

Human Prefrontal Cortex Development Data

CSIC 5011 Final Project

HUANG Xinrui^{1*}, ZENG Yeqin^{1*}, GU Yanwu^{1*}

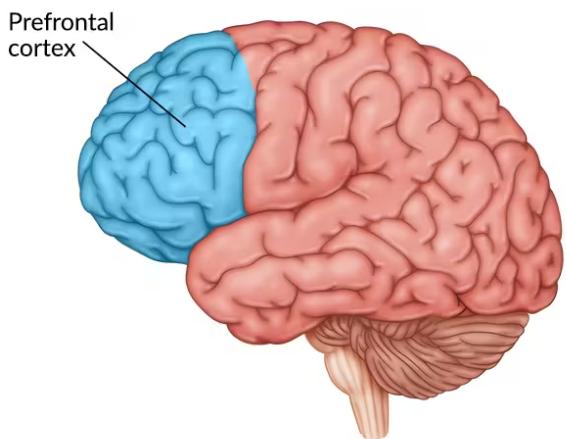
¹: Department of Mathematics, HKUST

^{*}: Equal Contribution



Introduction

- The prefrontal cortex (PFC), located at the front of the frontal lobe in mammals, plays a crucial role in memory, emotion, cognitive behavior, decision-making, and social behavior [1,2]. **Identifying cell types** and their **developmental features** has been challenging yet important.
- We analyzed over 24,153 genes of 2,309 single cells from the developing human prefrontal cortex (gestational weeks 8-26) using RNA sequencing provided by [3].



Key questions

- Do different cell types exhibit **distinct patterns**? What is the **optimal method** for visualizing data projections of human PFCs?
- How can different **subgroups** within a main cell type be identified?
- How can the **developmental trajectories** across different gestational weeks (GWs) be traced? Which key genes play a significant role in regulating developmental trajectories? What are the differences between various analytical approaches.

Methodology

- We used six different topological methods including **UMAP**[4], **T-SNE**[5], **PCA**[6], **MDS**[7], **ISOMAP**[8], **LLE**[9] to classify 2,300 cells from six main cell type to 27 subgroups and compared their effects. We also visualized the expression of marker genes of different samples.
- **Monocle**[10] and **Monocle3**[11] were two tools for analyzing single-cell RNA sequencing data, primarily used for trajectory inference, analyzing dynamic changes in cell states, and constructing cell lineages, which we used to analysis the developmental trajectories of PFC and then compared.

