



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

CSIC 5011 Final project

2023/05/17

MA Ruochen, TANG Jihong, RUAN Yuyan, HUANG Zhi

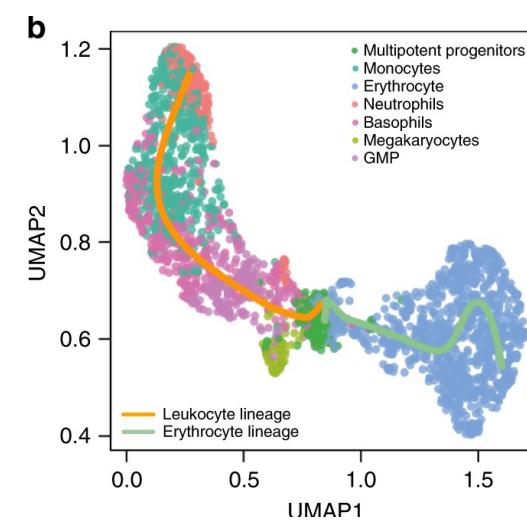
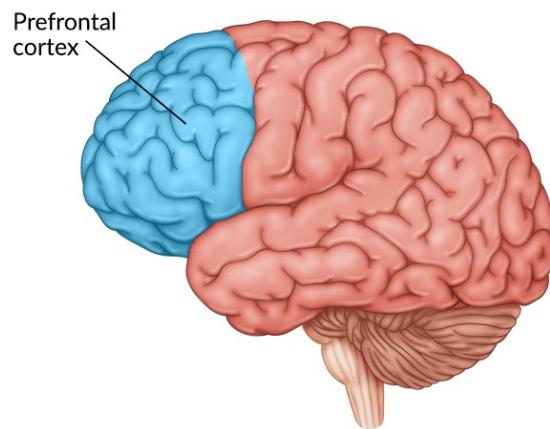
# Contents

- Introduction
- Aim & Dataset
- Part 1: Clustering analysis and cell type annotation
- Part 2: Trajectory analysis
- Part 3: Single-cell topological data analysis
- Conclusion
- Reference

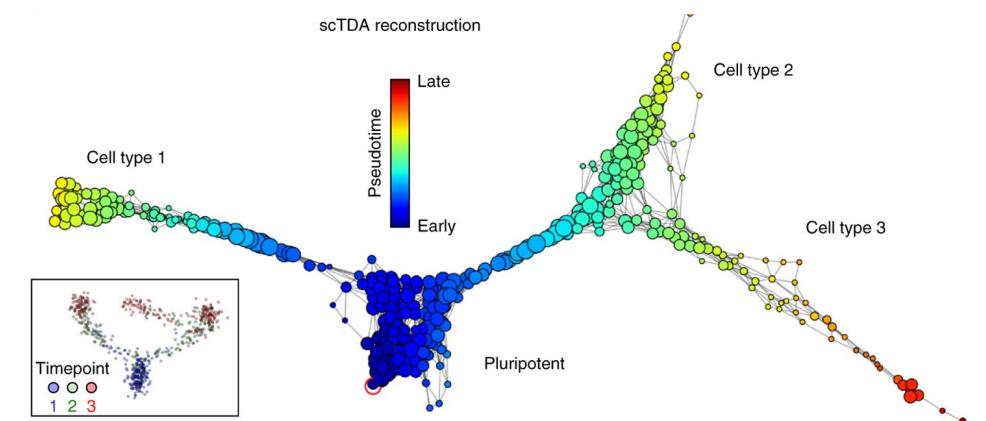
# Introduction



- Human prefrontal cortex (PFC)
  - plays a crucial role in higher-order cognitive functions
- Single cell sequencing enables developmental trajectory analysis



Van den Berge, et al., Nature Communications, 2020



Rizvi, et al., Nature biotechnology, 2017

# Aim & Dataset

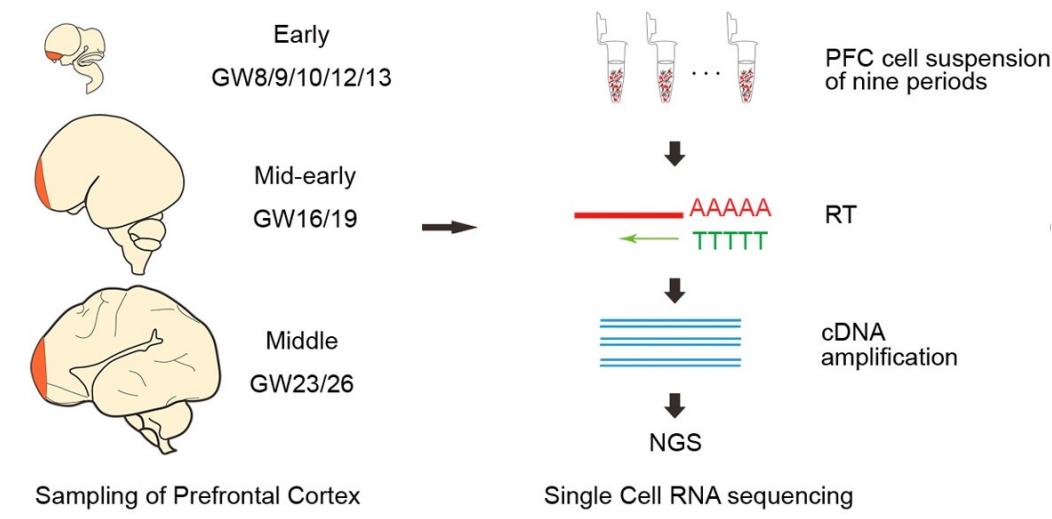


- Aim

- Identify subgroups of cells
- Trace the developmental trajectory

- Dataset

- 2394 single cells from gestational weeks (GW) 8 to 26
- Expression of 24153 genes in TPM



Zhong, et al, Nature, 2018

	GW8	GW9	GW10_01	GW10_02	GW10_03	GW12	GW13	GW16	GW19	GW23_01	GW23_02	GW26	Sum
Gender	Female	Female	Male	Female	Female	Male	Female	Female	Female	Male	Female	Female	
Sequenced cells	23	88	48	95	48	88	24	789	120	143	181	747	2,394
Filtered cells	23	88	47	92	47	85	24	776	120	132	176	699	2,309

