Developmental Trajectory Analysis of Single Cells in the Human Embryonic Prefrontal Cortex

Ruochen MA[1], Jihong TANG[1], Yuyan RUAN[2] , Zhi HUANG[2] {rmaam, jtangbd, yruanaf, zhuangdq}@connect.ust.hk
[1]: Division of Life Science, HKUST [2]: Department of Chemical and Biological Engineering, HKUST

Introduction

The prefrontal cortex (PFC) is one of the most complex and important regions of the brain, as it plays a crucial role in a wide range of cognitive processes that underpin human behavior and decision-making [1]. The PFC is located at the front of the brain's frontal lobe and is involved in a variety of functions, including working memory, attention, planning, decision-making, and social behavior. Dysfunction of the PFC has been linked to a number of neuropsychiatric disorders, including schizophrenia, depression, and ADHD. This suggests that understanding the underlying biology of the PFC is crucial for developing effective treatments for these disorders. Recent advances in single-cell sequencing technology have provided researchers with an unprecedented level of details about the PFC. By analyzing individual cells in the PFC, researchers can gain insights into the genetic and molecular mechanisms that underlie the function of this complex brain region. Moreover, these technologies enable researchers to investigate the developmental trajectories of PFC cells at the gene level, providing a more comprehensive picture of how the PFC develops and functions over time [2].

In this project, we used single-cell sequencing data from 2394 single cells consisting of 21453 genes at nine different gestational weeks [3]. We first conducted hierarchical clustering to identify subgroups of PFCs. Next, to uncover the developmental trajectories of PFCs, we applied dimensionality reduction methods and calculated the pseudotime. Lastly, we conducted single-cell topological data analysis (TDA) which is a topology-based computational analyses and evaluated its performance.

Data Description

The dataset includes 2394 single cells spanning from gestational weeks 8 to 26, with transcripts per million (TPM) of 24153 genes for each cell. We utilized several algorithms including Slingshot (SLS), TSCAN, Monocle3 (MNC) and scTDA. The programs were written in R, and Python.

Results

1. Cell type subgroups of prefrontal cortex (PFCs)

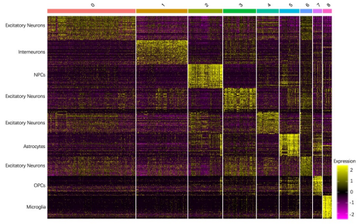

Fig. 1A. Hierarchical clustering results in 9 subgroups.

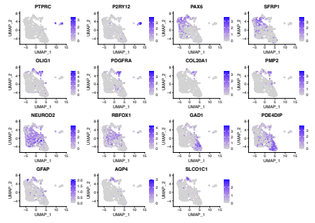

Fig. 1B. Marker gene expression in cell clusters.

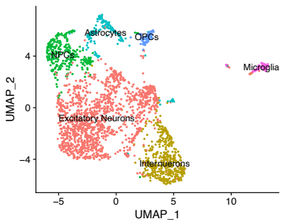

Fig. 1C. Cell type annotation of 9 clusters.

Nine different subgroups of cells, including excitatory neurons, interneurons, NPCs, astrocytes, OPCs, and microglia cells, were identified through hierarchical clustering using Seurat packages (as shown in Figure 1A). Each cell subgroup was distinguished based on cell-specific expressed marker genes, and detailed marker gene expression in cell clusters is presented in Figure 1B. The U-MAP visualization in Figure 1C displays the annotation of cell types of the PFC, which correspond to nine different clusters in the heatmap.

2. Comparison and Interpretation of Trajectory Inference Results Using Different Methods

Our goal was to identify the best method for accurately reconstructing cell developmental trajectories. To assess the performance of each method, we first calculated the Pearson correlation between the week time and the inferred pseudo time values obtained from each method. Our findings indicate that the UMAP1 dimension is inversely correlated with the week time, suggesting a negative relationship between the two. However, Monocle exhibits a strong correlation with the week time, indicating that its inferred trajectories closely align with the ground truth timeline. On the other hand, Slingshot performed the worst among the methods we evaluated. It did not exhibit significant correlation with either UMAP1 or the week time, suggesting that its inferred trajectories significantly deviate from both the underlying data structure and the ground truth timeline. Based on these results, in the following analysis, we regard Monocle as the preferred method for reconstructing cellular developmental trajectories.

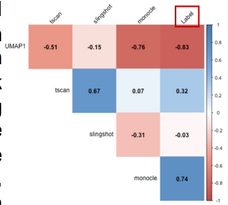

Fig. 2A Correlation between UMAP1 and pseudo time.