Main Result

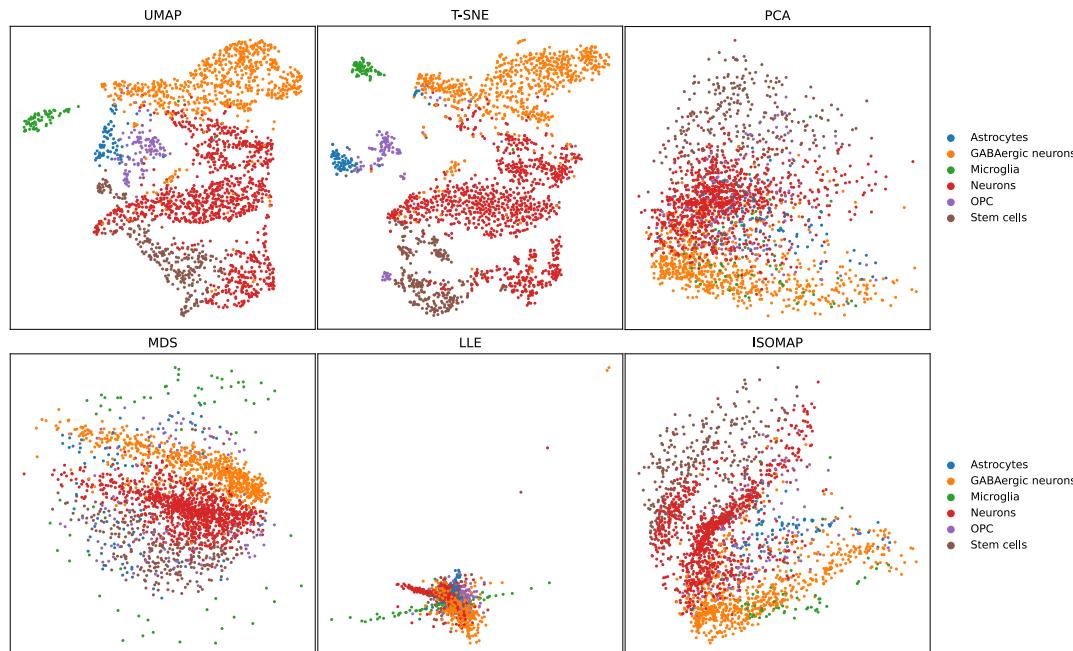


Figure 1: The comparison of different topological methods w.r.t
six main cell types[3]

- Figure 1 compares different dimension reduction methods. We observed that **UMAP and T-SNE** outperform the other methods, as they clearly **separate distinct cell types**, unlike other approaches where the cell types are more intermixed.

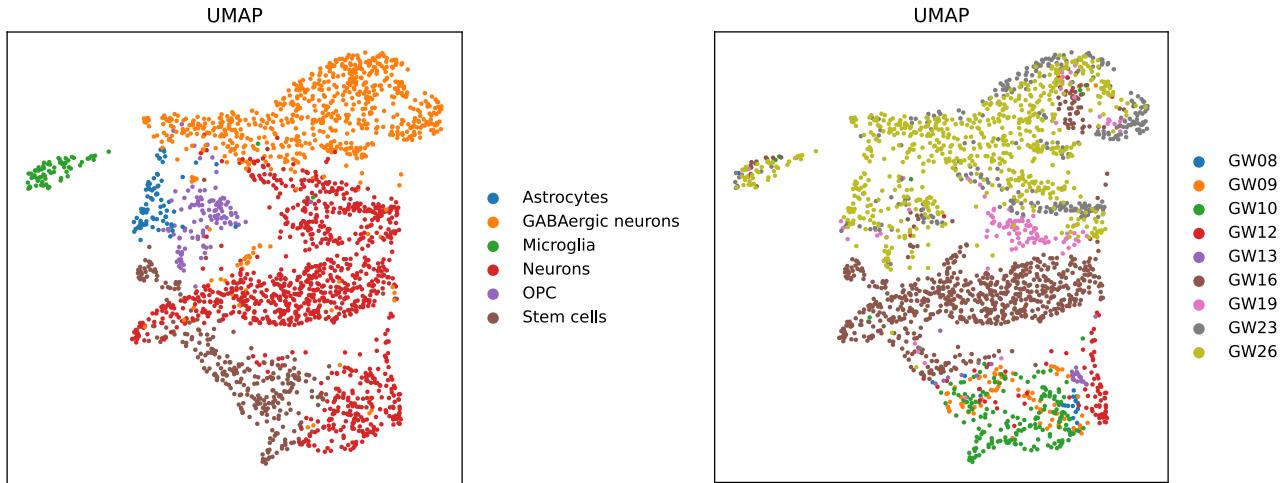


Figure 1(a): UMAP of cells w.r.t main cell types

Figure 2: UMAP of cells w.r.t. GWs

- Based on these findings, we selected UMAP to perform dimensionality reduction for further analysis. Figure 2 shows the UMAP embedding **colored by developmental stages**, revealing that the proportions of cell types vary significantly across different gestational weeks, with each stage dominated by distinct cell types.

Figure 3 illustrates the expression patterns of six marker genes, showing that regions with high expression correspond precisely to the specific cell types identified in Figure 1. These results confirm that the selected marker genes are uniquely and highly expressed in their respective cell types, making them reliable identifiers for cell type classification.

