

Temporal Dynamics of Gene Expression in Parkinson's Disease

Sowmya Janmahanthi, Behrouz Delfanian, Barak Landsman, Vani Pant,
Costin-Andrei Taulescu

University of Luxembourg

Abstract—Parkinson's disease (PD) is a neurodegenerative disorder closely linked to mitochondrial dysfunction, with mutations in the PINK1 gene impairing mitochondrial quality control in dopaminergic neurons. In this study, we used time-resolved single-cell RNA sequencing (scRNA-seq) to investigate how PINK1 loss affects human induced pluripotent stem cell (iPSC) differentiation into dopaminergic neurons. Control and PINK1 mutant cells were profiled across five key developmental stages, from pluripotency to mature neurons. Comprehensive computational analyses—including data integration, dimensionality reduction, clustering, differential gene expression, trajectory analysis, and functional enrichment—revealed that PINK1 mutation does not block dopaminergic differentiation but reshapes its timing and transcriptional programs. PINK1 mutant cells exhibited early mitochondrial stress, reduced proliferative signaling, metabolic rewiring toward glycolysis, and precocious neuronal differentiation during progenitor stages. Despite these alterations, mature PINK1 neurons showed strong dopaminergic marker expression alongside Parkinson's disease–relevant changes, including elevated SNCA and altered stress-response genes. Together, these findings indicate that PINK1 loss drives developmental heterochrony and compensatory metabolic adaptations that preserve differentiation outcomes but may increase late-stage neuronal vulnerability, providing mechanistic insight into PINK1-associated Parkinson's disease pathogenesis.

Index Terms—Parkinson's disease, PINK1, single-cell RNA-seq, dopaminergic neurons, mitochondrial dysfunction, iPSC differentiation

I. INTRODUCTION

A. Proteins

A protein is a large biomolecule composed of a chain of amino acids. They are made through a very complicated process often referred to as the "Central dogma of molecular biology". To understand it however, one must first comprehend what DNA is;

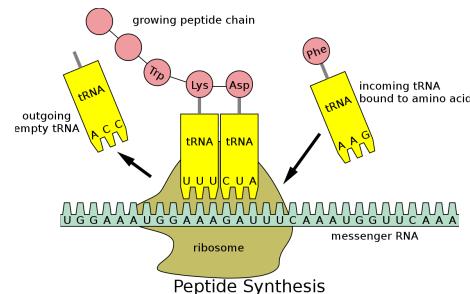
1) *Pre-Dogma*: Deoxyribonucleic acid (not to be confused by ribonucleic acid), more commonly known as DNA (and not RNA) is the genetic code of a cell. It dictates how the cell operates at any given moment. In Eukaryotic cells, such as most human and animal cells, the DNA is contained within the cell nucleus. Meanwhile, in Prokaryotic cells which have no nucleus, the genetic material is suspended in the cell's cytoplasm.

2) *The central dogma of molecular biology*:

a) *Transcription*: The process begins with a promoter region in the DNA called a "TATA box" for its chemical composition, involving an alternating pattern of Thymine and

Adenine molecules. This attracts the **polymerase** protein to bind to the DNA molecule and begins the **transcription** process by first pulling the two strands of the molecule apart before section by section, it ejects the complementary nucleotides of the strand, eventually resulting in a long string of messenger RNA (mRNA).

b) *Translation*: The mRNA proceeds to make its way outside the nucleus until it arrives at the **ribosome**, another imperative protein to the process. Within the ribosome, the mRNA gets split and matched with triplets of the complementary nucleotides (A with U, C with G). Those triplets are called **codons**, they are part of the **translation RNA** (tRNA for short) which have the nucleotides on one side of the molecule and amino acids on the other;



Once the tRNA binds to the mRNA, it gets into a structure which allows the amino acids to also bind to each other. They are what then creates the primary structure of a protein.

c) *Protein structure*: In short, proteins are divided into four structural categories which they go through while getting constructed within the ribosome.

Primary structure: The first structure a protein takes, it is simply a one-dimensional long string of amino acids.

Secondary structure: After a certain point, sections of the protein begin folding, creating what are referred to as α -helices and β -pleated sheets (or simply β sheets).

Tertiary structure: Once the protein gets to a specific size, it begins folding recursively, first locally and then globally, creating a three-dimensional structure of a group of α -helices and β -sheets.

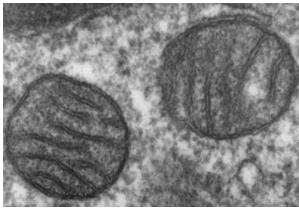
Quaternary structure: Lastly, several tertiary structures fold onto one another and create a "four-dimensional" structure composed of mostly tertiary structures.

B. What makes proteins so important?

While cells are the building blocks of life, proteins are the building blocks of cells. Every protein is a relatively large, specialized molecule the structure of which determines its functionality. For example, the polymerase is a specialized protein which transcribes DNA in the beginning of protein synthesis. Another good example is the PINK1 protein, which binds to old mitochondria and labels them for recycling.

C. Mitochondria

As many know, the mitochondria is the powerhouse of the cell. It is an *organelle* which exists to supply energy in the form of a molecule called Adenosine Tri-Phosphate (ATP for short). Briefly, it works very similarly to a wind or power generator. By using diffusion of positively-charged ions, it turns Adenosine Di-Phosphate into Adenosine Tri-Phosphate, in other words, spending energy to create potential energy, or spending ATP to create ADP and releasing ions. This is often seen in the flagella of bacteria, using potential energy to create motion. The mitochondria is built to maximize surface area as to allow for as much room for ATP-synthase enzymes to operate in parallel.

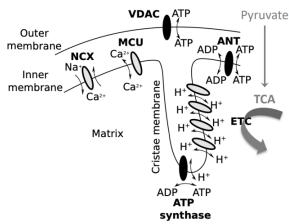


D. Parkinson's Disease

Parkinson's Disease is a neuro-degenerative condition caused by the accumulation of dopaminergic neurons in the brain.

It is important to discuss Parkinson's openly and freely as it is a very real condition affecting many people in order to increase awareness in the general, healthy public.

1) *What causes Parkinson's?*: Due to a variety of factors such as lifestyle, pre-existing conditions or genetics, the dopaminergic neurons, the neurons in the brain responsible for a variety of functions such as motor control, certain senses and emotions die off due to a mutation in the PINK1 gene. The specific part in the DNA responsible for PINK1 protein synthesis. Because the PINK1 protein is damaged, it binds to old mitochondria but they do not undergo mitophagy (recycling) due to the PINK1 protein not being recognized. Dopaminergic neurons are the most vulnerable to this because of their high energy requirements. Neurons use electric signals to communicate, but dopaminergic neurons are especially energy intensive because Parkinson's causes a degradation of their axons, making them even less energy-efficient, which causes a cascade effect.



II. METHODOLOGY

We applied an integrative single-cell RNA sequencing (scRNA-seq) analysis to study transcriptional changes during the differentiation of control and PINK1 mutant human iPSC-derived dopaminergic neurons. Rigorous quality control, normalization, data integration, and unsupervised clustering ensured high data quality and reliable cross-condition comparisons. Differential expression, temporal trajectory, and functional enrichment analyses identified key genes and pathways associated with mitochondrial dysfunction and Parkinson's disease. This comprehensive approach enabled robust identification of biologically meaningful molecular mechanisms and potential therapeutic targets.

A. Experimental Design and Data Acquisition

Single-cell RNA sequencing (scRNA-seq) data were obtained from control (CTL) and PINK1 mutant human induced pluripotent stem cell (iPSC) lines undergoing directed differentiation toward dopaminergic neurons. The scRNA-seq dataset was collected longitudinally across multiple time points over a 57-day period, enabling the characterization of dynamic transcriptional changes during disease progression. Five critical time points spanning dopaminergic differentiation were sampled to enable analysis of Parkinson's disease progression: Day 0 (pluripotent stage), Day 18 (neural progenitors), Day 25 (dopaminergic progenitors), Day 37 (early neurons) and Day 57 (mature dopaminergic neurons). This temporal resolution enabled comprehensive tracking of transcriptional changes throughout the differentiation trajectory.

B. Single-Cell RNA-Seq Data Processing

1) *Quality Control and Normalization*: Quality control measures were applied to remove low-quality cells and ensure the reliability of downstream analyses by filtering data based on key quality metrics. Count data were normalized using log-normalization (scale factor = 10,000) via the `NormalizeData` function. Both raw (RNA) and imputed (RNA_imputed) assays were normalized for downstream analyses. These quality control steps ensured retention of high-quality cells while removing technical noise, providing a robust foundation for downstream analyses. Later time points exhibited more uniform quality metrics, indicating improved technical consistency. Normalization corrected for differences in sequencing depth and RNA content and reduced gene-level technical variability, enabling meaningful cross-cell comparisons and reliable clustering and differential gene expression analyses.

2) *Data Integration*: To enable robust cross-condition comparisons and account for batch effects between CTL and PINK1 samples, the integrated assay served as the foundation for all dimensionality reduction and comparative analyses.

C. Dimensionality Reduction and Visualization

1) *Principal Component Analysis*: Dimensionality reduction was an initial step in the analysis to simplify the high-dimensional scRNA-seq data. Principal Component Analysis

(PCA) was first applied to identify the major sources of variation and retain the most informative features for downstream analyses. PCA was performed on scaled, integrated data using RunPCA with 50 principal components.

2) *UMAP Projection*: For visualization, Uniform Manifold Approximation and Projection (UMAP) was subsequently used to embed cells into a low-dimensional space based on transcriptional similarity, enabling intuitive visualization of cellular relationships and potential structure within the data. UMAP was computed using RunUMAP with parameters optimized for trajectory analysis: `dims = 1:50`, `n.neighbors = 50`, and `seed.use = 18` (for reproducibility). UMAP projections were visualized with cells colored by differentiation day or genotype using DimPlot.

D. Unsupervised Clustering

Unsupervised clustering was then performed to group cells according to their gene expression profiles. This approach enabled the identification of distinct cell populations, including subsets enriched for dopaminergic neuron markers and signatures of mitochondrial dysfunction, providing insight into cellular heterogeneity and Parkinson's disease-related progression. Cell clustering was performed using the Louvain algorithm. A shared nearest neighbor (SNN) graph was constructed using `FindNeighbors` with PCA dimensions 1-20. Clusters were identified at resolution 0.4 using `FindClusters`, yielding distinct cell populations. Cluster-specific marker genes were identified using `FindAllMarkers` with parameters: `only.pos = TRUE`, `min.pct = 0.25`, `logfc.threshold = 0.25`, and `random.seed = 42`.

E. Differential Gene Expression Analysis

1) *Day-Wise DEG Identification*: Differential expression analysis was conducted separately for each time point comparing CTL versus PINK1 conditions using the Wilcoxon rank-sum test (`test.use = "wilcox"`) implemented in Seurat's `FindMarkers` function. Parameters included: `logfc.threshold = 0.25`, `min.pct = 0.05`, and `densify = TRUE`. Genes were considered significantly differentially expressed at absolute \log_2 fold change > 0.5 .