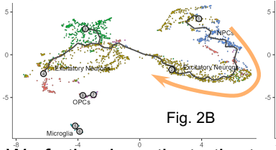

Fig. 2B

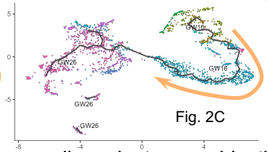

Fig. 2C

Fig. 2B shows the trajectory generated by MNC, which is well-aligned with the ground truth time labels (Fig. 2C). In addition, there are several branches within the excitatory neuron cells in the inferred trajectory, which also corresponds with the cell lineage tree reported in the paper [3].

We further investigate the temporally variant genes identified by different algorithms. A GAM [4] model for predicting each gene expression using pseudo time as the independent variable and the two most significant genes were further analyzed. The expressions of these genes versus the first component of the UMAP embeddings are plotted in Fig. 2D. We can see that astrocytes, excitatory neurons and microglia cells can be roughly separated by all algorithms. However, NPCs are clearly separated from other cells in MNC. For TSCAN, NPCs are partially mixed with neurons. As in SLS, they are hardly identified. This may account for the great performance difference since NPCs are stem cells, and precisely recognizing the original point may be more important. This indicates that correctly identifying the root cells can help improve the accuracy of trajectory inference.

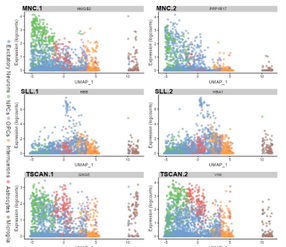

Fig. 2D. Visualization of the top 2 temporally variant genes

Results

3. Developmental trajectory of PFCs revealed by topological data analysis

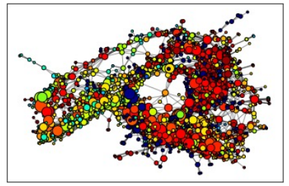

Fig. 3A. Rooted topological representation

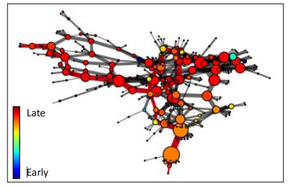

Fig. 3B. Skeleton of the network

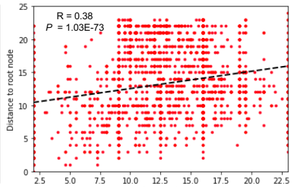

Fig. 3C. Pseudo-time based on topological representation

In addition, we took advantage of the topological data analysis technique to help reveal the developmental trajectory of PFCs. After transforming the data into the longitudinal format, the rooted topological representation of the PFCs (Figure 3A.) could be constructed using scTDA[5]. The skeleton of the network in Figure 3B seems to capture some patterns of the differentiation process. The pseudo-time calculated based on the topological representation shared a significant correlation with the actual time point in Figure 3C. However, further work on data engineering may lead to better construction of the topological representation.

Conclusion

In conclusion, our study investigated the developmental trajectories of prefrontal cortex (PFC) cells using single-cell sequencing data. Hierarchical clustering identified distinct cell subgroups, and marker gene expression confirmed their identities. Evaluating different methods, we found that Monocle demonstrated a strong correlation with the actual timeline, indicating its accuracy in reconstructing cellular developmental trajectories. Slingshot, on the other hand, deviated significantly from both the data structure and ground truth timeline. Incorporating topological data analysis provided additional insights into the PFC's differentiation process. Our findings emphasize the importance of accurately identifying stem cells and highlight Monocle as a preferred method for precise trajectory inference in PFC developmental studies.

Reference

[1] Teffer, K., & Semendeferi, K. (2012). Human prefrontal cortex: evolution, development, and pathology. Progress in brain research, 195, 191–218.
[2] Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., Dudoit, S., & Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. Nature Communications, 11(1).
[3] Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., Xu, X., Tang, F., Zhang, J., Qiao, J., & Wang, X. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature, 555(7697), 524–528.
[4] Hastie, T. J. (2017). Generalized additive models. Statistical Models in S, 249–307.
[5] Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., & Rabadan, R. (2017). Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. Nature Biotechnology, 35(6), 551–560.

Contribution

Ruochen MA: Seurat, PCA, UMAP, subgroups; poster; presentation
Jihong TANG: Topological data analysis; poster; presentation
Yuyan RUAN: Monocle, TSCAN, temporally variant genes; poster; presentation
Zhi HUANG: Slingshot, monocle, Correlation and comparison, poster, presentation