Single-Nucleus Transcriptomics Identifies Coordinated Changes in Upper-Layer Excitatory Neurons in Major Depressive Disorder

Haitham Abdulaziz Alahmdi, Hugo Tavares

haitham.ab.alahmadi@gmail.com, hm533@cam.ac.uk

Abstract

Using single-nucleus RNA sequencing data from the dorsolateral prefrontal cortex of 17 individuals with major depressive disorder (MDD) and 17 matched controls, we identified upper-layer excitatory neurons (Ex-10-L2-4) as showing the strongest and most consistent cell-type-specific transcriptional signal associated with MDD among 26 cell-type clusters tested. Pseudobulk differential expression analysis (DESeq2) identified 253 significantly altered genes in this population (FDR<0.05), 91.7% of which were downregulated. Direction-specific pathway enrichment analysis converged on three coordinated biological themes among these downregulated genes: glutamate transport (SLC1A2, SLC1A3), microtubule cytoskeletal organization (TUBB family, STMN3/4), and mitochondrial Complex I function (NDUFS2, NDUFS3, NDUFS8). A nucleus-level mixed-effects statistical model, applied independently to two focal clusters, confirmed these findings were not attributable to our data-processing choices. Together, these results point to a coordinated pattern of reduced energy metabolism, impaired structural maintenance, and altered synaptic glutamate handling in upper-layer excitatory neurons in MDD.

Date: August 24, 2026

Contents

1 Introduction **2**

2 Related Work **2**

3 Methodology **2**

3.1 Method details 2

3.2 Evaluation Procedure 3

4 Results **3**

5 Discussion **6**

6 Limitations and Future Work **6**

7 Conclusion **7**

References **7**

1 Introduction

Major depressive disorder (MDD) affects an estimated 5% of adults worldwide and is a leading cause of disability, yet the specific brain cell types driving its underlying molecular pathology remain poorly characterized. Post-mortem studies consistently implicate the prefrontal cortex, but this region contains dozens of distinct cell types, and bulk tissue analyses cannot determine which of these populations drive the transcriptional changes observed in MDD – a signal from one affected cell type can be diluted or masked entirely when averaged across an entire tissue sample.

Single-nucleus RNA sequencing addresses this limitation by measuring gene expression at cellular resolution, enabling cell-type-specific differential expression analysis rather than tissue-level averages. This approach relies on the key assumption that post-mortem transcriptional differences between MDD cases and matched controls reflect disease-associated biological signal rather than solely reflecting confounds such as agonal state, medication history, or post-mortem interval – a limitation inherent to case-control post-mortem designs that is addressed through careful covariate matching but cannot be fully eliminated.

We performed differential expression analysis across 26 cell-type clusters in dorsolateral prefrontal cortex (Brodmann area 9) samples from 17 MDD cases and 17 matched controls, to determine which cell populations show the most robust evidence of altered gene expression in MDD. Among these clusters, upper-layer excitatory neurons (Ex-10-L2-4) showed by far the strongest signal, motivating a deeper investigation into the specific genes and biological pathways driving this pattern.

2 Related Work

Prior single-nucleus studies of the dorsolateral prefrontal cortex in MDD have identified cell-type-specific differential expression concentrated in deep-layer excitatory neurons and oligodendrocyte precursor cells [1]. Subsequent work has extended this approach to female cohorts, chromatin accessibility profiling, and metabolic pathway analysis in the same brain region, consistently confirming that MDD-associated transcriptomic changes are highly cell-type-specific rather than uniform across the tissue. More broadly, selective cellular vulnerability — where specific cell populations show disproportionate molecular disruption relative to others — has been reported across several neuropsychiatric and neurodegenerative conditions, motivating closer examination of which cell types are most and least affected in MDD.

A practical challenge in working with this class of data is that published methodology sections often omit implementation-level detail relevant to reproducing intermediate processing steps (e.g., the exact statistical definition of an expression-detection threshold). Where such detail was missing, we benchmarked multiple plausible implementations against the original study's own reported numbers to select a well-justified approach, rather than assuming a single interpretation without verification.

