

# Trajectory inference

Mohammed Charrouf

Leiden Computational Biology Center, LUMC  
Delft Bioinformatics Lab, TU Delft

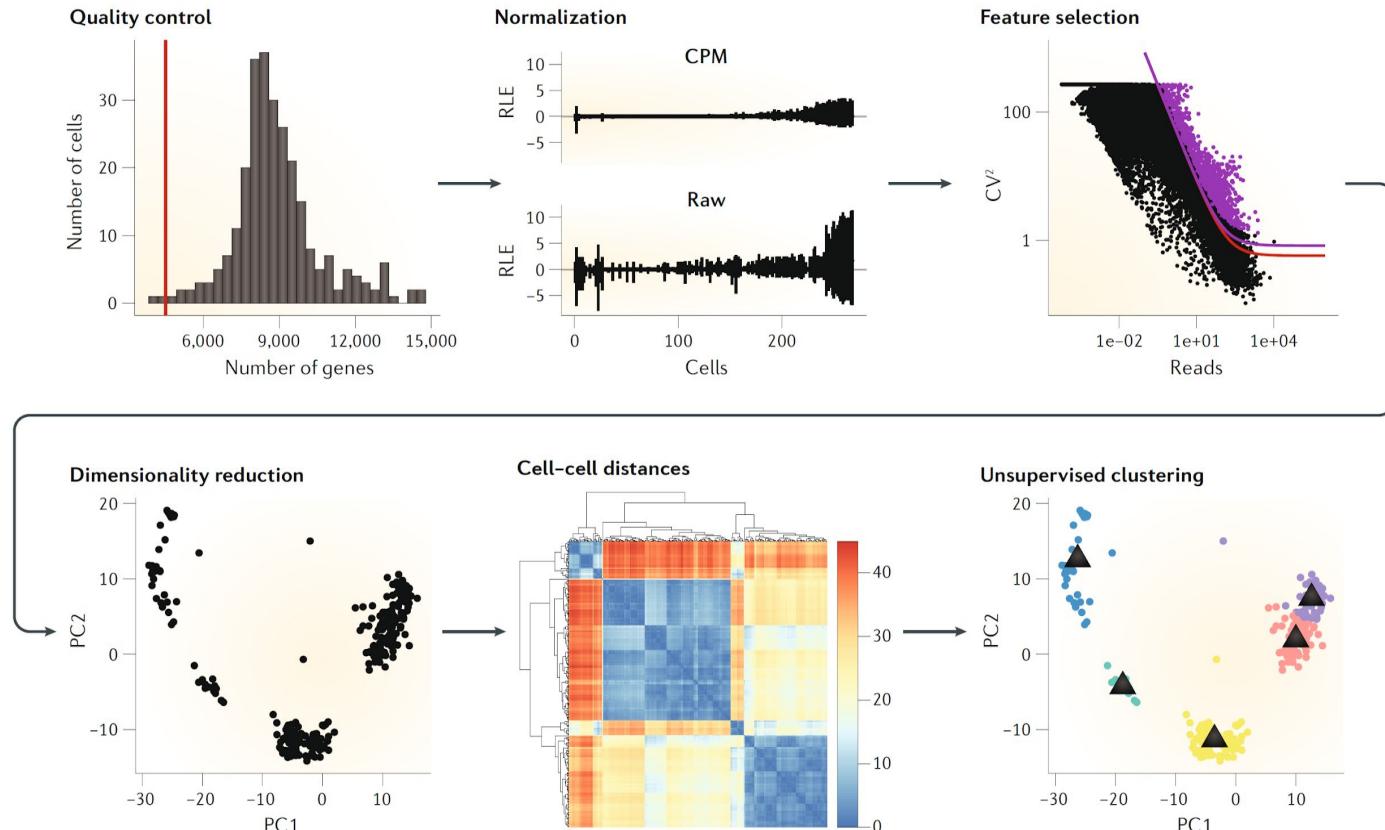


Leiden  
Computational Biology Center

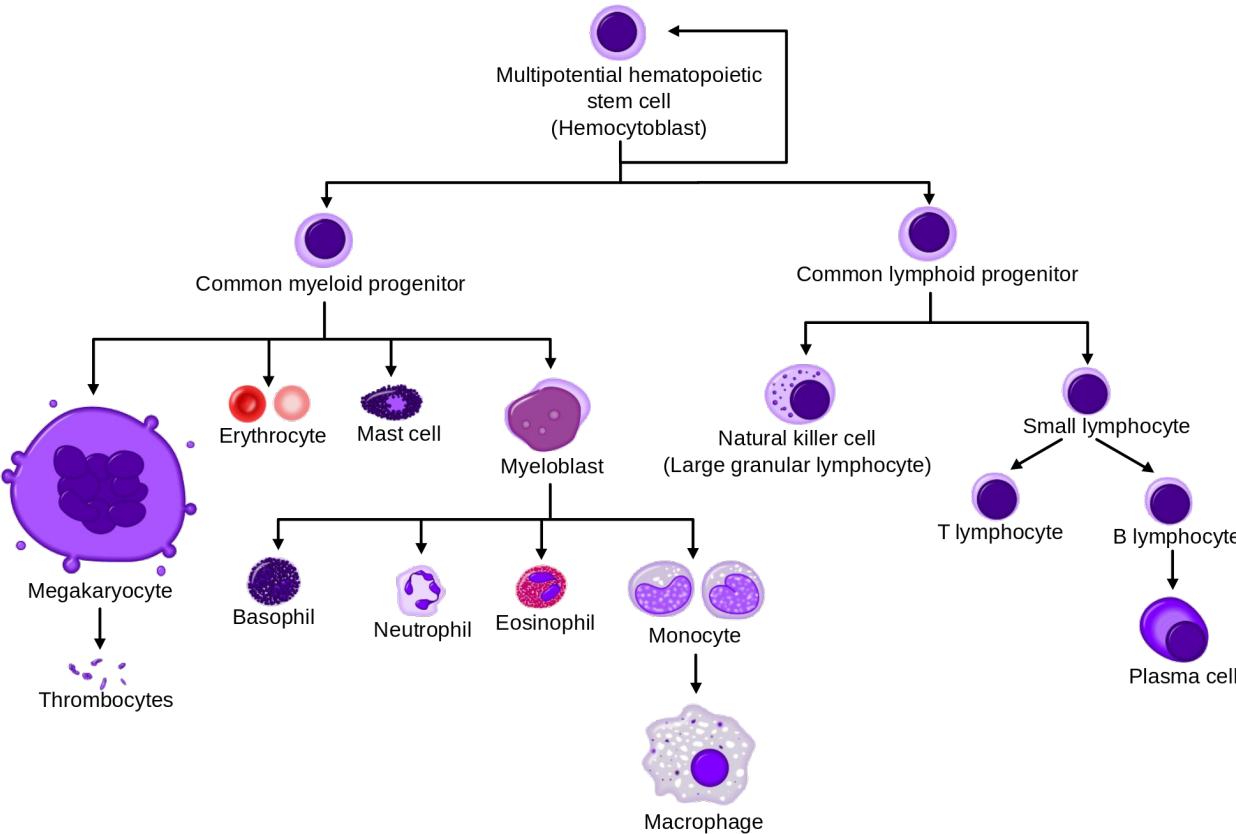


# Single cell RNA-seq workflow

Kiselev et al. (<https://doi.org/10.1038/s41576-018-0088-9>)

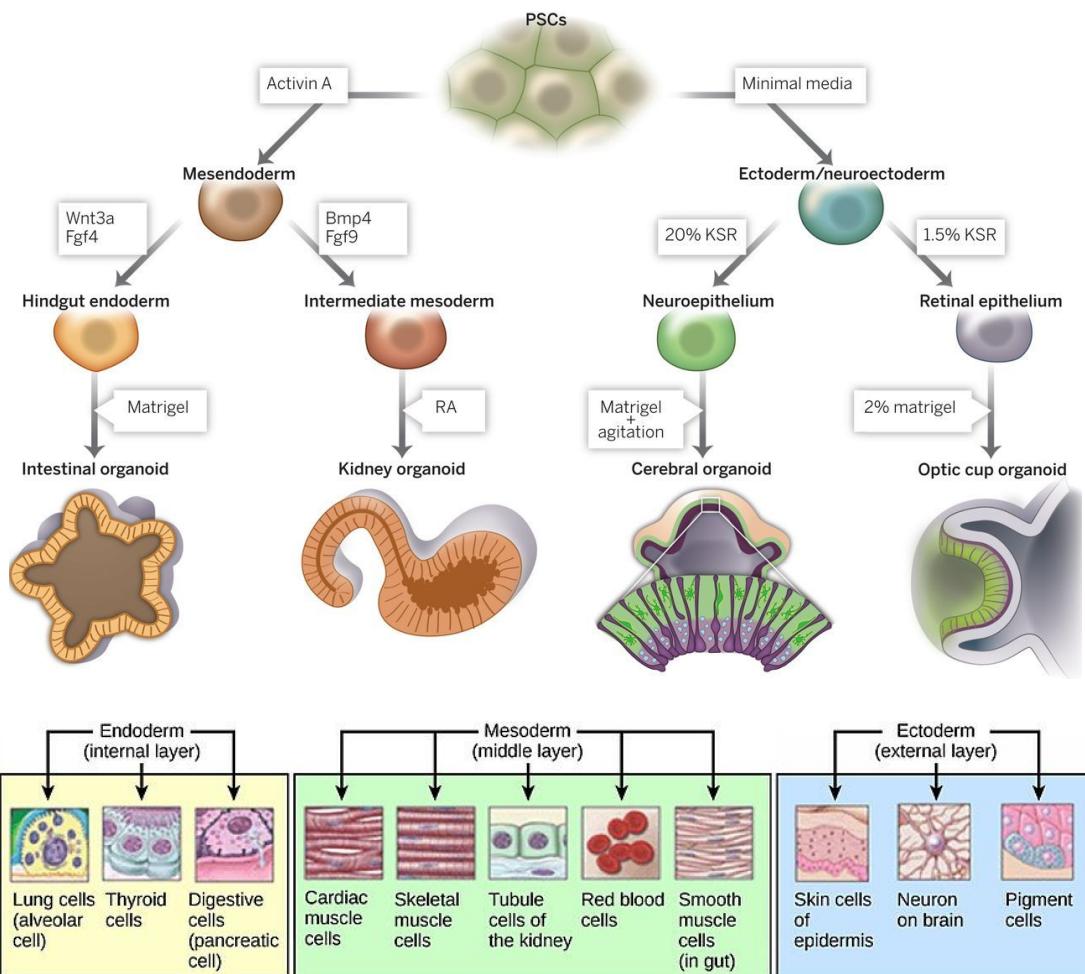


# Differentiating cells

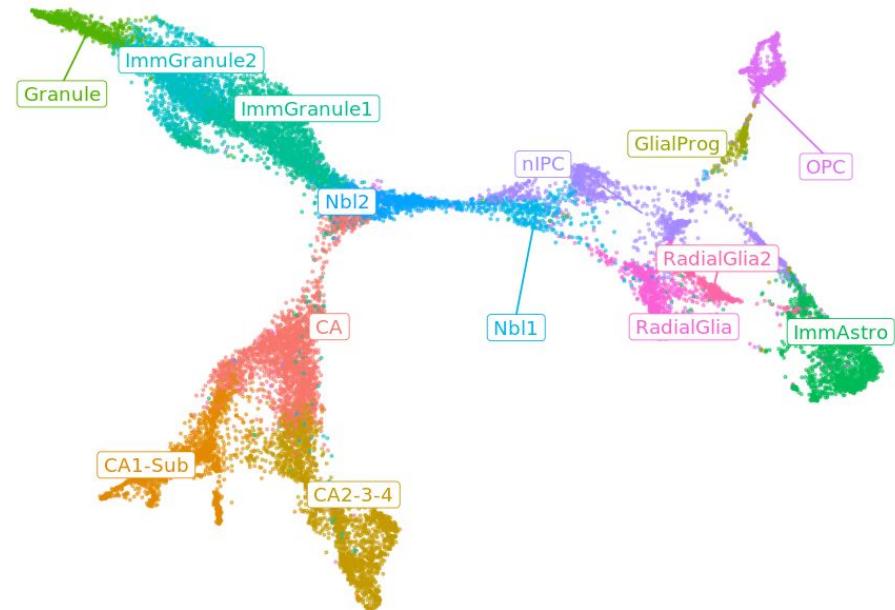
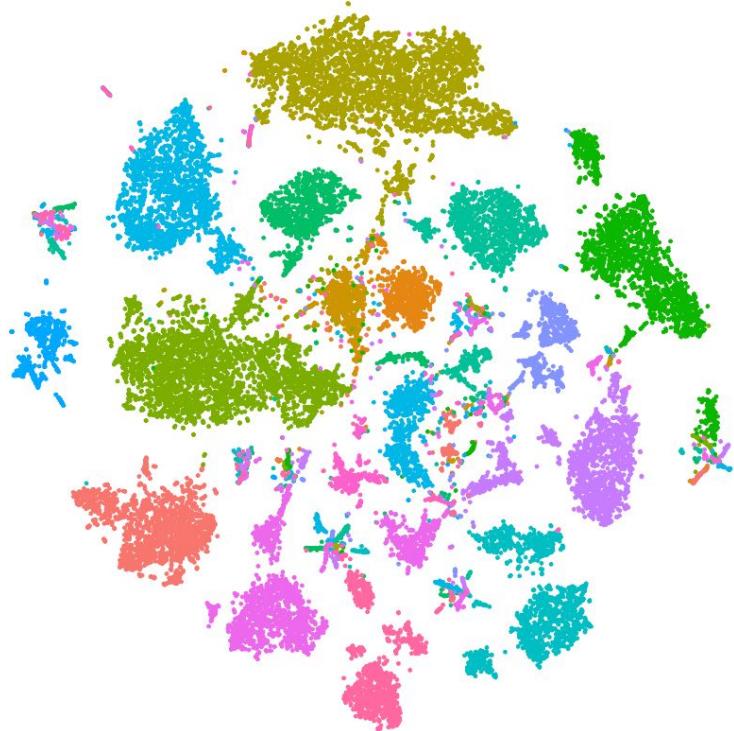


