Lecture 6

Inferring cell pseudotime from scRNAseq data

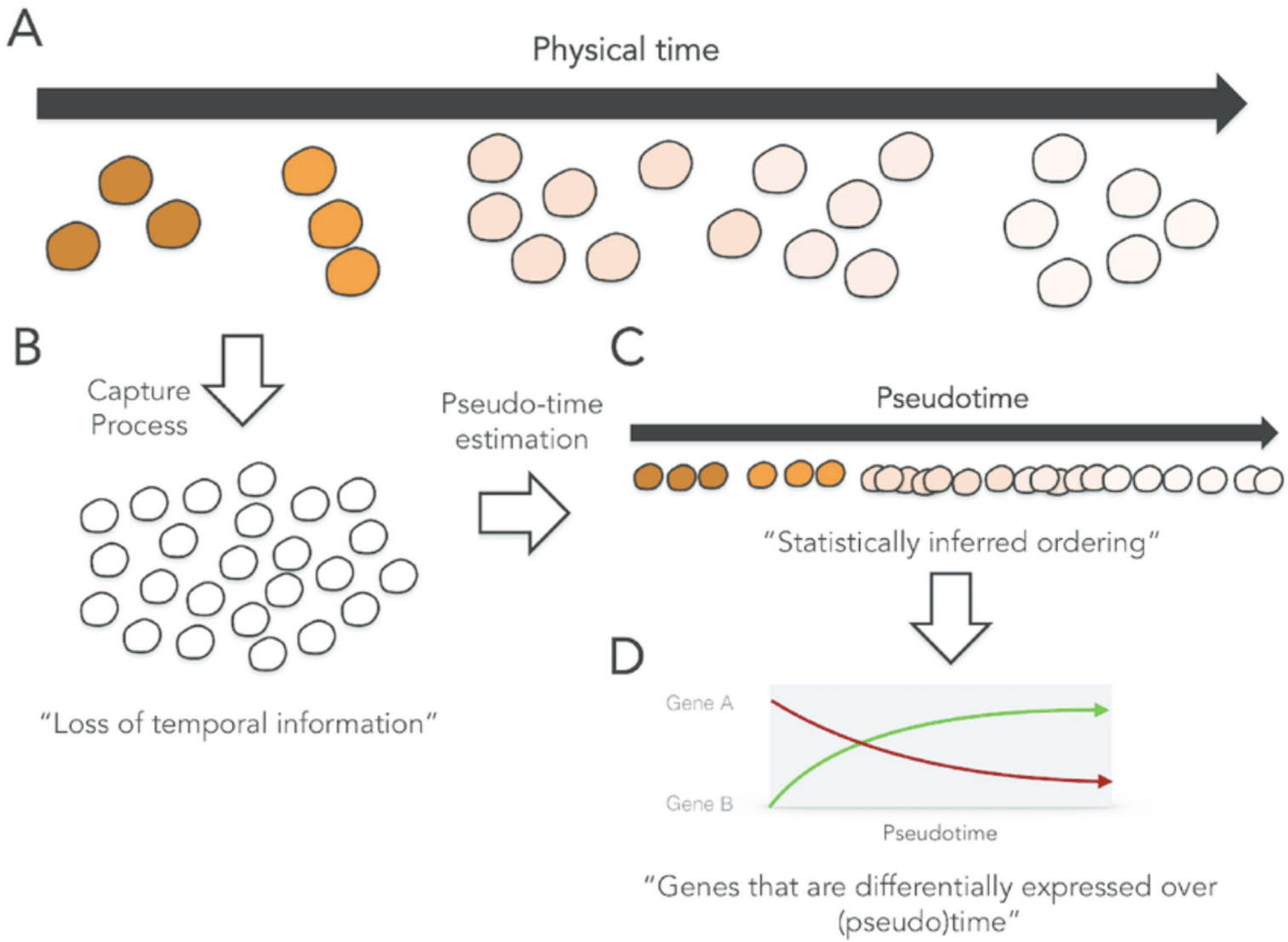
Physalia course 2023

—

Single-cell RNA-seq with R/Bioconductor

Instructors: Orr Ashenberg & Jacques Serizay

What is pseudotime?



What is pseudotime?

- Pseudotime is an abstract unit of progress: it's simply the distance between a cell and the start of the trajectory, measured along the shortest path

- The trajectory's total length is defined in terms of the total amount of transcriptional change that a cell undergoes as it moves from the starting state to the end state.

- **Pseudotime is a measure of how much progress an individual cell has made through a process (such as cell differentiation).**

http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories

What is pseudotime in single-cell RNA-seq?

- In single-cell expression studies of processes such as cell differentiation, captured cells might be widely distributed in terms of progress.

- That is, in a population of cells captured at exactly the same time, some cells might be far along, while others might not yet even have begun the process.

- **By ordering each cell according to its progress along a learned trajectory, pseudotime inference alleviates the problems that arise due to asynchrony.**

http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories

What is pseudotime in single-cell RNA-seq?

- In single-cell expression studies of processes such as cell differentiation, captured cells might be widely distributed in terms of progress.

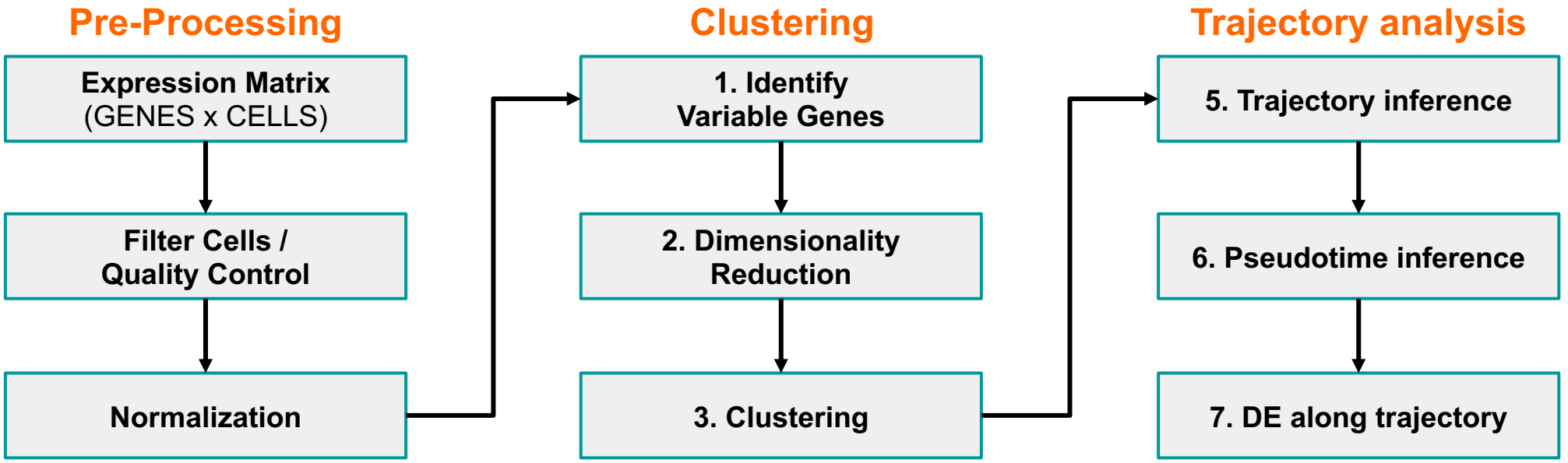
- That is, in a population of cells captured at exactly the same time, some cells might be far along, while others might not yet even have begun the process.

- **By ordering each cell according to its progress along a learned trajectory, pseudotime inference alleviates the problems that arise due to asynchrony.**

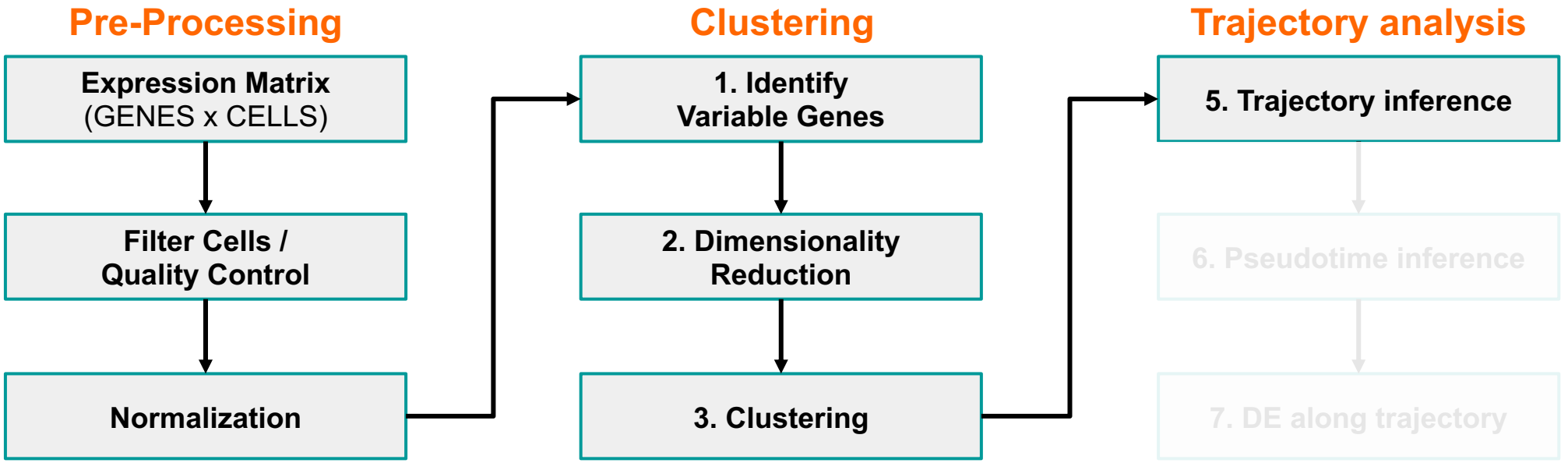
This is why pseudotime and trajectory inference are largely overlapping.

http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories

Analysis workflow



Analysis workflow

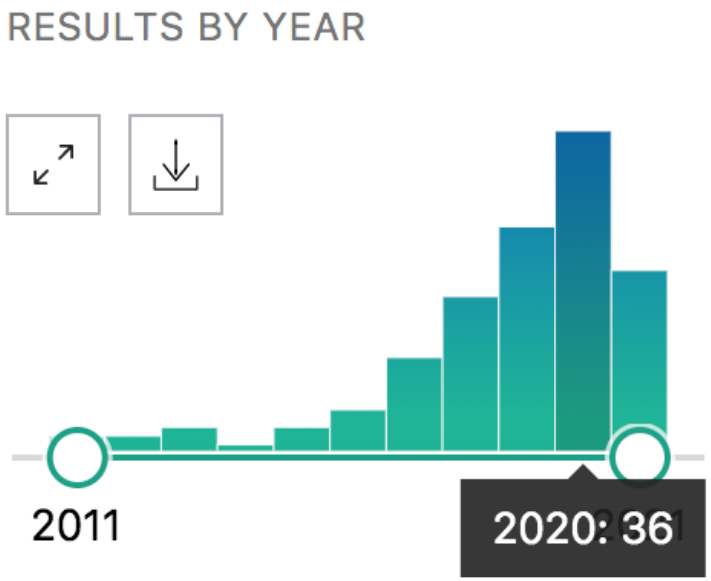


https://github.com/agitter/single-cell-pseudotime

Trajectory inference (TI) tools

A new algorithm comes out ~ every other week…

Pubmed results for “**trajectory inference method single-cell**”:



Trajectory inference (TI) tools

A new algorithm comes out ~ every other week…

Pseudocell Tracer-A **method** for **inferring** dynamic **trajectories** using scRNAseq and its application to B cells undergoing immunoglobulin class switch recombination.

Reiman D, Manakkat Vijay GK, Xu H, Sonin A, Chen D, Salomonis N, Singh H, Khan AA.

PLoS Comput Biol. 2021 May 3;17(5):e1008094. doi: 10.1371/journal.pcbi.1008094. eCollection 2021 May.