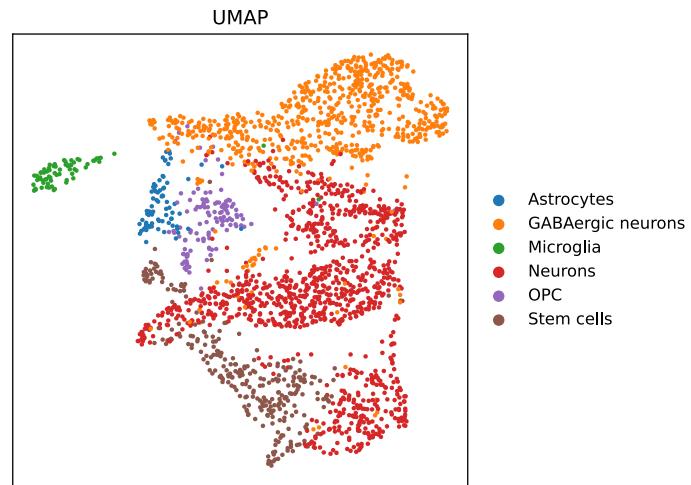


Figure 1(a): UMAP of cells w.r.y main cell types

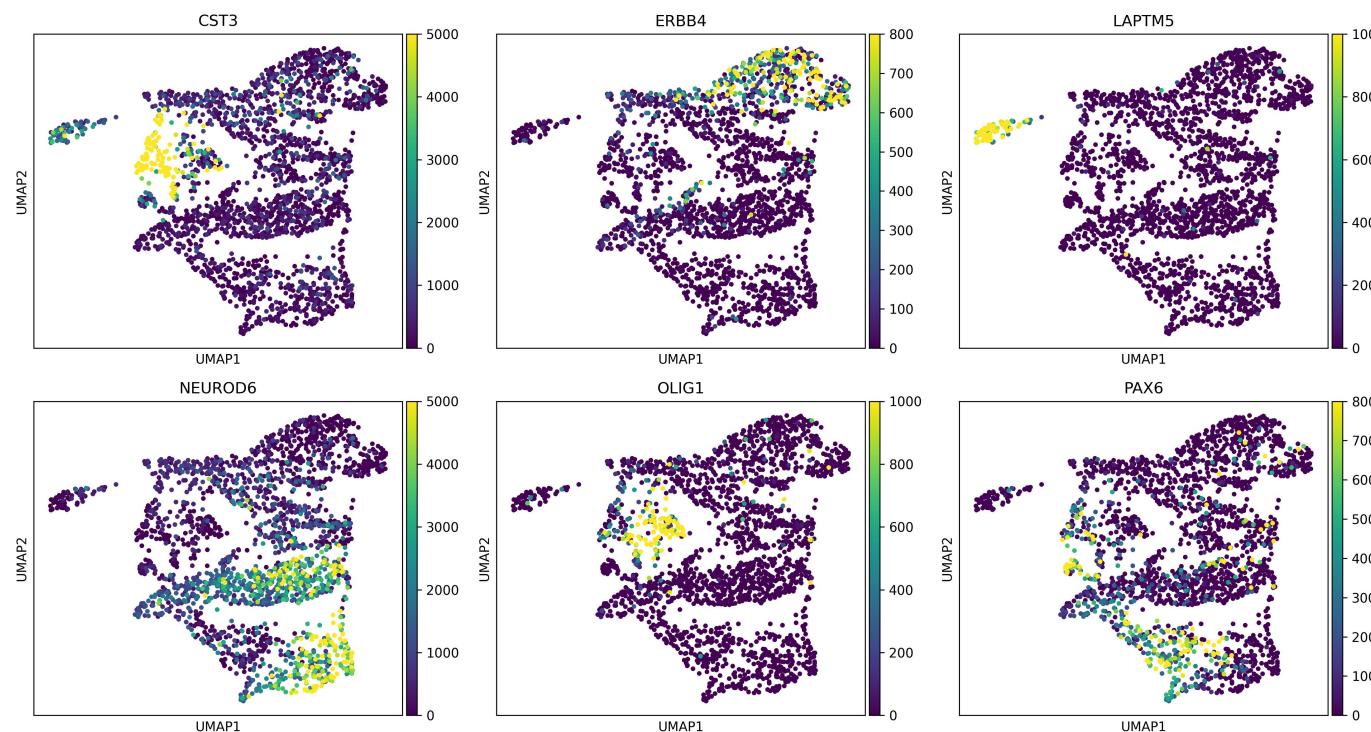


Figure 3: Gene expressions of 6 marker gene under UMAP

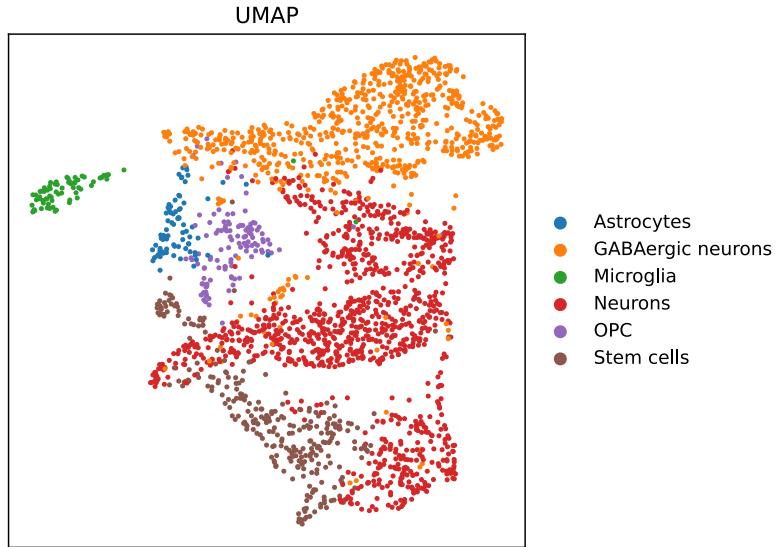


Figure 1(a): UMAP of cells w.r.y main cell types