3 Methodology

3.1 Method details

Preprocessing. Raw count data (approximately 80,000 nuclei, GEO accession GSE144136) were loaded into a Seurat object, annotated with per-nucleus metadata (cell type, individual, condition, sequencing batch), normalized, and scaled, regressing out total UMI count and batch as technical

covariates. Following normalization, seven clusters showing uneven contribution across individuals were excluded, yielding 26 final cell-type clusters for downstream analysis.

Nuclei purification. To ensure high-confidence cell-type assignments prior to differential expression analysis, low-confidence nuclei (e.g., doublets or contaminated droplets) were identified and removed from each cluster using a correlation-based approach: a median expression profile was computed per cluster over a set of highly variable "purification genes," and each nucleus's correlation to its cluster's median profile was calculated. A two-component Gaussian mixture model was then fit to the resulting correlation distribution per cluster to separate high- and low-confidence nuclei, retaining nuclei with high posterior probability ($P > 0.75$) of belonging to the high-correlation component and low probability ($P < 0.25$) of belonging to the low-correlation component.

Gene expression filtering. Following nuclei purification, genes were filtered per cluster based on a minimum expression-detection threshold, consistent with excluding genes detected in fewer than 10% of a cluster's nuclei, applied after purification and independently per cluster.

Pseudobulk differential expression. For each of the 26 purified, gene-filtered clusters, raw counts were aggregated (summed) across all nuclei belonging to the same individual within that cluster, yielding one expression profile per individual per cluster. Differential expression between MDD cases and controls was tested using DESeq2 with the design formula `~batch + condition`, controlling for sequencing batch while testing the effect of diagnostic condition. Effect sizes (log2 fold changes) were shrunk using the `ashr` method to stabilize estimates for lowly expressed genes.

Nucleus-level robustness check. To confirm that the strength of the signal identified in the most affected cluster was not an artifact of the pseudobulk approach, a nucleus-level mixed-effects model was independently applied to two focal clusters: gene expression was modeled as a function of condition and per-nucleus UMI count (fixed effects), with individual and sequencing batch as random effects. All upstream processing was held identical; only the statistical model was changed.

Pathway enrichment analysis. Direction-specific over-representation analysis was performed on significantly downregulated genes using Enrichr, querying the GO Biological Process, GO Cellular Component, GO Molecular Function, and KEGG 2021 Human gene set libraries, to identify coordinated biological themes among the affected genes.

3.2 Evaluation Procedure

Tools and libraries. R (Seurat, DESeq2, ashr, lme4, lmerTest, mixtools, sparseMatrixStats, dplyr), enrichR, Quarto for reproducible, literate reporting.

Reference / ground truth. Statistical significance was assessed using the Benjamini-Hochberg false discovery rate (FDR) correction within each cluster (FDR < 0.05). Pathway enrichment significance was assessed using Enrichr's adjusted p-values within each queried gene set library.

Metrics. For each cluster, the number of significant genes (FDR < 0.05), their direction of change (up/downregulated), and log2 fold change were reported. For pathway enrichment, gene ratio, overlap count, and adjusted p-value were reported per enriched term.

4 Results

Across the 26 cell-type clusters tested, upper-layer excitatory neurons (Ex-10-L2-4) showed by far the strongest transcriptional signal associated with MDD: 253 genes were differentially expressed (FDR < 0.05), compared to a small number or none across the remaining clusters (Figure 1).

