TOGGLE identifies fate and function process within a single type of mature cell through single-cell transcriptomics.

Junpeng Chen[1,2,17+], Zhouweiyu Chen[4+], Eric Jiang[3+], Kaiqing Liu[1,2+], Tianda Sun[5], Yibing Nong[1], Tao Yuan[9], Pawan Daga[10], Charles C. Dai[11,12], Yexing Yan[14], Jinwen Ge[7,8], Haihui Wu[15], Tong Yang[7], Shanshan Wang[7], Zixiang Su[16], Tian Song[18], Changping Li[13]*, You Li[3]*, Kailin Yang[7,8]*, Richa A. Singhal[17]*, Lu Cai[2*], Alex P. Carll[1,17]*

1 Address: Department of Physiology, University of Louisville School of Medicine, Louisville, Kentucky,40204, USA

2 Address: Pediatric Research Institute, Department of Pediatrics, University of Louisville School of Medicine, Louisville, Kentucky 40202, USA

3 Address: Kebei Hospital of Shaoxing, Shaoxing City 312081, China

4 Address: Xiangtan University, Xiangtan 411105, China

5 Address: Department of Computer Science, University of York, Heslington, York YO10 5DD, United Kingdom

7 Address: Key Laboratory of Hunan Province for Integrated Traditional Chinese and Western Medicine on Prevention and Treatment of Cardio-Cerebral Diseases, School of Integrated Chinese and Western Medicine, Hunan University of Chinese Medicine, Changsha, China

8 Address: Hunan Academy of Chinese Medicine, Changsha, China

9 Address: College of Mechanical Engineering and Automation, Huaqiao University, Xianmen, China, yuan__t@163.com

10 Address: Department of Internal Medicine, University of Louisville, Kentucky, USA

11 Address: Department of Oral and Maxillofacial Surgery, School of Dentistry, University of Maryland, Baltimore, Maryland 21201, USA.

12 Address: Fischell Department of Bioengineering, A. James Clark School of Engineering, University of Maryland, College Park, Maryland 20742, USA

13 Address: School of mechanical engineering and automation, Fuyao University of Science and Technology, Fuzhou, China

14 Address: Department of Psychology, Daqing Hospital of Traditional Chinese Medicine, Daqing, China

15 Address: Pinghu Hospital of Traditional Chinese Medicine, Pinghu, Zhejiang Province, China

16 Address: University of Minnesota, College of Biological Sciences, St. Paul, MN 55108, USA, suzixiang.mitchell@gmail.com
17 Address: Center for Cardiometabolic Science, Division of Environmental Medicine, Christina Lee Brown Envirome Insttitute, University of Louisville, Louisville, KY 40202, United States.
18 Address: Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing, China

+ Junpeng Chen (drchenjunpeng@hotmail.com), Eric Jiang (jyjiang@link.cuhk.edu.hk), Zhouweiyu Chen (yahoo_con_cn@163.com) and Kaiqing Liu are co-first authors who have contributed an indispensable part to this article and should be considered to have made equal contributions. Their names are listed without any specific order.

*Corresponding author should be: Changping Li (15073200235@163.com), You Li (liyou@cjgs23.wecom.work), Kailin Yang (yklab2014@hotmail.com) , Richa A. Singhal (richa.singhal@louisville.edu), Lu Cai (lu.cai@louisville.edu) and Alex P. Carll (Alex.Carll@louisville.edu). Alex P. Carll is the project and Lab leader.

Abstract

Almost inevitably, mature cells meet their demise through specific processes of phenotypic change such as apoptosis. Although single-cell RNA sequencing (scRNA-seq) technology can reconstruct processes like life cycles and disease progression, identifying specific types of cell death in dying cells remains challenging due to the complexity of related signaling pathways and marker genes. To address this, our study introduces a fate prediction method, TOGGLE, which reconstructs complete programmed cell fates from untracked data. Based on a corollary of Takens' theorem, TOGGLE enables analysis of highly similar programmed cell death data with indistinct fate boundaries. This approach circumvents the need for live sample tracking and serial assessments. In hematopoiesis and reprogramming, TOGGLE reconstructs full fate biases and predicts cell differentiation from early development to maturity. In brain tissue from a murine model of cerebral infarction, TOGGLE predicts the fate trajectories of neurons undergoing post-infarction programmed cell death (especially ferroptosis) and identifies genes expressed differentially between dying and healthy neurons. Subsequently, functional transitions and differences were discussed using cells from the heart.

1 Introduction

The phenotype of terminally differentiated cells inevitably morphs toward one of many pathways of cell death. Whole tissues may bear cells following distinct molecular patterns that dictate separate fates; for example, upon cerebral infarction, neurons in ischemic stroke may succumb to to programmed cell death such as ferroptosis, apoptosis or pyroptosis, at different stages of pathogenesis and under different conditions[1–4]. Given the complexity of signaling pathways and marker genes involved in programmed cell death, identifying the specific type of cell death occurring in cells remains challenging. Ferroptosis poses a powerful example of the value for distinguishing between types of programmed cell death. As a crucial form of programmed cell death in mature cells, ferroptosis is involved not only in embryonic development but also in tumor growth, immune evasion, and muscle atrophy, and catalyzed by iron [5–8]. In ischemic stroke (cerebral infarction), a cerebrovascular disease with high rates of mortality and disability, ferroptosis significantly contributes to neuronal death [9,10]. Regulating ferroptosis is essential for rescuing dying neurons and restoring the function of at-risk neurons in affected brain areas[9,10]. Despite the importance of ferroptosis, its mechanisms remain poorly understood[7]. With advancements in technology, single-cell RNA sequencing (scRNA-seq) has become a valuable tool for studying ferroptosis in brain cells during ischemic stroke. However, due to limitations in analytical methods, most research has focused on neuroglia cells[11,12], with only a few using scRNA-seq to study neurons[13]. In the few studies that have analyzed scRNA-seq data from neurons, existing tools can identify neuron types (e.g.,glutamatergic neurons)[13] but cannot distinguish between different neuronal states, including whether they are undergoing necrosis or their stage of programmed cell death. Characterizing the contributions of different modes of cell death across neuronal cell populations after cerebral infarction could facilitate the development of new therapies. Indeed, in virtually any model of tissue injury, a fate-prediction tool that leverages scRNA-seq data to map trajectories of programmed cell death, differentiate cells undergoing programmed death from those undergoing alternate processes, and specify type of programmed death for each cell would elucidate the types of programmed death that mediate disease progression and potentially open new avenues for therapeutic development. This is not merely a prediction process; it should enable the distinction of stages in mature cells.

While there are numerous tools available for analyzing scRNA-seq data, few are designed to predict the fate of Mature cells, such as neurons. Current computational methods include Weighted Optimal Transport (WOT)[14] and Cospar[15]. WOT constructs lineage information based on dynamic data trajectories but tends to misidentify similar transition trajectories. Cospar utilizes transition matrix principles to infer state transitions between nodes. However, it is not suitable for multi-stage classifications where correspondences are not unique; it cannot classify static, differentiated cell data, such as Mature cells, nor can it distinguish between different programmed stages within the same cell type. The primary challenges in fate prediction, illustrated in **Figure 1**, include:

(1) Endpoint Loss. When using the above tools, information on both the starting point and endpoint of cell fate is required. For instance, it is necessary to have progenitor cells like Granulocyte-Monocyte Progenitors (GMP) that can give rise to progeny clones, and to capture samples of descendant clones such as Neutrophils and Monocytes. However, in certain diseases, it is impossible to obtain samples of cells that have reached their endpoint. For example, in ischemic stroke, neurons that have completed programmed cell death cannot be captured.

(2) Time Loss. The above algorithm requires sampling at multiple time points to piece together the continuous and gradual changes in cells, thereby reconstructing the trajectory of transition states during progenitor cell growth. However, in clinical practice, it is difficult to obtain samples at planned time intervals, such as post-surgical tumor tissues, hematomas after brain hemorrhage, or samples from deceased individuals or aborted embryos. Moreover, ethical and logistical concerns often repeated biopsies on humans at different time points. There is thus a compelling need for sensitive and accurate analytical approaches to predict cell fate from samples collected at a single

time point. This requires first distinguishing cells at various stages of fate progression with high classification accuracy, followed by using these precise classifications to predict cell fate.

(3) Sample Imbalance. When collecting samples at multiple time points, it is possible to capture the starting point, endpoint, and intermediate processes of cell fate, along with dynamic information about the changes in proportion. Sample imbalance can be a significant issue not only at single timepoint but also across multiple timepoints. This problem arises due to the inherent limitations of single-cell technologies, such as uneven cell capture, sequencing depth variability, and the over- or underrepresentation of certain cell types. At a single time point, it is difficult to balance the number of cells in different stages of fate progression, leading to sparse data and uneven sample distribution. This usually ver-represent certain cell stages relative to the true distribution, because uneven cell numbers impacts both functional and type classification results. Cells with smaller populations and higher similarity are often misclassified by algorithms as the same type, this issue commonly arises in the identification of rare death types, complex immune functional subgroups, or rare subpopulations. This requires modifying the algorithm's principles, shifting from multi-time-point state transition construction to reconstructing continuous trajectory information from a single time point.

(4) Lack of Fate Boundaries. There is a distinct fate boundary between progenitor cells and Neutrophil/Monocyte cells, where transcriptional differences cause divergent fates. However, ferroptosis and other forms of programmed cell death are often unidirectional and gradual fate progressions lacking distinct boundaries. Thus, cells undergoing programmed death are often distributed across a spectrum of gene expression patterns that diverge from cells that are healthy, making the construction of the fate trajectory more challenging.

We named this method "Trajectory of Gene Generated Lineage Estimation (TOGGLE)." Based on the above issues, this study developed a new algorithm—TOGGLE. Compared with Cospar, Super OT, WOT, and GAN-based OT algorithms, TOGGLE offers the following advantages:

(1) This addresses a novel type of tracing problem. Except for immune cells, mature cells generally do not undergo cloning, making lineage tracing based on cell division unsuitable for mature cells. Mature cells exhibit highly similar temporal states, and their capture is often limited to a single stage within the programmed fate of cells.

(2) This method enables the tracing of mature cell fates or functions by constructing the continuous fate progression of mature cells (e.g., neurons) in highly similar datasets and grouping them according to their fate progression, even when clear fate boundaries are absent in the data.

(3) Unlike previous studies that rely solely on the cells at the starting point of a fate, TOGGLE can reconstruct the entire fate trajectory in complete datasets with unbalanced sample sizes, without requiring pre-identification of progenitor and descendant cells.

(4) A corollary of Takens' theorem is proposed, and the resulting method is termed modal entropy embedding. This represents a novel dynamic reconstruction theory based on chaos phase space reconstruction. Hence, this corollary is named the TOGGLE corollary.

(5) This algorithm not only establishes the developmental and differentiation lineage of cells (e.g., hematopoietic stem cells) and the reverse induction lineage of regeneration (e.g., reprogrammed cells), but also constructs the programmed fate lineage of mature cells (e.g., ferroptosis), functional transformations (e.g., cardiac pathway differences induced by e-cigarettes), functional lineages (e.g., cardiomyocytes), and functional subgroups (e.g., fibroblasts). TOGGLE thus enables the construction of the complete growth process of cells from development and differentiation to becoming mature cells. Furthermore, our novel tool facilitates the establishment of a comprehensive fate-tracking process for mature cells, including functional differences, functional transformations, and programmed cell death.

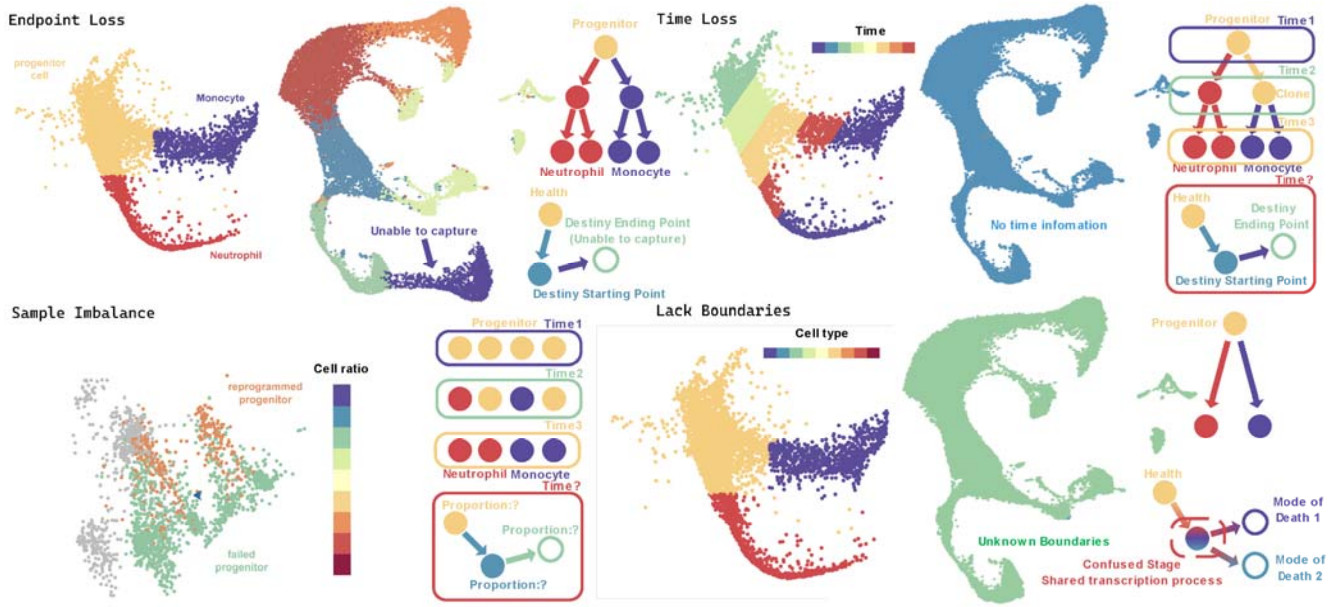
Endpoint Loss
progenitor cell
Monocyte
Neutrophil
Unable to capture
Progenitor
Neutrophil
Monocyte
Health
Destiny Ending Point
(Unable to capture)
Destiny Starting Point
Time Loss
Time
No time infomation
Progenitor
Time1
Time2
Clone
Time3
Neutrophil
Monocyte
Time?
Health
Destiny
Ending Point
Destiny Starting Point
Sample Imbalance
reprogrammed progenitor
failed progenitor
Cell ratio
Progenitor
Time1
Time2
Time3
Neutrophil
Monocyte
Time?
Proportion:?
Proportion:?
Proportion:?
Lack Boundaries
Cell type
Unknown Boundaries
Progenitor
Health
Confused Stage
Shared transcription process
Mode of Death 1
Mode of Death 2

Figure 1. Challenges in Cell Fate prediction. Endpoint Loss: Current tools typically requires capturing cells at the endpoint of differentiation, whereas fate prediction may not capture this final state; Time Loss: Current tools require data from multiple time points, whereas fate prediction often relies on data from a single time point; Sample Imbalance: Fate prediction models frequently encounter imbalanced sample sizes; Lack of Boundaries: Previous tools depends on distinct boundaries between cell types, whereas fate prediction of neurons lacks clear boundaries between cell states. The sample data for the drawing are from GSE140802, GSE99915 and GSE232429.

2 Results

We used hematopoietic and fibroblast reprogramming datasets to reconstruct cell lineages. TOGGLE successfully identified progenitor cells from the entire dataset, demonstrating its capability to classify functionally continuous cells into distinct types. Furthermore, TOGGLE predicted the developmental trajectory of progenitor cells, indicating its ability to perform intra-class classification within the same cell type. Previous algorithms were used for subtype classification of cells or binary classification of differentiation in developing cells, which required extensive preprocessing. However, TOGGLE can directly distinguish between progenitor cells, what do progenitor cells develop into, and differentiated cells, thereby eliminating the need for cumbersome identification and preprocessing steps.