# Differentiating cells

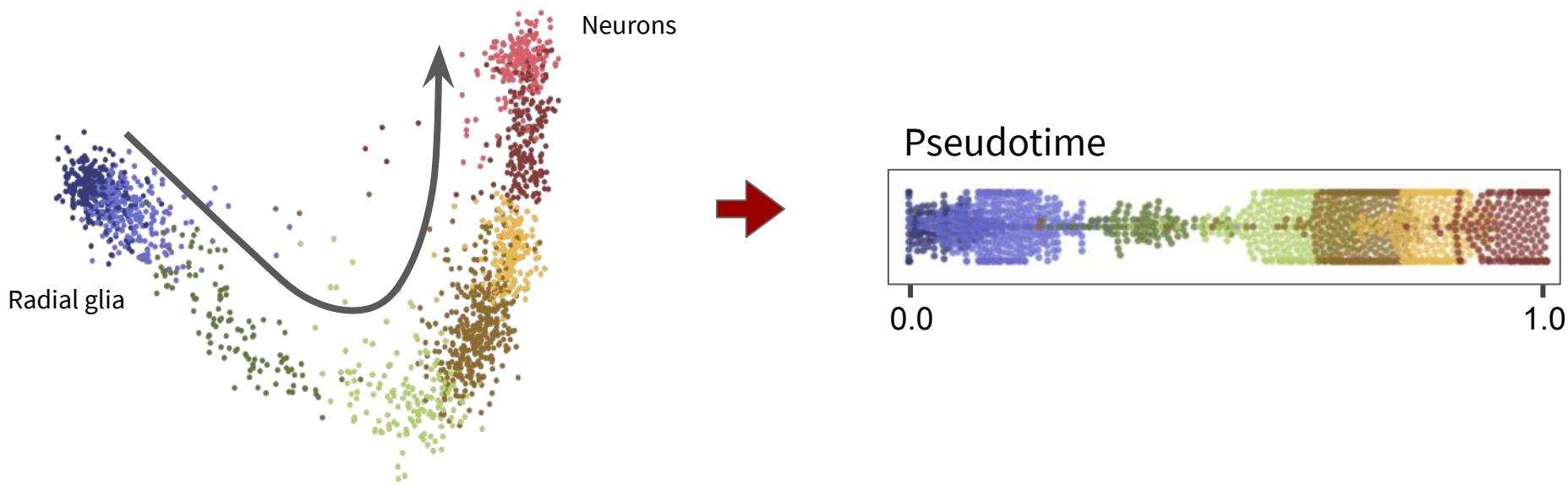
Lancaster et al.  
(<https://doi.org/10.1126/science.1247125> )



# Clustering of differentiating cells

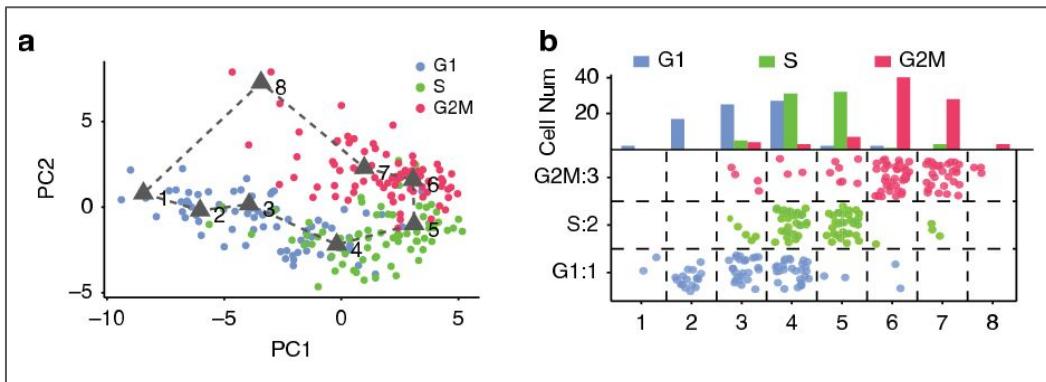
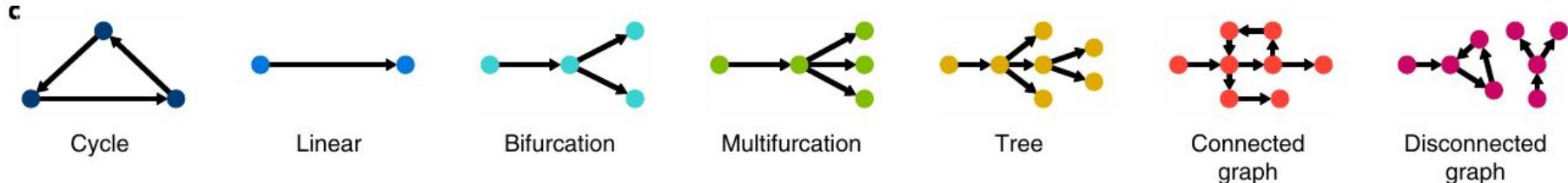


# Trajectory inference / pseudotime inference



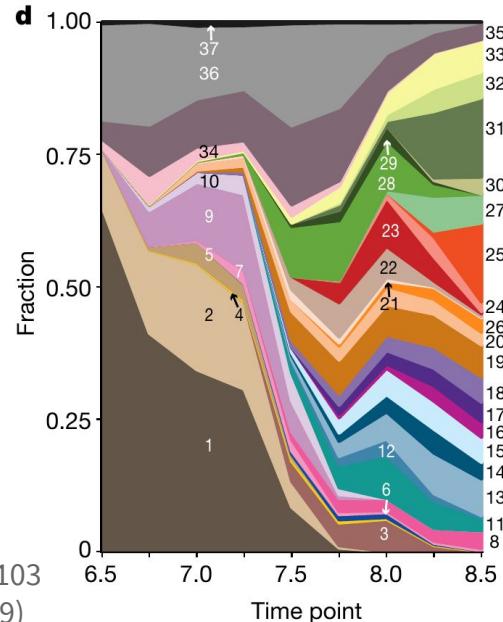
# Trajectory structure

Saelens et al. (<https://doi.org/10.1038/s41587-019-0071-9>)



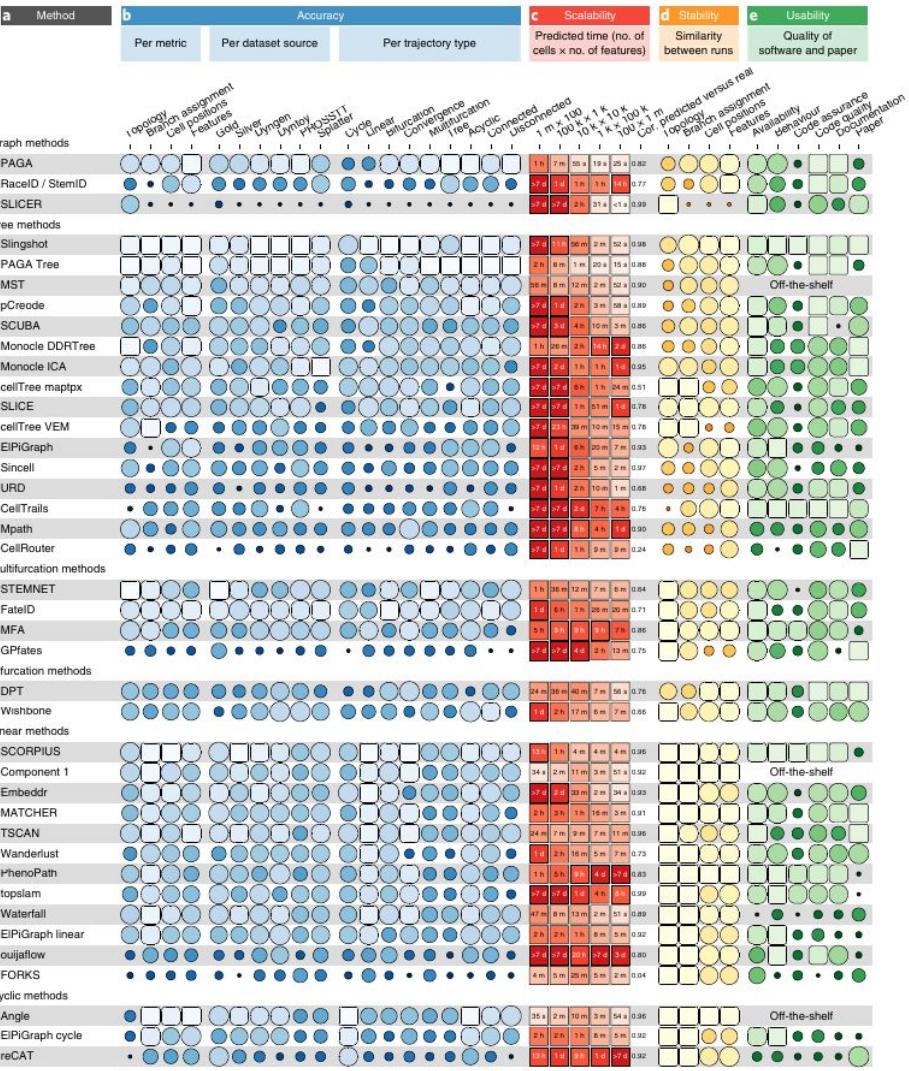
Liu et al. (<https://doi.org/10.1038/s41467-017-00039-z>)

Pijuan-Sala et al.  
(<https://doi.org/10.1038/s41586-019-0933-9>)



# Many many trajectory inference methods

- Comprehensive evaluation by Saelens et al.
  - Accuracy (per metric, data source, trajectory type)
  - Run time per # cells
  - Usability
  - And way more
- Interactive website to select best tool for your data
  - <http://guidelines.dynverse.org/>