Zhong, et al, Nature, 2018

# Part 1: Clustering analysis and cell type annotation

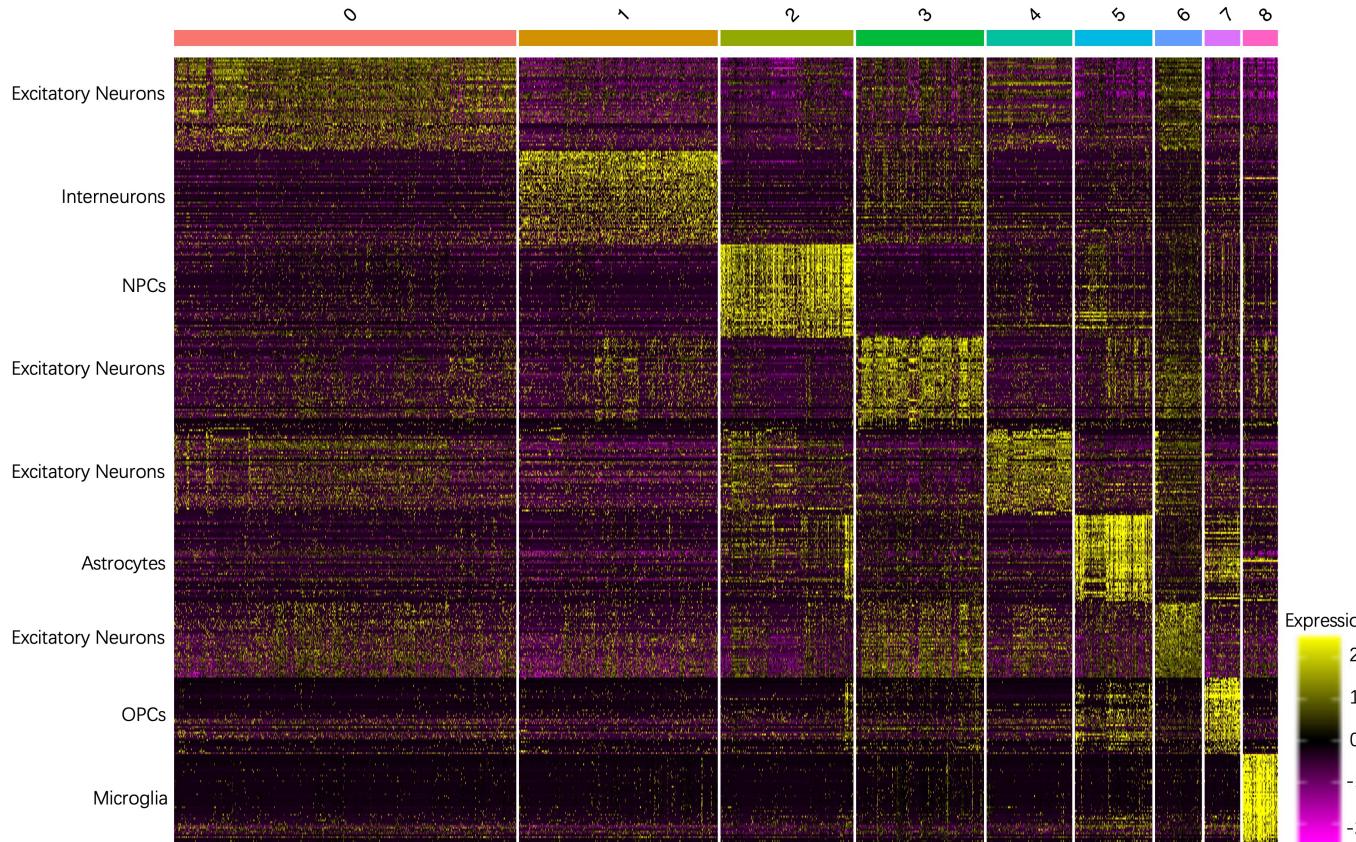


Figure 1a. Hierarchical clustering results in 9 subgroups.

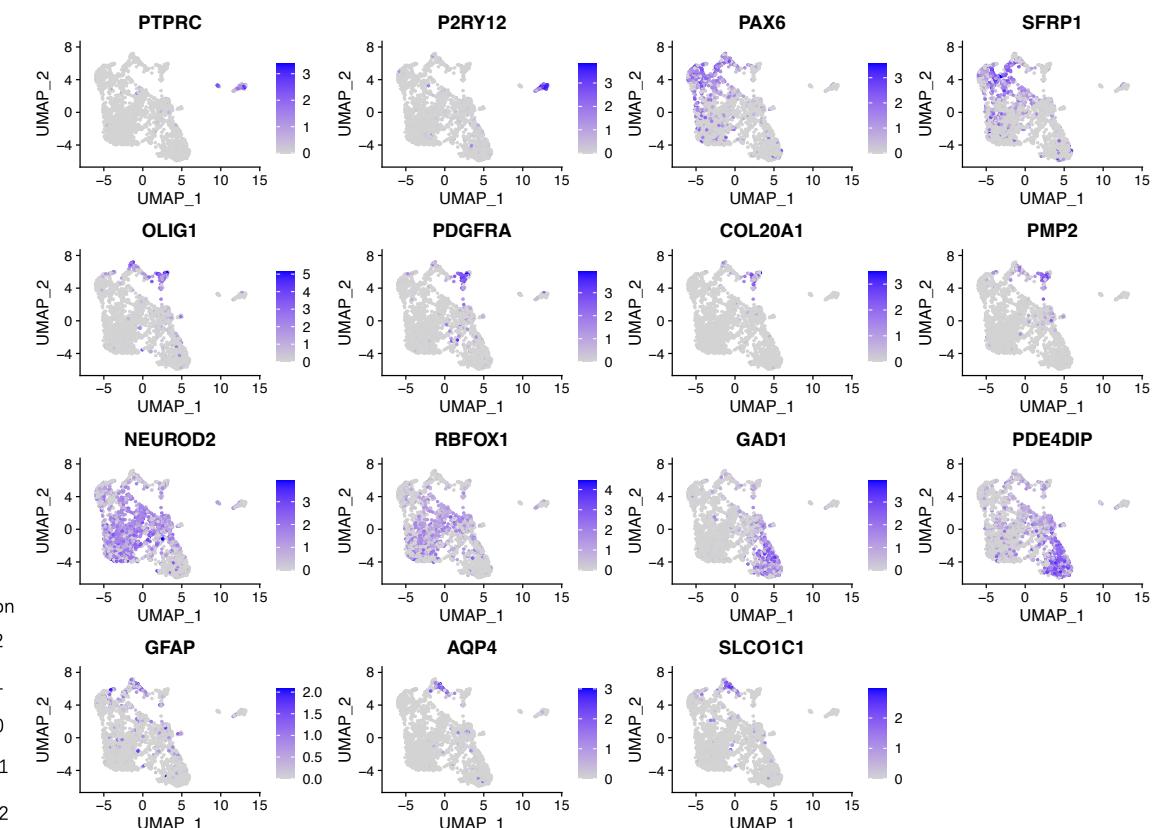


Figure 1b. Marker gene expression in cell clusters.

# Part 1: Clustering analysis and cell type annotation

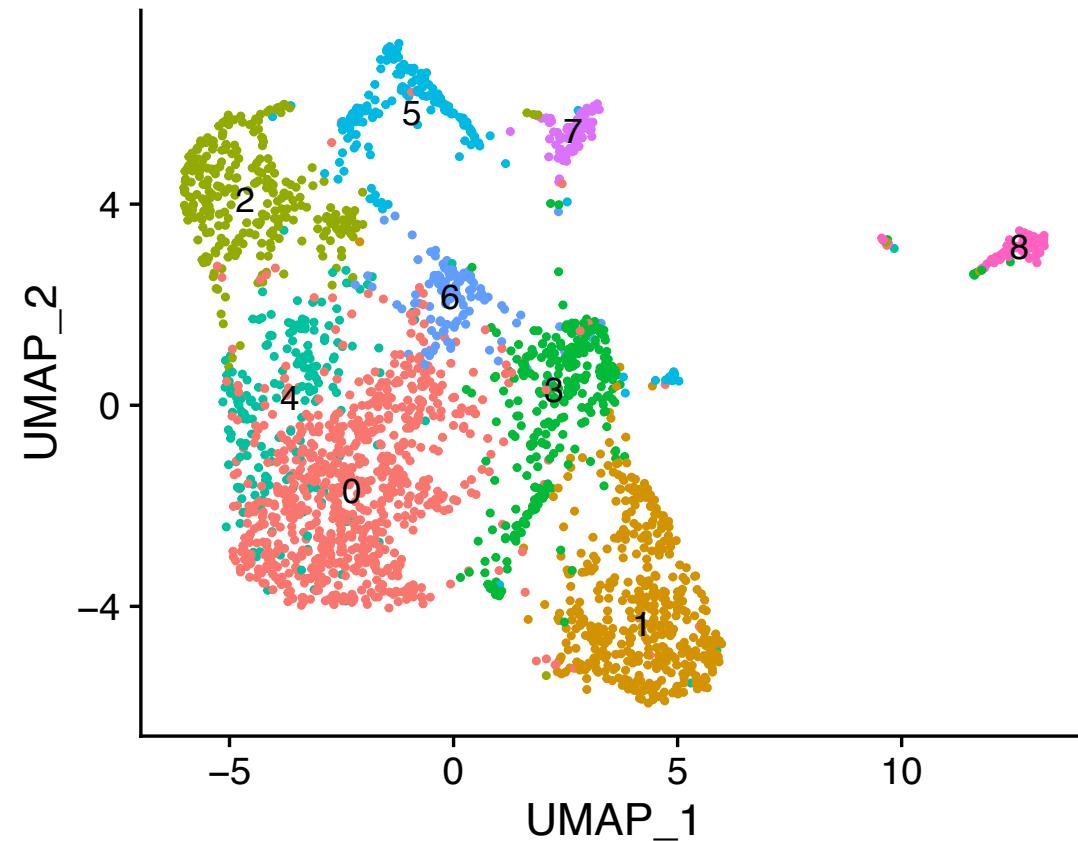


Figure 1c. UMAP visualization of the 9 cell subgroups.