Programmed cell fate is driven by RNA transcriptional pathways, and some studies focus on how these pathways are transformed. For example, with the increasing attention on RNA pathway alterations among public health specialists, we utilized datasets on e-cigarette and nicotine exposures (GSE183614) to analyze changes in cellular RNA pathways, These changes are not caused by genetic factors and may reflect the influence of non-epigenetic factors and long-term metabolic changes. The same computational process was applied to the differential genes in these datasets. Through long-term alterations in RNA pathways, cells may ultimately change their function. Therefore, identifying cell fate also requires distinguishing the execution of cellular functions as a crucial step.

Subsequently, we used neurons undergoing ferroptosis as sample to establish the programmed fate lineage of mature cells. Following ischemic stroke, cells within ischemic areas generally undergo various forms of programmed cell death, including apoptosis, autophagy, pyroptosis, and ferroptosis[16–18]. Conventional single-cell sequencing analysis tools face limitations in examining these multifaceted, progressive neuronal death processes. This is due to the challenge of capturing data from neurons that have already died, and because programmed cell death lacks a clear "decision boundary," making the construction of fate trajectories complex. Previous studies focused on cell subtype classification and cell type classification. Building on this foundation, TOGGLE can further distinguish different functional states within the same cell subtype, as well as differentiate the programmed fate stages within the same cell subtype.

A functional comparison table is provided in the Discussion.

2.1 Trace the cell lineage of hematopoietic development

To assess the efficacy of TOGGLE and make comparisons with other methods, we applied it to publicly available datasets with known lineage obtained from the Cospar software. The first dataset traced the differentiation of mouse bone marrow hematopoietic progenitor cells (HPCs) in a controlled culture environment (GSE140802) over four timepoints[9]. We identified cells using the method in Weinreb et al.'s work[9] and focused our analysis on three cell types: progenitor cells, monocytes, and neutrophils: progenitor cells, monocytes, and neutrophils. Progenitor cells exhibit a dual fate bias, as defined in the original Cospar research paper[15], where a portion differentiates into monocytes (referred to as monocyte progenitors), while another portion differentiates into neutrophils (referred to as neutrophil progenitors)[10]. Our objective is to trace the fate bias of progenitor cells, specifically to identify monocyte progenitors and neutrophil progenitors within the progenitor cell population.

TOGGLE is an unsupervised method. TOGGLE first generates a similarity matrix described by graph distances(Chapter 4.2). Cells that are relatively close are considered to have the same fate preferences and are clustered into a unified group in the same iteration round. The algorithm segments the distance matrix through binarization(Chapter 4.3), initially obtaining a boundary, then iteratively generating a second boundary based on the first(Chapter 4.4). Each classification results in a positively correlated group and a negatively correlated group, with the boundary between them forming a dividing line. Within the positively correlated group, a secondary positively correlated group and a secondary negatively correlated group are identified. Similarly, within the negatively correlated group, a secondary positively correlated group and a secondary negatively correlated group are identified. This process creates a total of two new boundaries at the intersections(Chapter 4.4). This process results in different groupings of cells, which are then sorted. After sorting, then deriving the required number of categorie according to the resolution set by number of target categories boundaries(Chapter 4.5). After completing the segmentation, samples within adjacent boundaries form a cluster. In this process, progenitor cells close to neutrophils are determined by TOGGLE to be neutrophil progenitors, while progenitor cells close to monocytes are determined to be monocyte progenitors (Figure 2G). (1)When analyzing descendants independently, by comparing with the true tracing labels in ground truth analysis, the accuracy of TOGGLE is found to be 77%, predict result as Figure 2H. Notably, the TOGGLE algorithm is implemented as a unsupervised method without prior training. In order to evaluate the performance of TOGGLE, we compared it to four widely used in silico algorithms: Cospar, Super OT, WOT, and GAN-Based OT. Cospar attained a rate of 65%, Super OT exhibited a slightly higher accuracy at 67%, WOT maintained a similar accuracy of 65%, and GAN-Based OT achieved 62%. (2)When both progenitor and descendant cells are selected for classification, TOGGLE achieves an accuracy of 82%. The accuracy of WOT is less than 50%, while Cospar cannot be used for this calculation.

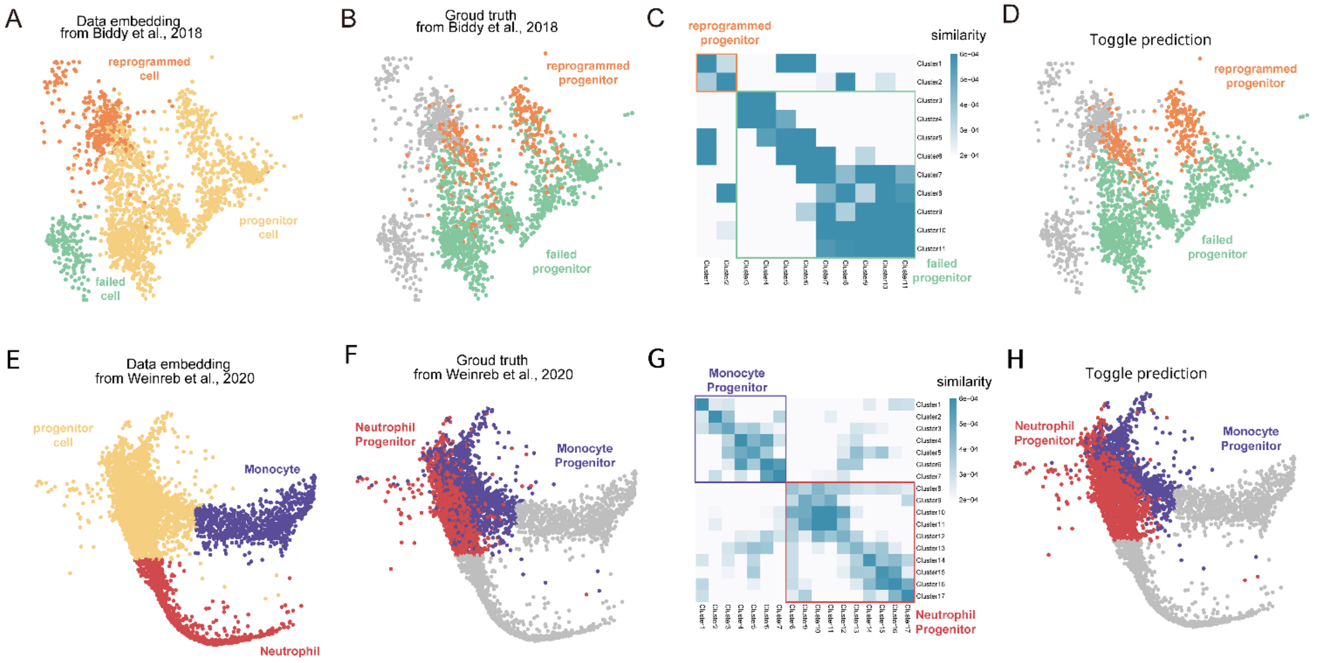

Figure 2. Tracing the lineage of fibroblast reprogramming[9] and hematopoietic cell[16] development. A: Data embedding of scRNA-seq data on experimentally-induced fibroblast reprogramming over six sampled timepoints in Biddy et al., 2018[19]; B: Ground truth of reprogrammed data; C: Similarity matrix of clustering results of reprogrammed data; D: TOGGLE's prediction result of reprogrammed data; E: Data embedding of neutrophil progenitor data from Weinreb et al. 2020[9]; F: Ground truth of neutrophil progenitor data; G: The similarity matrix of clustering results of neutrophil progenitor data; H: The TOGGLE's prediction result of neutrophil progenitor data.

2.2 Trace the cell lineage of fibroblast reprogramming

We also validated the performance of TOGGLE on a dataset from mouse embryonic fibroblasts sampled at eight timepoints across transcription factor (TF) direct lineage reprogramming to induced endoderm progenitor cells (GSE99915)[19]. In this experiment, cells initially expanded, but many were lost due to aging or cell death. By day 28, 30% of the cells were successfully reprogrammed[19]. Building on the work of Biddy et al.[19], we identified cells using genes such as *Spint2*, *Apoa1*, *Peg3, and Col1a2*, which are strongly correlated with successful or failed reprogramming. For further analysis, we focused on progenitor cells (fibroblasts), reprogrammed cells, and failed cells. Induced progenitor cells display a dual fate bias, wherein a fraction differentiates into reprogrammed cells, referred to as reprogrammed progenitors, while another fraction differentiates into failed-reprogrammed cells, referred to as failed progenitors[19]. Here, we investigated the fate bias of fibroblasts undergoing transcription factor-induced reprogramming to progenitors, specifically by identifying the successfully reprogrammed progenitors and failed progenitors within the progenitor cell population.

We employed TOGGLE to reconstruct the complete tracing of the reprogramming process of fibroblasts to stem cells, methods same as chapter 2.1. After completing the segmentation, samples within adjacent boundaries form a cluster. In this process, progenitor cells close to Reprogrammed cells are identified by TOGGLE as Reprogrammed progenitor cells, while those close to Failed cells are identified as Failed progenitor cells (Figure 2C). (1)When analyzing descendants independently, by comparing with the true tracing labels in the ground truth, TOGGLE achieved an accuracy of 91.32%, predict result as Figure 2D. TOGGLE is a unsupervised algorithm, and we benchmarked it against the two most common in silico algorithms, with Cospar's accuracy being 67%, and WOT's accuracy being 31%, showing a significant accuracy advantage. (2)When both progenitor and descendant cells are selected for classification, TOGGLE achieves an accuracy of 84%. The accuracy of WOT is less than 50%, while Cospar cannot be used for this calculation.

2.3 Fate Differentiation of Mature Cells

2.3.1 Changes in RNA Pathways

TOGGLE was applied to analyze the effects of e-cigarette aerosol inhalation on mouse cardiac tissue using dataset GSE183614[20]. Taking the differentially expressed genes of adult mice exposed to e-cigarette smoke and fresh air as an example, they were divided into 6 groups(Chapter 4.5). Except for Group 2, Group 3, Group 4 and Group 5, the remaining groups did not show significant enrichment in biological processes or KEGG signaling pathways and require further investigation (Figure 3). The enrichment results of each gene group and the grouping results of differentially expressed genes for other intervention methods are shown in Supplementary Material 7. The results indicate that exposure to e-cigarette aerosol significantly altered RNA pathways in the hearts of mice. These changes are reflected in the grouping of differentially expressed genes and their enrichment in biological functions and signaling pathways. More pronounced immune and inflammatory responses (e.g., TNF signaling) and cell cycle-related changes among young mice compared to mature adults suggested that the impact of e-cigarette use is greater at younger ages.

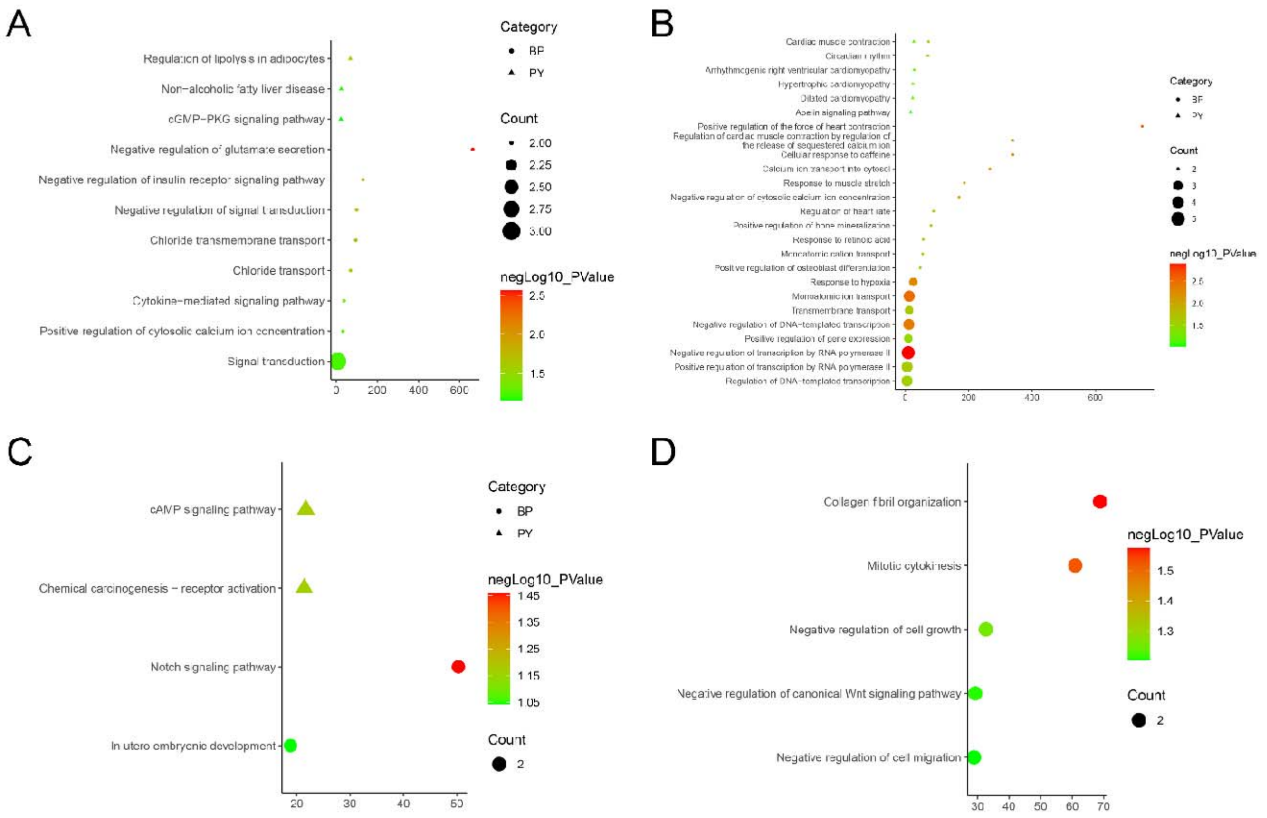

Figure 3. The Enrichment Results of Genes in Each Gene Group. A: Biological processes and pathways of Group 2; B: Biological processes and pathways of Group 3; C: Biological processes and pathways of Group 4; D: Biological processes of Group 5. (BP: biological processes; PY: signaling pathways. X-axis is fold enrichment)

2.3.2 Functional Differentiation of Mature Cells

TOGGLE was applied to analyze single-nucleus RNA sequencing (snRNA-seq) and bulk RNA sequencing (bulk-RNA seq) data from a rat myocardial infarction (MI) model (dataset: GSE253768). After classifying the snRNA-seq data into various cell types, including fibroblasts, myofibroblasts, macrophages, T cells, and cardiomyocytes (Figure 4A), TOGGLE divided the fibroblasts and myofibroblasts into seven groups (Figure 4B). The MI group was predominantly represented by FibR1-G5, FibR1-G6, and FibR1-G7 (Figure 4C). The results of the inferred ligand-receptor interactions between these fibroblasts and T cells or macrophages (cell-cell communication) revealed differences in the strength and types of interactions, suggesting functional variations among these fibroblast subgroups (Figures 4D and 4E). Further analysis of the highly variable genes in the MI fibroblasts also identified eight distinct functional groups (Figure 4F). Additionally, TOGGLE was used to map bulk RNA-seq data onto the snRNA-seq dataset, revealing that the gene expression pattern of fibroblasts in the bulk RNA-seq MI group closely matched that of FibR2-G6 (highlighted in the red box in Figure 4G). This finding suggests that the bulk RNA-seq fibroblasts predominantly exhibit functions similar to those of FibR2-G6, which are associated with cell migration, adhesion, angiogenesis, and ECM-receptor interactions (Figure 4H).

Thus, TOGGLE can effectively combine bulk RNA-seq data with snRNA-seq to contextualize each, and which cell types and specific phenotypes undergo nuclear transcriptional changes that most closely represent overall tissue-level changes in transcription.