2) *Temporal DEG Dynamics and Overlap Analysis*: DEGs were categorized as "Up in CTL" or "Up in PINK1" based on fold change direction. Temporal dynamics were quantified by counting DEGs per category at each time point. Overlap analysis computed the number of shared DEGs between all time point pairs to assess temporal consistency of transcriptional changes.

F. Gene Expression Trajectory Analysis

For key marker genes and Parkinson's disease-associated genes, mean expression and standard error were calculated per condition and time point using the imputed RNA assay. Expression trajectories were visualized as line plots with error bars representing standard error of the mean (SEM).

G. Gene Ontology and Pathway Enrichment Analysis

1) *g:Profiler Enrichment*: Gene Ontology (GO) enrichment analysis was performed using `gprofiler2`. For each time point, genes significantly upregulated in CTL or PINK1 (adjusted $p < 0.05$, $|\log_2\text{FC}| > 0.5$) were tested for enrichment in GO Biological Process (GO:BP), Cellular Component (GO:CC), and Molecular Function (GO:MF) terms. Enrichment significance was assessed at FDR-corrected $p < 0.05$ using the `g:SCS` algorithm. Mitochondrial-related pathways were identified by filtering GO terms containing keywords: "mitochondri", "oxidative", "respiratory", "electron transport", or "ATP".

2) *STRING Protein-Protein Interaction Network Analysis*: Protein-protein interaction (PPI) network analysis was conducted using the STRING database v12.0. The top 100 genes upregulated in CTL and PINK1 at Day 25 (peak divergence) were submitted to STRING for network construction. Analysis parameters: `organism = Homo sapiens`, `minimum interaction score = 0.90` (highest confidence), `network type = full STRING network`. Network statistics (nodes, edges, clustering coefficient, PPI enrichment p -value) were extracted. Functional enrichment within networks was assessed using STRING's built-in tools for GO Biological Process, KEGG pathways, and Reactome pathways (FDR < 0.05).

H. Statistical Analysis and Visualization

All statistical analyses were performed in R Version 4.4.1. Visualizations were generated using `ggplot2`, `pheatmap`, and `patchwork`. A consistent color scheme was applied throughout: green (#4DAF4A) for CTL and pink/magenta (#E91E8C) for PINK1. High-resolution figures (300 DPI) were exported as PNG files. Statistical significance was defined as adjusted $p < 0.05$ unless otherwise stated.

I. Computational Resources

All analyses were executed on a high-performance computing cluster (8 CPU cores, 64 GB RAM) using parallel processing enabled via the `future` package (`plan("multicore", workers = 8)`).

III. RESULTS

A. Overview of Dataset

We analyzed 27,455 high-quality single cells (CTL: 13,905 cells; PINK1: 13,550 cells) spanning five critical differentiation time points from pluripotent stem cells to mature dopaminergic neurons. Quality control ensured robust detection of gene expression patterns with balanced representation across genotypes and time points.

B. Differentiation Trajectories Show Preserved Architecture with Subtle Divergence

UMAP analysis (Fig. 2) demonstrates that both CTL and PINK1 cells follow a continuous developmental trajectory, validating the differentiation protocol. Early pluripotent cells (Day 0) occupy distinct regions, while later time points

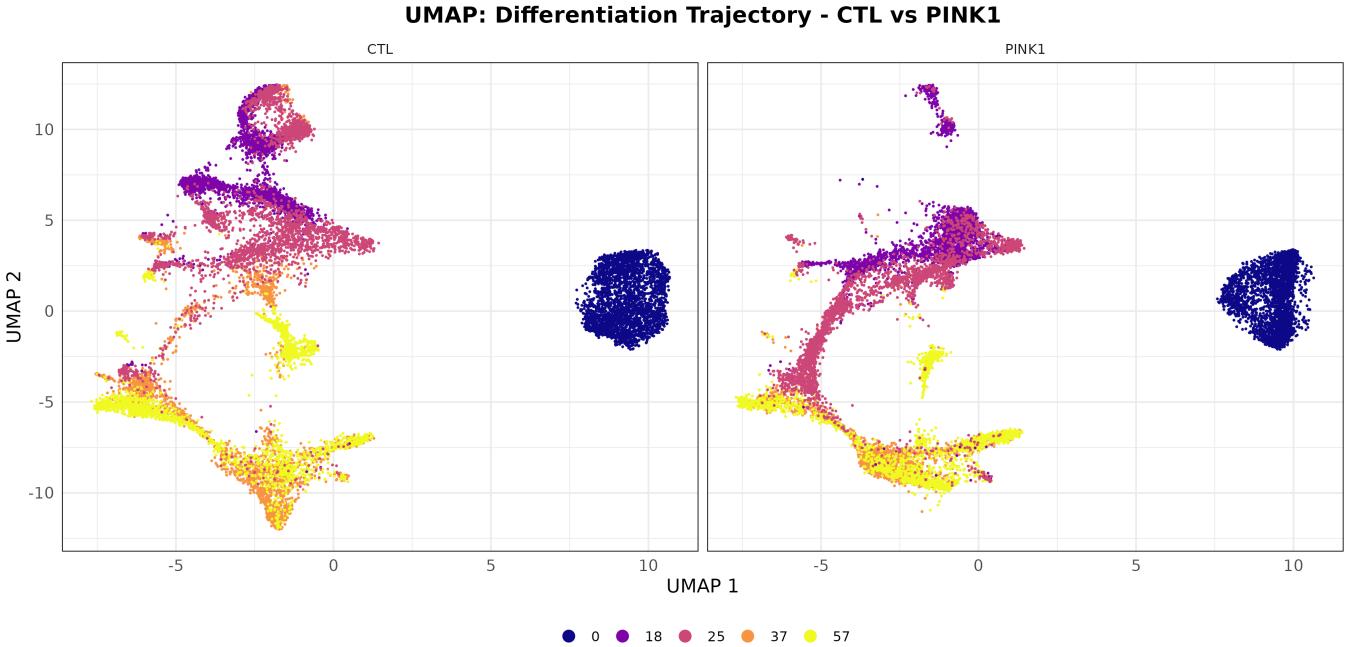


Fig. 2: UMAP of CTL and PINK1 cells across five differentiation days, colored by day.

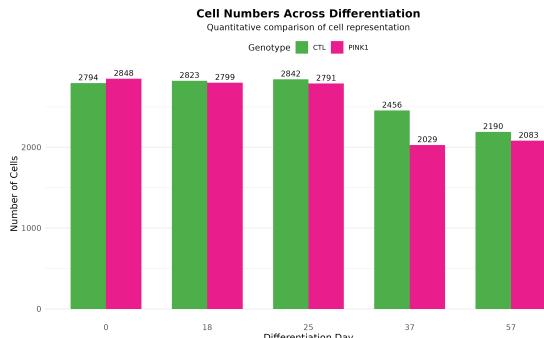


Fig. 3: Number of CTL and PINK1 cells per differentiation day

progressively separate along the UMAP axes, reflecting transcriptional remodeling during differentiation. Critically, Day 57 mature neurons form discrete clusters in both genotypes, confirming terminal differentiation. The preservation of overall trajectory structure in PINK1 mutants indicates that the differentiation program remains fundamentally intact despite PINK1 loss, though subtle distributional differences hint at altered differentiation kinetics.

C. Cell Survival and Proliferation Are Unaffected by PINK1 Mutation

Cell composition analysis (Fig. 3) reveals no significant genotype-dependent differences in cell numbers, indicating that PINK1 mutation does not cause catastrophic cell loss during the 57-day differentiation protocol. The slight reduction at later stages (Days 37-57) occurs equivalently in both

genotypes and likely represents physiological developmental pruning during neuronal maturation. This finding is critical as it demonstrates that any observed transcriptional changes reflect primary PINK1-dependent effects rather than secondary consequences of cell death or proliferative defects.

D. Differential Expression Dynamics and Overlap

The temporal DEG analysis (Fig. 4a) unveils a striking pattern of transcriptional divergence. The presence of substantial DEGs at Day 0 (1,063 total) demonstrates that PINK1 mutation affects even the pluripotent state, potentially priming cells for altered developmental responses. DEG counts peak during the neural progenitor stage (Day 25: 1,409 total DEGs), suggesting that this developmental window is particularly sensitive to mitochondrial dysfunction. Most remarkably, the dramatic asymmetry with CTL consistently showing more upregulated genes implies that PINK1 loss leads to transcriptional attenuation rather than aberrant activation. By Days 37-57, PINK1-upregulated genes drop to minimal levels (16-21), potentially reflecting exhaustion of compensatory mechanisms or terminal differentiation constraints.

DEG overlap analysis (Fig. 4b) reveals the temporal structure of transcriptional divergence. The clustering of Days 18-37 with high mutual overlap (505-1,121 shared DEGs) identifies a "critical window" where PINK1-dependent transcriptional differences are maximal and persistent. The relative isolation of Day 0 (low overlap with later stages) suggests that pluripotent-stage alterations represent a distinct transcriptional state, potentially reflecting metabolic differences in stem cells versus differentiated progeny. Day 57's reduced overlap with earlier stages indicates that mature neurons establish

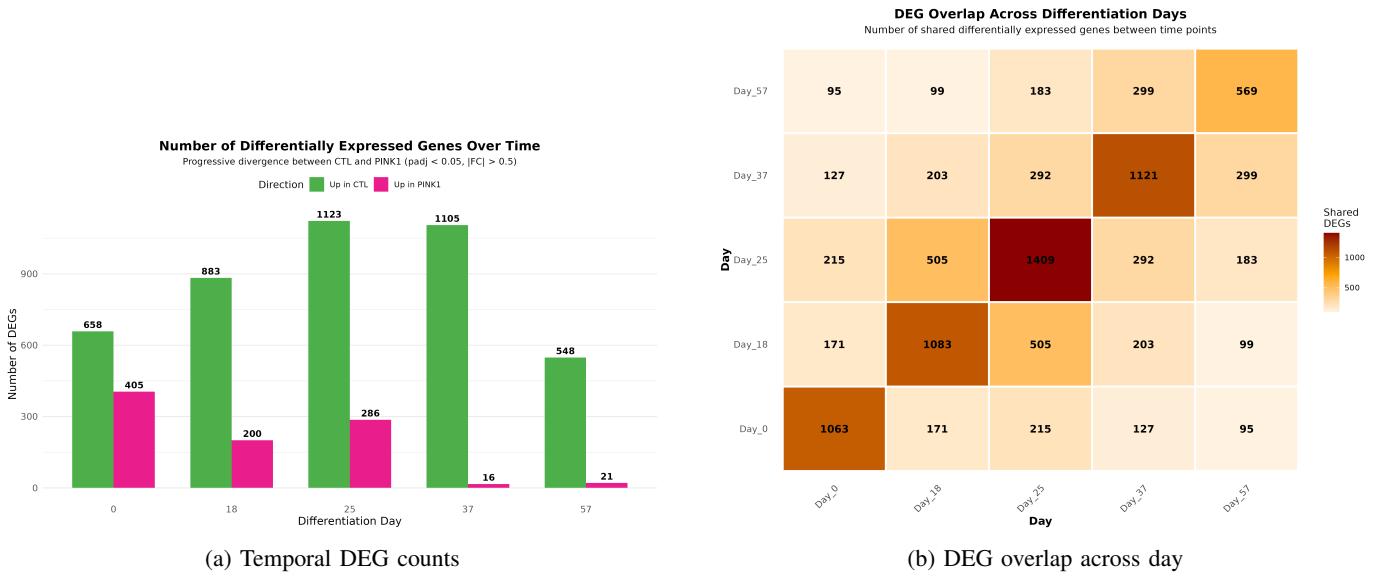


Fig. 4: Temporal structure of differential expression across differentiation.

a unique transcriptional profile where many earlier DEGs resolve, consistent with the dramatic reduction in PINK1-upregulated genes.