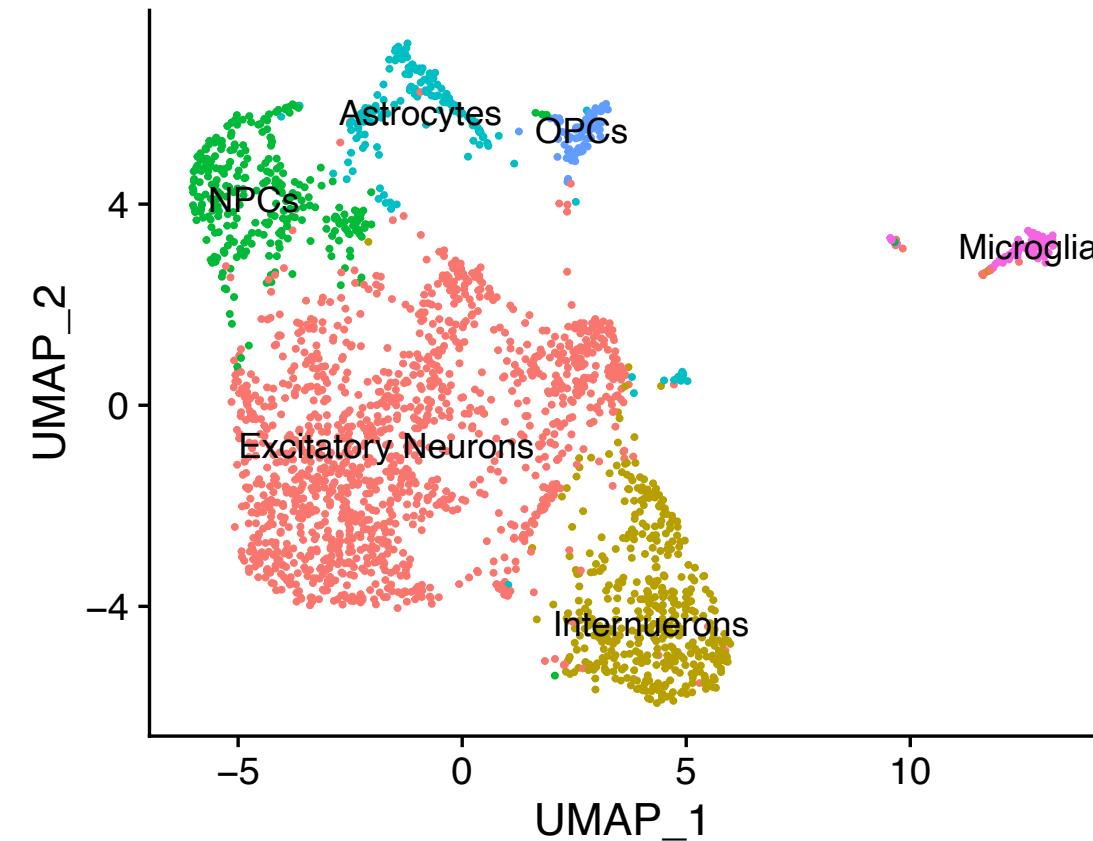


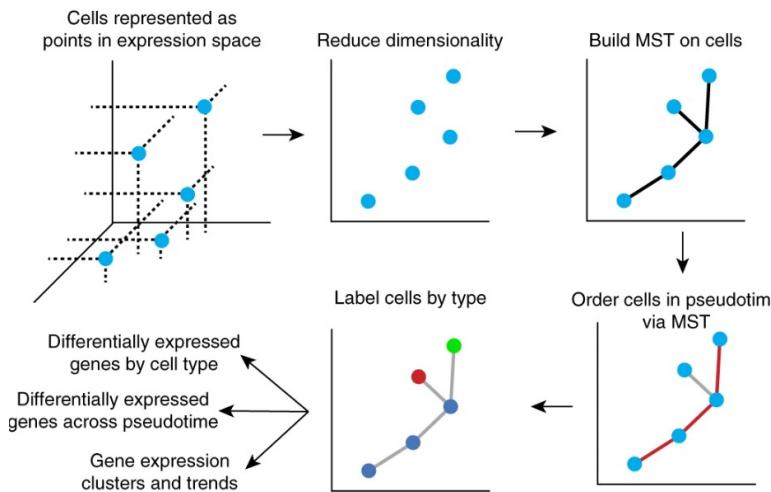
Figure 1d. Cell type annotation of the subgroups

# Part 2 Comparison of Analysis methods

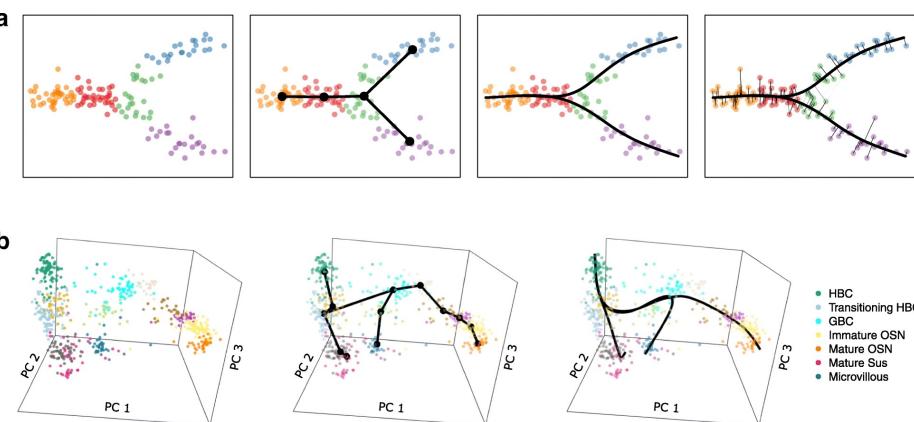


## Introduction to the trajectory Analysis methods

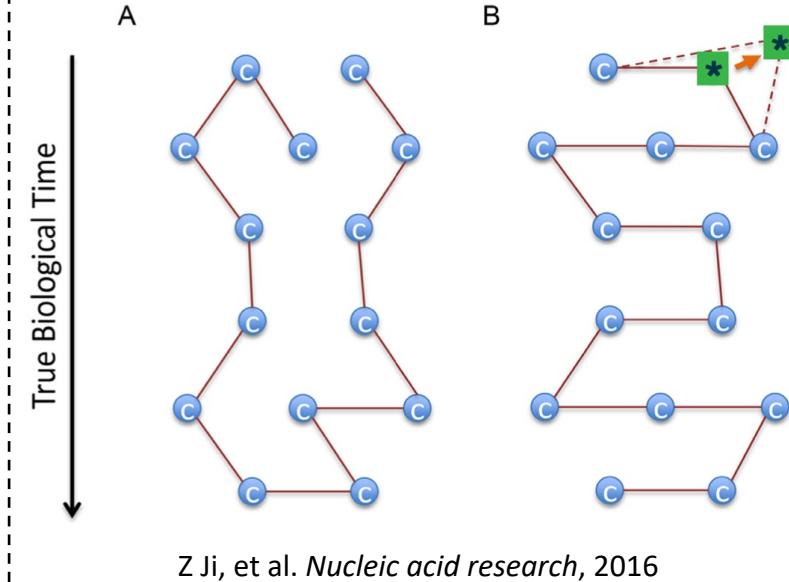
### Monocle



### slingshot



### TSCAN



Order cells based on their progress through a **biological process** (cell expression differentiation), allowing users to find genes that change as a function through the process

Use **cluster-based minimum spanning trees** (MSTs) to initialize pseudo time trajectories and fit simultaneous principal curves

Also construct a **minimum spanning tree** (MSTs) but estimate the number of different cell states(clusters)