PMID: 33939691 **Free PMC article.**

Single cell RNA sequencing (scRNAseq) can be used to **infer** a temporal ordering of cellular states. Current **methods** for the **inference** of cellular **trajectories** rely on unbiased dimensionality reduction techniques. ...

Slingshot: **cell** lineage and pseudotime inference for **single-cell** transcriptomics.

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S.

BMC Genomics. 2018 Jun 19;19(1):477. doi: 10.1186/s12864-018-4772-0.

PMID: 29914354 **Free PMC article.**

RESULTS: We introduce **Slingshot**, a novel method for inferring **cell** lineages and pseudotimes from **single-cell** gene expression data. ...CONCLUSIONS: **Slingshot** is a uniquely robust and flexible tool which combines the highly stable techniques neces ...

Trajectory inference (TI) tools

A new algorithm comes out ~ every other week…

Pseudocell Tracer-A **method** for **inferring** dynamic **trajectories** using scRNAseq and its application to B cells undergoing immunoglobulin class switch recombination.

Reiman D, Manakkat Vijay GK, Xu H, Sonin A, Chen D, Salomonis N, Singh H, Khan AA.

PLoS Comput Biol. 2021 May 3;17(5):e1008094. doi: 10.1371/journal.pcbi.1008094. eCollection 2021 May.

PMID: 33939691 **Free PMC article.**

Single cell RNA sequencing (scRNAseq) can be used to **infer** a temporal ordering of cellular states. Current **methods** for the **inference** of cellular **trajectories** rely on unbiased dimensionality reduction techniques. ...

Slingshot: **cell** lineage and pseudotime inference for **single-cell** transcriptomics.

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S.

BMC Genomics. 2018 Jun 19;19(1):477. doi: 10.1186/s12864-018-4772-0.

PMID: 29914354 **Free PMC article.**

RESULTS: We introduce **Slingshot**, a novel method for inferring **cell** lineages and pseudotimes from **single-cell** gene expression data. ...CONCLUSIONS: **Slingshot** is a uniquely robust and flexible tool which combines the highly stable techniques neces ...

Trajectory inference (TI) tools

Pseudocell Tracer-A **method** for **inferring** dynamic **trajectories** using scRNAseq and its application to B cells undergoing immunoglobulin class switch recombination.

Reiman D, Manakkat Vijay GK, Xu H, Sonin A, Chen D, Salomonis N, Singh H, Khan AA.

PLoS Comput Biol. 2021 May 3;17(5):e1008094. doi: 10.1371/journal.pcbi.1008094. eCollection 2021 May.

PMID: 33939691 **Free PMC article.**

Single cell RNA sequencing (scRNAseq) can be used to **infer** a temporal ordering of cellular states. Current **methods** for the **inference** of cellular **trajectories** rely on unbiased dimensionality reduction techniques. ...

Slingshot: **cell** lineage and pseudotime inference for **single-cell** transcriptomics.

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S.

BMC Genomics. 2018 Jun 19;19(1):477. doi: 10.1186/s12864-018-4772-0.

PMID: 29914354 **Free PMC article.**

RESULTS: We introduce **Slingshot**, a novel method for inferring **cell** lineages and pseudotimes from **single-cell** gene expression data. ...CONCLUSIONS: **Slingshot** is a uniquely robust and flexible tool which combines the highly stable techniques neces ...

Stick to **standards**!!

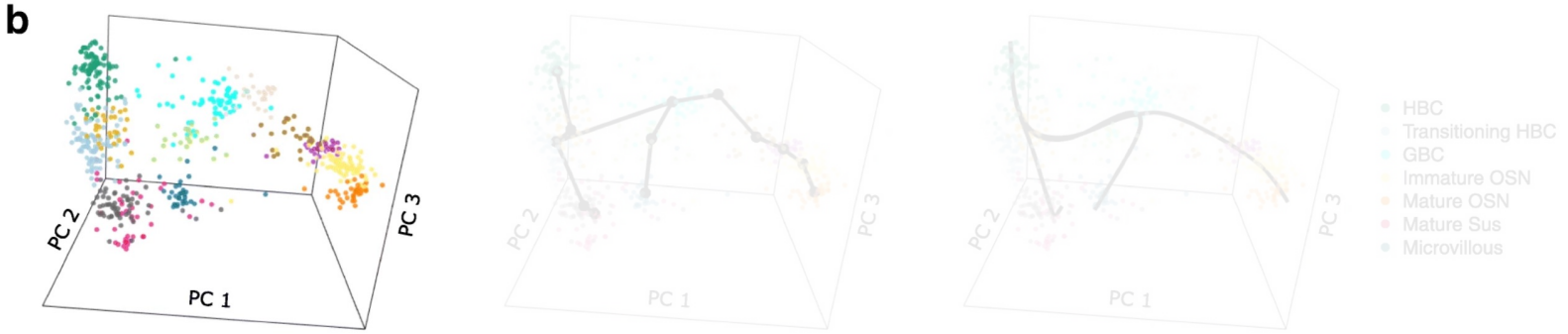
QCed, troubleshooted, optimized and **generic**

Slingshot

Slingshot is one of the most widely used and robust approaches to infer trajectory in relatively simple datasets.

<u>It works by:</u>

0. Get clustered data in a low-dimensional space



Street et al., 2018

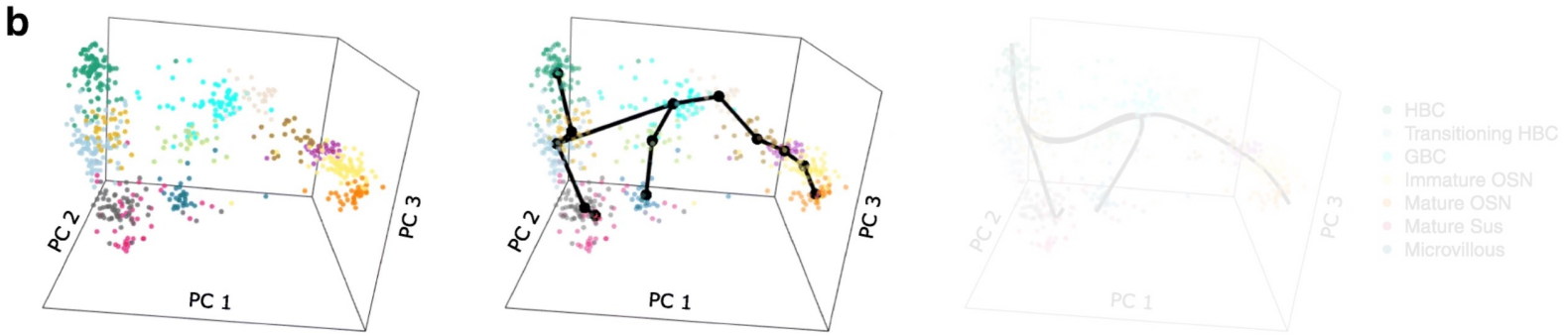
Slingshot

Slingshot is one of the most widely used and robust approaches to infer trajectory in relatively simple datasets.

It works by:

0. Get clustered data in a low-dimensional space

1. Building a **minimum spanning tree** on the clusters



Street et al., 2018

Slingshot

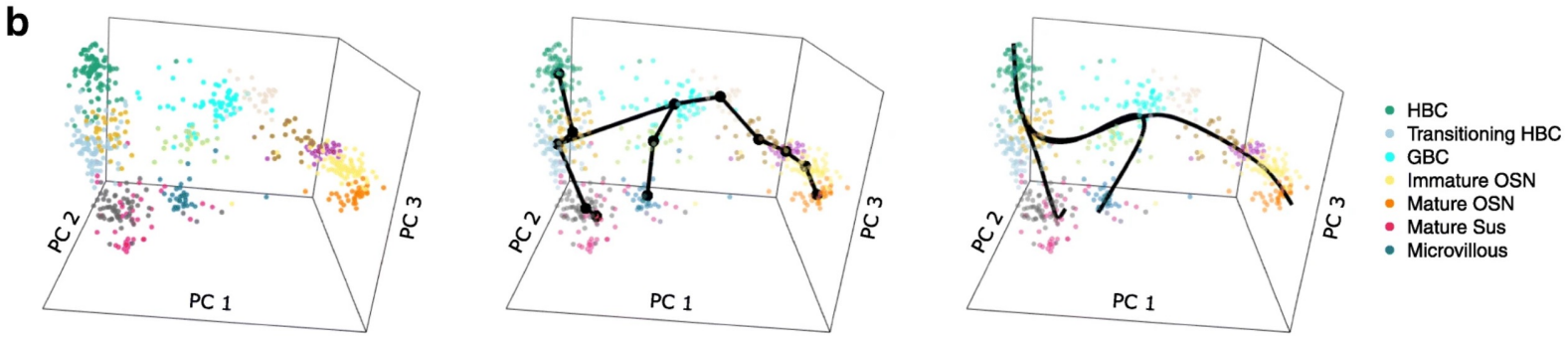
Slingshot is one of the most widely used and robust approaches to infer trajectory in relatively simple datasets.

It works by:

0. Get clustered data in a low-dimensional space

1. Building a **minimum spanning tree** on the clusters

2. Fit **principal curves** through the MST

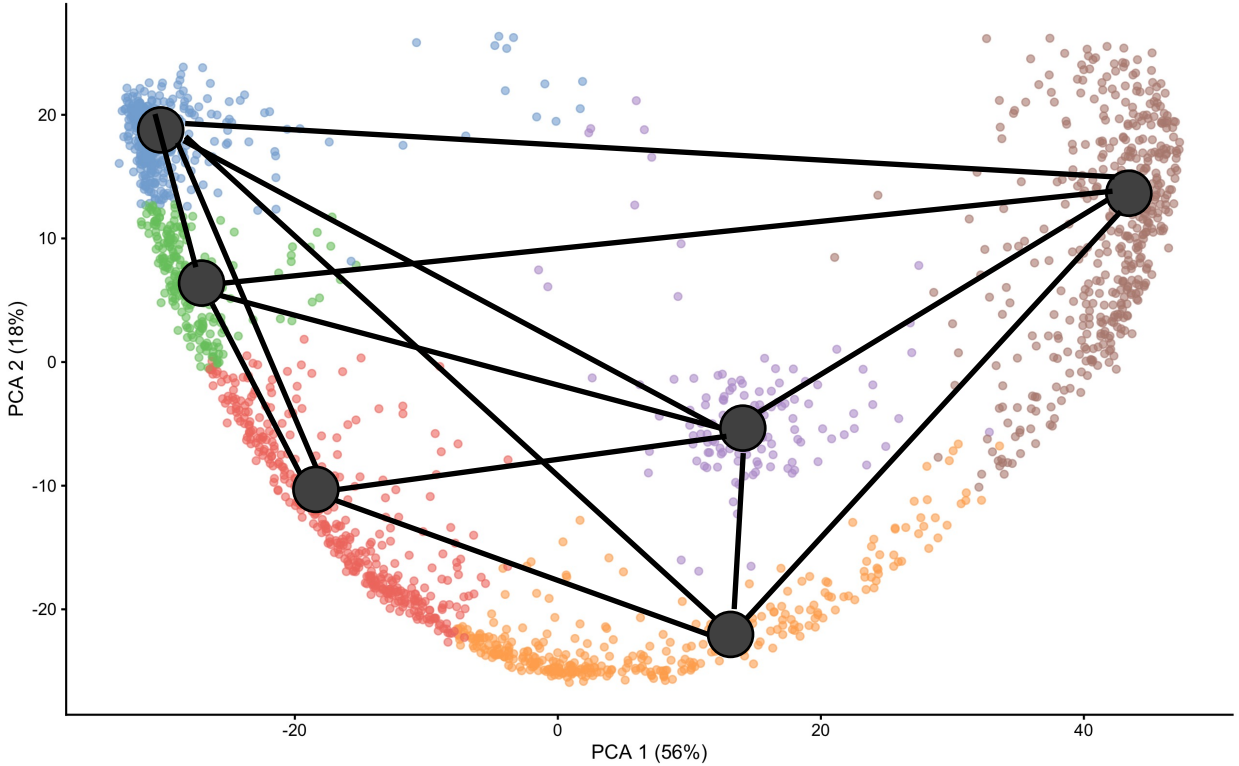


Street et al., 2018

Minimum spanning tree?!?

