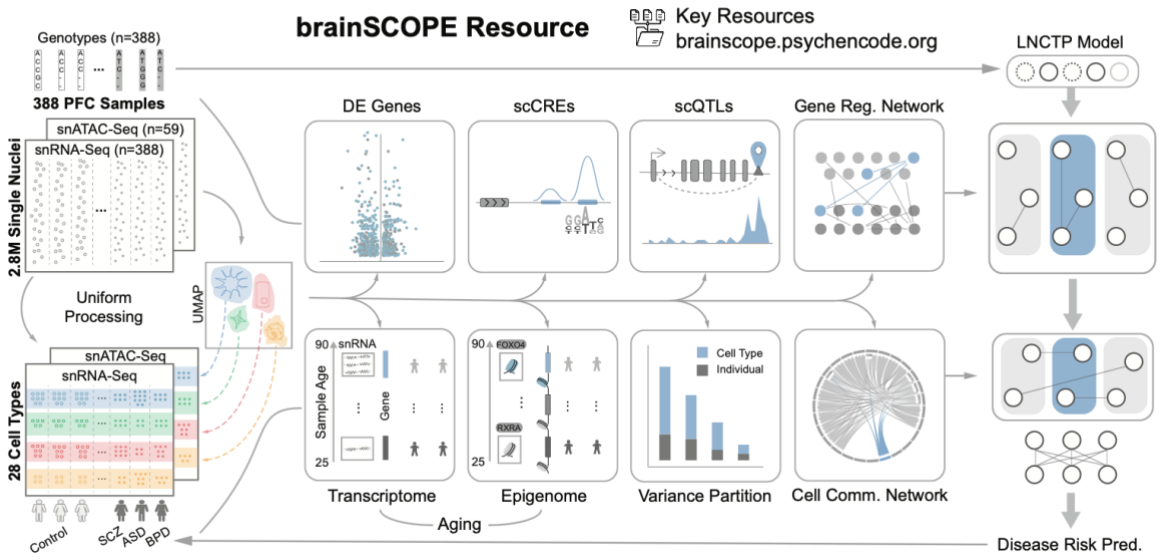
Summary Figure:



Background:

Genetic variants associated with neuropsychiatric disorders impact brain function at multiple levels, from gene expression in individual cells to complex intercellular brain circuits. Previous large-scale studies such as GTEx [1] and ROSMAP [2] have gathered large cohorts of samples from populations and generated extensive brain eQTL catalogs at the bulk-tissue level. However, the lack of cellular resolution has hindered our ability to fully understand the cellular mechanisms underlying brain disorders.

Recently, advancements in single-cell genomics have enabled precise assessments of gene expression and regulation at the single-cell level, enhancing our potential to refine our understanding of brain-related traits. Despite these technological breakthroughs, the challenge remains in assembling sufficiently large and diverse cohorts, and in developing state-of-the-art computational strategies that can support robust analyses and facilitate the development of comprehensive models of brain gene regulation.

Project Description:

Our project filled this gap by leveraging multiple datasets, including those from numerous single-cell experiments (single-nucleus RNA-Seq, ATAC-Seq, Multiome, and genotyping) conducted by the PsychENCODE [3] and AMP-AD consortia [4], to develop an integrative brain resource. This resource, comprising millions of nuclei from approximately 400 individuals, leverages a harmonized cell-annotation scheme aligned with established atlases such as those from the Brain Initiative Cell Census Network (BICCN) [5] to facilitate comparative analyses. Subsequently, we conducted in-depth transcriptomic and epigenetic analyses such as expression variance partitioning, QTL calculation, and developed predictive models for aging and Alzheimer's disease that are tailored to our expansive brain datasets utilizing advanced Bayesian statistical and deep learning methodologies.

We identified single-cell cis-regulatory elements (scCREs) from the chromatin datasets in our resource and validated their functionality through targeted STARR-seq. Furthermore, we compared these scCREs with our generated reference set of over 400,000 open-chromatin regions derived from bulk DNase data. Our snATAC-seq data have enhanced our understanding of the relationship between regulatory regions and brain phenotypes.

We have identified cell type-specific eQTLs by integrating single-cell expression and genotype data from large cohorts, revealing patterns that cannot be observed in bulk gene expression studies. Given the intrinsic sparsity of snRNA-seq data which reduces statistical power especially in rarer cell types, we developed a Bayesian linear mixed-effects model. This model enhances the detection of eQTLs in these less abundant cell types, thereby providing an additional call set. To address the limitations of current cell type-specific eQTLs, which remain at the pseudo-bulk level and lack resolution in showing effect size changes along cell lineages, we further developed a dynamic single-cell eQTL model with a Poisson mixed-effects approach, incorporating a continuous trajectory and a pseudotime-genotype interaction term. This enables the estimation of dynamic eQTLs that vary in effect size along the pseudotime trajectory. Preliminary experiments conducted on IT Neurons have clearly demonstrated the capability of dynamic eQTLs to exhibit varying effects as the neurons undergo pseudo-temporal changes.

We constructed gene-regulatory networks (GRNs) for cell types in the prefrontal cortex by integrating data from our scQTLs, snATAC-seq, transcription-factor-binding sites, and gene co-expression analyses. This has enabled us to elucidate complex network-rewiring across different cell types. Experimental validation such as CRISPR knockouts has also been performed. Additionally, to deepen our understanding of cellular signaling and regulation, we developed a cell-to-cell communication network utilizing publicly available ligand-receptor pairs in conjunction with our snRNA-seq data. This approach allows us to investigate how communication patterns between cells are altered in disease states. Our preliminary findings indicate significant alterations in these communication networks within individuals diagnosed with schizophrenia and bipolar disorder, suggesting a fundamental shift in cellular interactions in these conditions.

Finally, we incorporated various preceding single-cell datasets and derived networks into a unified framework, named the Linear Network of Cell Type Phenotypes (LNCTP), to model and interpret the connections between genotypes and phenotypes. The LNCTP is designed as a conditional energy-based model that encapsulates the joint distribution of key visible variables, conditioned on the genotype, and incorporates additional latent layers. It enables (1) imputation of cell-type-specific and bulk tissue gene expression based on genotypes; (2) prediction of the disorder risks of disorders; (3) identification of key genes and pathways that drive specific phenotypes; (4) simulation of genetic perturbations and quantification of their effects on gene expression or the propensity for certain traits.

Conclusion:

Our population-scale single-cell brain resource has the potential to enhance precision medicine strategies for neuropsychiatric disorders by identifying and prioritizing genes and potential drug targets within specific cell types. This approach aims to tailor therapeutic strategies more effectively to individual needs.

Reference:

[1] GTEx Consortium et al. “Genetic effects on gene expression across human tissues.” Nature vol. 550,7675 (2017): 204-213. doi:10.1038/nature24277

[2] De Jager, Philip L et al. “A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research.” Scientific data vol. 5 180142. 7 Aug. 2018, doi:10.1038/sdata.2018.142

[3] Wang, Daifeng et al. “Comprehensive functional genomic resource and integrative model for the human brain.” Science (New York, N.Y.) vol. 362,6420 (2018): eaat8464. doi:10.1126/science.aat8464

[4] Wan, Ying-Wooi et al. “Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models.” Cell reports vol. 32,2 (2020): 107908. doi:10.1016/j.celrep.2020.107908

[5] BRAIN Initiative Cell Census Network (BICCN). “A multimodal cell census and atlas of the mammalian primary motor cortex.” Nature vol. 598,7879 (2021): 86-102. doi:10.1038/s41586-021-03950-0