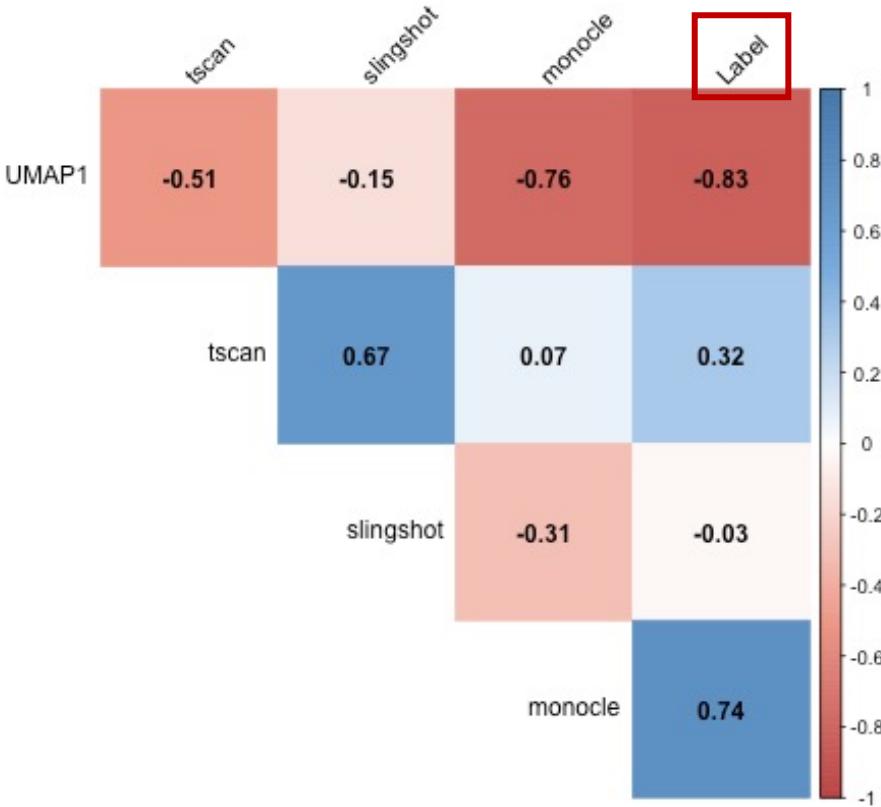
# Part 2 Comparison of Analysis methods



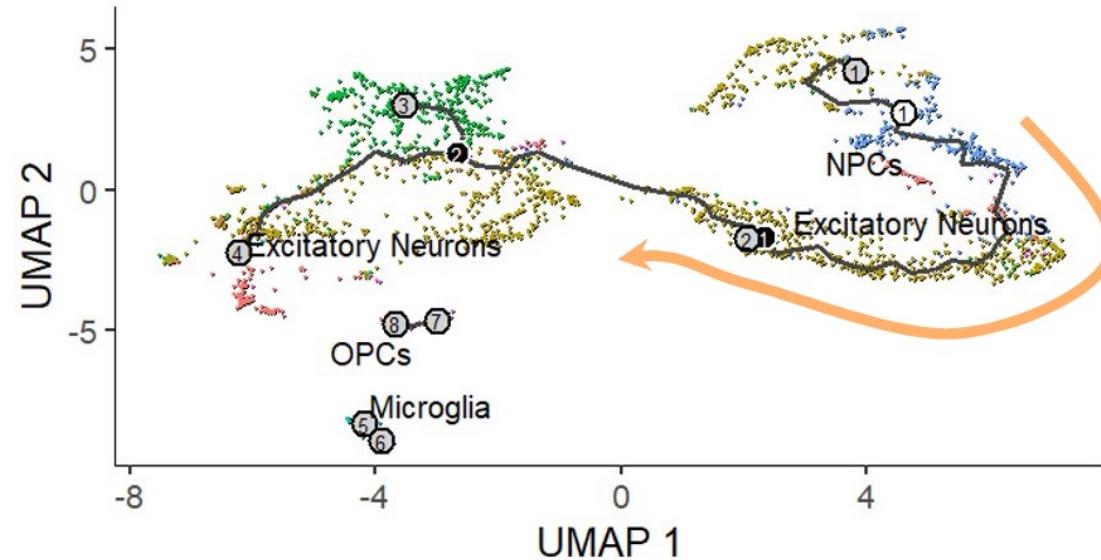
## Result

Correlation between the Ground truth (Week Time)  
and pseudo time result and UMAP1

A



B Monocle Analysis result



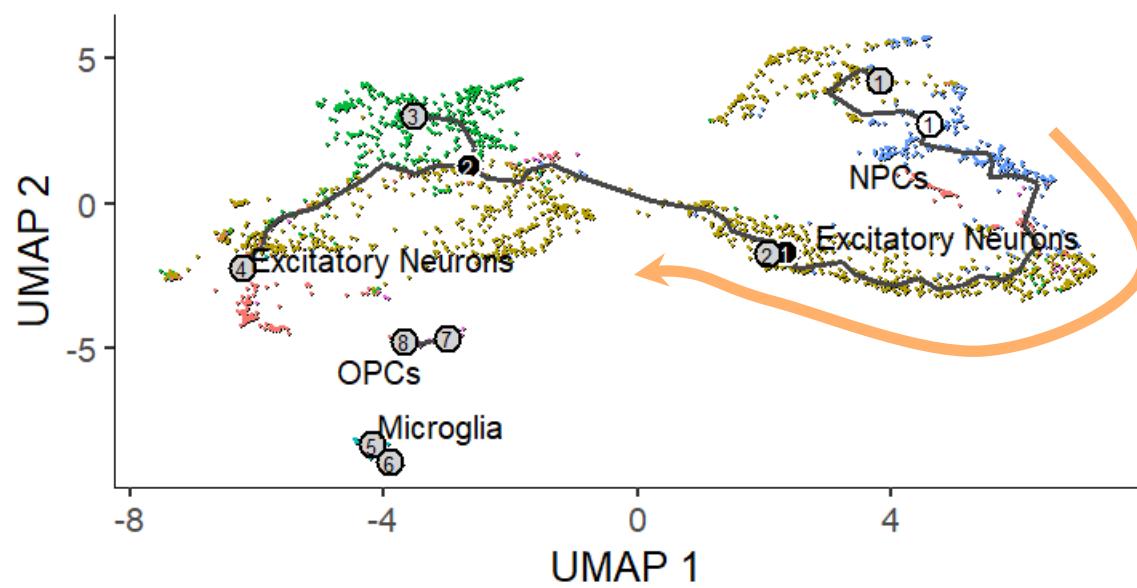
## Conclusion:

1. UMAP1 is inversely correlated to the week time.
2. Monocle is strongly correlated with week time (Ground Truth)
3. Slingshot preformed worst, which not correlated with UMAP and Week time

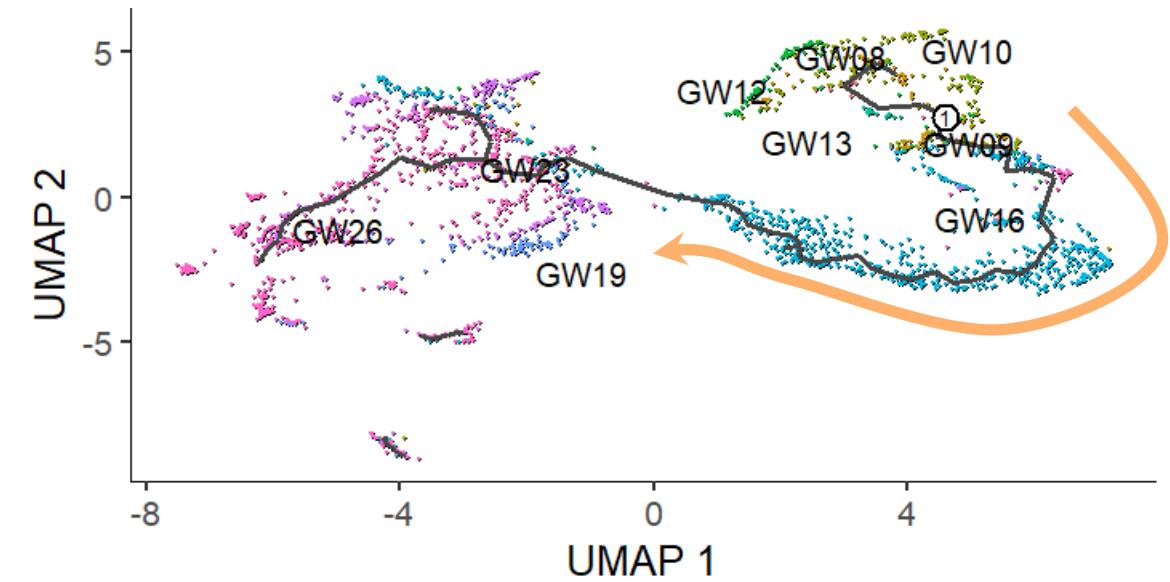
# Part 2: Trajectory Analysis: Detailed Look of Monocle3)



