

Identification of *trans* regulatory network from multi-condition multi-omics quantitative trait loci with applications in Alzheimer's Disease



Memorial Sloan Kettering
Cancer Center

Xuewei Cao^{1,2}, Yifei Wang¹, Tabassum Fabiha^{2,3}, Yunshen Bai⁴, Zining Qi⁴, Lude Franke^{5,6}, Ellen A. Tsai⁷, Harm-Jan Westra^{5,6},

The Alzheimer's Disease Functional Genomics Consortium, Gao Wang^{1,8*}, Kushal K. Dey^{2,3*}

¹Center for Statistical Genetics, The Gertrude H. Sergievsky Center, ⁴Department of Biostatistics, ⁸Department of Neurology, Columbia University, New York, NY, USA

²Sloan Kettering Institute, ³Gerstner Sloan Kettering Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, USA

⁵Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, ⁶Oncode Institute, Groningen, The Netherlands

⁷Translational Biology, Research and Development, Biogen Inc., Cambridge, MA, USA

Highlights

- Importance:** Over 70% heritability of complex diseases and gene expression is driven by *trans* effects downstream of GWAS genes[1].
- Challenges:** Limited sample size and smaller causal effects of associations make it challenging to decode *trans*-regulatory architecture[2, 3].
- Improvement:** Integration of multiple molecular QTL (xQTL) studies can aggregate *trans* effect gene programs downstream of multiple *trans*-eSNPs and gene perturbations across various tissues and cell types.