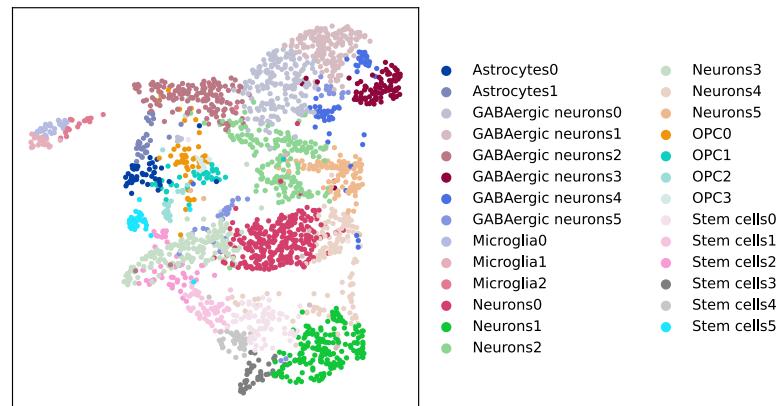


Figure 4: Subtype result of UMAP

Figure 4 presents the UMAP embedding, now colored to reflect the clustering of cells into 27 subtypes derived from the 6 main cell types using Louvain algorithm, enabling further exploration and experimental validation.

