

# Trajectory Inference

Ahmed Mahfouz

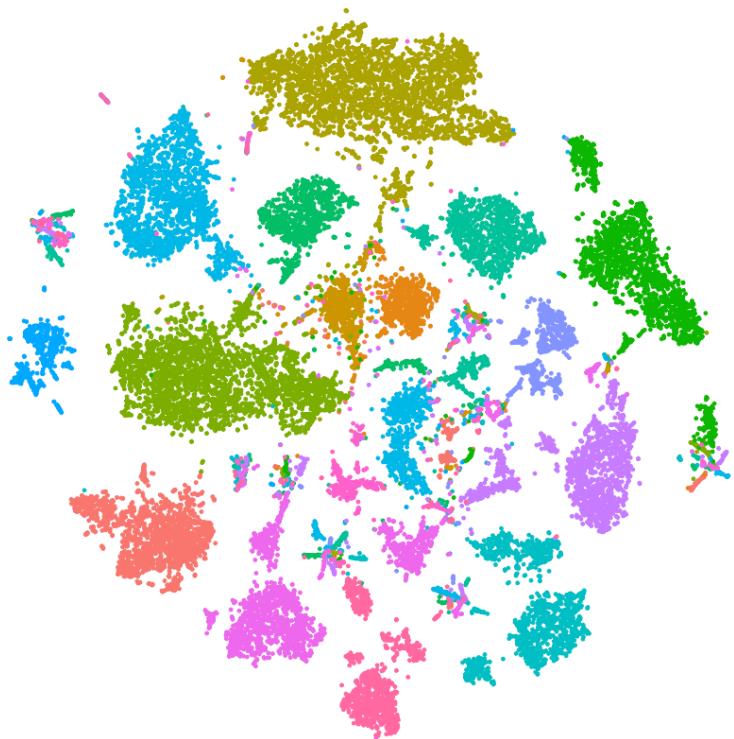
Department of Human Genetics, Leiden University Medical Center  
Pattern Recognition and Bioinformatics, TU Delft

mahfouzlab.org  
@ahmedElkoussy

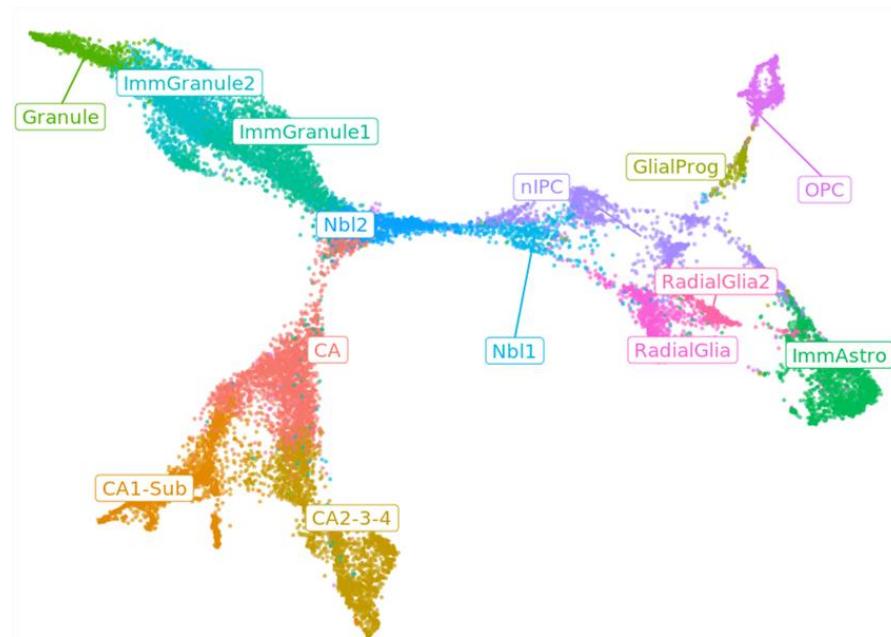
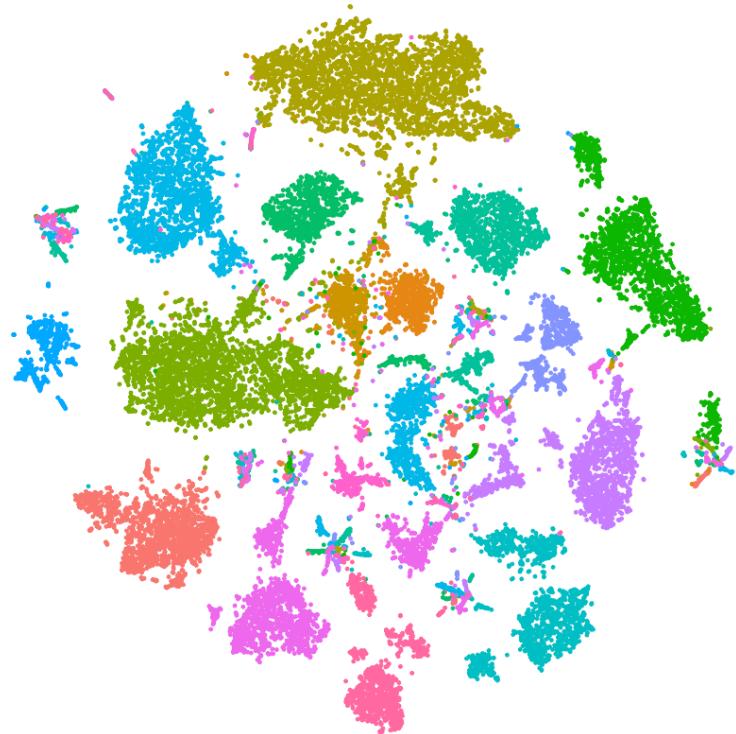
(Slides by Mo Charrouf)



# Clustering of differentiating cells

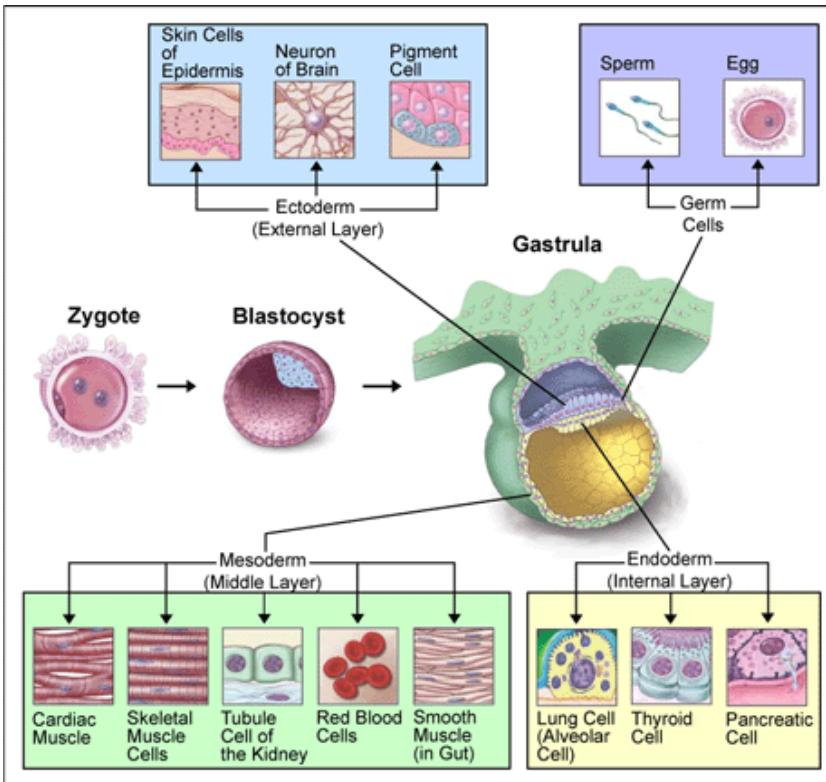


# Clustering of differentiating cells



# When do continuous structures pop up?

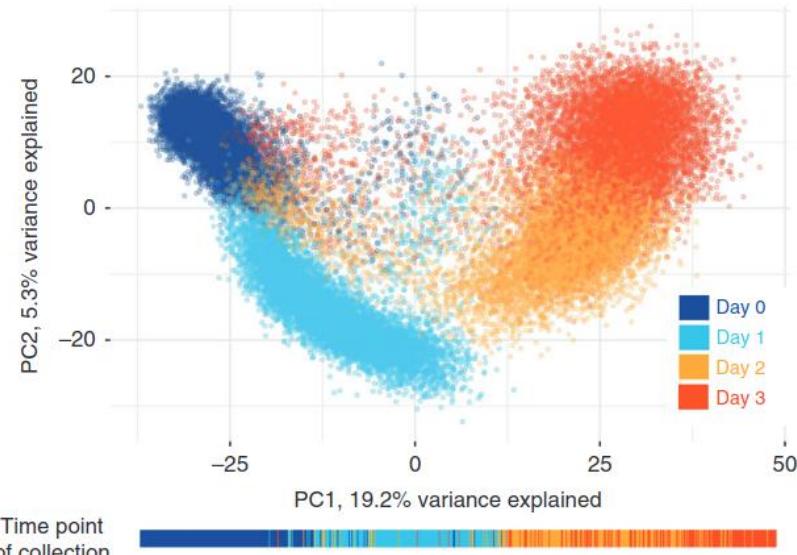
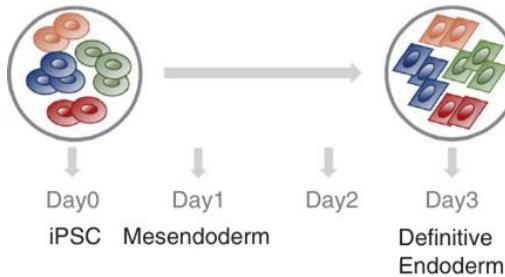
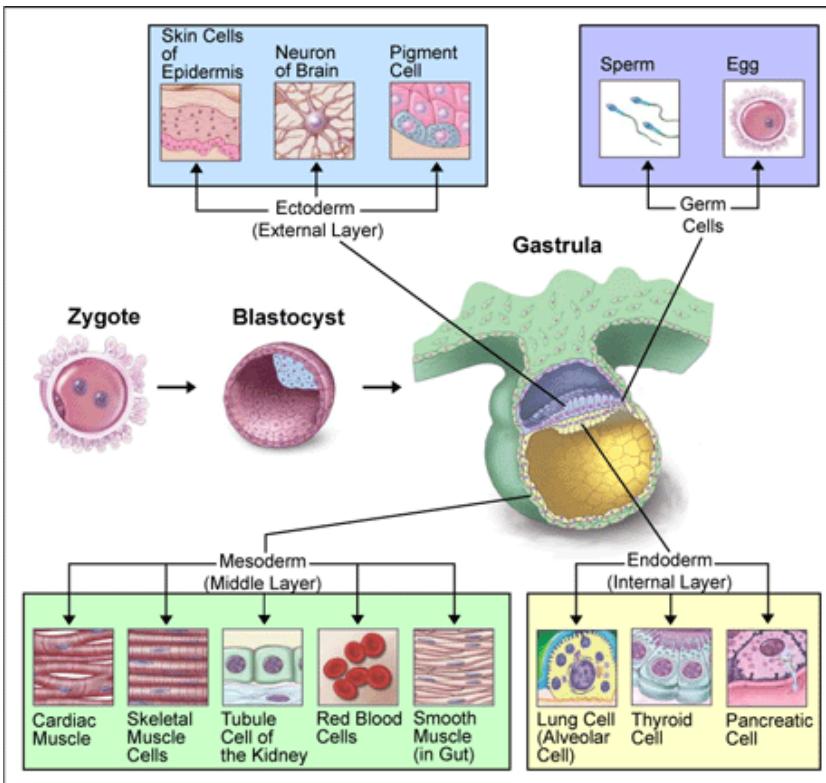
## Embryonic development



# When do continuous structures pop up?

Cuomo et al., 2020

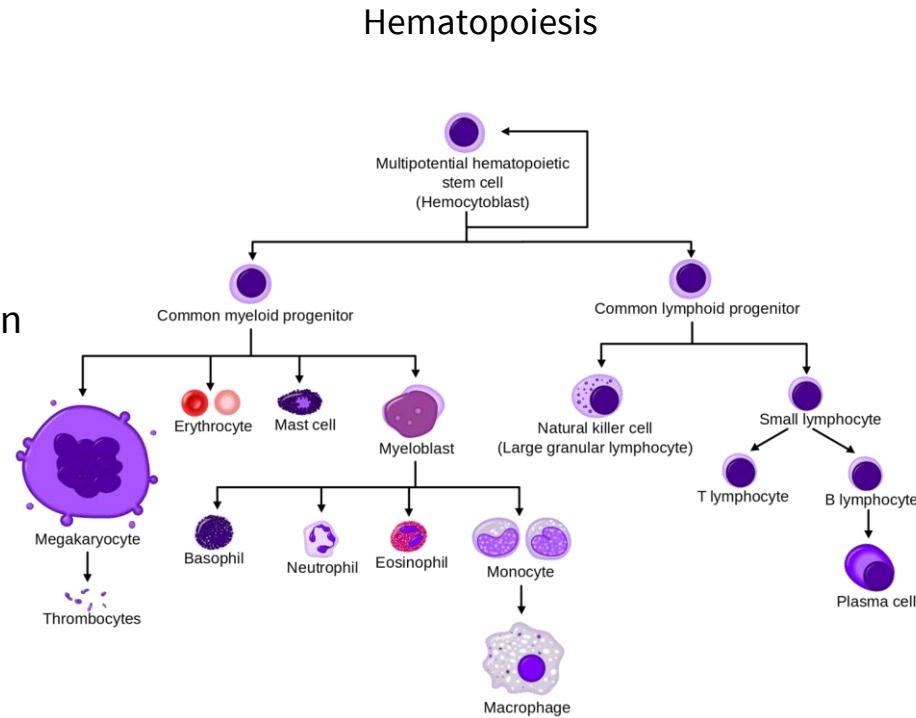
## Embryonic development



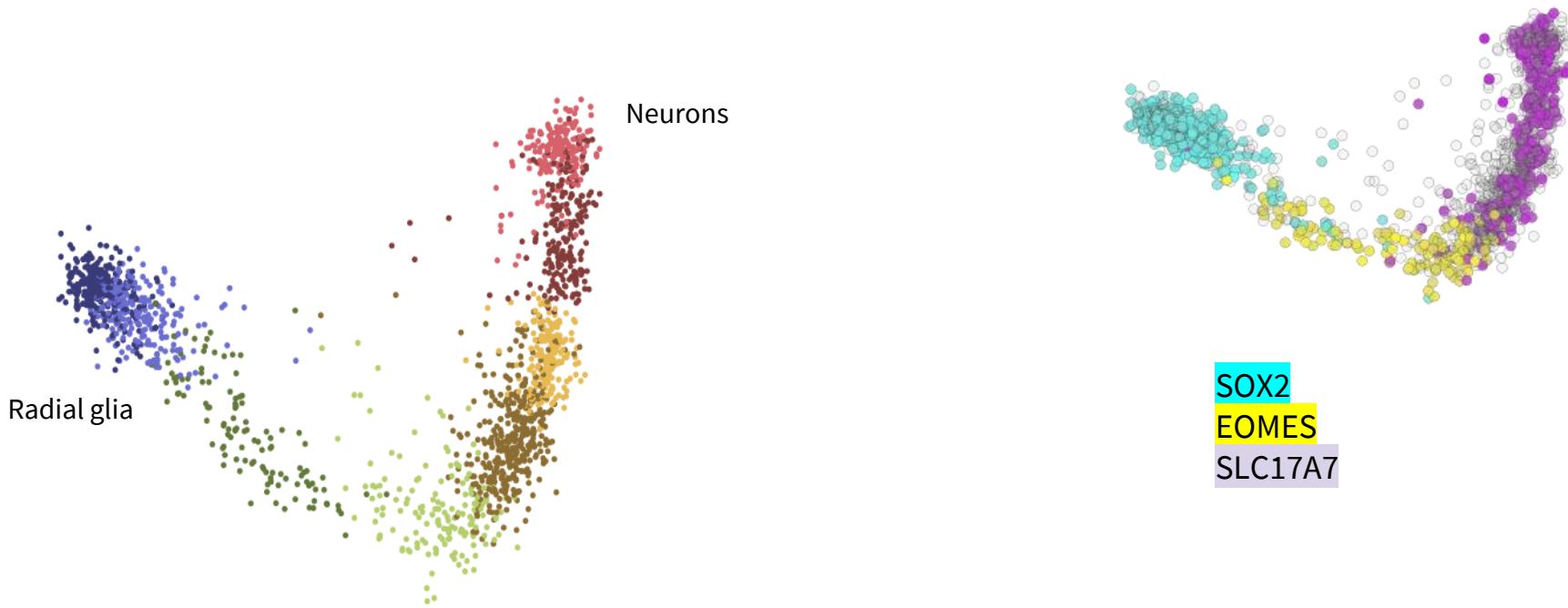
# When do continuous structures pop up?