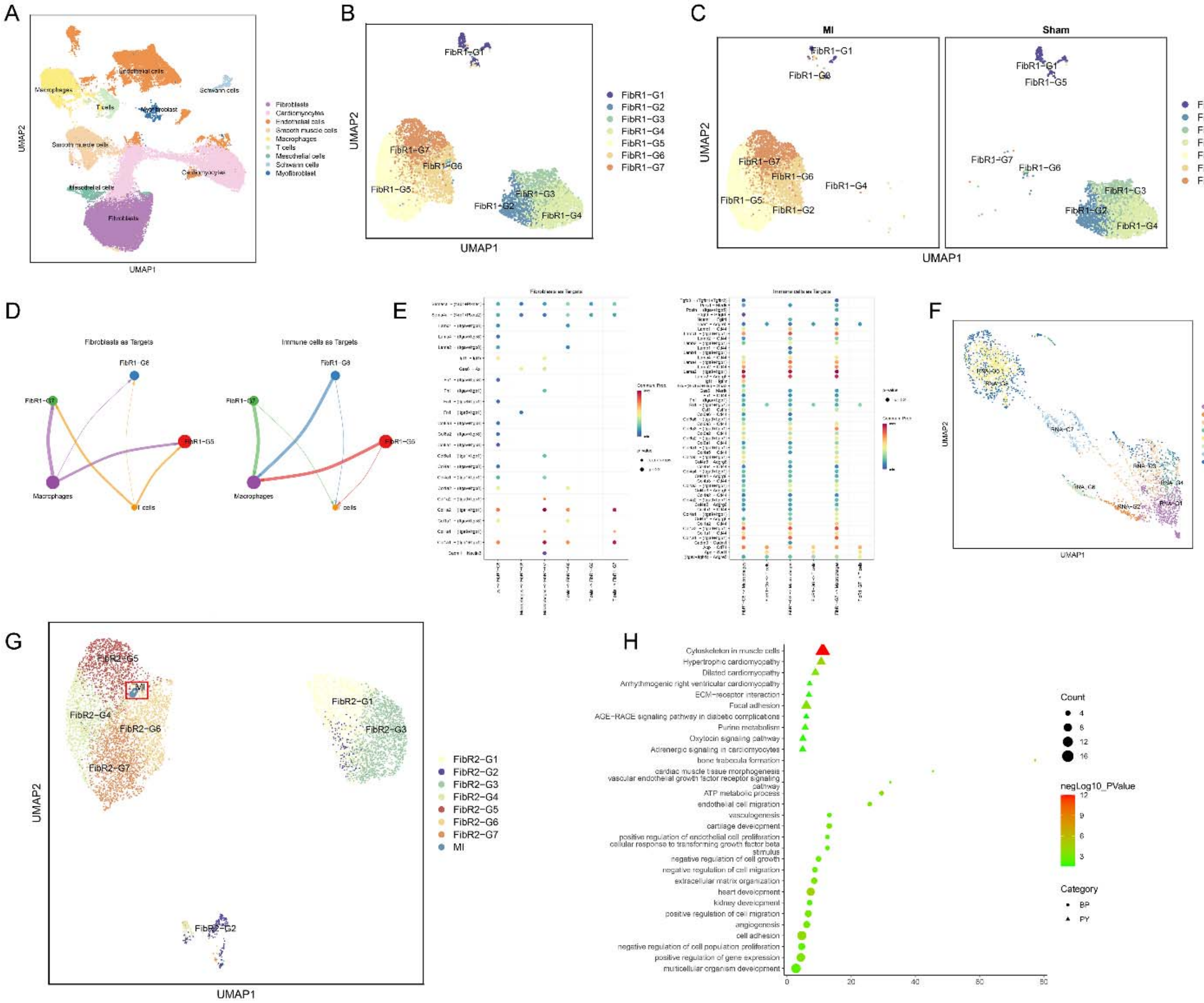

Figure 4 Analysis results of GSE253768 set. A: Cell types of snRNA-seq; B: Fibroblasts and myofibroblasts were divided into 7 groups after TOGGLE analysis; C: Distribution of 7 cell types in MI cohort and Sham cohort; D: Weight/strength diagram of cell communication between fibroblasts and immune cells; E: Bubble diagram of cell communication between fibroblasts and immune cells; F: RNA distribution of fibroblasts in the MI group; G: Mapping MI group bulk RNA-seq data onto UMAP of fibroblast snRNA-seq results; H: Enrichment results of FibR2-G6 (BP: biological processes; PY: signaling pathways. X-axis is fold enrichment).

2.4 Programmed Cell Death Trajectories Prediction in Neurons

2.4.1 Preliminary Prediction of Ischemic Stroke-Induced Neuronal Fate

The above results confirm the reliability of TOGGLE for predicting cell fate. In cases of ischemic stroke, neurons often undergo programmed cell death, with ferroptosis playing a significant role in neuronal loss. However, ferroptosis frequently overlaps with other forms of programmed cell death, making it challenging for existing algorithms to effectively distinguish ferroptosis from other cell death mechanisms. Therefore, we applied TOGGLE to predict and analyze neuronal fate in a rat model of ischemic stroke, distinguishing between healthy and dying neurons to forecast programmed cell death, particularly ferroptosis. And we further verified the expression of apoptosis genes and ferroptosis genes in animal experiments. Ischemic stroke samples [middle cerebral artery occlusion (MCAO) cohort] and healthy samples (Sham cohort) from the GSE232429 dataset were included in the analysis[13]. The neuronal scRNA-seq data were then analyzed using TOGGLE and initially divided into five groups by dimension reduction using UMAP: Group R1-1, Group R1-2, Group R1-3, Group R1-4, and Group R1-5 (Figure 5A).

Cohorts were analyzed for differentially expressed genes, which we cross-referenced with programmed cell death-related genes[3] that we then counted in each comparison group (Figure 3D). Group R1-1 vs. Group R1-5 had fewer programmed cell death-related genes, while comparisons between Group R1-1 and other groups, as well as Group R1-5 and other groups, had a higher number of programmed cell death-related genes. This suggests that Group R1-1 and Group R1-5 exhibit similar states, potentially representing consecutive stages in a continuum. Additionally, the cell counts of Group R1-1 and Group R1-5 in the sham cohort were high (Figure 5B and 5C), suggesting that both groups consisted of cells that had not entered a programmed cell death state. Given that Group R1-5 was abundant in both MCAO cohort and sham cohort, we inferred that it represented healthy cells. Furthermore, Group R1-2 cells were more prevalent in the MCAO cohort compared to sham and had a significantly more differentially expressed programmed cell death-related genes compared to other cell groups, indicating that Group R1-2 consists of cells that have entered a programmed cell death state. Given the moderate number of differentially expressed programmed cell death-related genes among cells in Group R1-3 and Group R1-4, they are inferred to be in intermediate states.

Based on CellMarker, the cells of Group R1-3 were identified as Mtss2+Gpr37l1+ glutamatergic neurons, and that of Group R1-4 were identified as Atp2b2+Grin2a+ glutamatergic neurons. As shown in Figure 5B and 3C, the cell counts of Group R1-3 and Group R1-4 showed little difference between the MCAO and Sham cohorts, which suggests that these two groups of cells had not undergone programmed cell death. Therefore, these groups were excluded from further analysis. The UMAP plots after excluding Groups R1-3 and R1-4 are shown in Figure 5E.

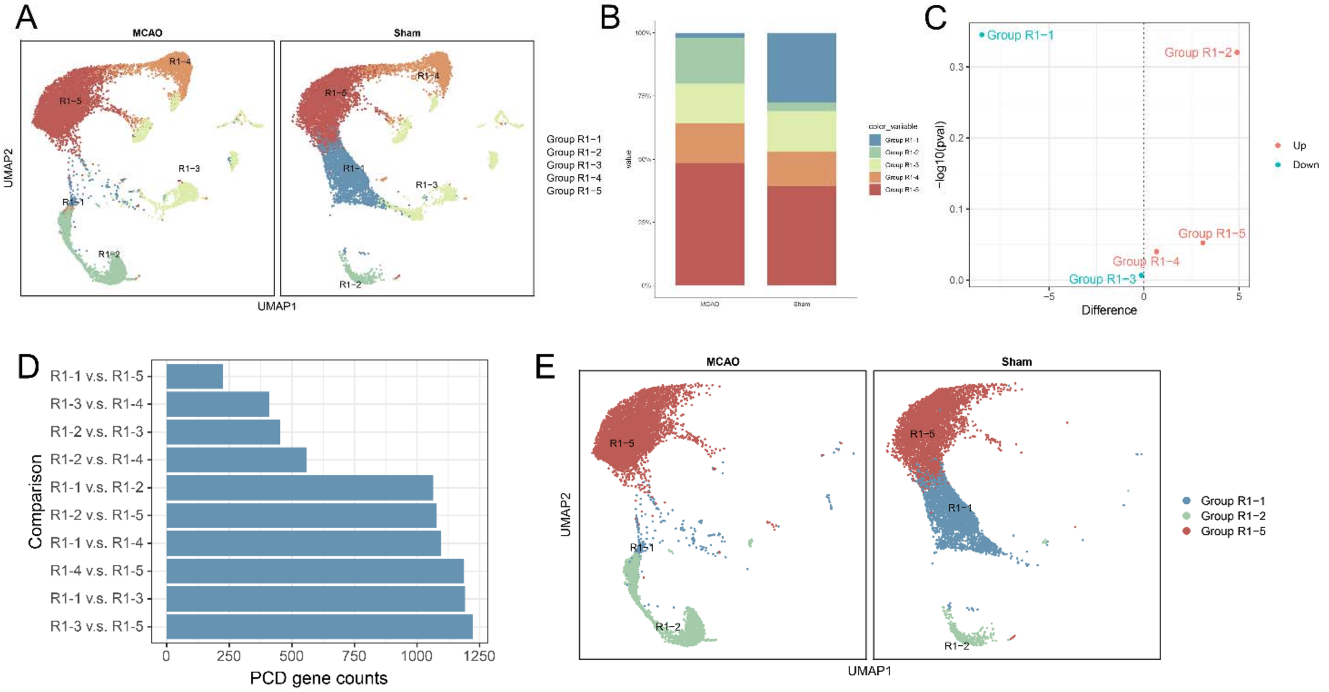

Figure 5. Preliminary Prediction of Neuronal Fate. A: The grouping results of the preliminary analysis; B: The proportion of each cell group in MCAO and Sham; C: Compared with Sham cohort, the changes in cell number in each cell group in MCAO cohort; D: Comparison of the number of differentially expressed programmed cell death genes between cell groups of all cohorts; E: UMAP graph without R1-3 and R1-4.

2.4.2 Prediction of Neuronal Programmed Cell Death Trajectories

The data of Group R1-1, Group R1-2, and Group R1-5 of the preliminary analysis results were further analyzed to predict their programmed cell death trajectories. Monocle3 was utilized to analyze their pseudotime (Figure 6A), it can be observed that the loss of temporal information is significant because the time restoration is not synchronized with the trajectory scale. And based on this data, TOGGLE identified 9 subgroups (Figure 6B and 6C). Based on the number of programmed cell death-related genes and the cell count within each group (Figure 6D), the following can be inferred: (1) Groups R2-1 are closely related to Group R2-6. (2) Groups R2-3 are distant from Groups R2-5, R2-6, and R2-7; similarly, Groups R2-1 are distant from Groups R2-5 and R2-7, consistent with the pseudotime analysis. (3) Group R2-2 is closely related to Group R2-4; however, since the cell counts of Group R2-2 in the sham cohort is larger, it is considered potential noise (Figure 6E, 6F, 6G). R2-9 cells are likely representative of healthy cells. R2-8 cells are predominantly observed in the MCAO cohort, while R2-1, R2-5, R2-6, and R2-7 are more frequent in the Sham cohort, suggesting they may represent intermediate states transitioning toward programmed cell death. In contrast, R2-2 and R2-3 cells show a marked increase in the MCAO cohort, indicating they are primarily composed of cells undergoing programmed cell death. The trajectory of programmed cell death may be Group R2-9→Group R2-8→Group R2-1→Groups R2-5 to R2-7→Group R2-2→Group R2-3.

In summary, we deduced that Group R2-9 is the most likely starting point and Group R2-3 the most plausible terminus. The FindMarkers function of Seurat was used to perform precise differential gene expression analysis. Group R2-1, Group R2-2, Group R2-3, Groups R2-5 to R2-7, and Group R2-8 were set as ident1, and Group R2-9 was set as ident2. The volcano plot of Group R2-9 v.s. Group R2-3 genes is shown in Figure 6H. The differentially expressed genes between these groups were imported into DAVID (ver. 2024q2, https://david.ncifcrf.gov) for Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway enrichment analysis, so as to obtain the number of genes in representative programmed cell death pathways (such as apoptosis and ferroptosis) (Figure 6I and 6J). It was found that as the trajectory moved away from Group R2-9, the number of apoptosis- and ferroptosis-related genes expressed in each group gradually increased, which was consistent with the inferred time trajectory. The results of enrichment analysis are shown in Table S1 (see supplementary materials), and the differentially expressed genes are shown in Table S2.

Among the differentially expressed genes of Group R2-3 vs. Group R2-9 (Table S2), several apoptosis- and ferroptosis-related genes were identified: *Gpx4* (Log2FC=-0.60, FDR<0.05), *Trf* (Log2FC=6.98, FDR<0.05), *Tfrc* (also known as *TFR*, Log2FC=3.38, FDR<0.05), *Smad1* (Log2FC=1.16, FDR<0.05), *Smad5* (Log2FC=1.70, FDR<0.05), *Acsl4* (Log2FC=1.39, FDR<0.05), *Rela* (also known as *p65*, Log2FC=1.08, FDR<0.05), *Bax* (Log2FC=1.46, FDR<0.05), and *Nfkbia* (also known as *IkBA*, Log2FC=4.00, FDR<0.05). These gene expression patterns are consistent with those observed during apoptosis and ferroptosis. However, several apoptosis- and ferroptosis-related genes showed expressions inconsistent with previous studies, including *Aifm2* (also known as *FSP1*, Log2FC=0.66, FDR<0.05), *Sirt1* (Log2FC=2.56, FDR<0.05), and *Casp3* (Log2FC=-0.40, FDR<0.05). This may be because cells in the early stages of programmed cell death are trying to "self-rescue." In addition, some positively expressed ferroptosis-driving genes were found in the differentially expressed genes. The ferroptosis-driving genes refer to FerrDb V2 (http://www.zhounan.org/ferrdb/current/)[21] (Table S2). The positions of these genes within relevant pathways are illustrated in Figure 7.