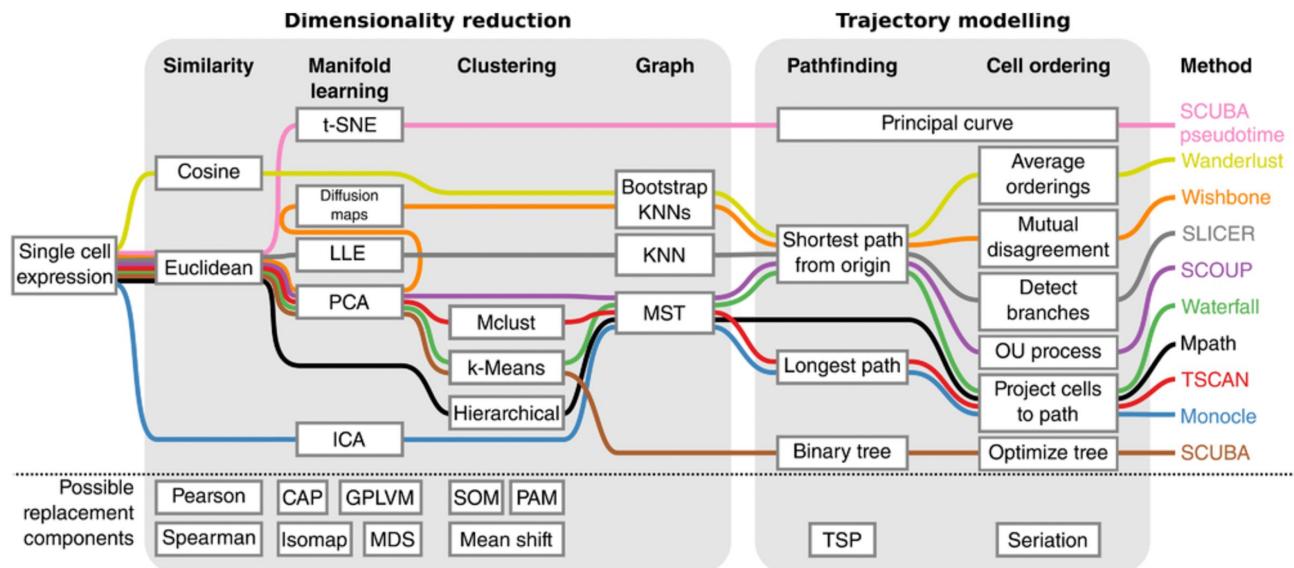
# Trajectory inference methods

Three methods will be shown

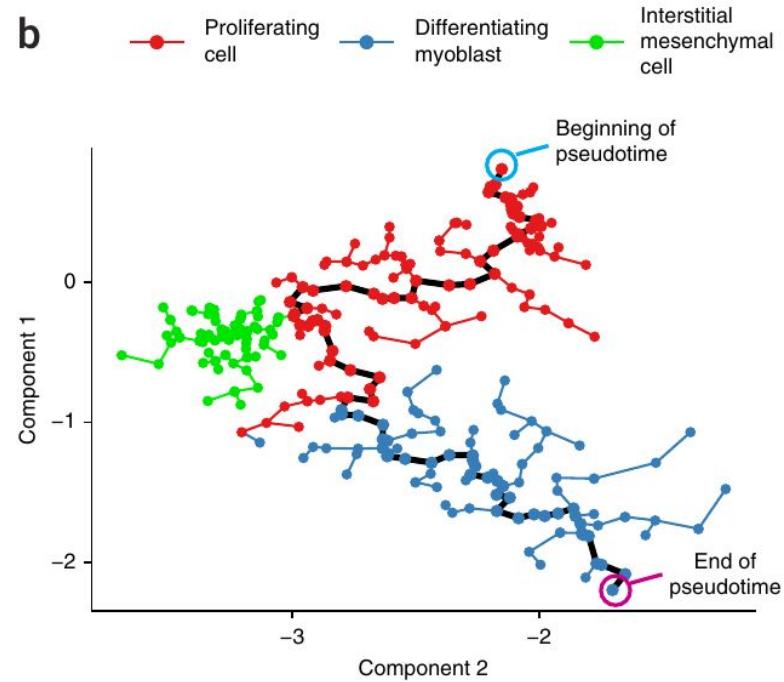
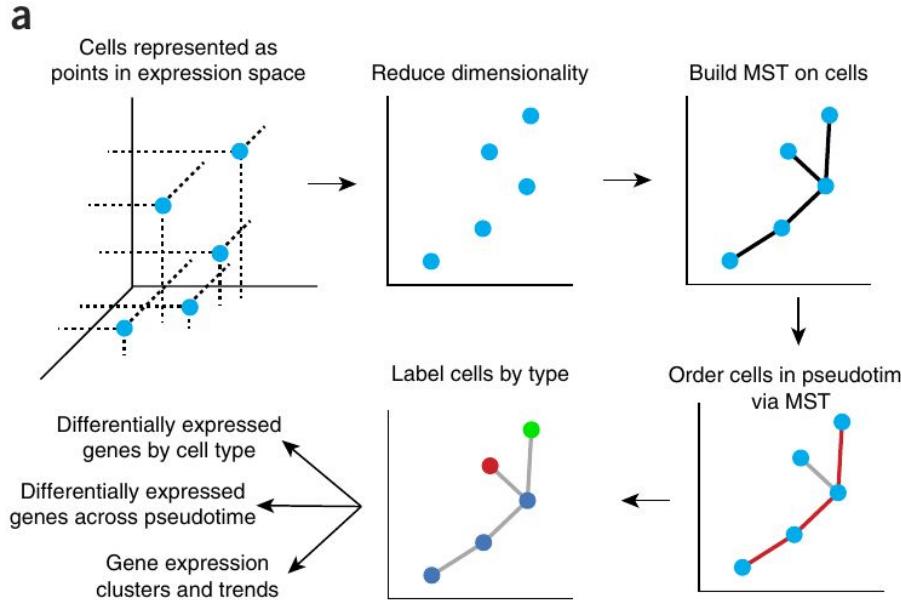
- “General method”, exemplified with **Slingshot** Street et al., 2018 (doi: [10.1186/s12864-018-4772-0](https://doi.org/10.1186/s12864-018-4772-0))
- **Monocle 2** Cole et al., 2017 (doi: [10.1038/nmeth.4402](https://doi.org/10.1038/nmeth.4402))
- **Ouija** Campbell et al., 2018

# General trajectory inference pipeline

1. Dimensionality reduction
2. Trajectory fitting
3. Pseudotime assignment

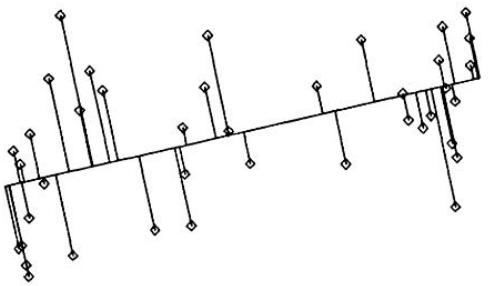


# Minimum spanning trees



# Principle curves

b Principal component



d Principal curve

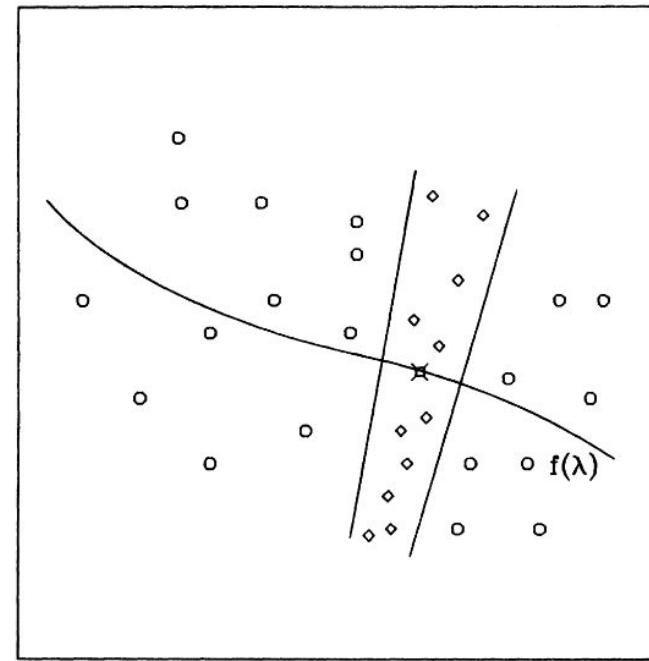
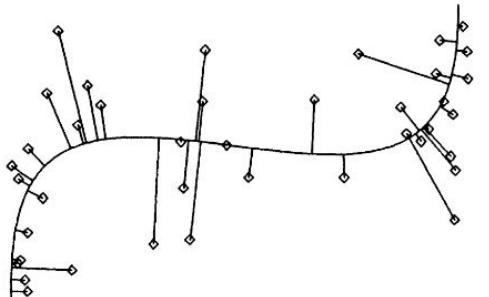
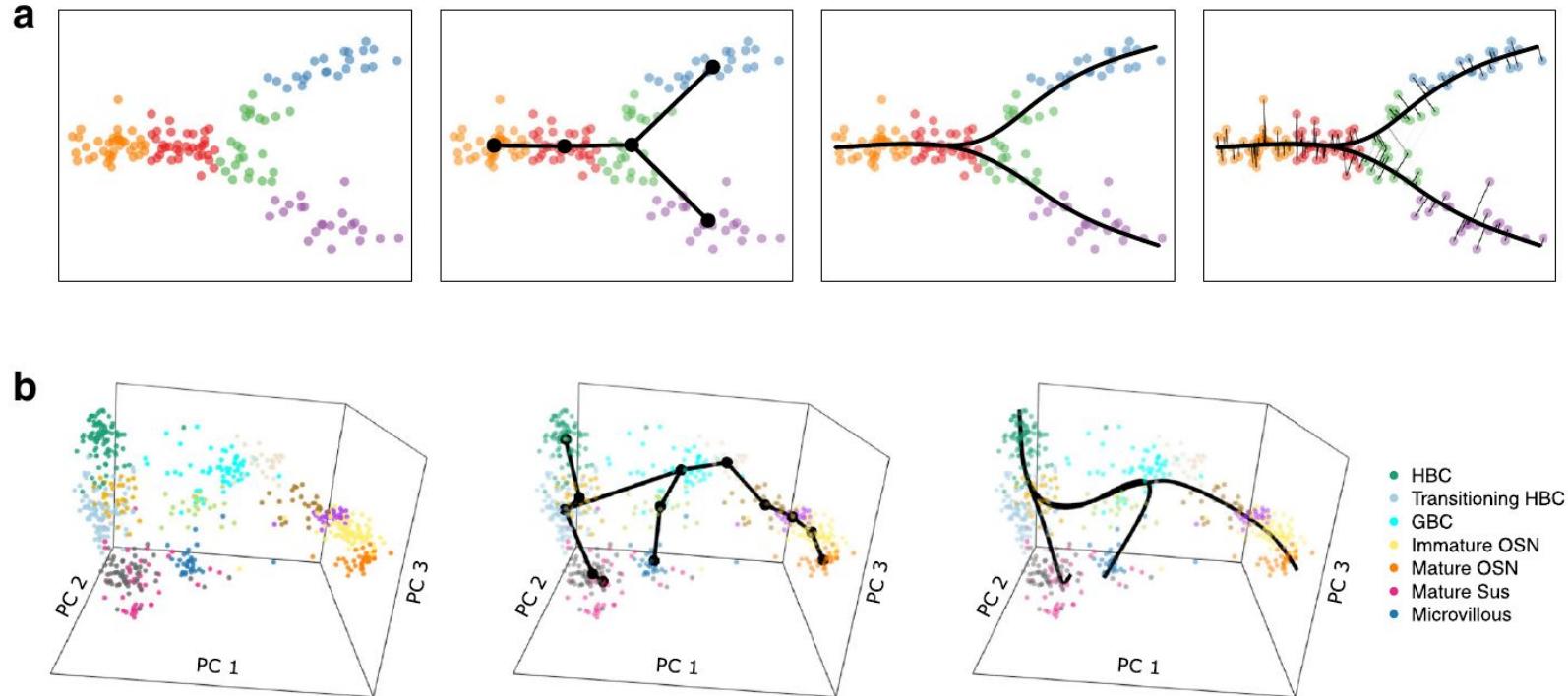


Figure 3. Each point on a principal curve is the average of the points that project there.

Hastie, Stuetzle, Principal curves, 1989

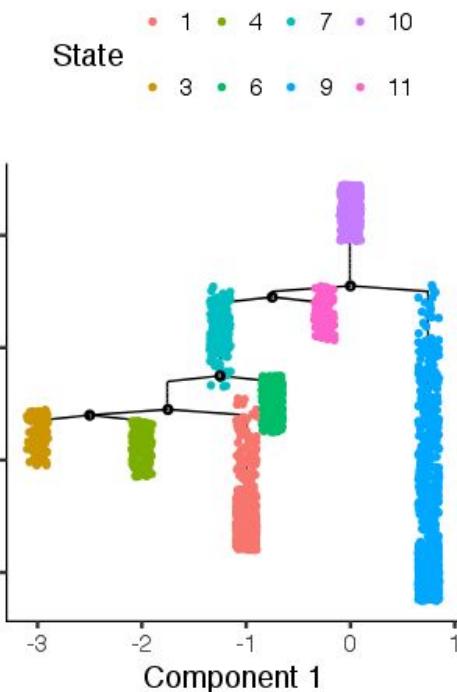
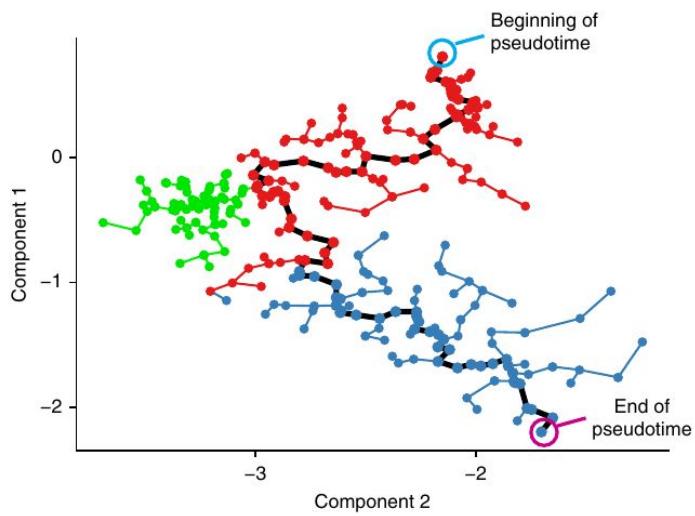
# Slingshot

