Project Research Plan

Microglial Single Cell Transcriptomics Reveal Drivers of Autism Severity - Aadhithya Ramesh

Rationale:

Affecting approximately 1 in 31 children, ASD carries significant lifelong impacts, and early diagnosis is strongly associated with better outcomes. ASD is highly heritable, and the genetic landscape is complex. There are hundreds of rare mutations, from single nucleotide variants to large copy number changes, that have been identified. Notably, transcriptomic studies have revealed reproducible patterns of dysregulated gene expression in ASD, including upregulation of immune and inflammatory pathways, downregulation of neuronal and synaptic markers, and a loss of normal region-specific cortical gene expression programs. By leveraging publicly available single-cell sequencing datasets, this project aims to identify genes and pathways that are differentially expressed in brain tissue and central regulators of disease development. By applying modern computational methods we aim to identify novel molecular signatures associated with ASD severity. This would help determine ASD more accurately, and it would let patients to focus on the level of treatment they need corresponding to their severity of ASD.

Research Question/Hypothesis(es)/Engineering Goal(s):

Individuals with ASD share a convergent set of transcriptomic perturbations that manifest across tissues and cellular compartments. Still, the magnitude and composition of those perturbations differ by ASD severity and is yet to be comprehensively elucidated. Given that microglia are involved in ASD development, it is hypothesized that they will demonstrate significant difference in multiple gene expression programs that evolve across low, intermediate, and high severity disease states.

Materials:

1. Source publications from which primary sequencing data will be downloaded for primary analysis: Wamsely et al., Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. Science 384, eadh2602 (2024). Used Processed single-cell RNA Sequencing Data from 33 ASD and 30 age, sex, time of death-matched healthy control pre-frontal cortex cells. Of the 35 identified cell types in this publication, the current project will focus on microglia,a key transcriptionally divergent cell type.

2. scRNA Seq Data Processing :Seurat (R) — Preprocessing, clustering, Differential Gene Expression Analysis

3. Hardware: Laptop with > 64 GB RAM, 200 GB Storage, R (?4.2) and Python (?3.9), Jupyter Notebook (Anaconda Web)

Procedures:

1. Data acquisition — processed single-cell data (excitatory neurons & microglia only)
a. From identified source publications download processed data objects that include cell metadata and sample/donor IDs.
b.Save Files: expression matrix (normalized counts), cell metadata (cell type, sample_id, diagnosis, age, sex, PMI/death cause, behavioral scores, genotype/15q label), and sample metadata table.

2. Sample & cell-level QC and cohort matching
a. Remove cells with extreme library size / gene counts (e.g., keep cells with 500–10,000 genes detected), high mitochondrial % (exclude >10–15%), and predicted doublets.
b. Exclude donors missing key metadata (diagnosis, age, sex). Compute per-donor cell counts per cell type; exclude donors with too few cells (e.g., <200 cells per cell type).
c. Cohort Creation and Matching: Low, Intermediate, High and beyond baseline clinical symptoms (low), presence of epilepsy only (or intermediate, or epilepsy and intellectual disability

3. Create pseudo-bulk expression matrices (per donor per cell type)
a. For each donor and for each cell type of interest (microglia), sum raw counts across cells to create donor-level

count matrices (genes × donors). This controls donor effects and is preferred for case/control DE. (Seurat: AggregateExpression).
b. Filter genes (retain genes with counts ? 10 in at least 5 donors. Normalize later in DE step.

4. Differential expression (DE) analyses — primary contrasts
a. Perform separate DE analyses per cell type (microglia) using pseudo-bulk counts. Primary comparisons are severity groups comparisons.
b. DE method & model with DESeq2 (R) and doing multiple test: Benjamini–Hochberg FDR; report genes with padj < 0.01 and |log2FC| ? 2.5

5. Pathway & enrichment analyses
a. For each DE gene list, run pathway enrichment (GO Biological Process) using clusterProfiler
b. Test enrichment of known ASD gene sets (SFARI) using Fisher's exact test. Also compute overlap statistics between idiopathic and 15q signatures (rank–rank hypergeometric overlap).

6. Classify Genes by Disease Severity Driver Status
a. Early drivers: Genes significant only in low to intermediate, but not in later transitions
b. Late drivers: Genes significant only in intermediate to high.
c. Cumulative Drivers: Genes significant only in low to how, but not stepwise.
d. Shared across all comparisons

Risk Assessment:

There are no risks, as everything is done online. The data is from past but recent datasets, and the project revolves around using a computer.

Data Analysis:

Seurat will be used to compare gene expression between groups (ASD vs healthy, idiopathic vs 15q duplication, and different age ranges) within excitatory neurons and microglia, and identify differentially expressed genes using standard statistics (reporting padj < 0.05 and |log2 fold change| ? 0.25). These results will be used in tables of genes and shown with graphs such as volcano plots, heatmaps, and bar graphs. Then, pathway enrichment analyses will be run, with GO Biological Process, and overlap tests with SFARI ASD genes using formulas like Fisher's test, and plot the top pathways as bar or dot plots. Classifying data through cohorts.

Bibliography:

1. Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I.-H., Brewster, S. J., Hanson, E., Harris, H. K., Lowe, K. R., Saada, A., Mora, A., Madison, K., Hundley, R., Egan, J., McCarthy, J., Eran, A., Galdzicki, M., Rappaport, L., & Kohane, I. S. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS ONE, 7(12), e49475. https://pubmed.ncbi.nlm.nih.gov/23227143/

2. Emani, P. S., Liu, J. J., Clarke, D., Jensen, M., Warrell, J., Gupta, C., Meng, R., Lee, C. Y., Xu, S., Dursun, C., Lou, S., Chen, Y., Chu, Z., Galeev, T., Hwang, A., Li, Y., Ni, P., Zhou, X., PsychENCODE Consortium, Bakken, T. E., … Gerstein, M. (2024). Single-cell genomics and regulatory networks for 388 human brains. Science, 384(6698), eadi5199. https://www.science.org/doi/10.1126/science.adi5199

3. Wamsley, B., Bicks, L., Cheng, Y, Kawaguchi, R., Quintero, D., Margolis, M., Grundman, J., Liu, J., Xiao, S., Hawken, N., Mazariegos, S., & Geschwind, D. H. (2024). Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. Science, 384(6698), eadh2602. https://doi.org/10.1126/science.adh2602

4. American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). American Psychiatric Publishing.

5. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators, & Centers for Disease Control and Prevention. (2012). Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveillance Summaries, 61(3), 1–19. https://pubmed.ncbi.nlm.nih.gov/22456193/

6. Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., Lee, L.-C., Rice, C.,

Giarelli, E., Kirby, R., Baio, J., Pinto-Martin, J., & Cuniff, C. (2009). Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. Journal of the American Academy of Child & Adolescent Psychiatry, 48(5), 474–483. https://doi.org/10.1097/CHI.0b013e31819b3848 https://pubmed.ncbi.nlm.nih.gov/19318992/

7. Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D. H., & Kriegstein, A. R. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. Science, 364(6441), 685–689. https://doi.org/10.1126/science.aav8130

Additional Project Information: