Exploration and Analysis of the Human Prefrontal Cortex Data

Jooran Lee[1*], Minghao WANG[2*]

[1]: Department of Life Science, HKUST [2]: Department of Chemical and Biological Engineering, HKUST

Supervised by Prof. Jiguang WANG, *: Equal Contribution

Introduction

The prefrontal cortex (PFC), located in the anterior section of the frontal lobe, is a critical area of the cerebral cortex. The maturation of cells within the PFC is vital for its functionality, which is essential for cognitive processes like decision-making and memory [1]. Single-cell RNA sequencing (scRNA-seq) technologies allow researchers to access the detailed transcriptional information of PFC cells in human embryonic prefrontal cortices from various gestational weeks (GWs). This capability facilitates the study of developmental trajectories and enables gene-level functional analysis of the PFC [2].

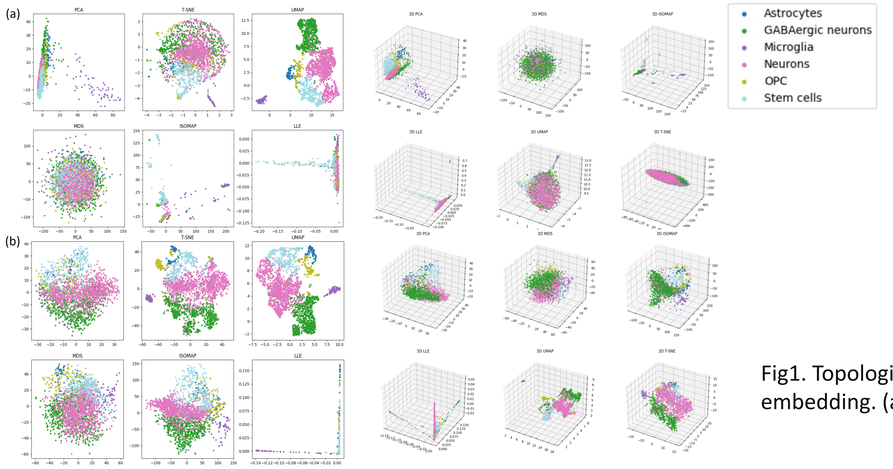
In this project, we analysed single-cell sequencing data from 2,309 individual cells, including 24,153 genes across 9 different gestational weeks (GWs) [3]. First, we use biological foundation models (BFM) [4] to integrate the original data into a unified space for automatically batch correction. Topological methods including UMAP, T-SNE, PCA, MDS, ISOMAP, LLE are compared, and cell-type subgrouping was achieved with graph-based clustering, visualized on a t-SNE plot. Trajectory analysis methods including Slingshot, Diffusion Map, and Monocle3 are performed to track the developmental trajectories of PFCs.

Data and Software

The dataset includes 2,309 single cells spanning from gestational weeks 8 to 26, with transcripts per million (TPM) of 24,153 genes for each cell. We utilized several algorithms including Diffusion Map, Slingshot, Monocle3, and Universal Cell Embedding (UCE). The programs were written in R, and Python.

Results

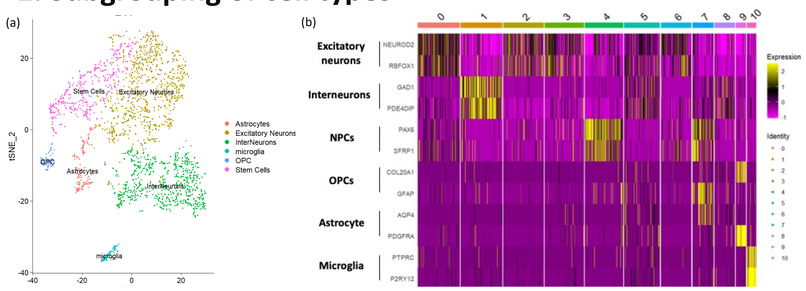
1. Topological analysis



From the 2D and 3D visualization, we can conclude that UMAP is the most suitable method for both original and BFM data. Besides, after data passing through BFM, the T-SNE result becomes better, which means after projection, even a random initialization of distribution (T-SNE setting) can also get good projection results.