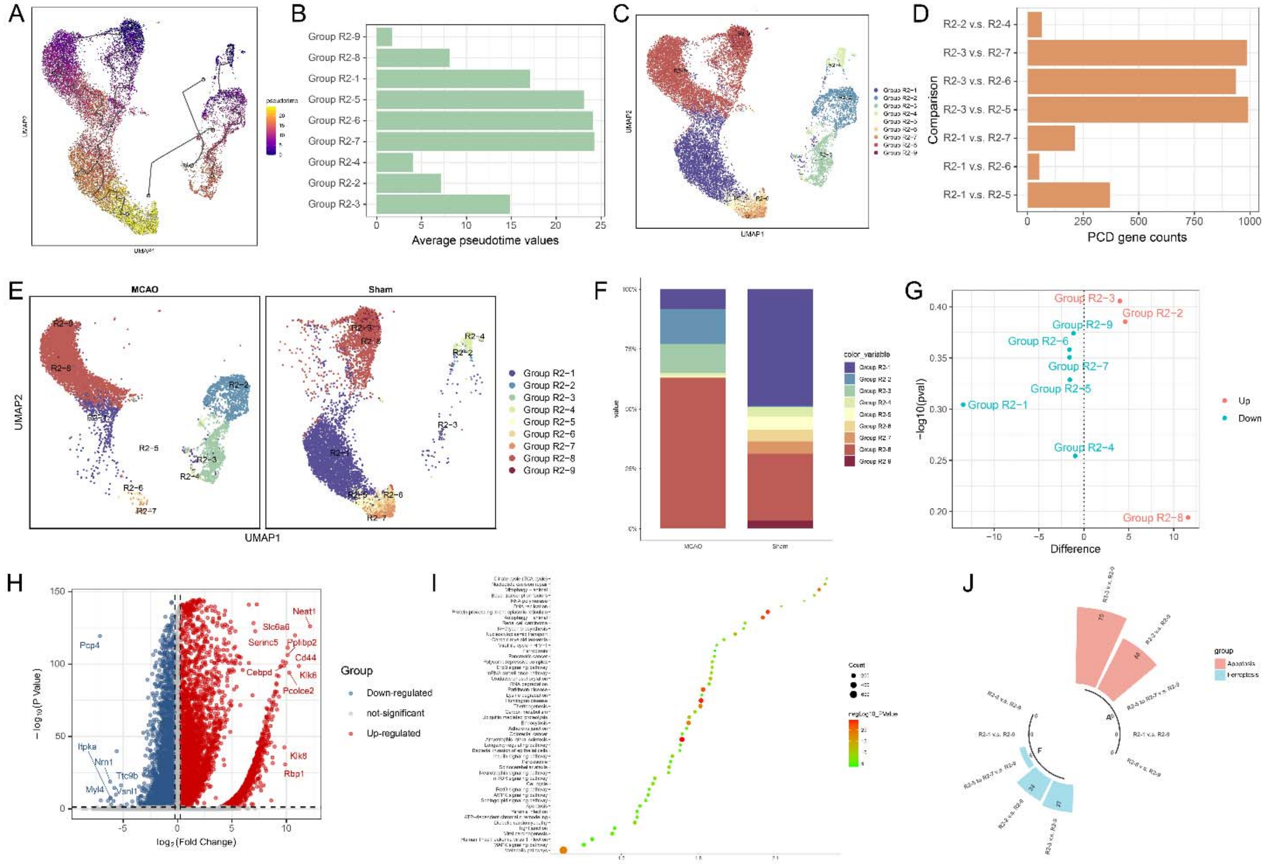

Figure 6. Programmed Cell Death Trajectories Prediction. A: Pseudo-time trajectory of cells; B: The average pseudotime values of groups; C: Cells in previous groups R1-1, R1-2, R1-5 further analyzed by TOGGLE that divided into 9 groups; D: The number of programmed cell death genes in each cell group; E: Distribution of cell groups in Sham and MCAO; F: The proportion of those 9 cell groups in MCAO and Sham; G: Compared with Sham, the changes in cell number in those 9 cell groups in MCAO; H: Volcano plot of Group R2-3 v.s. Group R2-9 genes; I: Top 50 pathways of Group R2-3 v.s. Group R2-9 genes; J: The number of genes enriched in apoptosis and ferroptosis. Pink genes were selected for experimental verification in follow-up studies (Figure 9 and Figure 10). PCD: programmed cell death.

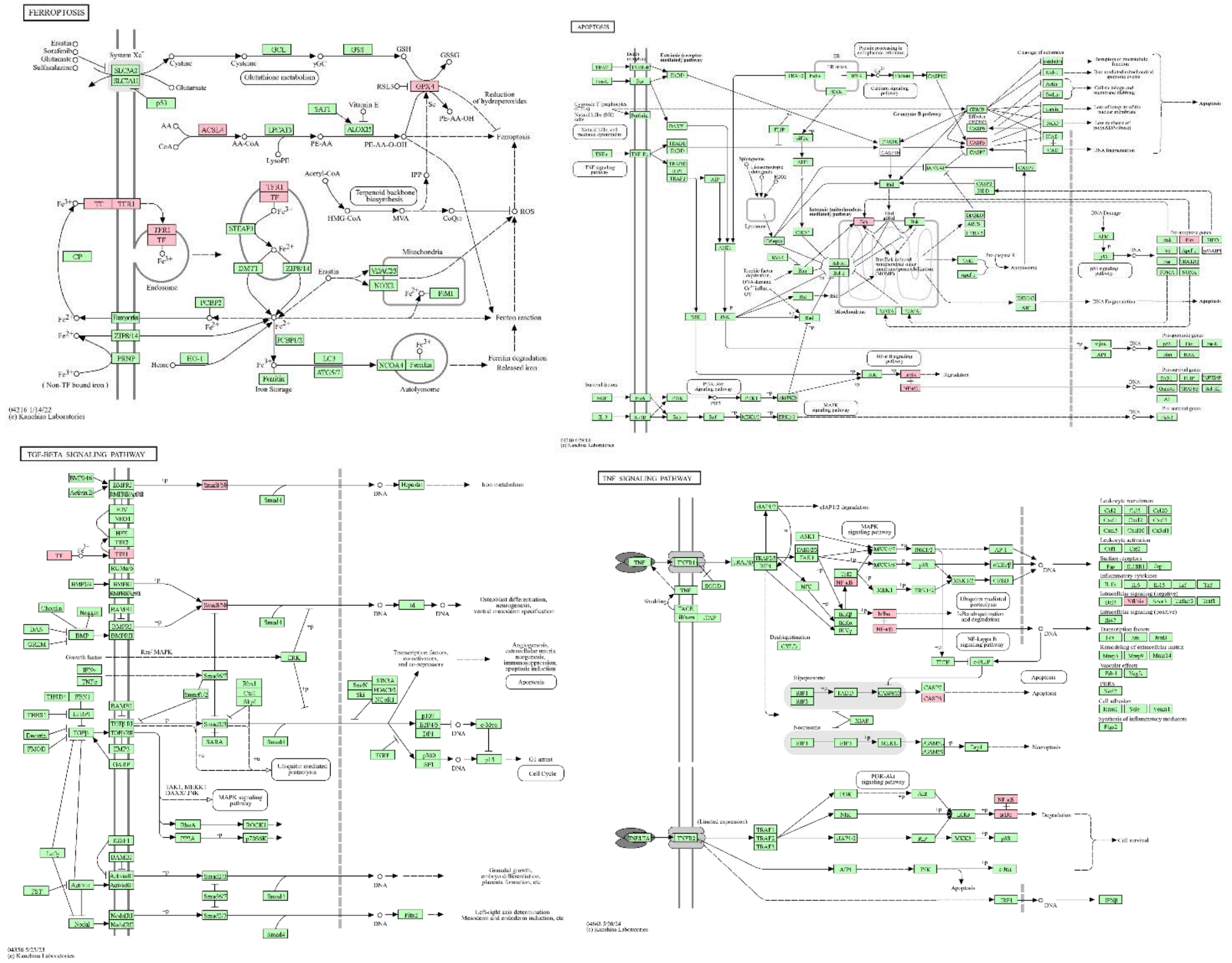

Figure 7. Signaling pathways related to ferroptosis and apoptosis. (modified from KEGG pathway mmu04216, mmu04210, mmu04350 and mmu04668[22])

To further analyze ferroptosis-dominant cell populations, TOGGLE was applied to Groups R2-2 and R2-3, and five groups (Figure 8A). Using the FindAllMarkers function of Seurat, genes uniquely and highly expressed in each group were identified with selection criteria of Log2FC ≥ 0.25 and FDR < 0.05. Ferroptosis driver genes were sourced from the FerrDb database (Ver. 2, http://www.zhounan.org/ferrdb/current/). Among the differentially expressed genes in Groups R3-4 and R3-5, ferroptosis driver and other programmed cell death genes were most prevalent. Furthermore, these groups displayed a higher number of highly expressed ferroptosis driver genes (Log2FC ≥ 2 and FDR < 0.05) compared to other groups, with over 70% of cells in Groups R3-4 and R3-5 showing ferroptosis markers relative to healthy cells (Figure 8B, 8C and 8D). The enrichment results showed that the pathways enriched by differentially expressed genes included ferroptosis (Figure 8E). These findings support identifying Groups R3-4 and R3-5 as ferroptosis-dominant cell populations.

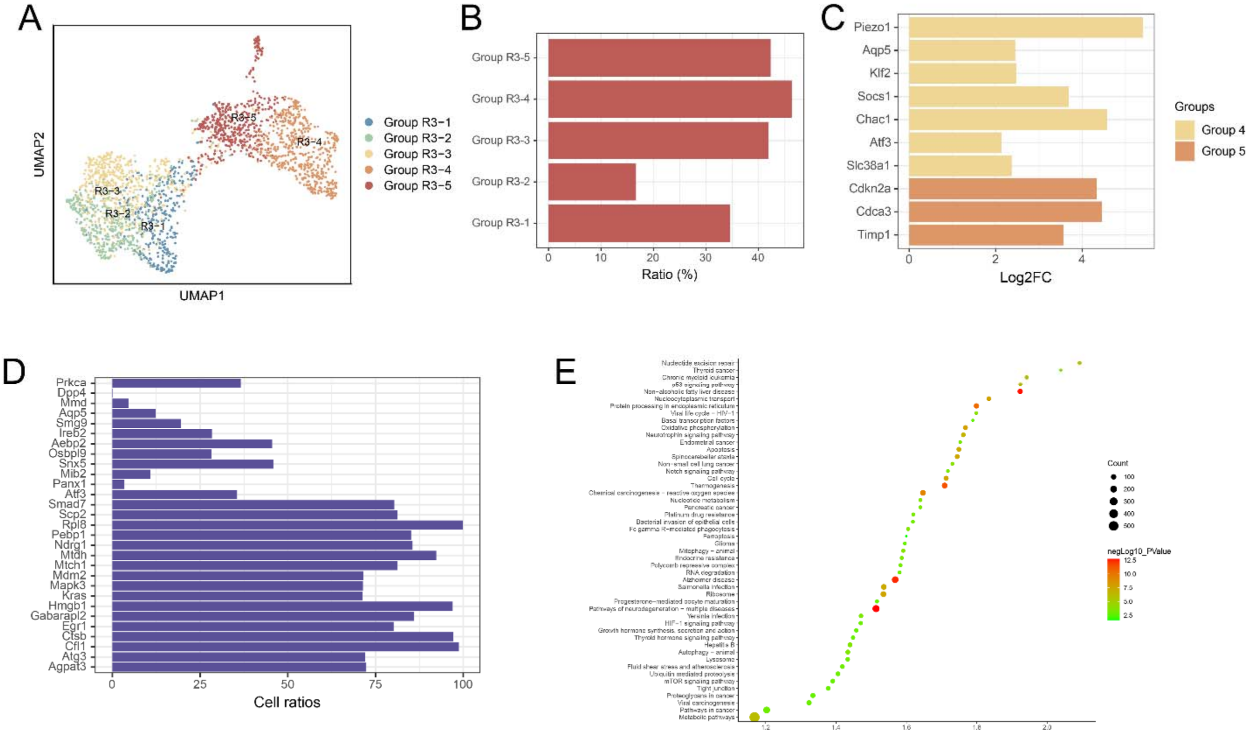

Figure 8. Ferroptotic Neurons Prediction. A: Cells in previous Groups R2-2 and R2-3 further analyzed by TOGGLE and divided into 5 groups; B: The proportion of ferroptosis-driving genes in each cell group; C: Log2FC values of highly expressed ferroptosis-driving genes; D: The proportion of cells expressing different ferroptosis-driving genes; E: Top 50 pathways of Group R3-4 and Group R3-5 genes

2.5 Verification of apoptosis- and ferroptosis-related genes

Our previous research demonstrated increased iron levels and neuronal iron accumulation in the infarcted brain tissue of the MCAO model. Deferiprone treatment in MCAO model effectively reduces iron content and diminishes iron accumulation[23]. Based on these findings, we used the MCAO model ± deferiprone to validate the predicted results of neuronal ferroptosis.

2.5.1 Pathological changes of brain tissue

In the Sham operation group, neurons exhibited normal morphology, with orderly arrangement, clear nuclear membranes, uniformly stained cytoplasm, and no evident pathological changes. In contrast, the MCAO group showed significant structural damage in the ischemic cortical area, with disorganized neuronal arrangement, widened intercellular spaces, and signs of cell swelling, rupture, and nuclear condensation, leading to vacuole formation. In the MCAO+deferiprone group, neuronal structural damage was mild, with minimal cell membrane rupture, slight nuclear condensation, and a significant reduction in vacuolar degeneration (Figure 9).

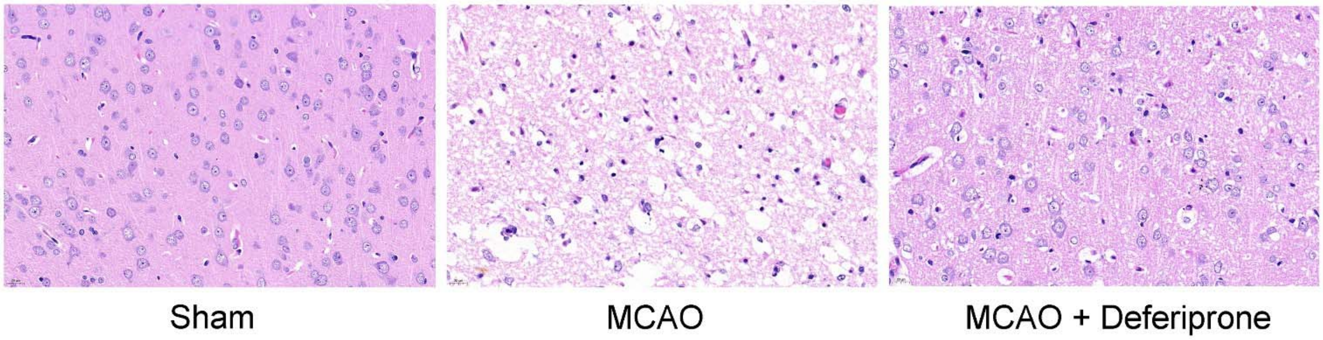

Figure 9 Pathological changes of brain tissue (HE staining, 40× magnification)

2.5.2 Ultrastructural changes of brain tissue

In the Sham group, mitochondria appeared structurally normal with clear features and intact membranes. In contrast, the MCAO group exhibited mitochondrial atrophy, reduced cristae, and increased membrane density, consistent with the morphological characteristics of ferroptosis. The MCAO+deferiprone group showed significant improvement in mitochondrial integrity compared to the MCAO group, with reduced damage (Figure 10).

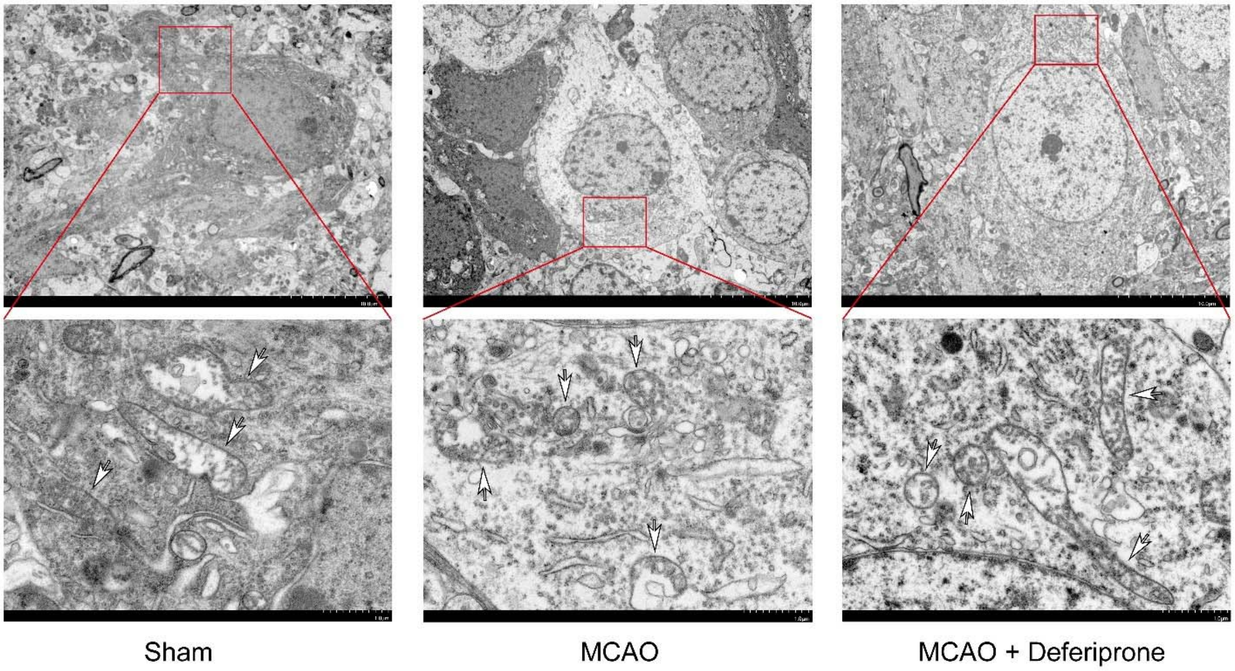

Figure 10 Ultrastructural changes of brain tissue (Transmission electron microscope. The upper images are at 1500× magnification, and the lower images are at 6000× magnification. Mitochondria are marked by white arrows.).

