CSIC 5011 Final-Project: Uncover Key Genes Regulating Human Embryonic Prefrontal Cortex Development by Single Cell Sequencing

Zhihan Zhu[1], Dong Song[2], Jiabao Li[2] and Zongchao Mo[2] {zzhuay, dsongad, jligm, zmoad}@ust.hk
[1]: Department of Chemical and Biological Engineering, HKUST [2]: Department of Life Science, HKUST

Introduction

The prefrontal cortex (PFC) is the cerebral cortex in the front part of the frontal lobe. The development of the prefrontal cortex cells matters in the function of the human prefrontal cortex [1], which is regarded as the center of cognitive functions such as decision-making, memory, etc. By single-cell sequencing technologies, the high-dimensional transcriptional information of the PFC cells is accessible in human embryonic PFCs of various gestational weeks (GWs), enabling the exploration of the developmental trajectories [2] and the functional analysis in gene-level for PFCs.

In this project, the single-cell sequencing data of 2394 single cells from 9 different GWs was collected [1]. Transcripts per million (TPM) of 21453 genes for each cell are counted and applied into the hierarchical clustering to discover the subgroups of the PFCs. By adopting software such as Slingshot, TSCAN, and Monocle3, the dimensional deduction methods such as PCA calculate the pseudo time [3] of the PFCs to uncover the developmental trajectories of the differentially expressed genes, whose gene functions elucidate the potential explanation of the PFC development.

Data & Software

There are 2394 single cells from gestational weeks (GW) 8 to 26. There are transcripts per million (TPM) of 24153 genes for each cell. Software used in this project is Slingshot, TSCAN, Monocle3, R, python.

Results

1. Developmental subgroups of prefrontal cortex (PFCs)

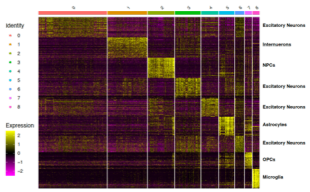


Fig 1

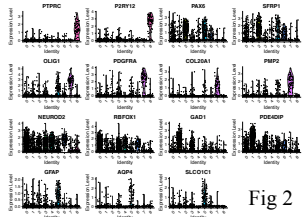


Fig 2

Via the hierarchical (shown in Fig 1) clustering, 9 different subgroups are identified using Seurat packages*. Based on cell specific expressed marker genes, cell subgroups are distinguished as excitatory neurons, interneurons, NPCs, astrocytes, OPCs and microglia cells. Detailed marker gene expression for the cell identification are presented in Fig 2.

U-MAP shows the PFCs cell types, corresponding to 9 different clusters in the heatmap. The distribution of the clustering indicates the relationship among cell types.

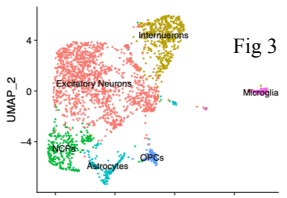


Fig 3

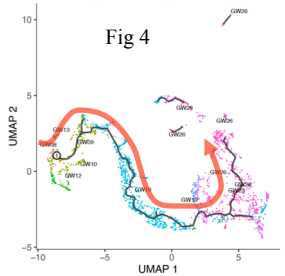


Fig 4

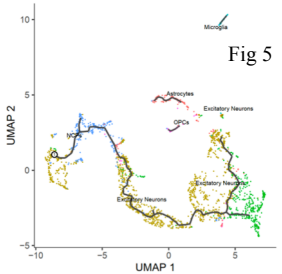


Fig 5

The trajectories of PFCs trace the gestational weeks (GW) of cells, with arrow indicating pseudo-time starting and ending point (Fig 4). Cell subgroup information is also projected into the trajectories to depict the developmental order (Fig 5).

3. Key genes regulating the prefrontal cortex in the developmental trajectories and their functions

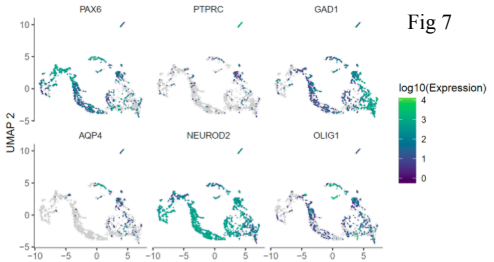


Fig 7

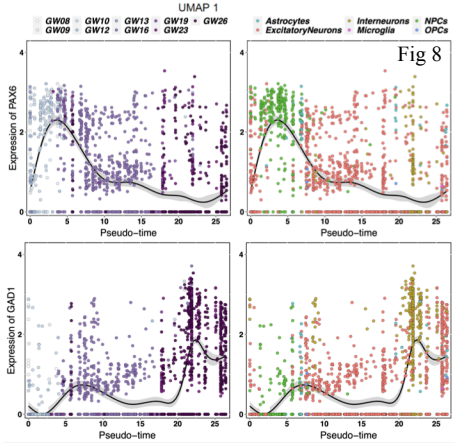
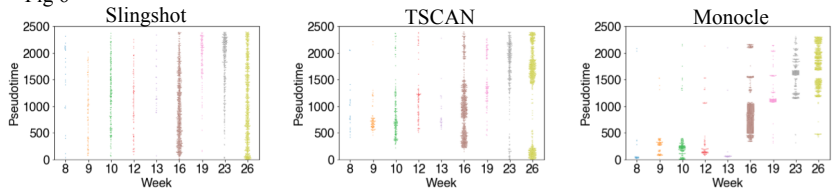


Fig 8

The developmental trajectories of predicted pseudo time for six marker genes (Fig 7) are selected to indicate the temporal expression in PFCs, regulating the development of the prefrontal cortex. For instance (Fig 8), expression of PAX6 increases at first and decreases with time going by. Expression of GAD1 rises fluctuatingly along the pseudo time axis. All of the genes are vital in the process of gene regulation, energy metabolism and other essential activities in cell development via the annotation of their gene functions.

2. Comparison among different Trajectories Inference Methods.

Fig 6



Correlation = 0.070 Correlation = 0.29 Correlation = 0.79

To compare the 3 trajectory inference methods*, we performed Spearman rank-order correlation test of the pseudo time and gestational week (Fig 6). As results, the Monocle3 algorithm has the highest correlation coefficient, indicating that Monocle3 can infer the time more properly in this dataset.

Discussion

In this projects, single cell expression of PFCs was subjected to clustering and cell type identification using Seurat. Developmental trajectories of single cells are uncovered using Monocle3 and UMAP, which outperforms other trajectory prediction methods. Significantly high correlation coefficient between predicted pseudo time and real GW timepoints indicate high consistency (r=0.79). Dynamic expression of marker genes further validated the cell trajectory inference.

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] Haghverdi, Laleh, et al. "Diffusion pseudotime robustly reconstructs lineage branching." Nature methods 13.10 (2016): 845.

Contribution

Zhihan Zhu: TSCAN, Correlation coefficient, Poster, Video
Dong Song: Trajectory, Visualization, Expression prediction, Slides, Video
Jiabao Li: Slingshot, Data processing, Poster, Video
Zongchao Mo: Seurat, Monocle3, Subgroups, Poster, Video

*For more information, please feel free to visit our GitHub webpage: https://github.com/jligm-hash/2021.csic5011.group1