Fit multiple principle curves simultaneously, ensuring a shared trunk



# Monocle 2

- Successor to Monocle 1
- End goal: Fit any arbitrary graph on the data

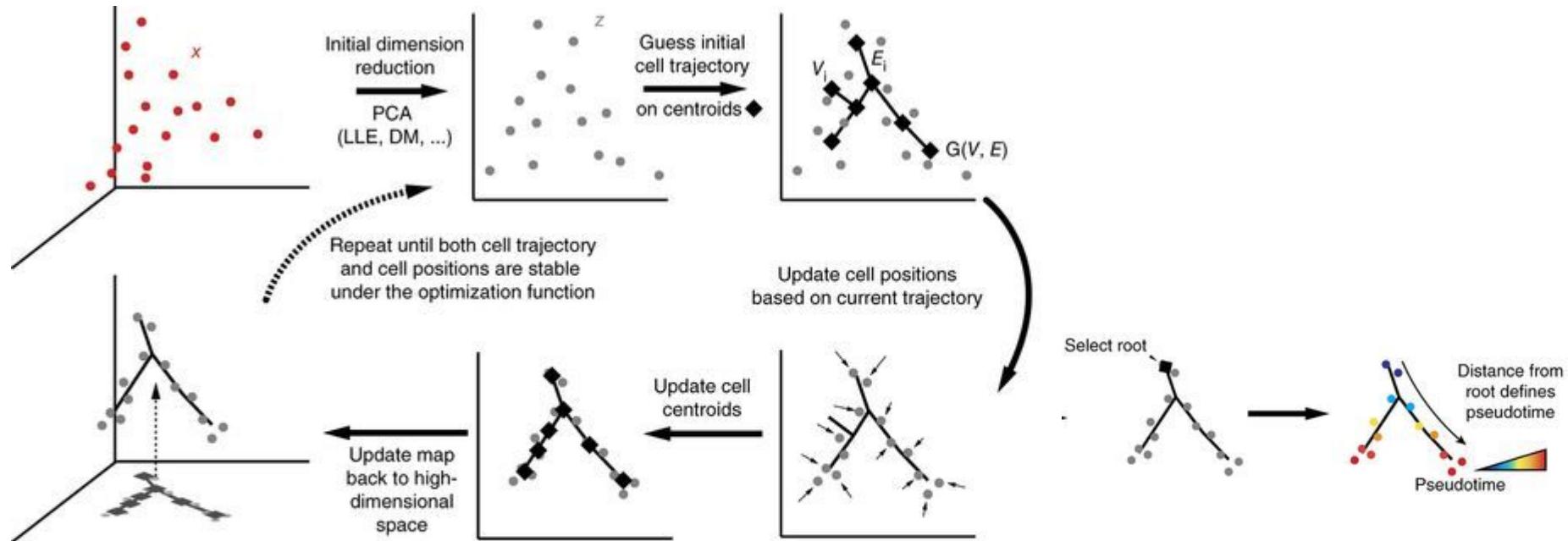


# Monocle 2

Optimize fit function

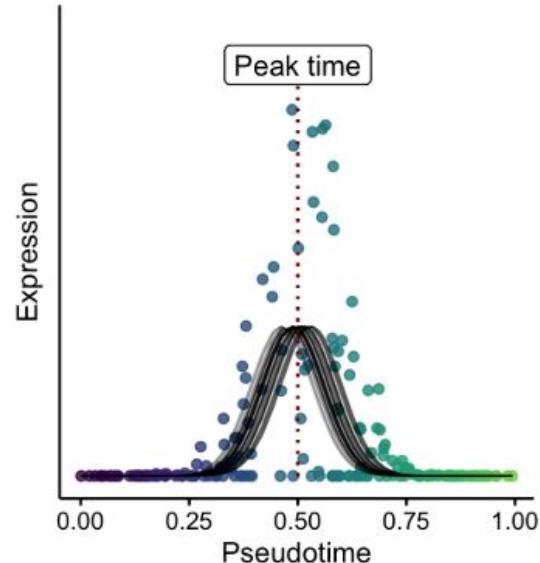
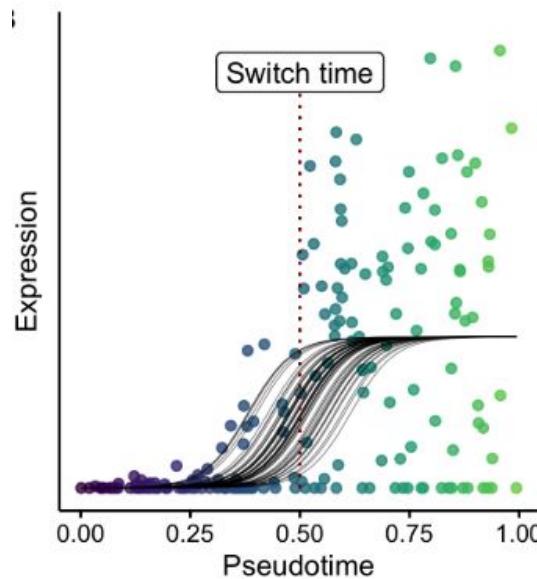
$$\min_{W, B, R, Y, Z} \sum_{i=1}^N \|x_i - Wz_i\|^2 + \frac{\lambda}{2} \sum_{k,k'} b_{k,k'} \|Wy_k - Wy_{k'}\|^2$$

$$+ \gamma \left[ \sum_{k=1}^K \sum_{i=1}^N r_{i,k} (\|z_i - y_k\|^2 + \sigma \log r_{i,k}) \right]$$



# Ouija

- Model a small set of marker genes instead of fitting trajectory on complete transcriptome
- Switch focus to interpretability



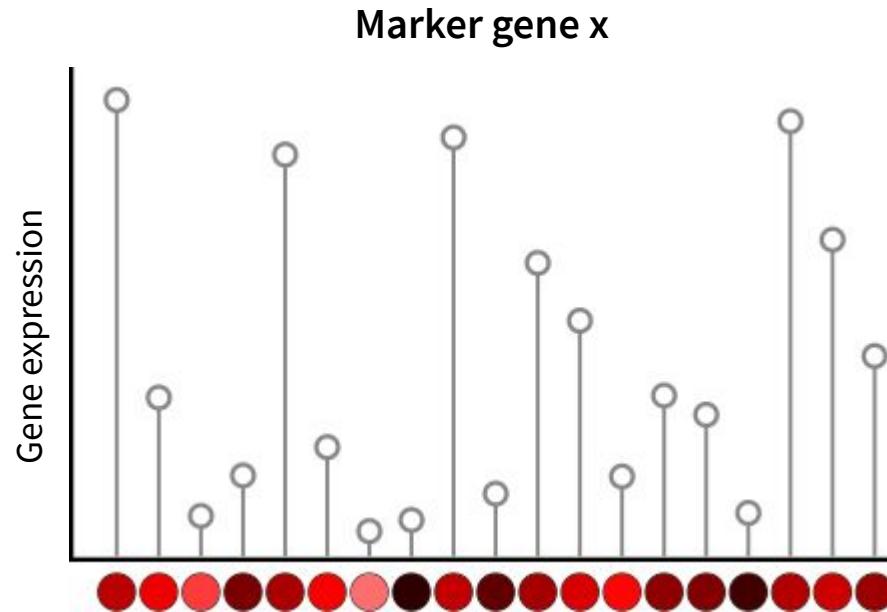
# Ouija intuition

True ordering:



Random cell ordering:

Goodness-of-fit: **low**



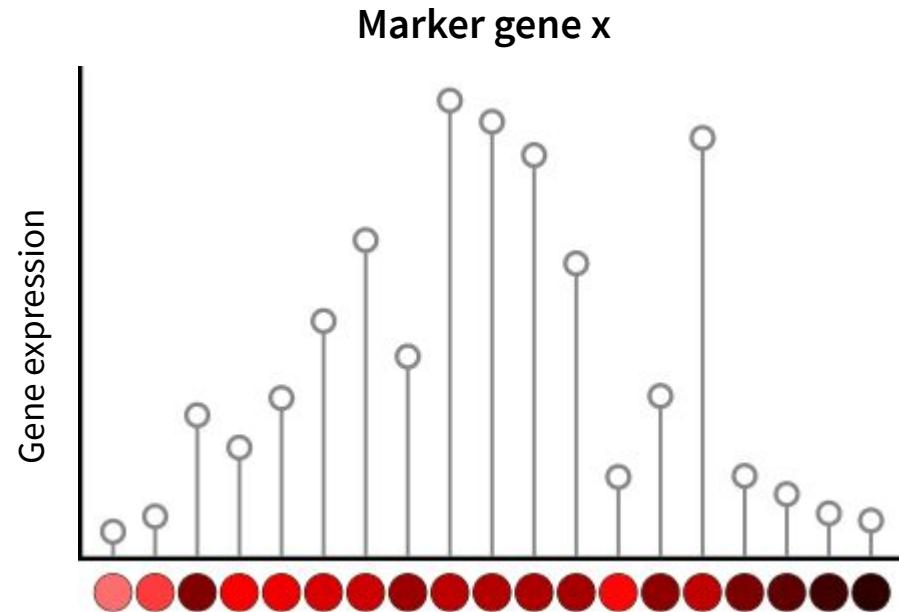
# Ouija intuition

True ordering:



Optimize iteration: 100

Goodness-of-fit: mid



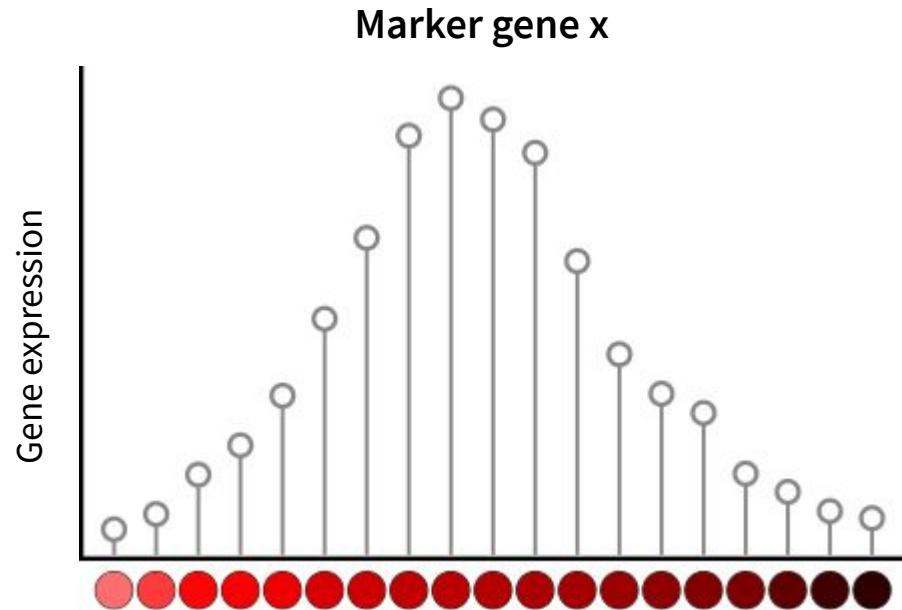
# Ouija intuition

True ordering:

