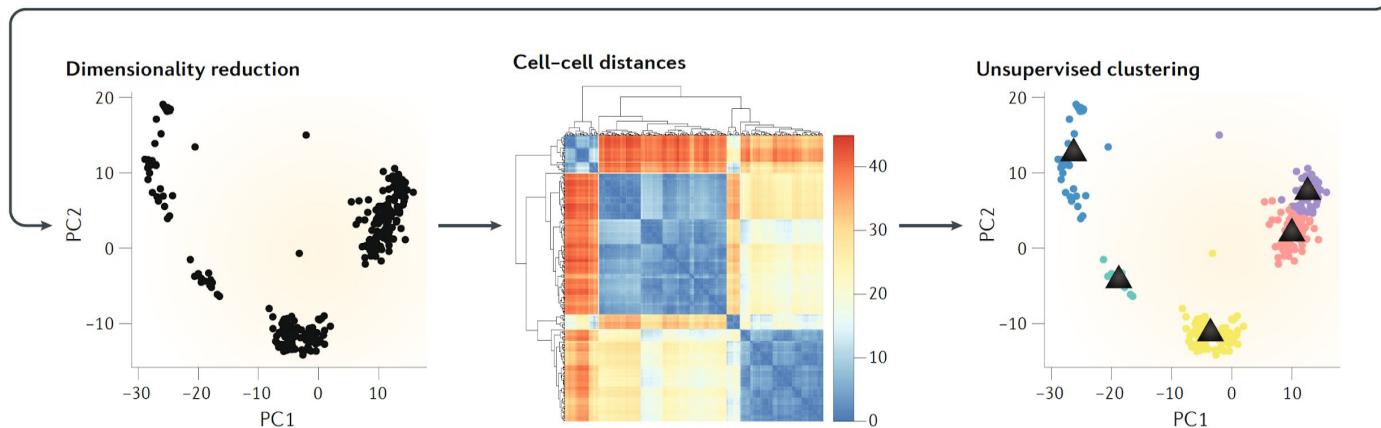
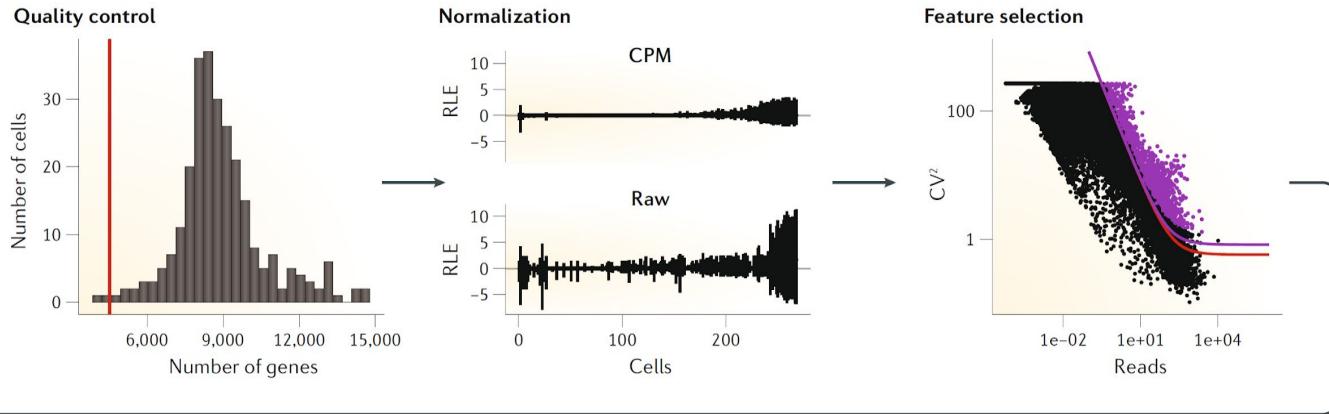


Trajectory inference

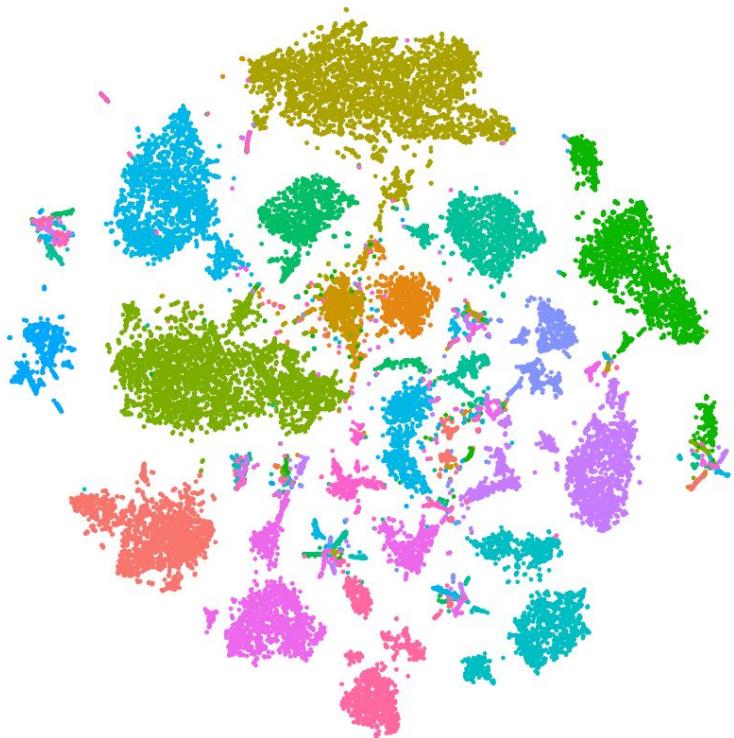
Mohammed Charrouf

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

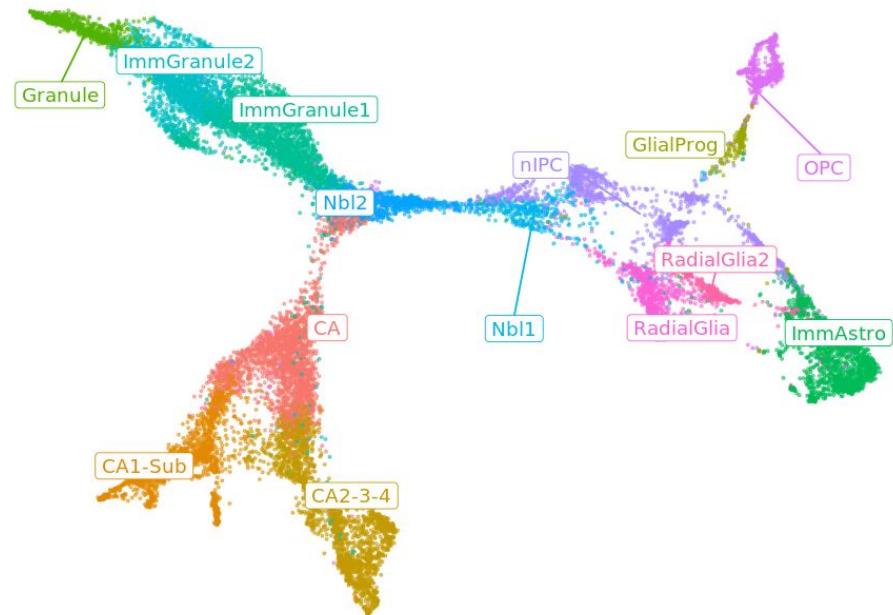
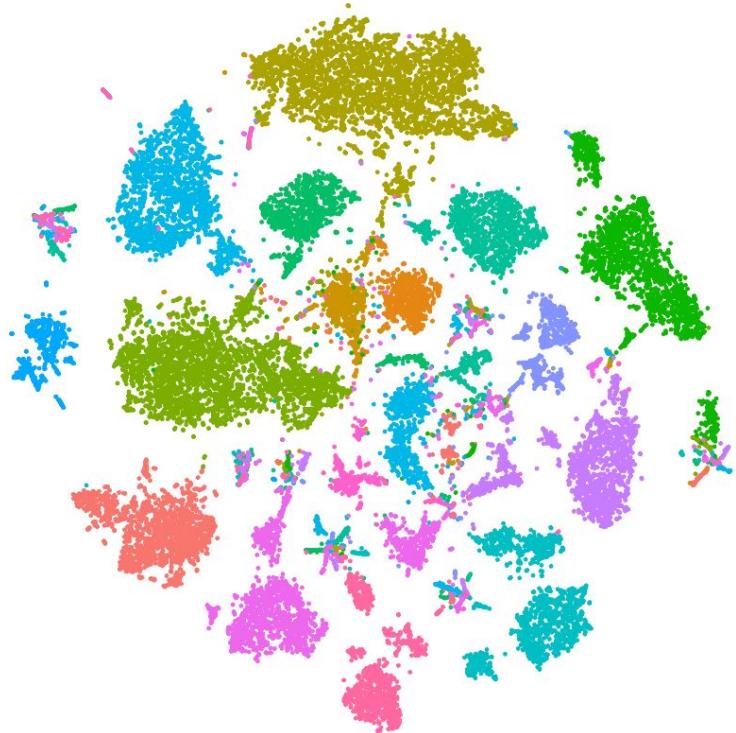
Single cell RNA-seq workflow



Clustering of differentiating cells

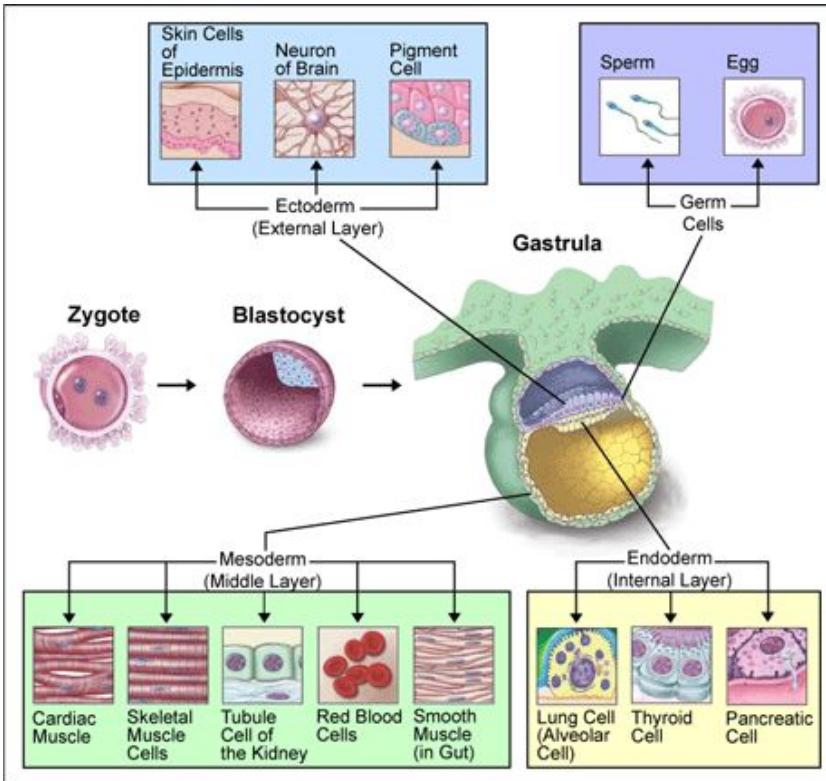


Clustering of differentiating cells



When do continuous structures pop up?

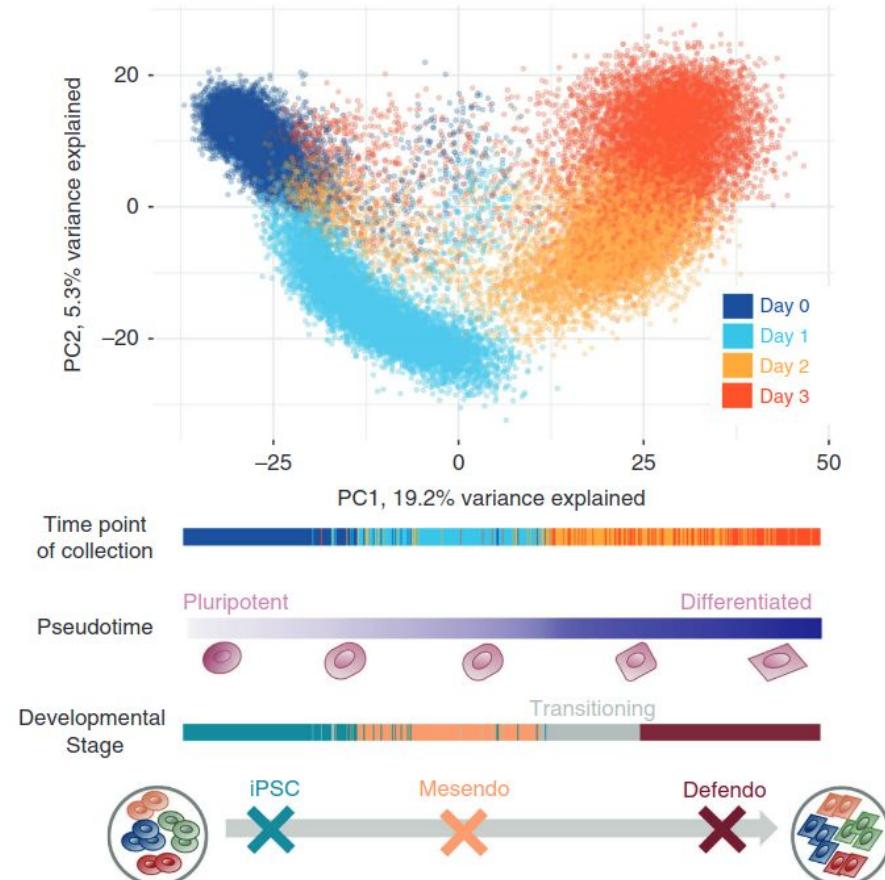
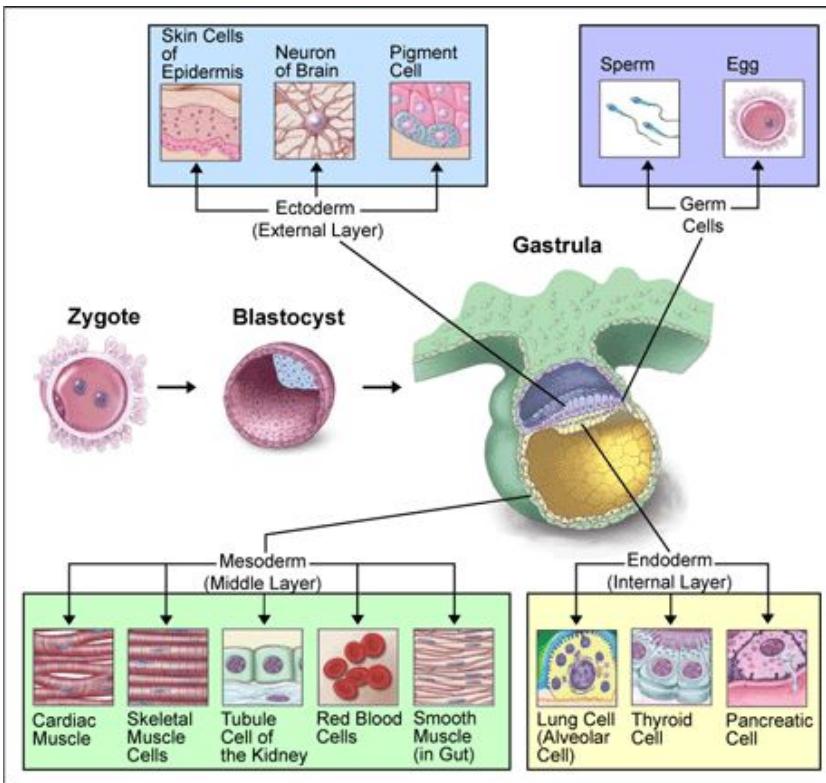
Embryonic development



When do continuous structures pop up?