## Cell differentiation

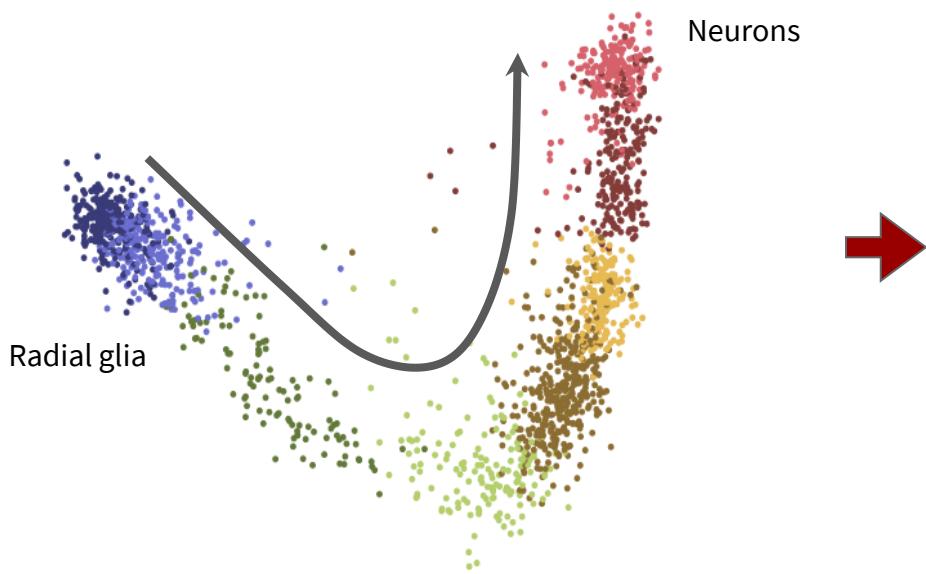
- Individual cells will differentiate in an unsynchronized manner
- Each cell is a snapshot along the differentiation trajectory



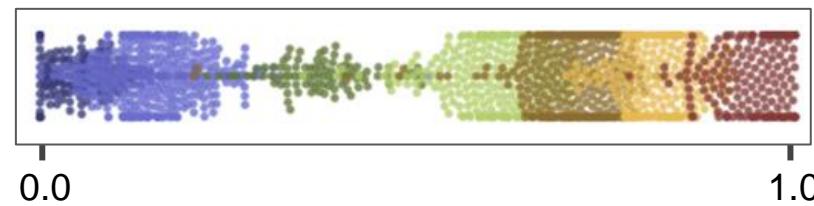
# Trajectory inference / pseudotime inference



# Trajectory inference / pseudotime inference

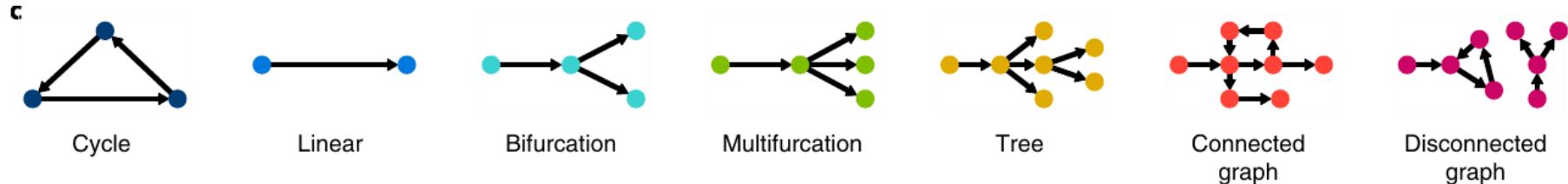


Pseudotime



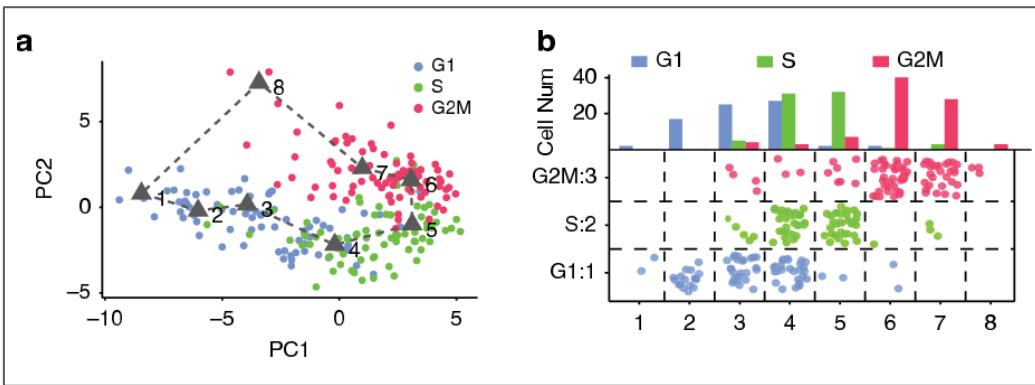
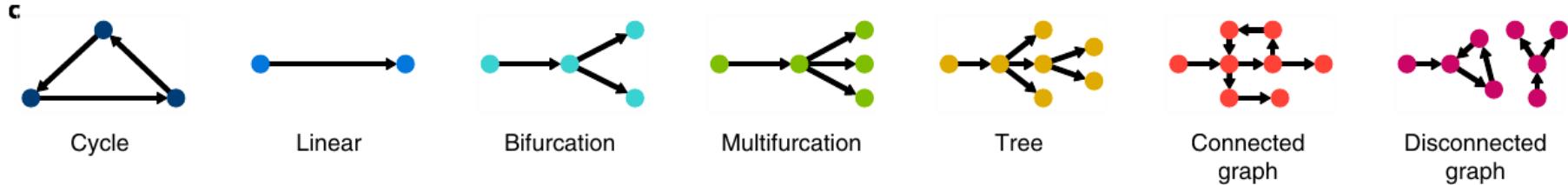
# Trajectory structure

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



# Trajectory structure

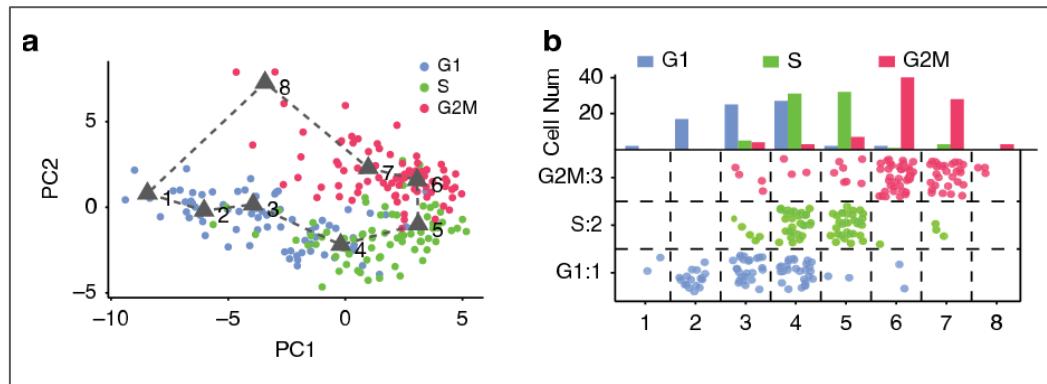
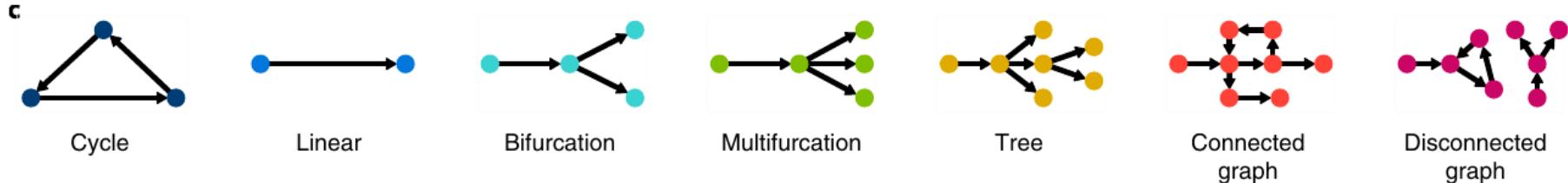
Saelens et al. (<https://doi.org/10.1038/s41587-019-0071-9>)



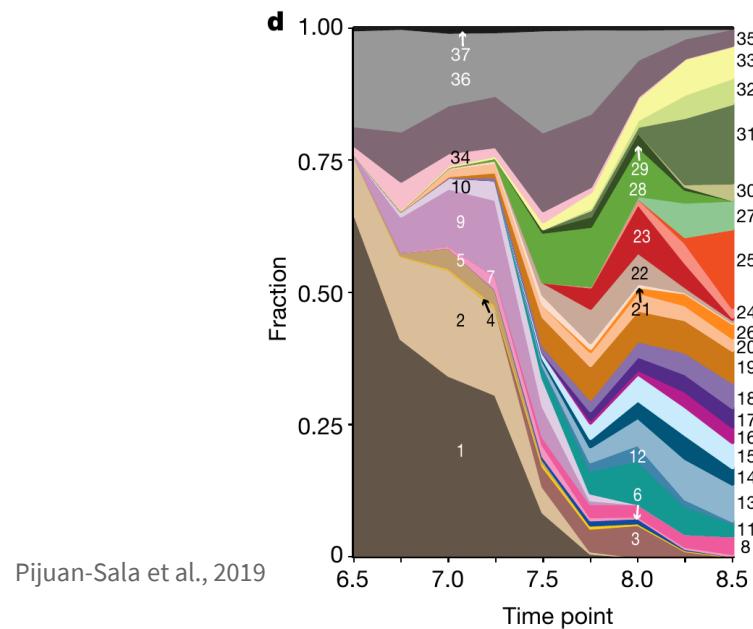
Liu et al. 2017

# Trajectory structure

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



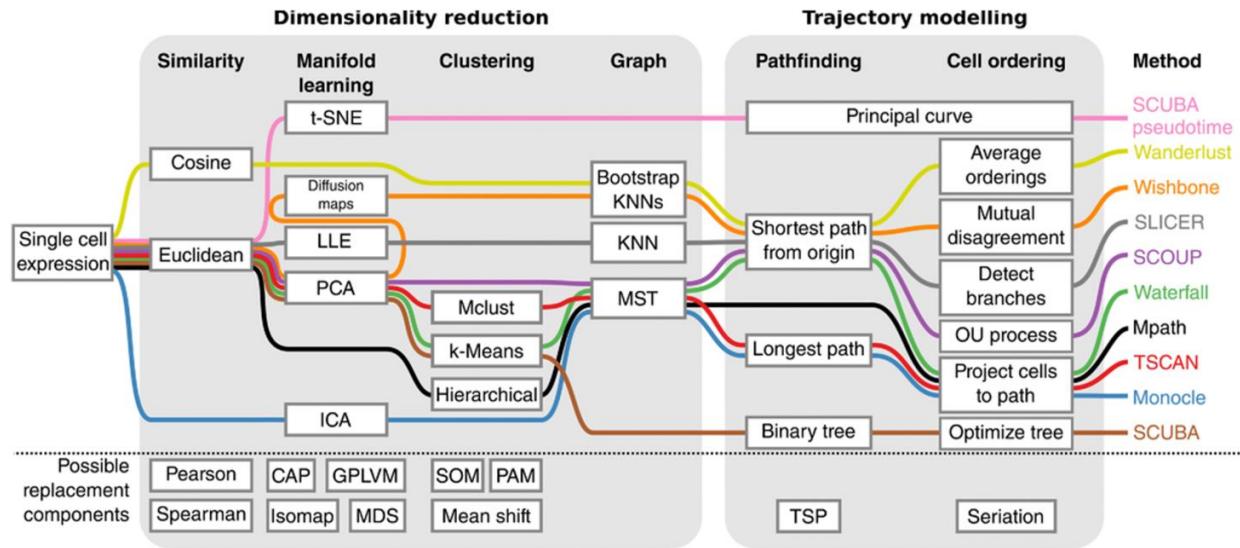
Liu et al. (<https://doi.org/10.1038/s41467-017-00039-z>)



# Methodology

# General trajectory inference pipeline

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



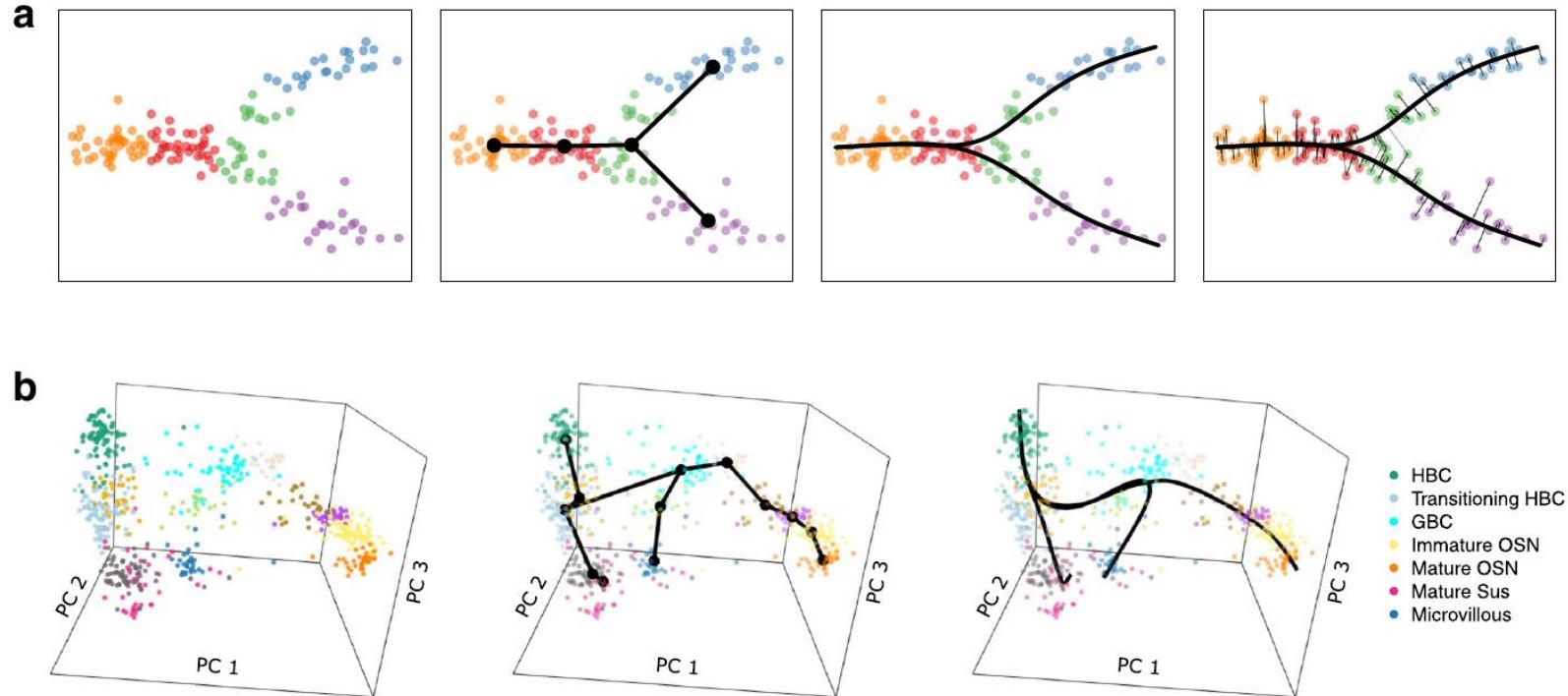
# Trajectory inference methods

Four methods will be discussed

- **Monocle 1, Slingshot** Trapnell et al., 2014, Street et al., 2018
  - Apply the general pipeline of dimensionality reduction, curve fitting, and pseudotime assignment
- **Monocle 2** Cole et al., 2017
  - A popular tool in trajectory inference
- **Ouija** Campbell et al., 2018
  - Gene-based fitting
- **RNA velocity** La Manno et al., 2018
  - Biologically-driven identification of trajectories

