigenetics without the need for cell sorting or single-cell biology. *BioRxiv*, 437368.

[2] Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS computational biology*, *8*(12), e1002822. doi:10.1371/journal.pcbi.1002822

[3] Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W., ... & Sullivan, P. F. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics*, *48*(3), 245.

Acknowledgements

- This work is supported by CSST program.
- The authors gratefully acknowledge the discussions with Elior Rahmani and Yue Wu.