Cuomo et al., 2020

Embryonic development

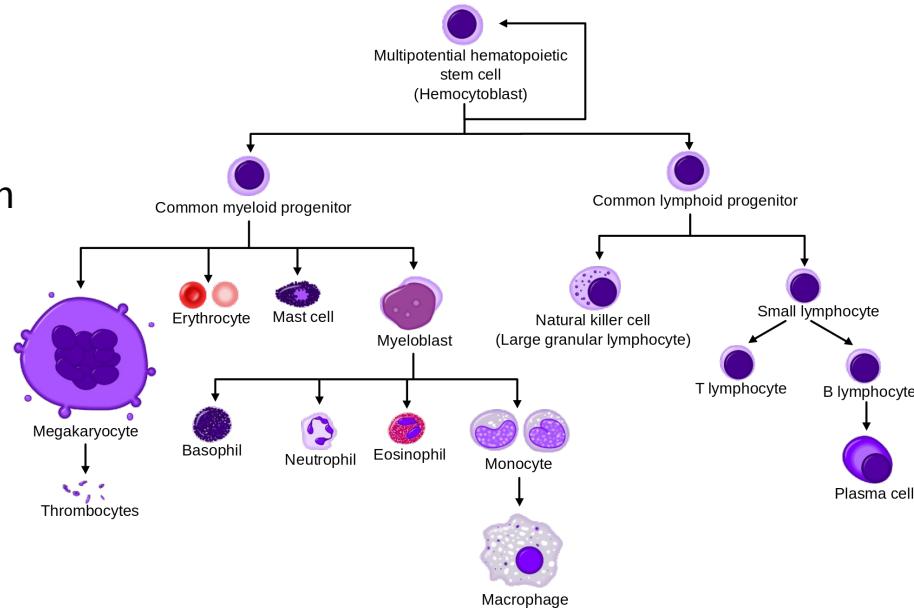


When do continuous structures pop up?

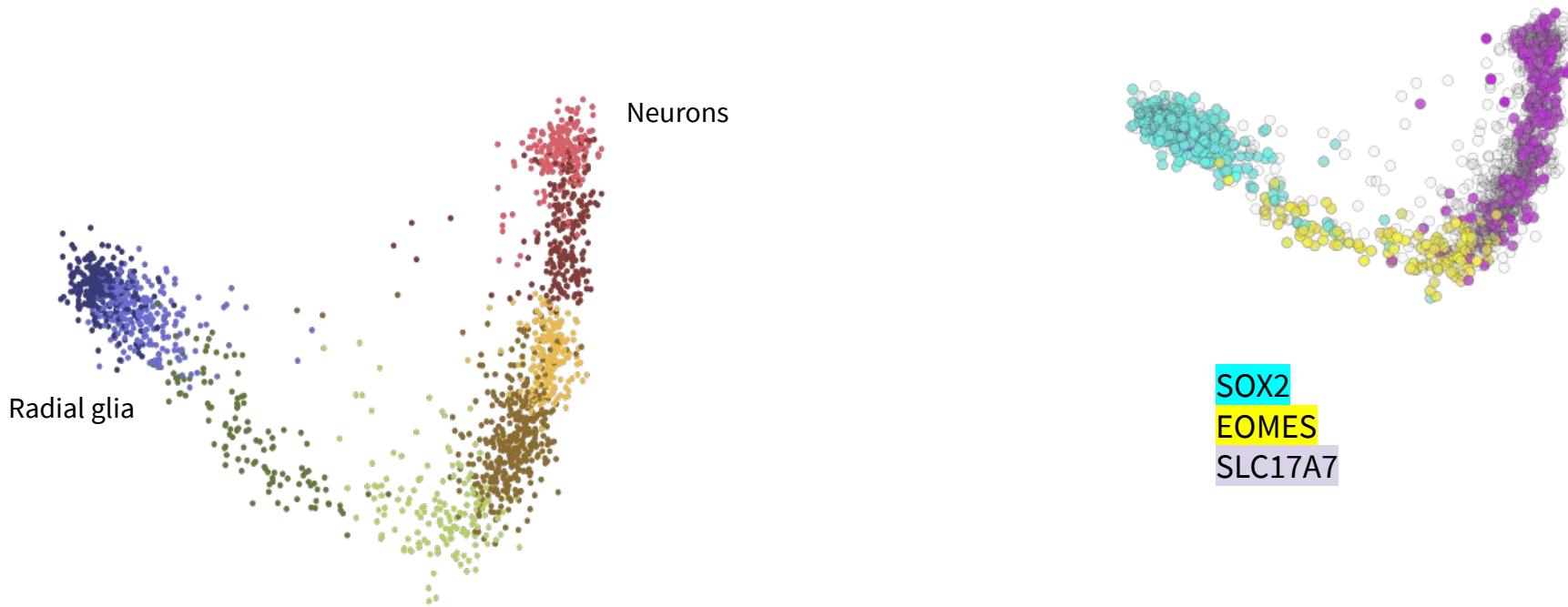
Cell differentiation

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

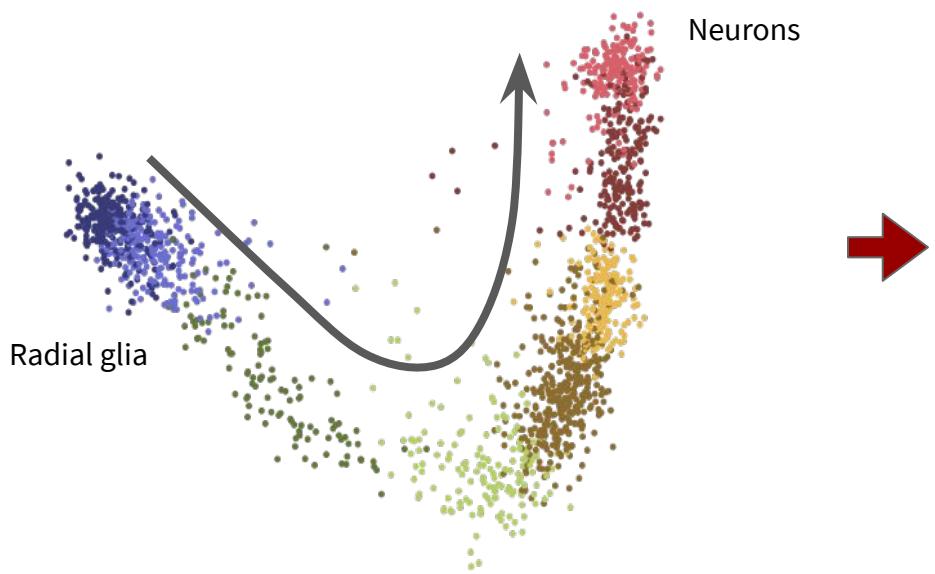
Hematopoiesis



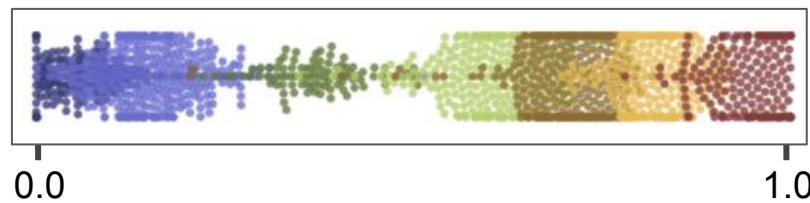
Trajectory inference / pseudotime inference



Trajectory inference / pseudotime inference

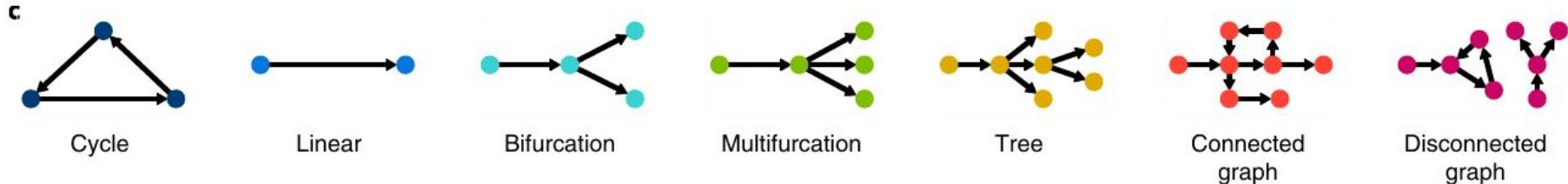


Pseudotime



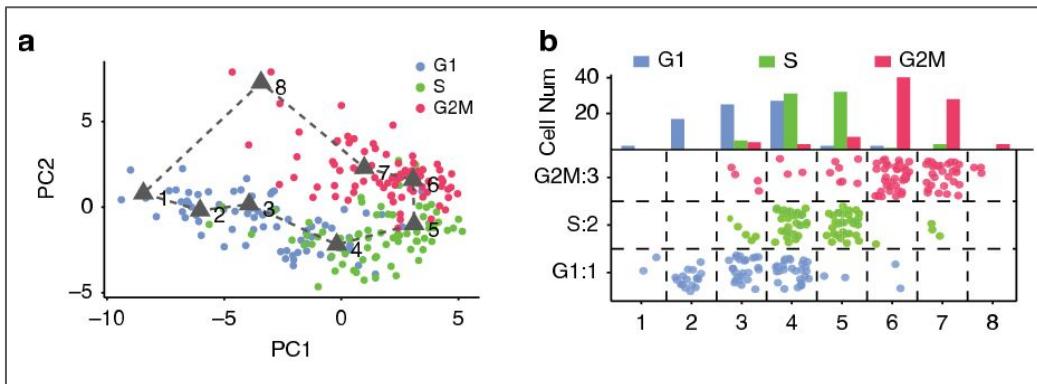
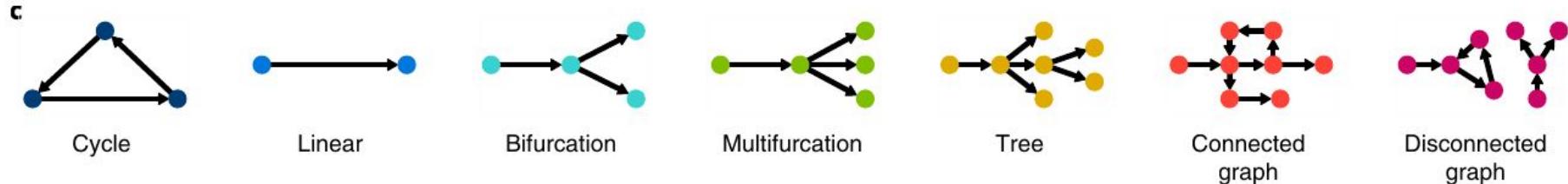
Trajectory structure

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



Trajectory structure

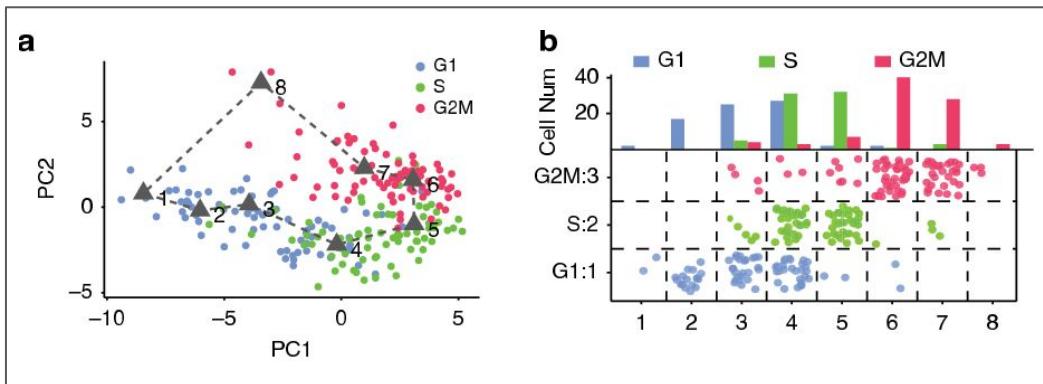
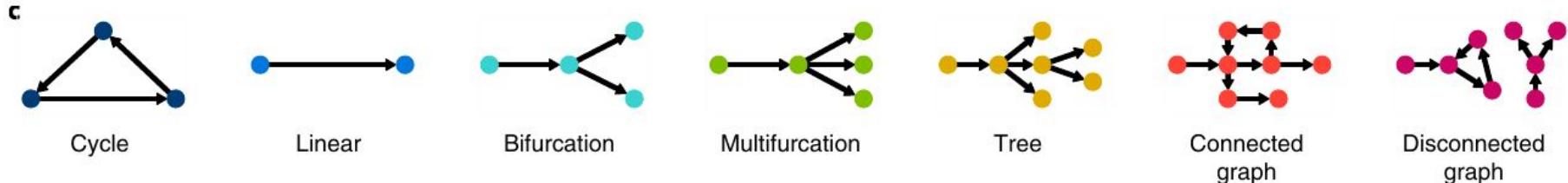
Saelens et al. (<https://doi.org/10.1038/s41587-019-0071-9>)