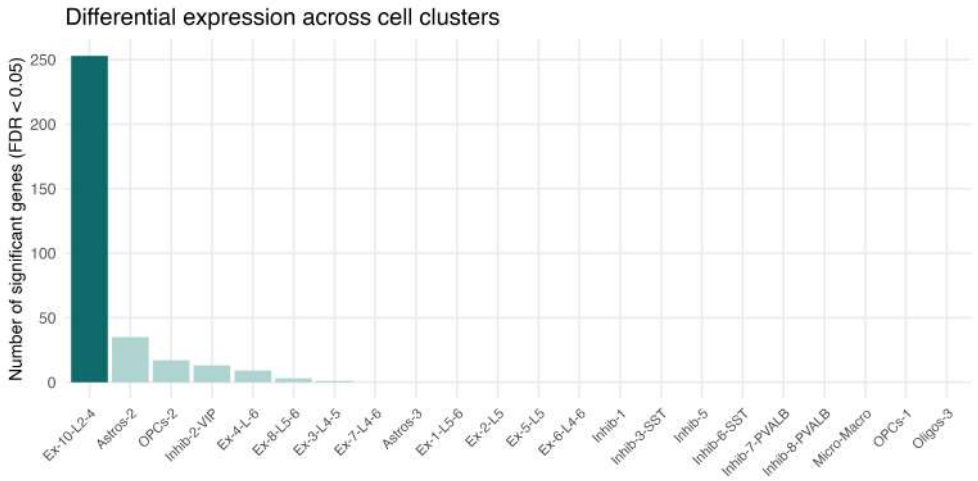


Figure 1 Number of significant genes per cluster (FDR < 0.05). Ex-10-L2-4 shows the largest number of differentially expressed genes among all clusters tested.

Of these 253 genes, 232 (91.7%) were downregulated in MDD relative to controls, indicating a pronounced directional bias (Figure 2).

To confirm that this signal was not an artifact of the pseudobulk approach or upstream data-processing choices, a nucleus-level mixed-effects model matching the original study's statistical specification was applied independently to Ex-10-L2-4 and a second cluster, Ex-7-L4-6, as a targeted cross-check (Table 1). For Ex-7-L4-6, this cross-check identified more significant genes (n=42) than the pseudobulk approach (n=13). For Ex-10-L2-4, it identified a smaller subset of 67 genes, 66 of which were also identified by the pseudobulk approach, indicating that the two gene sets are highly consistent rather than contradictory. This cross-check was conducted solely to confirm that the pseudobulk findings were not attributable to upstream data-processing choices; all reported findings and biological interpretation in this study are based on the pseudobulk analysis alone.

Table 1 Significant genes identified per cluster: pseudobulk analysis versus mixed-effects model cross-check.

Cluster	Pseudobulk (DESeq2)	Mixed-effects cross-check	Overlap
Ex-10-L2-4 (upper-layer excitatory)	253	67	66
Ex-7-L4-6 (deep-layer excitatory)	13	42	8

Direction-specific Gene Ontology enrichment analysis of the 232 downregulated genes in Ex-10-L2-4 identified 22 significant terms (FDR < 0.05) across the Biological Process, Cellular Component, and Molecular Function ontologies (Figure 3), converging on three biological themes: (1) glutamate transmembrane transport (SLC1A2, SLC1A3, TTYH1, PRAF2), (2) microtubule cytoskeletal organization (TUBB, TUBB4A, TUBB4B, TUBA4A, STMN3, STMN4, CDK5), and (3) mitochondrial Complex I / NADH dehydrogenase activity (NDUFS2, NDUFS3, NDUFS8). KEGG pathway analysis independently converged on overlapping gene sets across pathway annotations related to proteasomal protein catabolism and oxidative phosphorylation.