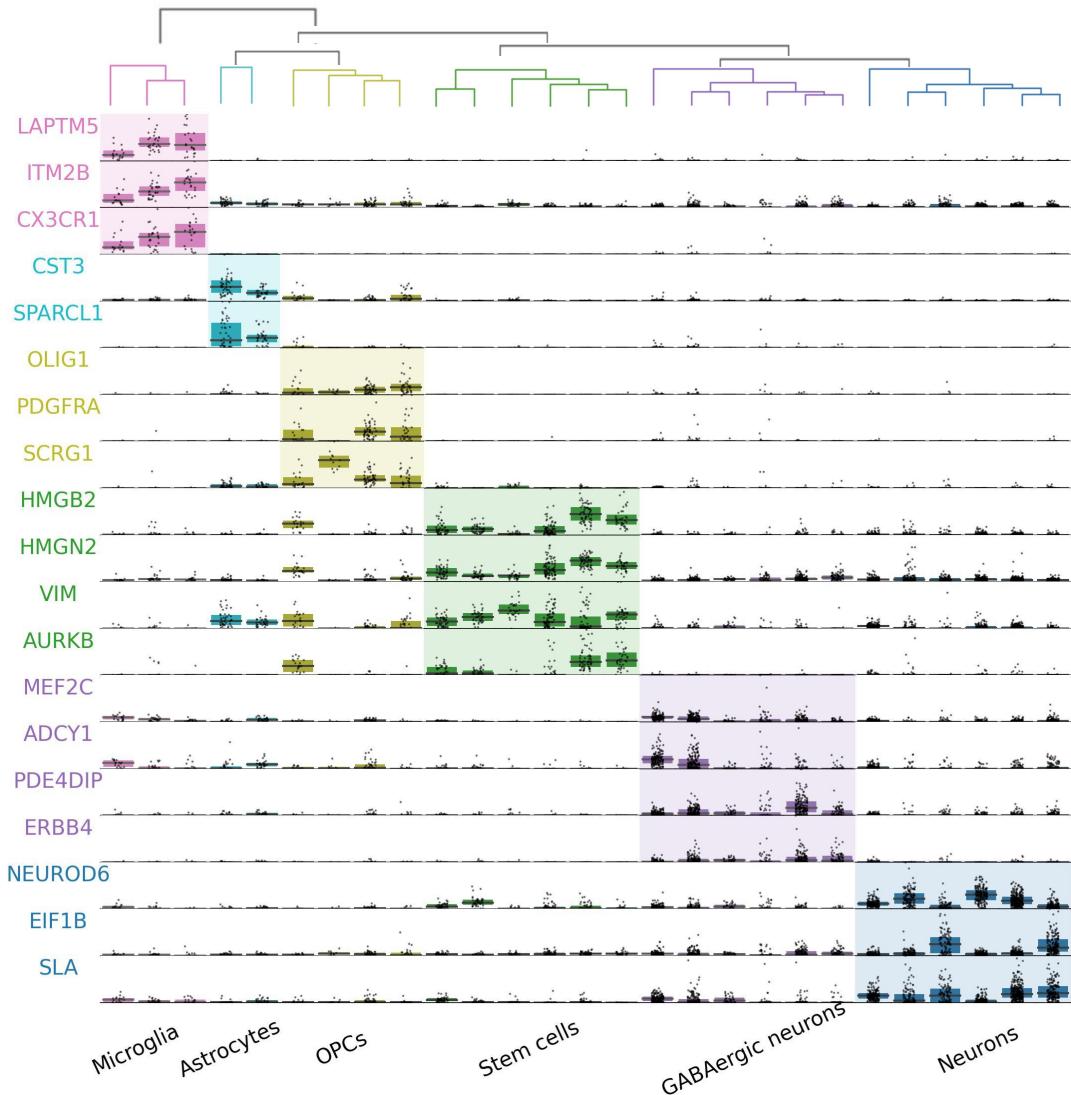


Figure 5: Hierarchical clustering analysis of 27 subtypes

Following the results presented in Figure 4, Figure 5 illustrates the hierarchical clustering analysis of the corresponding 27 subtypes, segregated from each main type. This analysis demonstrates that the expression patterns of the marker genes not only serve to effectively distinguish main cell types, as reflected in the distinct clustering of subtypes within each main type, but also reveal substantial differences in gene expression across subtypes, highlighting the heterogeneity within each group.