Fig1. Topological analysis on the PFC data and BFM embedding. (a): original PFC, (b): BFM embedding.

2. Subgrouping of cell types



Cell-type subgrouping was achieved through graph-based clustering in Seurat and visually represented in a T-SNE plot. The distinct cell types were labelled according to their marker gene expression markers, and total of six cell clusters were identified, where majority consisting of excitatory neurons and interneurons.

Fig2. (a) Clustering result with cell-type label. (b) Heatmap showing the expression levels of the marker genes.

3. Trajectory analysis

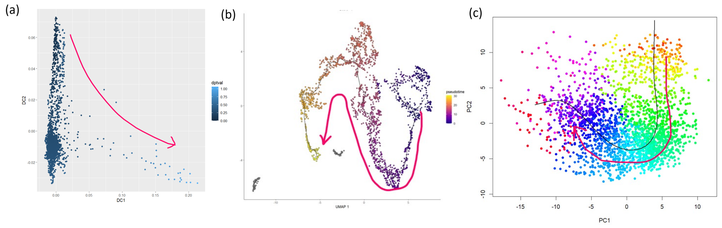


Fig3. Trajectory analysis results. (a)Diffusion map, (b)Monocle3, (c)Slingshot.

The trajectory analysis was performed via Diffusion map(a), Monocle3(b), and Slingshot(c). The Spearman correlation coefficients were determined to assess the relationship between pseudo-time and week number for each method. The correlations were found to be R=0.426 (a), R=0.543 (b), and R=0.803 (c), respectively, all of which were statistically significant with p-values less than 2.2e-16. Both clustering and pseudo-time prediction was best performed by the Monocle3, thereby further downstream analysis were performed by using the pseudo-time calculated by the Monocle3.

The genes showing the differential expression along the PFC development were found, where most were significantly associated with the neuron differentiation, and neuronal development. For instance, the expression of NEUROD6 (Fig4) rises initially during the early developmental period (GW 08-12) while gradually decreases as the cells become mature in the later stage.

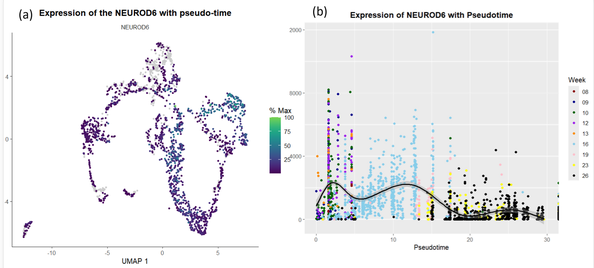


Fig4. (a) Feature Plot of NEUROD6 (b) Expression of NEUROD6 along the pseudo-time during PFC development

Discussion

In this project, BFM, clustering methods, and the marker gene expressions of the human PFCs were used to identify their cell types. Different methods of the developmental trajectories of single cells were used, and Monocle3 was identified to outperform compared to other trajectory prediction methods. Several genes were identified to be the prediction markers of the pseudo-time of the PFCs, However, their expressions should be further experimentally validated.

References

[1] Zhong, Suijuan, et al. "A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex." Nature 555.7697 (2018): 524-528.

[2] Van den Berge, Koen, et al. "Trajectory-based differential expression analysis for single-cell sequencing data." Nature communications 11.1 (2020): 1-13.

[3] Zhong, S., et al. "A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. " Nature, 555(7697), 524–528.

[4] Yanay Rosen, et al. "Universal Cell Embeddings: A Foundation Model for Cell Biology. " bioRxiv 2023.11.28.568918

Contributions

Jooran LEE: Cell subgrouping, Trajectory analysis

Minghao WANG: BFM embedding, Topological analysis