Trajectory inferred by the Monocle3



Ground truth time labels



Main trend is aligned with the ground truth time labels

# Part 2: Trajectory Analysis: Detailed Look of Monocle3)



Black circles indicate the branching nodes

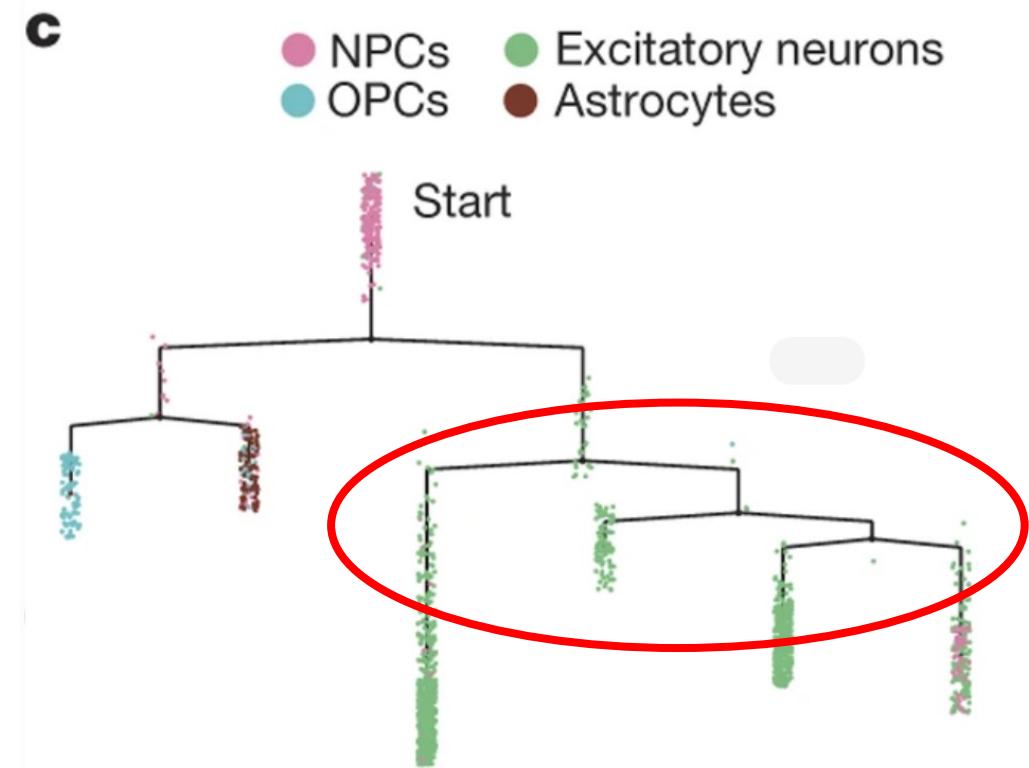
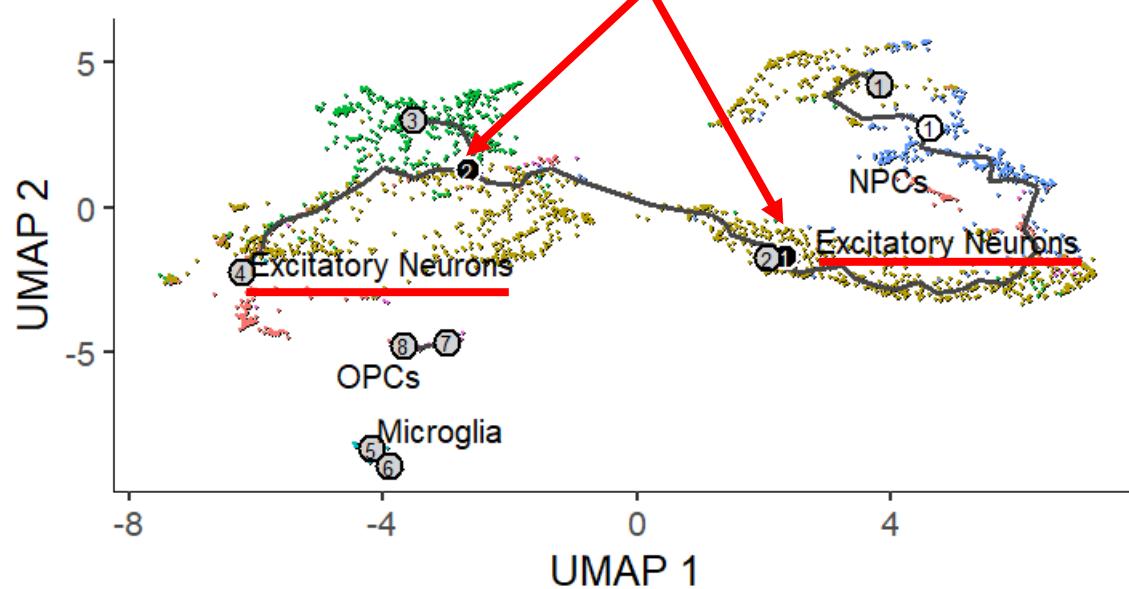


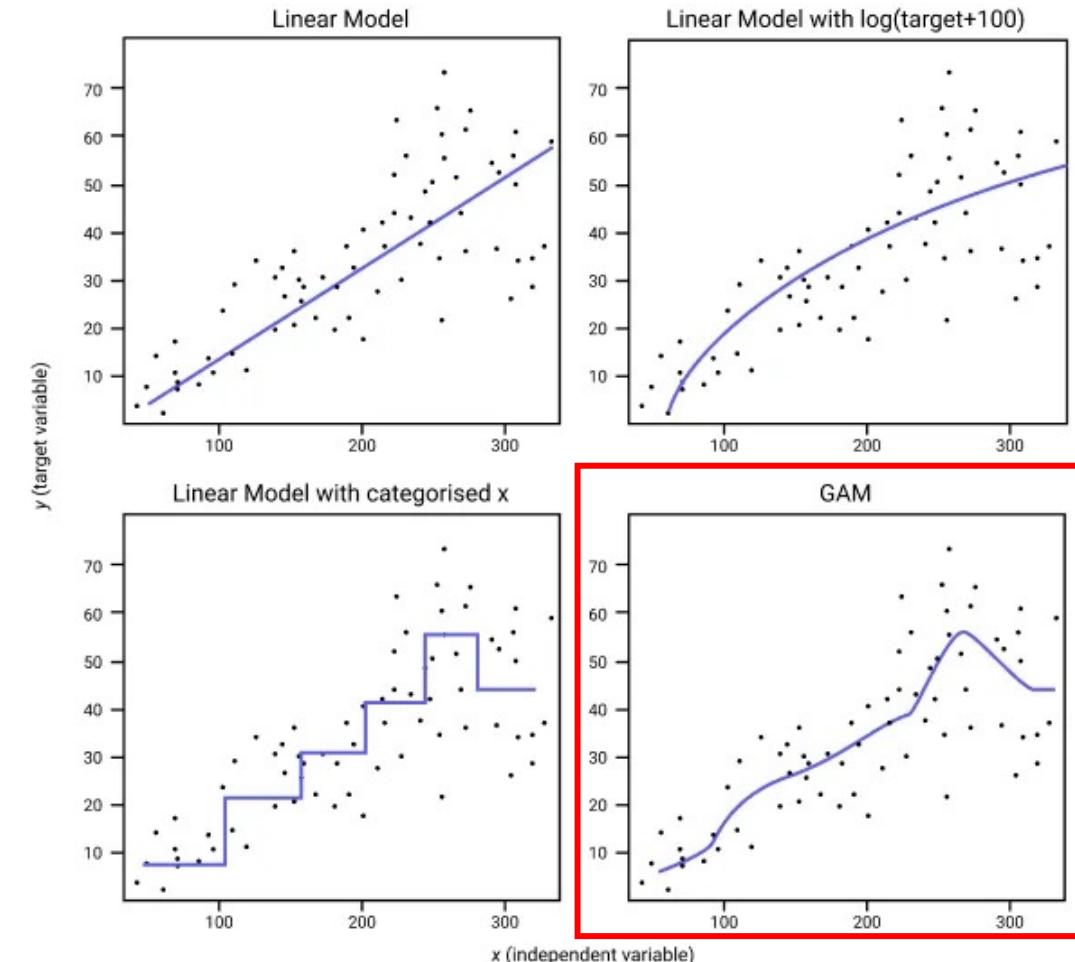
Fig 1. (c) of [1]