E. Volcano Plots Reveal Stage-Specific DEG Signatures

Volcano plot analysis (Fig. 5) provides gene-level resolution of temporal transcriptional dynamics. Day 0 shows significant baseline alterations including glycolytic genes (LDHA) upregulated in PINK1, suggesting early metabolic adaptation. Days 18–25 display the widest fold-change distributions, with CTL showing strong upregulation of proliferation markers (MKI67, TOP2A) and PINK1 showing neuronal differentiation genes (DDC, ASCL1, STMN2). By Days 37–57, the volcano plots become increasingly asymmetric, with few genes significantly upregulated in PINK1, reinforcing the transcriptional dampening phenotype.

F. Hierarchical Clustering Reveals Distinct Temporal Gene Expression Modules

The DEG heatmap (Fig. 6) provides a comprehensive view of temporal gene expression dynamics. Hierarchical clustering identifies several key gene modules: (1) **Proliferation module**: CTL-enriched genes (MKI67, BIRC5, HHEX, DDX3Y) showing peak expression Days 18–37, confirming active cell cycle progression; (2) **Metabolic adaptation module**: PINK1-enriched glycolytic genes (PDK4, MTRNR2L1) expressed throughout differentiation, reflecting compensatory metabolic rewiring; (3) **Neuronal differentiation module**: PINK1-enriched neuronal genes (NR4A2, PCDHGA3, COL3A1) with late-stage expression, indicating enhanced terminal differentiation capacity; (4) **Developmental regulator module**: Transcription factors and signaling molecules (MECOM, SIX3, NR5A2) with complex temporal patterns. The modular organization suggests that PINK1 mutation affects multiple coordi-

nated transcriptional programs rather than scattered individual genes.

G. Marker and PD Gene Trajectories

Gene trajectory analysis (Fig. 7a) reveals nuanced kinetic alterations rather than complete differentiation failure. Pluripotency markers show largely appropriate downregulation, though SOX2's delayed decline in PINK1 may reflect extended progenitor states or delayed neuronal maturation. The earlier LMX1A peak in PINK1 (Day 18 vs Day 25) suggests that specification pathways are intact and possibly accelerated, contrasting with the overall transcriptional dampening observed. Most remarkably, the dramatic upregulation of TH and NR4A2 in PINK1 at Day 57 represents a striking compensatory response: despite mitochondrial dysfunction, PINK1 neurons achieve and potentially exceed normal dopaminergic identity. This may represent an adaptive mechanism where enhanced dopaminergic transcription compensates for metabolic limitations, or could reflect dysregulation of feedback mechanisms normally restraining dopaminergic gene expression.

Analysis of PD-associated genes (Fig. 7b) reveals disease-relevant molecular changes. The most striking finding is SNCA (α -synuclein) upregulation in PINK1 at Days 37–57. Given that α -synuclein aggregation is the pathological hallmark of PD and SNCA multiplication causes familial PD, this elevation could represent a direct pathogenic mechanism linking PINK1 dysfunction to sporadic PD. The divergent trajectory of PARK7 (DJ-1) sharply declining in PINK1 suggests impaired oxidative stress responses, as DJ-1 is a critical oxidative damage sensor. ATP13A2 elevation in PINK1 may reflect compensatory lysosomal/autophagic upregulation attempting to clear damaged mitochondria. VPS35's transient upregulation suggests retromer pathway engagement, potentially related to mitochondrial quality control or synaptic vesicle trafficking.

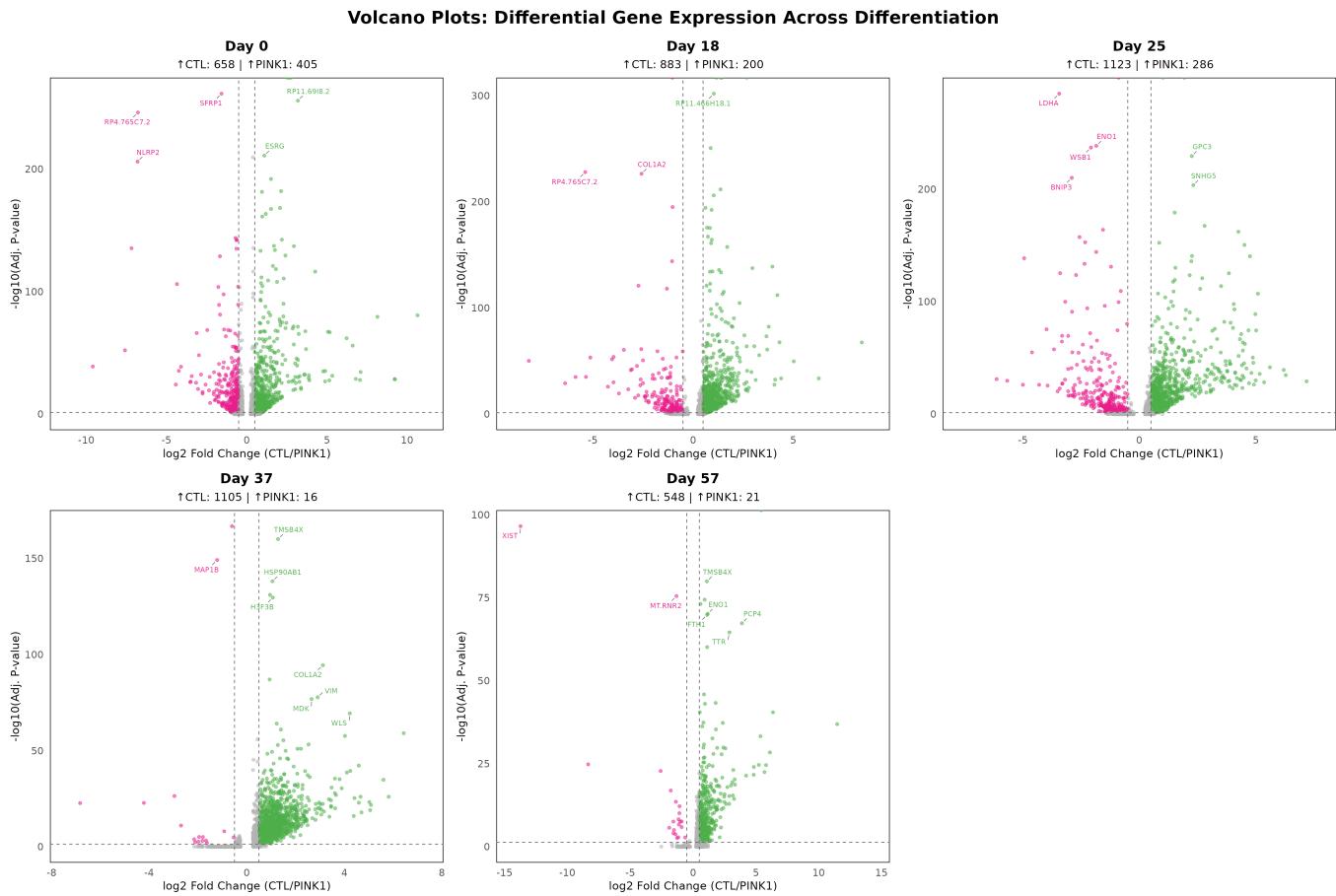


Fig. 5: Volcano plots of CTL vs PINK1 DEGs at each differentiation day.

H. GO Enrichment Reveals Divergent Biological Process Engagement

GO Biological Process enrichment (Fig. 8a) reveals fundamentally different cellular priorities between genotypes. CTL cells at Days 18-25 are overwhelmingly focused on cell cycle progression, with mitotic processes enriched to extraordinary significance levels ($p < 10^{-20}$, 40+ genes). This indicates robust progenitor expansion a hallmark of normal neural development. In stark contrast, PINK1 cells at the same stages show enrichment of neuronal differentiation processes (neurogenesis, neuron differentiation), suggesting premature exit from the cell cycle and precocious commitment to neuronal fate. By Day 37, CTL cells transition to energy metabolism pathways, indicating metabolic maturation concurrent with late-stage differentiation. PINK1 cells at late stages show enrichment of morphogenetic processes (neuron projection development), consistent with terminal differentiation. This temporal mismatch PINK1 cells "skip" the proliferative phase and enter differentiation earlier provides a mechanistic explanation for the observed transcriptional patterns and may reflect adaptive responses to mitochondrial stress favoring differentiation over division.

Cellular Component enrichment (Fig. 8b) provides spatial

context for the observed transcriptional changes. The strong mitochondrial enrichment in PINK1 cells at Day 0 confirms that PINK1 loss causes baseline organellar alterations detectable even before differentiation begins. CTL cells show predominant nuclear enrichment at Days 18-25, consistent with active gene regulation during proliferation. In contrast, PINK1 cells show cytoplasmic and junction-related enrichment at the same stages, aligning with earlier neuronal differentiation. By Days 37-57, PINK1 cells show extensive enrichment of neuronal-specific compartments synapses, axons, growth cones demonstrating that despite mitochondrial dysfunction and altered kinetics, PINK1 neurons successfully establish complex neuronal architectures.