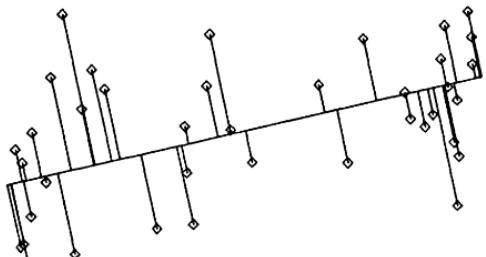
# Slingshot

Fit multiple principle curves simultaneously, ensuring a shared trunk



# Principle curves

b Principal component



d Principal curve

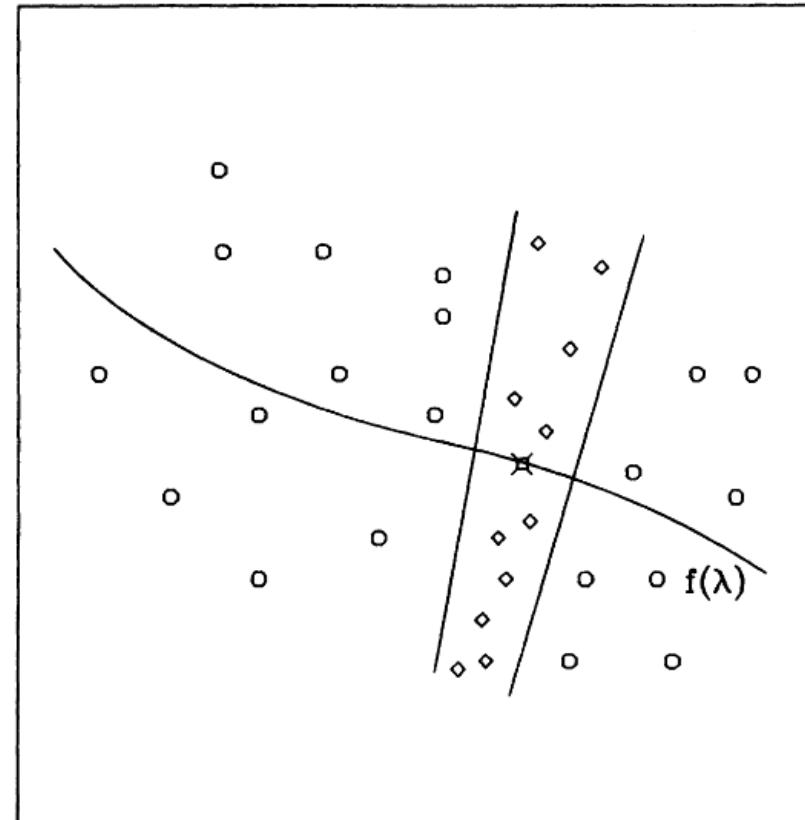
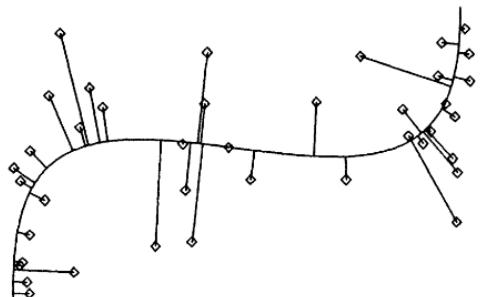
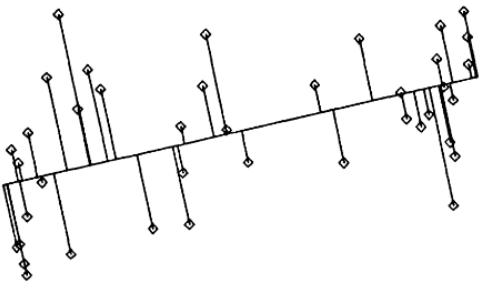


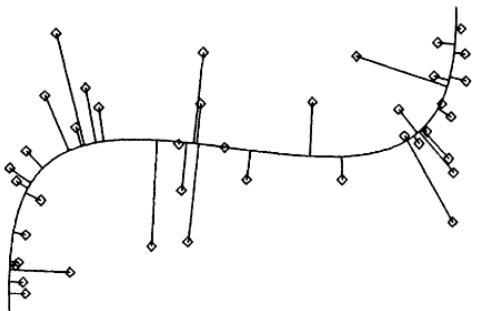
Figure 3. Each point on a principal curve is the average of the points that project there.

# Principle curves

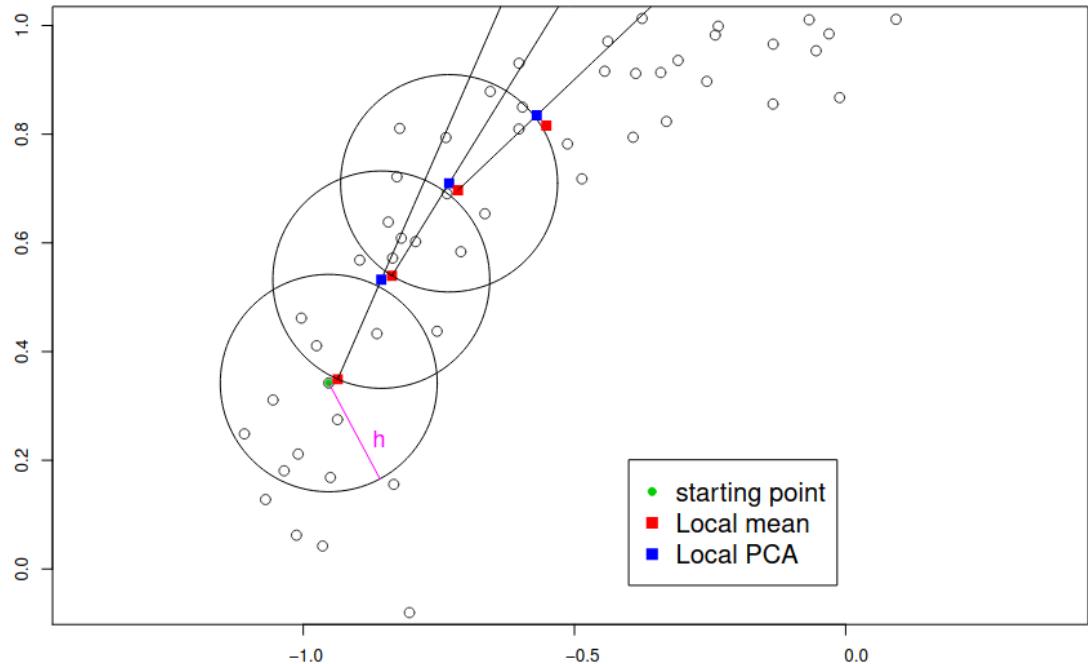
b Principal component



d Principal curve

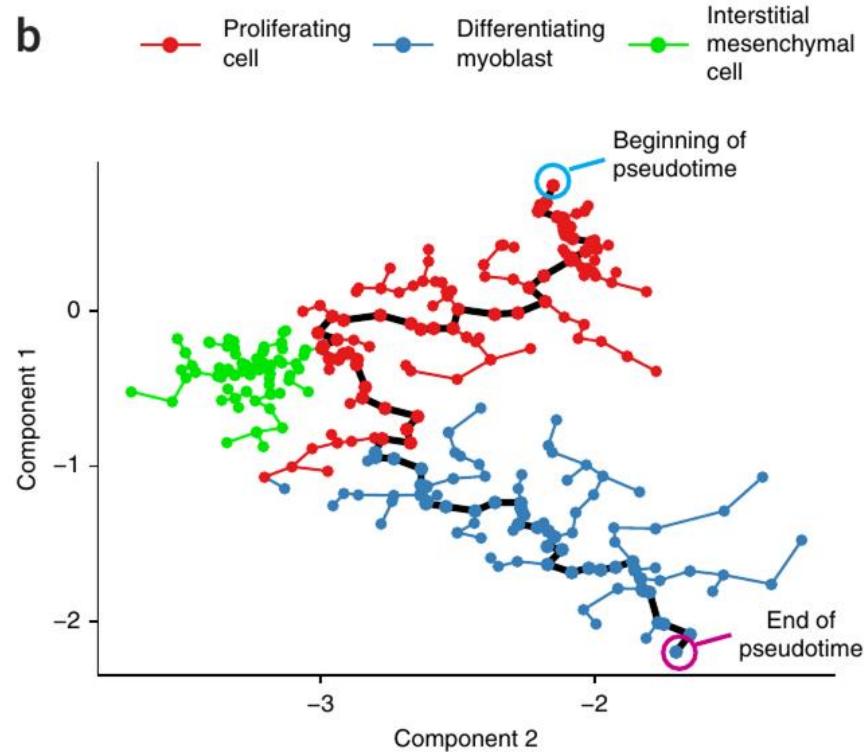
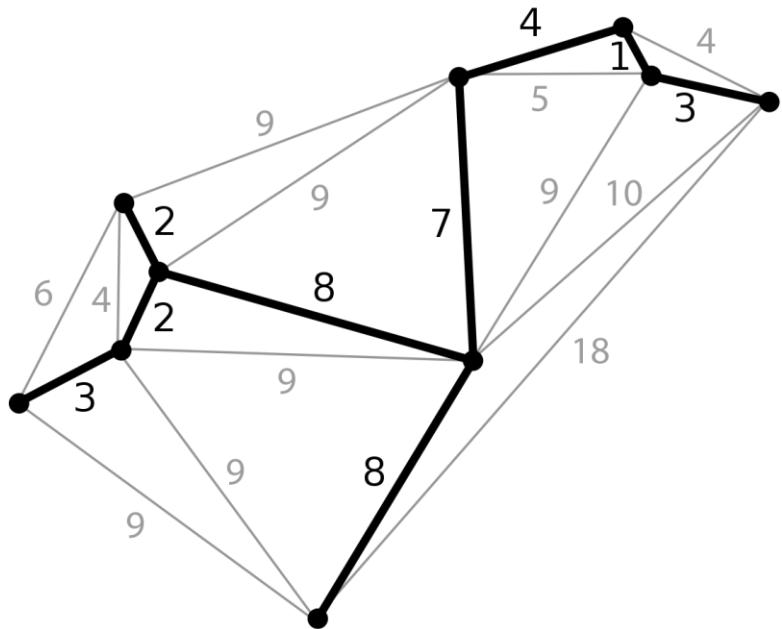


Local principal curve (LPC) algorithm



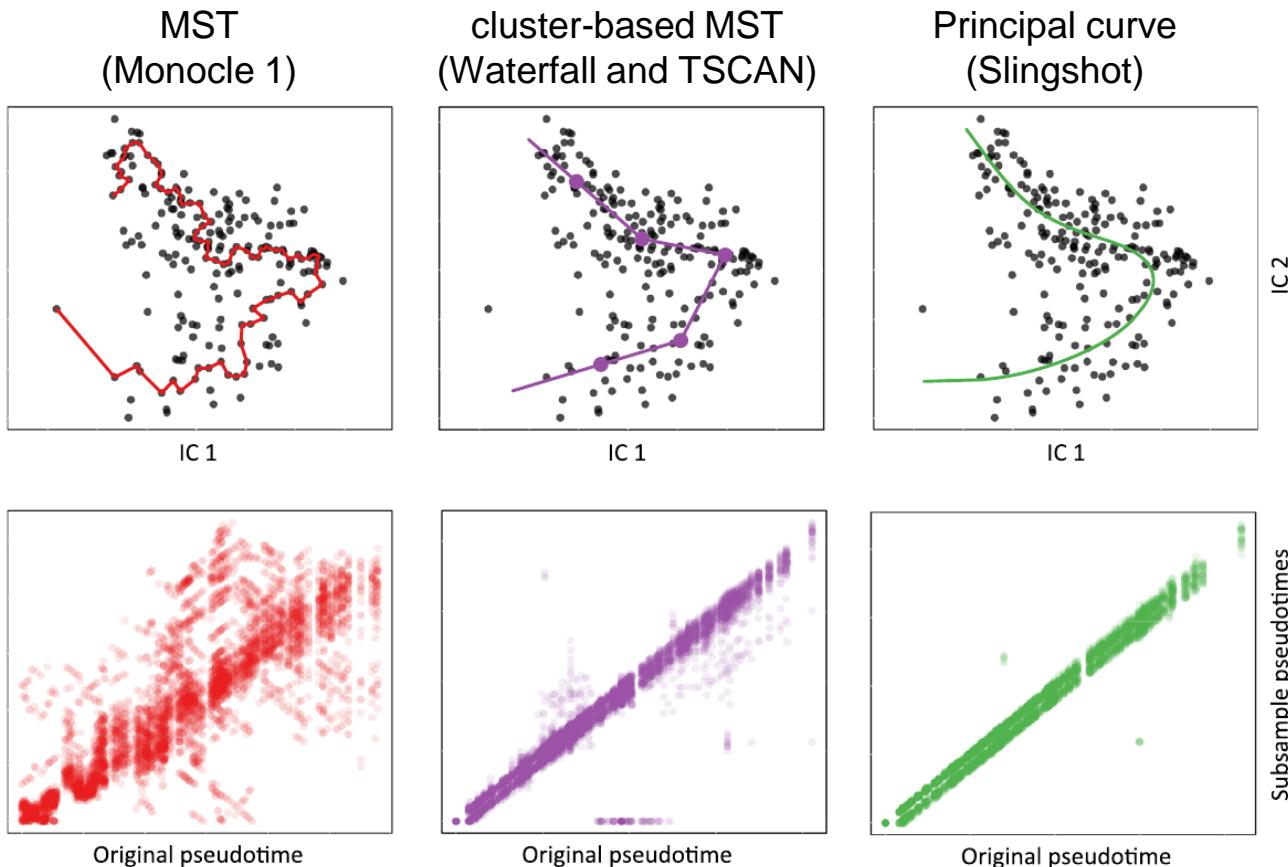
# Monocle 1

Minimum spanning tree (MST)



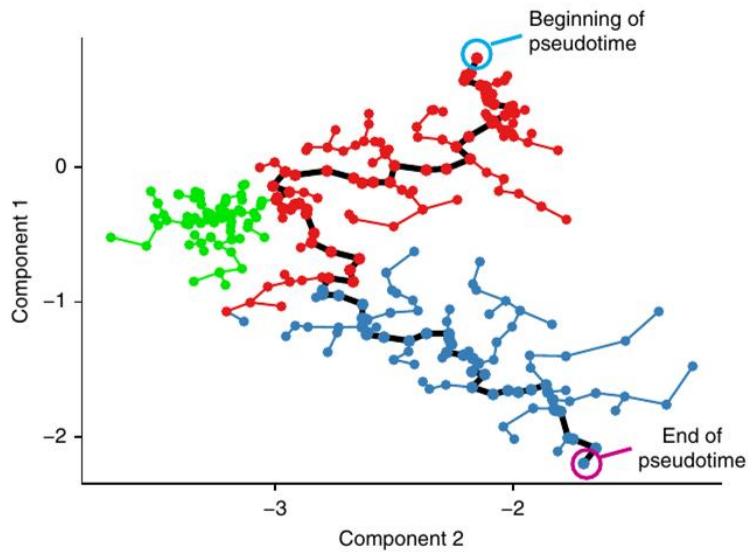
Trapnell et al., 2014

# Robustness: principal curves vs MST



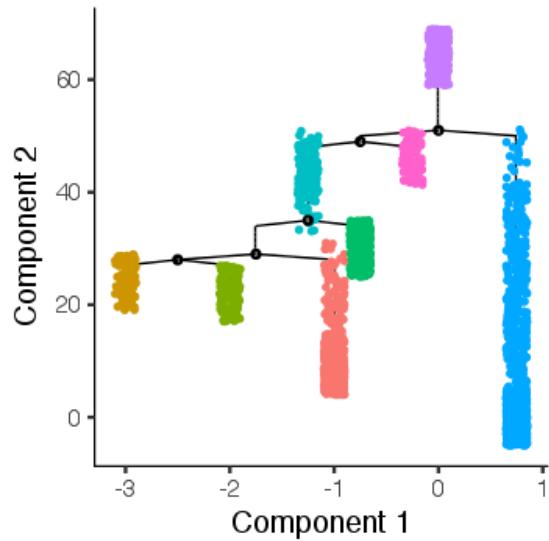
# Monocle 2

- Successor to Monocle 1
- End goal: Fit any arbitrary graph on the data
  - Curves, bifurcations, loops