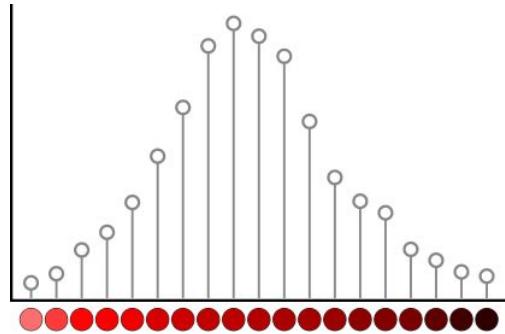


Optimize iteration: 500

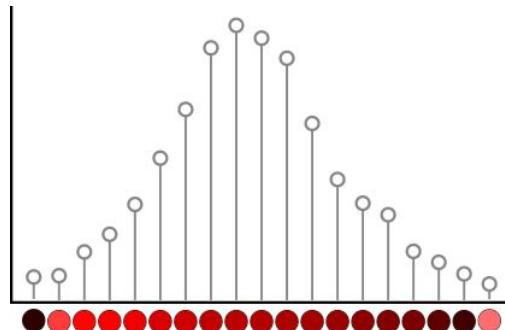
Goodness-of-fit: high



## Marker gene x

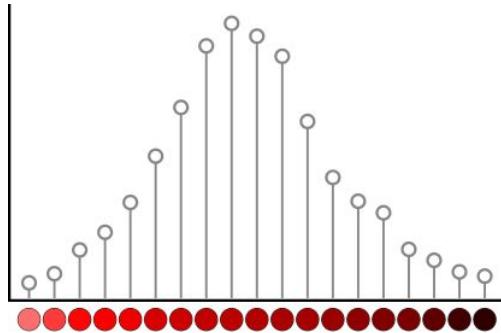


Goodness-of-fit: **high**



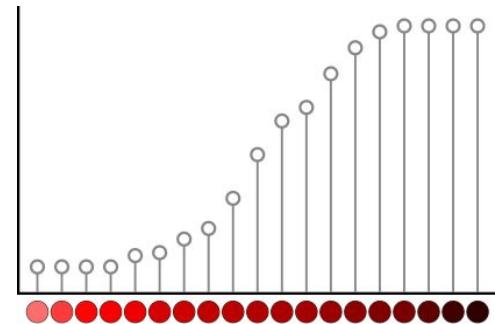
Goodness-of-fit: **high**

Marker gene x

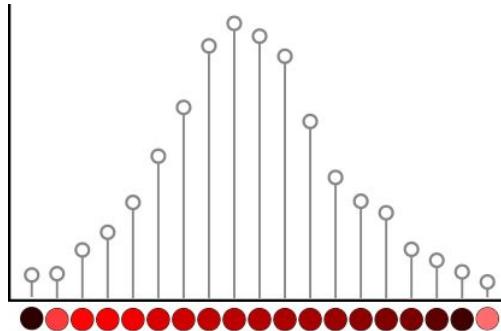


Goodness-of-fit: **high**

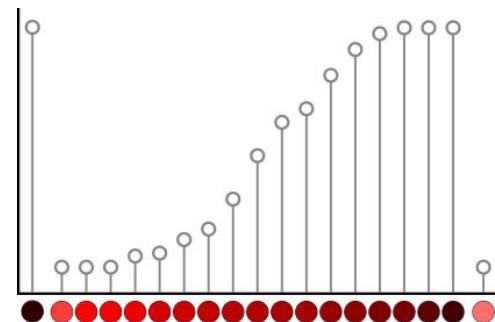
Marker gene y



Goodness-of-fit: **high**



Goodness-of-fit: **high**

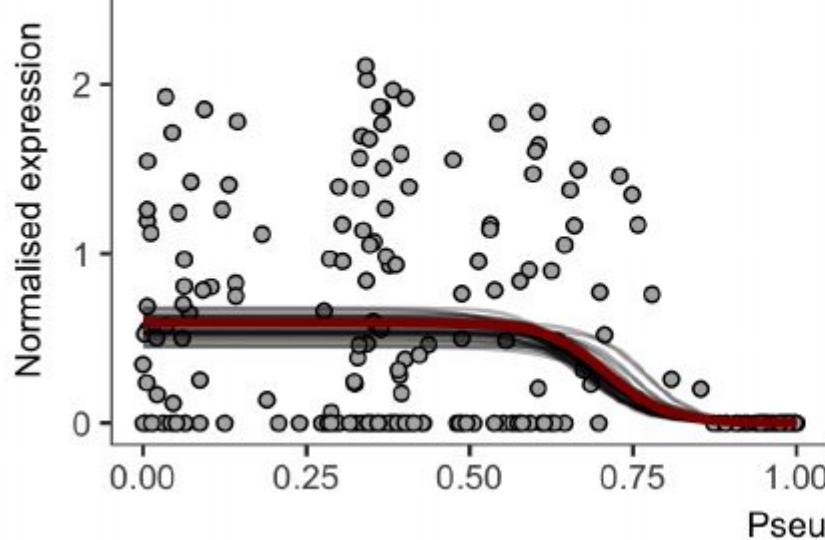


Goodness-of-fit: **mid**

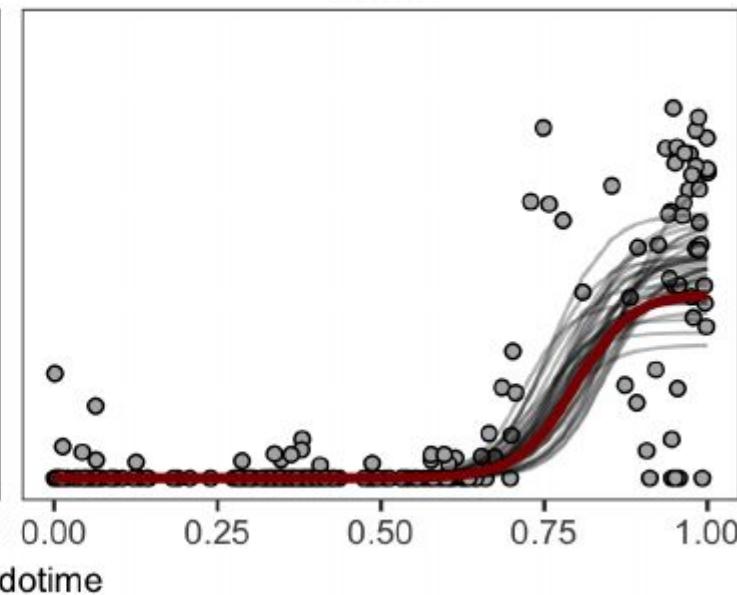
# Ouija probabilistic modelling

C

ID1

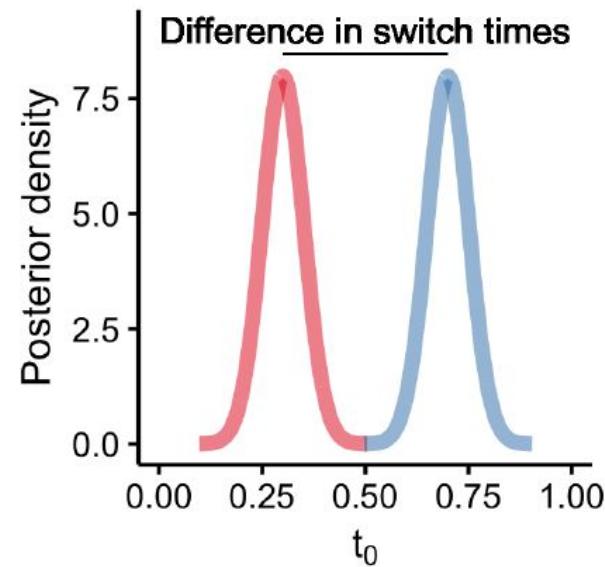
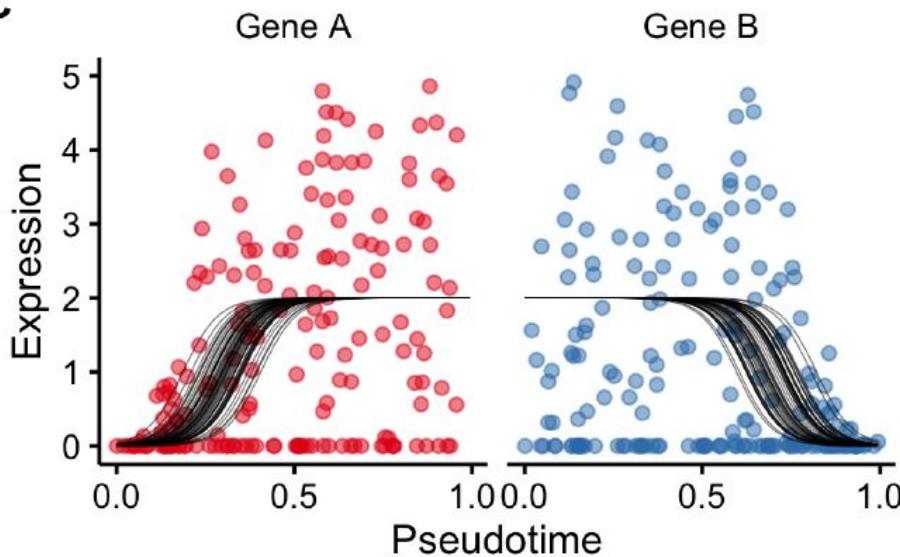


MYOG

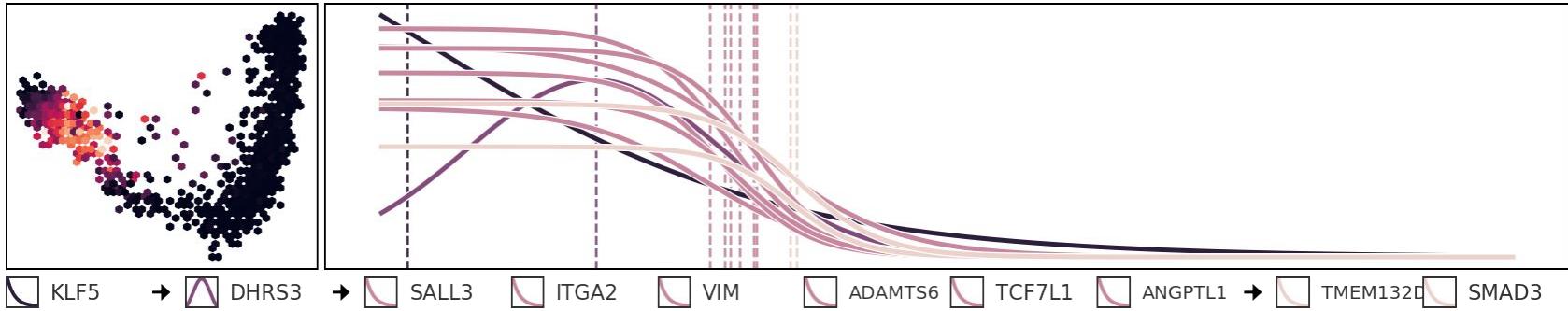


# Ouija probabilistic pseudotime assignment

C



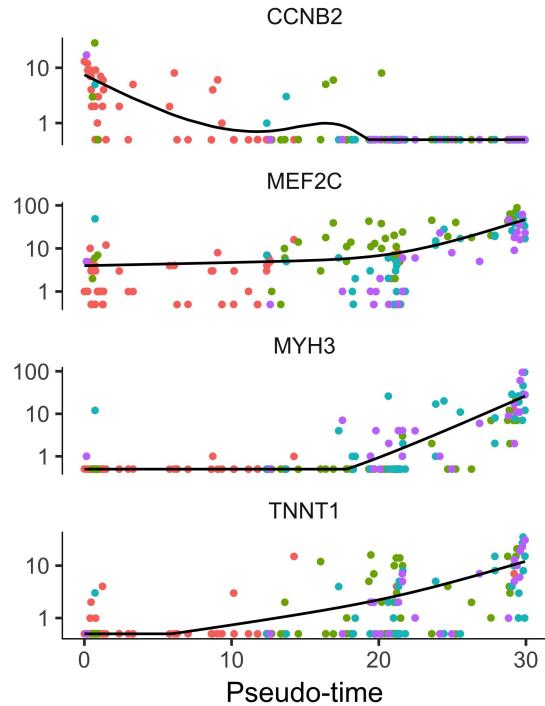
# Gene ordering



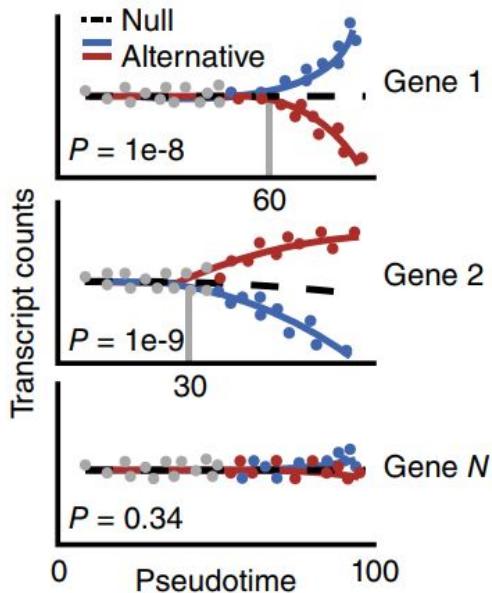
- KLF5 member of Kruppel-like family of transcription factors
  - Repressor of neurite growth, down-regulation linked to cell cycle arrest
- VIM, highly variable gene
  - Known marker of gliogenesis