Liu et al. 2017

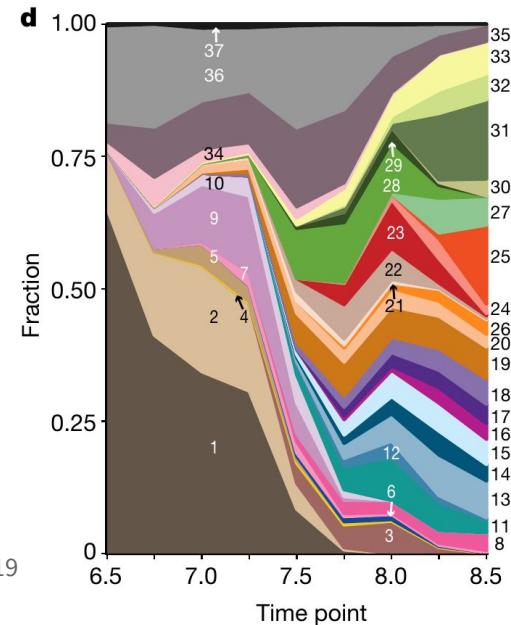
Trajectory structure

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



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

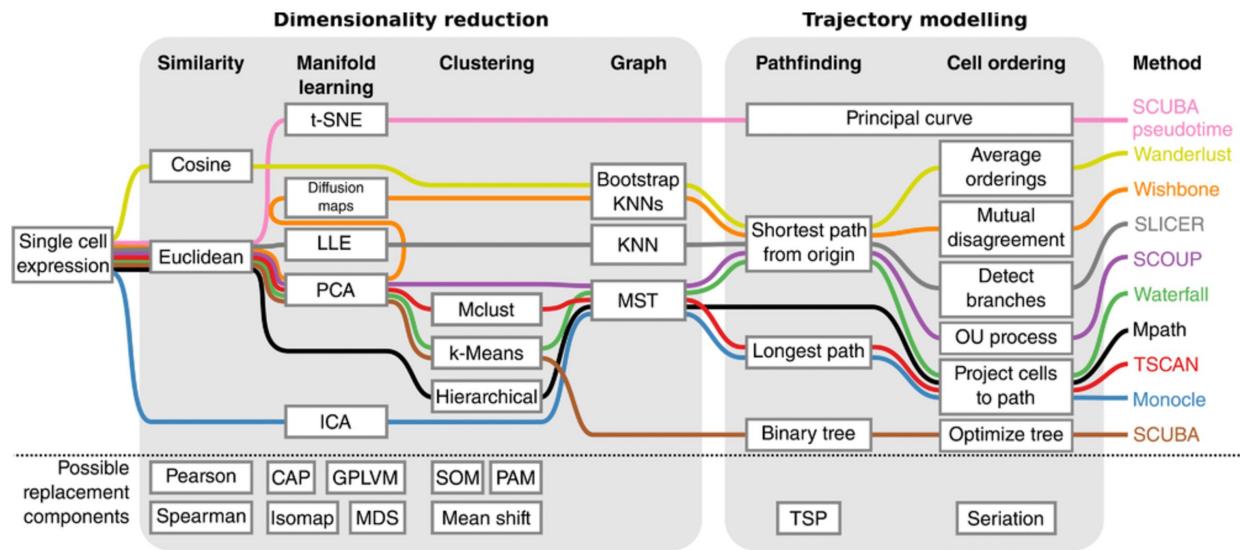
Pijuan-Sala et al., 2019



Methodology

General trajectory inference pipeline

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



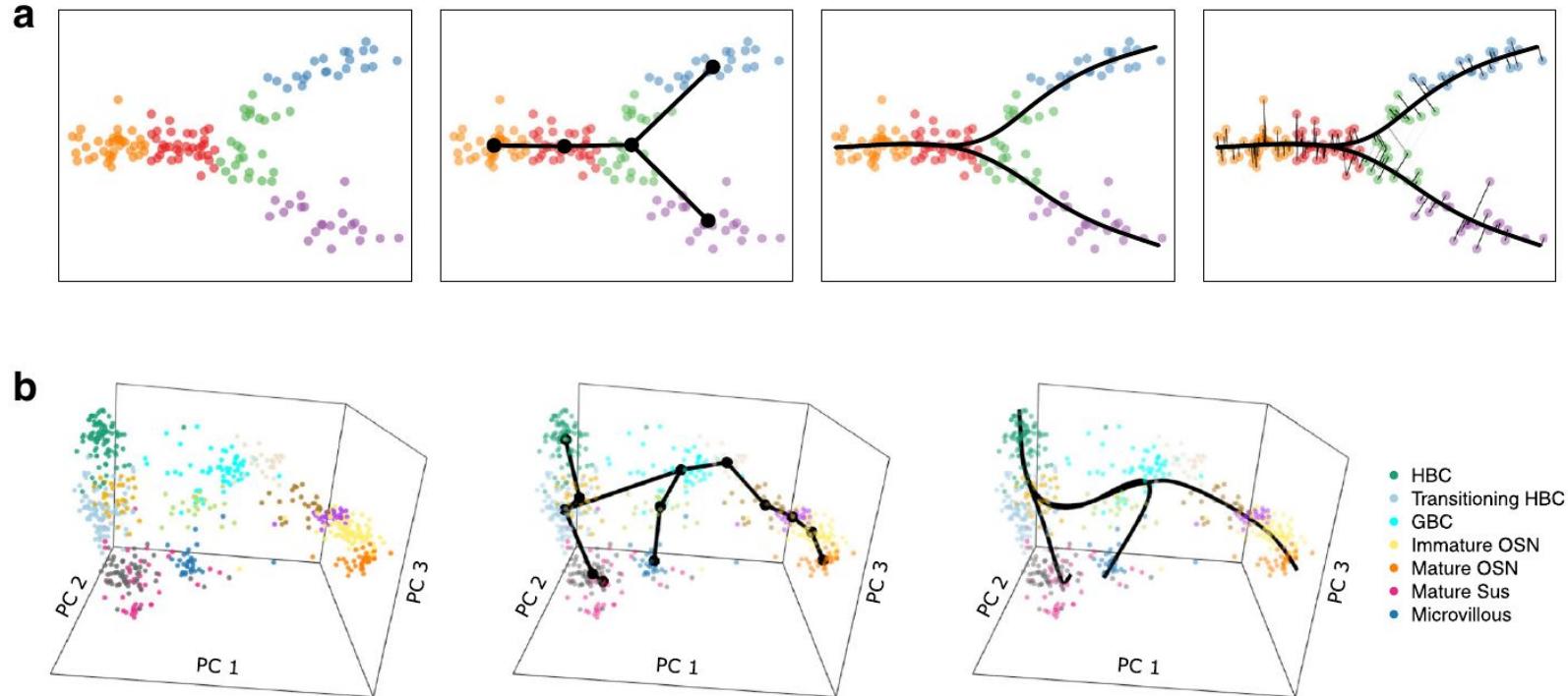
Trajectory inference methods

Four methods will be shown

- **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
 - Most 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

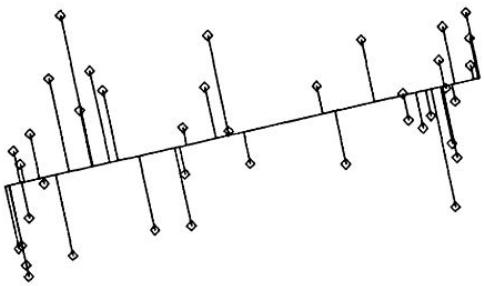
Slingshot

Fit multiple principle curves simultaneously, ensuring a shared trunk



Principle curves

b Principal component



d Principal curve

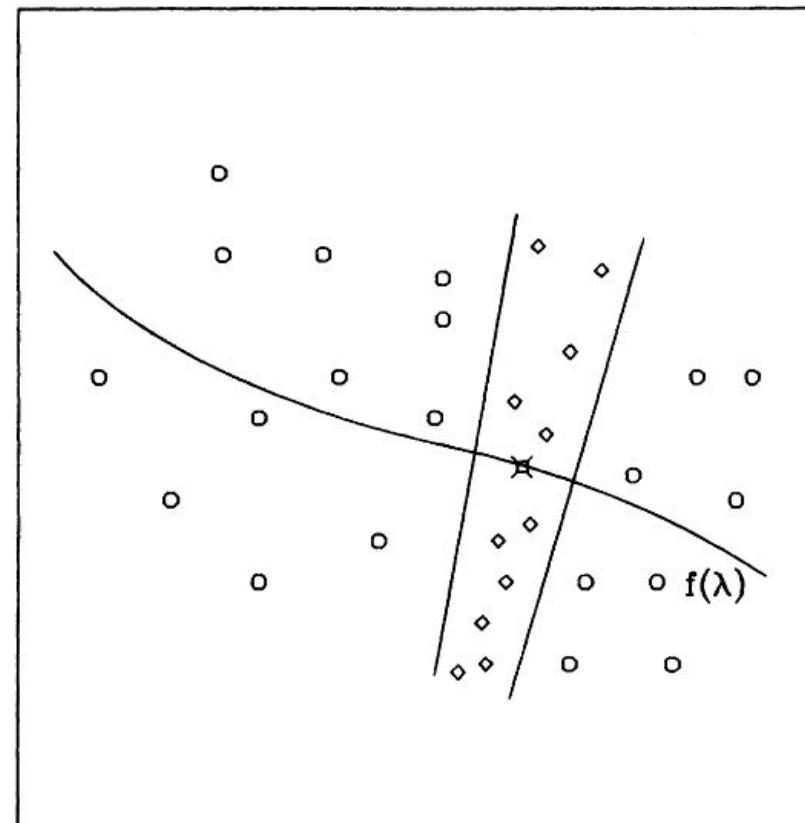
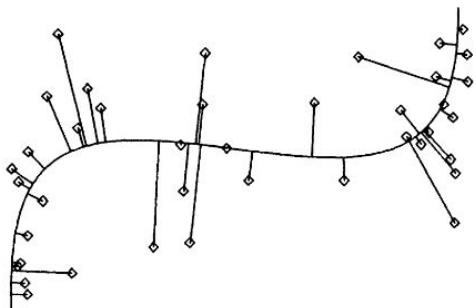
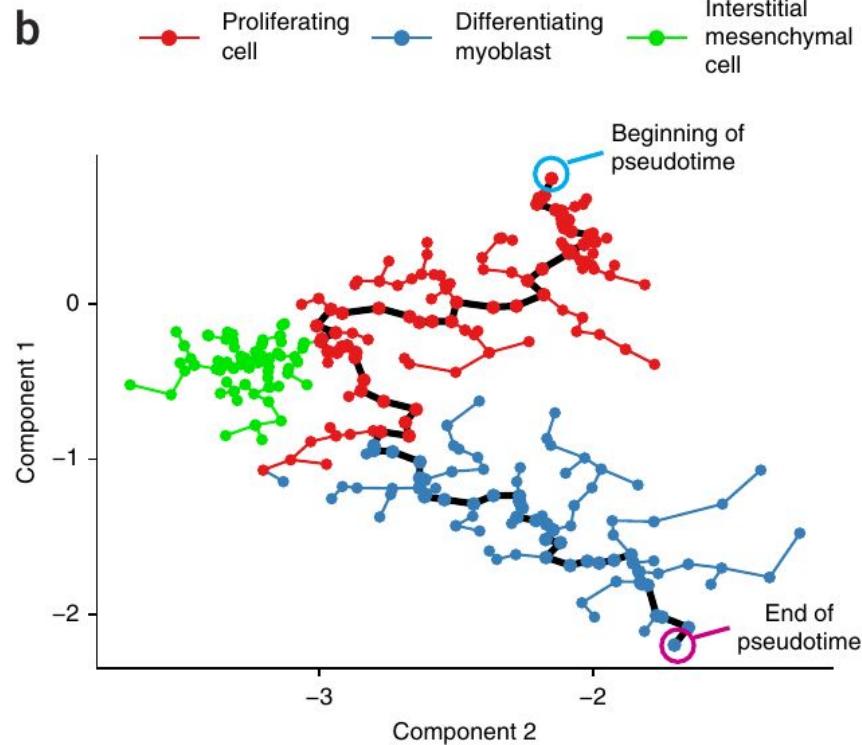
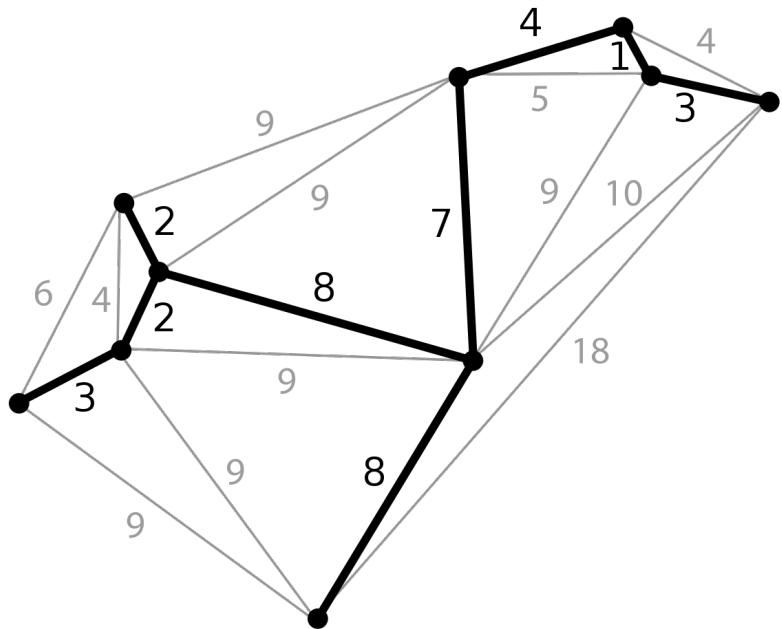


Figure 3. Each point on a principal curve is the average of the points that project there.

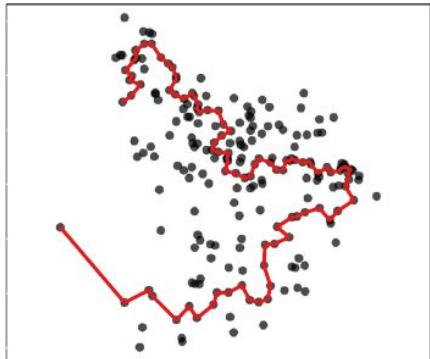
Monocle 1

Minimum spanning tree (MST)

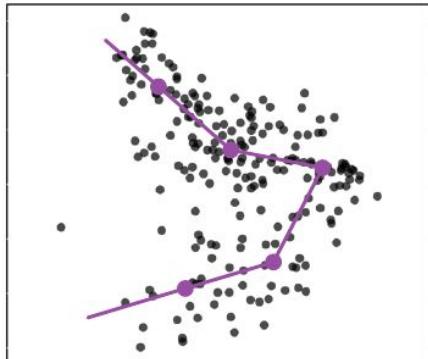


Trapnell et al., 2014

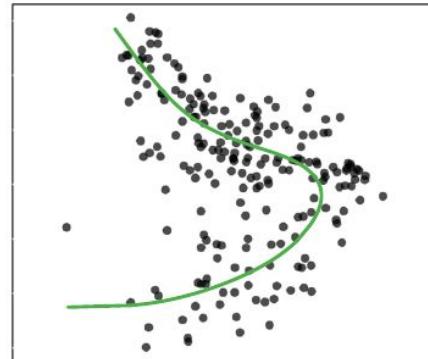
Principal curves vs MST



IC 1

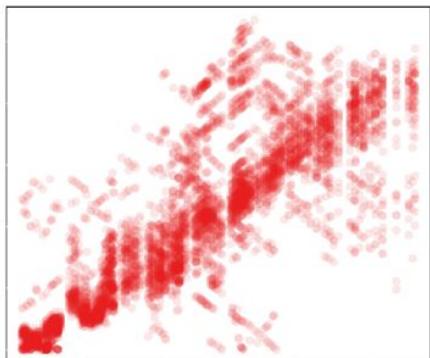


IC 1

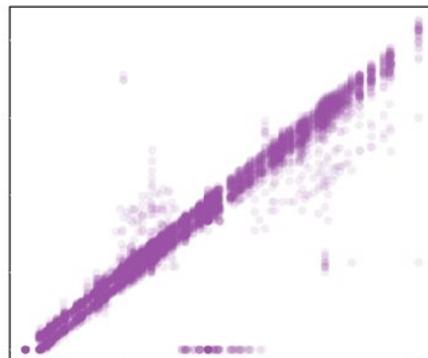


IC 1

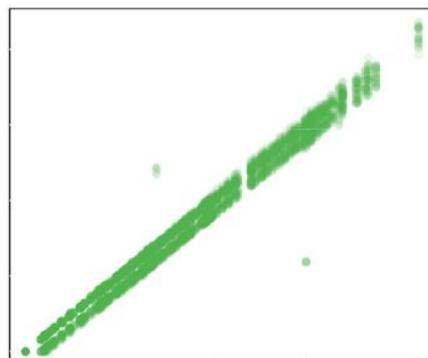
IC 2



Original pseudotime



Original pseudotime

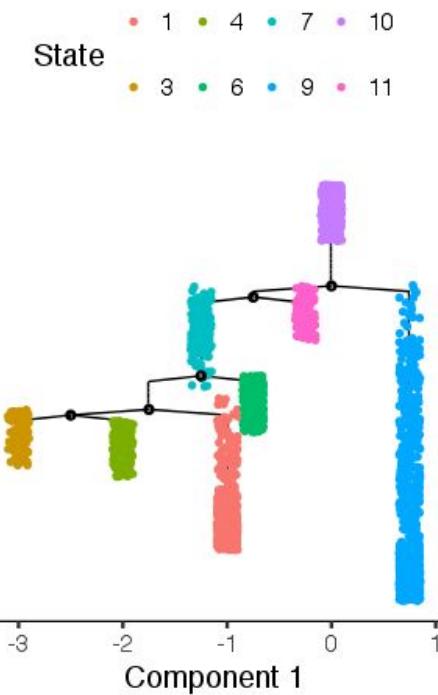
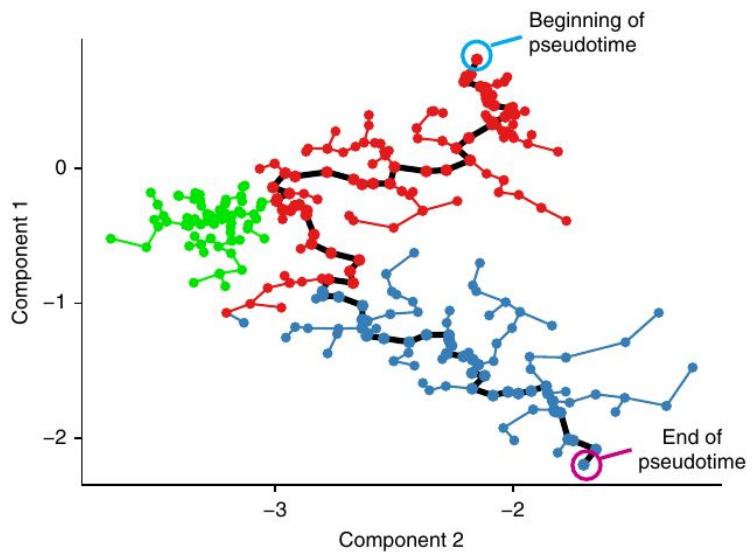