2.5.3 Expression levels of *Acsl4, Gpx4* and *Fsp1* proteins detected by immunohistochemistry

Compared to the Sham operation group, the MCAO group showed an increase in Acsl4 levels ($P<0.05$) and a decrease in Gpx4 levels ($P<0.05$). These findings are consistent with the differential gene expression results from single-cell sequencing. Meanwhile, compared to the Sham operation group, the MCAO group showed an decrease in Fsp1 levels ($P<0.05$). Compared with the MCAO group, the MCAO+deferiprone group showed a decrease in Acsl4 levels and Fsp1 levels ($P<0.05$), and a increase in Gpx4 levels ($P<0.05$). (Figure 11).

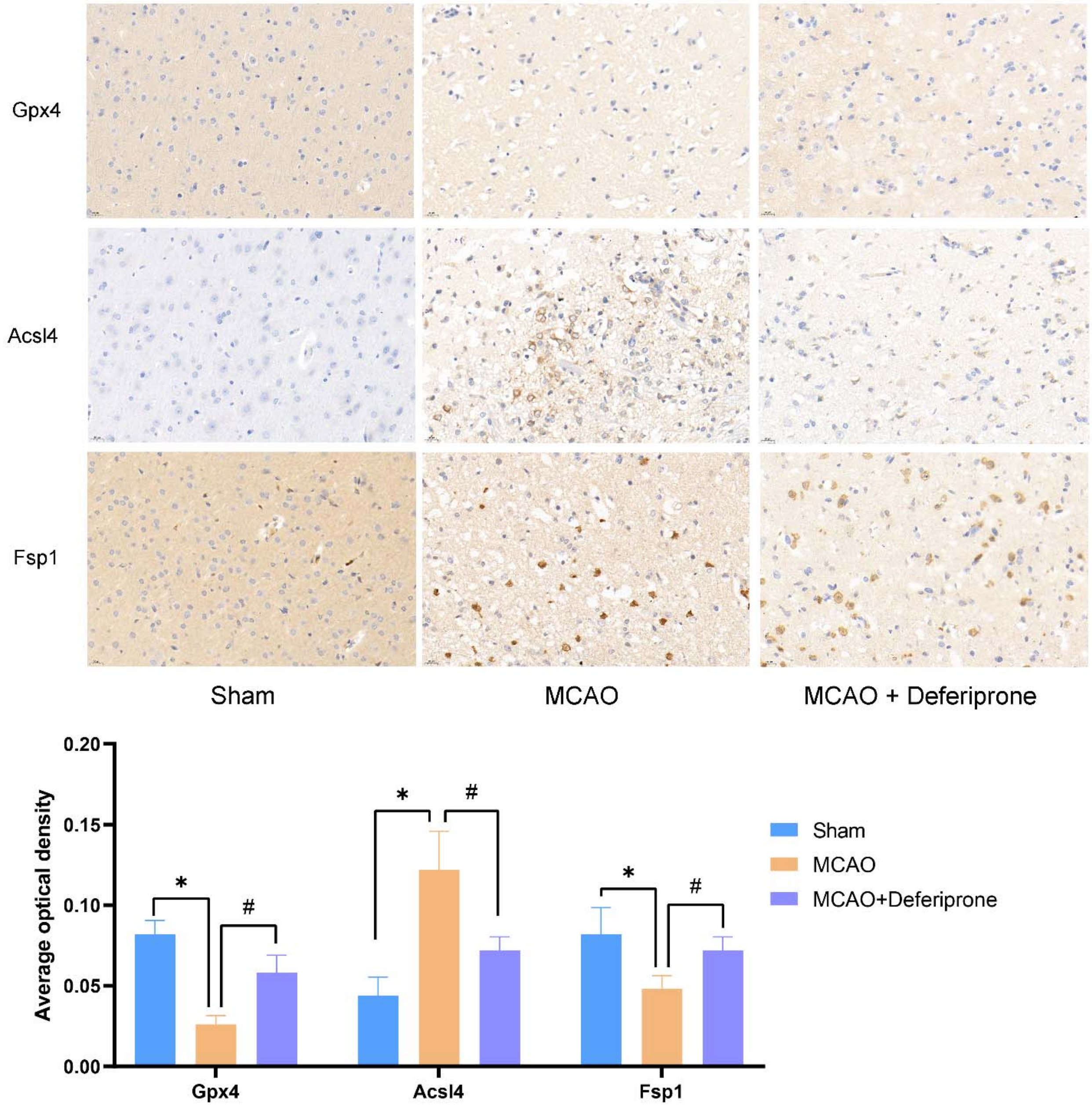
Gpx4
Acsl4
Fsp1
Sham
MCAO
MCAO + Deferiprone
Average optical density
0.20
0.15
0.10
0.05
0.00
*
#
*
#
*
#
Gpx4
Acsl4
Fsp1
Sham
MCAO
MCAO+Deferiprone

Figure 11 Expression levels of Acsl4 and Gpx4 proteins. (Immunohistochemistry, 40× magnification. *compared with the sham group, $P<0.05$; #compared with MCAO group, #$P<0.05$. n=3 per group)

2.5.4 Expression levels of *Bax, IkBa, Casp3, p-Smad, p65(Rela), Tfrc, Trf and Sirt1* proteins detected by immunofluorescence staining

Compared to the Sham operation group, the MCAO group exhibited significantly elevated levels of Bax, IκBα, p-Smad, p65 (Rela), Tfrc, and Trf proteins ($P<0.05$). These results are consistent with the differential gene expression findings from the single-cell sequencing analysis. Meanwhile, Compared to the Sham operation group, the Casp3 protein level was increased, while the Sirt1 protein level was decreased in the MCAO group ($P<0.05$) (Figure 12).

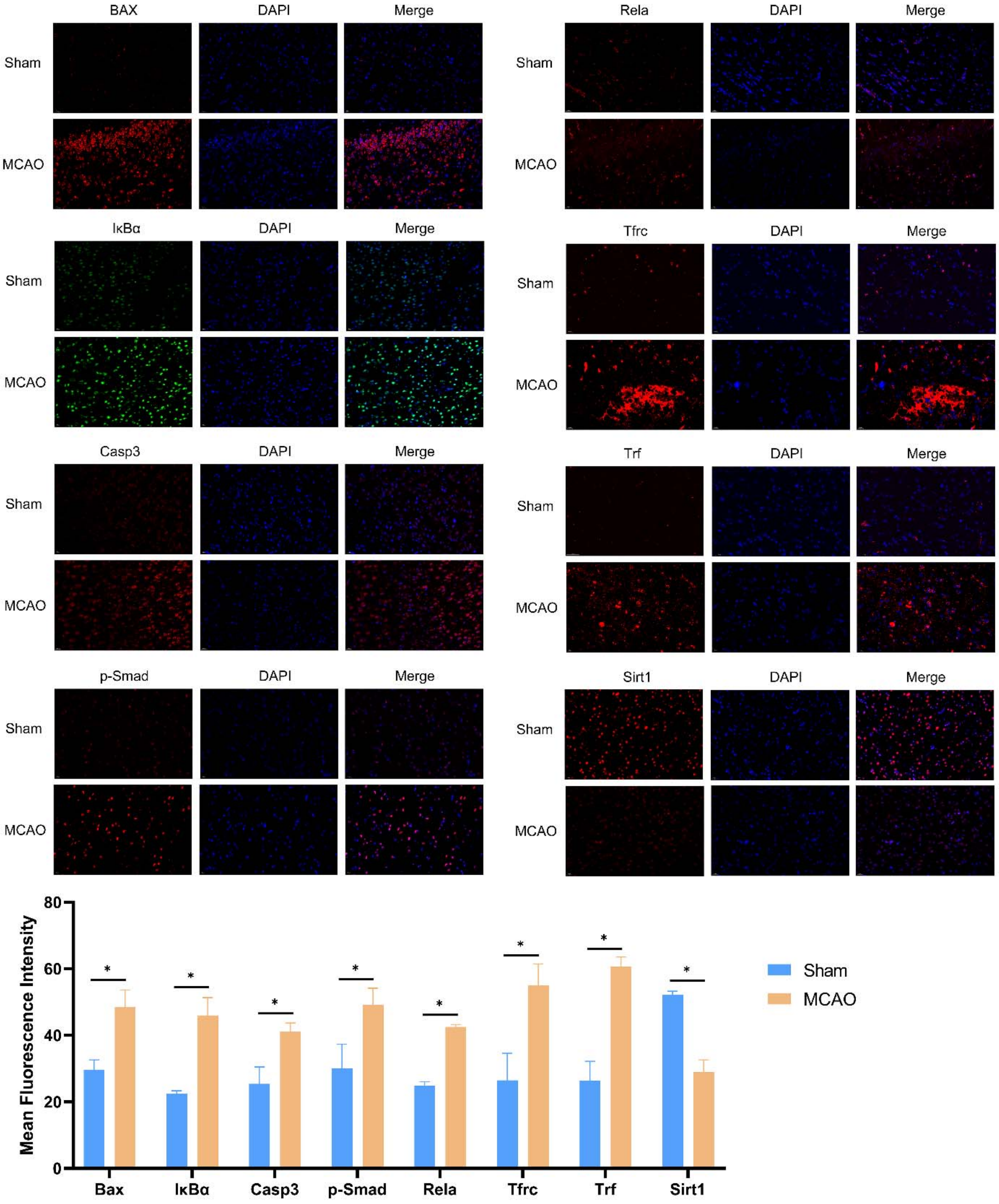
BAX
DAPI
Merge
Sham
MCAO
Rela
DAPI
Merge
Sham
MCAO
IκBα
DAPI
Merge
Sham
MCAO
Tfrc
DAPI
Merge
Sham
MCAO
Casp3
DAPI
Merge
Sham
MCAO
Trf
DAPI
Merge
Sham
MCAO
p-Smad
DAPI
Merge
Sham
MCAO
Sirt1
DAPI
Merge
Sham
MCAO
Mean Fluorescence Intensity
80
60
40
20
0
*
Bax
IκBα
Casp3
p-Smad
Rela
Tfrc
Trf
Sirt1
Sham
MCAO

Figure 12 Expression levels of Bax, IkBa, p-Smad, p65, Tfrc, and Trf proteins (Immunofluorescence, 40× magnification. Compared with the sham group, *P<0.05. n=3 per group).

3 Discussion

This research developed TOGGLE, a unsupervised learning method for predicting cell fate using scRNAseq data. TOGGLE allows researchers to obtain a graph distance matrix between cells, which can be interpreted as a measure of cell similarity. This matrix reveals the approximate relationships between cells. One of the key features of TOGGLE is its flexible method for calculating the similarity matrix. This flexibility allows for easy updates to the algorithm to meet specific requirements, without the need for extensive modifications to subsequent steps. As a result, the scalability of the TOGGLE model is enhanced. Once the similarity matrix is obtained, TOGGLE facilitates the unsupervised clustering of cells, grouping them based on their states. The algorithm's resolution can be adjusted during this step to achieve a desired number of clusters, tailored to the specific needs of the analysis. Cells that are clustered together not only share significant similarity in terms of cell type, but also likely have consistent fate determinations. This ensures their coherence in pseudotemporal order. After determining the cell clusters, TOGGLE can be further utilized to acquire pseudotime ordering results, which enables the reconstruction of the complete dynamical process of the cells.

We established a lineage tracing model using the development and differentiation of hematopoietic stem cells (GSE140802) and then utilized reprogrammed cells (GSE99915) to construct the lineage of regenerative processes to demonstrate TOGGLE’s utility in elucidating the single-cell functions of forward differentiation and reverse differentiation. We conducted a benchmark evaluation of TOGGLE using two scRNA-seq datasets that include known lineage information. In addition, we constructed an RNA functional map using bulk RNA-seq of the effects of e-cigarette on the myocardium (GSE183614), and constructed a fibroblast functional map and an MI cohort fibroblast RNA functional map using snRNA-seq data of MI (GSE253768), further expanding the application scope of TOGGLE. In this evaluation, we compared the performance of TOGGLE with widely used algorithms. The results reveal that TOGGLE consistently outperformed both WOT and cospar in accurately predicting the fate of progenitor cells and effectively constructing the cell development process through pseudotime calculation. Compared with mainstream algorithms, our method demonstrates the following advancements (Table 1):

Table 1: Advanced algorithms

Previous	**Our Advantages**	**Reference**
Cell Rank	Used for grouping cell subtypes. TOGGLE not only enables subtype grouping but also distinguishes functional stages and fate stages.	Marius Lange et al.[24] *Nature Methods*

DeepCell	Pre-training is required. TOGGLE does not require pre-training. In Fig. 1, we introduce Problems 1 and 4, which are typically not applicable to pre-training scenarios.	Dylan Bannon et al.[25] *Nature Methods*
scGen	Cell perturbation and infection responses across different cell types. TOGGLE can be used to analyze cell perturbation and functional differences within the same cell type under different states.	Mohammad Lotfollahi et al.[26] *Nature Methods*
CellTypist	Previous studies focused on cell subtype classification and cell type classification. Building on this foundation, TOGGLE can further distinguish different functional states within the same cell subtype, as well as differentiate the programmed fate stages within the same cell subtype.	C. Domínguez Conde et al.[27] *Science*
SCENIC		Sara Aibar et al.[28] *Nature Methods*
pySCENIC		Bram Van de Sande et al.[29] *Nature Protocols*
Waddington-OT	Cell subtype classification. TOGGLE can perform developmental and functional differentiation, as well as programmatic fate stage distinction within the same subtype.	Geoffrey Schiebinger et al.[14] *Cell*
Markov chain Models (MARGARET)	Used to identify how many developmental outcomes a cell has. TOGGLE not only distinguishes multiple developmental outcomes but also performs subgroup classification, developmental prediction, program fate stage categorization, and differentiation of cells with different functions within the same subtype.	Brian K. Chu[30] *BMC Systems Biology*
Cospar	Used for binary classification of progenitor cells. TOGGLE can be applied to multi-class classification of both progenitor and mature cells.	Wang et al.[15] *Nature Biotechnology*

By utilizing TOGGLE to analyze neuronal data after ischemic stroke (GSE232429), we successfully mapped the trajectory from healthy neurons to those undergoing programmed cell death. Differential gene expression analysis revealed downregulation of *Gpx4* and upregulation of *Trf, Tfrc, Smad1, Smad5, Acsl4, Rela, Bax,* and *Nfkbia* in neurons experiencing programmed cell death compared to healthy counterparts. This trend was validated in animal models of ischemic stroke, further confirming TOGGLE's utility in constructing fate trajectories. In addition, several apoptosis- and ferroptosis-related genes showed expressions inconsistent with previous studies. For example, Aifm2 and *Sirt1* are generally downregulated during ferroptosis[31,32] but were upregulated in dying neurons in the GSE232429 dataset, whereas *Casp3* is upregulated during apoptosis[33], but was downregulated in dying neurons in GSE232429. Our experiments also found that Aifm2 and Sirt1 were downregulated, while Casp3 was upregulated in the MCAO group. We infer that the reason for this phenomenon may be that cells in the early stages of programmed cell death are trying to "self-rescue". Future tests are needed to capture dying cells to confirm this hypothesis. In the future, TOGGLE may provide insights into genetic differences between cells at various stages of programmed cell death, such as apoptosis and necroptosis. This could lead to the identification of novel biomarkers and pathways, potentially guiding the development of targeted intervention drugs aimed at modulating these processes.