I. Top Consistently Regulated Genes Identify Key Drivers

Analysis of consistently regulated genes (Fig. 9) identifies potential key drivers of phenotypic divergence. CTL-upregulated genes are enriched for transcription factors involved in developmental patterning (PITX2, SIX3, HHEX, ZIC1/4, EMX2), suggesting that CTL cells maintain broader developmental potential or engage distinct patterning programs. The enrichment of metabolic genes (MRPL23, RRM2) aligns with higher proliferation rates. PINK1-upregulated genes show striking enrichment of non-coding RNAs (XIST,

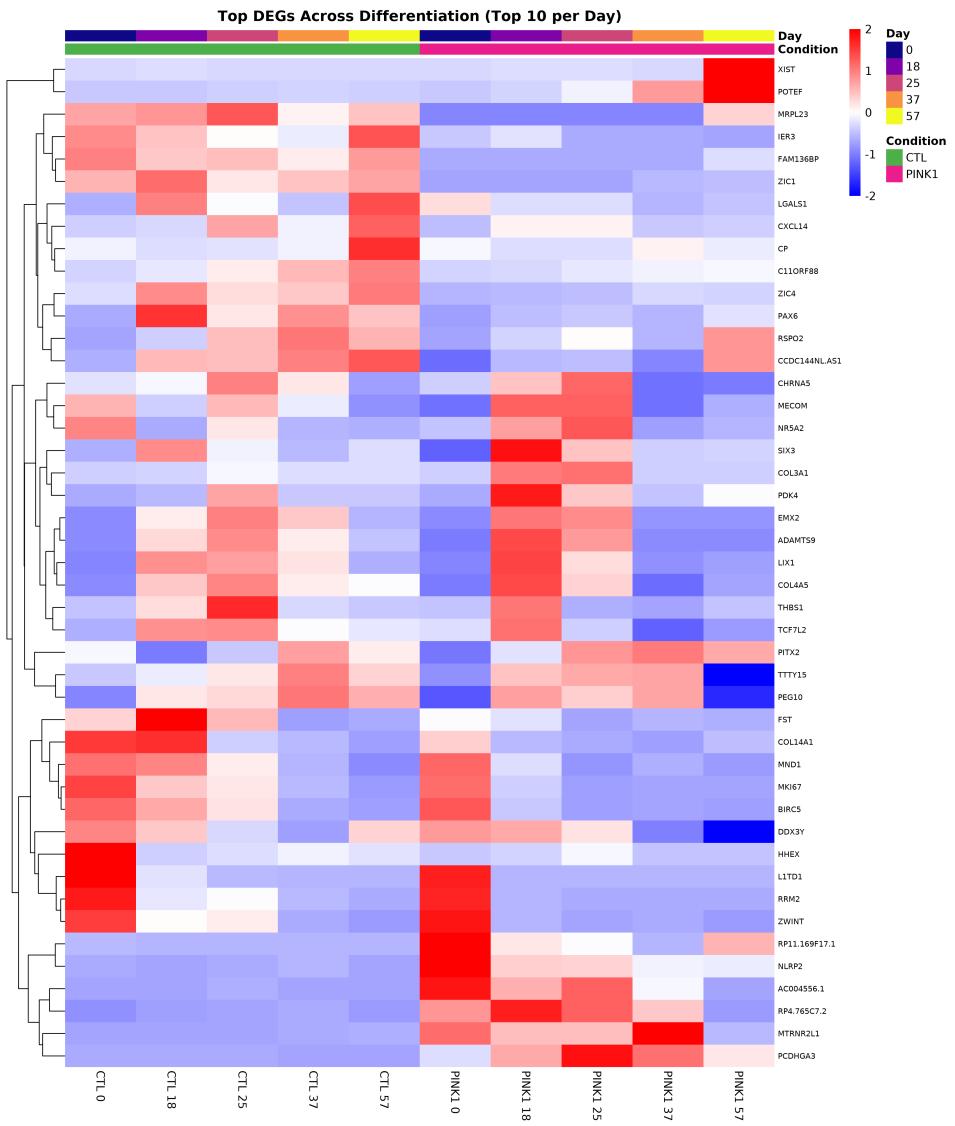


Fig. 6: Heatmap of top DEGs (50 genes) across days and genotypes, Z-score normalized and clustered.

MTRNR2L1, MIR7.3HG), suggesting post-transcriptional regulatory adaptations. Critically, MTRNR2L1 encodes humanin-like peptides with known neuroprotective and mitochondrial protective functions, potentially representing a direct compensatory mechanism. The presence of NR4A2 (NURR1) and PCDHGA3 among top PINK1 genes reinforces the enhanced dopaminergic differentiation phenotype. PDK4 upregulation confirms metabolic rewiring toward glycolysis, a classical response to mitochondrial dysfunction.

J. Clustering and Cluster Marker Structure

Unsupervised clustering (Fig. 10a) identified 14 transcriptionally distinct cell populations, demonstrating substantial cellular heterogeneity throughout differentiation. The spatial organization of clusters along the UMAP axes reflects developmental progression, with early clusters occupying one region and late clusters occupying another. The presence of

multiple clusters within broad developmental stages (e.g., several progenitor clusters) could reflect: (1) asynchronous differentiation where cells progress at different rates; (2) cell cycle-dependent transcriptional variation; (3) stochastic gene expression creating transient states; or (4) genuine sublineage specification (e.g., distinct subtypes of dopaminergic neurons). Smaller clusters (8, 11, 12) may represent rare transitional states or specialized populations. Notably, both CTL and PINK1 cells are represented across all clusters, indicating that PINK1 mutation does not eliminate specific cell types but rather may alter the proportions or transcriptional states within clusters.

The cluster marker heatmap (Fig. 10b) provides molecular definitions for each cell population. The stark enrichment of histone genes in specific clusters unambiguously identifies S-phase cells undergoing active DNA replication, likely representing the proliferative progenitor pool. Neuronal clus-

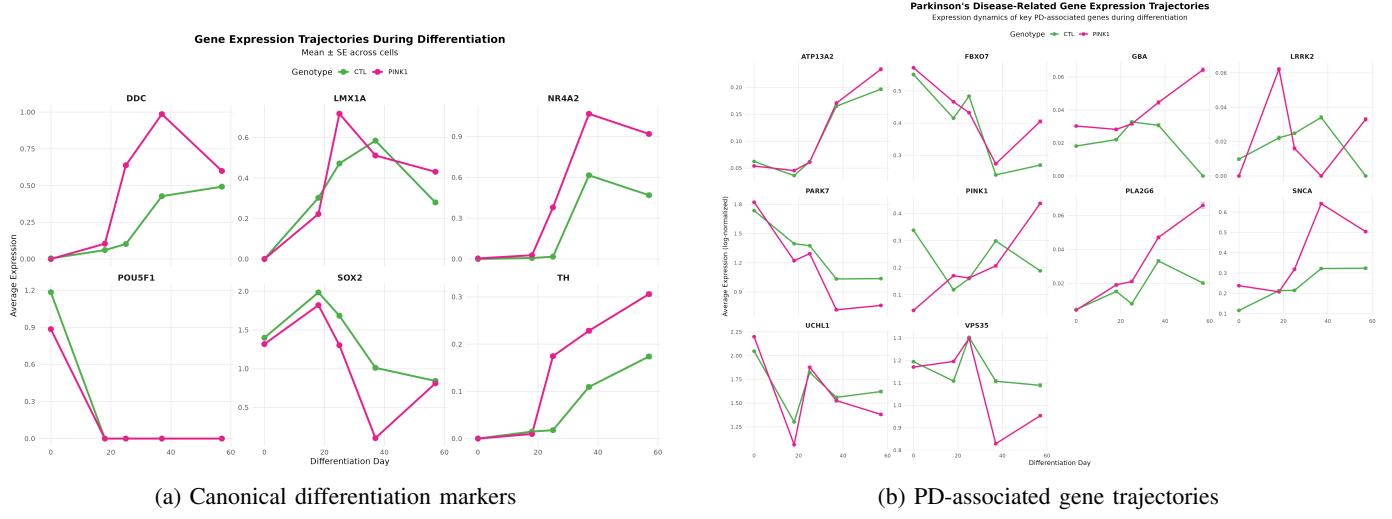


Fig. 7: Gene expression trajectories for developmental and PD-associated genes.

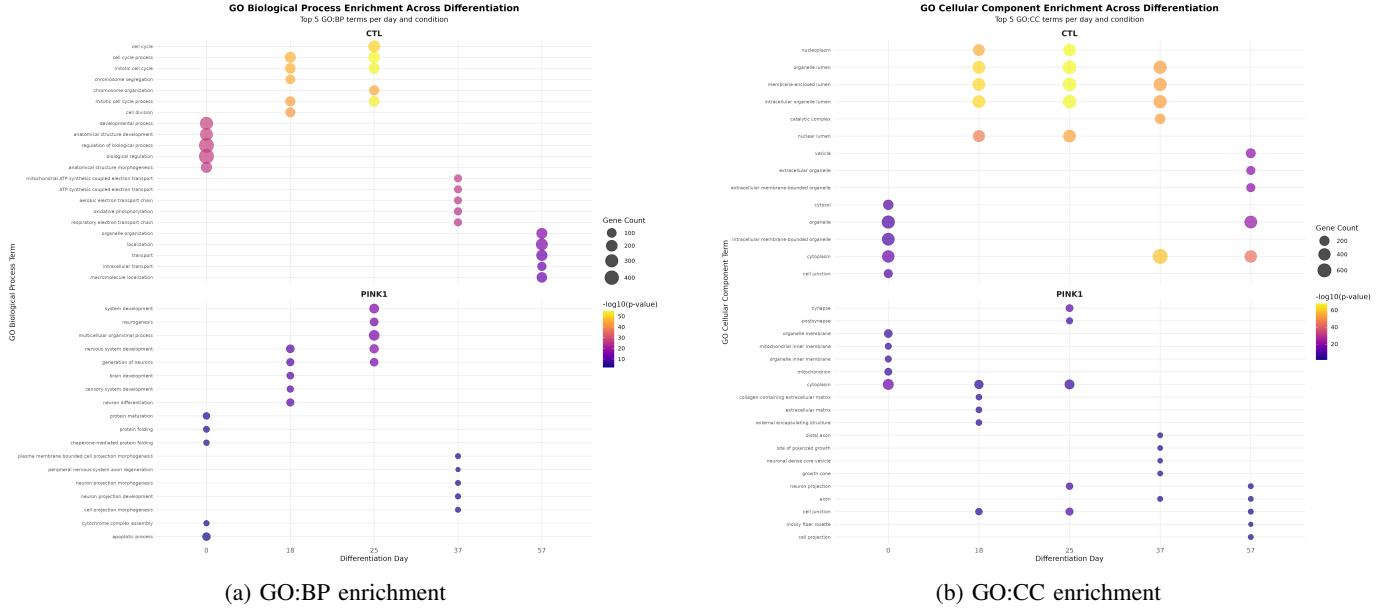


Fig. 8: GO Biological Process enrichment and GO Cellular Component enrichment

ters show canonical markers (MAP1B, CALM2, MALAT1) confirming neuronal identity. The presence of cluster-specific mitochondrial gene expression is particularly relevant for PINK1 studies, as it suggests that different cell populations have different metabolic profiles and may respond differently to mitochondrial dysfunction. Clusters with specialized markers (SULF1, CHGA, TFF3) potentially represent functionally distinct neuronal subtypes.

The dot plot (Fig. 10c) adds quantitative rigor to cluster characterization. The near-universal expression of housekeeping genes like MALAT1 and mitochondrial RNAs validates data quality while the sharp restriction of proliferation markers confirms cluster identities. Critically, the variable expression levels (color intensity) of ubiquitously present genes suggests

TABLE I: STRING Network Statistics Comparison

Parameter	CTL	PINK1
Nodes	95	92
Edges	176	34
Avg. degree	3.71	0.739
Clustering coef.	0.412	0.21
Expected edges	16	3
PPI p-value	$< 10^{-16}$	$< 10^{-16}$
Fold enrich.	11.0×	11.3×

underlying biological heterogeneity even among cells expressing the same marker.