# Pseudotime analysis

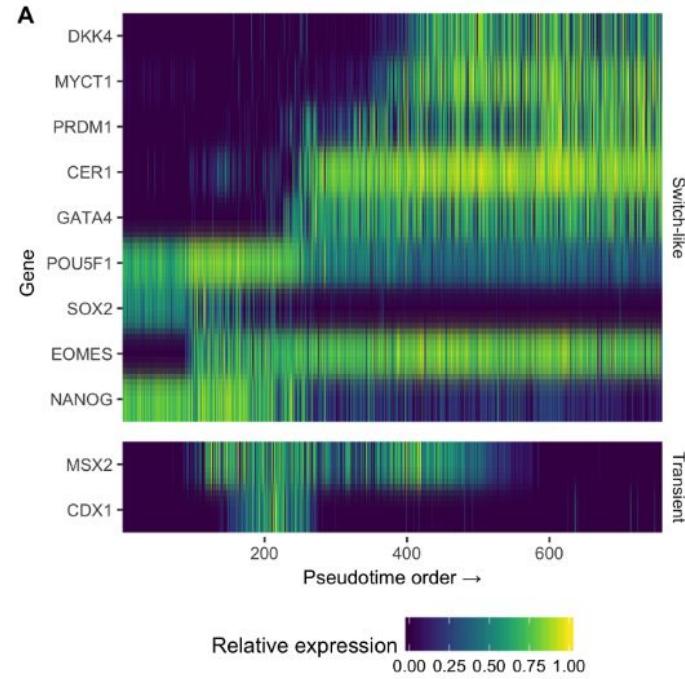
## Pseudotime-gene relation



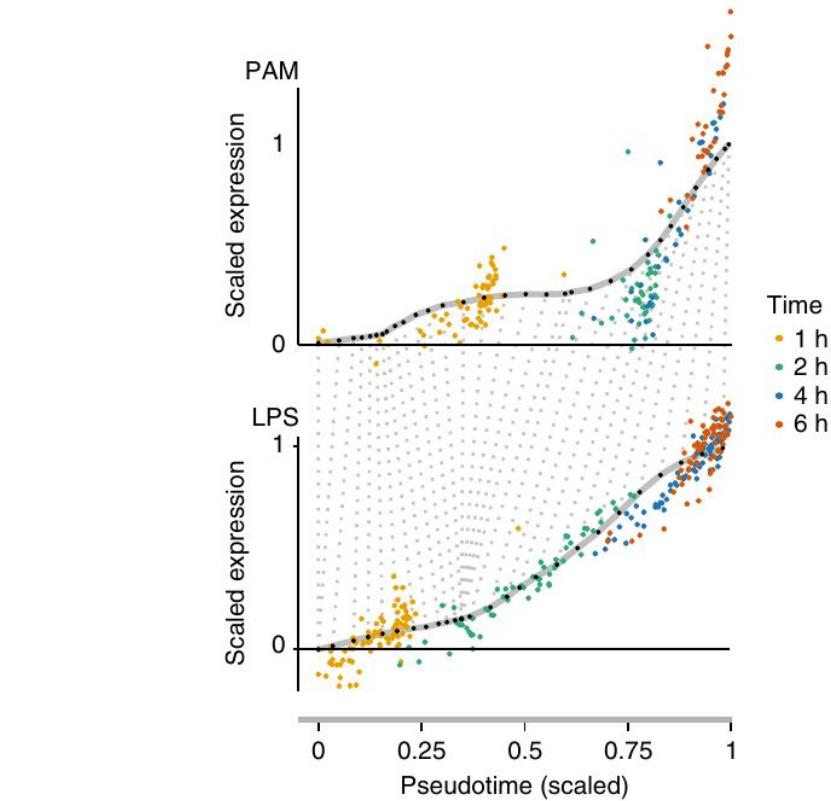
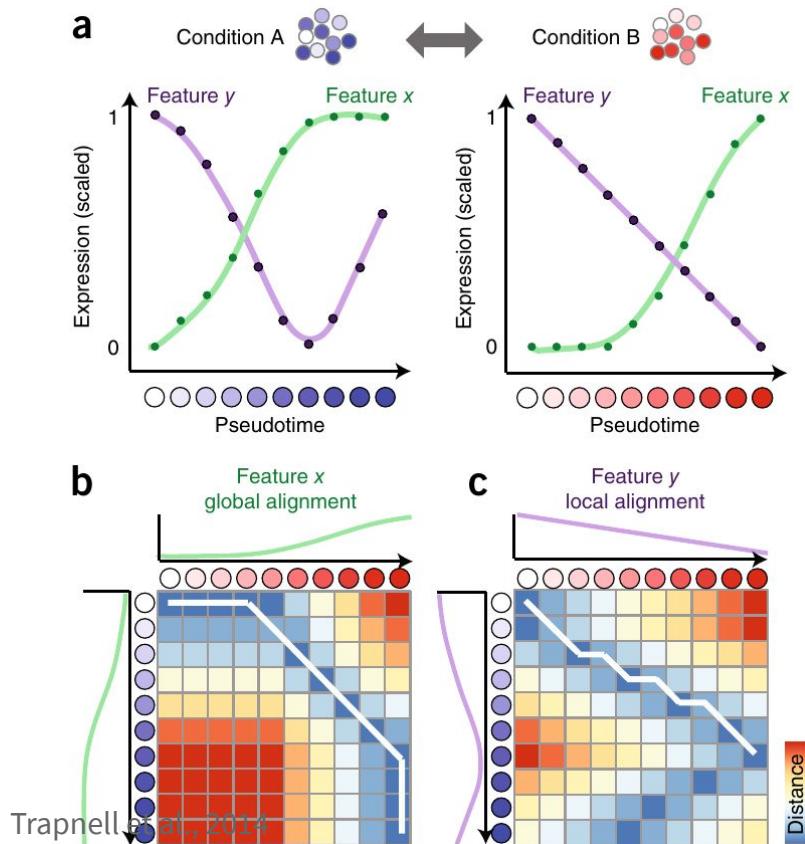
## Detect branching genes



## Interpretation of gene behaviour



# Comparing trajectories



“CellAlign”, Alpert et al., 2018

# Is pseudotime inference a solved problem?

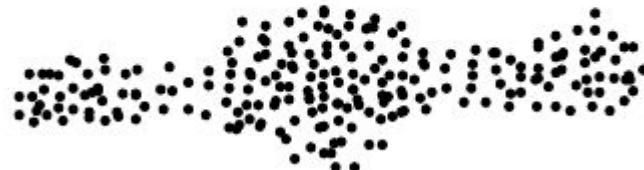
no.

## Fundamental limits on dynamic inference from single-cell snapshots

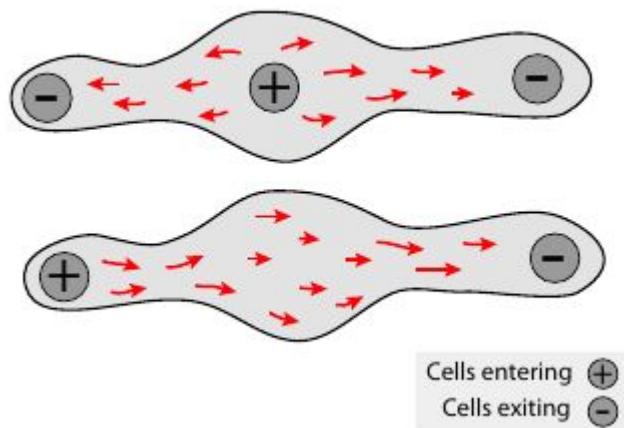
Caleb Weinreb<sup>a</sup>, Samuel Wolock<sup>a</sup>, Betsabeh K. Tusi<sup>b</sup>, Merav Socolovsky<sup>b</sup>, and Alon M. Klein<sup>a,1</sup>

“The general challenge, even with perfect data, is that many regulatory mechanisms can generate the same dynamic process, and many dynamic processes can give rise to the same distribution.”

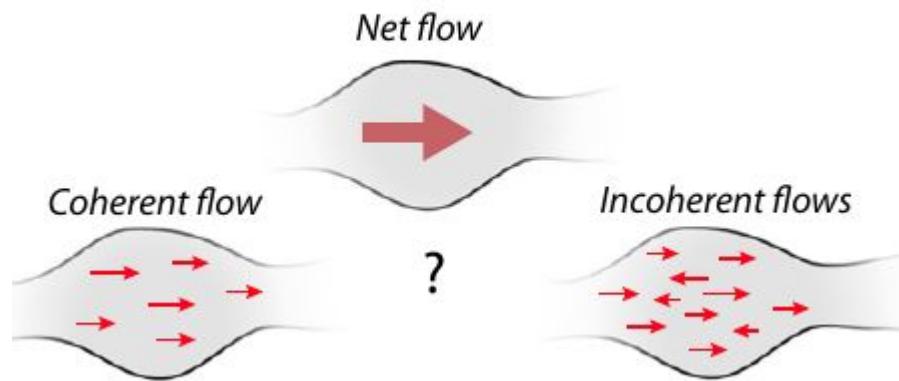
# No unique solution



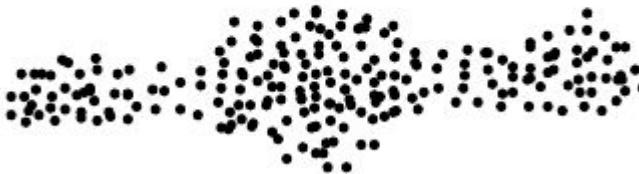
Entry and exit points direct the flow of cells



Net velocity may not equal actual velocity

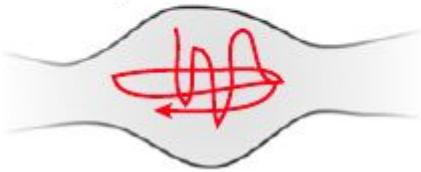


# No unique solution

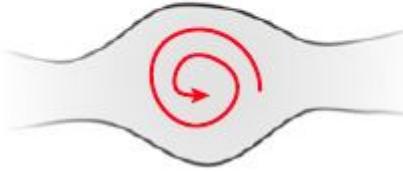


**Rotations in state space do not alter cell density**

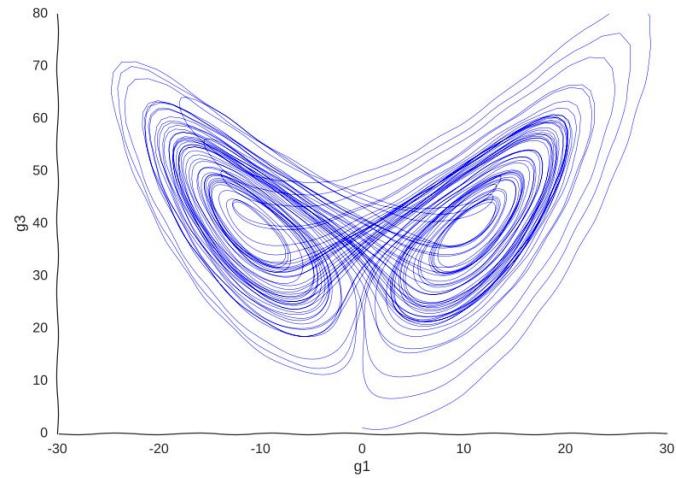
*Simple fluctuations*



*Periodic oscillations*

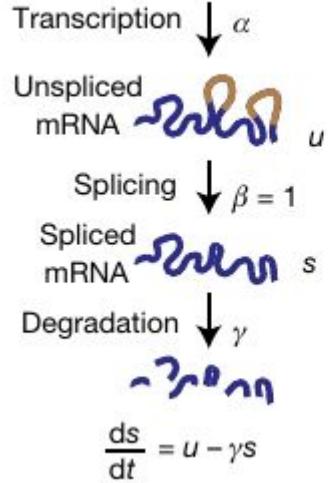


Two-state cyclical transition

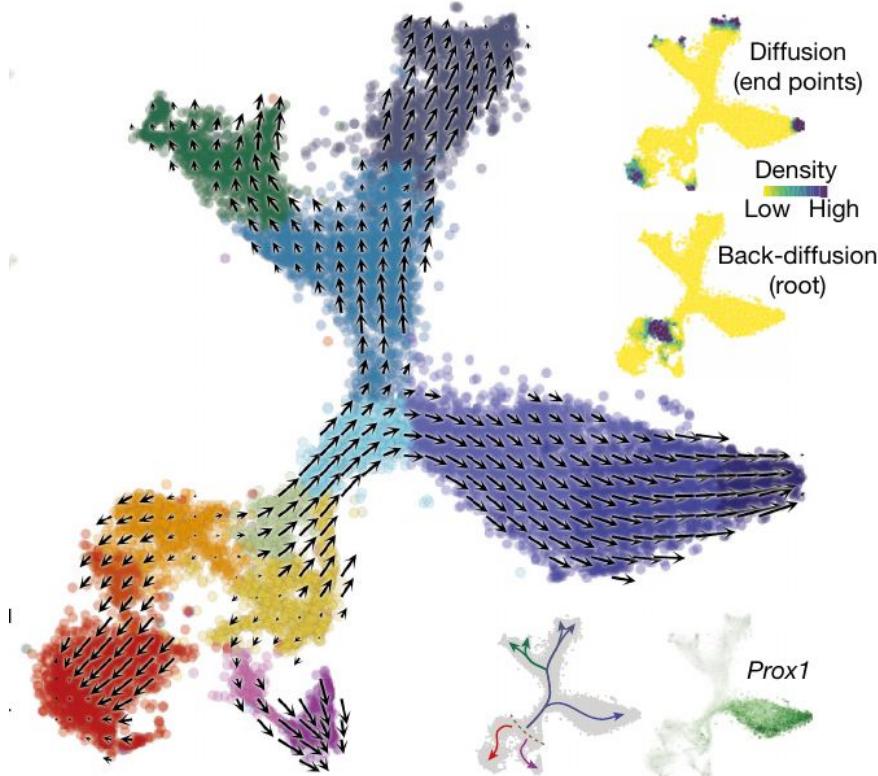
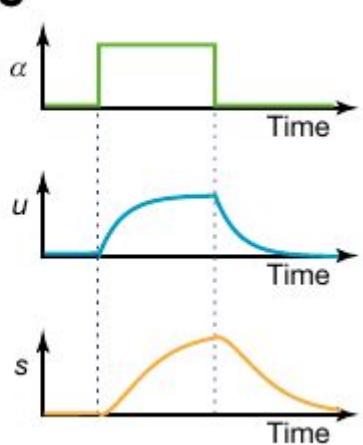