Or minimum <u>weight</u> spanning tree

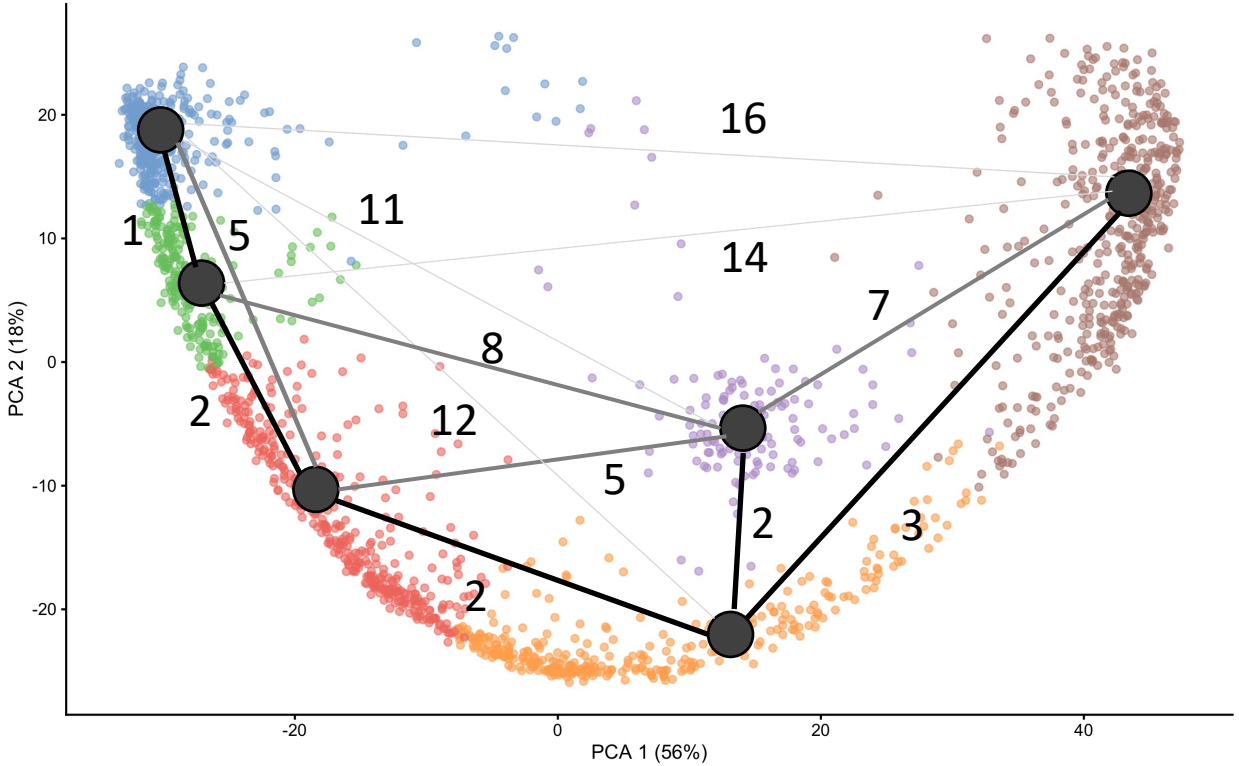
1. Build a graph with edges between each pair of clusters



Minimum spanning tree?!?

Or minimum <u>weight</u> spanning tree

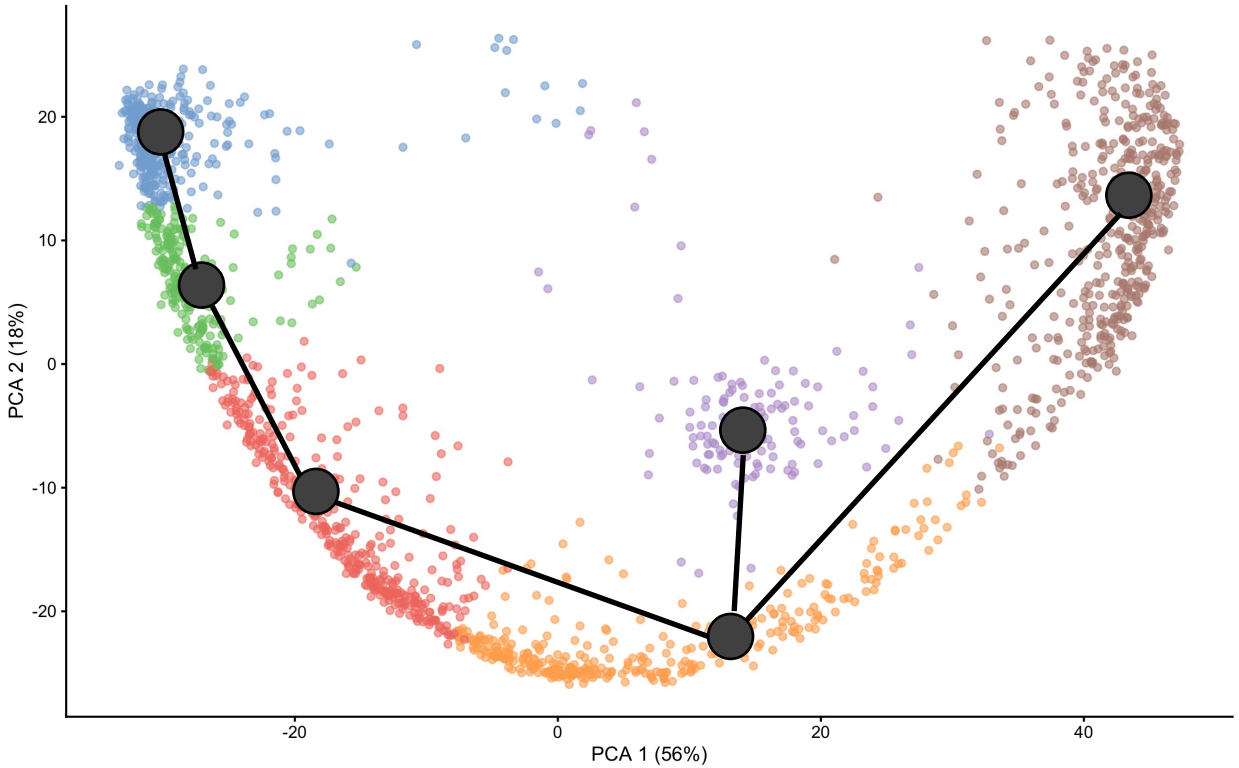
1. Build a graph with edges between each pair of clusters
2. Add weights to each edge according to proximity of the two clusters



Minimum spanning tree?!?

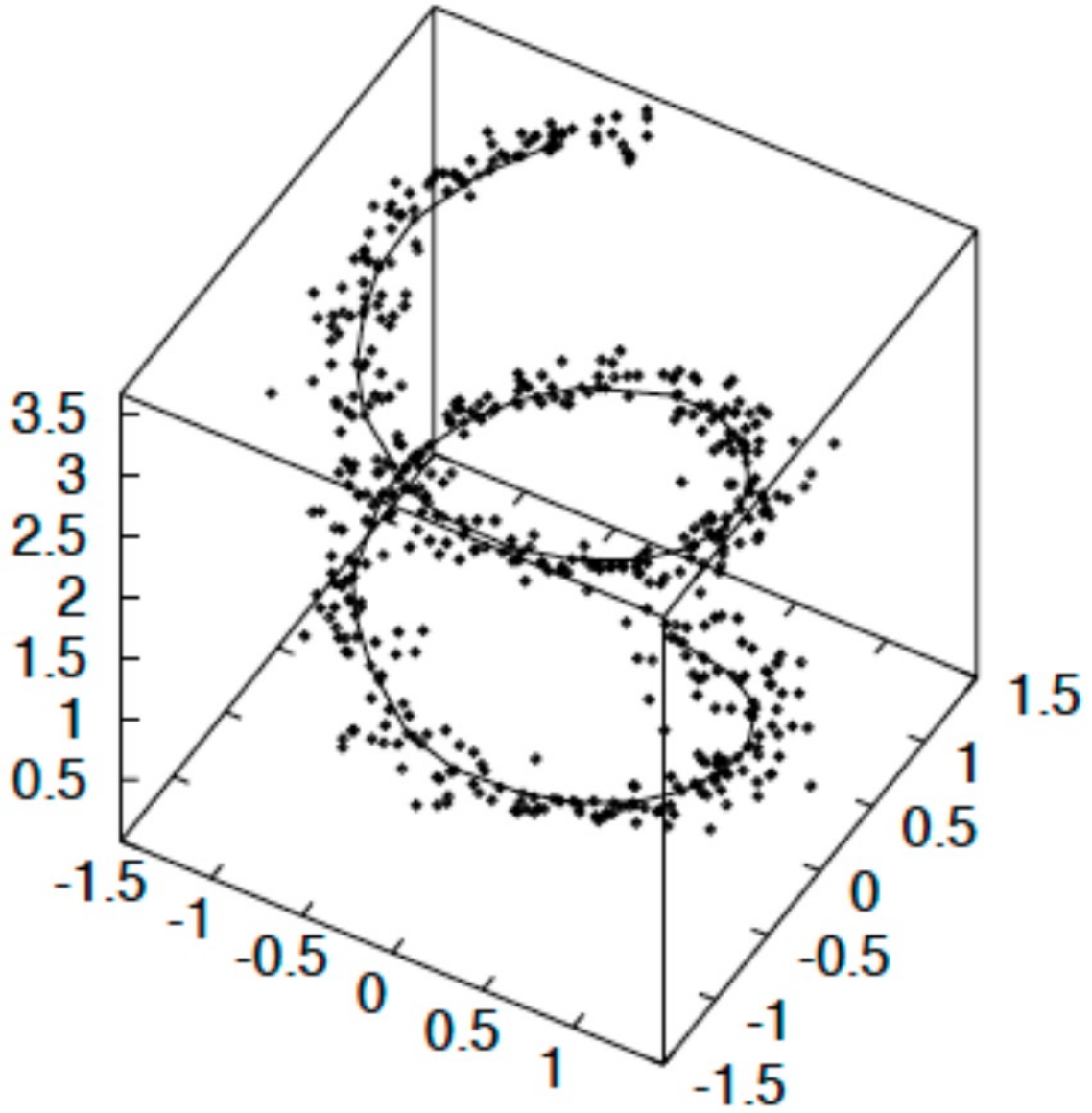
Or minimum <u>weight</u> spanning tree

1. Build a graph with edges between each pair of clusters
2. Add weights to each edge according to proximity of the two clusters
3. Find the shortest path between clusters