K. STRING Network Analysis Reveals Contrasting Functional Modules

Top 15 Consistently Up/Down Regulated Genes

Genes showing strongest average fold change across all time points ($\text{padj} < 0.01$)

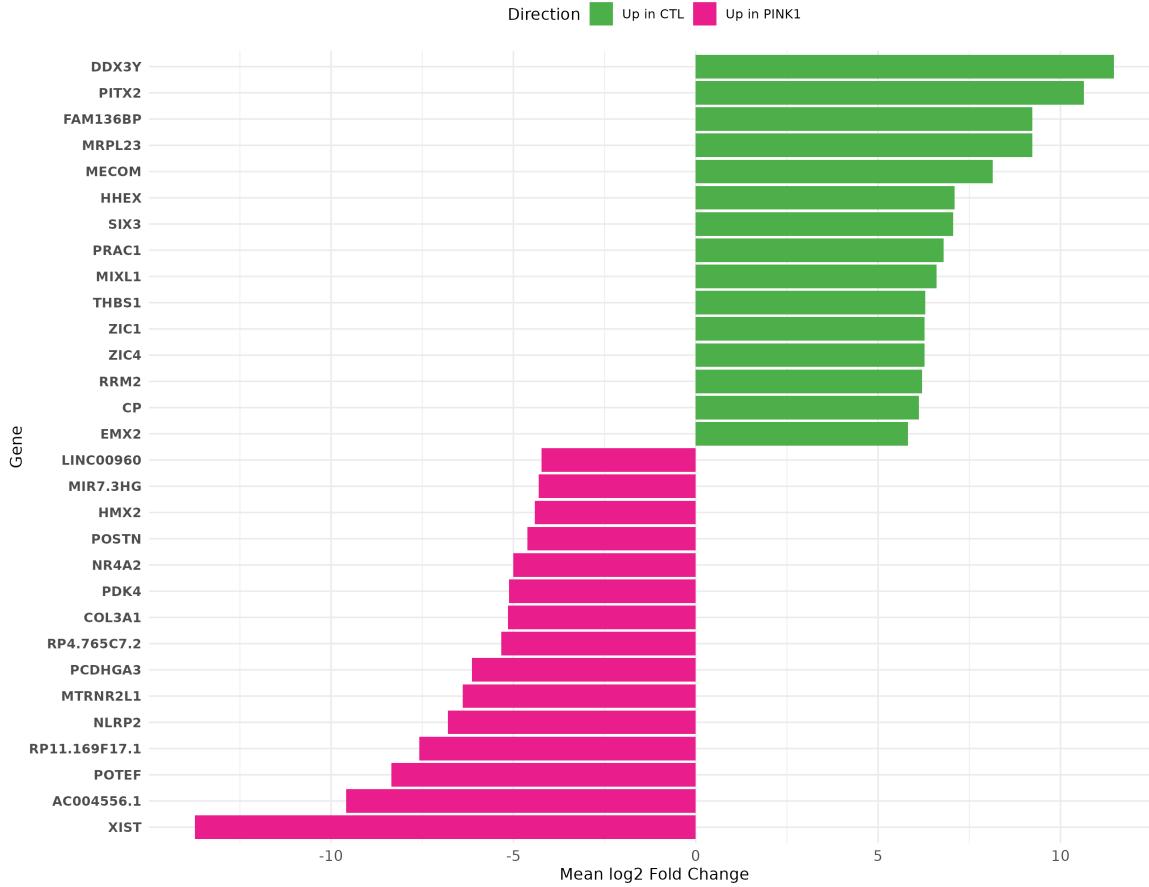
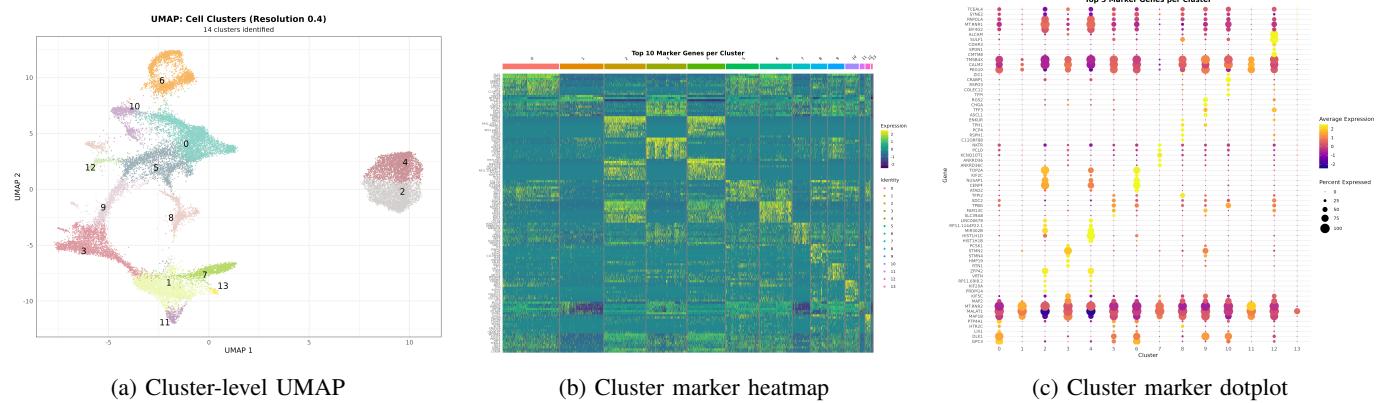


Fig. 9: Top 15 consistently regulated genes reveal persistent transcriptional biases. Genes ranked by mean absolute fold change across all time points (adjusted $p < 0.01$).



(a) Cluster-level UMAP

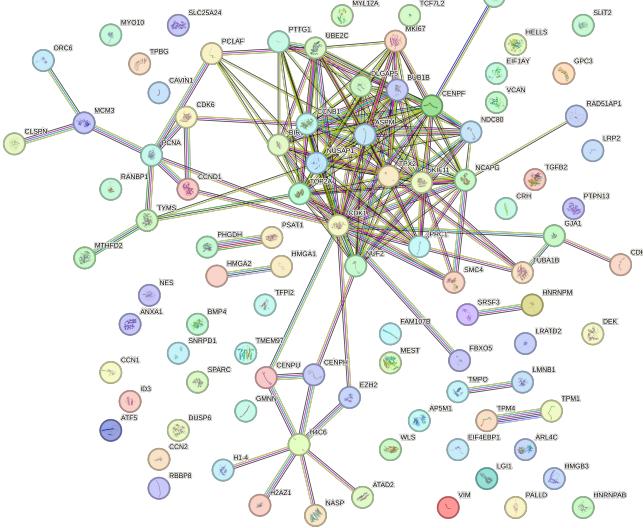
(b) Cluster marker heatmap

(c) Cluster marker dotplot

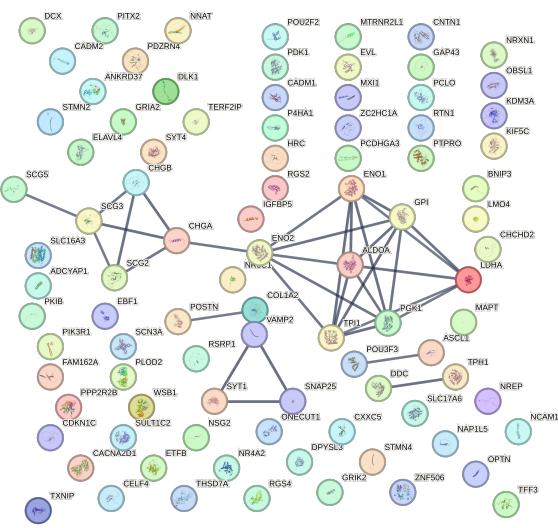
Fig. 10: Unsupervised clustering and marker structure reveal 14 transcriptionally distinct cellular states.

1) CTL Network: Highly Interconnected Cell Cycle Module: The CTL STRING network (Fig. 11a, Table I) reveals a highly integrated cell cycle machinery. With 176 edges connecting 95 proteins (average degree = 3.71), the network is 11-fold more connected than expected by chance ($p < 10^{-16}$). This extraordinary connectivity indicates functional coordina-

tion: these proteins don't work independently but form a synchronized proliferation program. The central hub containing MKI67, CCNB1, CDK1, TOP2A, and BIRC5 represents the core mitotic machinery. Peripheral modules provide metabolic support, chromatin remodeling, and structural components necessary for rapid cell division.



(a) CTL STRING network



(b) PINK1 STRING network

Fig. 11: STRING PPI networks for CTL- and PINK1-upregulated genes at Day 25.

2) PINK1 Network: Sparse Glycolytic and Neuronal Module: The PINK1 STRING network (Fig. 11b) presents a dramatically different architecture despite equivalent statistical enrichment significance. With only 34 edges connecting 92 proteins (average degree = 0.739), the network is 5-fold less connected than CTL, indicating sparse, modular organization rather than integrated coordination. The central glycolytic module-containing nearly complete representation of the glycolytic pathway (ENO1/2, LDHA, ALDOA, GPI, TPI1, PGK1) demonstrates a coherent metabolic adaptation to mitochondrial dysfunction. This "Warburg-like" shift from oxidative phosphorylation to glycolysis is a classical compensatory response when mitochondria are impaired. The chaperonin/secretogranin cluster indicates enhanced vesicular compartmentalization. Critically, many neuronal differentiation genes remain unconnected, suggesting they are independently upregulated rather than coordinated through protein-protein interactions.

3) STRING Functional Enrichment Analysis: STRING functional enrichment analysis (Fig. 12) provides systems-level interpretation of the networks. The CTL enrichment graph is visually dominated by cell cycle terms, with hierarchical relationships clearly visible. The extreme enrichment significance ($FDR \sim 10^{-20}$) indicates that nearly every gene in the CTL network participates in proliferation. In striking contrast, the PINK1 enrichment graph shows two distinct modules: (1) **Metabolic module** centered on glycolysis, representing the compensatory metabolic shift; (2) **Neuronal module** centered on neurotransmitter secretion and synaptic function, representing terminal differentiation. The moderate enrichment significance ($FDR \sim 10^{-6}$) and dual module structure reflect the sparse network architecture - these are separate adaptive responses rather than integrated programs.