Among the forms of programmed cell death, ferroptosis—a process driven by iron-dependent lipid peroxidation—has emerged as particularly relevant[34]. Ferroptosis is now recognized as a key mechanism underlying ischemic stroke, with inhibition of this pathway offering a promising therapeutic approach. However, the incomplete understanding of ferroptosis mechanisms has limited advances in treatment exploration[35]. Our study not only mapped the trajectory of neuronal programmed cell death following ischemic stroke but also identified neuron populations primarily undergoing ferroptosis. We also found some positively expressed ferroptosis-driving genes among the differentially expressed genes, such as *Ctsb, Mtdh, Ndrg1, Smad7, Egr1, Agpat3, Mapk3, Mdm2, Cd82, Lgmn, Elavl1, Nras, Mdm4, Lpin1, Lig3, Idh1, Por, and Acsl1* (Table S2). Targeted intervention of these genes in the future may reduce the number of neuronal deaths and alleviate the sequelae of ischemic stroke. Meanwhile, our prior work demonstrated that elevated iron levels and neuronal iron accumulation in infarcted brain tissuecould be suppressed by deferiprone,[23] which mitigated neuronal death and reversed infarction-induced changes in key ferroptosis-related proteins (Gpx4 and Acsl4). Transmission electron microscopy further revealed ferroptosis-specific neuronal changes in the MCAO group, which were alleviated by deferiprone. These findings strongly indicate the presence of neuronal ferroptosis in ischemic stroke and further validate the reliability of the TOGGLE algorithm. Our work thus offers a new framework and toolset for investigating neuronal ferroptosis mechanisms and potential therapeutic targets. Moving forward, TOGGLE could be applied to map programmed cell death trajectories across various tissues and organs, supporting in-depth exploration of specific programmed cell death mechanisms.

Due to financial constraints and limitations of experimentation, the establishment of a stable and commercially accessible technology for in vivo cell fate prediction remains a challenge. Consequently, investigating the transcriptional diversity of cellular fate presents a difficulty. Nevertheless, our

TOGGLE algorithm offers a computational approach to accomplish fate prediction, thereby significantly reducing the expenses associated with experimentation and facilitating research in developmental biology and cancer biology. Fate-specific changes in gene expression are accompanied by substantial epigenetic remodeling, which enhances the accessibility of the genome to fate-specifying transcription factors[9]. Therefore, the integration of scRNA-seq data and single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) data to construct a multi-modality model is hypothesized to outperform a model solely based on scRNA-seq[36].

In previous multimodal research, it is typically necessary to first perform non-one-hot encoding on the expression data of a single type, followed by employing optimal transport optimization theory to process these combined expression data. This aligns with the type of data TOGGLE is capable of handling, thus TOGGLE can also be applied within various multimodal models. This application is able to enhance the accuracy of non-one-hot encoding and optimize the optimal transport results of combined data, thereby facilitating multimodal cell fate prediction[37].

The limitations of this study is that when the expected number of categories has already been estimated, the desired number of categories can be obtained by adjusting the target number of categories. While the algorithm can automatically identify and provide the estimated number of categories and can reduce resolution to produce more category results, it cannot independently verify the actual number of true categories. Further comparison and calculation of differential genes are required for validation.

4 Methods

4.1 Pipeline introduction

The algorithm automatically assigns a presumed label by itself to the samples and these labels are used for learning. Consequently TOGGLE as an unsupervised algorithm. In initializing the inter-class metrics of cells, it is necessary to consider the phenomenon of metric center offset. When evaluating the similarity distribution S between cell A and surrounding cells with cell A as the reference center, one cannot infer that a similarity distribution like S exists when cell B is the reference center, meaning the similarity between cell A and cell B is typically not equal to the similarity between cell B and cell A. By modeling similarity metrics using the asymmetry of graph distances, it is possible to effectively represent this irreversibility of cell similarity distances. As shown in Figure 13, "1.2 Neighbor Activation", Input a gene expression data firstly, TOGGLE uses the graph diffusion in order to get the graph diffusion matrix(GDM) which has been proven effective successfully. After that, TOGGLE uses learned features to keep the multiple kernels density consistency[38] , as illustrated in Figure 13, "Consistent Density by unsupervised Learning".

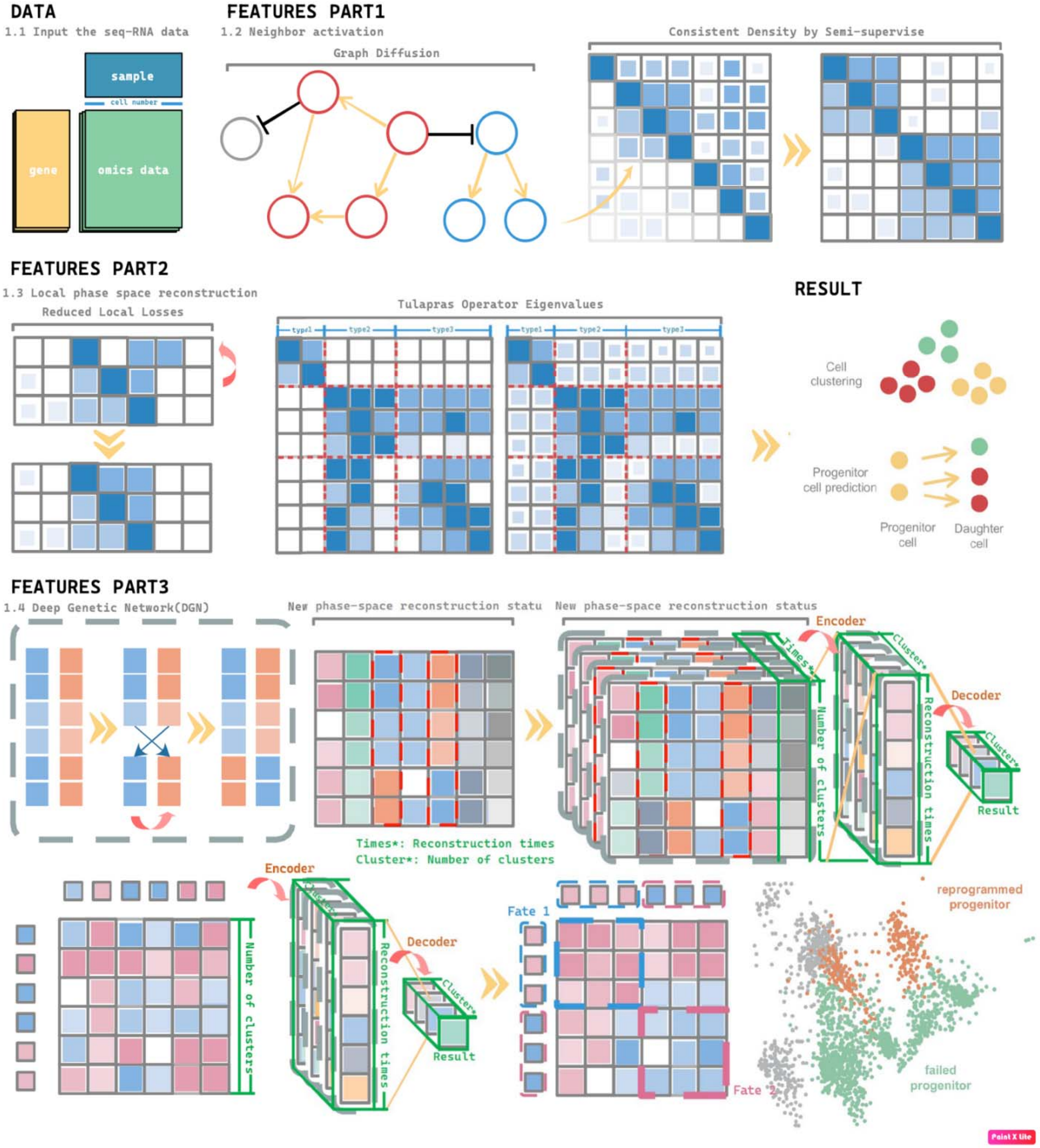
DATA
1.1 Input the seq-RNA data
sample
cell number
gene
omics data
FEATURES PART1
1.2 Neighbor activation
Graph Diffusion
Consistent Density by Semi-supervise
FEATURES PART2
1.3 Local phase space reconstruction
Reduced Local Losses
Tulapras Operator Eigenvalues
type1
type2
type3
RESULT
Cell clustering
Progenitor cell prediction
Progenitor cell
Daughter cell
FEATURES PART3
1.4 Deep Genetic Network(DGN)
New phase-space reconstruction statu
New phase-space reconstruction status
Times*
Number of clusters
Encoder
Cluster*
Reconstruction times
Decoder
Result
Times*: Reconstruction times
Cluster*: Number of clusters
Fate 1
Fate 2
reprogrammed progenitor
failed progenitor
Paint X Ule

Figure 13 Part 1.1: Original high-dimensional data. Part 1.2, left of the yellow double arrow: Use Graph Diffusion to generate a distance matrix. Right of the yellow double arrow: Calculate correlation metrics and perform cell sorting. Part 1.3, left: Sort the compressed high-dimensional data to maximize the density of the effective region, a concept known as phase space reconstruction. Right: Obtain cells with similar transcriptional states. The distance matrix here can be used to identify similarities, also referred to as a similarity matrix. Result: Clusters are obtained using a heuristic algorithm. Part 3, first row: Perform fitting with DGN, encode the results, and overlay them. Second row: Decode the encoded results to obtain the classification of clusters.

Following the completion of similarity measurement, it is necessary to perform sample sorting within clusters and boundary delineation. Unlike traditional algorithms that utilize absolute distances for sample sorting and defining inter-cluster boundaries, TOGGLE does not employ maximal eigenvalues or statistical features in multi-classification problems. Instead, it constructs relative distances using unsupervised principles and an iterative binarization hypothesis. At the start of the iteration, based on the singular value results after decentralization processing of the similarity matrix, the relative correlation index of all samples is calculated. Then, through the iterative process to reduce local losses, as shown in Reduced Local Losses in Figure 13, the binarization results are used as label hypotheses for unsupervised clustering. The initial clusters of this selfsupervised clustering are treated as the environmental response for unsupervised learning. By minimizing the loss function of environmental response, the hypothesis space internally re-differentiates relevance and repeats the iteration process. Using TOGGLE to complete the sorting of clusters within the similarity matrix achieves higher hierarchical clustering accuracy. As a result, cells will be clustered into subgroups, as well as different cells fate heterogeneity were clustered into subgroups with group ID.

In Figure 13-PART2, cells with similar states are clustered together, allowing several subgroups to be identified among progenitor cells, where cells within each subgroup share similar states. However, different subgroups may follow the same fate trajectory; for example, subgroups A and E could both represent early progenitor subpopulations, necessitating their grouping. In Figure 13-PART2, the sparse clustering of individual cells has been transformed into a non-sparse clustering of subgroups. We designed the Deep Genetic Network, an algorithm that maintains density consistency through phase space reconstruction. After each mutation using a genetic algorithm, a result is obtained, and each result is labeled as a code. By representing all possible results with these codes, the complete phase transition process in the feature space is reconstructed. Decoding the results clusters subgroups with similar states, such as grouping early progenitor subgroups A and E together. A heuristic algorithm is then used to determine the boundaries of the subgroups. Additionally, an extra test is conducted by mixing progenitor and progeny cells at different stages for classification, to demonstrate the algorithm's capability to classify mixed cells across categories.

Additionally, an extra test will be conducted by mixing progenitor and progeny cells at different stages for classification, in order to demonstrate the algorithm's ability to classify mixed cells across categories. To validate the performance of TOGGLE, we have applied it to two single-cell datasets: Hematopoiesis and Fibroblast Reprogramming[10,19]. Upon analyzing the data, TOGGLE demonstrated its successful cell fate prediction capability and achieved state-of-the-art performance.

4.2 Use Graph Diffusion to generate distance matrix

For the database, we used the expression data of scRNA-seq as the TOGGLE's input data. The datasets of scRNA-seq are from Gene Expression Omnibus database (accession No. GSE140802, GSE99915 and GSE99915)[9,13,39–42]. For the *simulated annealing* process for boundaries, we used built-in functions in MATLAB, because it's faster to deal with float and double data. All the data sets are available together with the open access database. Because the graph diffusion matrix exhibits asymmetry, the upper and lower halves of the matrix obtained by diagonal splitting contain different information. Moreover, graph diffusion utilizes inter-cellular graph distances for similarity measurement, which incorporates spatial distribution information. Each cell's transcriptional data affects the relative positioning of cells in the graph space, rather than relying on simple sequence alignments to determine similarity.

To obtain an RNA-seq expression matrix (M) of size $X \times Y$, transcriptional results for Y RNA molecules from X cells were retrieved from the NCBI database. Cospar was then employed to compute the cell's graph diffusion matrix, resulting in a non-symmetric graph diffusion matrix of size $X \times X$. This matrix provides information about cell similarity. Extracting the $i-th$ row and $j-th$ column of the matrix yields M_{ij}, representing the one-way graph distance from the $i-th$ cell to the $j-th$ cell. Due to the asymmetry of the scatterplot distances, M_{ji} represents the one-way graph distance from the $j-th$ cell to the $i-th$ cell, typically not equal to M_{ij}.

Based on a common assumption that cells use the similarity matrix as a measure of class when cell populations are distributed around clustering centers denoted by d , clusters D are formed, centered around d. Ideally, the center d_i of cluster D_i is represented on the diagonal of the similarity matrix. Cluster D represents a highly similar group of cells within a subpopulation, generating label assumptions denoted as D' . TOGGLE utilizes these generated label assumptions D' for unsupervised learning.

4.3 Calculate correlation metrics

To generate predictive label assumptions D' , it is necessary to use inter-cluster distances. First, we construct a decentered matrix $\overline{D_\iota}$ for the similarity matrix M, allowing the data to be measured around the assumed virtual origin point.

$$\overline{D_\iota} = D_i - \frac{1}{n}\sum_{j=1}^{n} D_{ij}$$

Where D_i is the vector corresponding to the $i-th$ row of input matrix D, with $1 \leq i \leq m$. We compute the covariance matrix cov_D of $\overline{D}$ to enhance the dissimilarity in the similarity matrix, thereby mapping the original data to a latent space with greater differences.

$$cov\_D = COV(\overline{D})$$

$COV(\overline{D})$ represents the covariance calculation matrix. Subsequently, we utilize the ideal feature matrix V, where the eigenvalues are distributed along the diagonal.

$$V = sparse((diag(cov\_D))^{-\frac{1}{2}})$$

$diag(cov\_D)$ denotes the extraction of diagonal elements from cov_D. The term "$sparse$" indicates the process of sparsifying the matrix. We achieve sparsification using "$sparse$" to obtain the sparsified diagonal matrix$(diag(cov_D))^{-\frac{1}{2}}$. Subsequently, we calculate the reconstruction matrix in the differential space to obtain the covariance similarity matrix pD.

$$pD = V \cdot cov\_D \cdot V^T = \mathrm{g}(D)$$

Where pD represents the enhanced covariance matrix emphasizing dissimilarity. Using the definition of pD, we obtain the latent space mapping result $pM = \mathrm{g}\ (M)$ for the similarity matrix M, represented in the form of the correlation assumption ∂. $¶_n$ represents the $n-th$ row vector in the matrix, with correlation coefficients for sample cells $-1 < ¶_n < 1$. Based on the numerical characteristics of the singular matrix, during the mapping process from the original object M to the mapped object M (M® pM), cells with higher similarity in the original data will acquire closely related correlation coefficients. This effectively separates samples in accordance with the label assumptions, resulting in the formation of the sample set D‘.

In simple terms, using $pM = \mathrm{g}\ (M)$ amplifies the differences within the similarity matrix M, making it easier to classify elements within the matrix.

4.4 Unsupervised perform cell sorting

We proposed the TOGGLE corollary to demonstrate why this algorithm is effective. According to the corollary in Supplementary information-Section 5, using local phase space reconstruction allows for maximum consistency in information density. When applied to this biological problem, its intuitive significance lies in achieving maximum local similarity through the iterative reconstruction process. This ensures that cells in the most similar fate stages are aligned together, thus capturing the maximum amount of information.