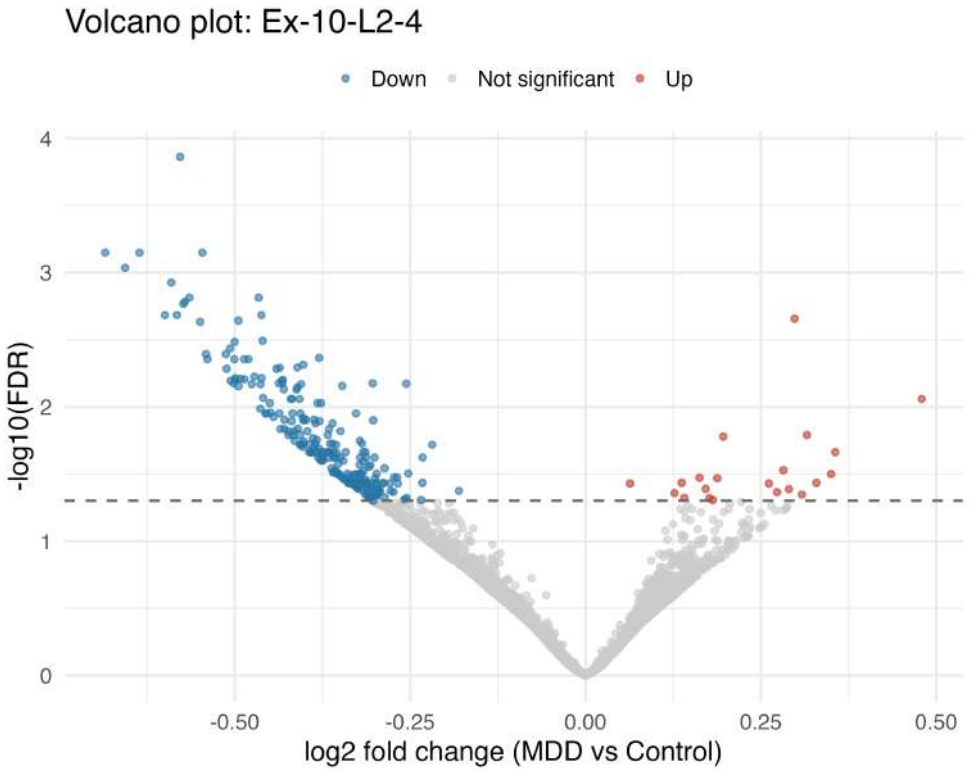


Figure 2 Volcano plot of differential expression in Ex-10-L2-4. 232 of 253 significant genes (91.7%) were downregulated in MDD.

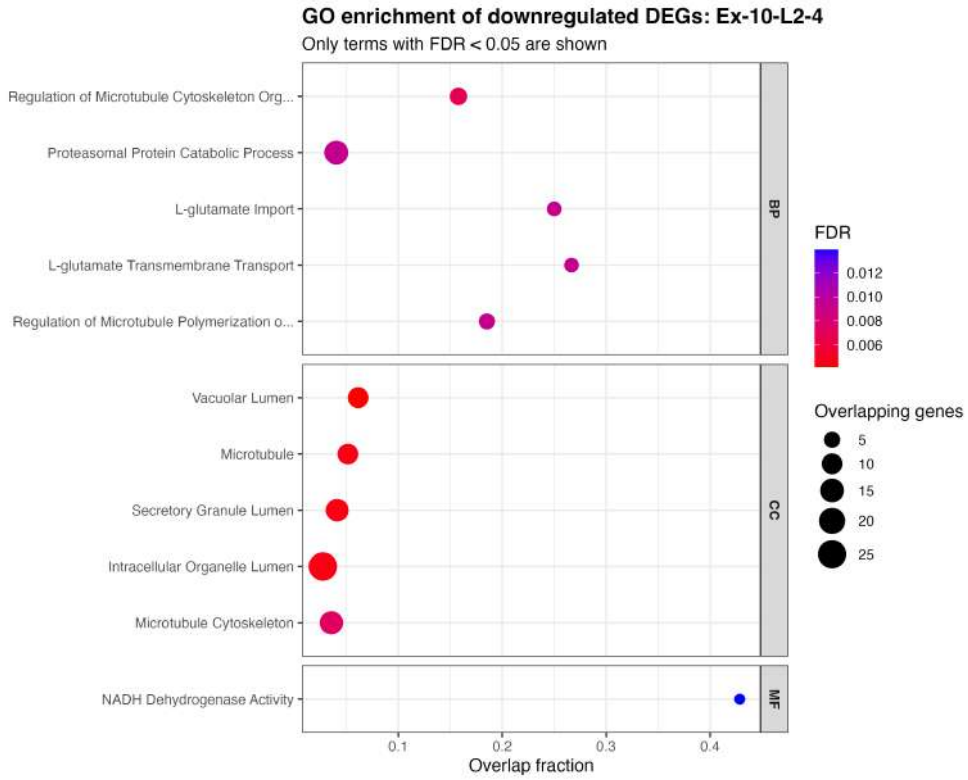


Figure 3 Gene Ontology enrichment of downregulated genes in Ex-10-L2-4 (FDR < 0.05), showing convergence on glutamate transport, cytoskeletal organization, and mitochondrial function.

5 Discussion

The convergence of three independent enrichment analyses (GO Biological Process, GO Cellular Component / Molecular Function, and KEGG) on overlapping gene sets strengthens confidence that the identified patterns reflect genuine coordinated transcriptional signals in upper-layer excitatory neurons, rather than artifacts of a single statistical method or gene list.

