

Human Prefrontal Cortex Development Data CSIC 5011 Final Project

 HUANG Xinrui¹, ZENG Yeqin¹, 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 [31]. Some key research questions emerged:

- 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] for dimensionality reduction and visualization, and utilized the Louvain algorithm to classify 2,309 cells from 6 main cell types into 27 subgroups.

Monocle [10] and Monocle3 [11] are tools for analyzing scRNA-seq data, mainly used for trajectory inference, studying dynamic changes in cell states, and constructing cell lineages. We used these tools to analyze the developmental trajectories of the PFC and compared the results.

Main Result

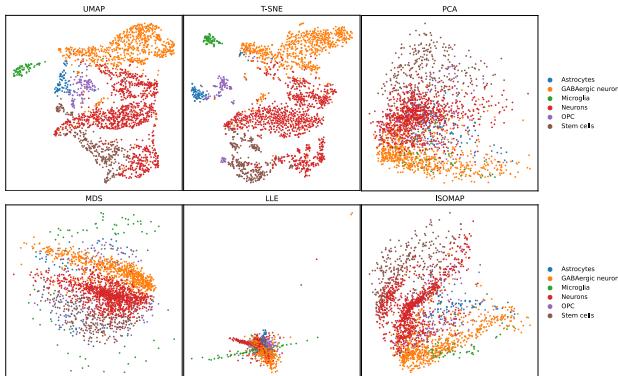


Figure 1: The comparison of different topological methods w.r.t six main cell types[3]

Figure 1 compares different dimension reduction methods. UMAP and T-SNE clearly outperform the other methods by better separating distinct cell types. Based on this, we used the result of UMAP in subsequent analyses. Figure 2 shows UMAP embeddings colored by developmental stages, revealing significant shifts in cell type proportions across gestational weeks, with each stage dominated by specific cell types. Figure 3 highlights the expression of six marker genes, showing that regions with high expression align closely with distinct cell types identified in Figure 1. This confirms that these marker genes are uniquely and strongly expressed in their respective cell types, making them effective markers for classification.

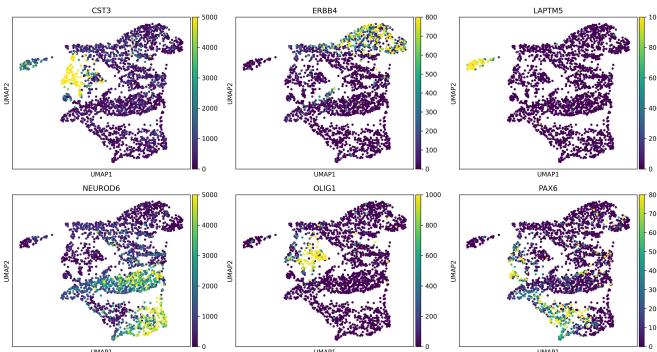


Figure 3: Gene expression of 6 marker genes under UMAP

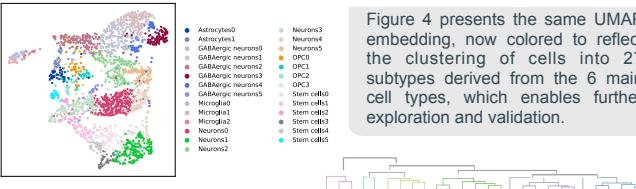


Figure 4 presents the same UMAP embedding, now colored to reflect the clustering of cells into 27 subtypes derived from the 6 main cell types, which enables further exploration and validation.

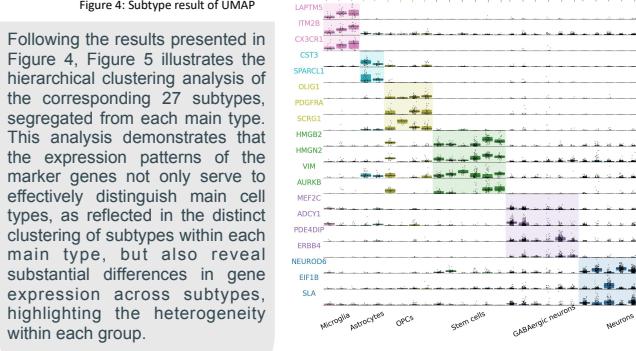


Figure 5: Hierarchical clustering analysis of 27 subtypes

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.

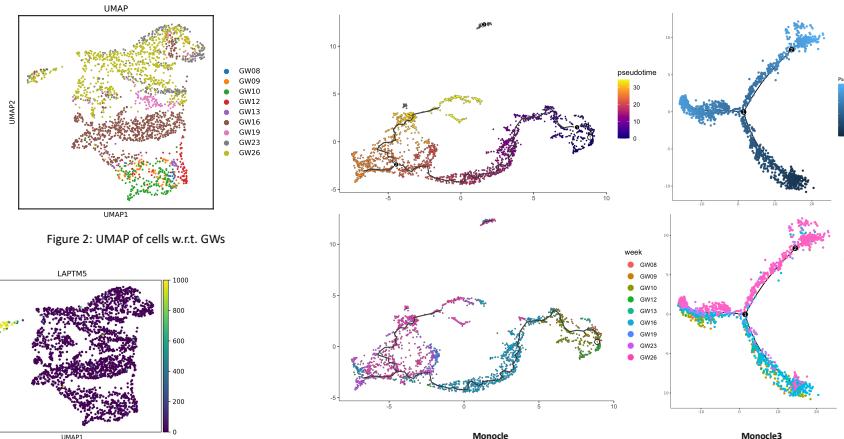


Figure 6. Comparison of developmental trajectories along weeks using Monocle and Monocle3

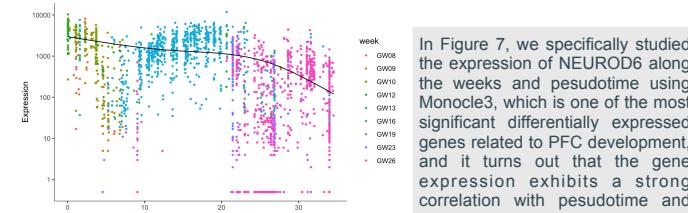


Figure 7. Expression of NEUROD6 along weeks and pseudo-time.

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.

References

- [1] Roth, G. & Dicke, U. Evolution of the brain and intelligence in primates. *Frog. Brain Res.*, 195, 413–430 (2012).

[2] O’Reilly, R. & Müller, F. Significant features in the early prenatal development of the human brain. *Ann. Anat.*, 190, 105–118 (2008).

[3] Zhong, S., Zhang, X., Fan, Y. & A, C. Single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*, 555, 524–528 (2018). <https://doi.org/10.1038/nature25908>

[4] McInnes, L. 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. <https://jmlr.org/papers/v9/vandermaaten08a.html>

[6] Mackiewicz 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 geodesic 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] Pérezsíki, Peter, Maciej Kuzniar, 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, Junyu, et al. "The single-cell transcriptional landscape of mammalian organogenesis." *Nature* 566.7745 (2019): 496-502.

Contributions

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