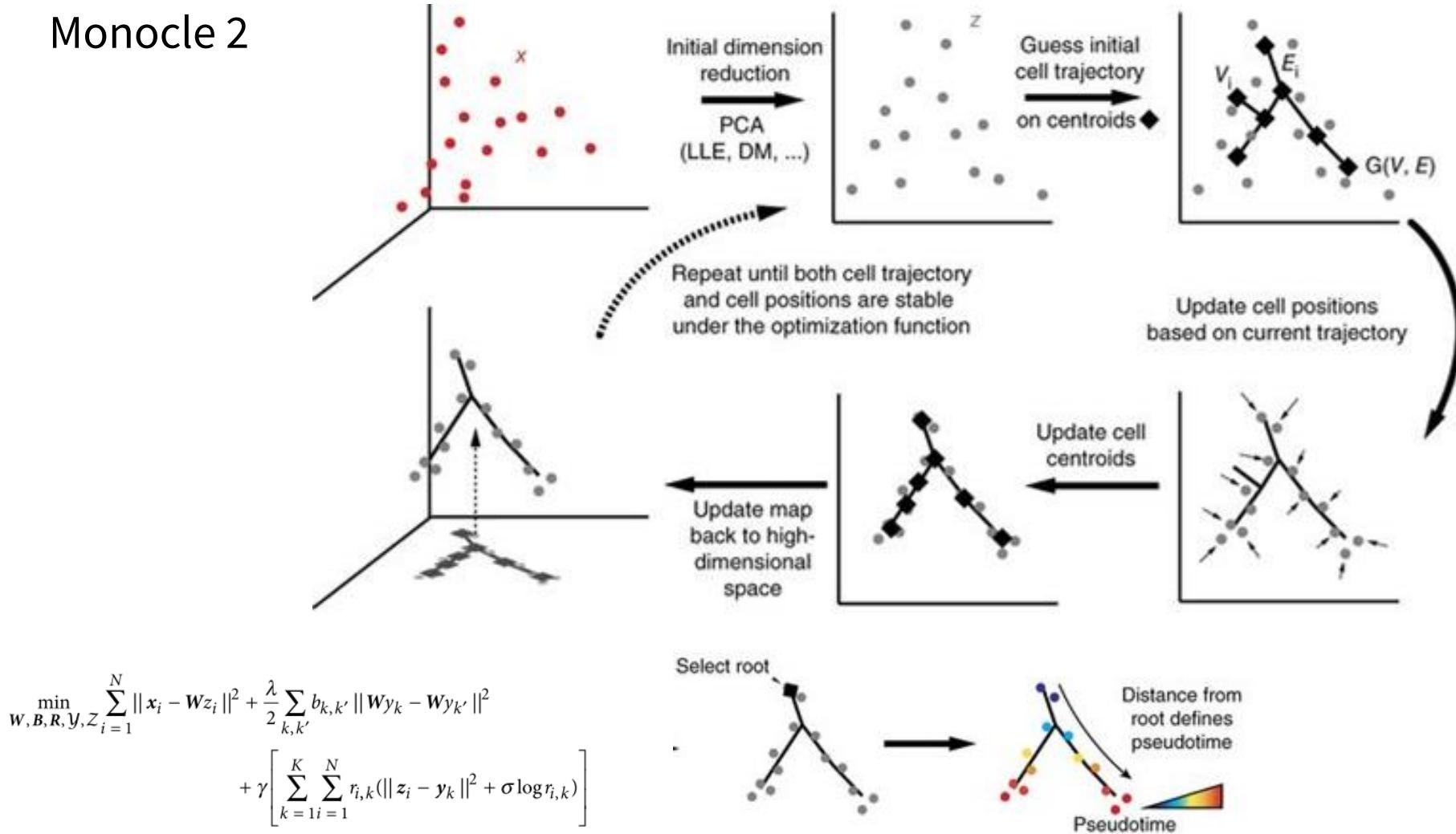


State

State	Color
1	Red
4	Green
7	Cyan
10	Purple
3	Yellow
6	Light Green
9	Blue
11	Pink

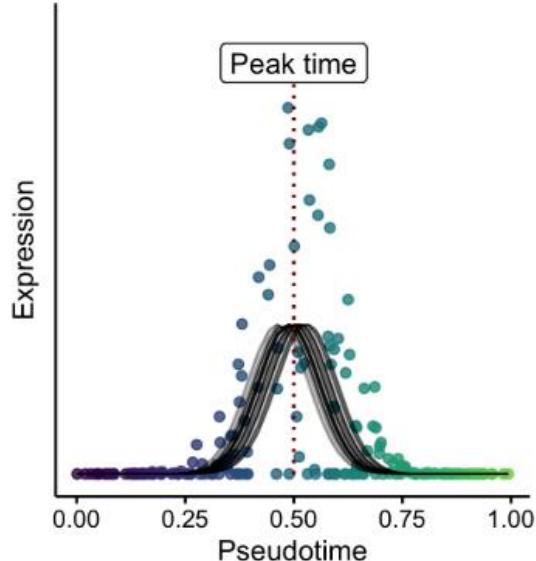
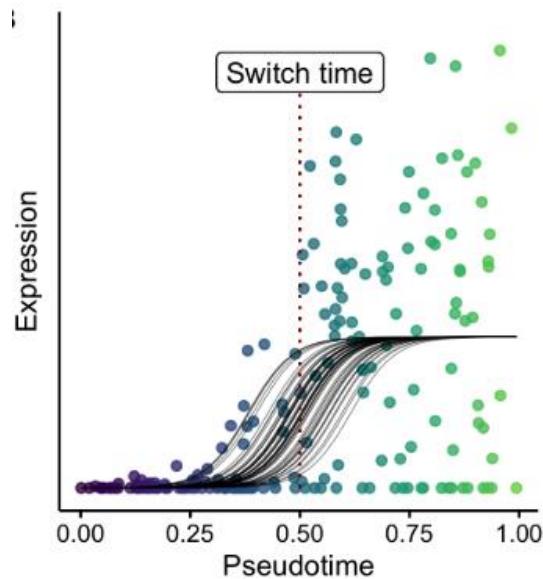


# Monocle 2

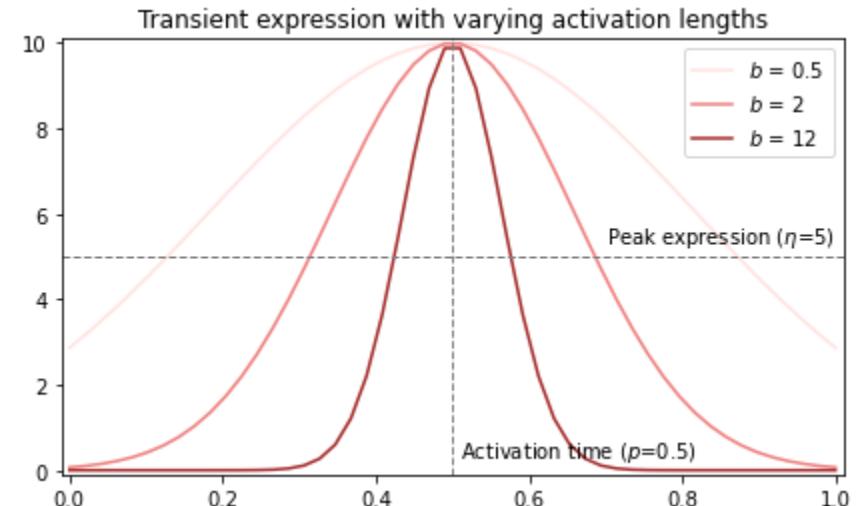
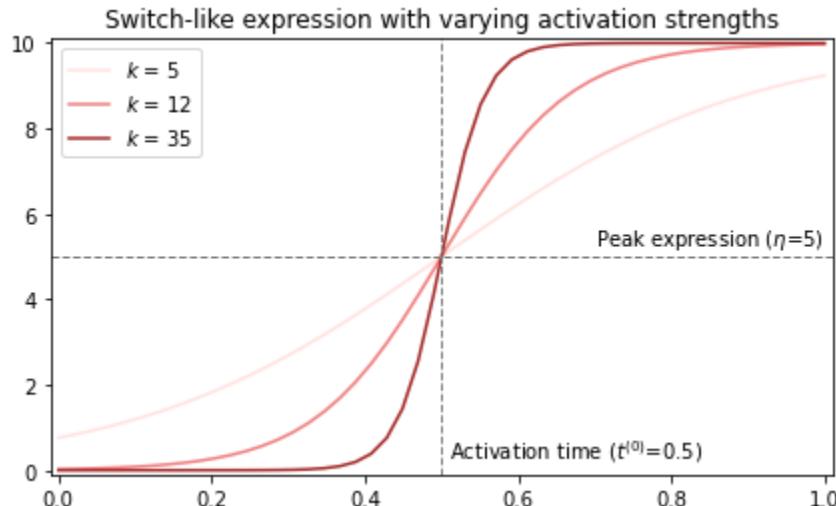


# Ouija

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



# Interpretable function parameters



$$f(t) = \frac{2\eta}{1+\exp(-k(t-t^0))}$$

$$f(t) = \eta \cdot \exp(b \cdot (t - p)^2)$$

# Ouija intuition

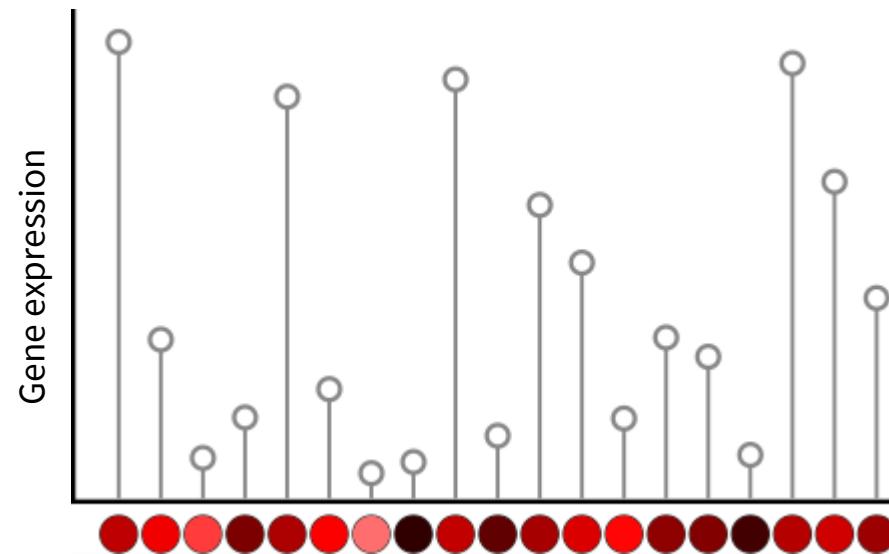
True ordering:



Marker gene x (Transient activation)

Random cell ordering:

Goodness-of-fit: **low**

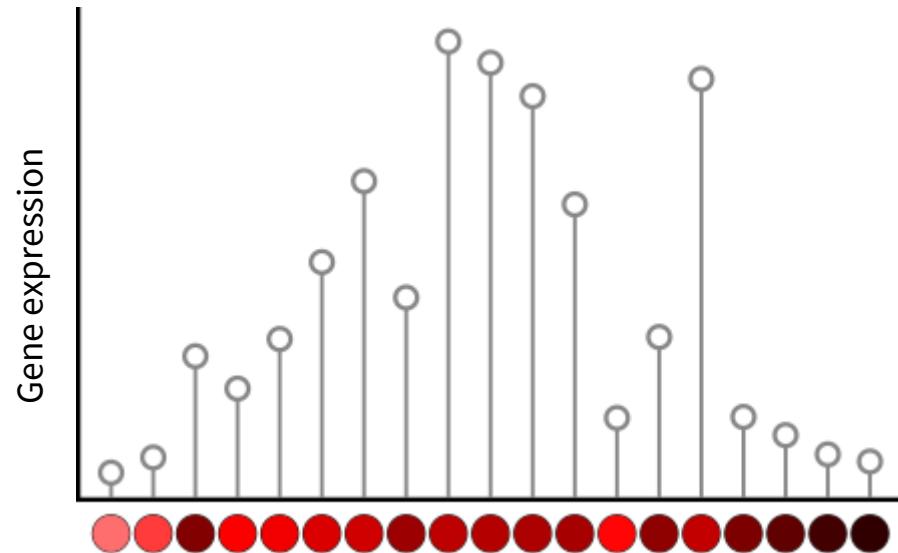


# Ouija intuition

True ordering:



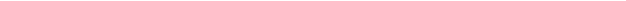
Marker gene x (Transient activation)



Optimize iteration: 100

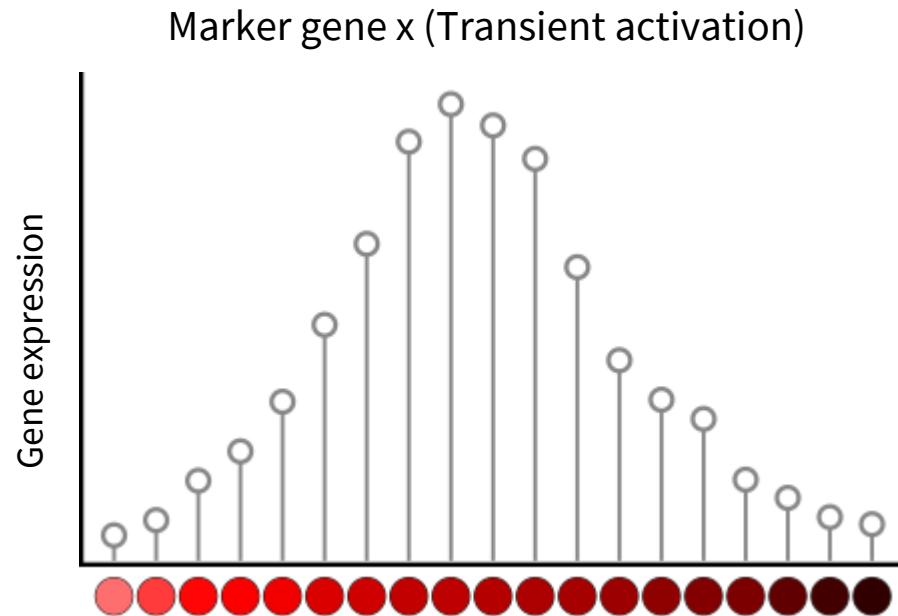
Goodness-of-fit: mid

# Ouija intuition

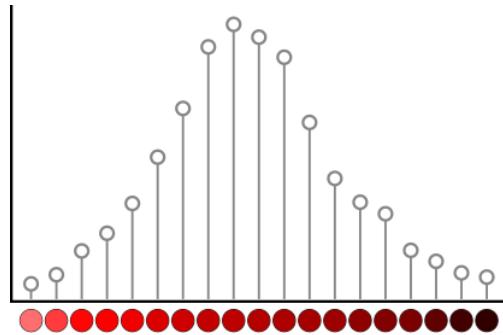
True ordering: 