Original pseudotime

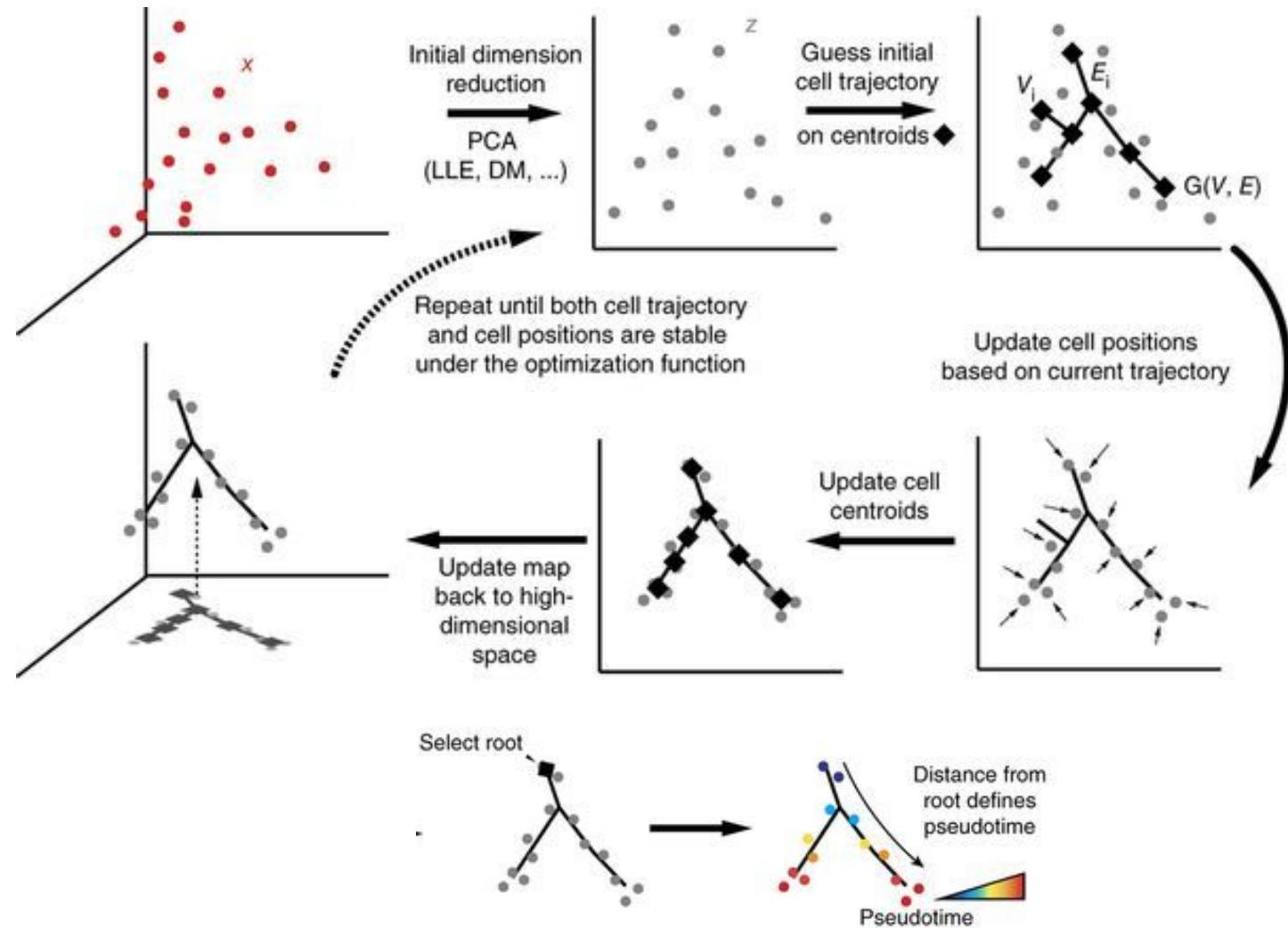
Subsample pseudotimes

Monocle 2

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

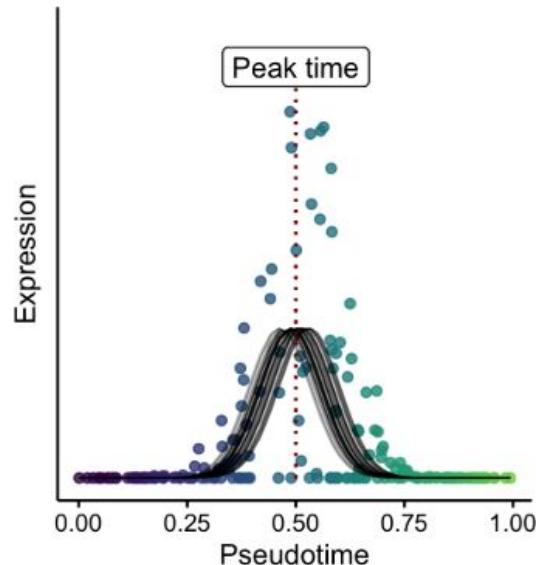
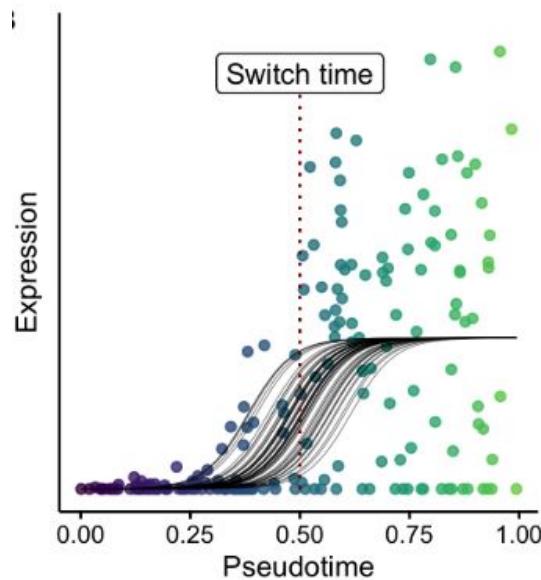


Monocle 2



Ouija

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

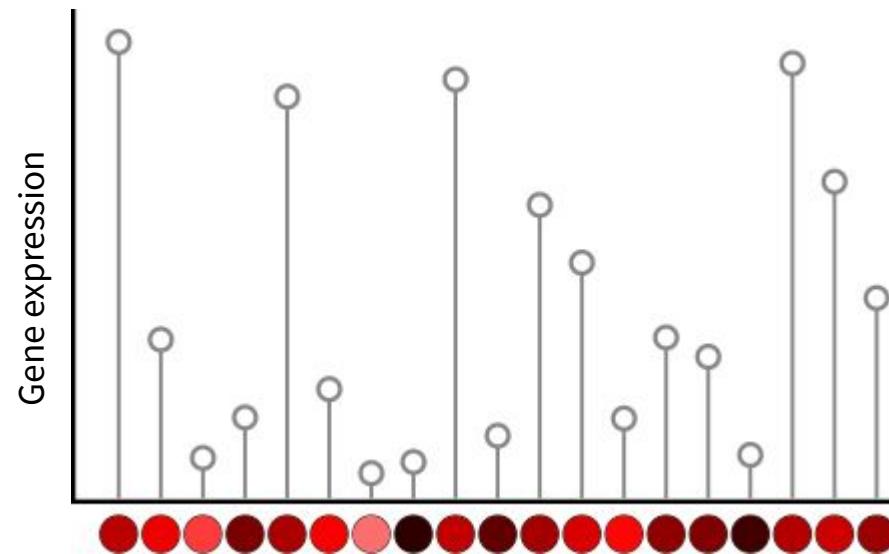


Ouija intuition

True ordering:



Marker gene x (Transient activation)



Random cell ordering:

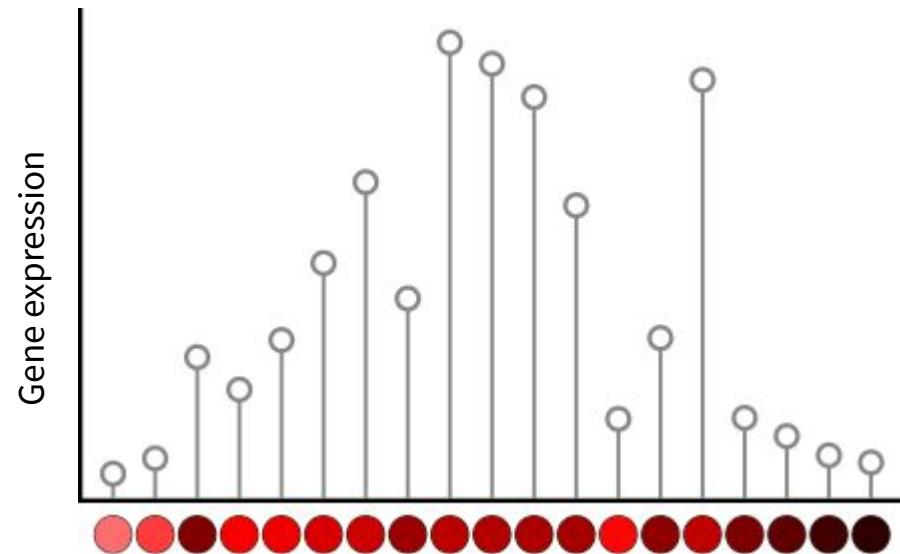
Goodness-of-fit: **low**

Ouija intuition

True ordering:



Marker gene x (Transient activation)



Optimize iteration: 100

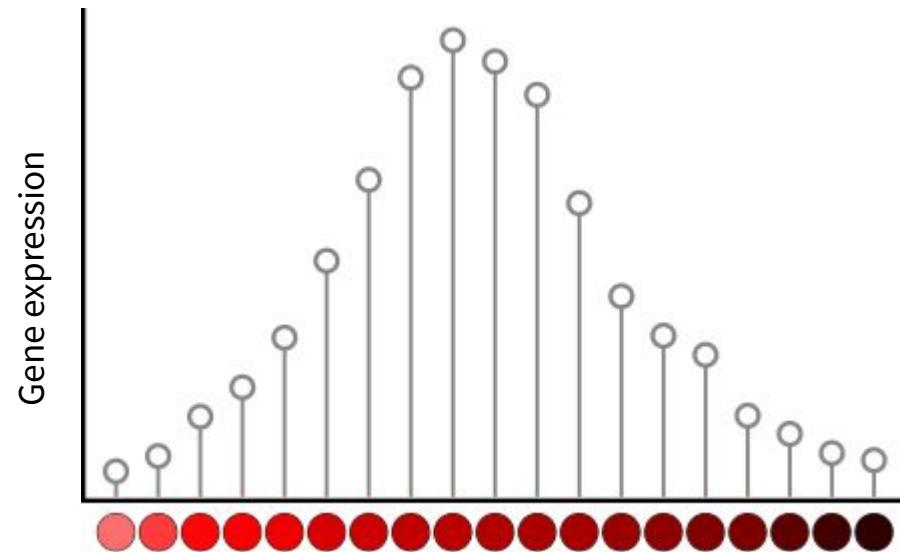
Goodness-of-fit: mid

Ouija intuition

True ordering:



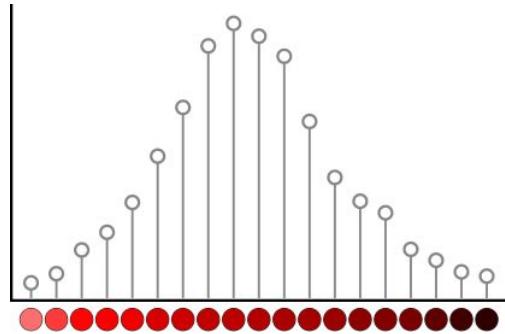
Marker gene x (Transient activation)