# RNA velocity of single cells

b

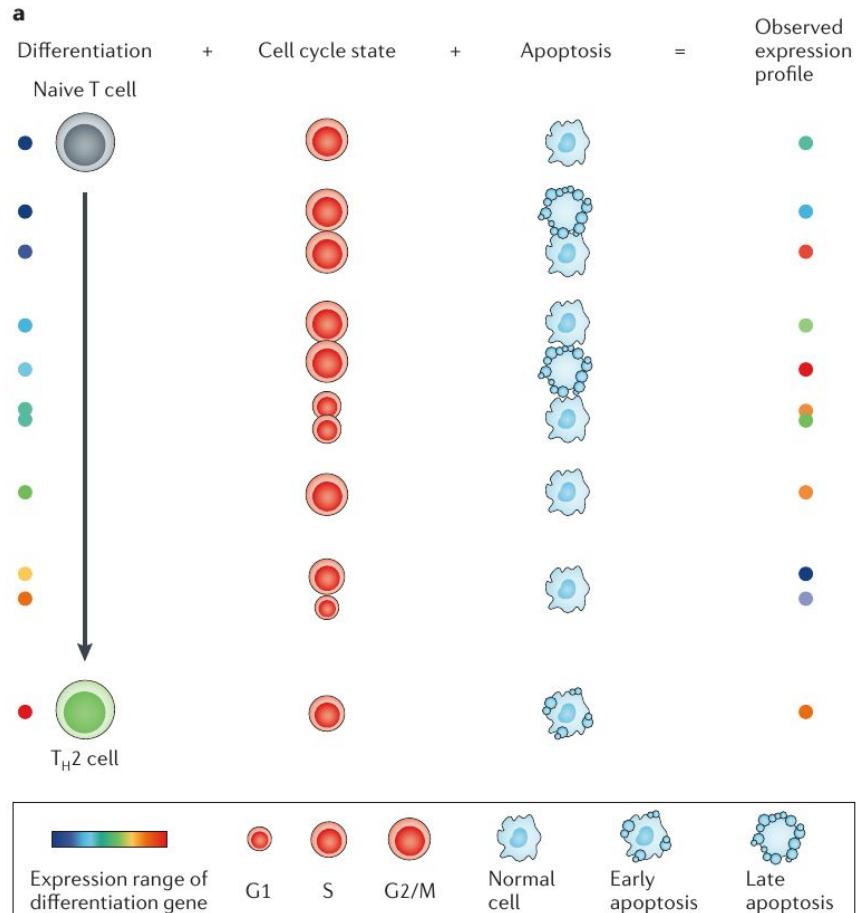


c



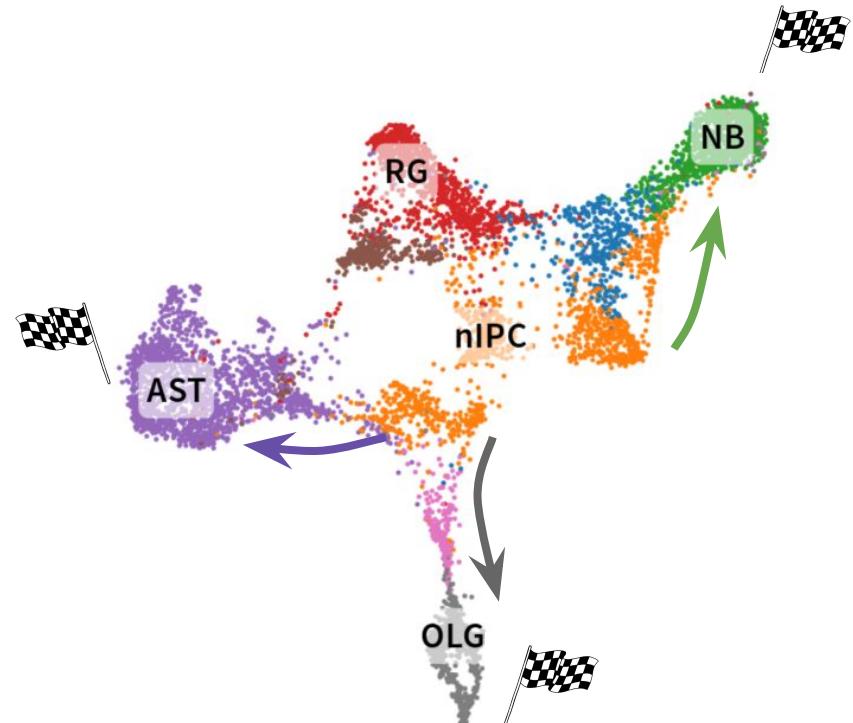
# Confounding sources of variation

What biological processes can lead to the measured transcriptional profile?

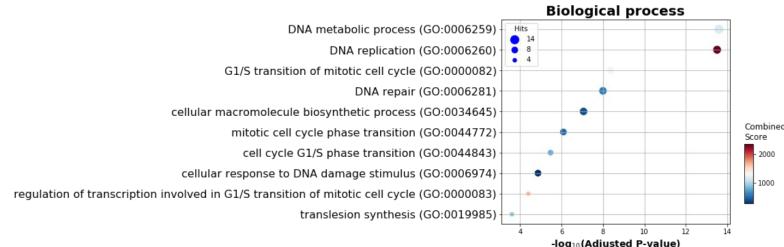
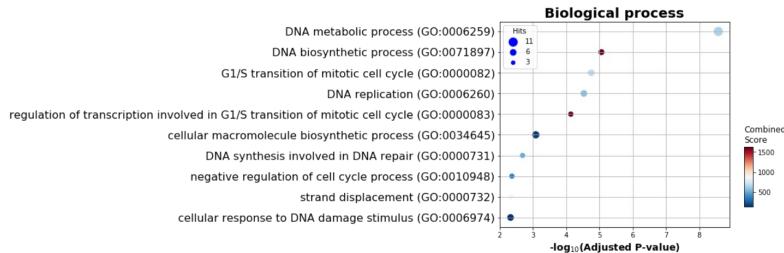
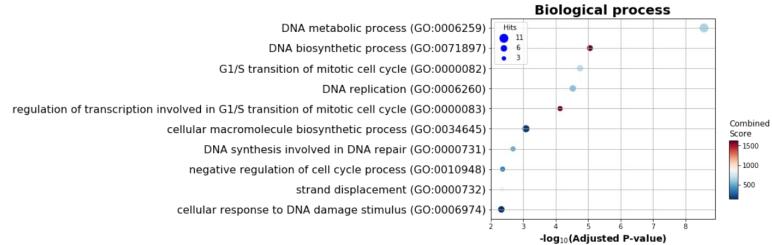
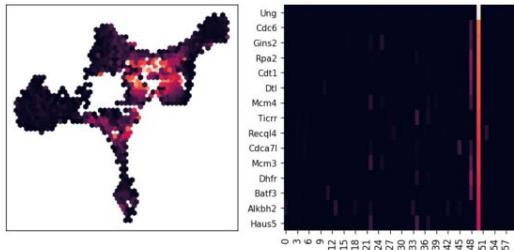
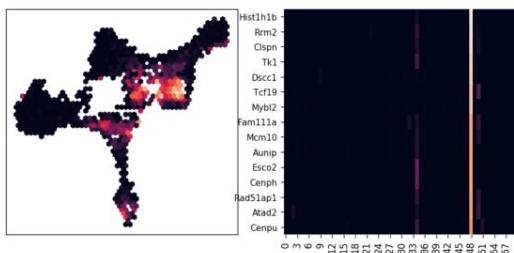
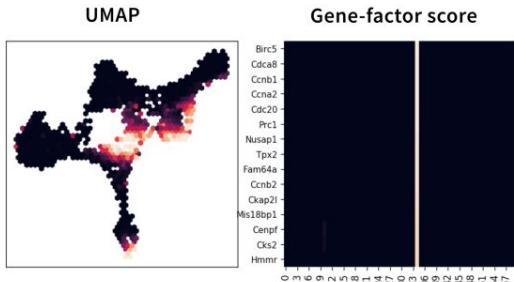


# Developing mouse hippocampus dataset

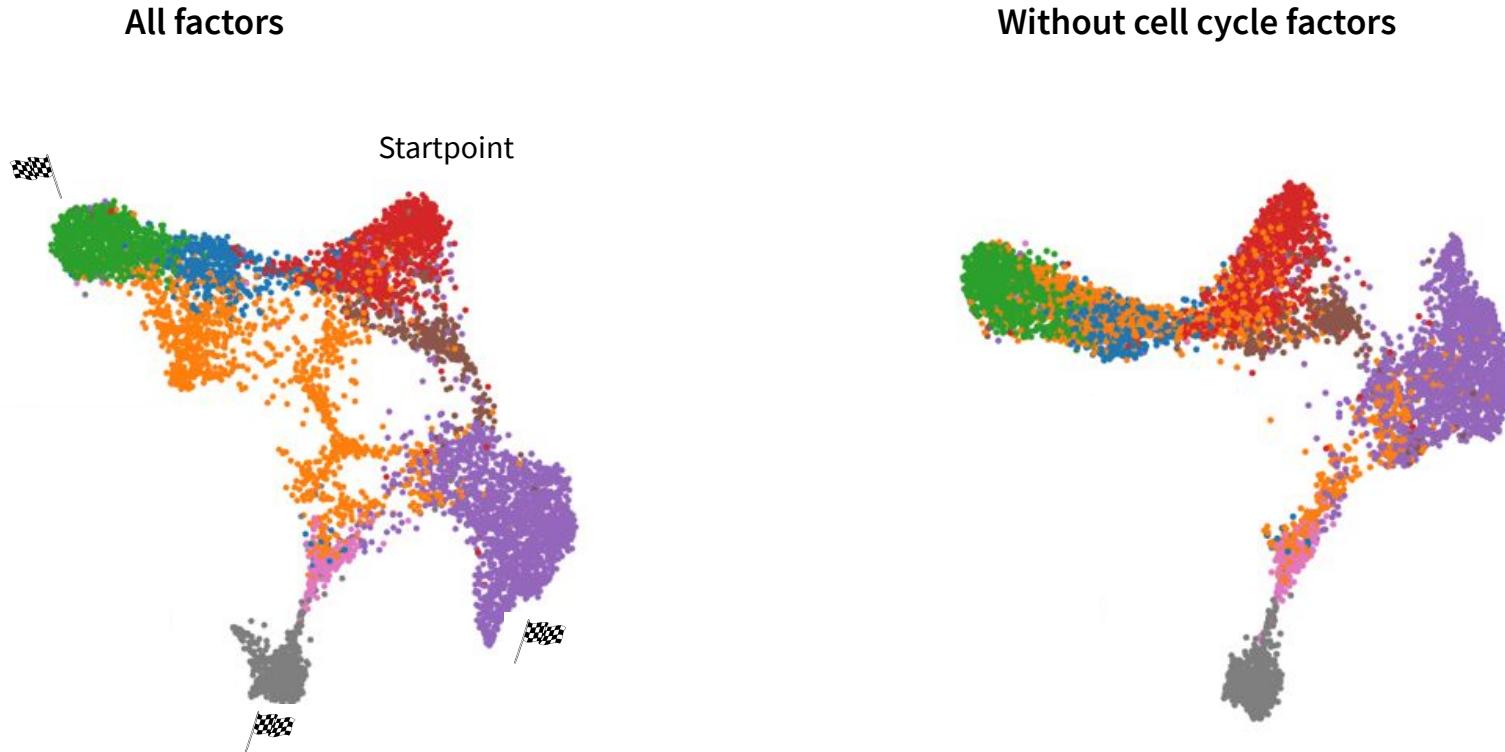
- Branching trajectory
- One common progenitor
  - Radial glia (RG)
- Intermediate cell state
  - Intermediate progenitor cells (nIPC)
- Multiple cell types
  - Neuroblasts (NB)
  - Oligodendrocytes (OLG)
  - Astrocytes (AST)



# Cell cycle as a confounding process



# Cell cycle as a confounding process



# Validation with RNA velocity



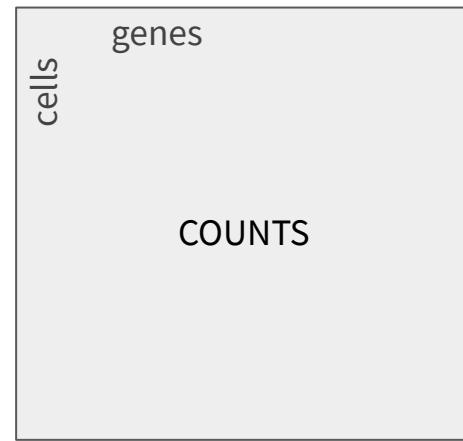
**Timepoint 0**



RNA velocity

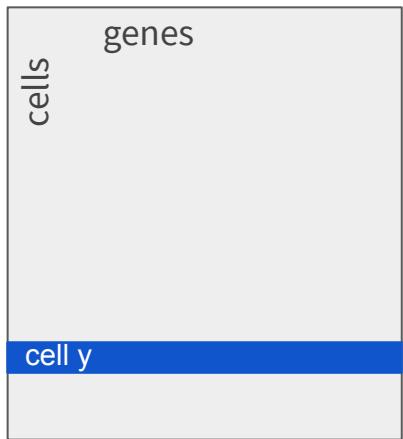


**Timepoint 1**

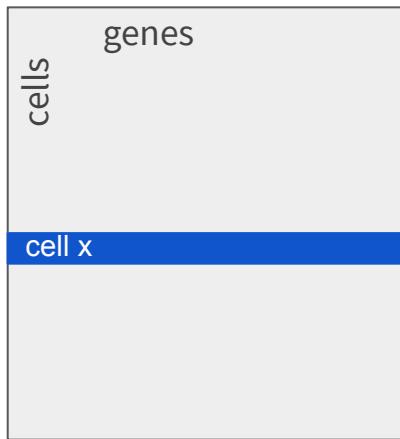


# Transition probabilities

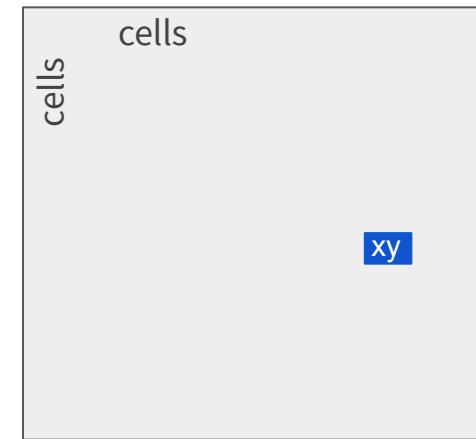
Counts t=0



Counts t=1



Transition probs.