# Optimize iteration: 500

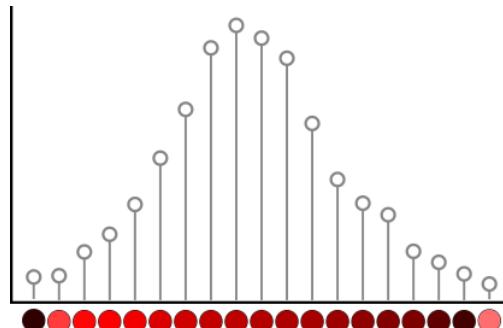
## Goodness-of-fit: **high**



## Marker gene x

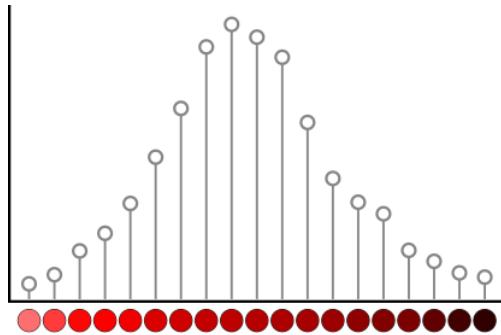


Goodness-of-fit: **high**

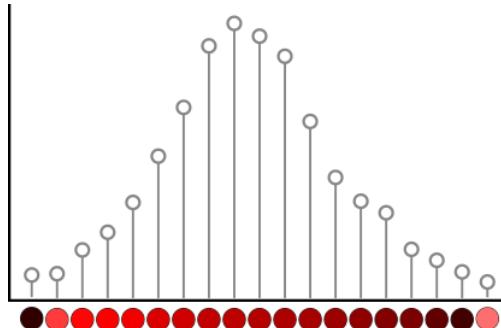


Goodness-of-fit: **high**

Marker gene x

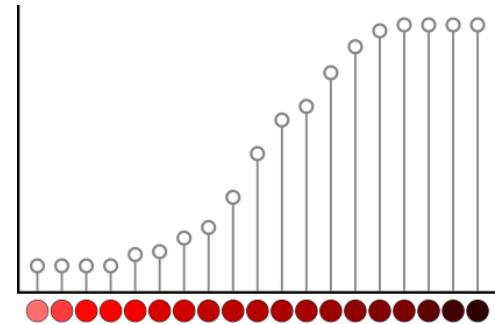


Goodness-of-fit: high

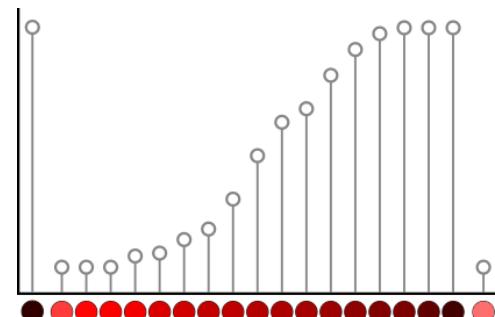


Goodness-of-fit: high

Marker gene y



Goodness-of-fit: high

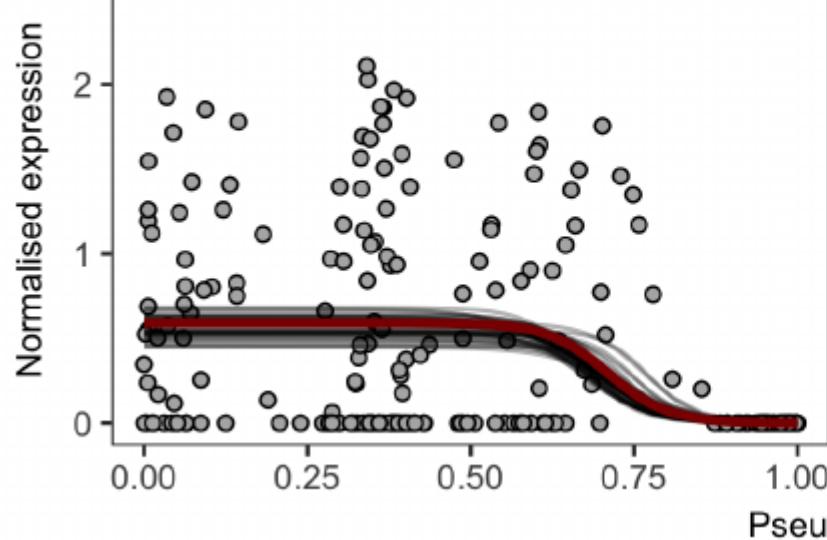


Goodness-of-fit: mid

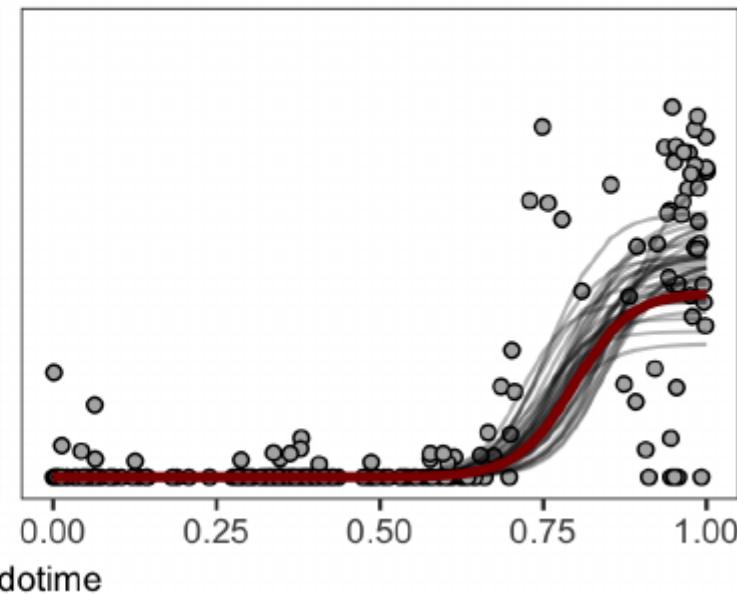
# Ouija probabilistic modelling

C

ID1



MYOG



# Fundamental limits on dynamic inference from single-cell snapshots

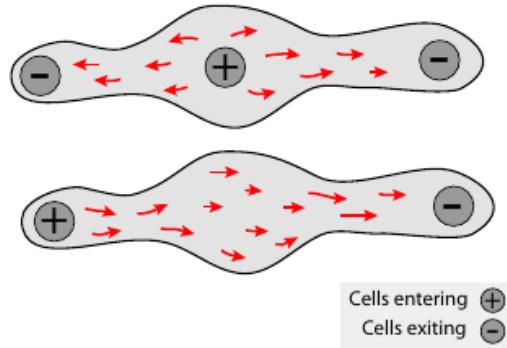
Caleb Weinreb<sup>a</sup>, Samuel Wolock<sup>a</sup>, Betsabeh K. Tusi<sup>b</sup>, Merav Socolovsky<sup>b</sup>, and Alon M. Klein<sup>a,1</sup>

“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



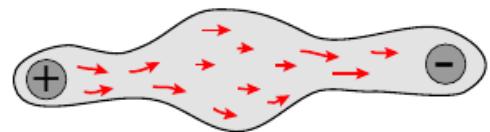
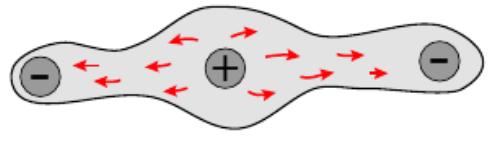
Entry and exit points direct the flow of cells



# No unique solution



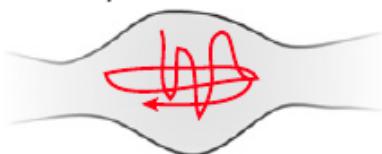
Entry and exit points direct the flow of cells



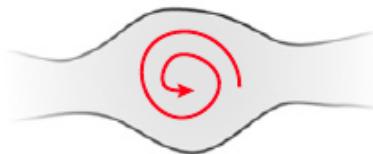
Cells entering  $\oplus$   
Cells exiting  $\ominus$

Rotations in state space do not alter cell density

*Simple fluctuations*



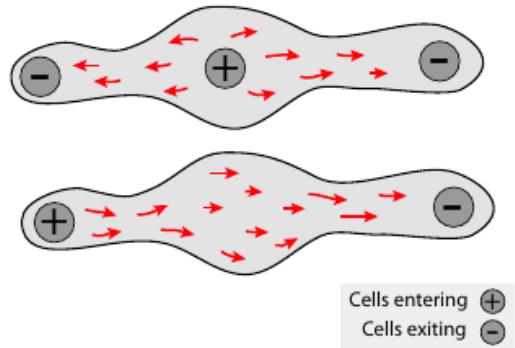
*Periodic oscillations*



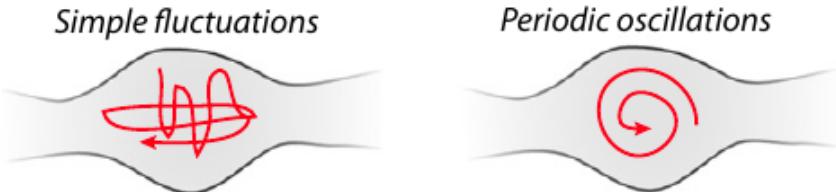
# No unique solution



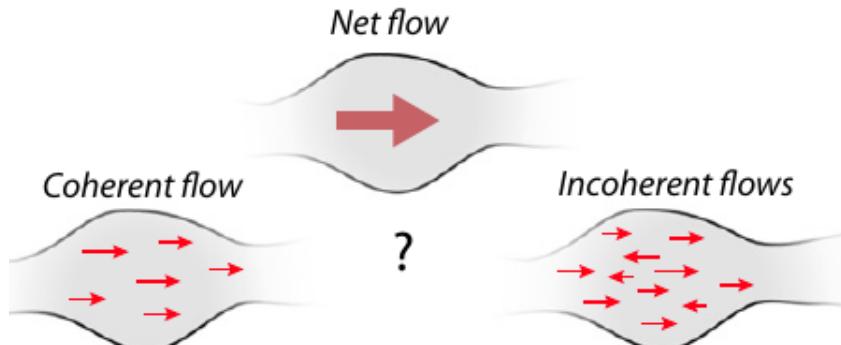
Entry and exit points direct the flow of cells



Rotations in state space do not alter cell density

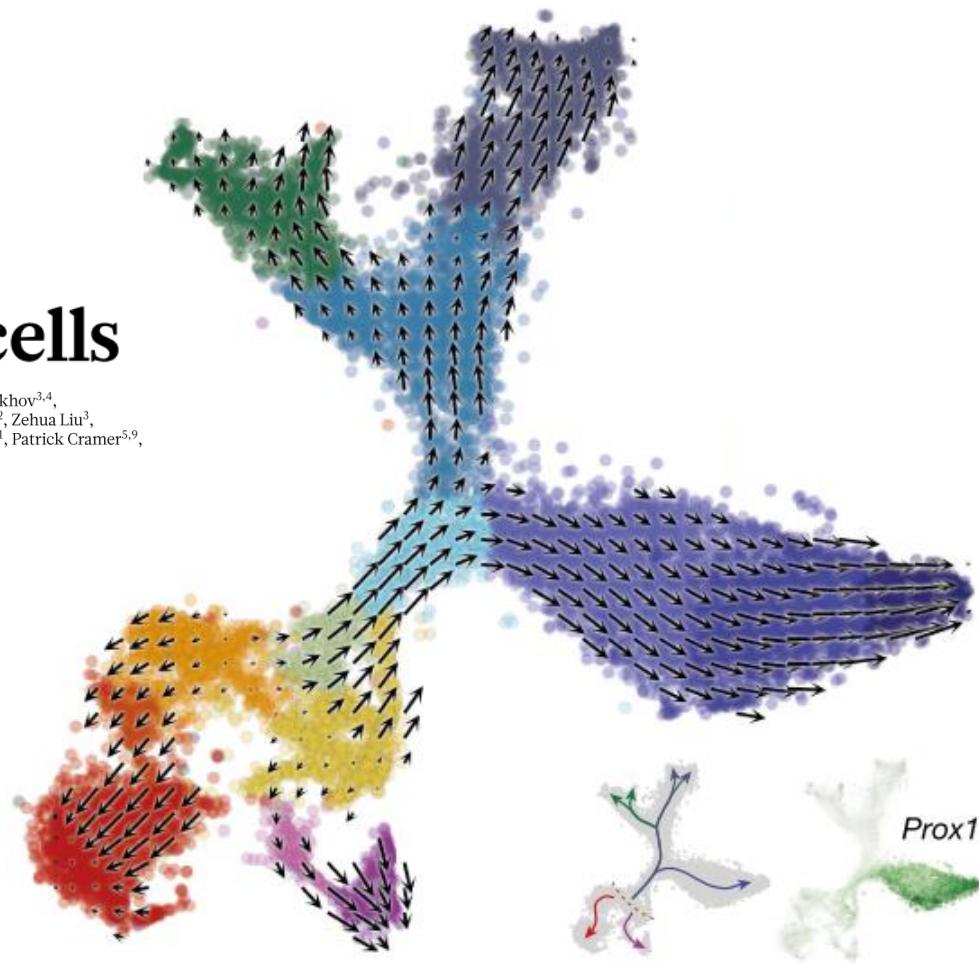


Net velocity may not equal actual velocity



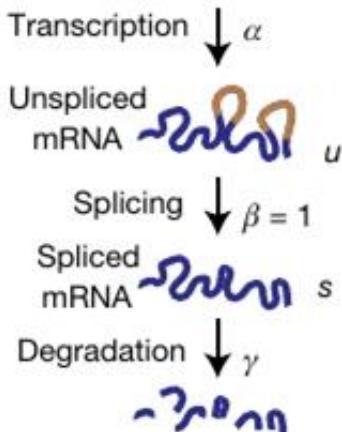
# RNA velocity of single cells

Gioele La Manno<sup>1,2</sup>, Ruslan Soldatov<sup>3</sup>, Amit Zeisel<sup>1,2</sup>, Emelie Braun<sup>1,2</sup>, Hannah Hochgerner<sup>1,2</sup>, Viktor Petukhov<sup>3,4</sup>, Katja Lidschreiber<sup>5</sup>, Maria E. Kastriti<sup>6</sup>, Peter Lönnberg<sup>1,2</sup>, Alessandro Furlan<sup>1</sup>, Jean Fan<sup>3</sup>, Lars E. Borm<sup>1,2</sup>, Zehua Liu<sup>3</sup>, David van Bruggen<sup>1</sup>, Jimin Guo<sup>3</sup>, Xiaoling He<sup>7</sup>, Roger Barker<sup>7</sup>, Erik Sundström<sup>8</sup>, Gonçalo Castelo-Branco<sup>1</sup>, Patrick Cramer<sup>5,9</sup>, Igor Adameyko<sup>6</sup>, Sten Linnarsson<sup>1,2\*</sup> & Peter V. Kharchenko<sup>3,10#</sup>



# RNA velocity of single cells

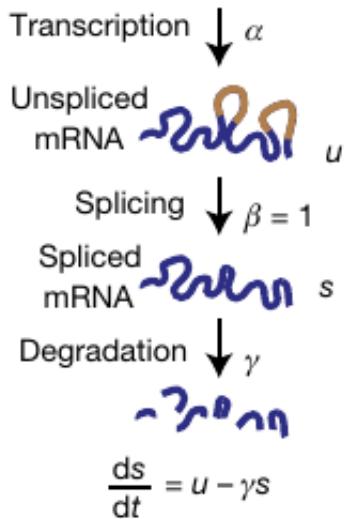
**b**



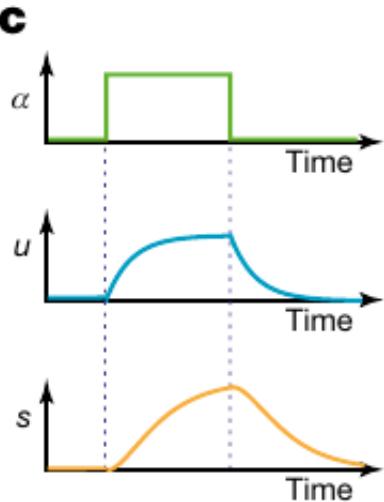
$$\frac{ds}{dt} = u - \gamma s$$

# RNA velocity of single cells

**b**

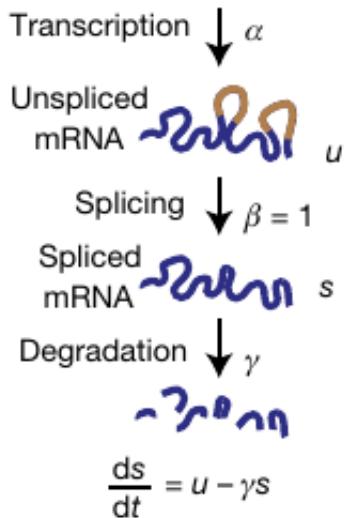


**c**

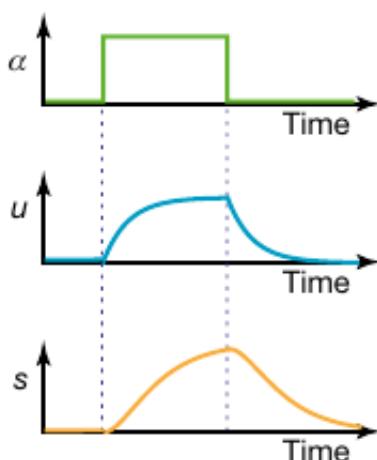


# RNA velocity of single cells

b



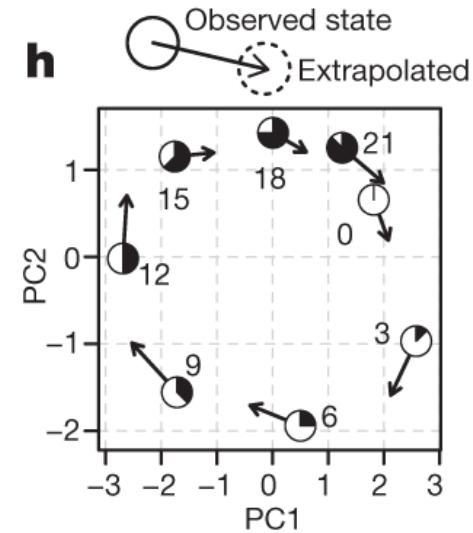
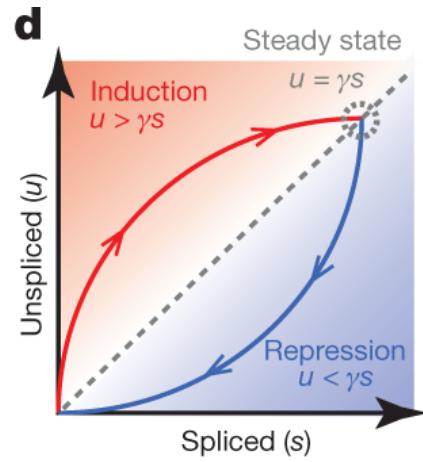
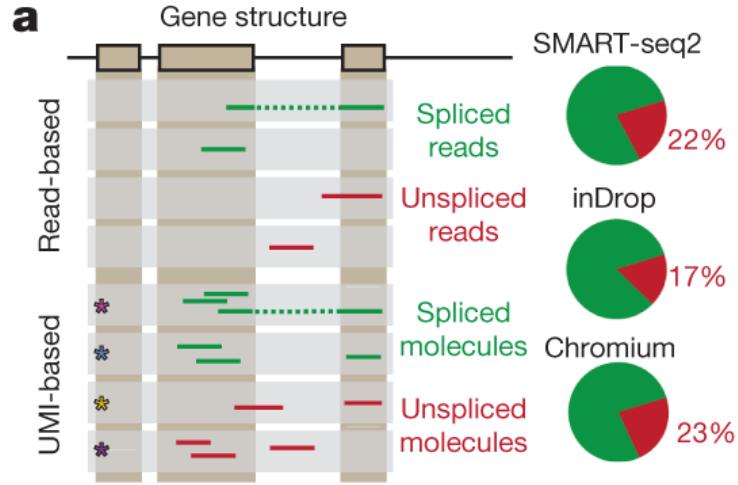
c



The ratio unspliced-to-spliced is proportional to length of (de)activation of a gene

- $u/s > 1$ : gene was recently activated
- $u/s < 1$ : gene was recently deactivated