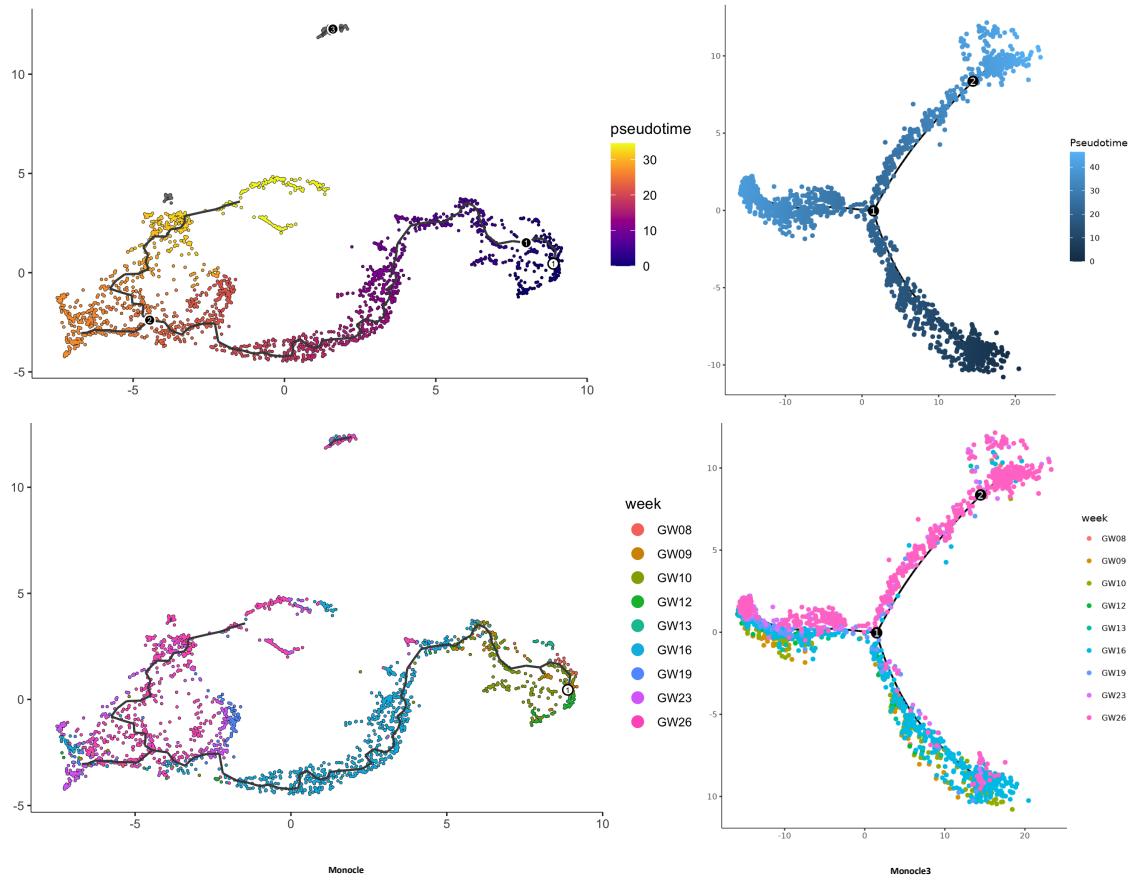


Figure 6. Comparison of developmental trajectories along GWs using Monocle and Monocle3

Figure 6 illustrates the single-cell trajectories analyzed using **Monocle** and **Monocle3**. Both methods effectively represent the progression in pseudo-time along the trajectory. However, when comparing to actual weeks, a significant number of samples were misclassified using **Monocle**, whereas **Monocle3** demonstrated improved accuracy in alignment. Despite this discrepancy, both methods maintained a coherent trend throughout the analysis.