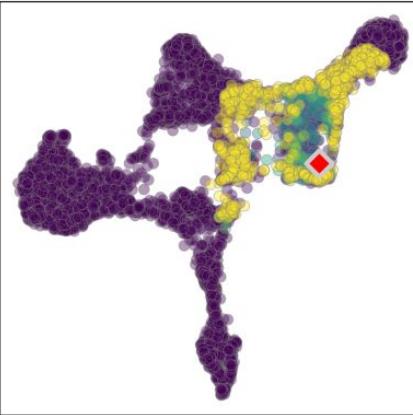


Precursor to

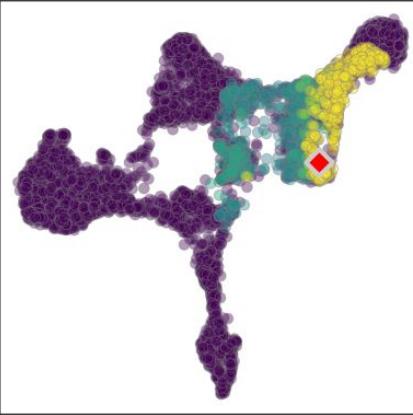
Cell cycle  
factors

Neuroblasts

Included



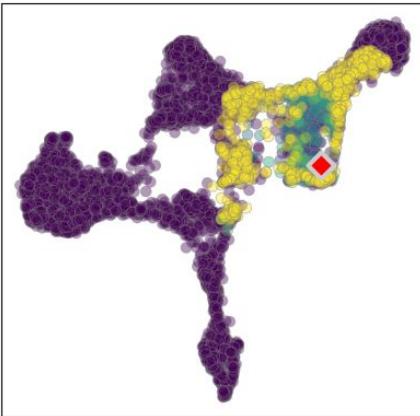
Excluded



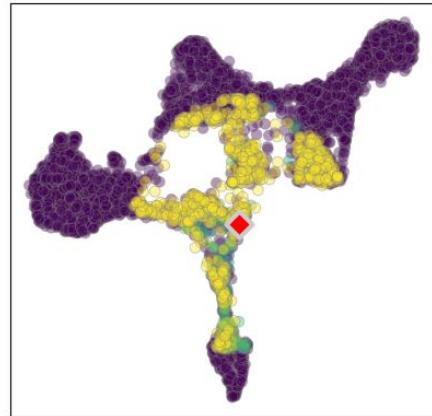
Precursor to

Cell cycle  
factors

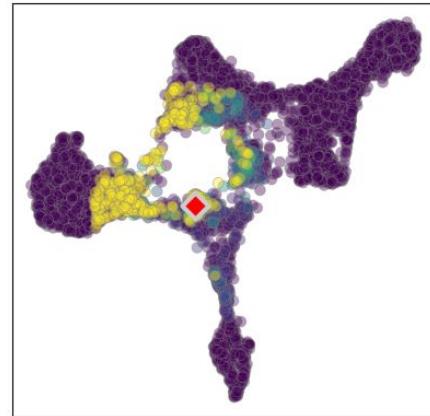
Neuroblasts



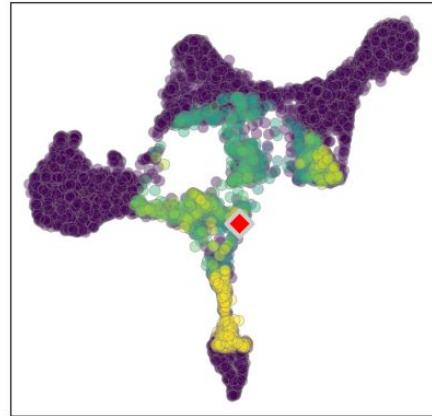
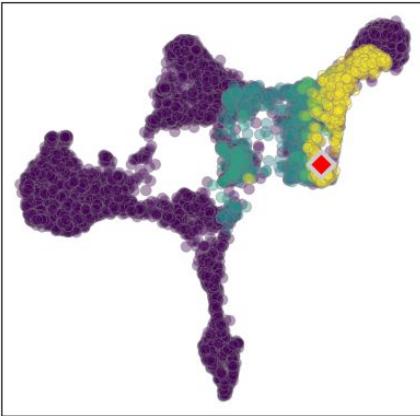
Oligodendrocytes



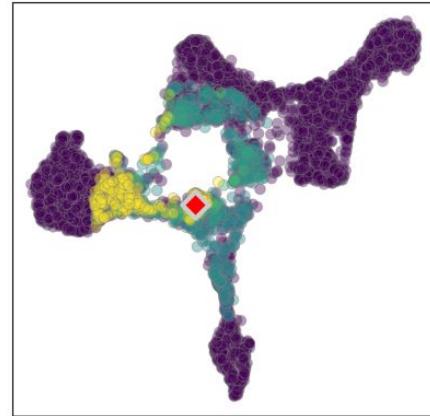
Astrocytes



Included



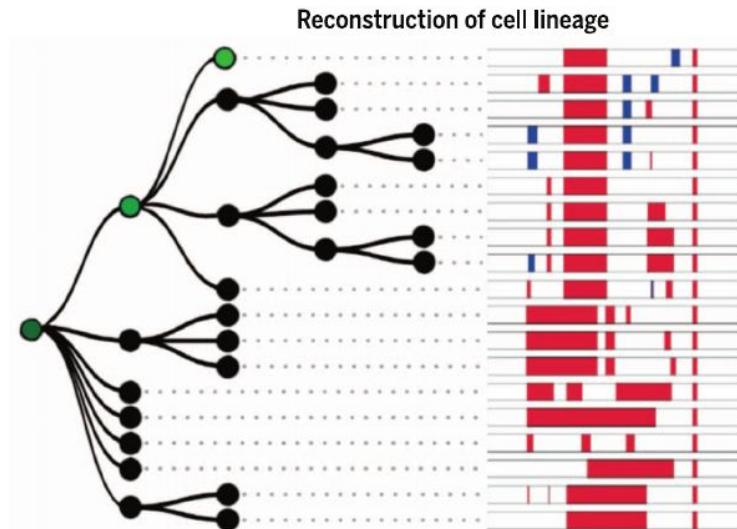
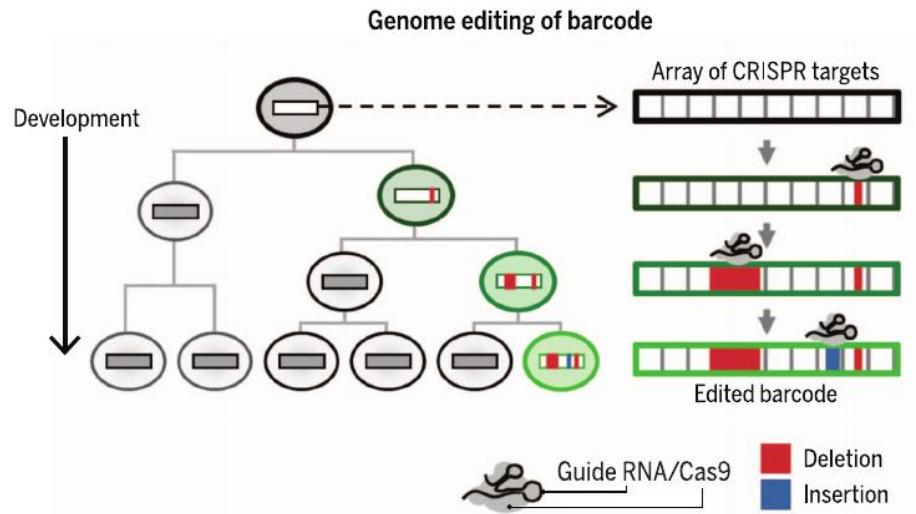
Excluded



# Lineage tracing

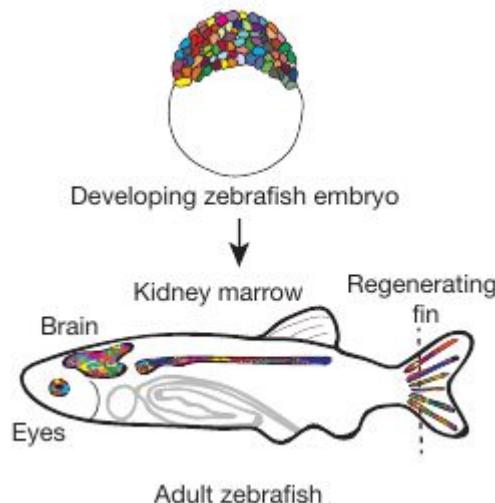
Pseudotime inference tracks cell differentiation “horizontally”

Lineage tracing tracks cell progeny:

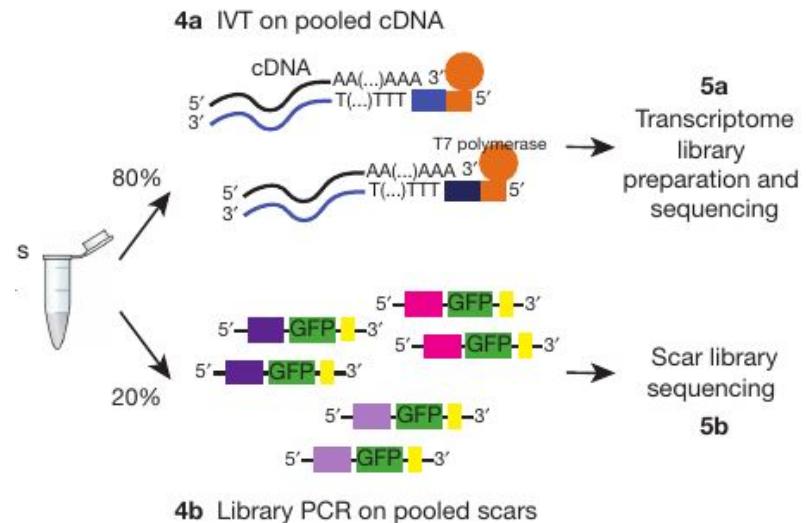


# ScarTrace: Combining lineage tracing and scRNA-seq

Lineage tracing of zebrafish development. Span: 10 hpf

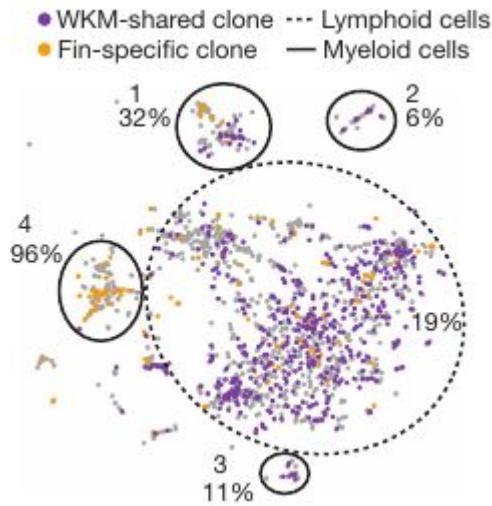


Cell material is split for scar library and transcriptome library preparation

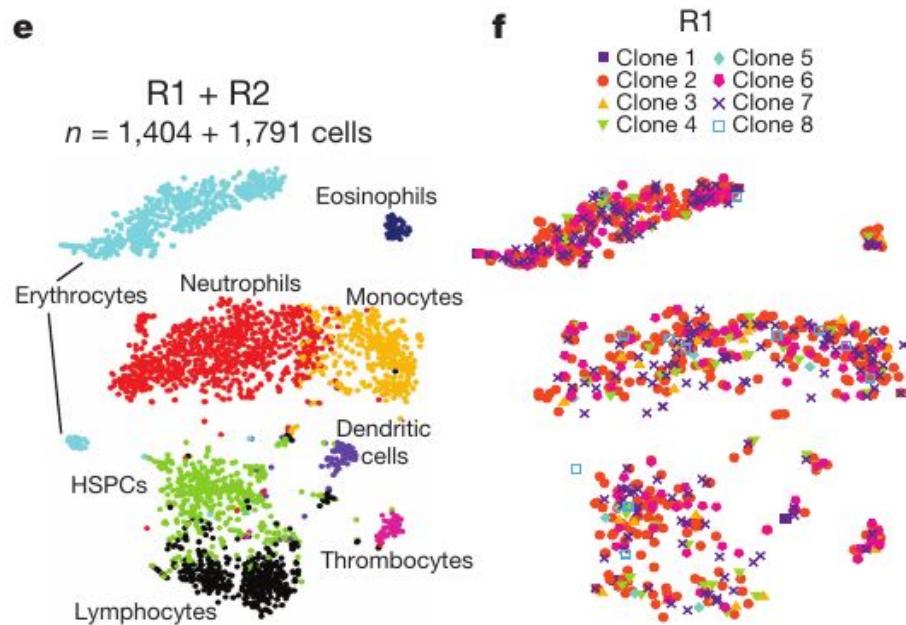


# ScarTrace: Combining lineage tracing and scRNA-seq

## Resident immune cells in fin



## Clonal overlap in kidney marrow



# Summary

- Single cell RNA-seq can yield continuous structures
- Clustering too arbitrary, infer a trajectory instead
- Cell is assigned a pseudotime instead of discrete cluster
- Many methods, but end goal is generally the same

# Trajectory inference practical

- Folder: session-trajectories
- Use Monocle 2 and Destiny for pseudotime inference