# RNA velocity of single cells

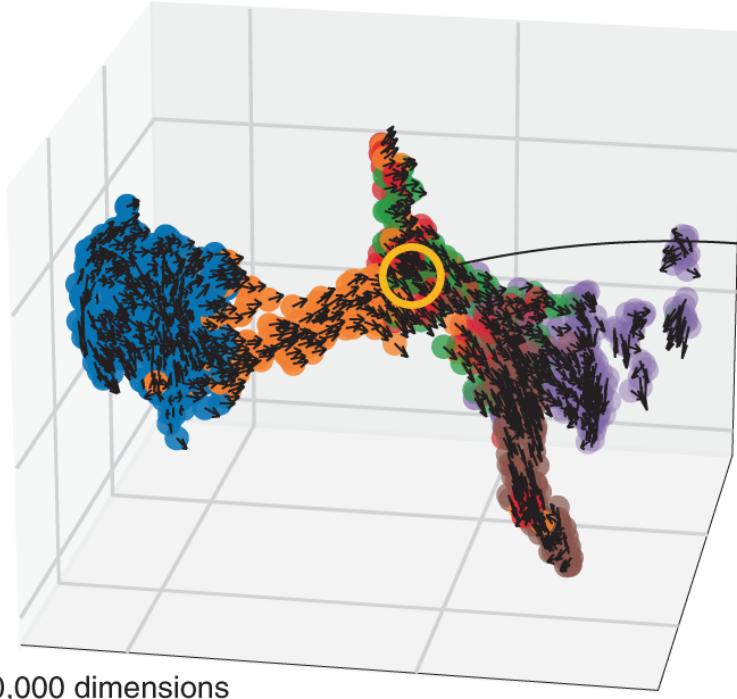


# Many RNA Velocity models/tools

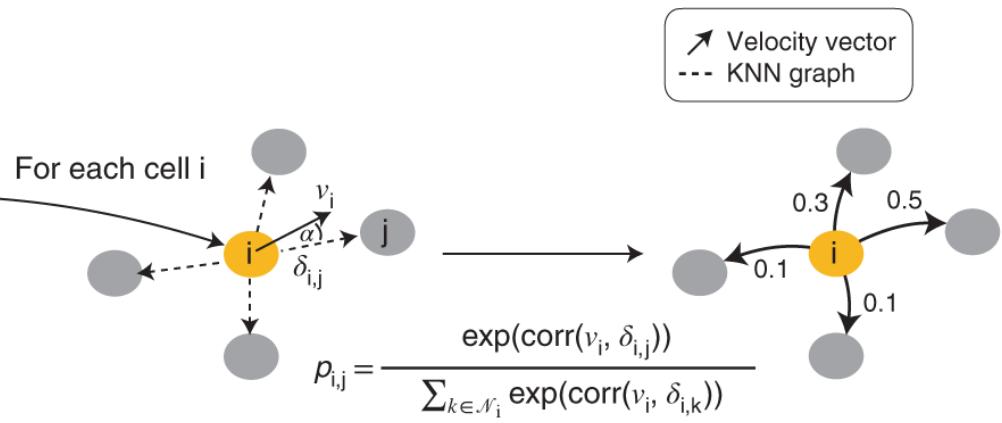
1. velocyto (2018) - La Manno et al., Nature
  2. scVelo (2020) - Bergen et al., Nature Biotechnology
  3. VeloSim (2021) - Zhang et al., Nature Computational Science
  4. kallisto bustools/kb-python (2021) - Melsted et al., Nature Biotechnology
  5. dynamo (2022) - Qiu et al., Nature Biotechnology
  6. VelocityTrain (2023) - Gorin et al., Nature Methods
  7. velovi (2023) - Jia et al., Nature Methods
- ...

# CellRank: Combining RNA velocity with cell–cell similarity

a



b



# Which method should you use?

<http://guidelines.dynverse.org>

Topology

Do you expect multiple disconnected trajectories in the data?

Yes I don't know No

Scalability

Number of cells: 1000

Number of features (genes): 1000

Time limit: 10s (1h)

Memory limit: 100MB (30GB)

Prior information

Are you able to provide the following prior information?

Start cell(s), End cell(s), # end states, # start states, # leaves, # states, Marker

Method selection

Benchmarking metrics

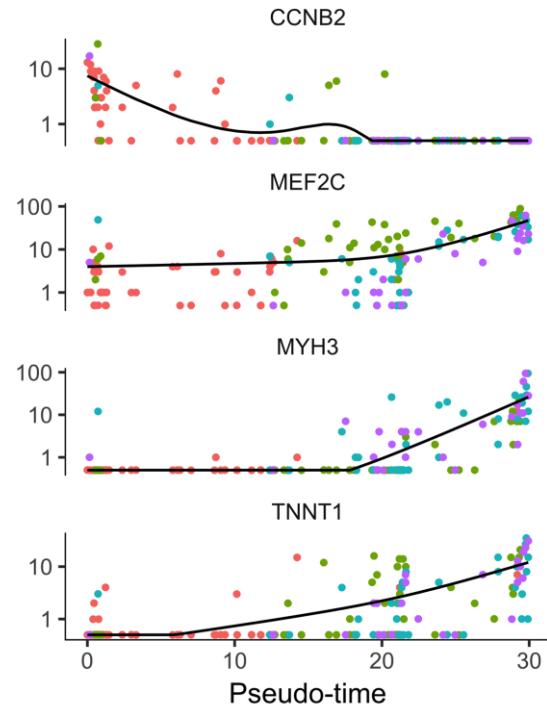
Show code ↗		Show/hide columns ⓘ		Options ⚙		Infer trajectories with dyno ↗			
Lenses	Default	Summary (Fig. 2)	Method	Scalability	Stability	Usability	Accuracy	Overall	Everything
Method						Accuracy	Scalability	Stability	
Slingshot	Name	Priors	Errors	Overall	8s	942MB			
PAGA Tree	Slingshot	X	100	19s	625MB	Unstable			
SCORPIUS	Slingshot	X	99	3s	507MB				
Angle	Slingshot	X	96	1s	308MB				
PAGA	Angle	X	92	15s	559MB	Unstable			
Embeddr	Angle	X	89	5s	591MB				
MST	Angle	X	89	4s	572MB	Unstable			
Waterfall	Angle	X	89	5s	369MB				
TSCAN	Angle	X	88	5s	476MB	Unstable			
Component 1	Angle	X	87	1s	516MB				
SLICE	Angle	X	83	16s	713MB				
EPIGraph linear	Angle	X	81	1m	573MB				
PhenoPath	Angle	X	79	5m	837MB				
pCreode	Angle	X	78	2m	444MB	Unstable			
Monocle ICA	Angle	X	78	1m	692MB	Unstable			
Wanderlust	Angle	X	78	51s	413MB				
MATCHER	Angle	X	77	43s	385MB				
Wishbone	Angle	X	76	1m	370MB				

# Pseudotime analysis

# Interpretation of gene behaviour

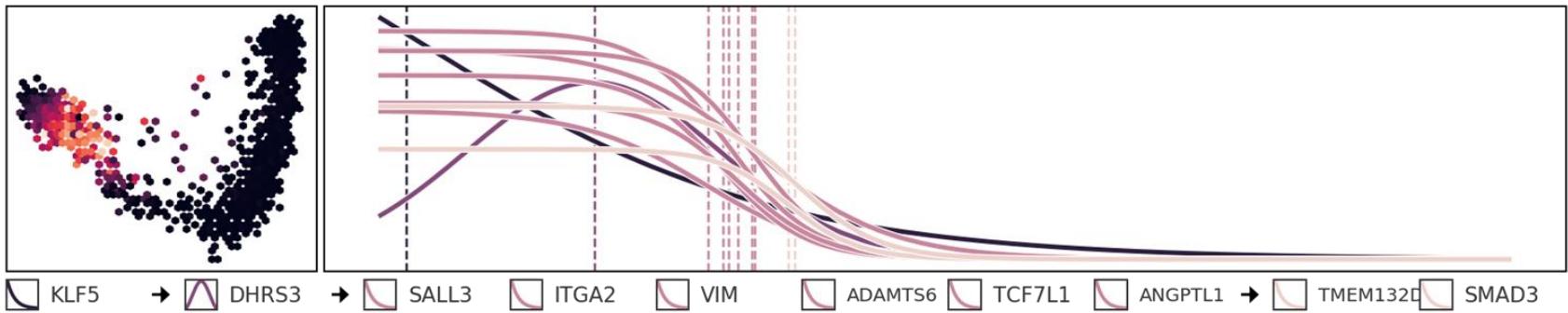
- Plot the gene expression as a function of pseudotime
- What do we see?

Pseudotime-gene expression pattern



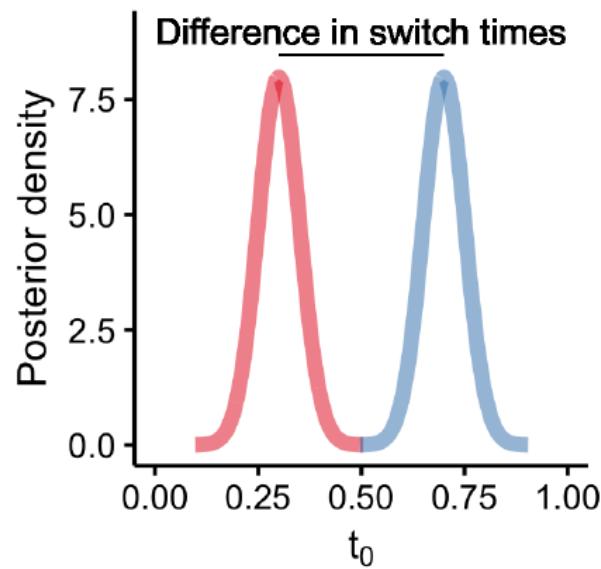
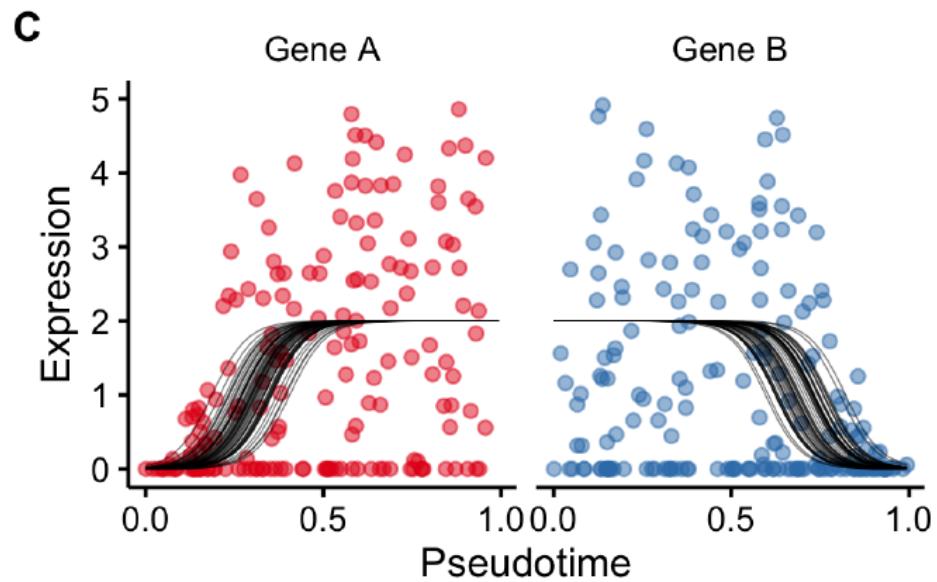
# Interpretation of gene behaviour

Charroud et al., 2020

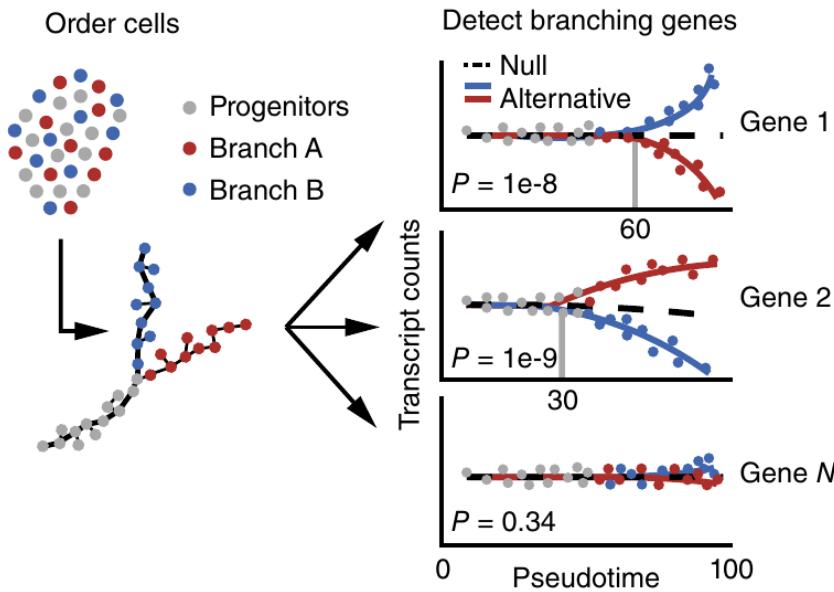


- KLF5 member of Kruppel-like family of transcription factors
  - Repressor of neurite growth, down-regulation linked to cell cycle arrest
- VIM, highly variable gene
  - Known marker of gliogenesis

# Differential activation testing



# Branch-dependent gene expression



Null model:

$$y \sim \text{pseudotime}$$

Alternative model:

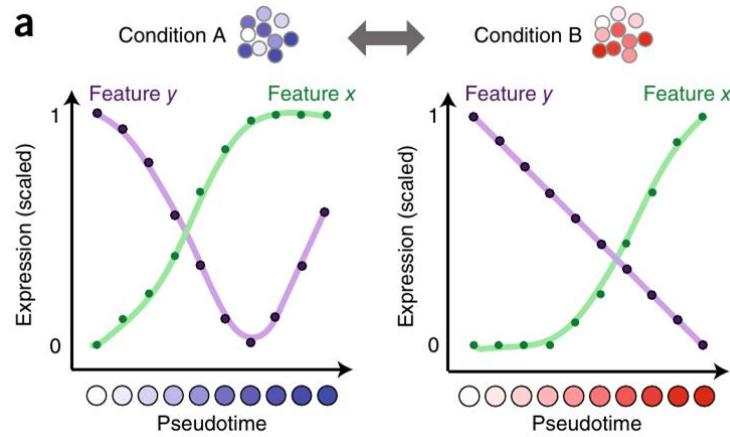
$$y \sim \text{pseudotime} + \text{branch} + \text{pseudotime:branch}$$

Likelihood ratio test:

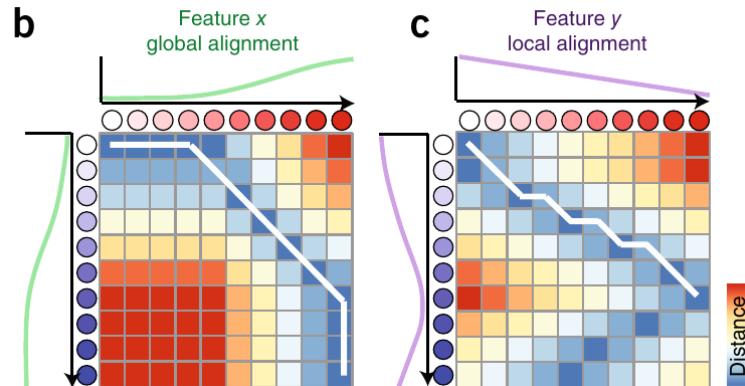
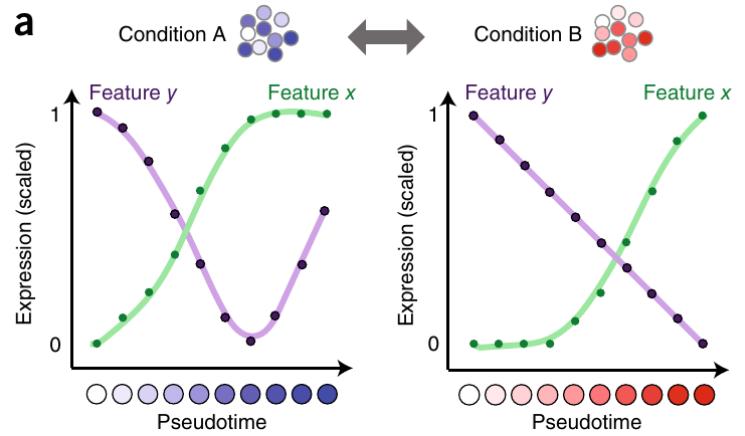
$$\text{ratio} = L(\text{alternative}) / L(\text{null})$$

if ratio  $\geq 0 \rightarrow$  branch-dependent expression

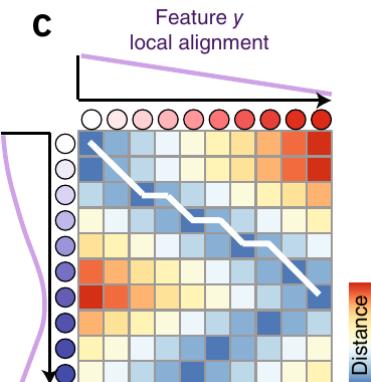
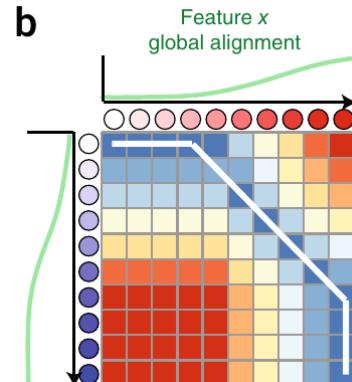
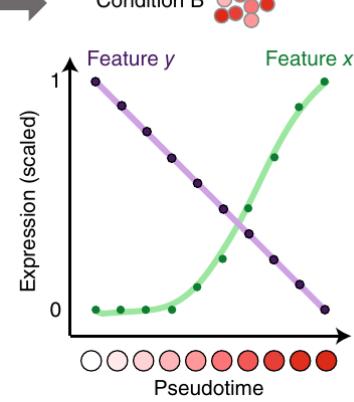
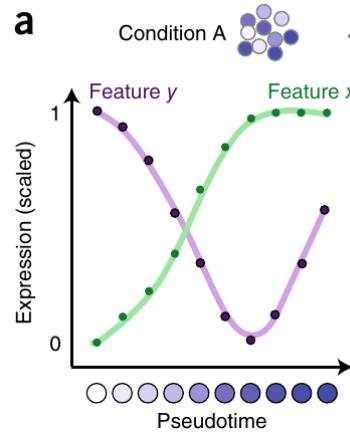
# Comparing trajectories