Excitatory neurons further evolves into several subtypes

# Part 2: Trajectory Analysis: Visualization of Temporally Varying Genes

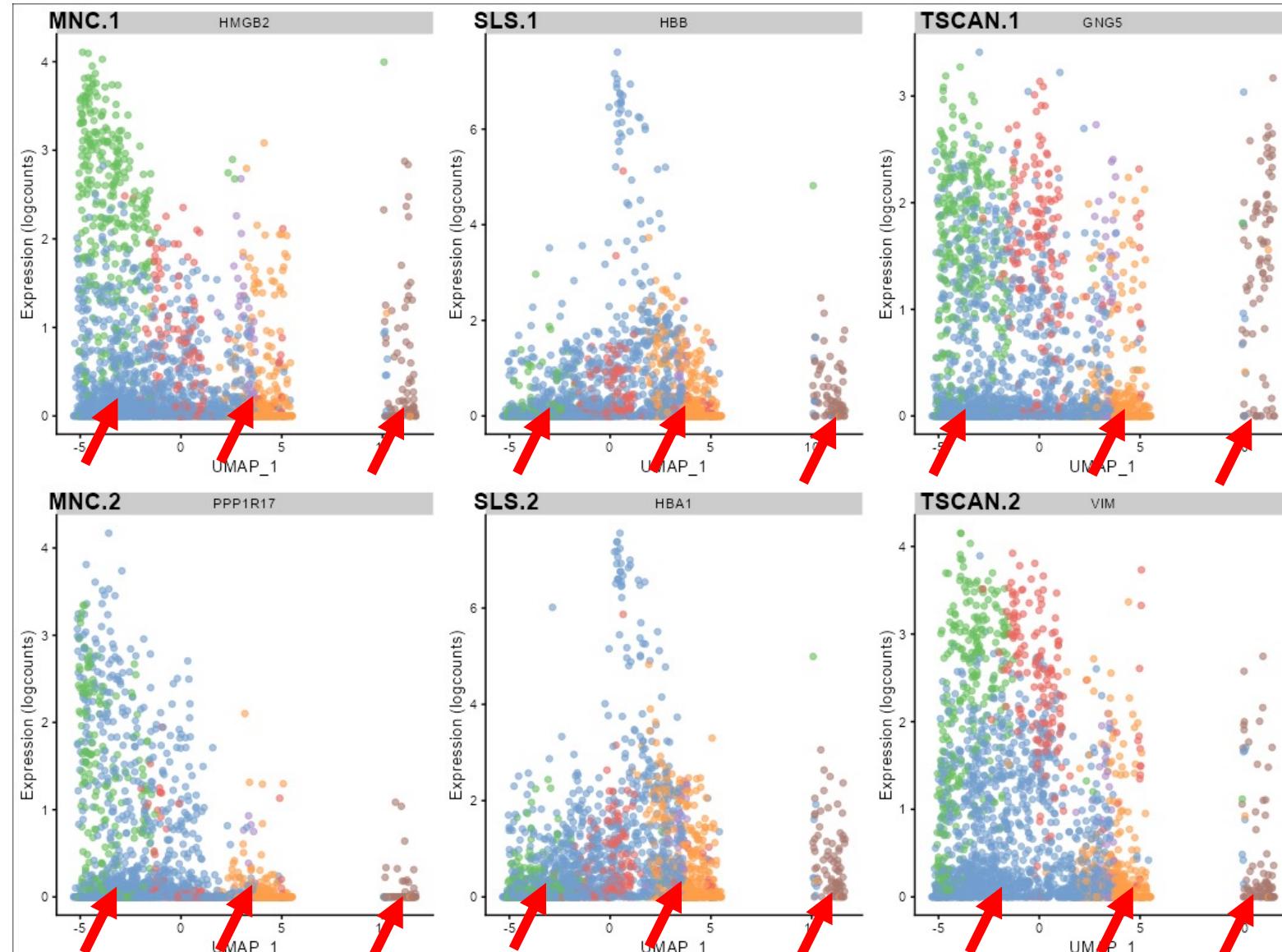


- Build a GAM model for every gene:  
Gene expression =  $f(\text{pseudo time})$
- Select top 2 genes with greatest statistical significance
  - Monocle3 (MNC): HMGB2, PPP1R17
  - Slingshot (SLS): HBB, HBA1
  - TSCAN: GNG5, VIM



Generalised Additive Model  
(GAM)

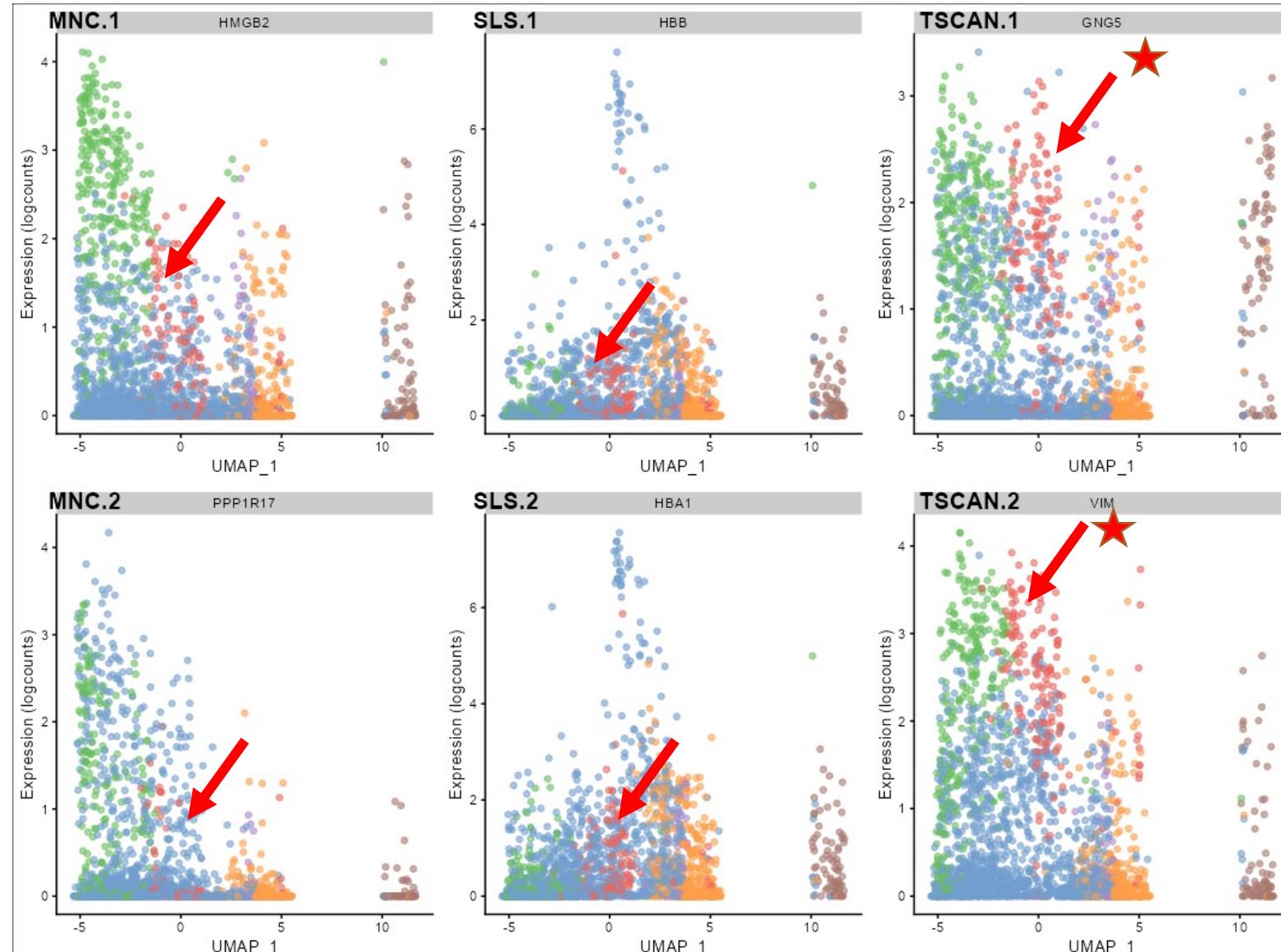
# Part 2: Trajectory Analysis: Visualization of Temporally Varying Genes



● Excitatory Neurons ● NPCs ● OPCs  
● Interneurons ● Astrocytes ● Microglia

**Correlation with ground truth labels**  
MNC: 0.74; TSCAN: 0.32; SLS: -0.03

# Part 2: Trajectory Analysis: Visualization of Temporally Varying Genes



● Excitatory Neurons ● NPCs ● OPCs  
● Interneurons ● Astrocytes ● Microglia

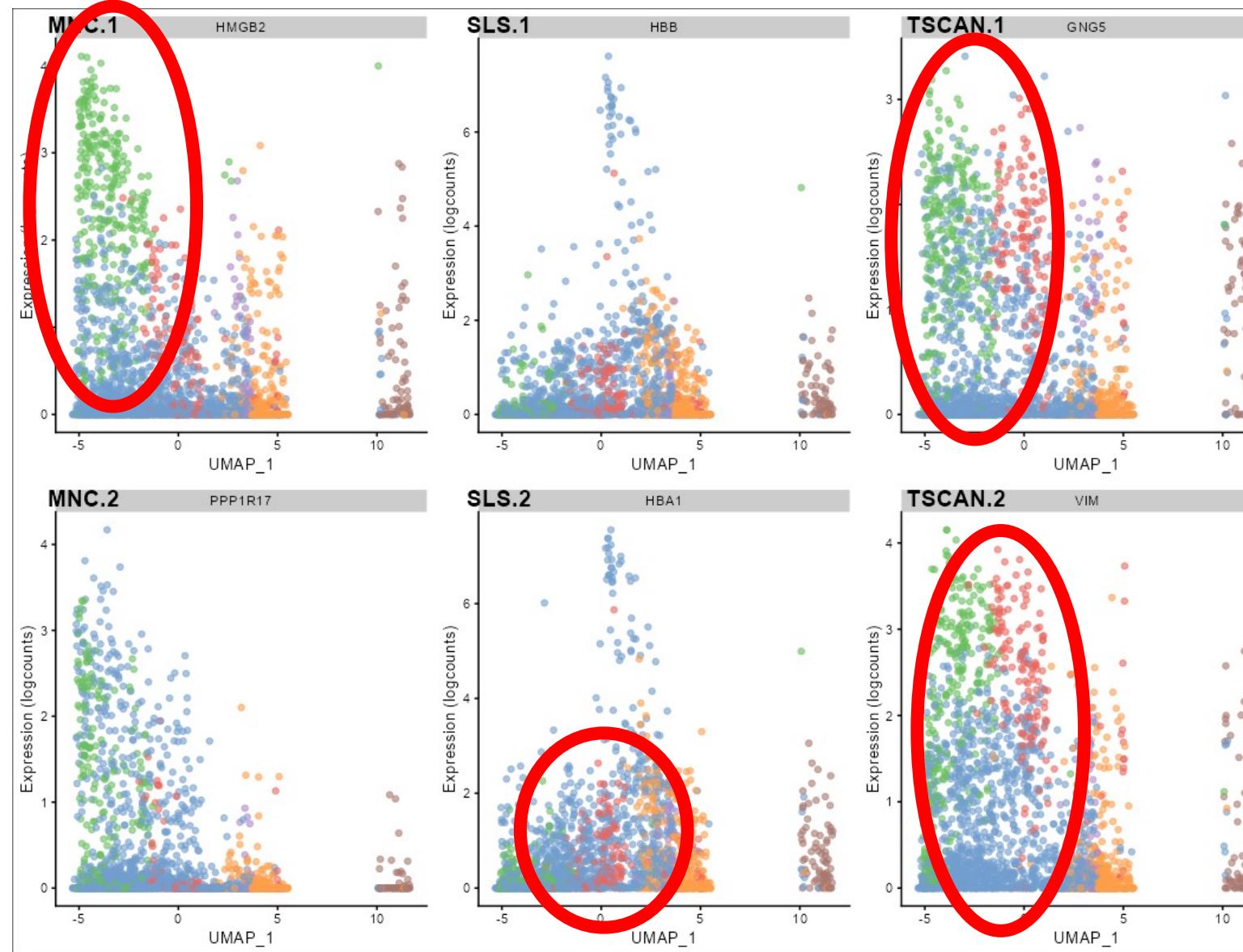
**Correlation with ground truth labels**  
MNC: 0.74; TSCAN: 0.32; SLS: -0.03