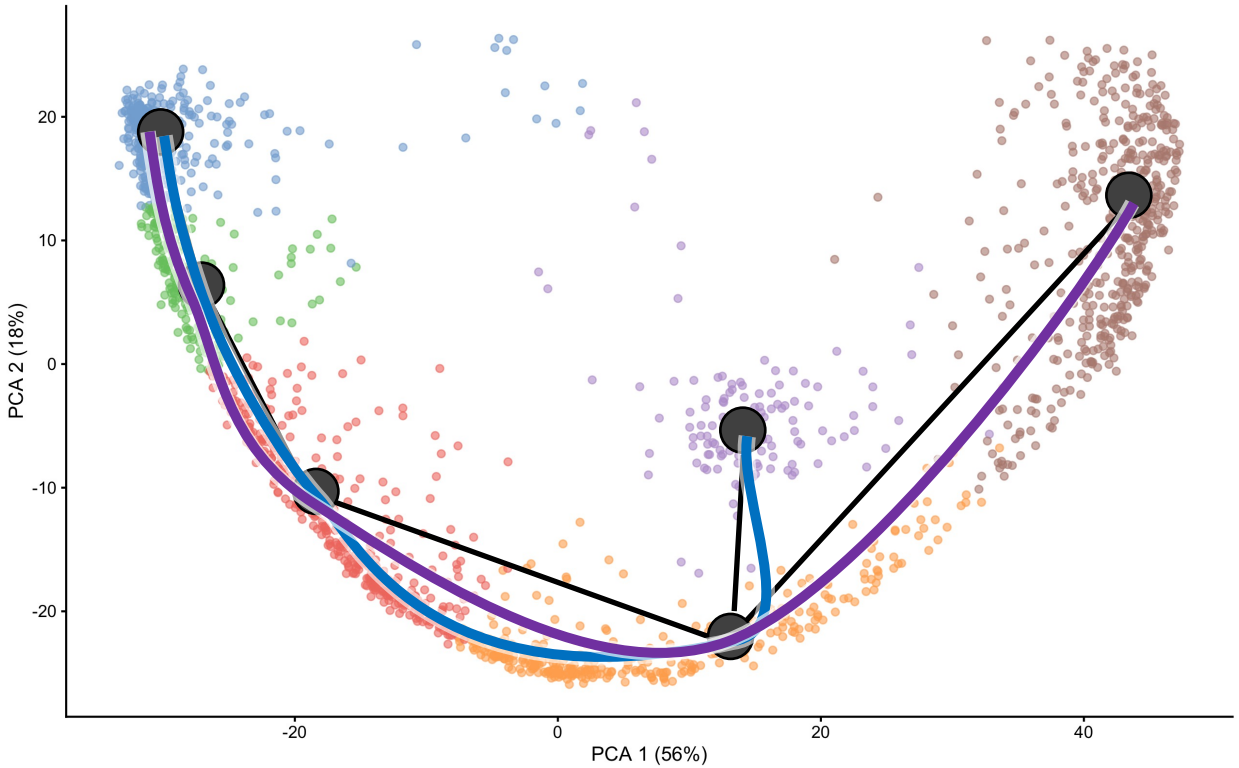
Principal curves?!?

A principal curve is a <u>smooth</u>, <u>one-dimensional, curve</u> that passes through the middle of a high-dimensional data set, providing a **<u>nonlinear summary of the data</u>**.



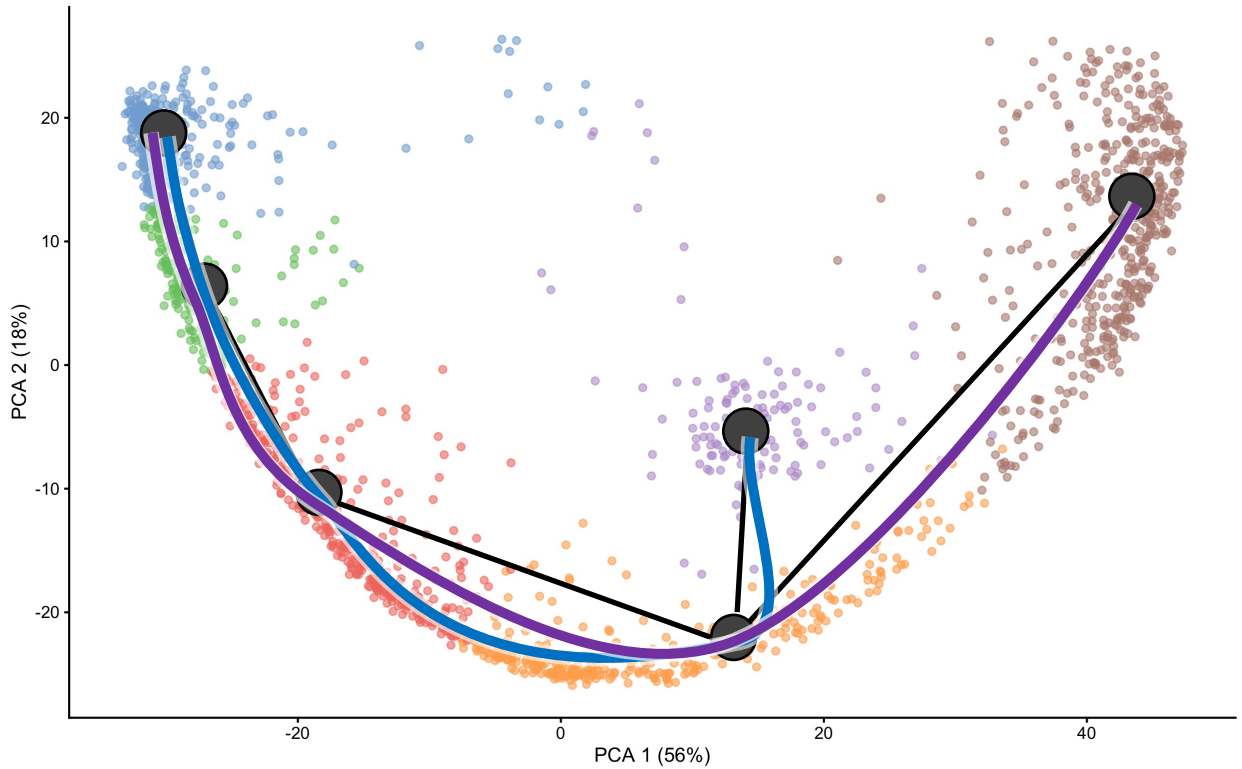
Principal curves?!?

A principal curve is a <u>smooth</u>, <u>one-dimensional, curve</u> that passes through the middle of a high-dimensional data set, providing a **<u>nonlinear summary of the data</u>**.



Principal curves?!?

A principal curve is a <u>smooth</u>, <u>one-dimensional, curve</u> that passes through the middle of a high-dimensional data set, providing a **<u>nonlinear summary of the data</u>**.

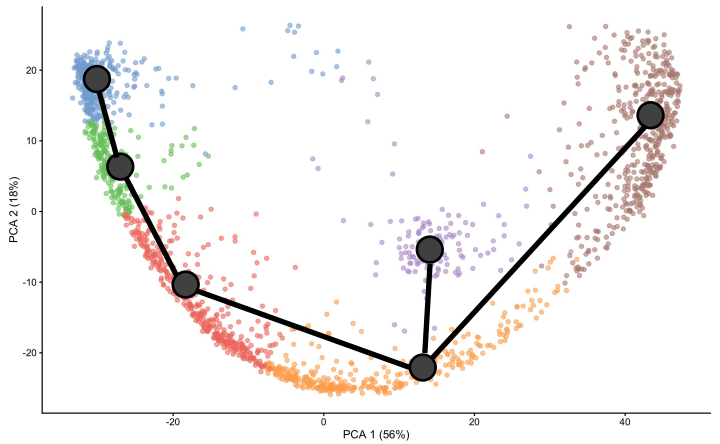


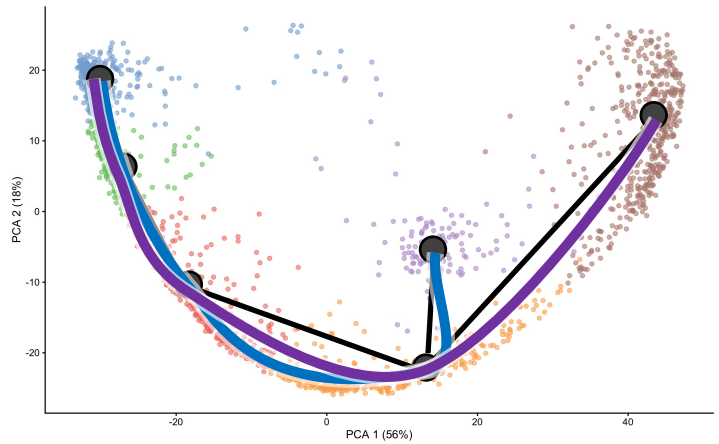
<u>CAREFUL:</u>

Once again, don't get tricked by the 2D visualization… Here, the principal curve is computed from 50 PCs, and subsequently embedded in only 2 PCs.

Why clustering quality matters for trajectory inference

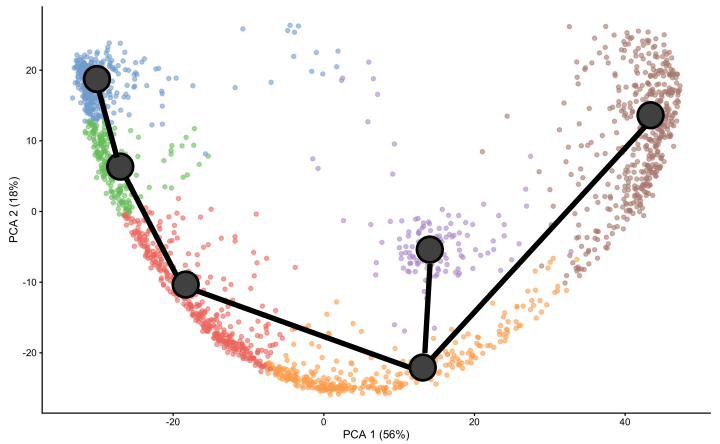
The MST is built from cluster centroids. If clustering changes, the MST will change and the fitted principal curves too.

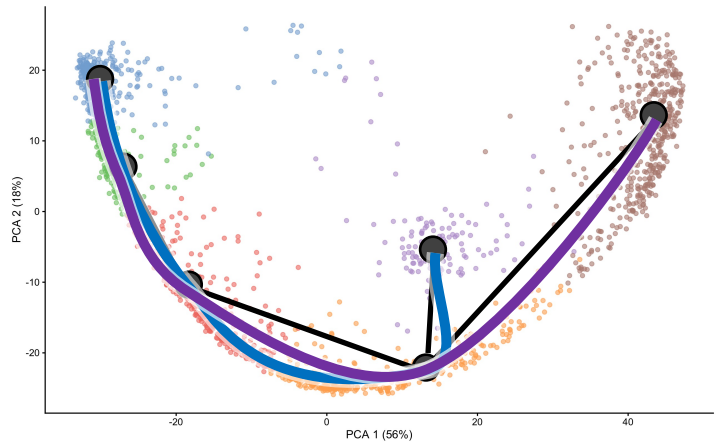




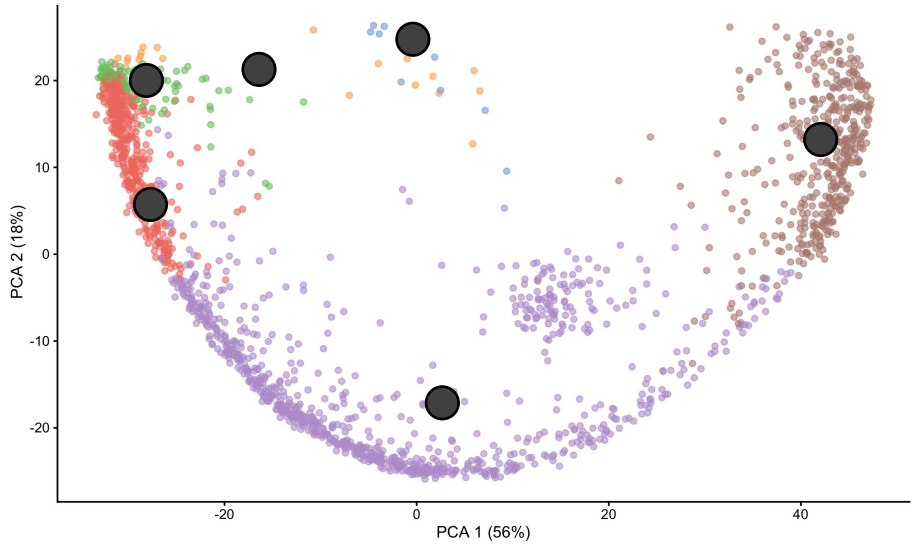
Why clustering quality matters for trajectory inference

The MST is built from cluster centroids. If clustering changes, the MST will change and the fitted principal curves too.