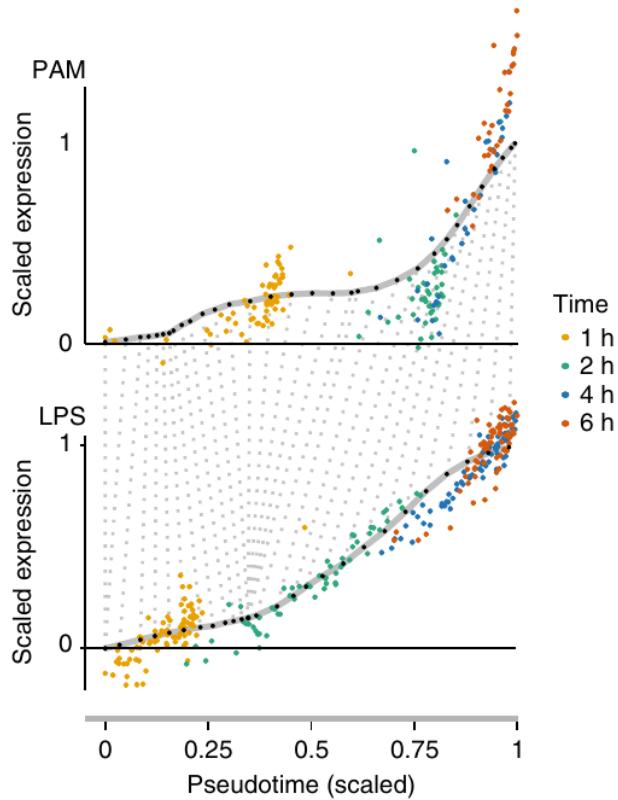
# Comparing trajectories



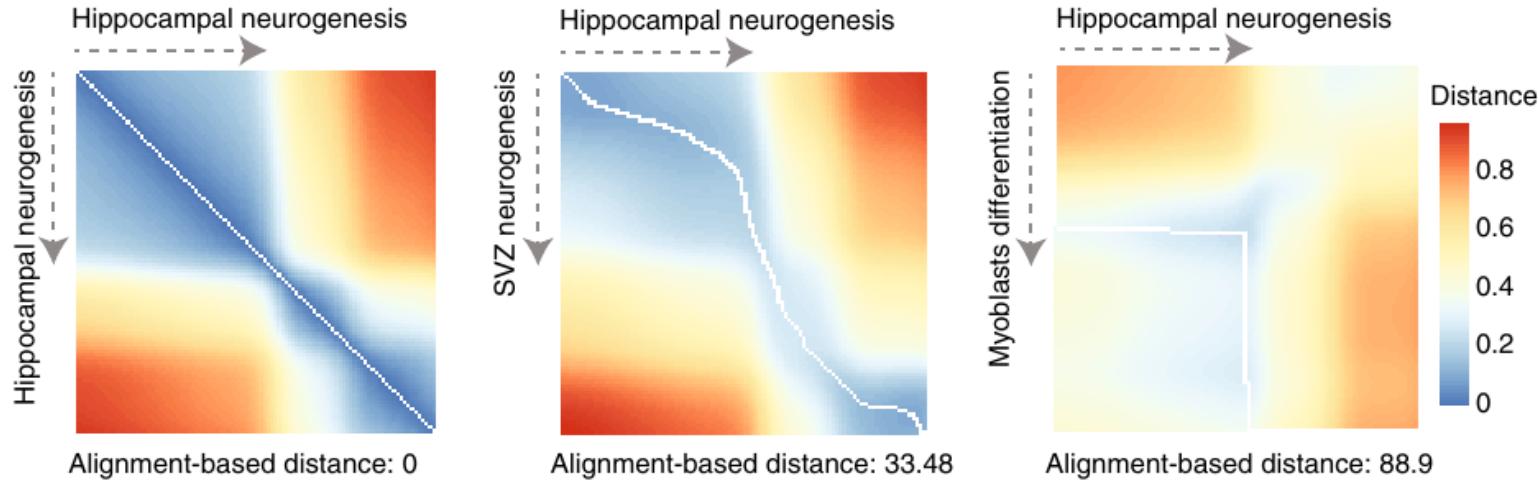
# Comparing trajectories



## Dynamic time warping



# Comparing trajectories



# Summary

- Trajectory inference can be used to order cells based on their transcriptional similarity
- Trajectory inference methods assign a pseudotime to each cell
- RNA velocity offers an alternative where we try to predict the how a cell will look like in the near future (no pseudotime)
- CellRank unifies trajectory inference (transcriptional similarity) and RNA velocity
- We can identify gene/features associated with particular trajectories

# Thank You!

 a.mahfouz@lumc.nl  
 mahfouzlab.org



[www.singlecell.nl](http://www.singlecell.nl)

Annual meeting: 3 April 2025 (Nijmegen)

# Spatial data analysis is challenging

NGS-based

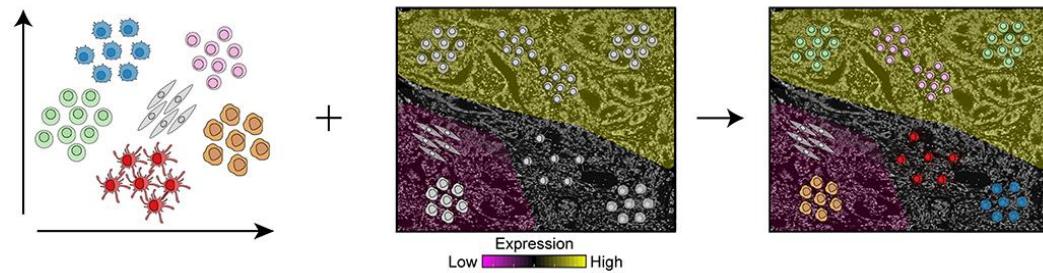
- + Full transcriptome
- Low resolution (spots)

In-situ  
sequencing/hybridization

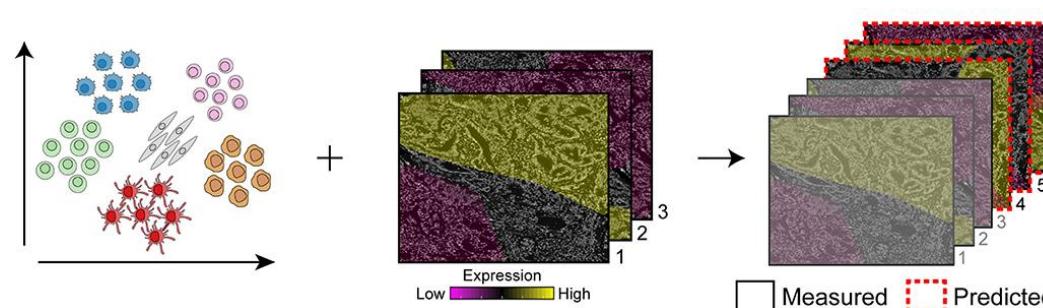
- + High resolution
- Limited genomic features

# Single-cell + spatial

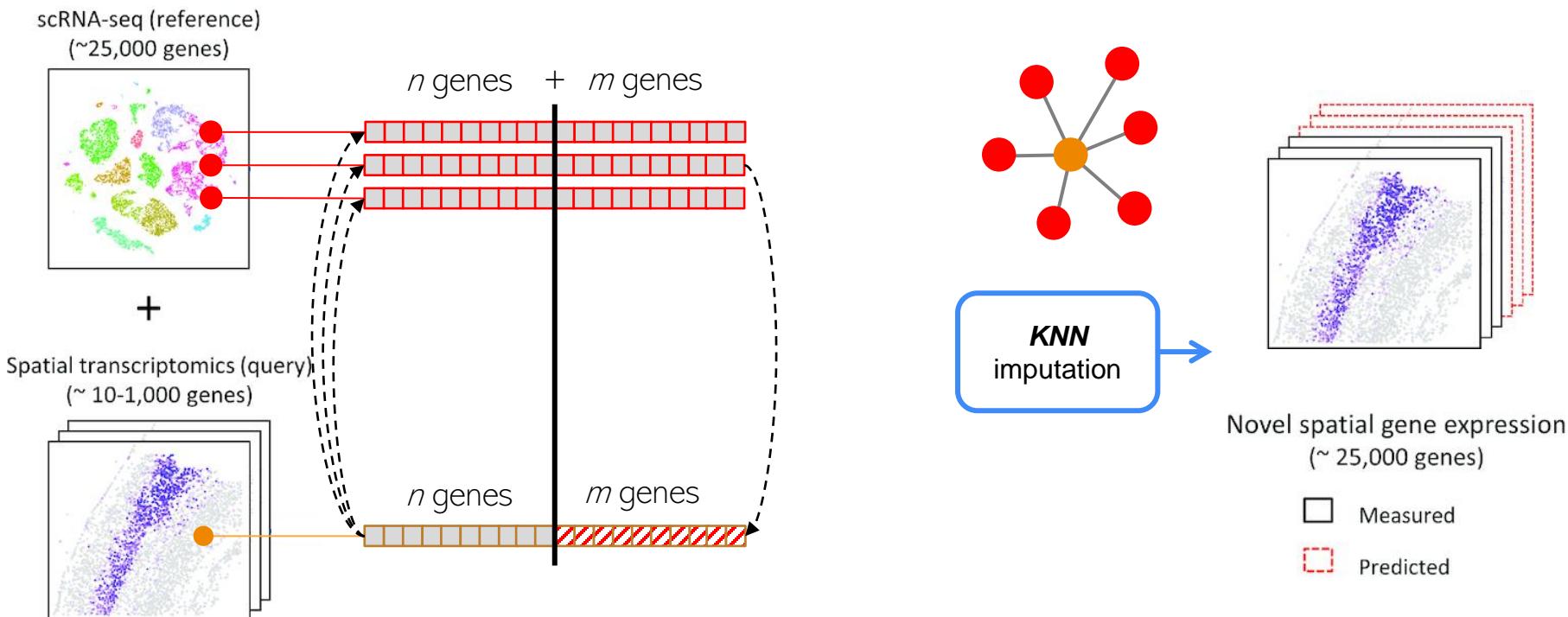
Predict the location of dissociated cells



Predicting spatial expression of unmeasured genes

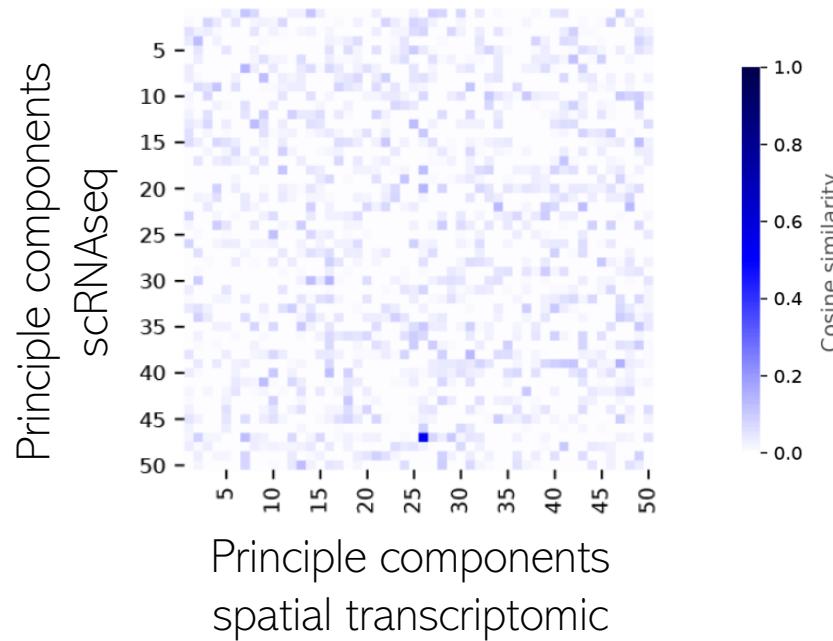


# SpaGE: Spatial Gene Expression Enhancement



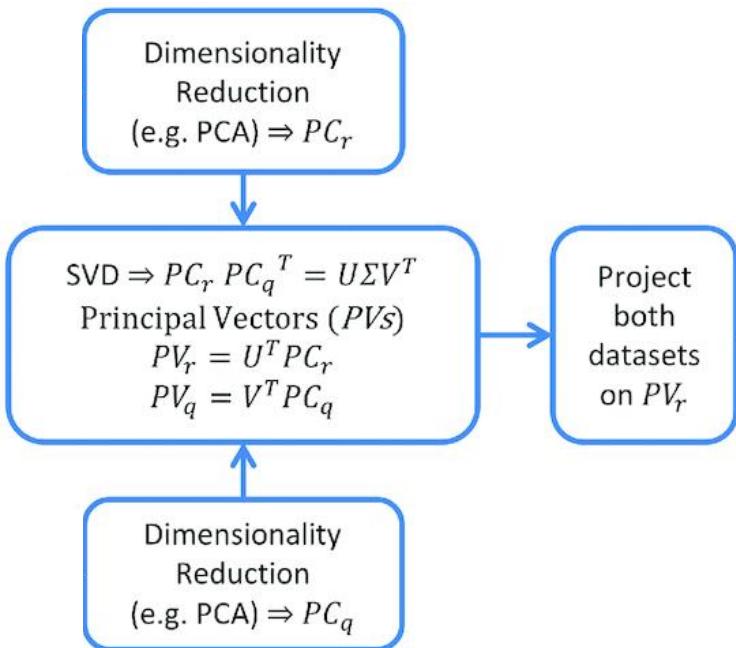
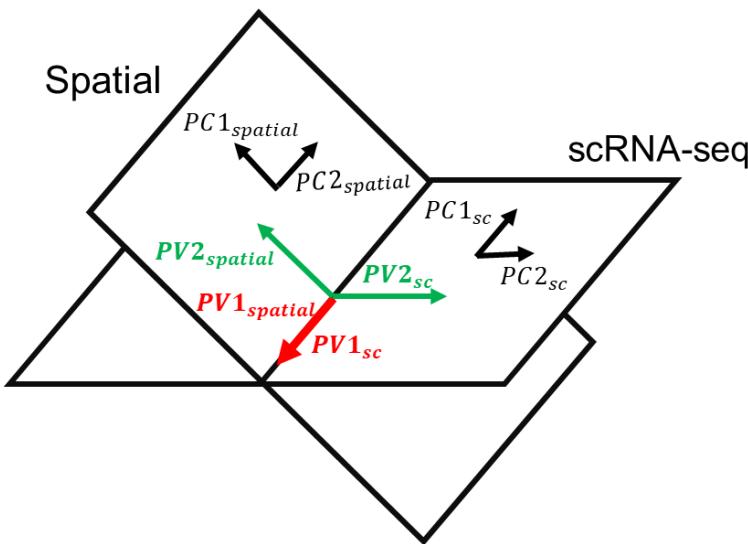
# Problem: single-cell and spatial data don't align