The detailed explanation of $loss(¶_n)$ can be found in Supplementary information-Section 6. When directly using correlation coefficients for classification, there is a presence of numerous boundary samples $¶_z$ between different clusters, which makes the boundaries less distinct or prone to being misclassified as independent clusters. Assuming that $\P_d$ is located at the clustering center of cluster D, when the convergence of cluster D within its hypothetical space Sp occurs, there exists a loss function, denoted as $loss(\P_n)$:

$$loss(\P_n) = \lim_{n \circledR d} |\P_n - \mathrm{d}_n| < \varepsilon$$

Where ε represents the theoretical threshold range. Here, d_n represents samples with the same label within cluster D. In an ideal cluster D, it holds that $D\{d_1 = d_2 = \ldots = d_n\}$. When applying binaryization assumptions to the samples using the parameter δ, the mapping process of any sample $\P_1 \sim \P_z$ in the space Sp, as defined in the label assumption set D', is as follows:

$$[\P_1 \sim \P_z, \P_\pi] \circledR [\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \ldots, \mathrm{d}_{n-1}, \mathrm{d}_n] = Sp$$

Where $\P_\pi$ represents non-boundary samples, which do not fall within the theoretically defined boundary vicinity interval $(\P_\pi - \varepsilon, \P_\pi + \varepsilon)$. In this scenario, the minimization of the loss function, denoted as $loss(\mathrm{d}_n)$, is achieved. Based on the convergence properties of the loss function $loss(\mathrm{d}_n)$, it is possible to minimize the inversion number of the space matrix Sp, denoted as $\mathrm{t}(\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \ldots, \mathrm{d}_{n-1}, \mathrm{d}_n)$. Starting from a correlation origin point of 0, clusters can be represented in the minimized form of the inversion number t as D^+ and D^-:

$$D^+ = \{0, {\P^+}_n, {\P^+}_n, \ldots, {\P^+}_{n+\alpha}\}$$

$$D^- = \{{\P^-}_m, {\P^-}_m, \ldots, {\P^-}_{m+\beta}, 0\}$$

Where ${\P^+}_n$ represents samples for which the binary assumption value d_n is positive, denoted as ${\P^+}_n$, and ${\P^-}_m$ m represents samples for which the binary assumption value d_m is negative, denoted as ${\P^-}_m$. Subsequently, TOGGLE can utilize the self-labeled binary assumption labels D^+ and D^- to complete the Markov decision process. The entire process can be represented as $\{D^+, D^-\} = g(D)$, when D satisfies the conditions of environmental response determination:

$$agent(D) = \exists\{\mathrm{t}\,[g(D)] \neq 0\}$$

Once the conditions for unsupervised learning are met, the reward mapping $D \circledR [D^+, D^-]$ is provided to indicate environmental responses. Similarly, when D^- satisfies the condition determined by $agent(D^-)$, a D^- branch is initiated as $\{D'^+, D'^-\} = g(D^-)$, until all samples in D^- are represented as a single mapping $D^- \circledR [D'^-]$. Each mapping process only adheres to the current state dictated by the environment response, ensuring that each iteration possesses a consistent invariance. It solely relies on the current hypothesis space Sp and is independent of the order of samples. In other words, during iterations, randomly shuffling a few samples does not affect the iteration results. Whenever a new branch is generated, the decision-making process $agent(D)$ depends on the inversion number $agent(D', D'', D''', ...)$ encompassing all mapping processes:

$$agent(D', D'', D''', ...) = \exists\{\mathrm{t}[g(D)] \neq 0, \mathrm{t}[g(\{D^{+})] \neq 0, \mathrm{t}[g(D^{-})] \neq 0, ...\}$$

During each iteration of the representation process $\{D^{+}, D^{-}\} = g(D)$, boundaries in the state of the Markov chain are generated. These boundaries, denoted as s, allow us to determine the boundaries of cluster D. As the iterations progress, sample boundaries within D^{+} and D^{-} can be obtained, denoted as s'. The number of boundaries required can be selected based on the resolution requirements.

In simple terms, the covariance similarity matrix pD exhibits square patterns along its diagonal, where the inside of each square has higher similarity values, defining what we refer to as cluster D. We employ the loss function $loss(¶_n)$ to assess the convergence of this cluster. Each cluster D has a central point $¶_d$ with the highest similarity, which corresponds to its true label d_n. When a cell sample $¶_n$ is assigned to cluster D, the cell's assumed label is also marked as d_n. However, $¶_d$ may have multiple potential labels, making it challenging to achieve uniformity within cluster D and potentially leading to the subdivision of cluster D into various smaller categories, making it difficult to measure the loss function value.

However, when $¶_n$ is transformed into binary values, i.e., solely discerning whether the sample $¶_n$ is positively correlated with the cluster's core $¶_d$ or not, if positively correlated, it is assigned to cluster D^{+}; otherwise, it goes to D^{-}. After a single iteration, the results of the iteration, D^{+} and D^{-}, are assumed to be the closest to the true labels. Based on these results, the convergence of samples within D^{+} and D^{-} is continuously evaluated. Taking D^{+} as an example, if not all samples within D^{+} are positively correlated with the center of D^{+}, D^{+} may be further subdivided into D'^{+} and D'^{-} based on their correlations, with D' representing a new iteration of D. Positive correlation between samples and cluster centers indicates convergence, while the presence of non-converging samples necessitates the delineation of boundaries between converging and non-converging samples. When computing samples within D', the knowledge reservoir used consists of D^{+} and D^{-}, which have already been transformed into unsupervised labels from test samples. When the entire matrix converges completely during the iterations, the sorting and segmentation process of the matrix is considered complete.

4.5 Deep Genetic Network

After completing the calculations in the previous chapter, the sorted matrix pM is obtained. The internal boundary of pM is $S = \{s', s'', s''' ... S^{n}{}'\}$, where $S^{n}{}'$ is the boundary generated during the n-th iteration. S is the predictive information obtained through unsupervised learning. After using S to segment pM, a differentiation matrix pM for cell fate stages can be obtained. At this point, the cell similarity matrix pM is transformed into a similarity matrix rM for fate stages. According to the TOGGLE theorem in Supplementary information-Section 5, setting $Y = rM$ allows the conclusion $E(Y_i, \varsigma) \Leftrightarrow E(Y_i)$ to be applied.

Reader need not read the complex formulas in Supplementary information-Section 5. Simply put, the results generated by the genetic algorithm have a certain randomness. After sufficiently iterating the algorithm, the complete distribution of results under fixed conditions can be obtained. nce the spatial distribution of conditions is sufficiently represented, the distribution of results will remain stable. The maximum distribution within the result distribution is taken as the final classification result, which proves the validity of the conclusion $(Y_i, \varsigma) \Leftrightarrow E(Y_i)$, as the maximum information density is achieved at this point.

Define the genetic algorithm as $Genetic$. In this context, a single result produced by the genetic algorithm$Genetic(rM)$ is an observation of the true result under specific conditions ς_i , denoted as $E(rM_i, \varsigma_i)$. $E(rM)$ is the ideal true fitted result. By encoding $E(rM_i, \varsigma_i)$, we can obtain $Ec[E(rM_i, \varsigma_i)] = R_i$. $Count(R_i, \varsigma)$ represents the count of occurrences of R_i under the condition set ς. Decoding the result yields

$$Ed(rM, \varsigma) = arg \underbrace{max}_{\varsigma i \in \varsigma} Count(R_i)。$$

Here Ec represents encoding, Ed represents decoding, and R_i represents a type of fitted result. For visualization purposes, we employ dimensionality reduction algorithms to project data into two or three dimensions, allowing us to intuitively depict the hidden structures within the data.

Subsequently, a heuristic algorithm is used to perform boundary partitioning for $Ed(rM, \varsigma)$. The valid sampling range for $Ed(rM, \varsigma)$ is $Ed(rM, \varsigma) \in (\alpha, \beta)$, with the sampling resolution P. Here set $\beta = \alpha + \gamma P$, resulting in a Bernoulli coding matrix, i.e., the valid information matrix:

$$Ed(rM, \varsigma)_{Boolean} = \begin{cases} Ed(rM, \varsigma) \in (\alpha, \alpha + iP) = 1 \\ Ed(rM, \varsigma) \notin (\alpha, \alpha + iP) = 0 \end{cases} \quad i \in (\alpha, \alpha + \gamma P)$$

The encoded result of $Ed(rM, \varsigma)$ is $E_{code} = \sum_{i=0}^{\gamma P} Ed(rM, \varsigma)_{Boolean}$, Decoding E_{code} yields the boundary $Decode\left[\sum_{i=0}^{\gamma P} Ed(rM, \varsigma)_{Boolean}\right]$, defined as follows:

$$Decode = \begin{cases} E_{code}(j, k) \neq E_{code}(j, k+1) \rightarrow Decode(j, k+1) = Decode(j, k+1) + 1 \\ E_{code}(j, k) \neq E_{code}(j+1, k) \rightarrow Decode(j+1, k) = Decode(j+1, k) + 1 \end{cases}$$

Using $Decode$, the frequency of boundary occurrences can be determined. Simulated annealing optimization is then applied to obtain the boundary. Using the same approach(Sledge & Príncipe, 2018), resulting in an $N \times B$ latent matrix Z, and assign hierarchical clustering labels to each cell. We employ a heuristic approach based on graph Laplacian decomposition to estimate the number of clusters C, with the reduced dimensionality B being equal to the desired number of clusters C, where C is also the aforementioned ranking-constrained parameter.

4.6 scRNA-seq Data Processing

Cells from the GSE232429 dataset[13] with nFeature_RNA > 200 and < 7000, nCount_RNA > 500 and < 25000, and percent.mt < 10% were retained for further analysis. DEsingle was used to perform a preliminary analysis of this dataset to identify differentially expressed genes. Differentially expressed genes were screened out using the criteria of $|\text{Log2FC}| \geq 0.25$ and FDR<0.05. Cellchat was used for cell-cell communication analysis[43].

4.7 Experimental Materials and Methods

4.7.1 Experimental Materials

(1) Experimental Animal

Fifty male specific pathogen free Sprague-Dawley (SD) rats, aged 6 to 8 weeks and weighing 250-280 g, were provided by the Experimental Animal Center of Hunan University of Chinese Medicine (Animal Quality Certificate No.: 430727211100405318). The rats were purchased from Hunan Sileike Jingda Experimental Animal Co., Ltd. [License No.: SCXK (Xiang) 2019-0004]. They were housed in the Experimental Animal Center of Hunan University of Chinese Medicine [License No.: SYXK (Xiang) 2019-0009] under standard conditions, with a room temperature of (24±1)°C, humidity of approximately 55%, and a 12-hour light/dark cycle. After a one-week acclimation period, the animals were used for the experiments. All animals' care and experimental procedures were approved by the Animal Ethics Committee of Hunan University of Chinese Medicine (Approve number: LL2019092009) and were in accordance with the National Institute of Health's Guide for the Care and Use of Laboratory Animals.

(2) Reagents

Paraformaldehyde (Batch No.: G1101) was purchased from Servicebio, and anhydrous ethanol (Batch No.: 10009259) from China National Pharmaceutical Group Chemical Reagent Co., Ltd. Electron microscopy fixative (Cat. No. G1102) was purchased from Wuhan Guge Biotechnology Co., Ltd. Transferrin (Trf) polyclonal antibody (Cat. No.: 17435-1-AP), Bax polyclonal antibody (Cat. No.: 50599-2-Ig) and IκBα polyclonal antibody (Cat. No.: 10268-1-AP) were obtained from Proteintech. Casp-3 antibody (Cat. No.: 9661L) was purchased from Cell Signaling. Sirt1 antibody (Cat. No. ab189494) was purchased from Abcam. Transferrin receptor (Tfrc) polyclonal antibody (Cat. No.: A5865), and Rela (p65) polyclonal antibody (Cat. No.: A2547) were sourced from ABclonal. P-smad (Cat. No.: AF8313) was acquired from Affinity Bioscience. Gpx4 antibody (Cat. No.: AF301385), Acsl4 polyclonal antibody (Cat. No.: AFW6826) and Fsp1 polyclonal antibody (Cat. No.: AFW21808) were purchased from AiFangbio. DAPI (Cat. No.: ab228549) was obtained from Abcam. Deferiprone tablets (specification: 50 mg/tablet, batch number: H20140379) was purchased form Apotex Inc.

(3) Instruments

The optical microscope (Axioscope-5) was purchased from Zeiss, Germany. The pathology microtome (RM2016) and Leica blades (LEICA-819) were both obtained from Leica Instruments Co., Ltd. The optical microscope (Axioscope-5) was purchased from Zeiss Inc. The pathology microtome (RM2016) and Leica blades (LEICA-819) were purchased from Leica Instruments Shanghai Co., Ltd. Transmission electron microscope (HT7700) was purchased from Hitachi Inc.

4.7.2 Animal Grouping, Modeling and Intervention

Fifty SD rats were randomly assigned into three groups: the Sham operation group, the MCAO group and the MCAO+Deferiprone group, with five rats in each group. The MCAO model was established using a modified reversible middle cerebral artery suture method[44] for the rats in the MCAO group and the MCAO+Deferiprone group. Rats were fasted for 12 hours prior to surgery. During surgery, the right common carotid artery and external carotid artery were exposed and clamped, and a suture was inserted into the internal carotid artery via the common carotid artery to block the origin of the right middle cerebral artery. The suture was advanced approximately 18.5±0.5 mm until resistance was felt, indicating the correct position. The suture was then fixed to both the common and internal carotid arteries, and the incision was closed. Throughout the procedure, care was taken to maintain body temperature, and after surgery, the animals were returned to normal feeding conditions. In the Sham group, no suture was inserted, and the animals were kept under normal conditions.

Successful model establishment was confirmed when the rats exhibited left-sided limb paralysis, instability when standing, or circling behavior when lifted by the tail. Neurological function was scored on a scale of 1 to 3, with non-successful models excluded. The neurological scoring system was based on the Zea Longa scale[44] as shown in Table S3.

Deferiprone tablets were dissolved in a 0.9% sodium chloride solution. Following model establishment, rats in the deferiprone group received oral administration of deferiprone at a dose of 125 mg/kg per day. Rats in the sham operation group and the MCAO group were given an equivalent volume of 0.9% sodium chloride solution. Treatments were administered once daily for three consecutive days.

4.7.3 Neuronal morphological changes observation by hematoxylin-eosin (HE) staining

The brain tissue was perfused with pre-cooled 0.9% sodium chloride solution, followed by fixation in 4% paraformaldehyde at 4°C. The tissue was then coronally sectioned into 4 μm thick slices. The sections were deparaffinized, rehydrated, and stained with hematoxylin for 5 minutes. After differentiation and rinsing with water, the sections were blued with distilled water, dehydrated through a graded ethanol series, and counterstained with eosin for 5 minutes. Following dehydration with ethanol and clearing with xylene, coverslips were applied and sealed with neutral balsam. Morphological observations and image acquisition were performed under a light microscope.

4.7.4 Transmission electron microscopy

The affected brain tissue was sectioned into 1 mm-thick slices. Samples were immersed in electron microscopy fixative at 4°C overnight, followed by fixation in 1% osmium tetroxide at 4°C for 2 hours. The samples were stained with 1% uranyl acetate for 1 hour, dehydrated, and embedded in epoxy resin. Ultrathin sections were prepared using an ultramicrotome, stained with uranyl acetate and lead citrate, and examined under a transmission electron microscope.

4.7.5 Immunohistochemical Staining for Acsl4, Gpx4 and Fsp1 Protein Expression Levels

Paraffin-embedded tissue sections (5 μm) were deparaffinized and rehydrated. Endogenous enzymes were inactivated by incubating the sections in 3% hydrogen peroxide at room temperature for 10 minutes, followed by two washes with distilled water. Antigen retrieval was performed by immersing the sections in 0.01 mol/L citrate buffer (pH 6.0) and heating in a microwave until boiling, then turning off the power and repeating the process after a 5-minute interval. After cooling, the sections were washed twice with PBS buffer (pH 7.2-7.6) for 3 minutes each. Blocking solution was added and incubated at room temperature for 20 minutes, with excess liquid removed without washing. Primary antibodies were applied and incubated at 37°C for 2 hours, followed by three washes with PBS buffer (pH 7.2-7.6) for 3 minutes each. Biotinylated secondary antibody solution was added and incubated at 20-37°C for 20 minutes, followed by another round of PBS washing. Streptavidin-HRP was applied and incubated at 20-37°C for 20 minutes, followed by washing in PBS as described earlier. DAB was used for color development, with reaction time controlled under a microscope (5-30 minutes). Sections were lightly counterstained with hematoxylin, dehydrated, cleared, and mounted with neutral balsam for microscopic observation of Acsl4 and Gpx4 protein expression. Images from five randomly selected fields of view from different samples were captured, and average optical density was analyzed using ImageJ software.