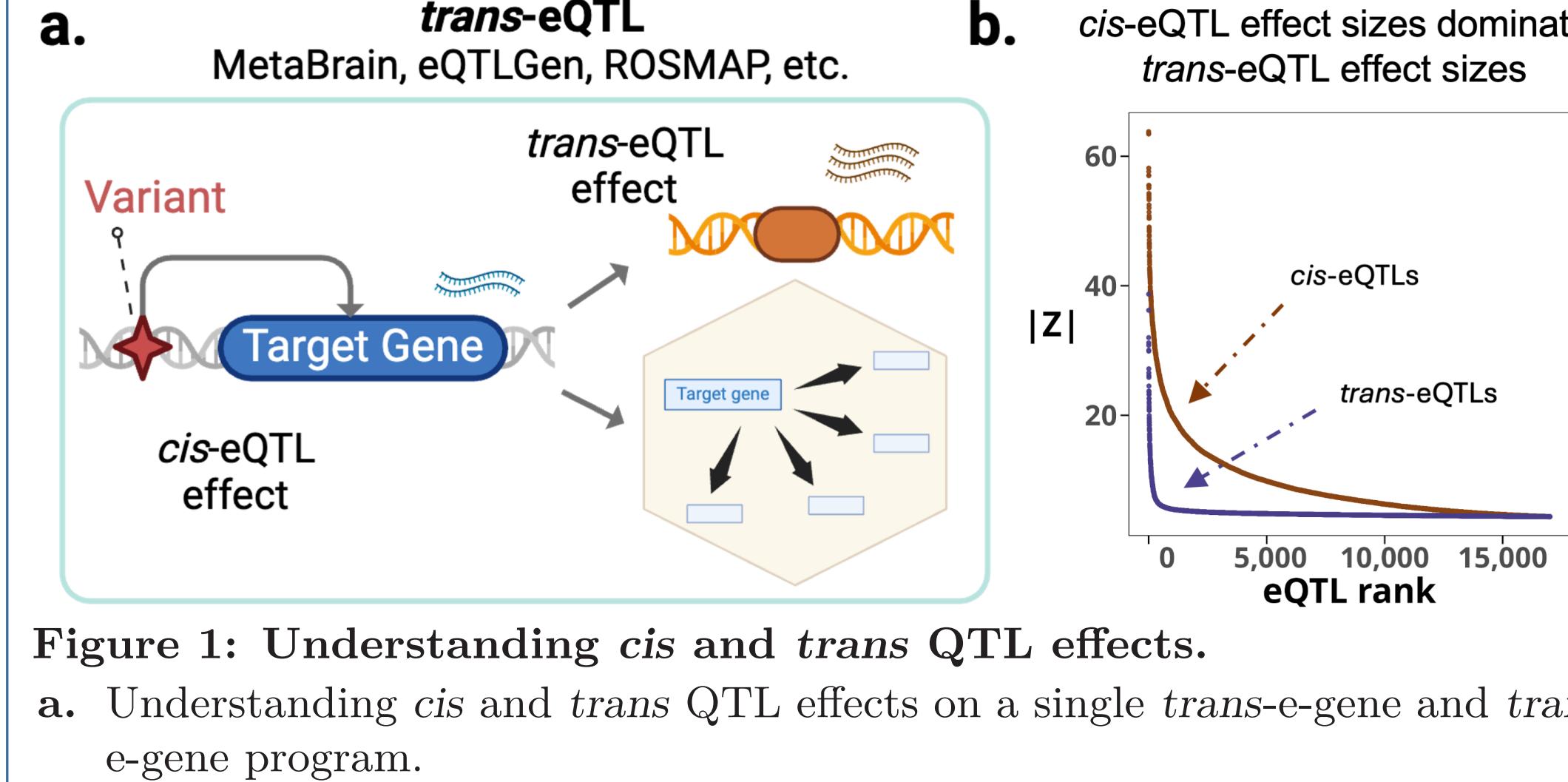


Figure 1: Understanding *cis* and *trans* QTL effects.

- Understanding *cis* and *trans* QTL effects on a single *trans*-e-gene and *trans*-e-gene program.
- Cumulative distribution of effects for the strongest *cis* and *trans*-eQTLs for each expressed gene in the dorsolateral prefrontal cortex brain region from ROSMAP dataset.