L. Synthesis: PINK1 Mutation Drives Precocious Differentiation with Metabolic Compensation

Integration of all analyses reveals a coherent model for PINK1 mutation effects on dopaminergic differentiation:

- 1) **Baseline mitochondrial dysfunction** (Day 0): PINK1 cells show mitochondrial GO enrichment and 1,063 DEGs even in pluripotent state, indicating pre-existing organelar stress.
- 2) **Altered developmental pacing** (Days 18-25): While CTL cells undergo robust progenitor expansion (cell cycle enrichment, 1,123 DEGs up), PINK1 cells show precocious neuronal differentiation (neurogenesis enrichment, earlier LMX1A peak). This represents a fundamental developmental shift from proliferation-prioritized to differentiation-prioritized strategy.
- 3) **Metabolic rewiring:** PINK1 cells engage glycolytic pathway upregulation (LDHA, ENO1/2, PDK4) as a compensatory response to mitochondrial dysfunction, representing a "Warburg-like" metabolic shift.
- 4) **Compensatory neuronal identity enhancement** (Day 57): Despite earlier challenges, PINK1 neurons show dramatically elevated expression of dopaminergic specification genes (TH, NR4A2, DDC), potentially representing feedback compensation where enhanced transcriptional identity counterbalances metabolic limitations.
- 5) **PD-relevant alterations:** Elevation of SNCA (α -synuclein) in PINK1 neurons provides a direct mechanistic link to sporadic PD pathogenesis, while PARK7 (DJ-1) downregulation suggests impaired oxidative stress responses.
- 6) **Network architecture differences:** The stark contrast between CTL's highly integrated cell cycle network and PINK1's sparse, modular network reflects fundamentally

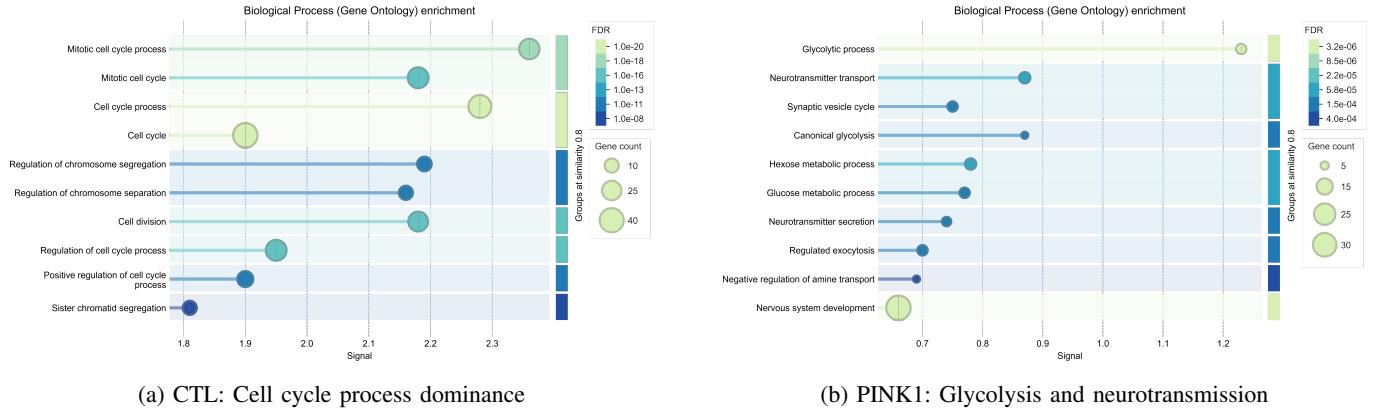


Fig. 12: STRING GO enrichment networks.

different cellular strategies coordinated proliferation versus independent adaptive responses.

- 7) **Preserved cellular diversity:** Despite profound transcriptional changes, PINK1 cells successfully generate all 14 identified cell clusters and achieve complex neuronal morphology (synaptic compartment enrichment), demonstrating remarkable developmental robustness.

This integrated model positions PINK1 loss as a driver of developmental heterochrony altered timing of developmental processes rather than developmental failure. The precocious differentiation may represent an adaptive response where cells “escape” the proliferative phase (which is metabolically demanding) to enter differentiation (which may be more tolerant of mitochondrial dysfunction). However, this adaptation comes at the cost of altered transcriptional programs, metabolic dependency on glycolysis, and potential vulnerability to proteostatic stress (α -synuclein elevation). These alterations provide molecular mechanisms linking PINK1 dysfunction to PD risk and suggest therapeutic strategies targeting metabolic compensation or proteostasis.

IV. DISCUSSION

This time-resolved scRNA-seq study shows that *PINK1* loss does not prevent differentiation of human iPSCs toward dopaminergic neurons, but instead reshapes the *timing* and *coordination* of transcriptional programs across development. Across genotypes, the global differentiation manifold is preserved (Fig. 2), and both conditions reach a mature neuronal state by Day 57. The largest divergence occurs during the progenitor-to-neuronal transition window (Days 18–37), where transcriptional differences peak (Fig. 4) and functional programs split between proliferation-dominant control (CTL) and differentiation/metabolic-adaptation dominant PINK1.

A. *PINK1* loss, mitochondrial stress, and compensatory metabolism

A key observation is that genotype-dependent differences are already present at Day 0 (Fig. 4), consistent with a baseline shift in cellular state before lineage commitment. PINK1-upregulated signatures repeatedly include metabolic regulators

and glycolytic enzymes (e.g., *LDHA*, *ENO1/ENO2*, *PDK4*; Fig. 5, Fig. 9), and the Day 25 PINK1 STRING network is organized around a coherent glycolytic core (Fig. 11). Together with GO patterns (Fig. 8), these results support a compensatory metabolic rewiring away from mitochondrial pyruvate oxidation and toward glycolysis, consistent with adaptation to impaired mitochondrial quality control.

B. Developmental pacing: proliferation versus precocious differentiation

During Days 18–25, CTL cells show strong enrichment of cell-cycle/proliferation programs (Fig. 8) and a densely connected Day 25 STRING network centered on canonical mitotic hubs (Fig. 11, Table I). In contrast, PINK1 cells at the same stages preferentially enrich neurogenesis and neuronal differentiation programs (Fig. 8) and show earlier specification kinetics for key markers (Fig. 7). This suggests that PINK1 mutation shifts the balance from progenitor expansion toward earlier commitment. A plausible interpretation is that proliferative states impose high energetic and biosynthetic requirements; if mitochondrial quality control is compromised, cells may minimize time spent in metabolically demanding proliferative programs and instead transition toward post-mitotic differentiation programs while sustaining energy production through glycolysis.

Importantly, total cell yields remain comparable between genotypes across time points (Fig. 3), arguing against catastrophic genotype-specific toxicity during the protocol. Differences are therefore more consistent with altered transcriptional priorities and cell-state pacing than with widespread cell death.

C. Transcriptional asymmetry and network architecture

A striking feature is the strong asymmetry in DEG directionality: CTL shows substantially more upregulated genes than PINK1 across multiple time points, while PINK1-upregulated genes become sparse by late stages (Fig. 4, Fig. 5). This is consistent with a global attenuation of transcriptional induction in PINK1 (or incomplete activation of specific developmental and proliferative programs), rather than broad aberrant activation. The network-level contrast supports this: CTL Day

25 DEGs form a highly integrated interaction network (176 edges), whereas the PINK1 Day 25 network is sparse and modular (34 edges) despite significant enrichment (Table I). This suggests that CTL engages a coordinated proliferative program, while PINK1 engages partially independent adaptive modules (metabolic rewiring and neuronal/vesicular programs) with weaker system-wide coordination.

D. Dopaminergic identity and late-stage vulnerability

Despite early divergence, PINK1 cultures reach a mature neuronal state and show strong dopaminergic marker expression at Day 57 (Fig. 7), indicating preserved terminal differentiation. However, dopaminergic neurons are particularly sensitive to mitochondrial dysfunction because synaptic transmission and electrophysiological signaling require sustained ATP-dependent ion gradients (e.g., Na^+/K^+ -ATPase activity and Ca^{2+} buffering), and oxidative stress management is critical in post-mitotic neurons. In this context, persistent glycolytic compensation may be insufficient under stress, potentially increasing vulnerability even when marker-based differentiation appears successful.

Disease-relevant trajectories support this vulnerability model: *SNCA* is elevated in PINK1 at late stages (Fig. 7), and additional PD-linked genes show condition-dependent dynamics, consistent with altered proteostasis, vesicle trafficking, and stress-response pathways. These changes provide a mechanistic link between PINK1 dysfunction and PD-relevant molecular phenotypes that could precede overt degeneration.

E. Heterogeneity and compositional effects

Clustering identifies multiple transcriptional states across differentiation (Fig. 10), and the overall developmental architecture is preserved between genotypes. Appendix analyses further show that major cluster classes are present across time points and conditions (Appendix Fig. 15), supporting the interpretation that PINK1 does not eliminate broad lineages but may alter the proportion of cells occupying specific states and/or the transcriptional program within those states. Disentangling composition effects from within-state regulation is therefore important for causal interpretation.

F. Limitations

- Replication and donor/batch structure:** The analysis compares available CTL and PINK1 samples across time points. Without explicit biological replication and donor-matched designs, some differences could reflect line or batch effects.
- Compositional confounding:** Day-wise comparisons (CTL vs PINK1 at the same nominal day) can be influenced by genotype-dependent differences in differentiation pacing, causing shifts in cell-state composition that may contribute to observed DEGs.
- Imputation effects:** Gene trajectories were computed using an imputed assay to stabilize trends; imputation can increase correlation structure and smooth expression dynamics, affecting peak timing and magnitude.

- DE testing/threshold harmonization:** The draft describes Wilcoxon-based DEGs with a fold-change threshold; if adjusted *p*-value thresholds were applied in practice, these should be stated consistently in Methods and used uniformly in Results/Discussion.
- Functional validation:** scRNA-seq provides transcriptional programs but does not directly measure mitochondrial respiration, ROS, mitophagy flux, electrophysiology, dopamine handling, or α -synuclein aggregation; thus functional and disease-mechanistic conclusions remain inferential.

G. Future directions

- Trajectory-aware inference:** Perform pseudotime and branch-aware analyses to compare genotypes along matched developmental progression and separate timing shifts from within-state regulation.
- Cluster-matched testing:** Compare CTL vs PINK1 within matched clusters and test for differences in cluster proportions across time points to quantify compositional shifts.
- Pathway/module scoring:** Score oxidative phosphorylation, glycolysis, ROS response, mitophagy/autophagy, and unfolded protein response programs at single-cell resolution to connect DEGs to mechanistic modules.
- Stress and rescue experiments:** Validate predicted vulnerability by applying mitochondrial or proteostasis stressors to Day 57 neurons and testing whether genetic/pharmacologic restoration of PINK1-related pathways mitigates phenotypes.

V. CONCLUSION

Across five differentiation stages, *PINK1* mutation preserves the overall ability of human iPSCs to generate mature dopaminergic-like neurons but alters the temporal coordination of transcriptional programs. The greatest divergence occurs during progenitor and early neuronal stages (Days 18–37), where CTL exhibits a highly coordinated proliferative program while PINK1 shows metabolic rewiring toward glycolysis and earlier engagement of neuronal differentiation pathways. By Day 57, PINK1 cultures display strong dopaminergic marker expression but also PD-relevant molecular changes, including elevated *SNCA* and altered stress-response trajectories. These findings support a model in which PINK1 loss drives developmental heterochrony and compensatory metabolism that preserves differentiation outcomes while potentially increasing late-stage energetic and proteostatic vulnerability.

ACKNOWLEDGMENTS

We thank the course instructor for their guidance throughout this project. Computational resources were provided by the institutional high-performance computing cluster.