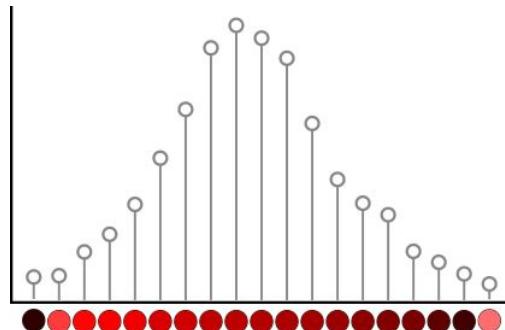
Optimize iteration: 500

Goodness-of-fit: high

Marker gene x

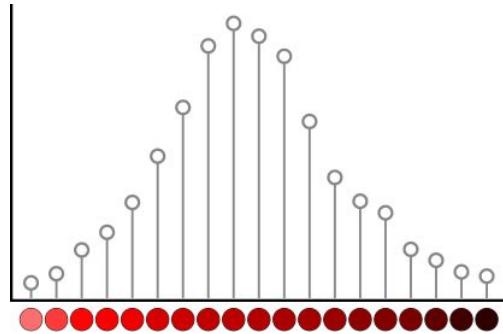


Goodness-of-fit: **high**



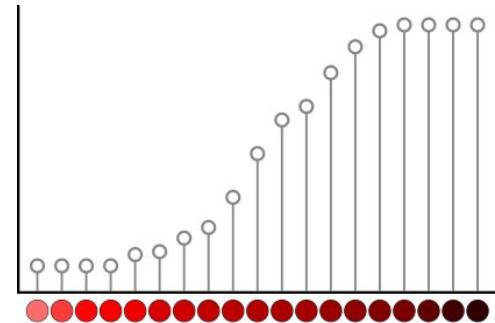
Goodness-of-fit: **high**

Marker gene x

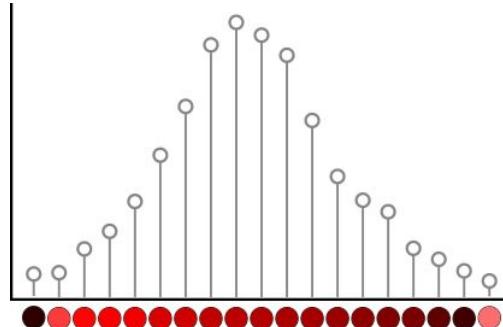


Goodness-of-fit: **high**

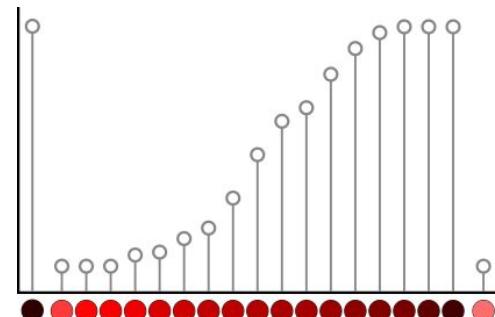
Marker gene y



Goodness-of-fit: **high**



Goodness-of-fit: **high**

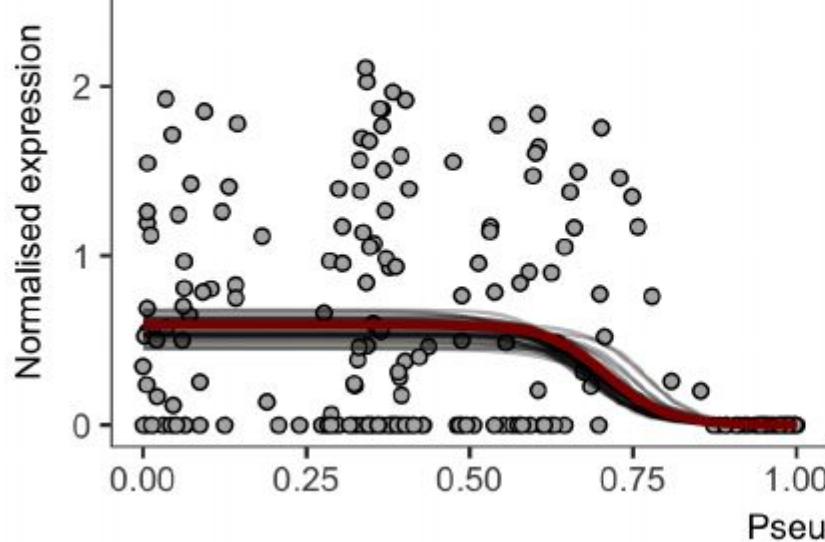


Goodness-of-fit: **mid**

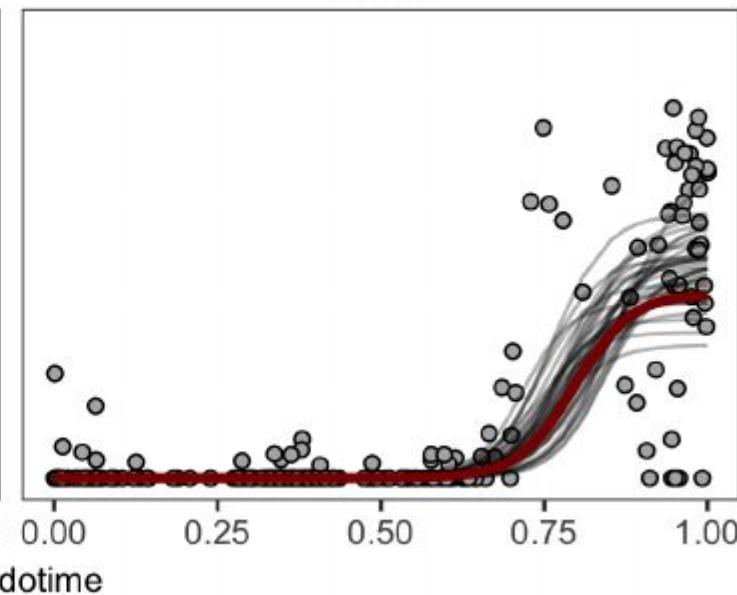
Ouija probabilistic modelling

C

ID1



MYOG

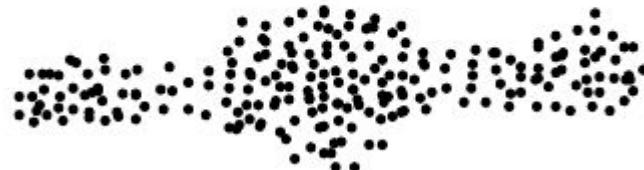


Fundamental limits on dynamic inference from single-cell snapshots

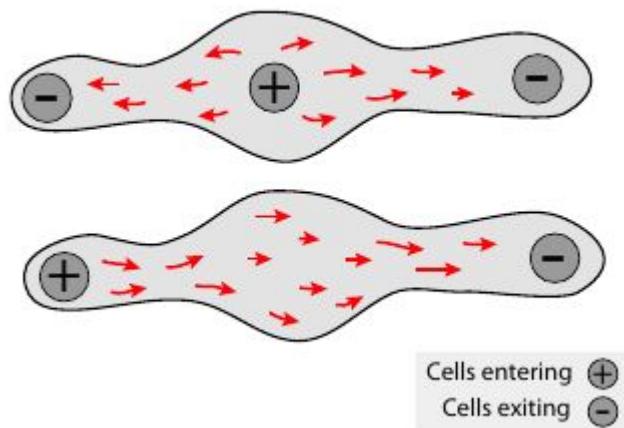
Caleb Weinreb^a, Samuel Wolock^a, Betsabeh K. Tusi^b, Merav Socolovsky^b, and Alon M. Klein^{a,1}

“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.”

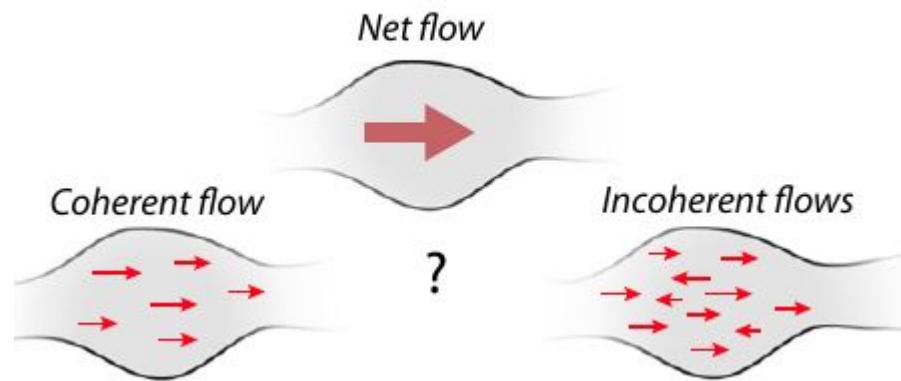
No unique solution



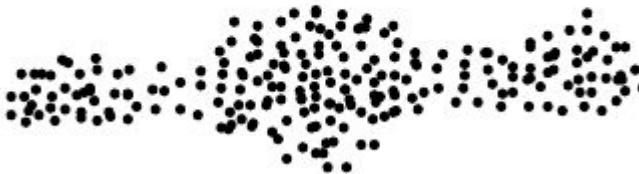
Entry and exit points direct the flow of cells



Net velocity may not equal actual velocity

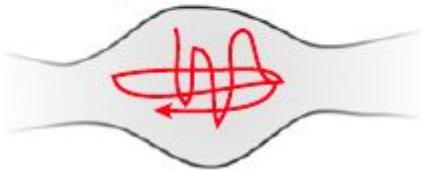


No unique solution

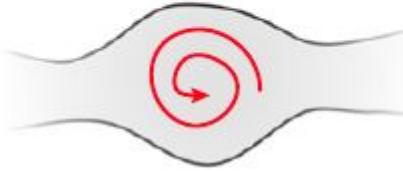


Rotations in state space do not alter cell density

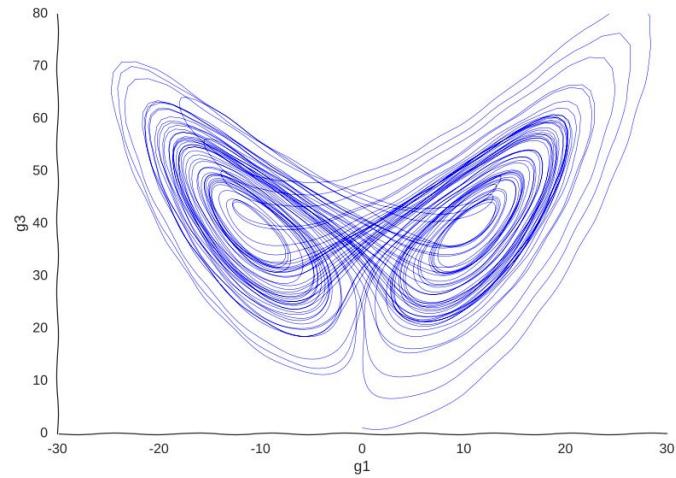
Simple fluctuations



Periodic oscillations

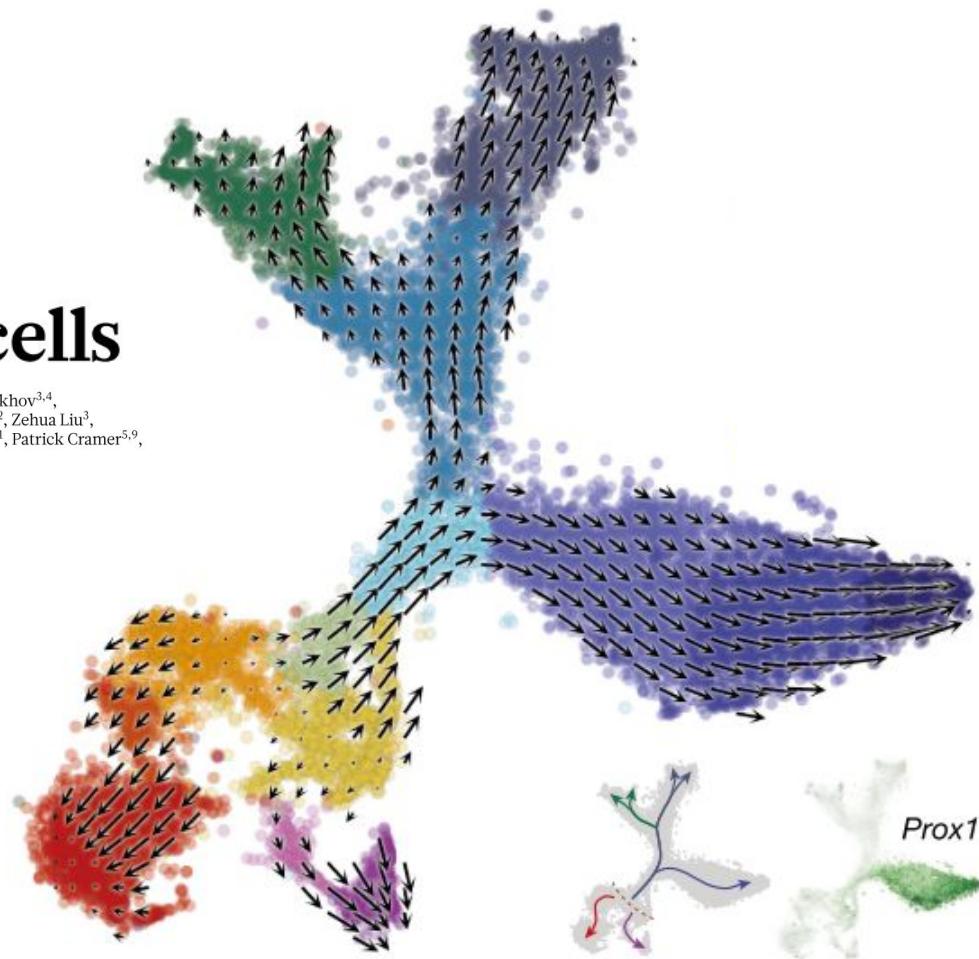


Two-state cyclical transition



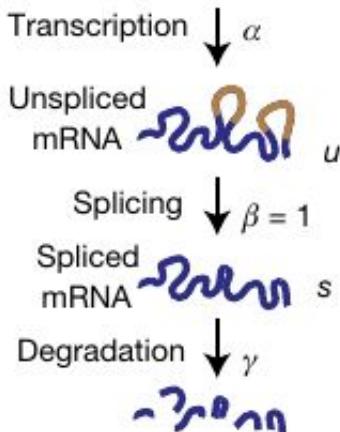
RNA velocity of single cells

Gioele La Manno^{1,2}, Ruslan Soldatov³, Amit Zeisel^{1,2}, Emelie Braun^{1,2}, Hannah Hochgerner^{1,2}, Viktor Petukhov^{3,4}, Katja Lidschreiber⁵, Maria E. Kastriti⁶, Peter Lönnberg^{1,2}, Alessandro Furlan¹, Jean Fan³, Lars E. Borm^{1,2}, Zehua Liu³, David van Bruggen¹, Jimin Guo³, Xiaoling He⁷, Roger Barker⁷, Erik Sundström⁸, Gonçalo Castelo-Branco¹, Patrick Cramer^{5,9}, Igor Adameyko⁶, Sten Linnarsson^{1,2*} & Peter V. Kharchenko^{3,10#}



RNA velocity of single cells

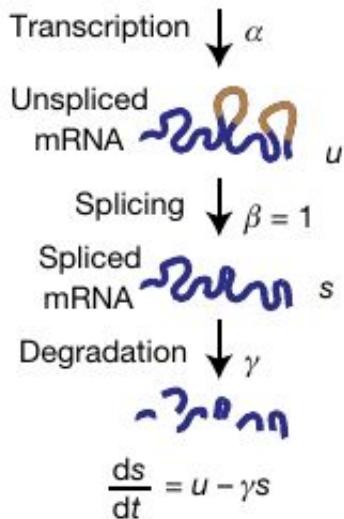
b



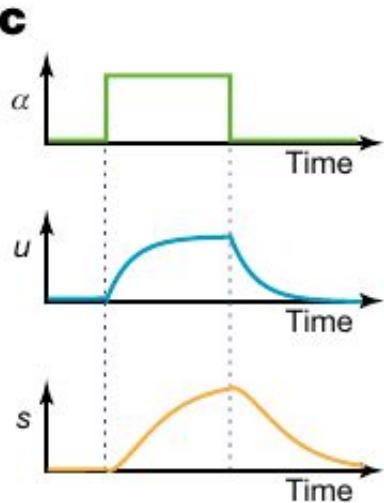
$$\frac{ds}{dt} = u - \gamma s$$

RNA velocity of single cells

b

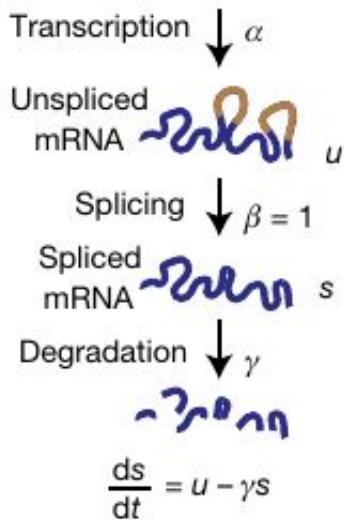


c

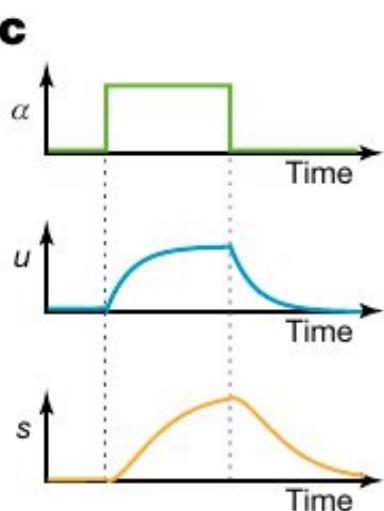


RNA velocity of single cells

b



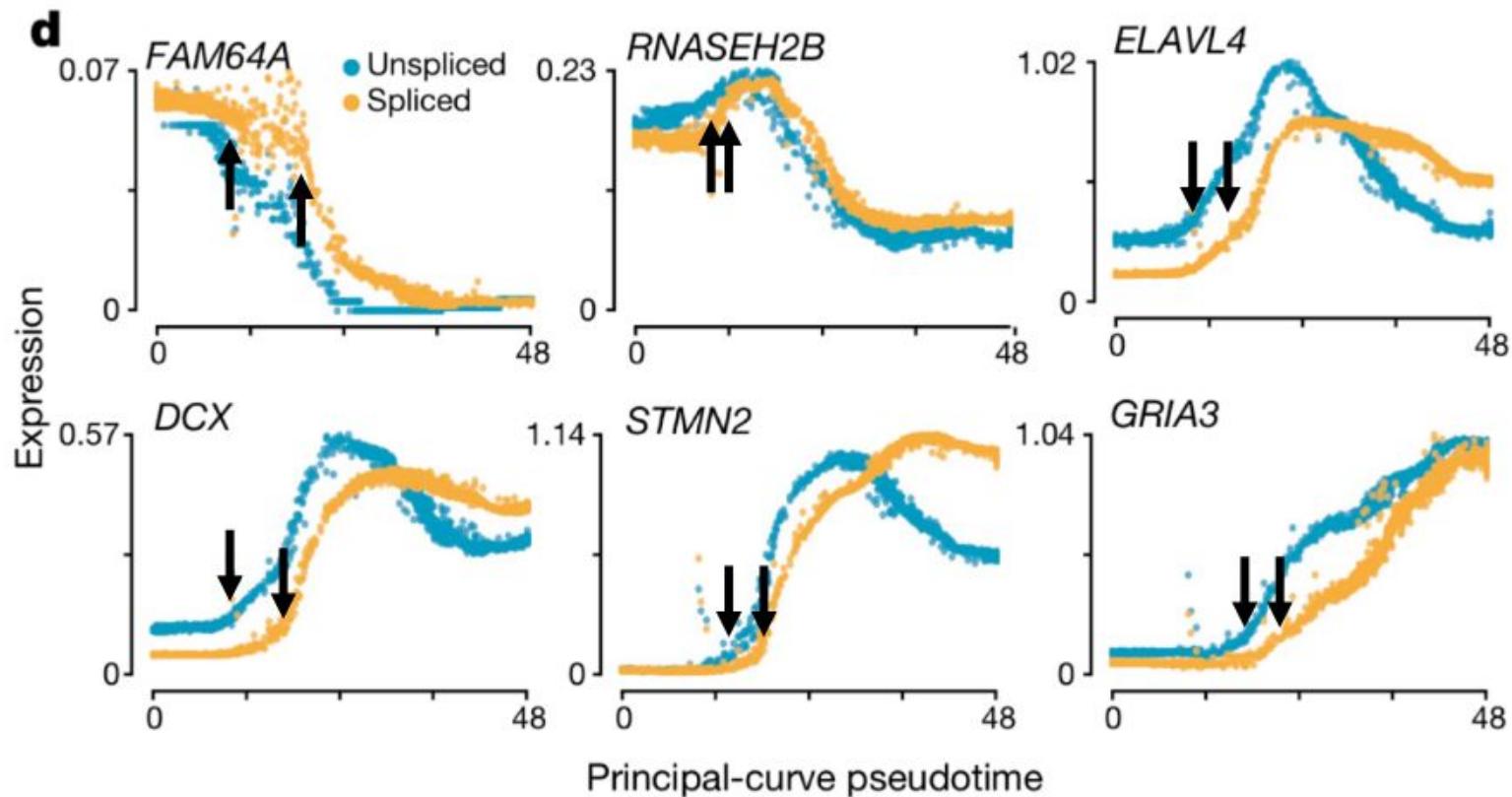
c



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

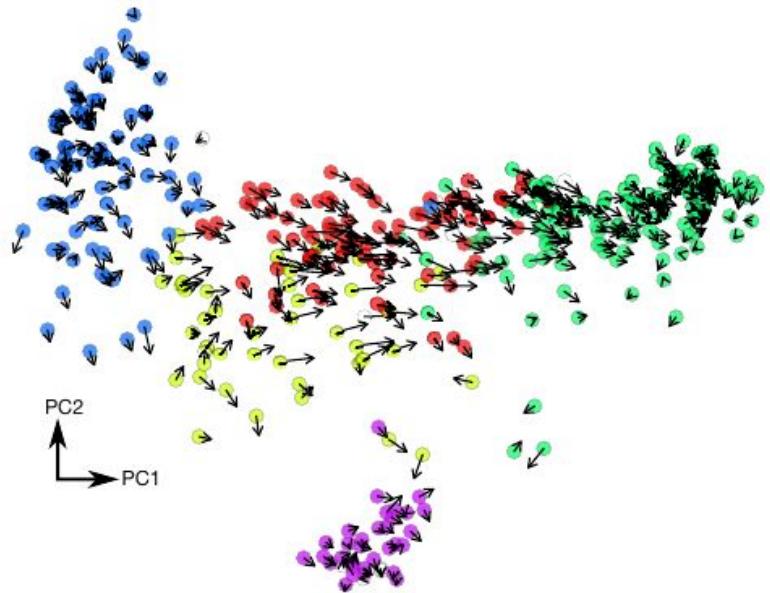
- $u/s > 1$: gene was recently activated
- $u/s < 1$: gene was recently deactivated

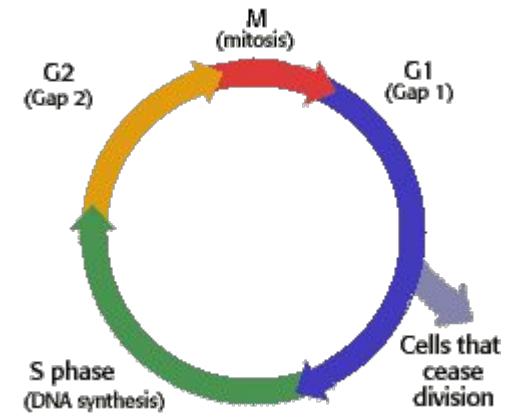
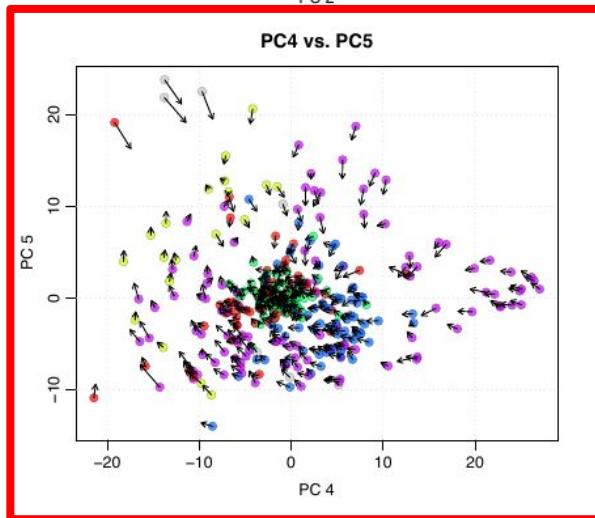
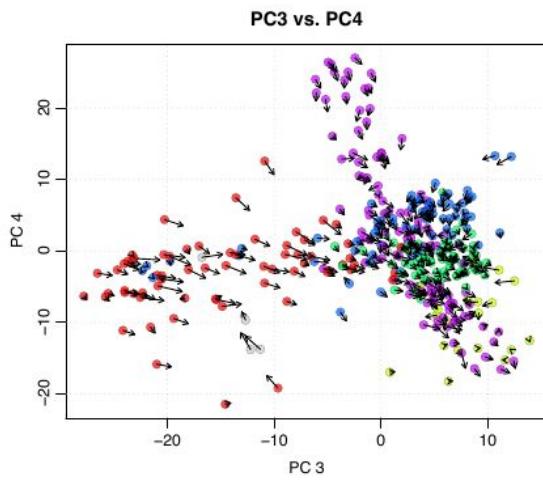
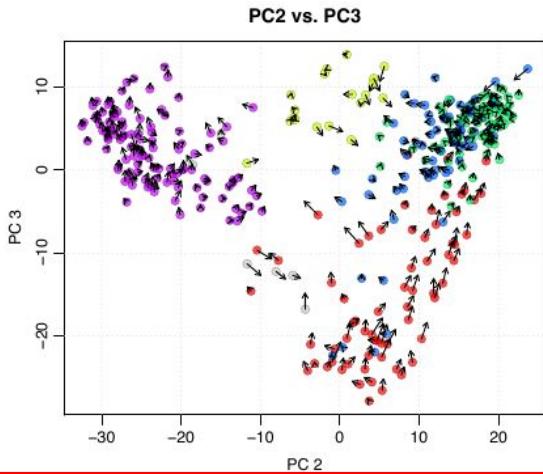
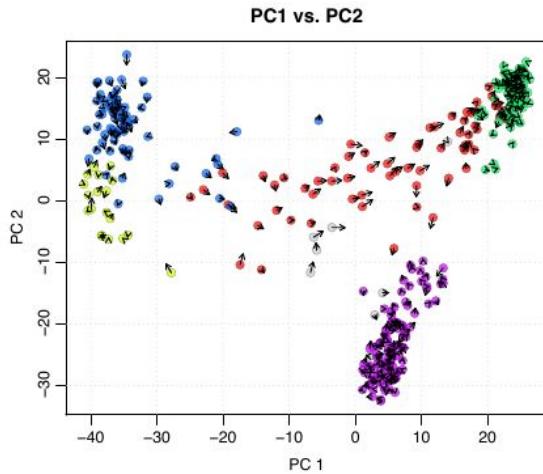
RNA velocity of single cells



RNA velocity of single cells

- The velocity of a gene is the predicted increase or decrease of its expression in the cell
- Used to extrapolate future state of a cell





Cyclic trajectories are also captured!

Which method should you use?

<http://guidelines.dynverse.org/>

Topology DEFAULT

Do you expect multiple disconnected trajectories in the data?

Yes I don't know No

Scalability COMPUTED

Number of cells
1000

Number of features (genes)
1000

Time limit
10s 1h

Memory limit
100MB 30GB

Prior information DEFAULT

Are you able to provide the following prior information?

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

Method selection DEFAULT

Benchmarking metrics DEFAULT

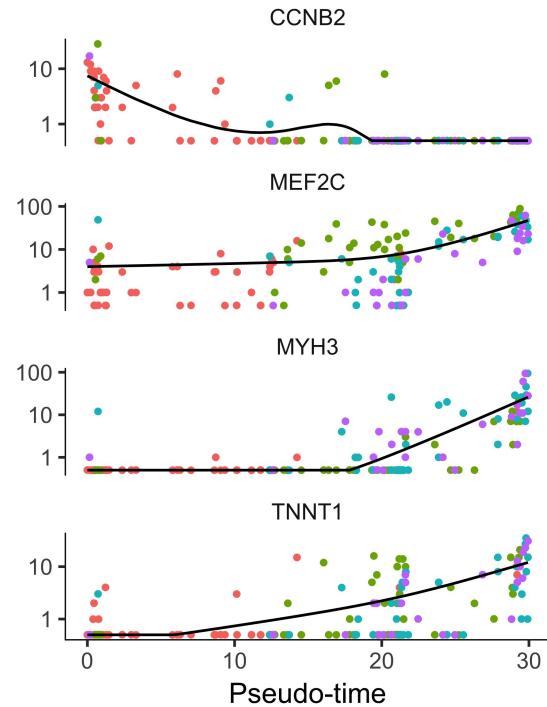
Show code ↗		Show/hide columns ⓘ		Options ⚙		Infer trajectories with dyno ↗			
Lenses	Default	Summary (Fig. 2)	Method	Scalability	Stability	Usability	Accuracy	Overall	Everything
Method						Accuracy	Scalability	Stability	
Name		Priors	Errors	Overall	Time	Memory	Unstable		
Slingshot	✓ ↗		100	8s	942MB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

Pseudotime analysis

Interpretation of gene behaviour

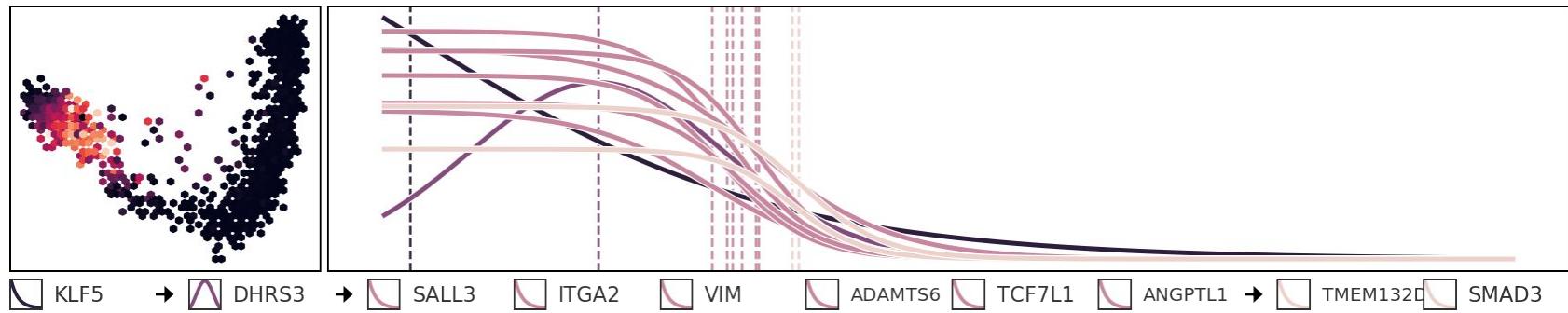
- Plot the gene expression as a function of pseudotime
- What do we see?

Pseudotime-gene expression pattern



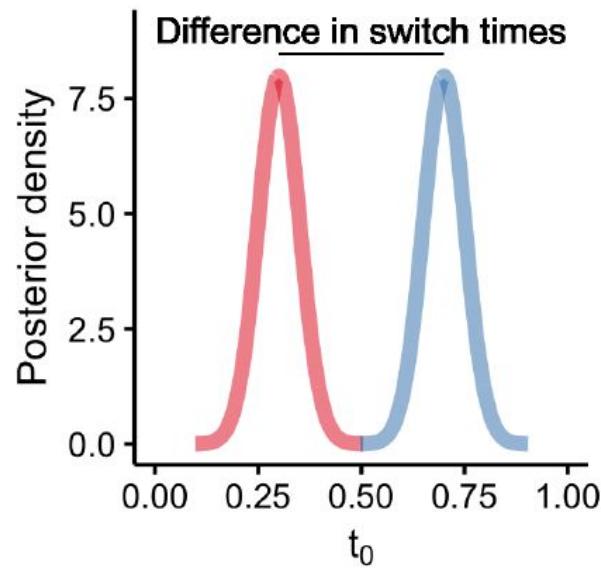
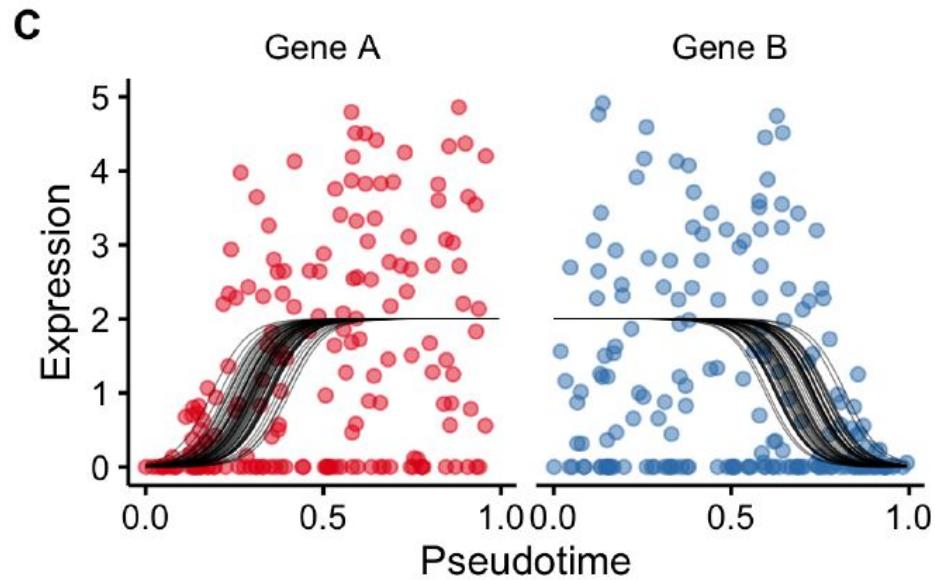
Interpretation of gene behaviour

Charrouet 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

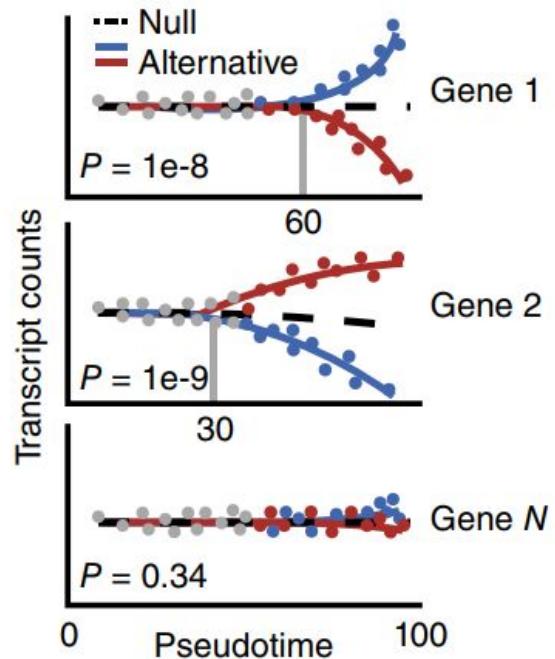
Differential activation testing



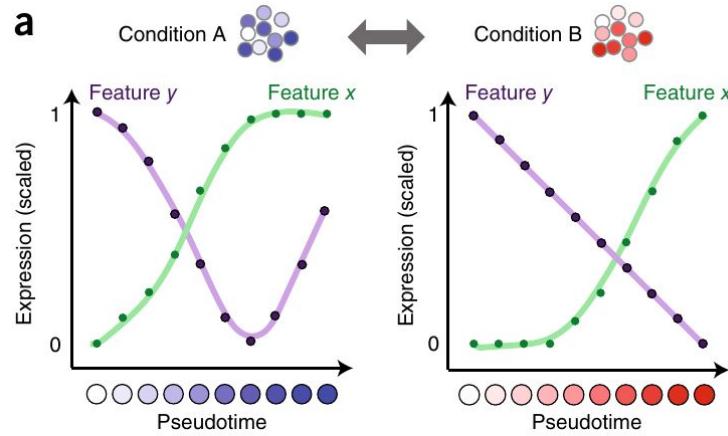
Branch analysis

Many analyses with branching trajectories:

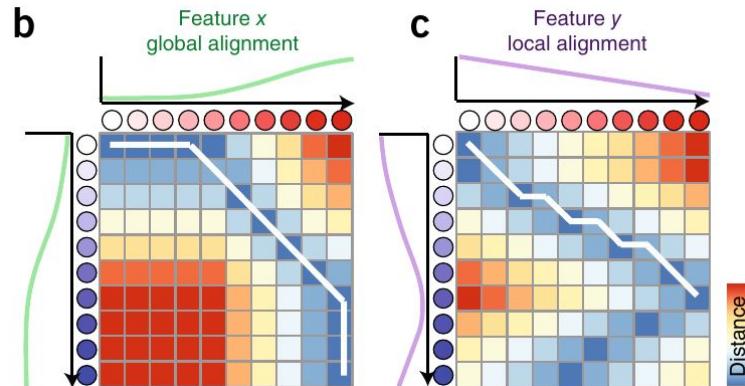
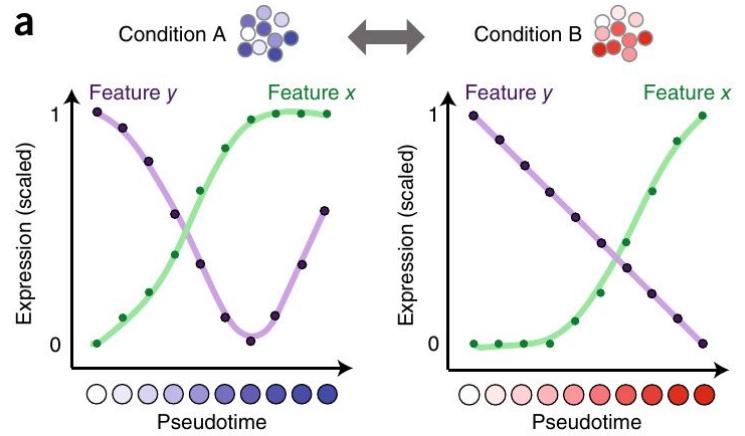
- Differential expression
 - Van de Berge et al. 2020
- Gene specific branching dynamics
 - Boukouvalas et al. 2018
- Branch-dependent expression
 - Monocle's BEAM



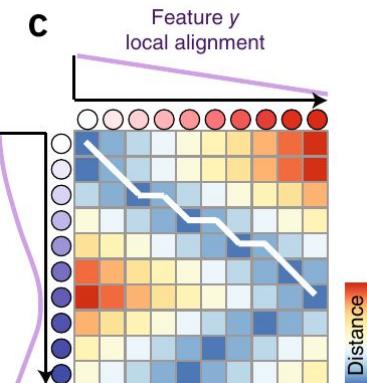
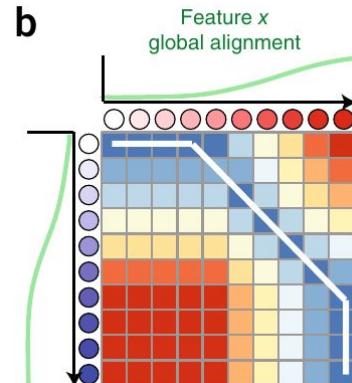
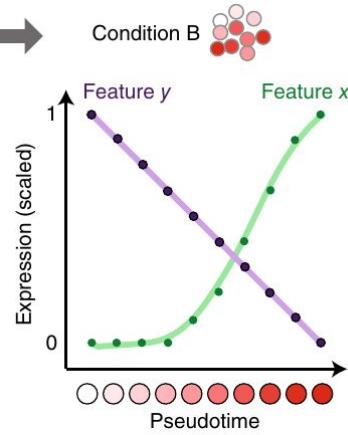
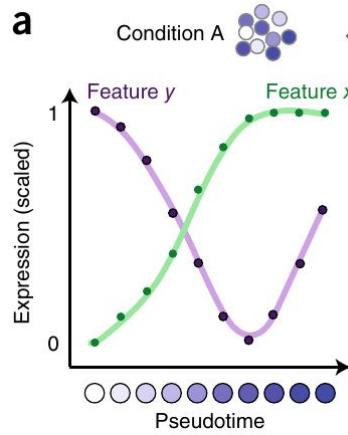
Comparing trajectories



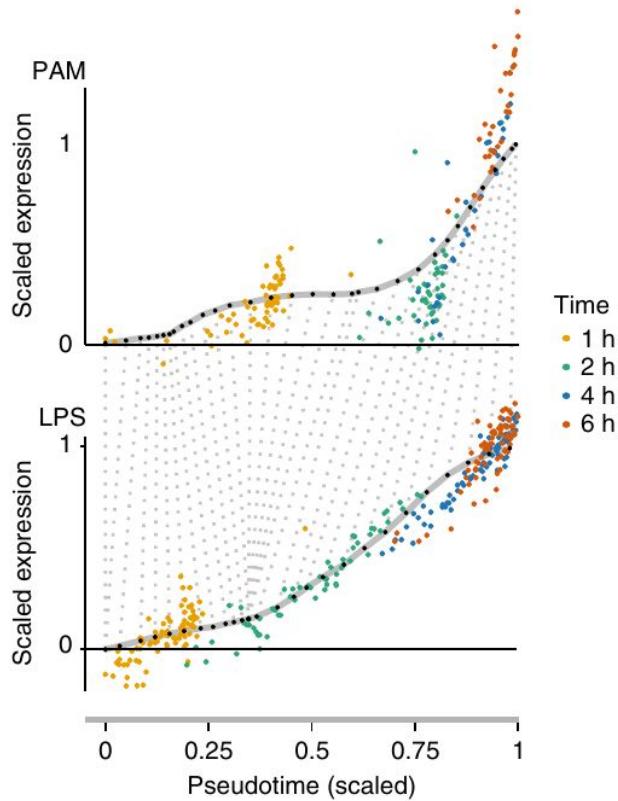
Comparing trajectories



Comparing trajectories



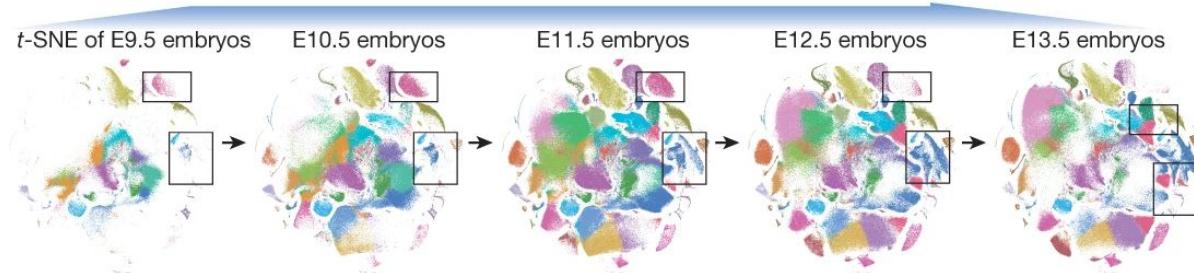
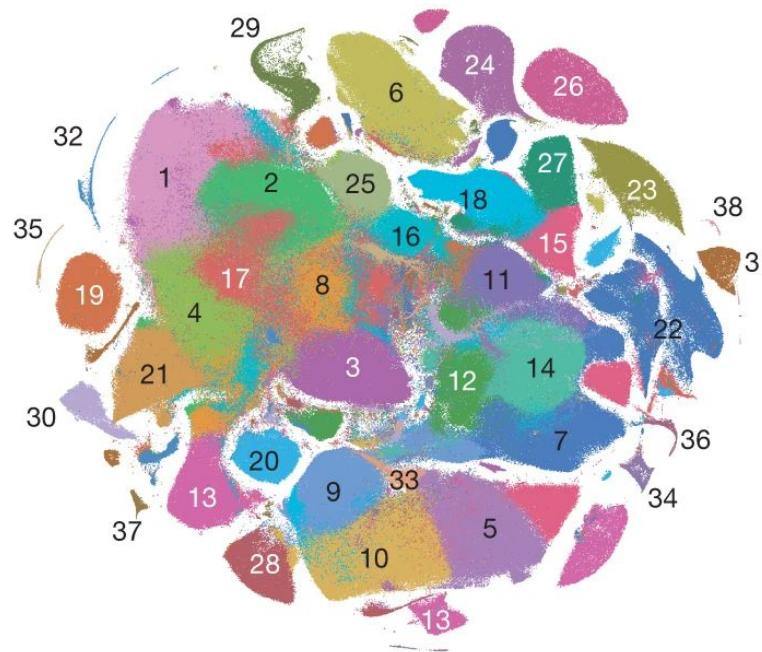
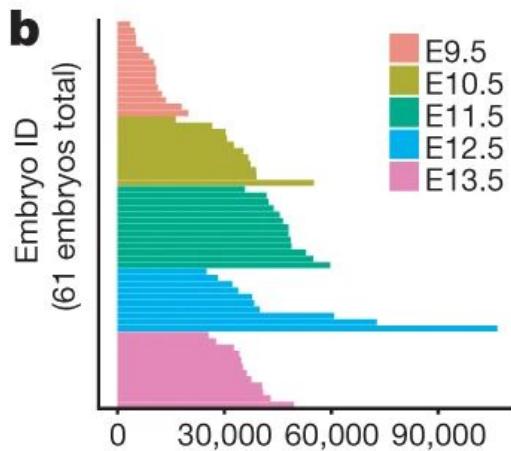
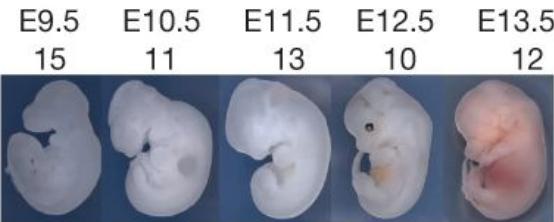
Dynamic time warping