Overview of transCCA

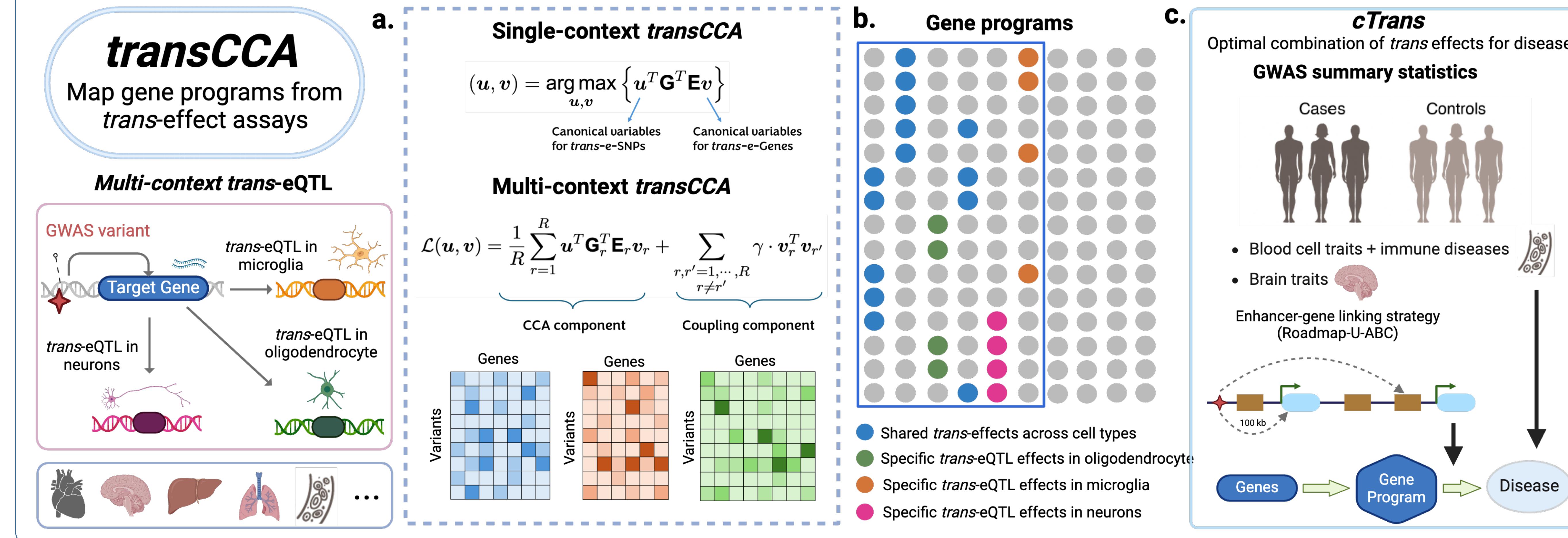


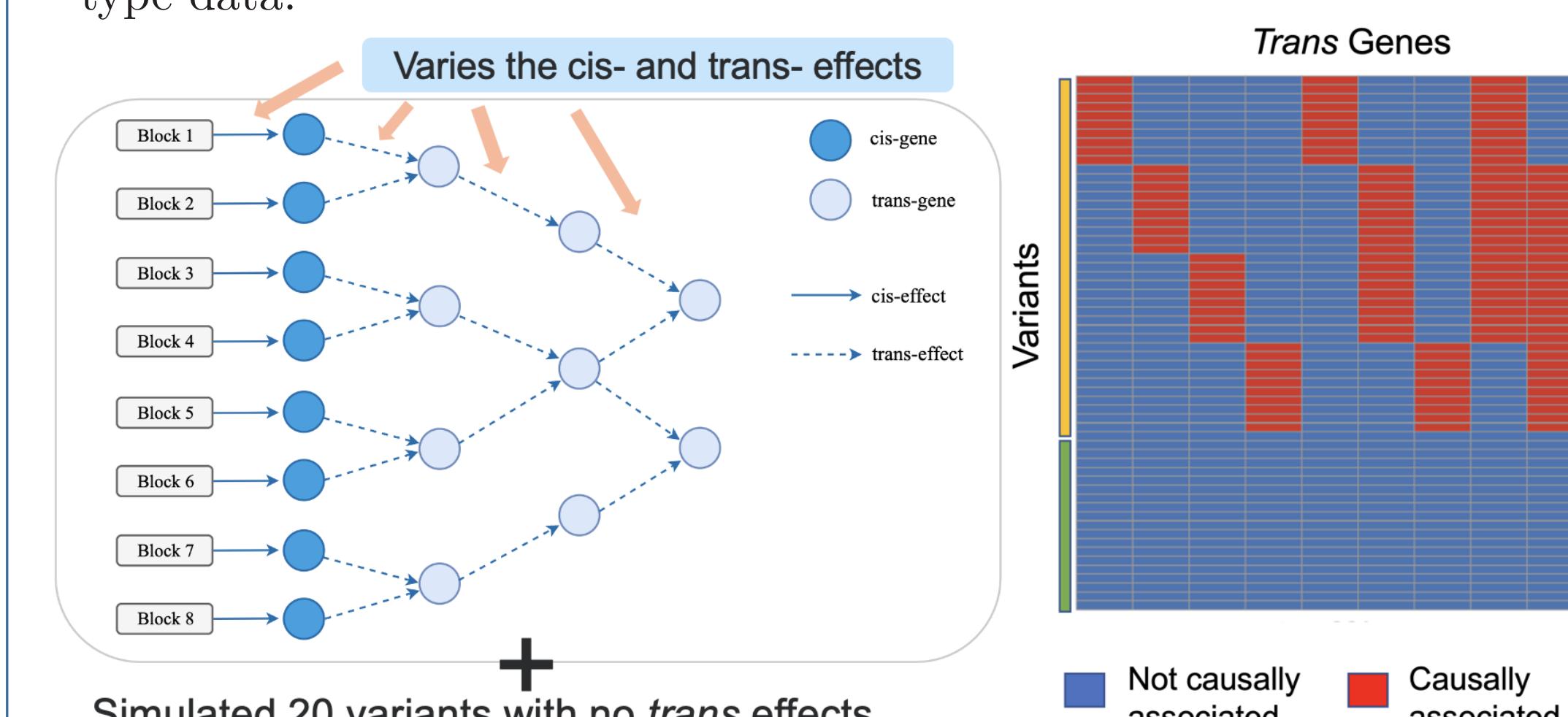
Figure 3: Framework of *transCCA* for mapping gene programs and disease-associated *trans*-effects.