AUTHOR CONTRIBUTIONS

- Barak Landsman** - Introduction
Costin-Andrei Tăulescu - Methodology
Sowmya Janmahanti - Results

Vani Pant - Discussion & Conclusion

Behrouz Delfinian - Appendix

AI STATEMENT

This manuscript was prepared with assistance from AI tools (e.g., Claude) for refining biological background explanations, suggesting interpretations of gene expression patterns, and editing text. All data processing, plotting, and final analytical decisions and conclusions were performed and approved by the authors.

APPENDIX

A. UMAP visualization of the integrated dataset

Unsupervised clustering (Louvain, resolution = 0.8) identified 18 distinct cell populations on the integrated UMAP. The embedding shows a predominant continuous differentiation trajectory, with early pluripotent states located left, progenitor/transitional populations in the center, and mature neuronal clusters occupying the right region.

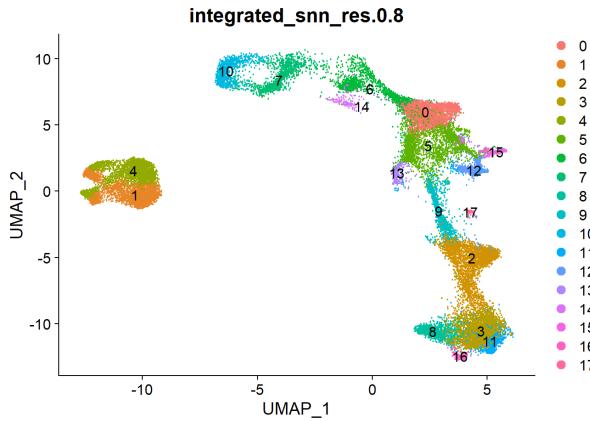


Fig. 13: Integrated UMAP colored by unsupervised clustering (Louvain, resolution = 0.8) showing 18 distinct cell populations.

B. Clustering stability across resolutions

Clustering stability was assessed by running Louvain clustering at multiple resolutions (0.2 to 1.0) on the integrated dataset. The dendrogram-like tree visualizes how clusters merge as resolution increases. Node size reflects the number of cells, edge thickness indicates the proportion of cells shared between parent and child clusters, and colors represent different resolution levels. This plot confirms robust separation of major developmental branches and shows the final 18-cluster solution (resolution 0.8) as a stable intermediate configuration.

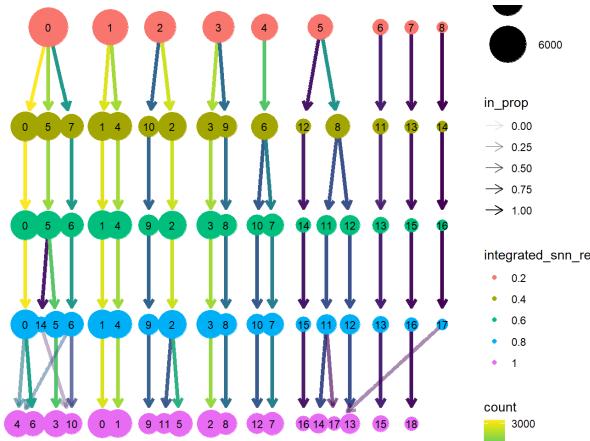


Fig. 14: Clustering tree showing stability and merging patterns across increasing Louvain resolutions (0.2–1.0).

C. Cell-type composition across time points and genotypes

Stacked bar plots show the proportional composition of the 18 identified cell clusters (cell types) across the five differentiation time points (Days 0, 18, 25, 37, 57) for both control (CTL) and PINK1 mutant conditions. Cluster proportions shift dramatically during differentiation, with early clusters dominating at Day 0, progenitor/transitional clusters peaking at intermediate days, and mature neuronal clusters becoming predominant at Day 57. The overall pattern is largely preserved between genotypes, with only subtle differences in timing and relative abundance.

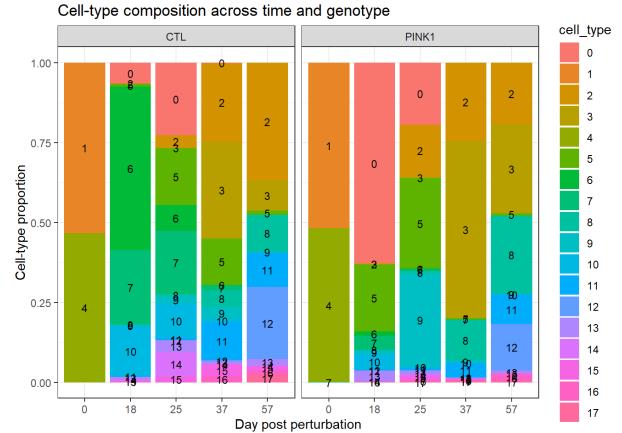


Fig. 15: Stacked bar plot of cell-type (cluster) proportions across differentiation time points and genotypes (CTL left, PINK1 right). Colors correspond to clusters 0–17 as shown in the legend.

D. Cluster distribution per genotype

UMAP projections at Day 18 (neural progenitor stage) highlight cluster distributions for CTL (left) and PINK1 (right). Both conditions show a broad progenitor-dominated landscape, with CTL exhibiting more spread across early clusters and PINK1 showing slightly more compact grouping in progenitor/mid-stage clusters.

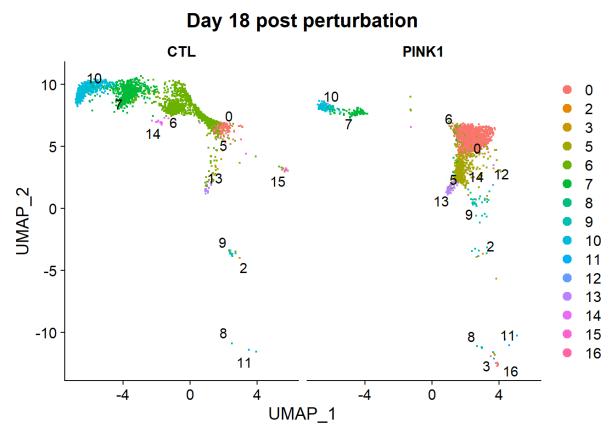


Fig. 16: Split UMAP visualization at Day 18 showing cluster distribution in CTL and PINK1 conditions.

UMAP projections at Day 37 post-differentiation show the distribution of the 18 clusters separately for control (CTL, left) and PINK1 mutant (right) conditions. At this early neuronal stage, both genotypes display a mix of progenitor and maturing neuronal clusters, with subtle shifts in density and cluster prominence between conditions (e.g., increased representation of certain late neuronal clusters in PINK1).

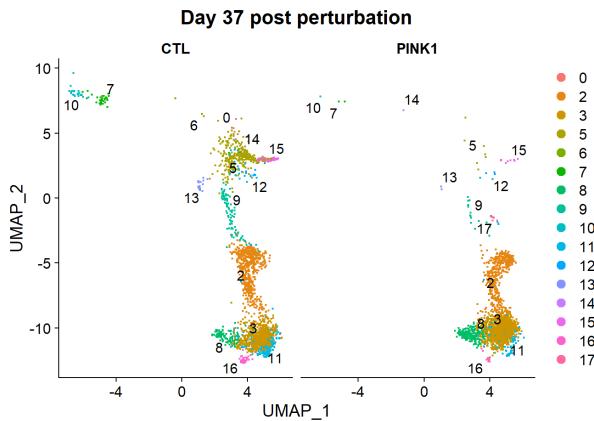


Fig. 17: Split UMAP visualization at Day 37 showing cluster distribution in CTL and PINK1 conditions.

E. Principal component analysis (PCA)

PCA was performed on the integrated dataset to identify major sources of variation. PC1 captures the primary differentiation trajectory from pluripotent/progenitor states (negative values) to post-mitotic neuronal states (positive values). Violin plots illustrate the distribution of cell embeddings along these components, while heatmaps display top contributing genes (loadings) for PCs 1–9. PC1-positive genes are enriched in late neuronal markers and activated at later time points, while differences along PC1 between CTL and PINK1 reflect altered differentiation kinetics rather than distinct cell identities.

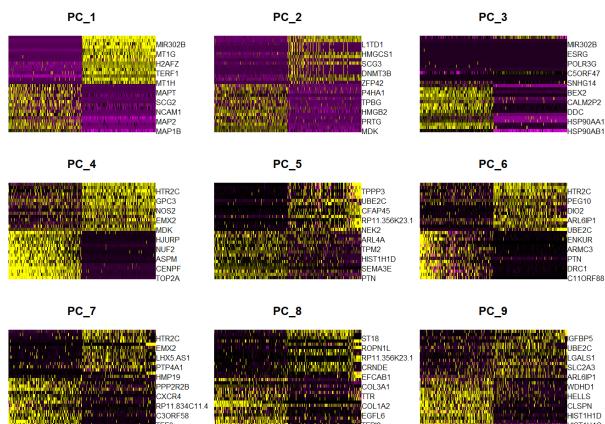


Fig. 18: Heatmap of top gene loadings across the first 9 principal components. Rows are genes, columns are PCs; color intensity reflects loading strength (yellow = positive, purple = negative).

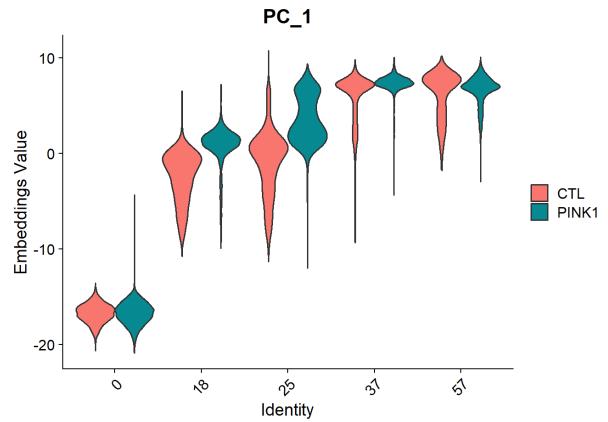


Fig. 19: Violin plot of cell embeddings along PC1 (differentiation trajectory) for CTL and PINK1.

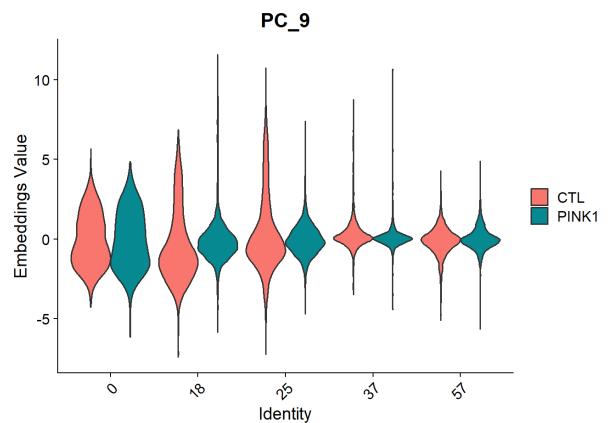


Fig. 20: Violin plot of cell embeddings along PC9 (genotype separation) for CTL and PINK1.