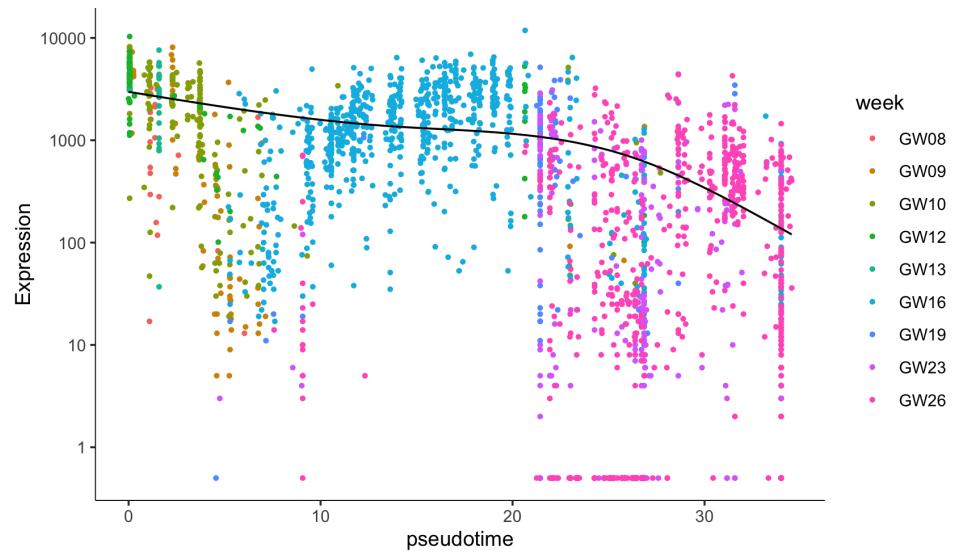


Figure 7. Expression of NEUROD6 along weeks and pseudo-time

In Figure 7, we specifically studied the expression of **NEUROD6** along the weeks and pesudotime using Monocle3, which is one of the most significant differentially expressed genes related to PFC development, and it turns out that the gene expression exhibits a strong correlation with pesudotime and weeks .

Conclusion

- In this project, we applied dimension reduction to analyze and visualize scRNA-seq data.
- It turns out that **UMAP** effectively preserved distinct cell types. Main cell types were further refined into **27 subtypes**, with their marker genes identified, providing insights into cellular heterogeneity.
- Different methods were employed to analyze the developmental trajectories of single cells, and **Monocle3** was identified to outperform other trajectory inference methods.
- Several genes were identified to be the **prediction markers** of the pseudo-time of the PFCs, further downstream analyses, such as pathway enrichment analysis, can be conducted to explore the roles these genes play in the biological processes associated with PFC development.

Reference

- [1] Roth, G. & Dicke, U. Evolution of the brain and intelligence in primates. *Prog. Brain Res.* 195, 413–430 (2012).
- [2] O'Rahilly, R. & Müller, F. Significant features in the early prenatal development of the human brain. *Ann. Anat.* 190, 105–118 (2008).
- [3] Zhong, S., Zhang, S., Fan, X. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex, *Nature* 555, 524-528 (2018). <https://doi.org/10.1038/nature25980>.
- [4] McInnes et al., (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861, <https://doi.org/10.21105/joss.00861>
- [5] Laurens, M. & Geoffrey H. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605, <http://jmlr.org/papers/v9/vandermaaten08a.html>
- [6] Maćkiewicz A, Ratajczak W. Principal components analysis (PCA)[J]. *Computers & Geosciences*, 1993, 19(3): 303-342.
- [7] Kruskal J B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis[J]. *Psychometrika*, 1964, 29(1): 1-27.
- [8] Tenenbaum, J. B., Silva, V. D., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500), 2319-2323.
- [9] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500), 2323-2326.
- [10] Perešíni, Peter, Maciej Kužniar, and Dejan Kostić. "Monocle: Dynamic, fine-grained data plane monitoring." *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. 2015.
- [11] Cao, Junyue, et al. "The single-cell transcriptional landscape of mammalian organogenesis." *Nature* 566.7745 (2019): 496-502.

Contribution

- This project was a collaborative effort with equal contributions from everyone.
- HUANG Xinrui primarily conducted the clustering and subgrouping analyses.
- ZENG Yeqing focused on the developmental trajectory analysis.
- GU Yanwu handled the remaining tasks and organized the poster.

**Thank you for
your listening!**