Similarity between principal components

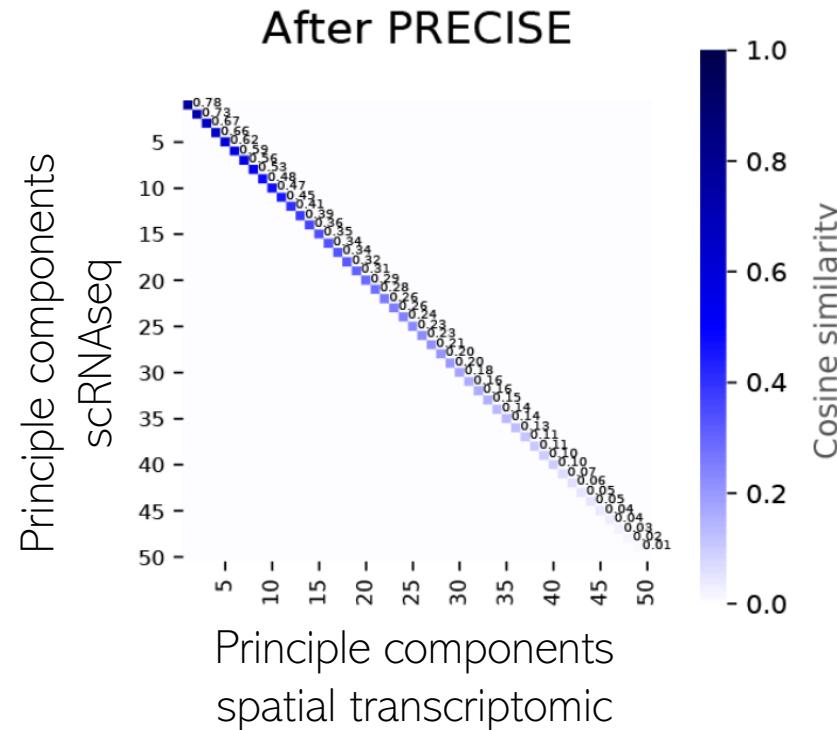
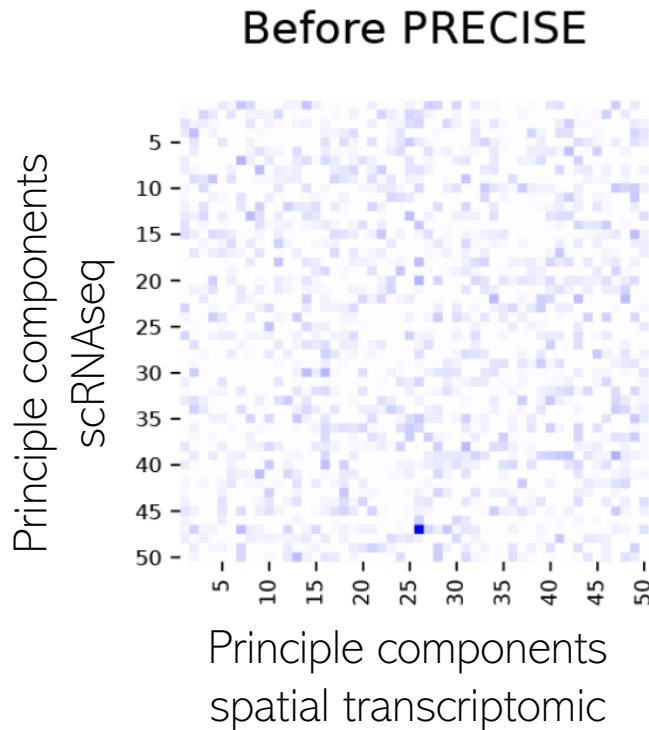


# Aligning single-cell and spatial data

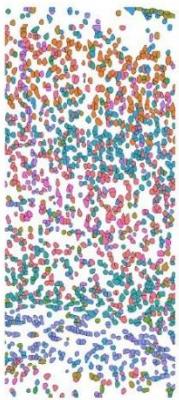
Domain Adaptation using PRECISE



# Aligning single-cell and spatial data

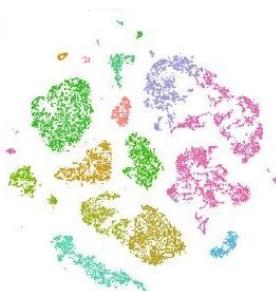


# SpaGE in primary visual cortex (V1Sp)

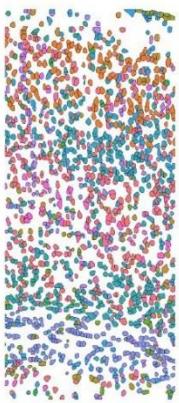


STARmap  
1,549 cells  
1,020 genes  
Wang et al. Science 2018

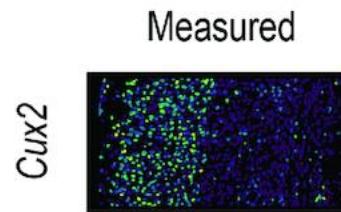
scRNA-seq  
14,249 cells  
34,617 transcripts  
Tasic et al. Nature 2018



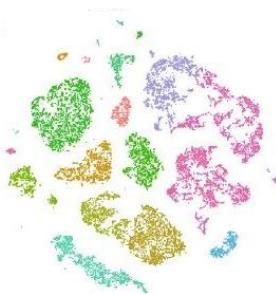
# SpaGE in primary visual cortex (V1Sp)



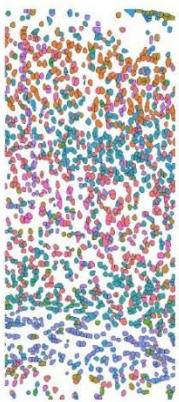
STARmap  
1,549 cells  
1,020 genes  
Wang et al. Science 2018



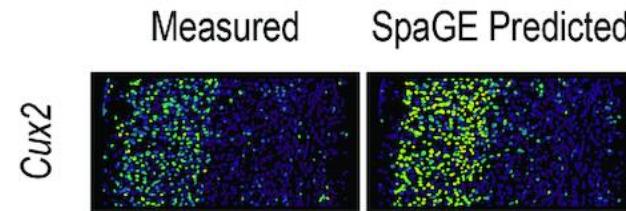
scRNA-seq  
14,249 cells  
34,617 transcripts  
Tasic et al. Nature 2018



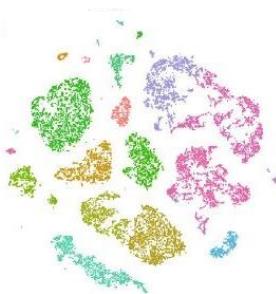
# SpaGE in primary visual cortex (V1Sp)



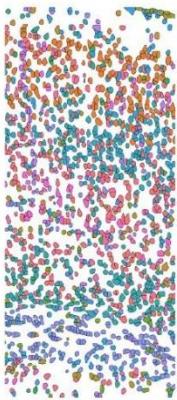
STARmap  
1,549 cells  
1,020 genes  
Wang et al. Science 2018



scRNA-seq  
14,249 cells  
34,617 transcripts  
Tasic et al. Nature 2018



# SpaGE in primary visual cortex (V1Sp)

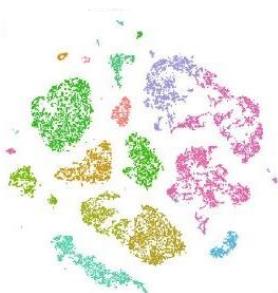


STARmap

1,549 cells

1,020 genes

Wang et al. Science 2018

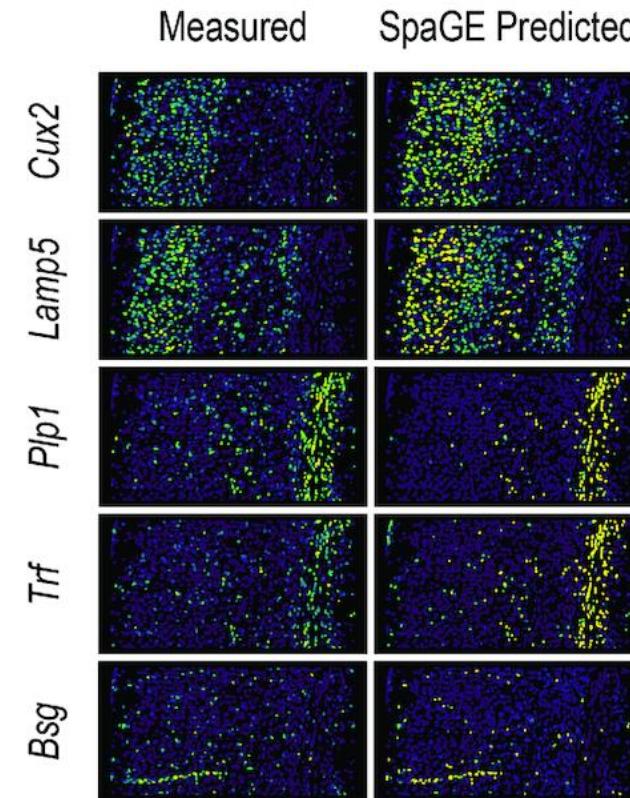


scRNA-seq

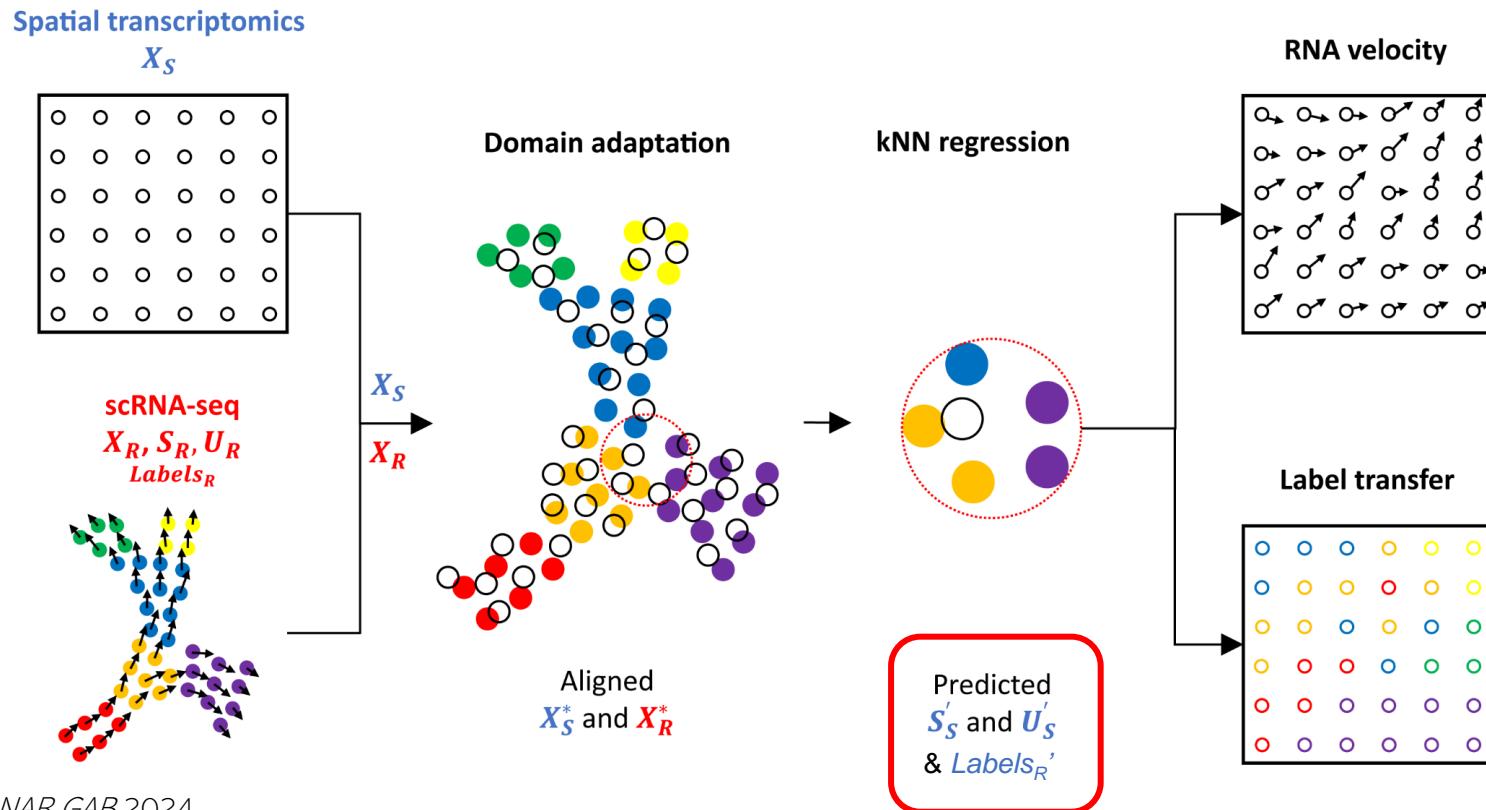
14,249 cells

34,617 transcripts

Tasic et al. Nature 2018

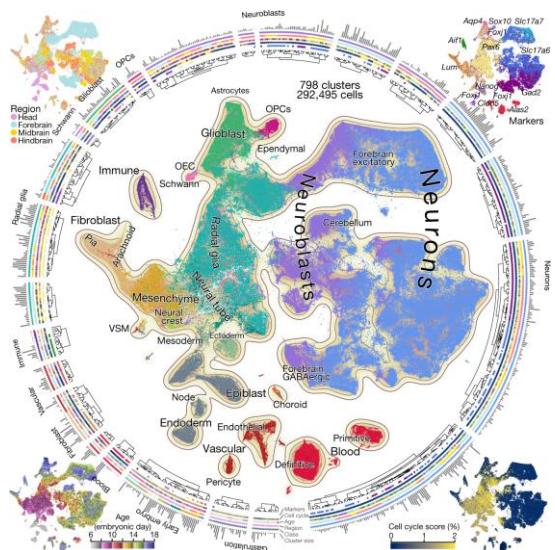


# SIRV: Spatially Inferred RNA Velocity

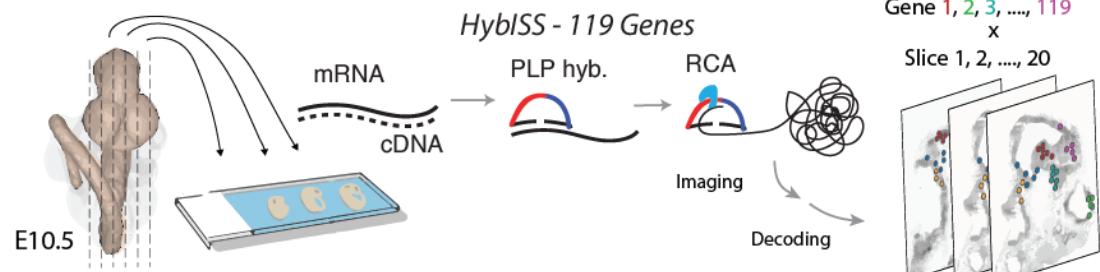


# Applying SIRV to the developing mouse brain

scRNA-seq atlas of the developing mouse brain

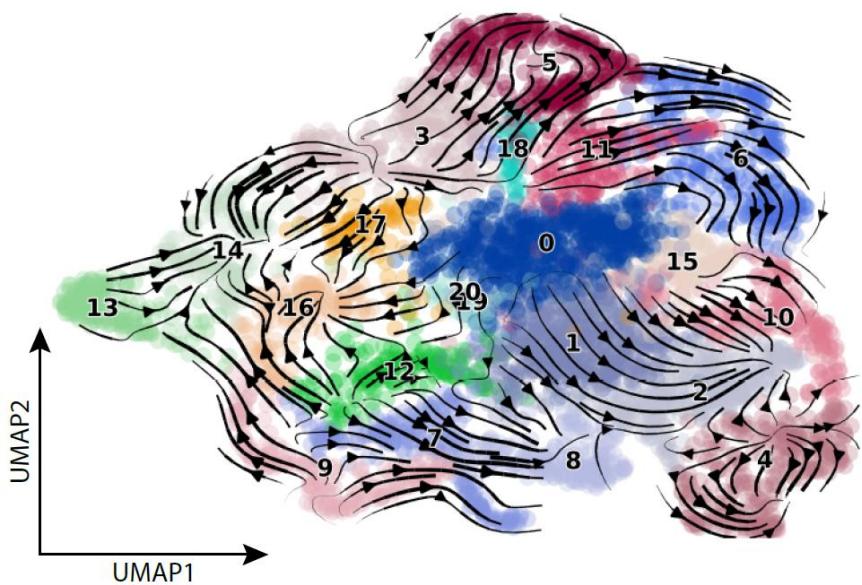


Matching HybISS data from E10.5  
(119 genes)

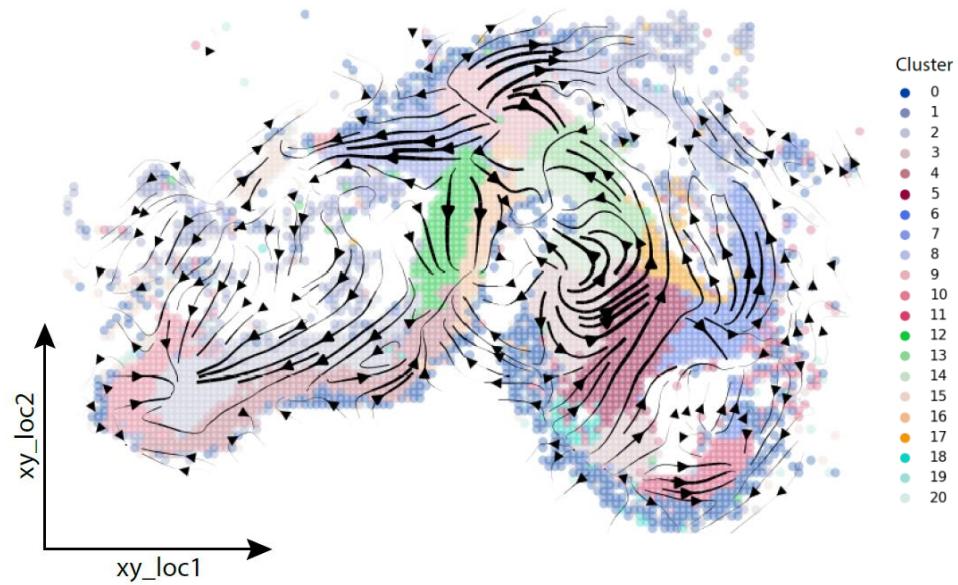


# SIRV infers RNA velocity of spatial data

Velocities on UMAP of spatial data



Velocities on HybLISS slide



Cluster

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20

# Velocity is driven by known differentiation markers