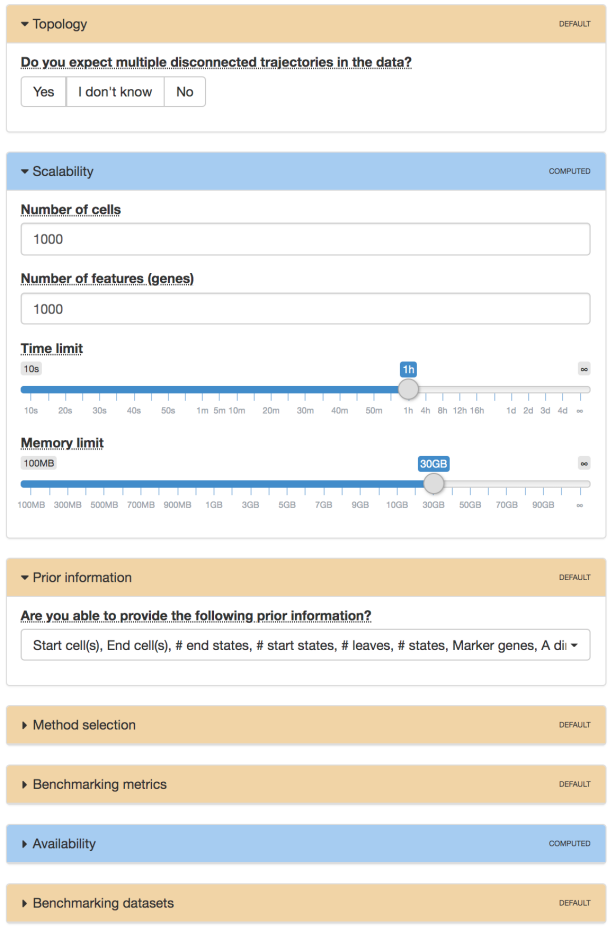


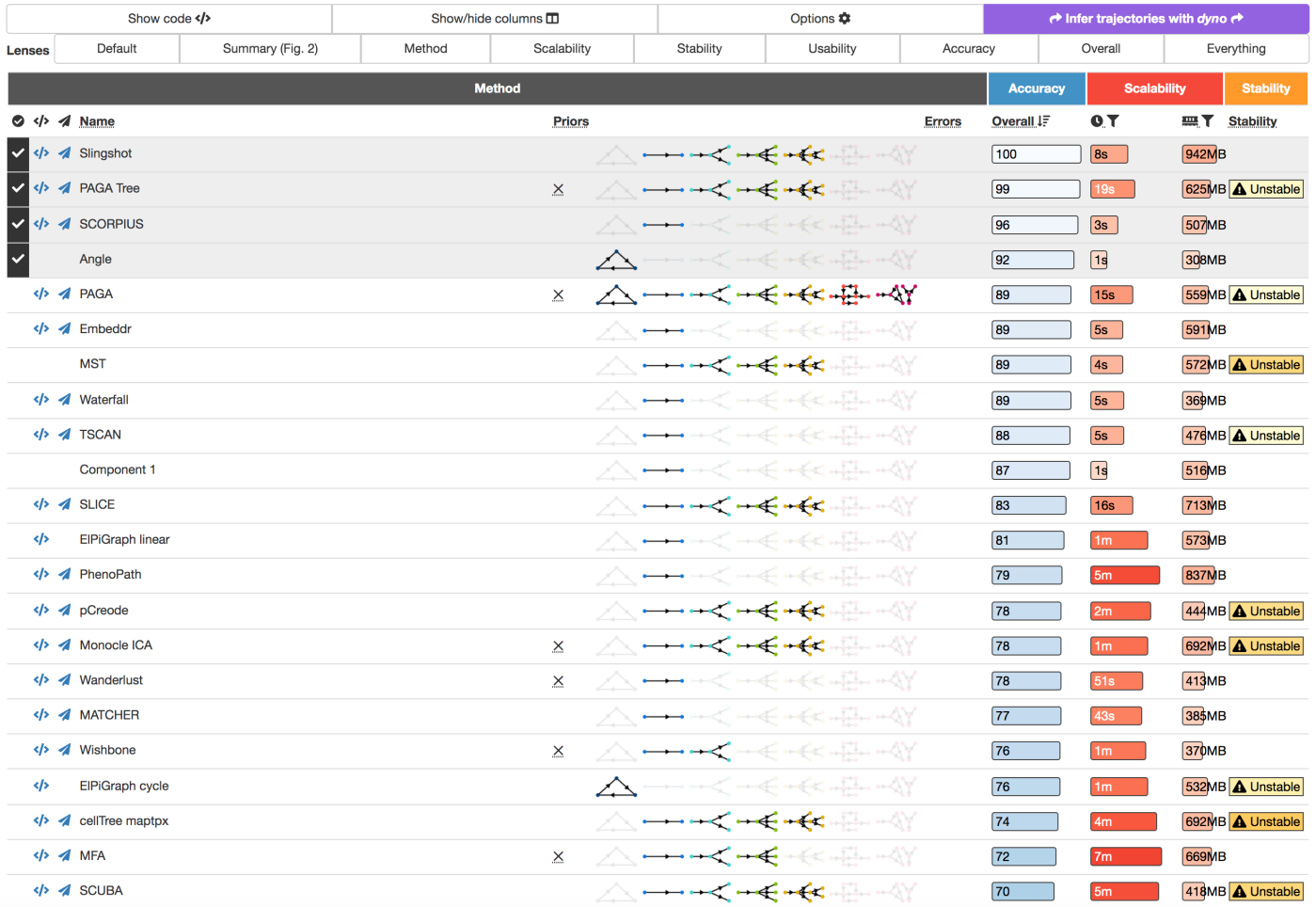
Only perform TI once you are confident your clusters represent a biologically relevant cell population.



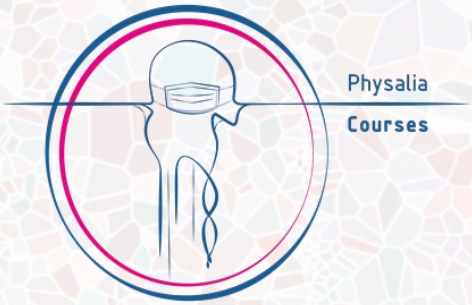
Making sense of all these different methods

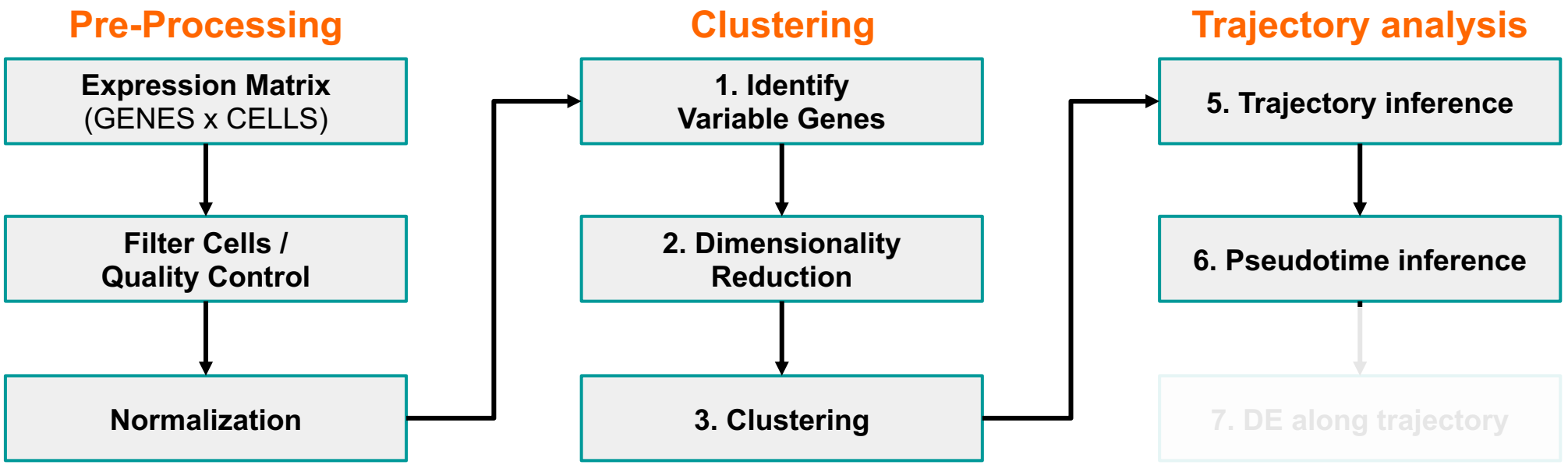
Dynverse project provides a handy exploratory/summarizing app to pick the best-suited TI algo.





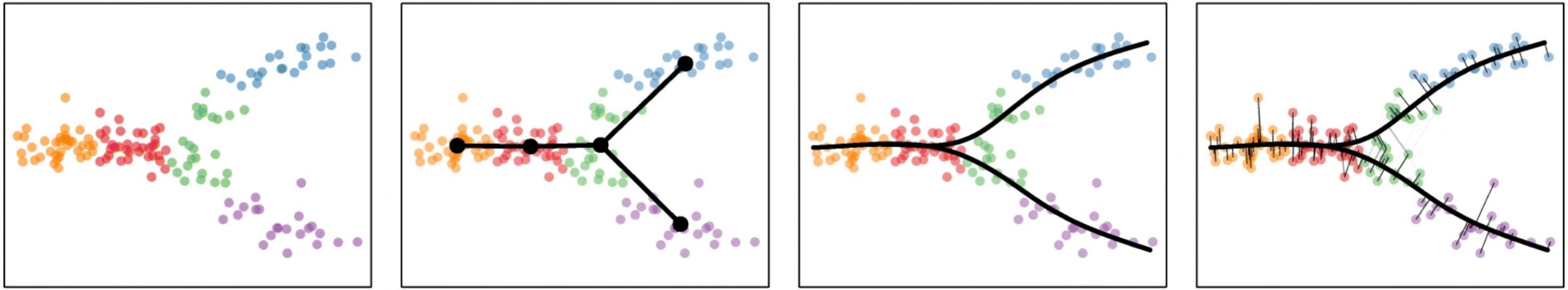
Analysis workflow





Pseudotime inference

In Slingshot, the pseudotimes values are estimated as the euclidean distance from each point to the closest principal curve (i.e. by **orthogonal projection of each point onto the curve**).

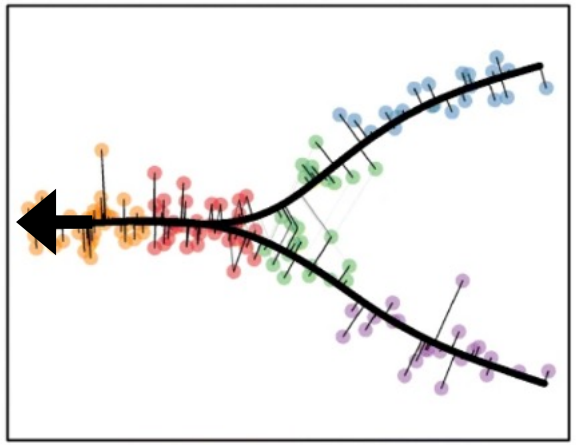


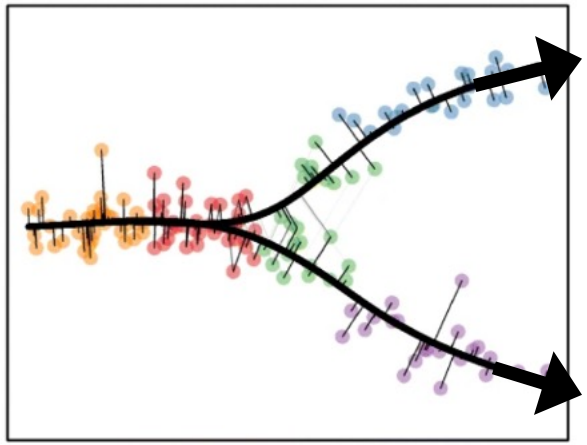
Street et al., 2018

But how are trajectories oriented???

A trajectory is not a **vector**: a trajectory is not **oriented**!