4.7.6 Immunofluorescence Staining for Bax, IκBα, Casp3, p-Smad, p65 (Rela), Tfrc, Trf, and Sirt1 Protein Expression Levels

Paraffin sections were deparaffinized and rehydrated, followed by membrane permeabilization with 0.3% Triton X-100 for 5 minutes. The sections were then blocked with 5% BSA in PBST at room temperature for 1 hour. Primary antibodies, diluted in PBST, were applied, and the sections were incubated overnight at 4°C. After washing with PBST, fluorescent secondary antibodies were added, and the sections were incubated in the dark at room temperature for 1 hour. The secondary antibodies were washed off, and the slides were mounted with anti-fade mounting medium containing DAPI. The next day, images were captured using a panoramic tissue cell quantification system at 40x magnification. Images from five randomly selected fields of view from different samples were captured and analyzed using ImageJ software to assess the average fluorescence intensity.

4.8 Statistical Analysis

All data are presented as Mean±SD, and statistical analysis and graphing were performed using GraphPad Prism 9.0. Non-parametric tests were used to analyze average fluorescence intensity and average optical density. A P-value of less than 0.05 was considered statistically significant.

Code and Data availability

Contribution

Acknowledgments

“***This research was supported in part by the U.S. National Science Foundation (NSF) under grant CNS***1828521 ***and the University of Louisville’s Research Computing team***.” *The University of Louisville provided computational equipment support for this research and assisted all researchers with equipment maintenance and update. These hardworking and dedicated staff members share this honor with us, including the following staff*: ***Harrison Simrall***, ***Tomas Felipe Llano Rios***, and ***Randy Vargas***.

References

---

1. Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., Andrews, D. W., Annicchiarico-Petruzzelli, M., Antonov, A. V., Arama, E., Baehrecke, E. H., Barlev, N. A., Bazan, N. G., Bernassola, F., Bertrand, M. J. M., Bianchi, K., Blagosklonny, M. V., Blomgren, K., Borner, C., Boya, P., Brenner, C., Campanella, M., Candi, E., Carmona-Gutierrez, D., Cecconi, F., Chan, F. K.-M., Chandel, N. S., Cheng, E. H., Chipuk, J. E., Cidlowski, J. A., Ciechanover, A., Cohen, G. M., Conrad, M., Cubillos-Ruiz, J. R., Czabotar, P. E., D’Angiolella, V., Dawson, T. M., Dawson, V. L., De Laurenzi, V., De Maria, R., Debatin, K.-M.,

DeBerardinis, R. J., Deshmukh, M., Di Daniele, N., Di Virgilio, F., Dixit, V. M., Dixon, S. J., Duckett, C. S., Dynlacht, B. D., El-Deiry, W. S., Elrod, J. W., Fimia, G. M., Fulda, S., García-Sáez, A. J., Garg, A. D., Garrido, C., Gavathiotis, E., Golstein, P., Gottlieb, E., Green, D. R., Greene, L. A., Gronemeyer, H., Gross, A., Hajnoczky, G., Hardwick, J. M., Harris, I. S., Hengartner, M. O., Hetz, C., Ichijo, H., Jäättelä, M., Joseph, B., Jost, P. J., Juin, P. P., Kaiser, W. J., Karin, M., Kaufmann, T., Kepp, O., Kimchi, A., Kitsis, R. N., Klionsky, D. J., Knight, R. A., Kumar, S., Lee, S. W., Lemasters, J. J., Levine, B., Linkermann, A., Lipton, S. A., Lockshin, R. A., López-Otín, C., Lowe, S. W., Luedde, T., Lugli, E., MacFarlane, M., Madeo, F., Malewicz, M., Malorni, W., Manic, G., Marine, J.-C., Martin, S. J., Martinou, J.-C., Medema, J. P., Mehlen, P., Meier, P., Melino, S., Miao, E. A., Molkentin, J. D., Moll, U. M., Muñoz-Pinedo, C., Nagata, S., Nuñez, G., Oberst, A., Oren, M., Overholtzer, M., Pagano, M., Panaretakis, T., Pasparakis, M., Penninger, J. M., Pereira, D. M., Pervaiz, S., Peter, M. E., Piacentini, M., Pinton, P., Prehn, J. H. M., Puthalakath, H., Rabinovich, G. A., Rehm, M., Rizzuto, R., Rodrigues, C. M. P., Rubinsztein, D. C., Rudel, T., Ryan, K. M., Sayan, E., Scorrano, L., Shao, F., Shi, Y., Silke, J., Simon, H.-U., Sistigu, A., Stockwell, B. R., Strasser, A., Szabadkai, G., Tait, S. W. G., Tang, D., Tavernarakis, N., Thorburn, A., Tsujimoto, Y., Turk, B., Vanden Berghe, T., Vandenabeele, P., Vander Heiden, M. G., Villunger, A., Virgin, H. W., Vousden, K. H., Vucic, D., Wagner, E. F., Walczak, H., Wallach, D., Wang, Y., Wells, J. A., Wood, W., Yuan, J., Zakeri, Z., Zhivotovsky, B., Zitvogel, L., Melino, G. & Kroemer, G. Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ* **25,** 486–541 (2018).

2. Moujalled, D., Strasser, A. & Liddell, J. R. Molecular mechanisms of cell death in neurological diseases. *Cell Death Differ* **28,** 2029–2044 (2021).
3. Qin, H., Abulaiti, A., Maimaiti, A., Abulaiti, Z., Fan, G., Aili, Y., Ji, W., Wang, Z. & Wang, Y. Integrated machine learning survival framework develops a prognostic model based on inter-crosstalk definition of mitochondrial function and cell death patterns in a large multicenter cohort for lower-grade glioma. *Journal of Translational Medicine* **21,** 588 (2023).
4. Zhang, Q., Jia, M., Wang, Y., Wang, Q. & Wu, J. Cell Death Mechanisms in Cerebral Ischemia–Reperfusion Injury. *Neurochem Res* **47,** 3525–3542 (2022).
5. Wang, X., Chen, X., Zhou, W., Men, H., Bao, T., Sun, Y., Wang, Q., Tan, Y., Keller, B. B., Tong, Q., Zheng, Y. & Cai, L. Ferroptosis is essential for diabetic cardiomyopathy and is prevented by sulforaphane via AMPK/NRF2 pathways. *Acta Pharm Sin B* **12,** 708–722 (2022).
6. Zeng, L., Yang, K., Yu, G., Hao, W., Zhu, X., Ge, A., Chen, J. & Sun, L. Advances in research on immunocyte iron metabolism, ferroptosis, and their regulatory roles in autoimmune and autoinflammatory diseases. *Cell Death Dis* **15,** 481 (2024).
7. Liao, M., Qin, R., Huang, W., Zhu, H.-P., Peng, F., Han, B. & Liu, B. Targeting regulated cell death (RCD) with small-molecule compounds in triple-negative breast cancer: a revisited perspective from molecular mechanisms to targeted therapies. *J Hematol Oncol* **15,** 44 (2022).
8. Qin, R., You, F.-M., Zhao, Q., Xie, X., Peng, C., Zhan, G. & Han, B. Naturally derived indole alkaloids targeting regulated cell death (RCD) for cancer therapy: from molecular mechanisms to potential therapeutic targets. *J Hematol Oncol* **15,** 133 (2022).
9. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367,** eaaw3381 (2020).
10. Moreno-Ayala, R. & Junker, J. P. Single-cell genomics to study developmental cell fate decisions in zebrafish. *Brief Funct Genomics* elab018 (2021). doi:10.1093/bfgp/elab018
11. Ma, H., Zhou, Y., Li, Z., Zhu, L., Li, H., Zhang, G., Wang, J., Gong, H., Xu, D., Hua, W., Liu, P., Zhang, X., Zhang, Y., Zhang, L., Hong, B., Zhou, W., Yang, P. & Liu, J. Single-Cell RNA-Sequencing Analyses Revealed Heterogeneity and Dynamic Changes of Metabolic Pathways in Astrocytes at the Acute Phase of Ischemic Stroke. *Oxid Med Cell Longev* **2022,** 1817721 (2022).

12. Liu, T., Bai, M., Liu, M., Li, T., Liao, Y., Zhao, C., Yao, M., Wang, J., Wen, A. & Ding, Y. Novel synergistic mechanism of 11-keto-β-boswellic acid and Z-Guggulsterone on ischemic stroke revealed by single-cell transcriptomics. *Pharmacol Res* **193,** 106803 (2023).
13. Nakamura, A., Sakai, S., Taketomi, Y., Tsuyama, J., Miki, Y., Hara, Y., Arai, N., Sugiura, Y., Kawaji, H., Murakami, M. & Shichita, T. PLA2G2E-mediated lipid metabolism triggers brain-autonomous neural repair after ischemic stroke. *Neuron* **111,** 2995-3010.e9 (2023).
14. Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A. & Lander, E. S. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176,** 928-943.e22 (2019).
15. Wang, S.-W., Herriges, M. J., Hurley, K., Kotton, D. N. & Klein, A. M. CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nat Biotechnol* **40,** 1066–1074 (2022).
16. Obeng, E. Apoptosis (programmed cell death) and its signals - A review. *Braz. J. Biol.* **81,** 1133–1143 (2021).
17. Bedoui, S., Herold, M. J. & Strasser, A. Emerging connectivity of programmed cell death pathways and its physiological implications. *Nat Rev Mol Cell Biol* **21,** 678–695 (2020).
18. Tuo, Q., Zhang, S. & Lei, P. Mechanisms of neuronal cell death in ischemic stroke and their therapeutic implications. *Medicinal Research Reviews* **42,** 259–305 (2022).
19. Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T. & Morris, S. A. Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564,** 219–224 (2018).
20. Neczypor, E. W., Saldaña, T. A., Mears, M. J., Aslaner, D. M., Escobar, Y.-N. H., Gorr, M. W. & Wold, L. E. e-Cigarette Aerosol Reduces Left Ventricular Function in Adolescent Mice. *Circulation* **145,** 868–870 (2022).
21. Zhou, N., Yuan, X., Du, Q., Zhang, Z., Shi, X., Bao, J., Ning, Y. & Peng, L. FerrDb V2: update of the manually curated database of ferroptosis regulators and ferroptosis-disease associations. *Nucleic Acids Research* **51,** D571–D582 (2023).
22. Kanehisa, M., Sato, Y. & Kawashima, M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci* **31,** 47–53 (2022).
23. Lan, B., Ge, J.-W., Cheng, S.-W., Zheng, X.-L., Liao, J., He, C., Rao, Z.-Q. & Wang, G.-Z. Extract of Naotaifang, a compound Chinese herbal medicine, protects neuron ferroptosis induced by acute cerebral ischemia in rats. *J Integr Med* **18,** 344–350 (2020).
24. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H. B., Pe’er, D. & Theis, F. J. CellRank for directed single-cell fate mapping. *Nat Methods* **19,** 159–170 (2022).
25. Bannon, D., Moen, E., Schwartz, M., Borba, E., Kudo, T., Greenwald, N., Vijayakumar, V., Chang, B., Pao, E., Osterman, E., Graf, W. & Van Valen, D. DeepCell Kiosk: scaling deep learning–enabled cellular image analysis with Kubernetes. *Nat Methods* **18,** 43–45 (2021).
26. Lotfollahi, M., Wolf, F. A. & Theis, F. J. scGen predicts single-cell perturbation responses. *Nat Methods* **16,** 715–721 (2019).
27. Domínguez Conde, C., Xu, C., Jarvis, L. B., Rainbow, D. B., Wells, S. B., Gomes, T., Howlett, S. K., Suchanek, O., Polanski, K., King, H. W., Mamanova, L., Huang, N., Szabo, P. A., Richardson, L., Bolt, L., Fasouli, E. S., Mahbubani, K. T., Prete, M., Tuck, L., Richoz, N., Tuong, Z. K., Campos, L., Mousa, H. S., Needham, E. J., Pritchard, S., Li, T., Elmentaite, R., Park, J., Rahmani, E., Chen, D., Menon, D. K., Bayraktar, O. A., James, L. K., Meyer, K. B., Yosef, N., Clatworthy, M. R., Sims, P. A., Farber, D. L., Saeb-Parsy, K., Jones, J. L. & Teichmann, S. A. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376,** eabl5197 (2022).

28. Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., Van Den Oord, J., Atak, Z. K., Wouters, J. & Aerts, S. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14,** 1083–1086 (2017).

29. Van De Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., Verbeiren, T., De Maeyer, D., Reumers, J., Saeys, Y. & Aerts, S. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15,** 2247–2276 (2020).

30. Chu, B. K., Tse, M. J., Sato, R. R. & Read, E. L. Markov State Models of gene regulatory networks. *BMC Syst Biol* **11,** 14 (2017).

31. Shi, H., Song, L., Wu, Y., Shen, R., Zhang, C., Liao, X., Wang, Q. & Zhu, J. Edaravone Alleviates Traumatic Brain Injury by Inhibition of Ferroptosis via FSP1 Pathway. *Mol Neurobiol* (2024). doi:10.1007/s12035-024-04216-2

32. Wang, X., Shao, N., Zhang, X., Chen, H., Chang, Z., Xie, D. & Zhang, J. Ferulic Acid Activates SIRT1-Mediated Ferroptosis Signaling Pathway to Improve Cognition Dysfunction in Wilson's Disease. *Neuropsychiatric Disease and Treatment* **19,** 2681–2696 (2023).

33. Liu, G., Wang, T., Wang, T., Song, J. & Zhou, Z. Effects of apoptosis□related proteins caspase□3, Bax and Bcl□2 on cerebral ischemia rats. *Biomedical Reports* **1,** 861–867 (2013).

34. Zhou, Y., Liao, J., Mei, Z., Liu, X. & Ge, J. Insight into Crosstalk between Ferroptosis and Necroptosis: Novel Therapeutics in Ischemic Stroke. *Oxidative Medicine and Cellular Longevity* **2021,** 9991001 (2021).

35. Yang, K., Zeng, L., Yuan, X., Wang, S., Ge, A., Xu, H., Zeng, J. & Ge, J. The mechanism of ferroptosis regulating oxidative stress in ischemic stroke and the regulation mechanism of natural pharmacological active components. *Biomedicine & Pharmacotherapy* **154,** 113611 (2022).

36. Jindal, K., Adil, M. T., Yamaguchi, N., Yang, X., Wang, H. C., Kamimoto, K., Rivera-Gonzalez, G. C. & Morris, S. A. Single-cell lineage capture across genomic modalities with CellTag-multi reveals fate-specific gene regulatory changes. *Nat Biotechnol* **42,** 946–959 (2024).

37. Sha, Y., Qiu, Y., Zhou, P. & Nie, Q. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nat Mach Intell* **6,** 25–39 (2023).

38. Wang, B., Zhu, J., Pierson, E., Ramazzotti, D. & Batzoglou, S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat Methods* **14,** 414–416 (2017).

39. Jiang, J., Ye, X., Kong, Y., Guo, C., Zhang, M., Cao, F., Zhang, Y. & Pei, W. scLTdb: a comprehensive single-cell lineage tracing database. *Nucleic Acids Research* gkae913 (2024). doi:10.1093/nar/gkae913

40. Jiang, J., Li, J., Huang, S., Jiang, F., Liang, Y., Xu, X. & Wang, J. CACIMAR: cross-species analysis of cell identities, markers, regulations, and interactions using single-cell RNA sequencing data. *Briefings in Bioinformatics* **25,** bbae283 (2024).

41. Jiang, J., Lyu, P., Li, J., Huang, S., Tao, J., Blackshaw, S., Qian, J. & Wang, J. IReNA: Integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles. *iScience* **25,** 105359 (2022).

42. Kong, Y., Jiang, J., Kong, W. & Qin, S. DRCTdb: disease-related cell type analysis to decode cell type effect and underlying regulatory mechanisms. *Commun Biol* **7,** 1205 (2024).

43. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat Protoc* (2024). doi:10.1038/s41596-024-01045-4

44. Longa, E. Z., Weinstein, P. R., Carlson, S. & Cummins, R. Reversible middle cerebral artery occlusion without craniectomy in rats. *Stroke* **20,** 84–91 (1989).