# Process Essay

## Microglial Single Cell Transcriptomics Reveal Drivers of Autism Severity

## Part One

**1) What inspired you to pursue this work?**

I was inspired to pursue this work because of my younger brother, who was diagnosed with severe ASD at age two. Supporting him at home made me want to understand why ASD presents so differently across individuals and what biology might explain differences in severity. I completed Rice University's *Genome Engineering: Changing the Future of Medicine* certificate, which introduced how molecular tools can translate into healthcare. Dr. Mike Rios, the physician who diagnosed my brother, allowed me to shadow him in the clinic. Over three months, I observed pediatric assessments and saw how early identification and clear guidance can change outcomes for families. These experiences motivated me to move from observation to investigation. I researched ASD mechanisms involving gut microbiome dysbiosis and cytokine signaling and published a review on that topic. I also presented this work at University of Texas Dallas organized by MEANT ("Gut as the New Brain") and earned a Top 5 presenter award, strengthening my ability to communicate complex science clearly. As I studied the literature, genetics repeatedly emerged as a driver of ASD heterogeneity, leading to this project: identifying severity-associated microglial gene programs using public transcriptomic data.

**2) How much time did it take you to complete this work?**

I worked on this project for over four months and spent more than 200 hours reading

primary literature, learning the analysis workflow, debugging code, running statistical comparisons, interpreting pathway results, and refining figures and the written report.

**3) How did you organize the work?**

I organized the work as a stepwise pipeline and tracked progress with a checklist. I broke the project into stages (data setup → quality control → pseudobulk aggregation → differential expression → pathway enrichment → driver classification → figures → writing), and checked off each stage after reruns validated results. I performed a secondary analysis of a publicly available single-cell RNA sequencing dataset from postmortem human prefrontal cortex. I restricted the analysis to microglia and tested whether microglial gene expression follows stepwise patterns across ASD severity transitions. To avoid treating thousands of cells from the same donor as independent samples, I constructed donor-level pseudobulk matrices by aggregating microglial raw counts within each donor. I then used DESeq2 to test low vs intermediate, intermediate vs high, and low vs high contrasts, followed by Gene Ontology Biological Process enrichment for each comparison. Finally, I integrated results across transitions and classified candidate genes as early, late, or cumulative contributors based on when they appeared during progression. Key parameters and thresholds were documented so the workflow can be reproduced using the same input files.

**4) What difficulties did you encounter and how did you handle them?**

Several challenges required careful decisions. Because clinical symptom reporting was incomplete for some donors, I excluded donors with insufficient information to assign severity consistently. To reduce confounding from subtype-specific signals, I excluded donors with the 15q duplication genetic subtype. Technical challenges included quality control and donor-level imbalances in cell counts, which can bias results. I applied standard single-cell QC filters and

used donor-level pseudobulk aggregation to ensure each donor contributed one profile per comparison. I also performed robustness checks: I reran differential expression with alternate thresholds, tested the stability of driver labels, performed leave-one-donor-out analyses, and repeated pseudobulk comparisons after applying minimum cell-count requirements and, when feasible, downsampling to a common cell count. These checks confirmed that the main conclusions and functional themes were stable.

## Part Two

**5) Where was the work completed?**

All work was completed at home on my personal computer in a home setting, entirely online using de-identified, publicly available postmortem transcriptomic data. No new human-subjects data were collected, and no interventions were performed.

**6) Who supervised your work?**

Dr. Pranav Murthy supervised this work as my mentor.

**7) Who helped you and what assistance did they provide? (equipment, ideas, critiques, materials, methodology, etc.)**

All work was completed at home on my personal computer using de-identified, publicly available postmortem transcriptomic data. No new human-subjects data were collected, and no interventions were performed. I performed the full analysis and wrote the paper independently. Mr. Pranav Murthy served as my mentor and provided oversight by reviewing the study plan, checking that methods and interpretations were scientifically reasonable, and suggesting edits for clarity and accuracy. This mentorship was unpaid, and no one else contributed to the analysis or writing.

**8) Was your project completed through a paid program/mentorship?**

This project was not completed through a paid program or mentorship. Mr. Pranav Murthy's mentorship was unpaid.