This is crucial when estimating pseudotime.



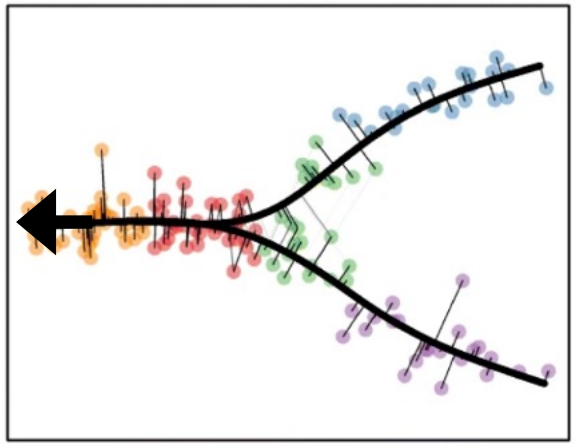


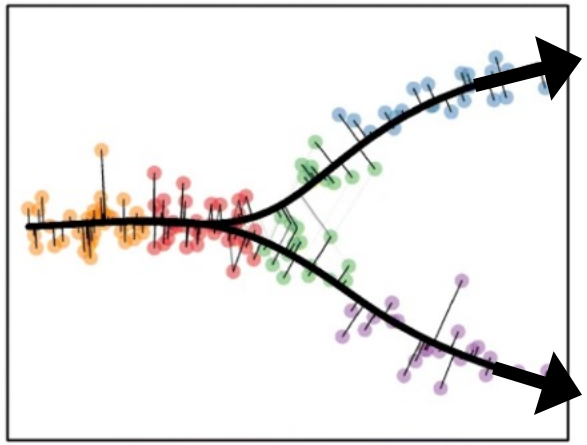
Street et al., 2018

But how are trajectories oriented???

A trajectory is not a **vector**: a trajectory is not **oriented**!

This is crucial when estimating pseudotime.





That is when one should use prior knowledge, ground-truth, or simply make a call.

Street et al., 2018

But how are trajectories oriented???

A trajectory is not a **vector**: a trajectory is not **oriented**!

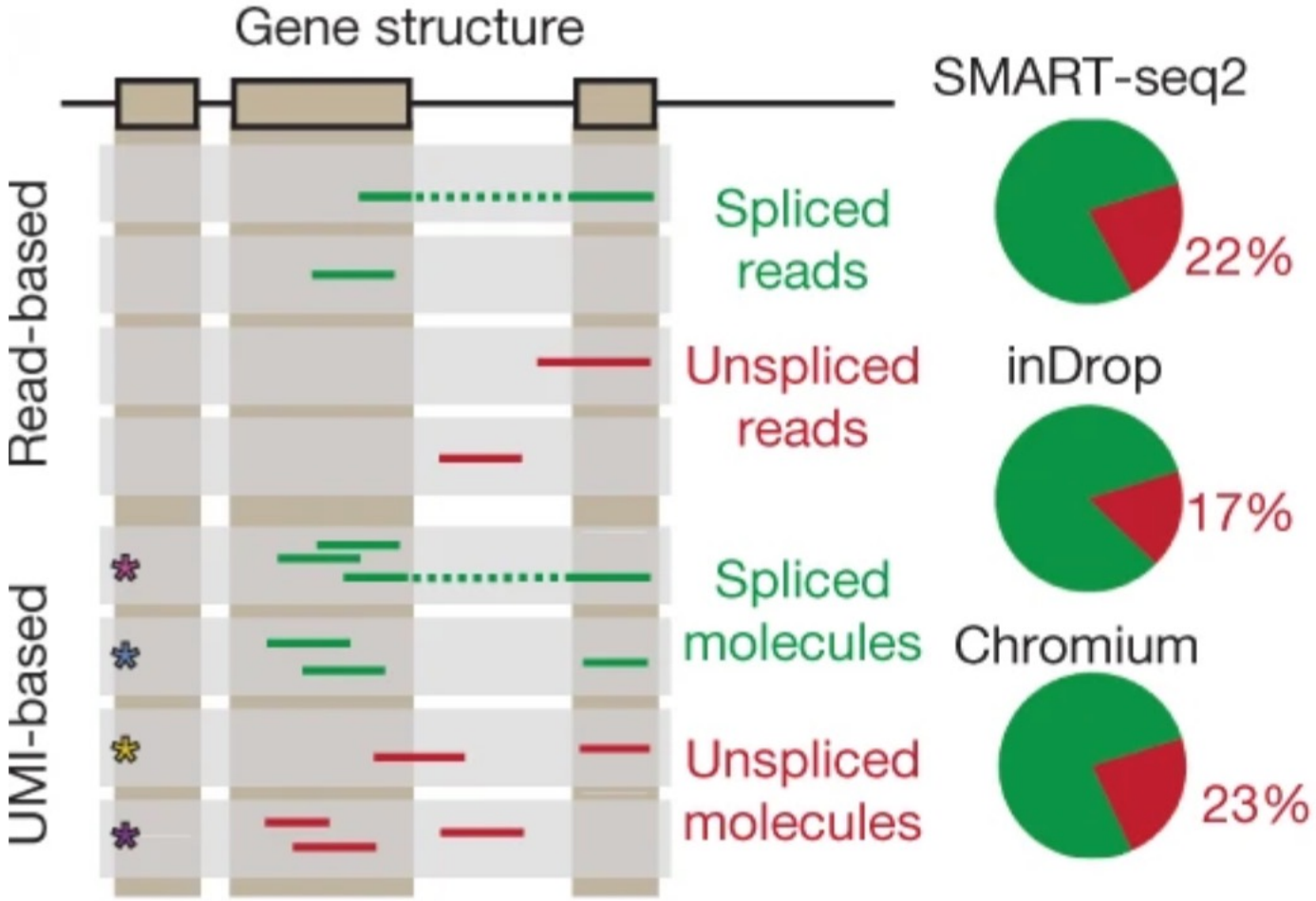
This is crucial when estimating pseudotime.

- Hematopoietic stem cell
- Lymphoid stem cell
- Committed lymphoid stem cell
- B cell
- T cell

Street et al., 2018

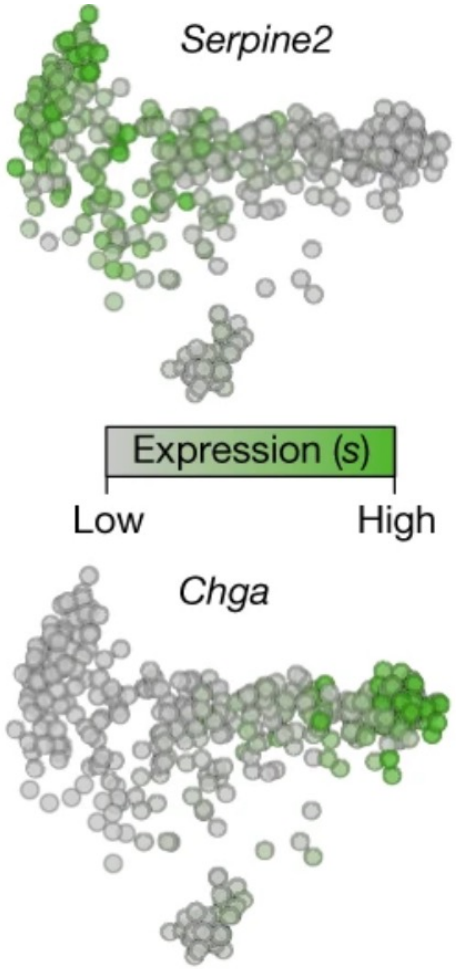
RNA velocity

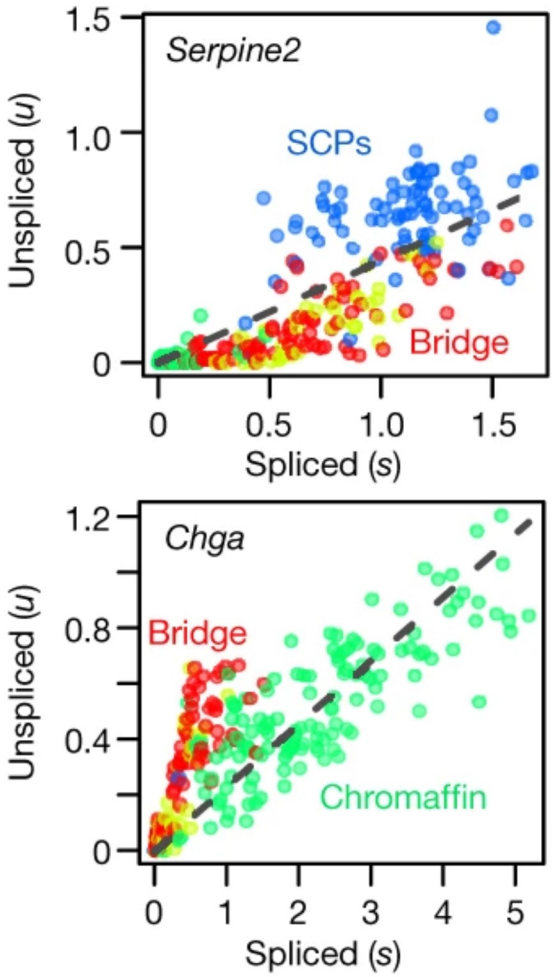
- Single-cell RNA-seq reads can be mapped onto exons or introns.

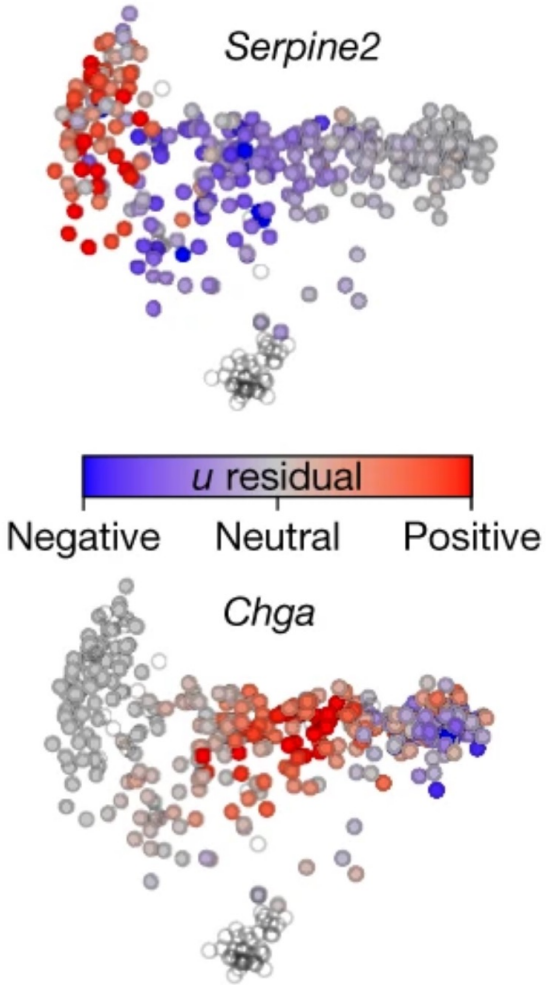


La Manno et al., 2018

RNA velocity



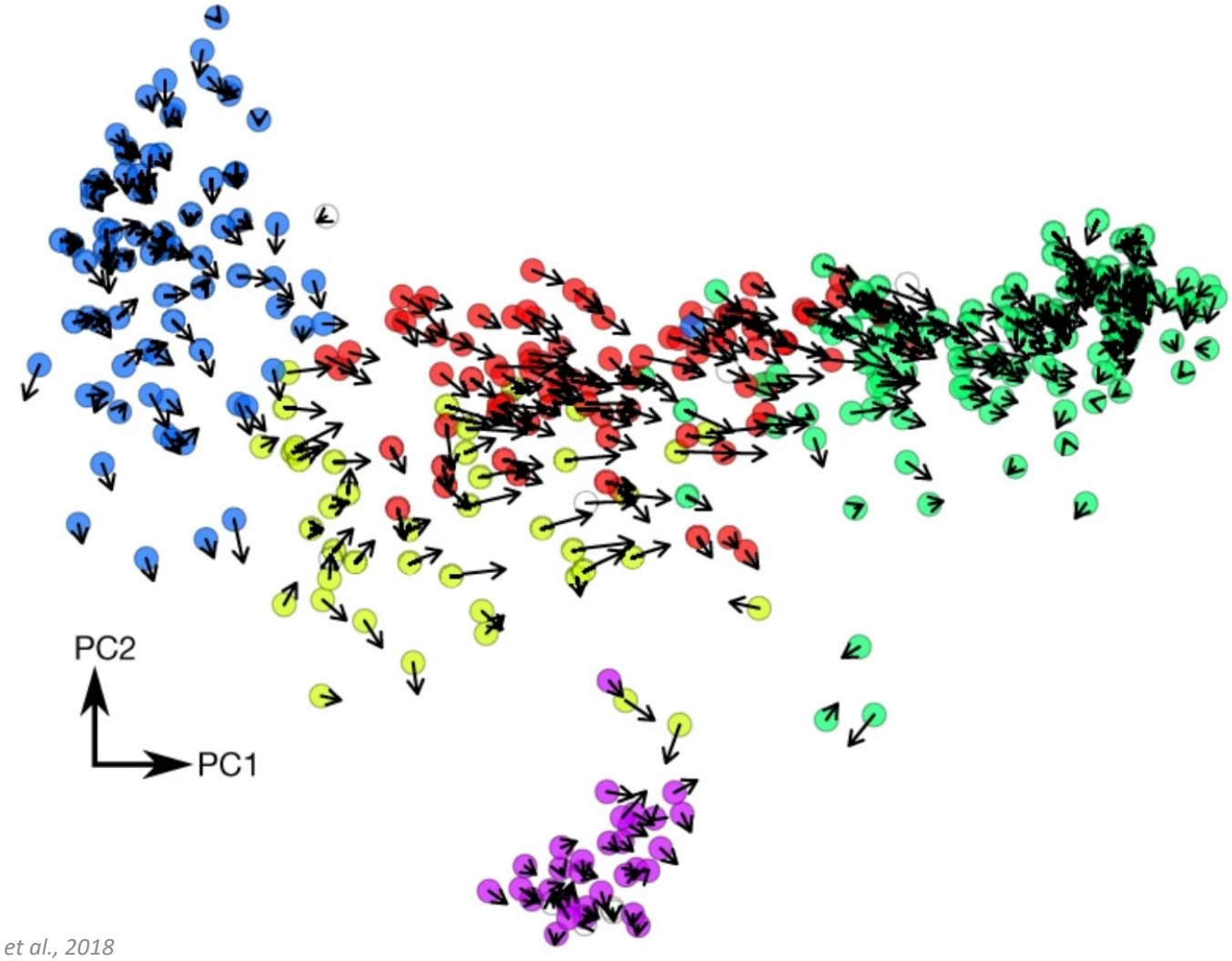




u residual represents whether the cell is far from an equilibrium state of expression

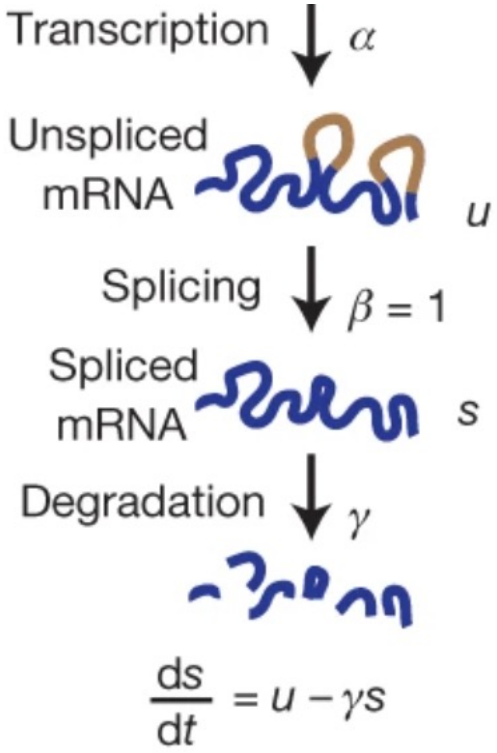
La Manno et al., 2018

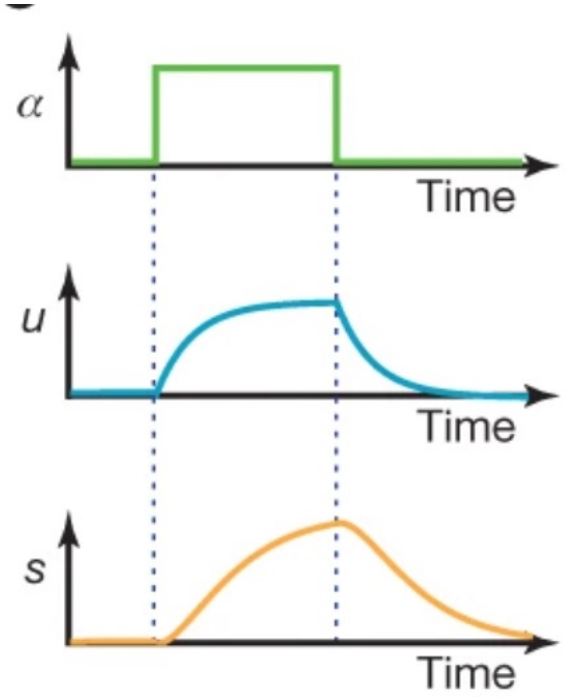
Generalizing this to all the detected (variable) genes, one can infer the future ”position” of each cell.



La Manno et al., 2018

Generalizing this to all the detected (variable) genes, one can infer the future ”position” of each cell.





CAREFUL!!!

This model relies on an important assumption:

that transcription is steady-state.

This is rarely true, even less true for differentiating cells!!

La Manno et al., 2018

RNA velocity generalized to non-steady-state cell populations

Article | Published: 03 August 2020

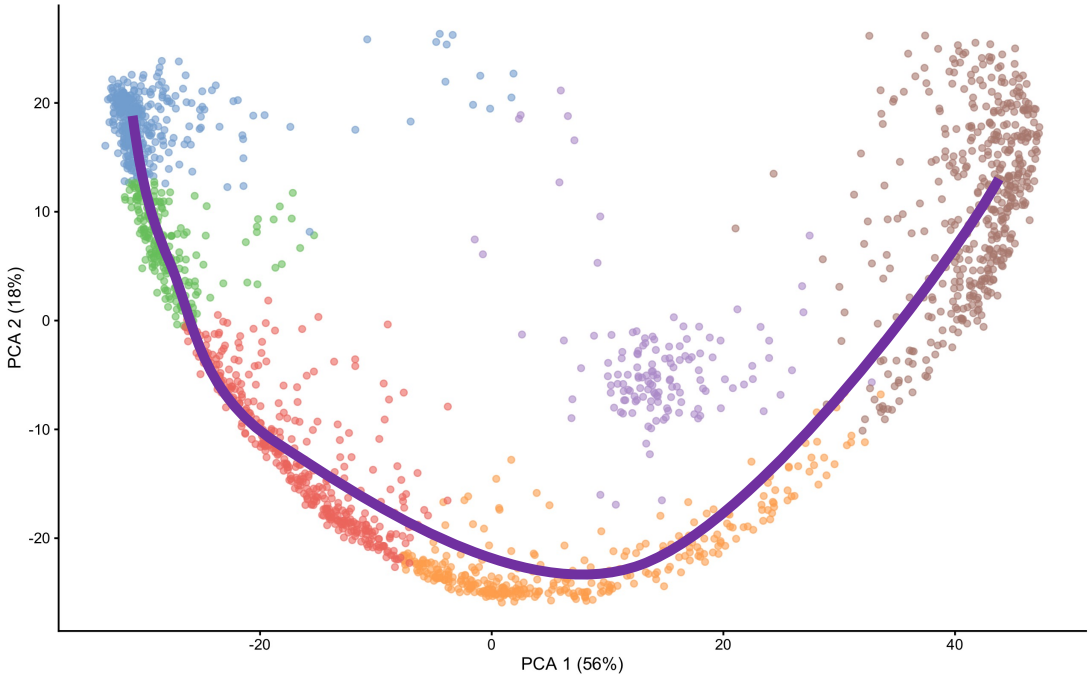
Generalizing RNA velocity to transient cell states through dynamical modeling

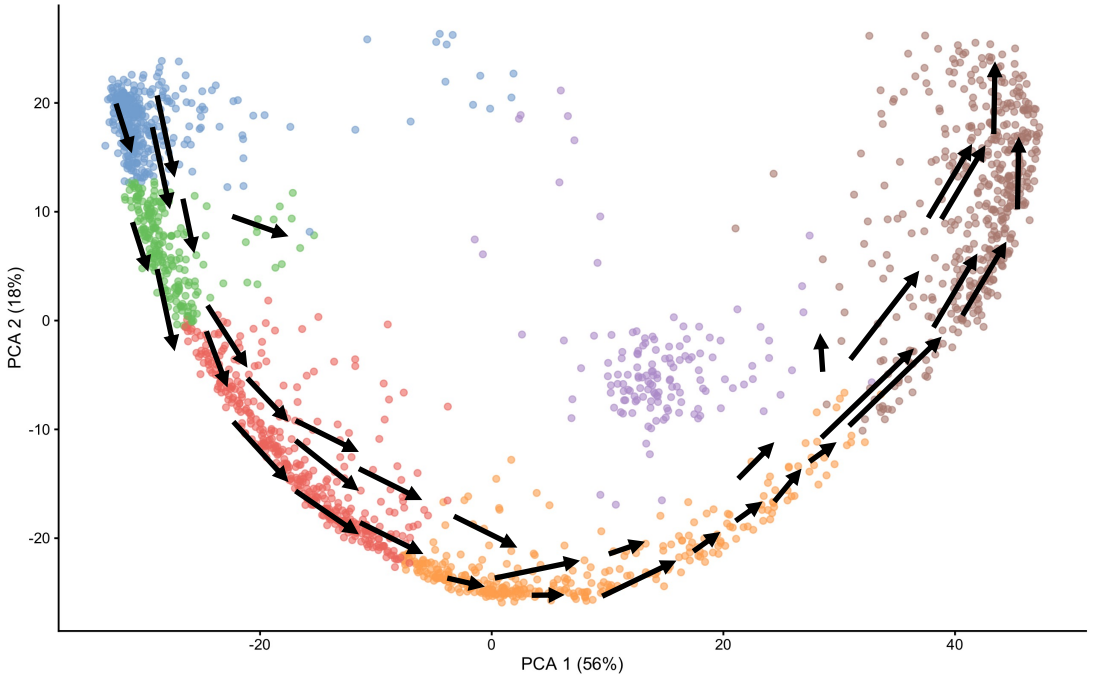
Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf ✉ & Fabian J. Theis ✉