- Overview of single- and multi-context *transCCA* for mapping gene programs from *trans*-effect assays using penalized matrix decomposition algorithm[4].
- Gene programs from *transCCA*, where each colored dot represents a gene grouped into distinct programs based on *trans*-effects.
- We developed an optimal combination of *trans*-effects (*cTrans*) tailored for each disease and complex trait. This approach includes: (i) defining gene programs by selecting the top 95% quantile of genes and assigning normalized probabilistic scores; (ii) annotating variants within gene programs, focusing on 100kb and enhancer-gene linked variants; and (iii) assessing disease informativeness by using the S-LDSC framework to measure heritability enrichment of variant annotations across complex diseases.

Numerical Studies & Comparisons

Consider the following *trans*-regulatory structure: For each block,

- there are 5 independent variants.
- there are M variants with LD in one gene region from ROSMAP genotype data.



Note: Variants have similar MAF as the variants in the ROSMAP cohort.

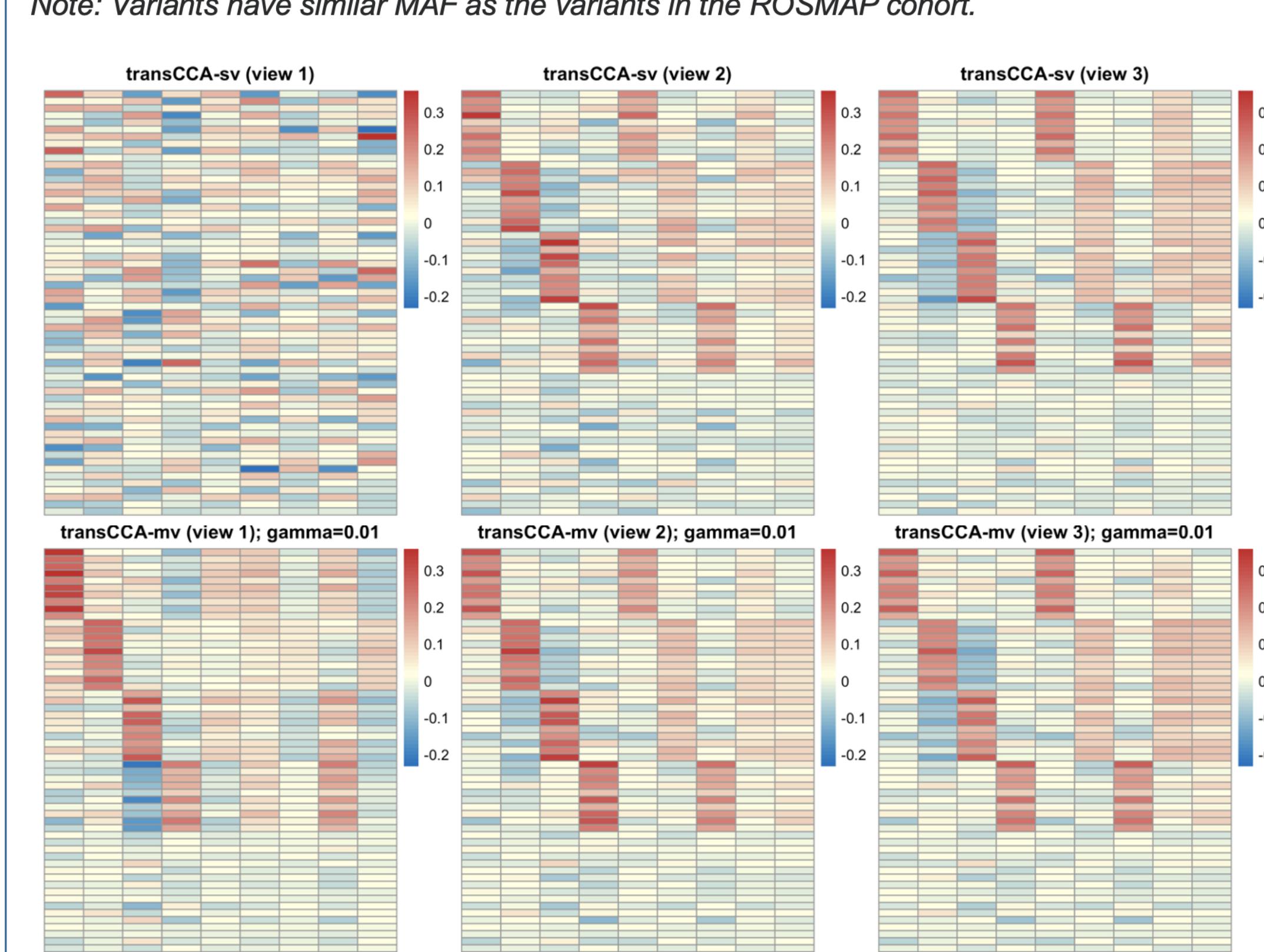


Figure 2: Multi-context *transCCA* can recover weaker sharing *trans*-regulatory associations.

Ongoing works in extensive simulation studies

- Evaluate the robustness of *transCCA* for leveraging the large-scale genetic variants in LD.
- Evaluate QTL calling based on *trans* gene program level phenotypes.

Applications to Public Available Data Resources, and the FunGen-xQTL Project with Alzheimer's Disease (AD) GWAS

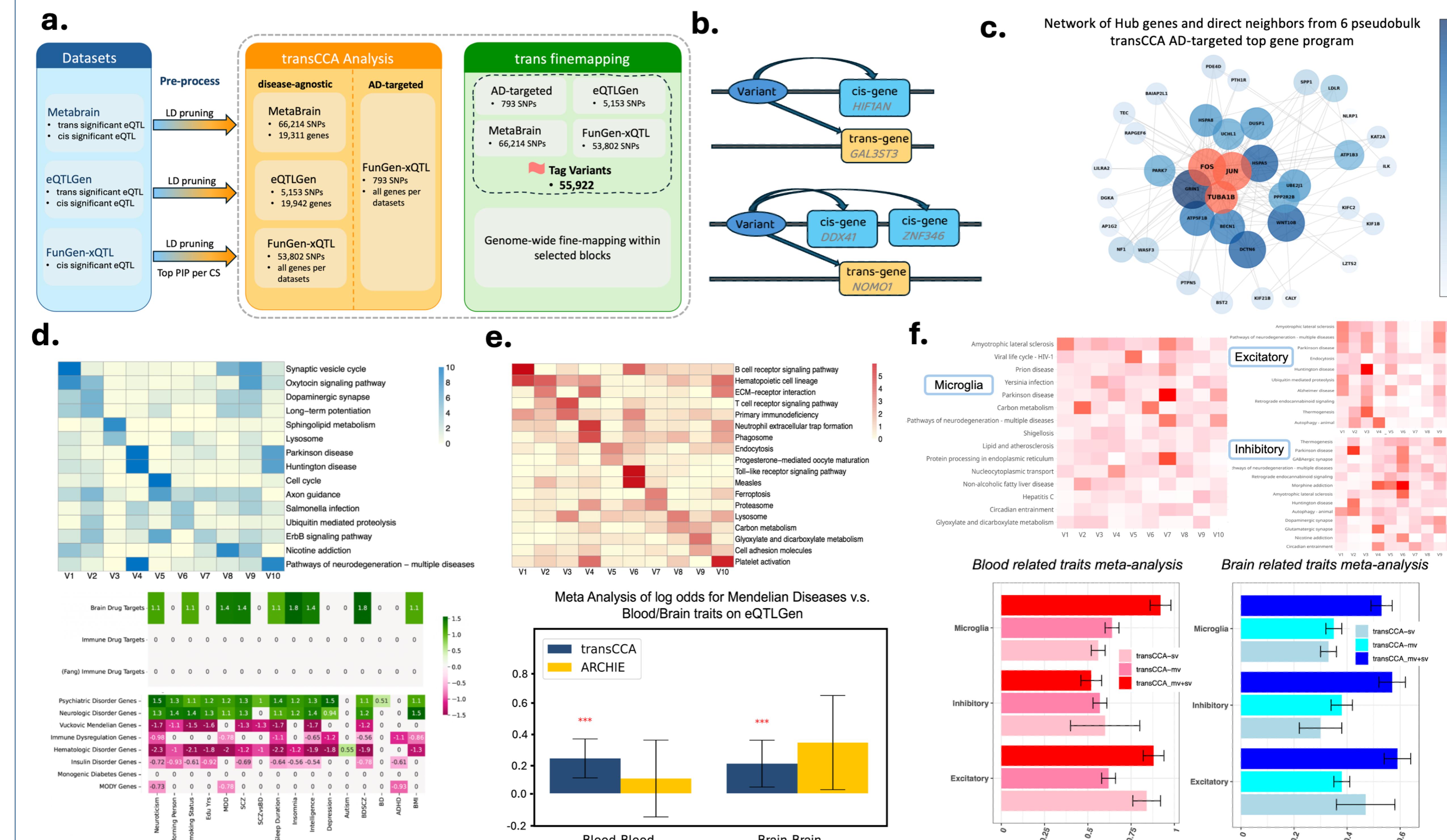


Figure 4: Application to the disease-agnostic and the AD-targeted *trans*-effects analyses.

- Overview of the data resources/preparations, *transCCA*, and *trans*-finemapping workflow.
- Examples of *trans*-finemapping results for AD-relevant genes. Two cases illustrated that the same variant has been identified by fine-mapping through AD-related *cis*-genes and AD-related *trans*-genes (*HIFIAN/GALS3TB*, (*DDX41,ZNF346/NOMO1*).
- AD-targeted *trans*-effects analysis in ROSMAP cell types by *transCCA*. Gene occurrence network in pseudobulk data, based on AD-targeted variants, revealed connections among genes within identified *trans*-gene programs that participate in the same pathway.
- Disease-agnostic *trans*-effects analysis for MetaBrain dataset. *transCCA* programs are enriched in several related neurological pathways (top); *cTrans* aggregation of the *trans*-effects for complex diseases are enriched in psychiatric/neurologic Mendelian disorders (bottom).
- Disease-agnostic *trans*-effects analysis for eQTL-Gene dataset. *transCCA* programs are enriched in several immune-related pathways (top); *cTrans* aggregation of the *trans*-effects from *transCCA* was more enriched in Mendelian Disease and Blood/Brain traits compared to the *trans*-effects from ARCHIE (bottom).
- Disease-agnostic *trans*-effects analysis for two brain neurons and microglia in ROSMAP dataset. *transCCA* programs for the specific brain cell types are enriched in the key neurodegeneration-related pathways (top); integrated *transCCA* programs showed higher disease heritability than both single-context and multi-context *transCCA* programs. Pathway enrichment analysis of excitatory neuron *transCCA* components using variants from FunGen-xQTL.

References

- Liu, X. *et al.* Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034 (2019).
- Wang, L. *et al.* Trans-eQTL mapping in gene sets identifies network effects of genetic variants. *Cell Genomics* **4** (2024).
- Dutta, D. *et al.* Aggregative trans-eQTL analysis detects trait-specific target gene sets in whole blood. *Nature communications* **13**, 4323 (2022).
- Witten, D. M. *et al.* A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).

Acknowledgement

We thank Drs. Masashi Fujita and Philip De Jager for collecting and curating the snRNA-seq data for brain cell-specific eQTL analysis.

Questions on *transCCA*: xc2270@columbia.edu (Xuewei Cao)