F. Integrated UMAP colored by condition and differentiation day

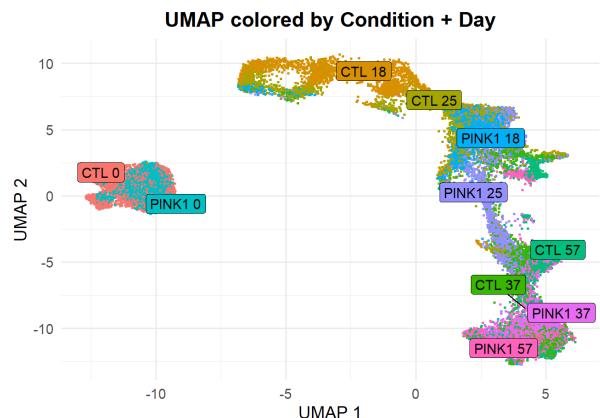


Fig. 21: Integrated UMAP colored by genotype (CTL/PINK1) and day post-differentiation (0, 18, 25, 37, 57). The trajectory progresses from early pluripotent states to mature neurons.

The integrated UMAP projection shows the overall structure of the dataset, with cells colored by a combination of genotype (CTL vs PINK1) and differentiation time point (Day 0, 18, 25, 37, 57). Both conditions follow a largely continuous developmental trajectory from pluripotent cells (left, Day 0) through progenitor stages (center) to mature neuronal populations (right, Day 57). Subtle shifts in density and progression timing are visible between CTL and PINK1, particularly at intermediate stages.

G. Dot plot of key differentiation markers across time points (PINK1 trajectory)

Dot plot showing the expression dynamics of selected dopaminergic differentiation and progenitor markers along the PINK1 knockout trajectory across the five time points (Days 0, 18, 25, 37, 57). Circle size represents the percentage of cells expressing each gene, while color intensity indicates average expression level (scaled). Early pluripotency markers (e.g., POU5F1, TDGF1) decrease rapidly, while mid/late dopaminergic specification genes (e.g., NR4A2, DDC, LMX1A) increase progressively, peaking at Days 37–57.

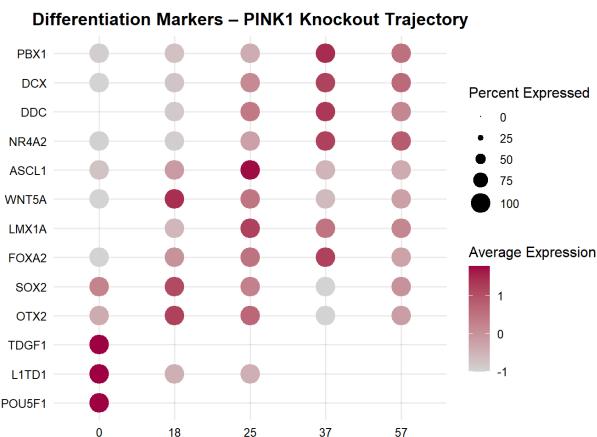


Fig. 22: Dot plot of key differentiation markers in PINK1 mutant cells across time points. Size = percent expressed, color = average scaled expression.

H. Enrichment analysis – significant DEGs upregulated in PINK1

Bubble/Manhattan-style plot from g:Profiler2 showing functional enrichment results for genes significantly upregulated in PINK1 vs CTL ($p < 0.05$). Analysis performed as over-representation analysis (ORA) on significant DEGs only (ordered_query = FALSE). X-axis groups ontology/pathway sources (GO:MF, GO:CC, GO:BP, KEGG, Reactome, WikiPathways). Bubble size reflects gene count, color intensity shows significance ($-\log_{10}(\text{adj. } p\text{-value})$), values > 16 capped at threshold. Strongest enrichment appears in GO:BP (metabolic and neuronal processes) and GO:CC (mitochondrial and synaptic terms).

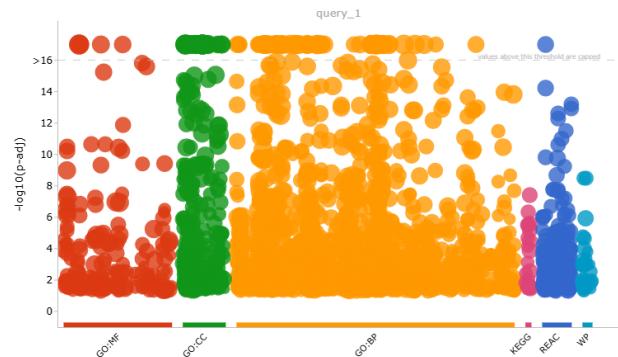


Fig. 23: g:Profiler enrichment plot for significant genes upregulated in PINK1 (query_1). Size = number of genes, color = $-\log_{10}(\text{adjusted } p\text{-value})$, values above 16 capped.

I. Top 20 enriched terms – significant DEGs

Bar plot showing the top 20 most significant enriched terms from g:Profiler2 over-representation analysis (ORA) on significant DEGs. Terms are ordered by adjusted p-value. X-axis shows $-\log_{10}(\text{adjusted } p\text{-value})$, bar fill indicates significance, and red circle size represents the number of intersecting genes (intersection_size). The strongest enrichments are in broad cellular component terms (cytoplasm, organelles, membranes) and general biological processes (regulation of metabolic/cellular processes).

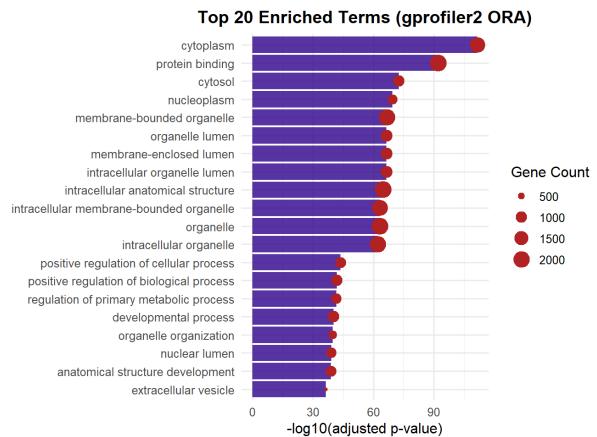


Fig. 24: Top 20 enriched terms from g:Profiler2 ORA on significant DEGs. Bars = $-\log_{10}(\text{adjusted } p\text{-value})$, red circles = gene count per term.

J. GSEA dotplot – GO Biological Process (top 20 enriched terms)

Dotplot from clusterProfiler GSEA analysis (gseGO) on ranked gene list (avg_log2FC, descending). X-axis shows GeneRatio (proportion of genes in the term among ranked list), dot size represents gene count, color indicates adjusted p-value. All top terms are strongly related to mitotic cell cycle, chromosome segregation, and phase transitions, confirming major enrichment of proliferation-related processes (likely driven by CTL-upregulated genes).

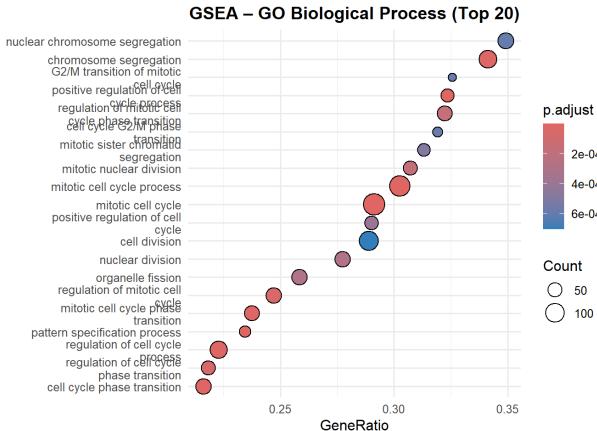


Fig. 25: GSEA results for GO Biological Process (top 20 terms). Dot size = gene count, color = adjusted p-value, x-axis = GeneRatio.

K. GSEA dotplot – Reactome Pathways (top enriched)

Dotplot from clusterProfiler GSEA (gsePathway) on ranked gene list (converted to ENTREZ IDs) for Reactome pathways. X-axis shows GeneRatio, dot size represents gene count, color indicates adjusted p-value (viridis scale). Top terms are strongly dominated by cell cycle/mitosis-related pathways (e.g., chromosome condensation, DNA replication, G2/M checkpoints, mitotic phases), reflecting major enrichment in proliferation and DNA maintenance processes.

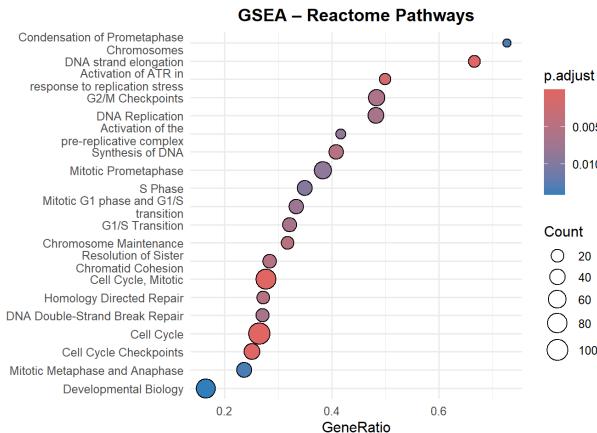


Fig. 26: GSEA results for Reactome pathways (top terms). Dot size = gene count, color = adjusted p-value, x-axis = GeneRatio.

L. GSEA dotplot – GO Cellular Component (top enriched terms for PINK1-upregulated genes)

Dotplot from clusterProfiler GSEA (gseGO, ont="CC") on ranked gene list of PINK1-upregulated genes (positive log2FC). X-axis shows GeneRatio, dot size represents gene count, color indicates adjusted p-value. Top terms highlight mitotic/chromosome-related cellular components (kinetochore, condensed chromosome, centromeric region) alongside some

neuronal features (neuron projection, cell body), suggesting coordinated enrichment in both proliferative and neuronal compartments.

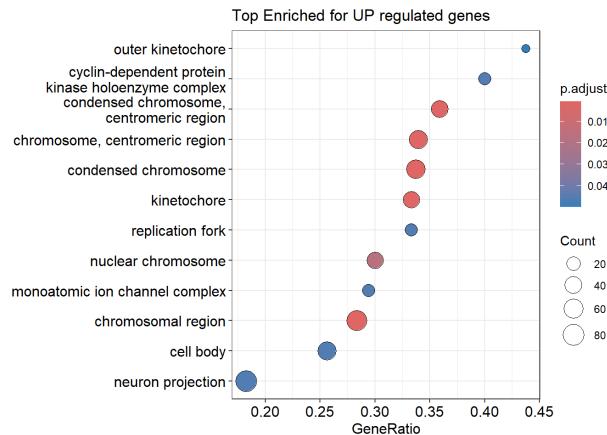


Fig. 27: GSEA results for GO Cellular Component on PINK1-upregulated genes (top terms). Dot size = gene count, color = adjusted p-value, x-axis = GeneRatio.