Nature Biotechnology **38**, 1408–1414 (2020) | Cite this article

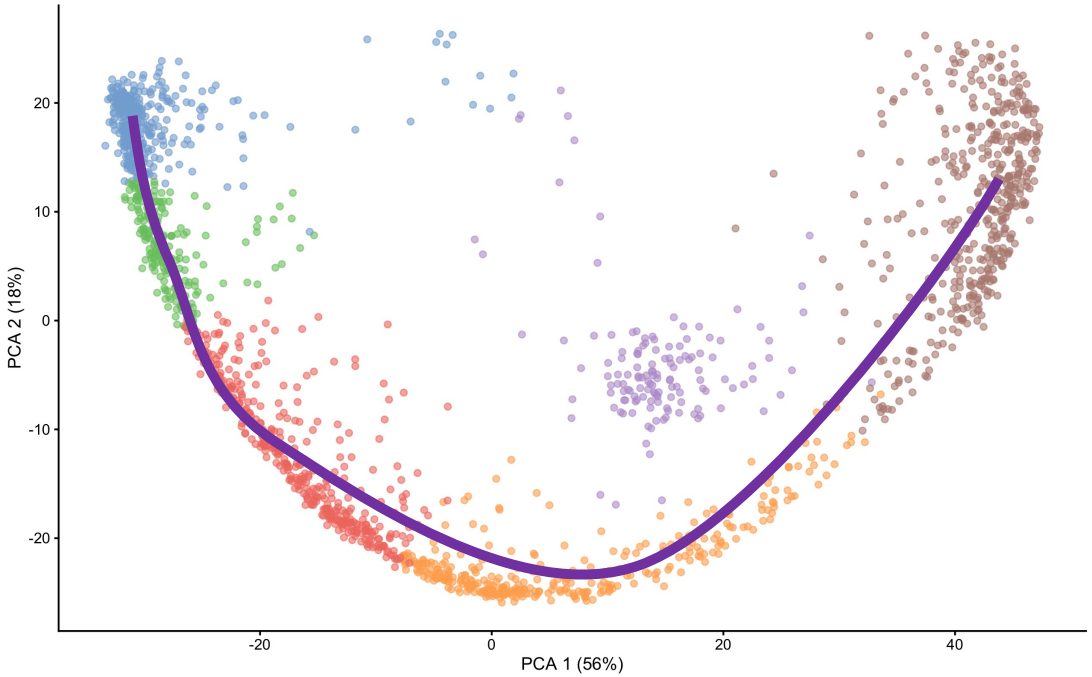
Bergen et al., 2020

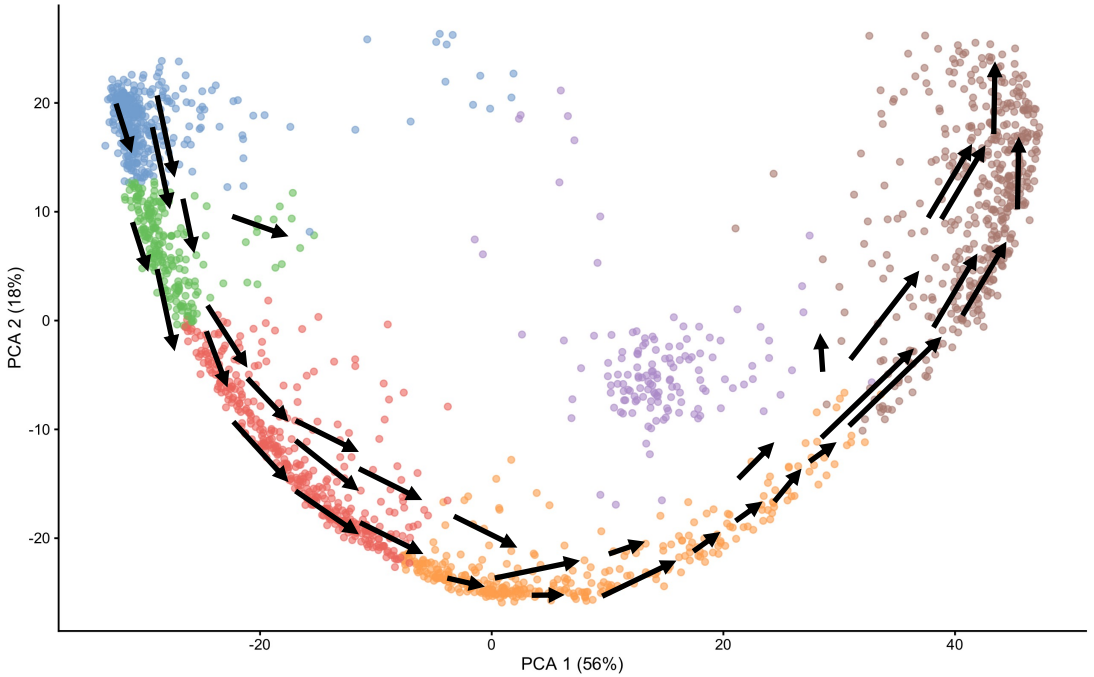
Using RNA velocity to infer directionality of the trajectory





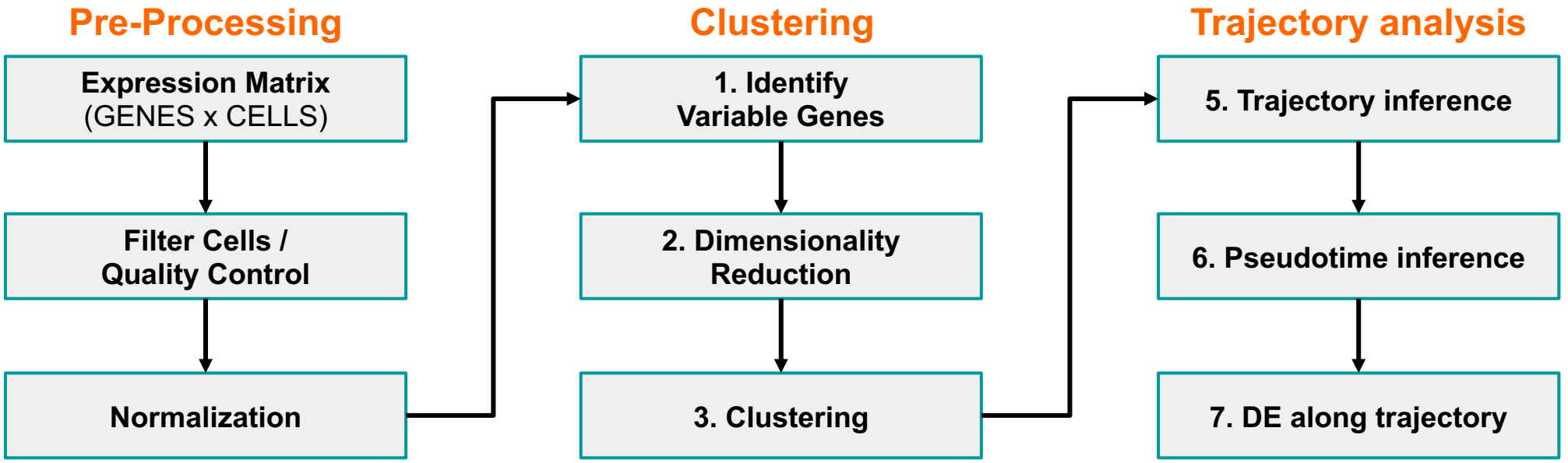
Using RNA velocity to infer directionality of the trajectory





RNA velocity implementations generally output a pseudotime value. However, they do not infer trajectories!! So they cannot capture a branching event, or cycles, etc….

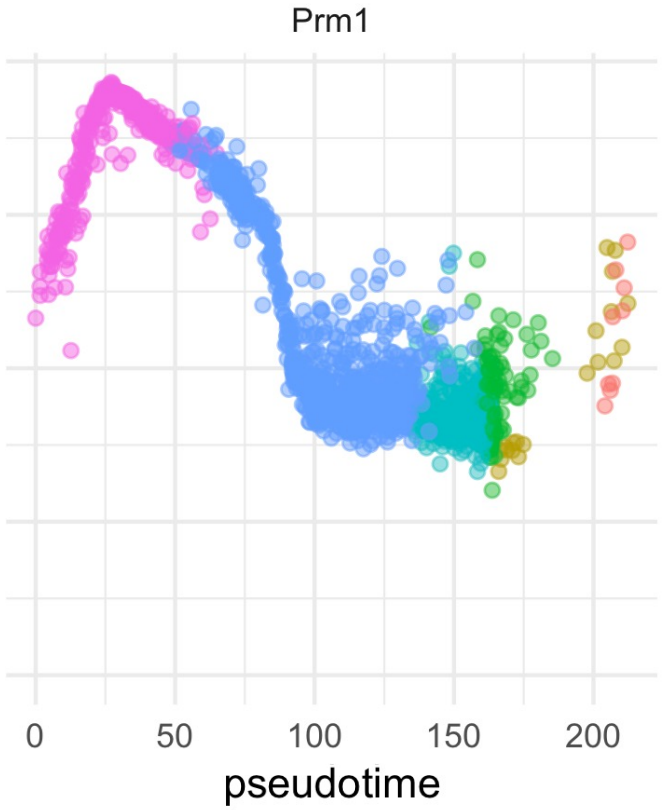
Analysis workflow



https://github.com/agitter/single-cell-pseudotime

Modeling gene expression along pseudotime

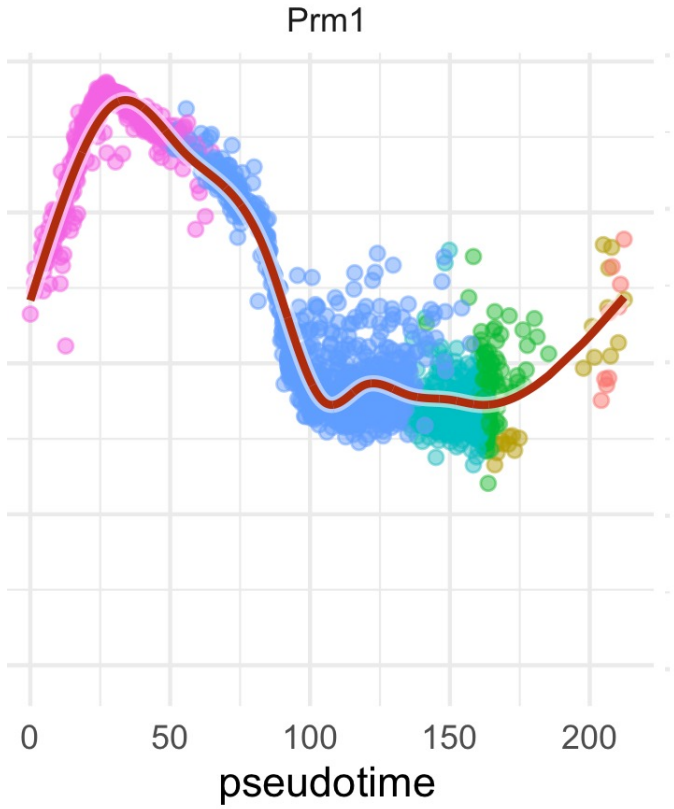
With a pseudotime value inferred to each cell, one can place all cells along an x axis, and plot gene expression on an y axis.



Modeling gene expression along pseudotime

With a pseudotime value inferred to each cell, one can place all cells along an x axis, and plot gene expression on an y axis.

Then time-dependent expression can be modelled, typically by fitting a GAM to the gene expression ~ pseudotime.



Performing temporal DE gene analysis

tradeSeq is an R package which facilitates GAM-fitting and gene DE analysis along and between trajectories.

It implements plug-and-play methods to use outputs from Slingshot, but most functions can be used directly with pseudotime values, without having to rely on trajectories specifically inferred with Slingshot.

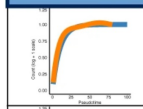
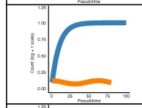
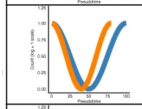
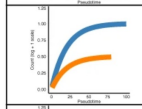
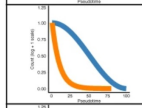
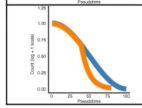
Article | Open Access | Published: 05 March 2020

Trajectory-based differential expression analysis for single-cell sequencing data

Koen Van den Berge, Hector Roux de Bézieux, Kelly Street, Wouter Saelens, Robrecht Cannoodt, Yvan Saeys, Sandrine Dudoit & Lieven Clement

Nature Communications **11**, Article number: 1201 (2020) | Cite this article

c

Lineages	Within the orange lineage		Between the orange and blue lineages		
	association test	startVsEnd test	diffEnd test	pattern test	earlyDE test
	DE	DE	Not DE	Not DE	Not DE
	Not DE	Not DE	DE	DE	DE
	DE	Not DE	Not DE	Not DE	Not DE
	DE	DE	DE	DE	Not DE
	DE	DE	Not DE	DE	DE
	DE	DE	Not DE	DE	Not DE

Van den Berge et al., 2020