The downregulation of SLC1A2 and SLC1A3 — the principal glutamate reuptake transporters (EAAT2/EAAT1) — is notable given the established role of glutamatergic dysregulation in MDD pathophysiology, most directly evidenced by the clinical efficacy of ketamine and its enantiomer esketamine, NMDA receptor antagonists approved for treatment-resistant depression [2]. Reduced expression of these transporters may reflect impaired clearance of synaptic glutamate, a mechanism broadly consistent with the rationale underlying glutamatergic antidepressant treatments, although this study does not test glutamate clearance or receptor activity directly.

The coordinated downregulation of microtubule-related genes (TUBB family, STMN3, STMN4, CDK5) suggests altered capacity for cytoskeletal remodeling in this cell population. Since microtubule dynamics underlie synaptic plasticity — the structural basis for learning, memory, and mood regulation — this finding is consistent with reports of reduced dendritic complexity and synaptic plasticity in postmortem MDD studies more broadly.

The mitochondrial Complex I findings (NDUFS2, NDUFS3, NDUFS8) point to reduced capacity for cellular energy production in this cell population. This result independently converges with separately published reports of oxidative phosphorylation dysregulation in the MDD dorsolateral prefrontal cortex, providing external support for this pattern beyond the present dataset.

Considered together, these three findings — reduced energy metabolism, impaired cytoskeletal maintenance, and altered glutamate handling — point to a coordinated, rather than isolated, pattern of transcriptional change in upper-layer excitatory neurons in MDD. Since both glutamate reuptake and synaptic remodeling are energy-dependent processes, reduced mitochondrial function may represent an upstream contributor linking the other two findings, though this remains speculative without direct measurements of cellular energy metabolism.

These findings should be interpreted as transcriptional associations rather than evidence of causal pathway dysfunction: postmortem case-control transcriptomic data cannot distinguish disease-driving changes from downstream consequences of illness, medication exposure, or other unmeasured confounds.

6 Limitations and Future Work

- Pseudosample size was small for several clusters (e.g., n=7–8 individuals), limiting statistical power to detect differential expression outside the two focal clusters examined in depth.
- These findings are based on postmortem case-control transcriptomic data and reflect statistical association, not causation. Reduced gene expression cannot be distinguished from a cause, a consequence, or an unrelated correlate of MDD using this study design alone.
- Medication status at time of death, while considered in the original study's analysis, was not included in the publicly available GEO dataset (GSE144136) and could not be incorporated as a covariate, representing a potential unaccounted confound.
- Pathway enrichment relied on over-representation analysis (ORA) applied to significantly differentially expressed genes; a ranked gene-set enrichment analysis (GSEA) using the full tested gene list was not performed due to time constraints and represents a natural extension of this work.
- The nucleus-level mixed-effects model used as a robustness check followed a linear model

specification rather than a negative-binomial formulation, which may be more statistically appropriate for count-based expression data; this was a deliberate choice made to isolate the effect of statistical method from model family in the robustness comparison.

- Findings are based on mRNA expression levels, not protein abundance or transporter activity directly; reduced transcription does not automatically confirm a corresponding reduction in protein function, and would benefit from orthogonal validation (e.g., proteomic or functional assays).
- A natural next step is to test whether the identified glutamate-handling, cytoskeletal, and mitochondrial gene signatures are specific to upper-layer excitatory neurons or are shared, to varying degrees, across other excitatory neuron subtypes in the same dataset.

7 Conclusion

This study identified upper-layer excitatory neurons (Ex-10-L2-4) as showing the most robust and consistent transcriptional signal associated with major depressive disorder among 26 cell-type clusters examined in the human dorsolateral prefrontal cortex. Pathway enrichment analysis revealed a coordinated pattern of downregulation across three biologically connected processes: glutamate transport, cytoskeletal organization, and mitochondrial energy metabolism. A nucleus-level statistical robustness check confirmed that this signal was not attributable to the choice of differential expression method, strengthening confidence in the finding. Together, these results point toward impaired synaptic glutamate handling, reduced structural plasticity, and diminished energy metabolism as a coordinated feature of upper-layer excitatory neuron dysfunction in MDD, motivating further investigation into whether this pattern extends to other excitatory neuron subtypes and whether it reflects a cause or consequence of the disorder.

References

[1] Corina Nagy et al. “Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons”. In: *Nature Neuroscience* 23 (2020), pp. 771–781.

[2] John H. Krystal et al. “Ketamine: A Paradigm Shift for Depression Research and Treatment”. In: *Neuron* 101.5 (2019), pp. 774–778.