Applications

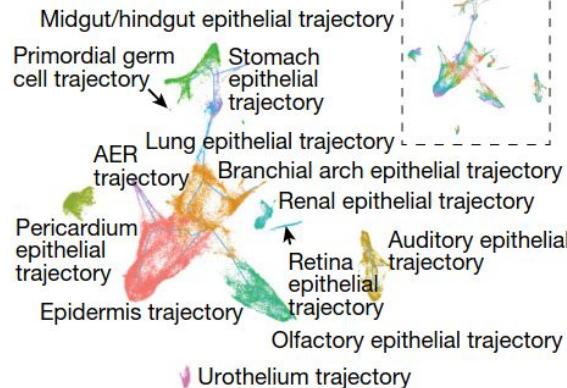
Mouse embryogenesis

Cao et al., 2019

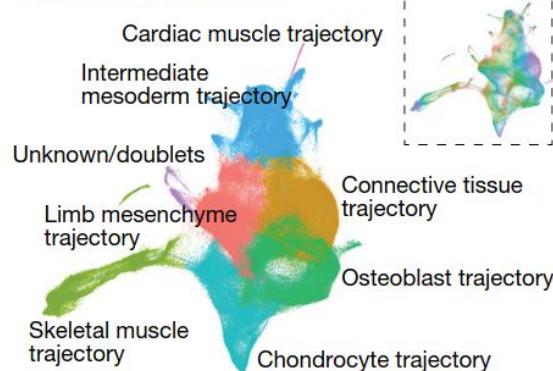


Mouse organogenesis

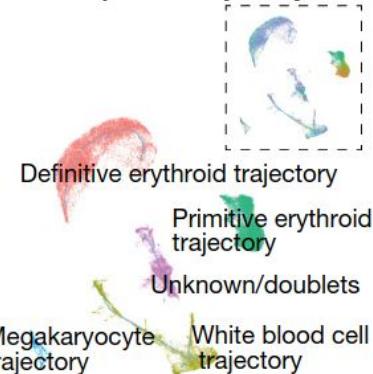
Epithelial trajectory



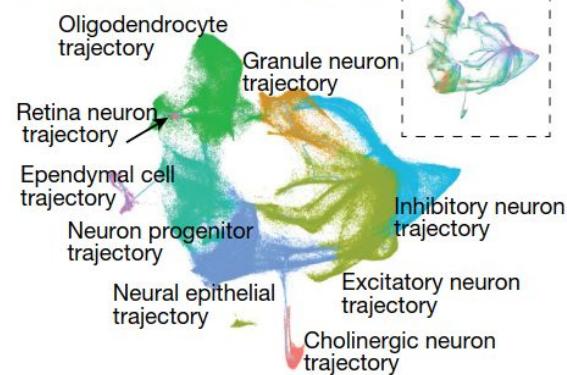
Mesenchymal trajectory



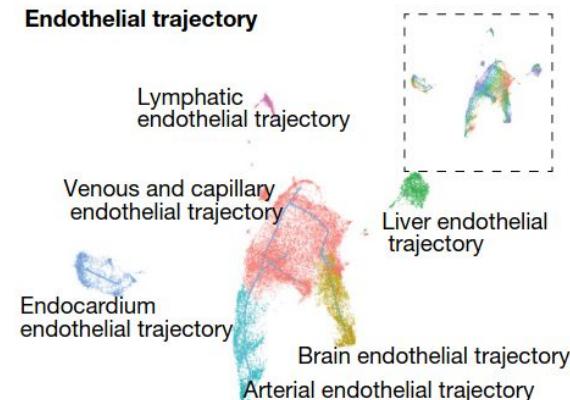
Haematopoiesis trajectory



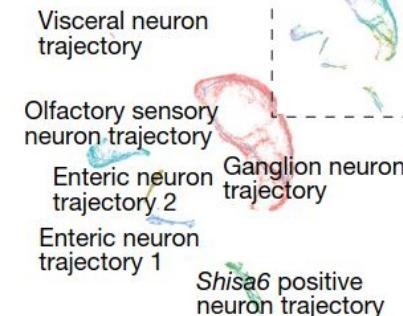
Neural tube/notochord trajectory



Endothelial trajectory



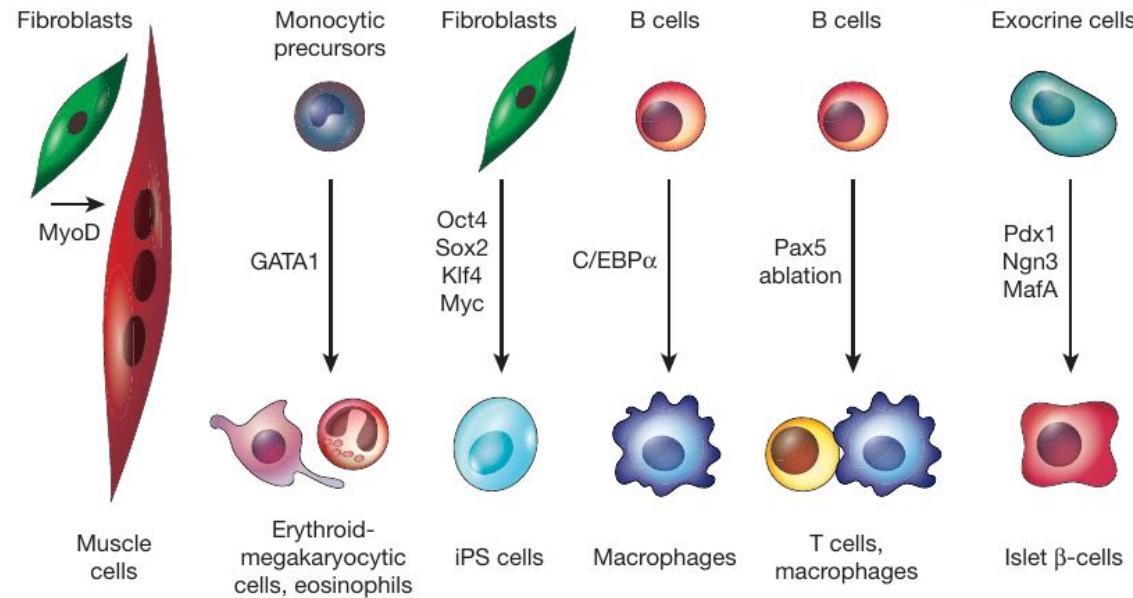
Neural crest (PNS neuron) trajectory 3



Induced transdifferentiation

Xia and Yanai, 2019

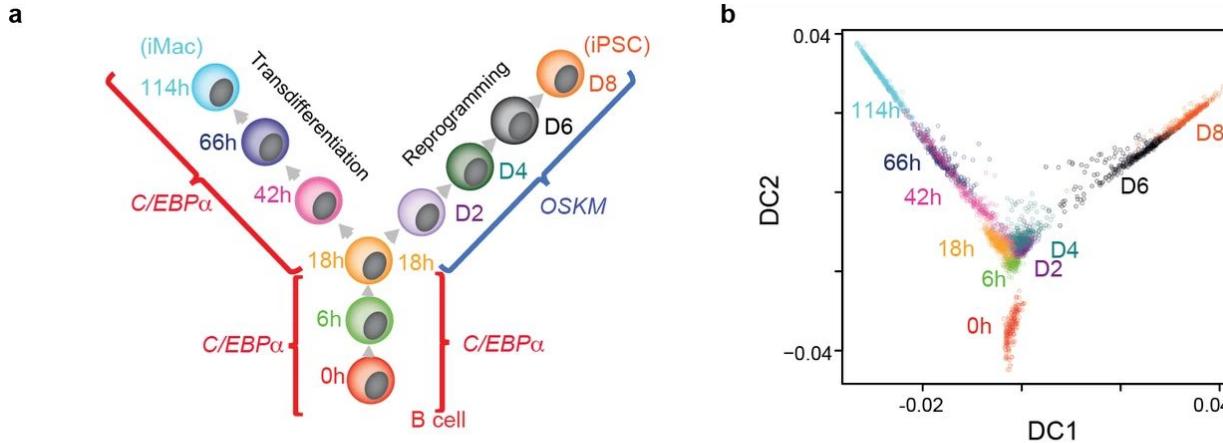
Transcription factors overexpression / silencing results in transdifferentiation:



Efficiency of transdifferentiation

Francesconi et al., 2019

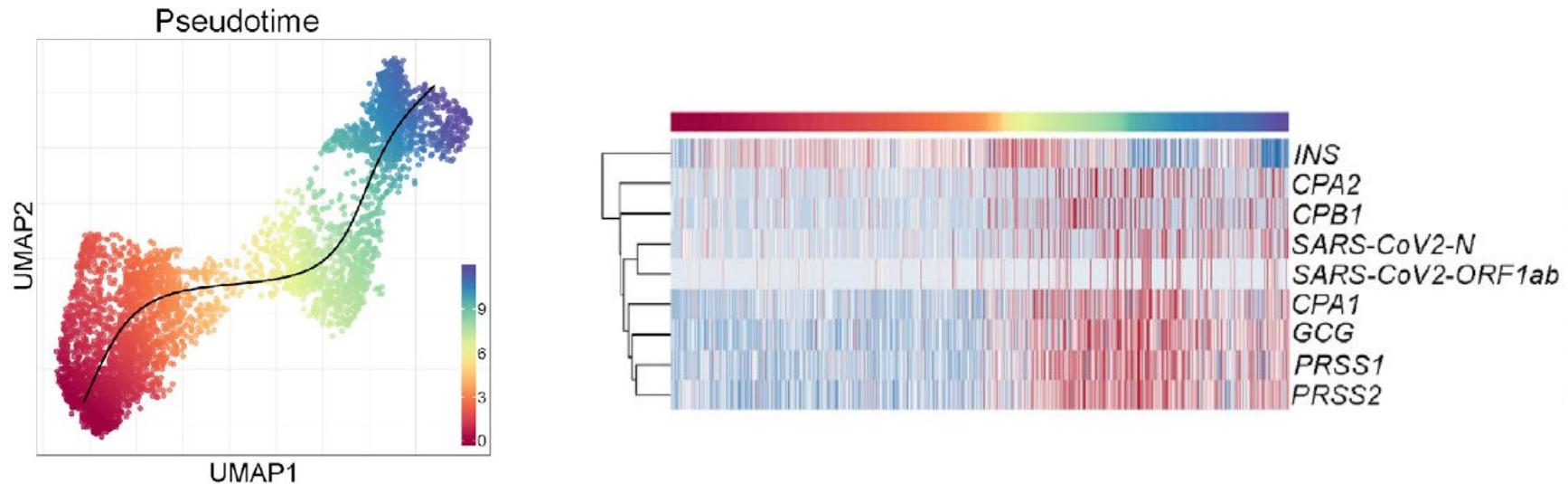
- Overexpression of C/EBPa (macrophages) and OSKM (iPSC) in B cells
- Homogenous final cell population, but variability in speed of differentiation
- Linked to Myc expression in initial state of the B cells



SARS-CoV induced beta-to-alpha cell trans-differentiation

Tang et al., 2021

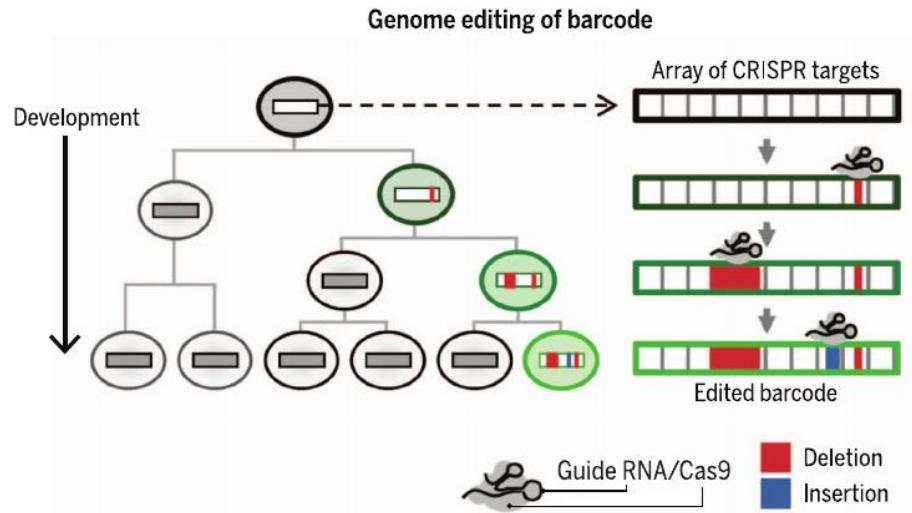
- COVID infected patients show a beta-to-alpha cell trans-differentiation in pancreatic islets



Lineage tracing

Pseudotime inference tracks cell differentiation “horizontally”

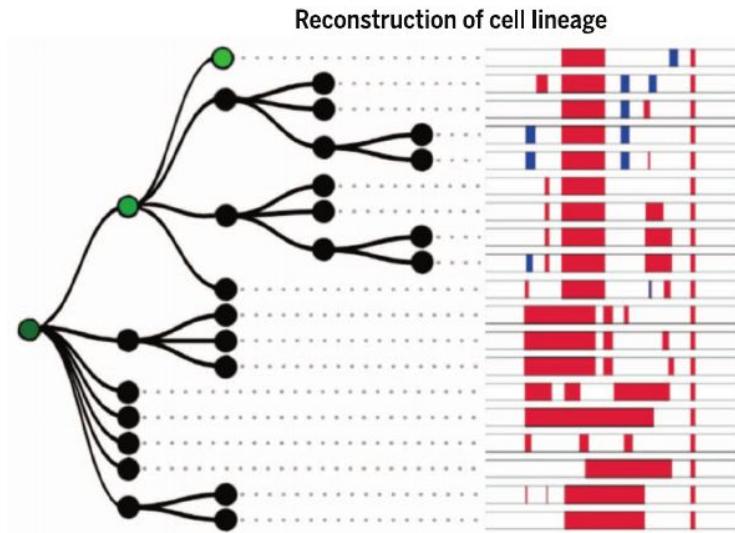
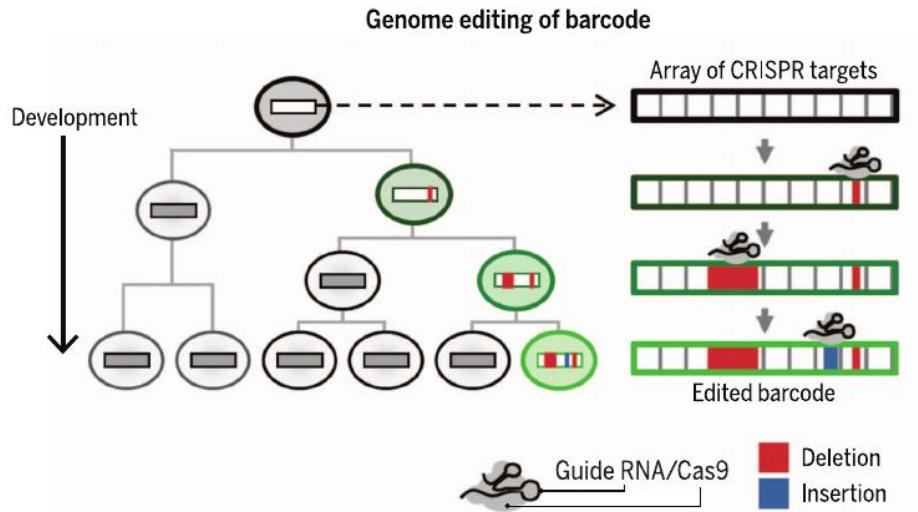
Lineage tracing tracks cell progeny:



Lineage tracing

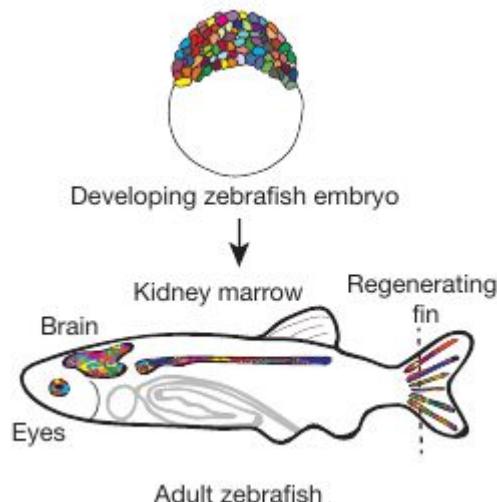
Pseudotime inference tracks cell differentiation “horizontally”

Lineage tracing tracks cell progeny:

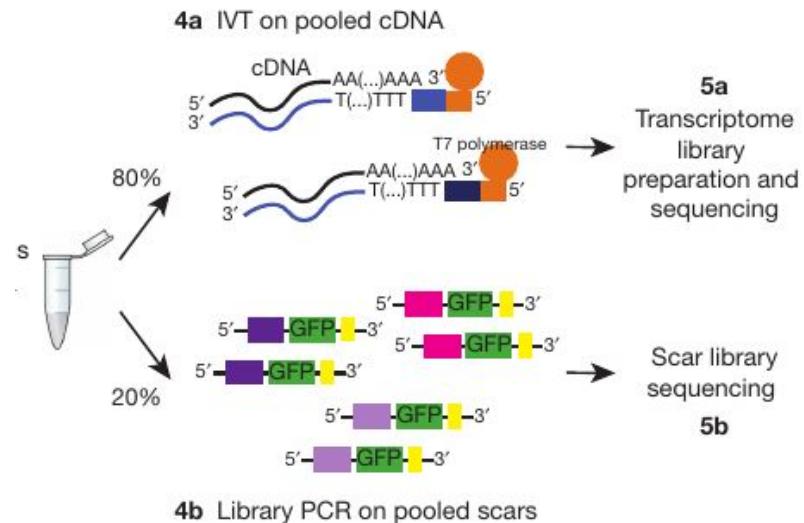


ScarTrace: Combining lineage tracing and scRNA-seq

Lineage tracing of zebrafish development. Span: 10 hpf

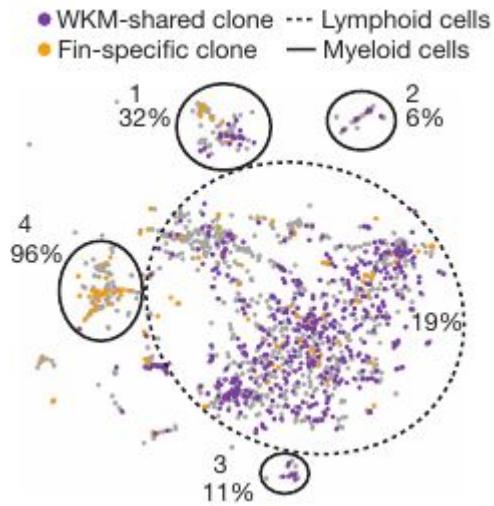


Cell material is split for scar library and transcriptome library preparation

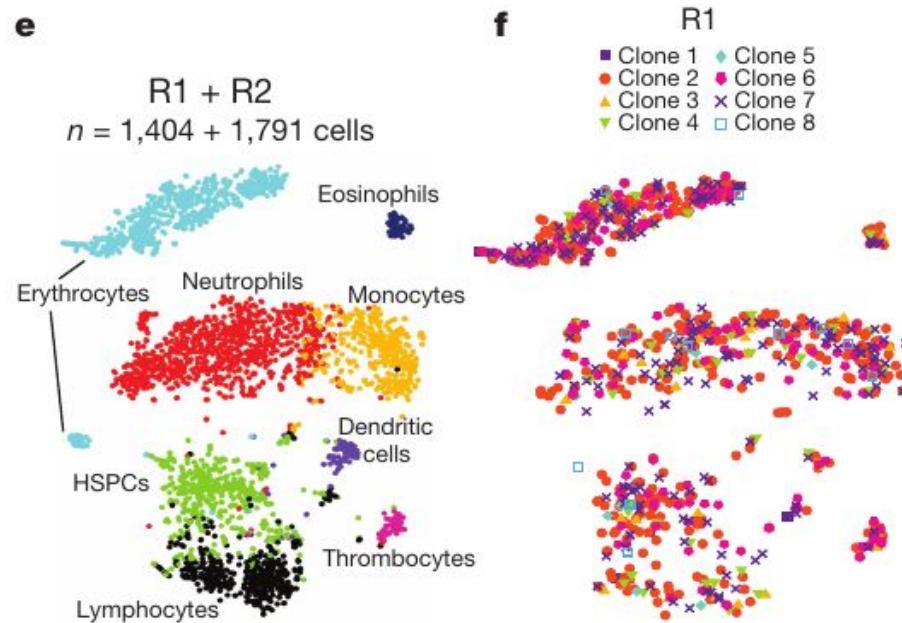


ScarTrace: Combining lineage tracing and scRNA-seq

Resident immune cells in fin



Clonal overlap in kidney marrow



Thank you!

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

m.charrout@tudelft.nl