## Common

- **Excitatory neurons, interneurons, and microglia** cells can be roughly separated.
- **Astrocytes** are mixed, and can be separated better in **TSCAN**

\* MNC, TSCAN and SLS stands for Monocle3, TSCAN and Slingshot

# Part 2: Trajectory Analysis: Visualization of Temporally Varying Genes



● Excitatory Neurons ● NPCs ● OPCs  
● Interneurons ● Astrocytes ● Microglia

## Correlation with ground truth labels

MNC: 0.74; TSCAN: 0.32; SLS: -0.03

## Difference: NPCs

MNC : clearly separated

TSCAN: partially mixed with neurons

SLS: hardly identified.

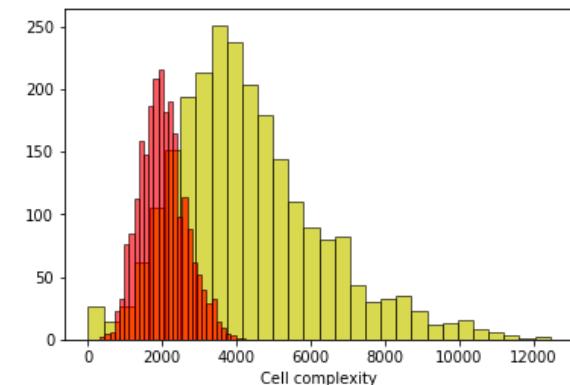
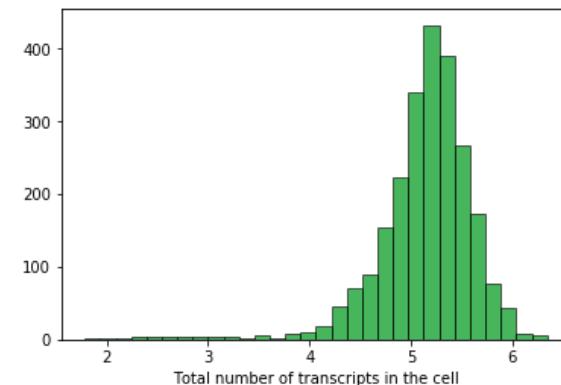
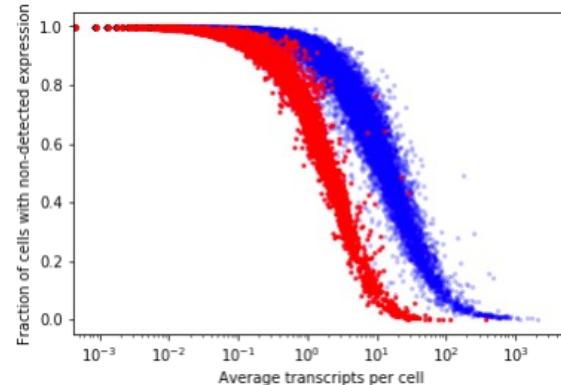
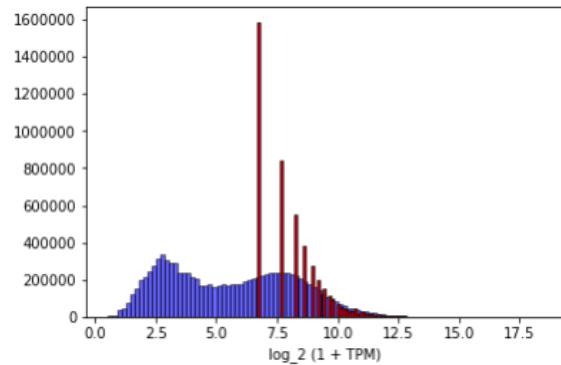
Correctly identifying NPCs (stem cells) is more beneficial and can improve the accuracy.

# Part 3 Topological Data Analysis

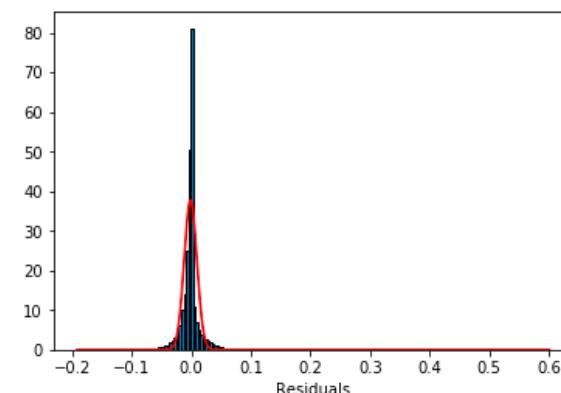
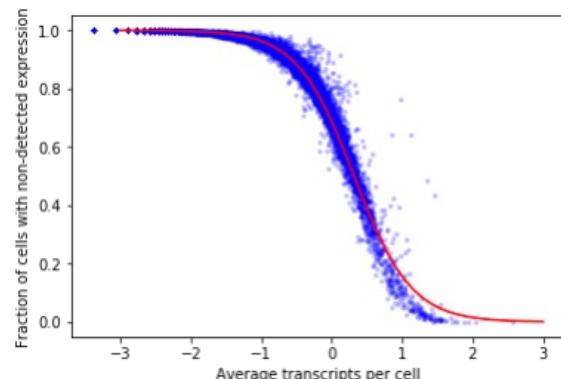


## 3.1: Data Preprocessing

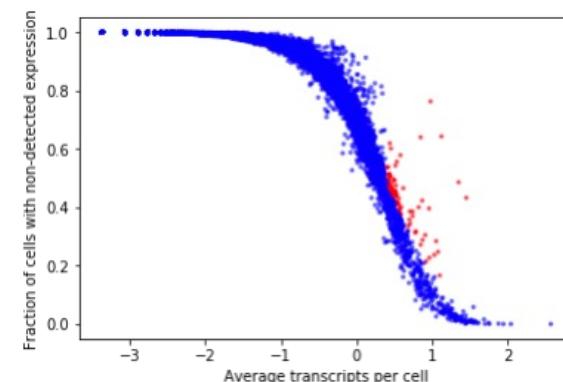
Subsample cells to pass all the criteria: Filter the cells



We can model the dependence of the dropout rate on the average gene expression by fitting a sigmoid function.



86 genes were selected



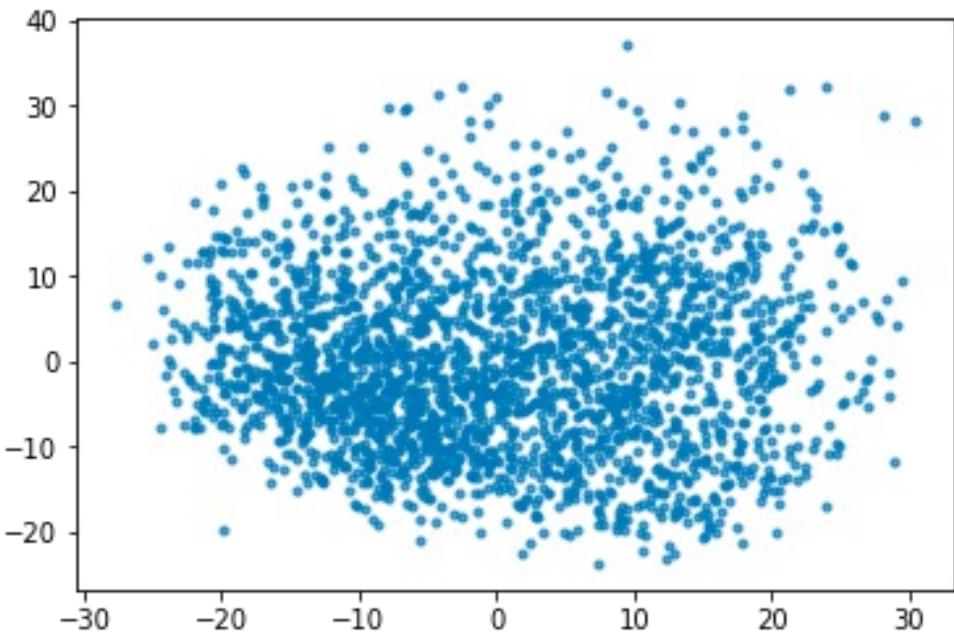
We can use the above sigmoidal fit to identify genes that have a low probability of drop-out and that have a higher variability than expected from purely technical noise.

# Part 3 Topological Data Analysis



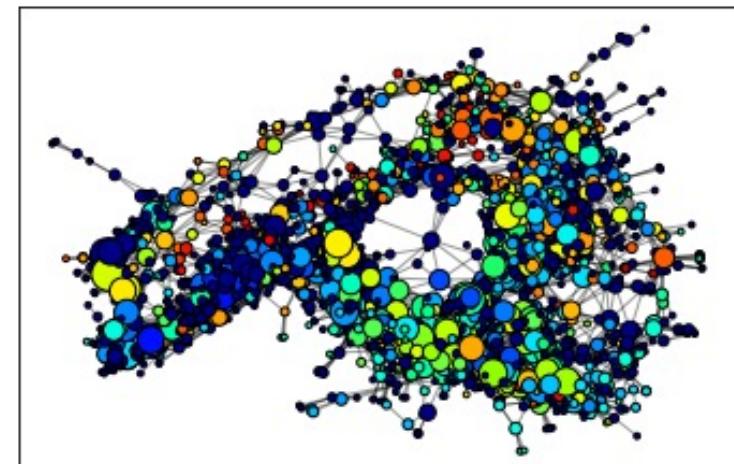
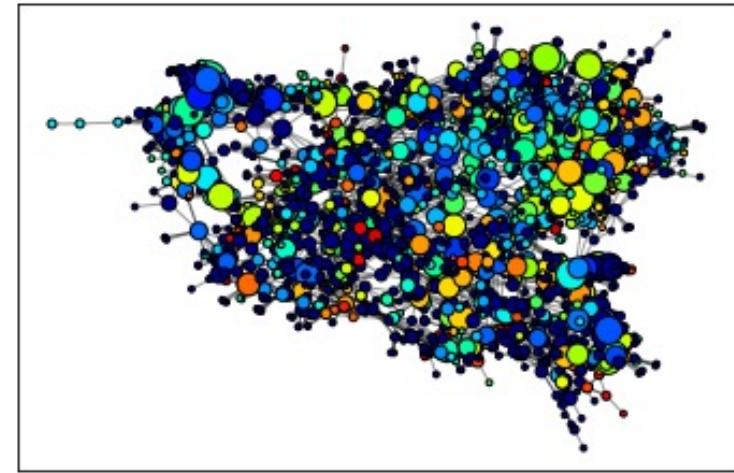
## 3.2: Topological Representation

- Dimensional reduction by PCA



Topological representation based on PC1 & PC2 using Mapper; Unrooted

Topological representation based on PC1 & PC2 using Mapper; Rooted



# Part 3 Topological Data Analysis



## 3.3: Developmental trajectory of PFCs revealed by topological data analysis

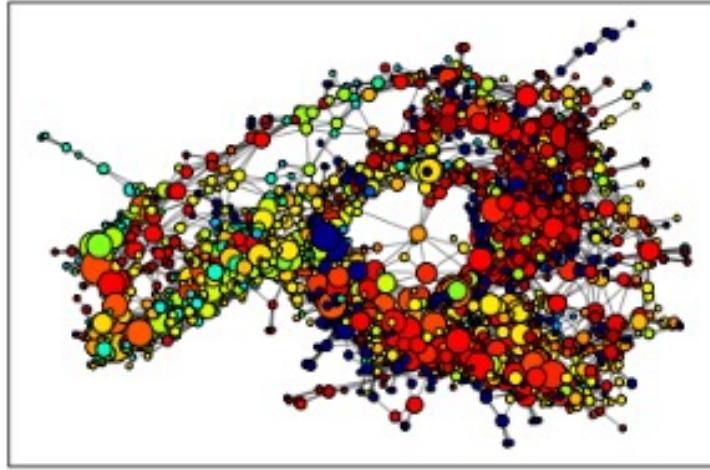


Fig. A Rooted topological representation

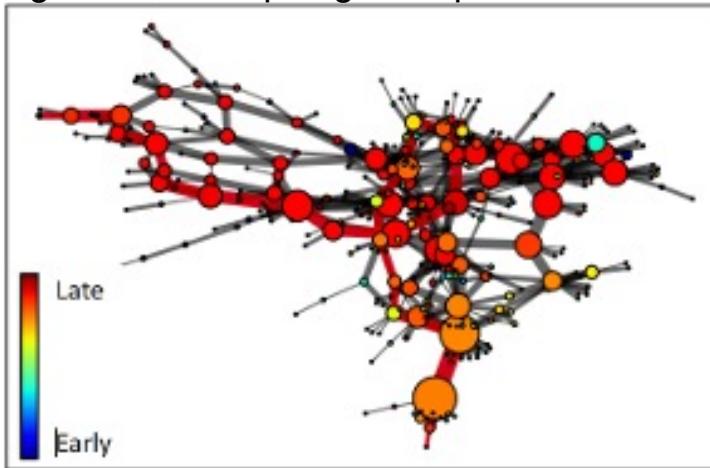


Fig. B Skeleton of the network

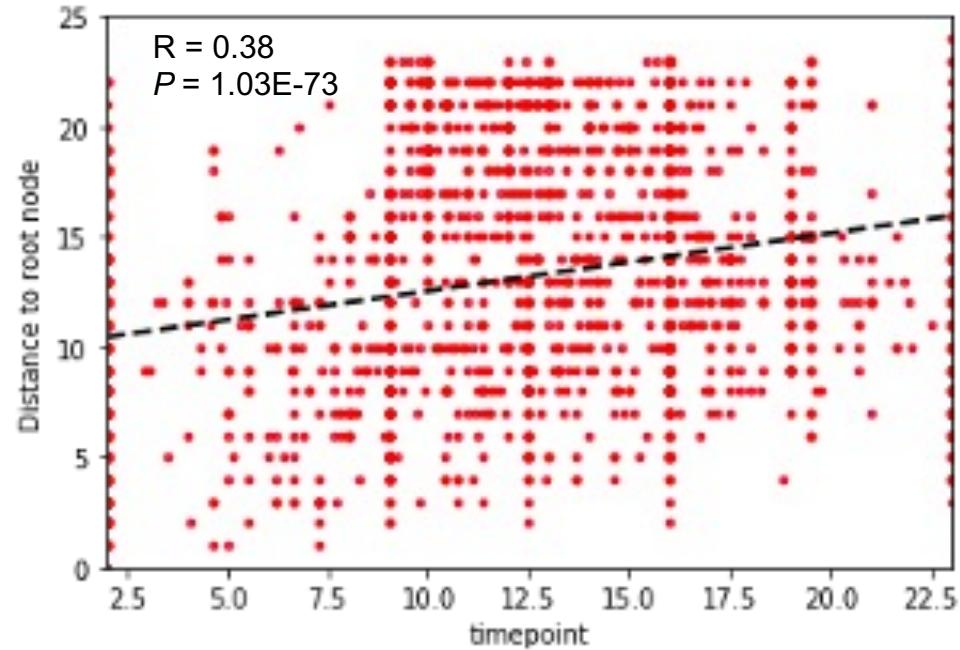


Fig. C Pseudo-time based on topological representation

Linear Regression Result:

- slope=0.3590
- Rvalue=0.3789
- Pvalue=1.0292e-73

However, further work on data engineering may lead to better construction of the topological representation.

# 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.
- 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

- 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.
- 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).
- 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.
- Hastie, T. J. (2017). Generalized additive models. *Statistical Models in S*, 249–307.
- 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.