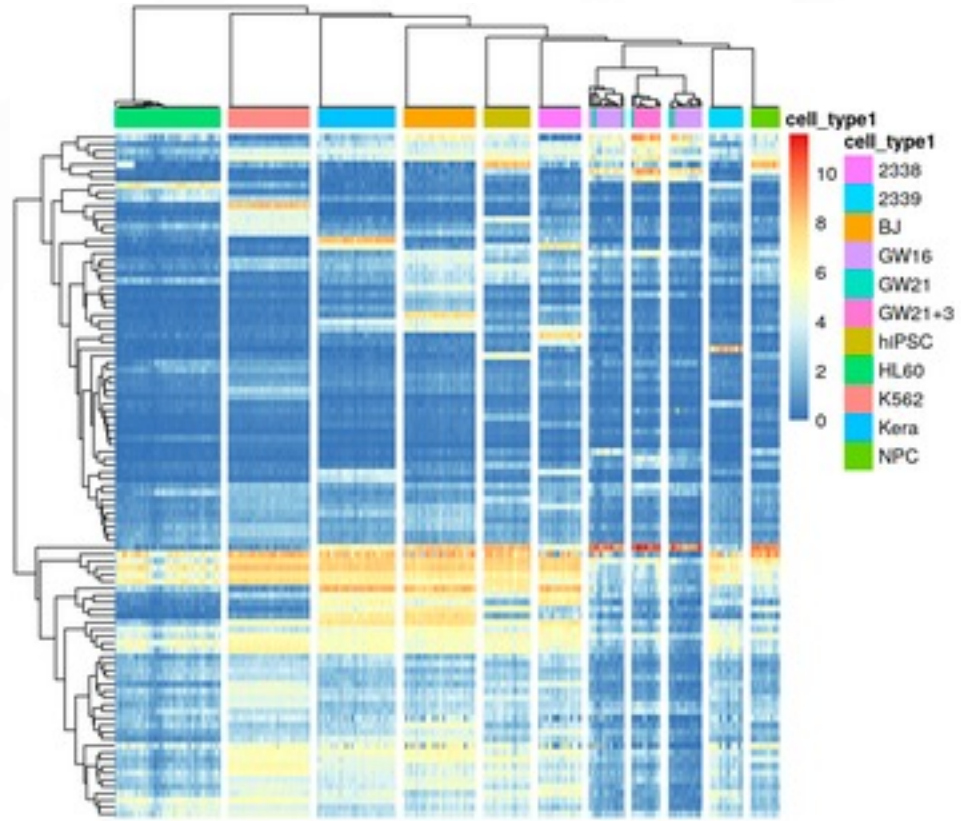
Inferring trajectories from scRNA-Seq data

Orr Ashenberg, Jacques Serizay, Fabricio Almeida-Silva

November 7, 2024

Physalia

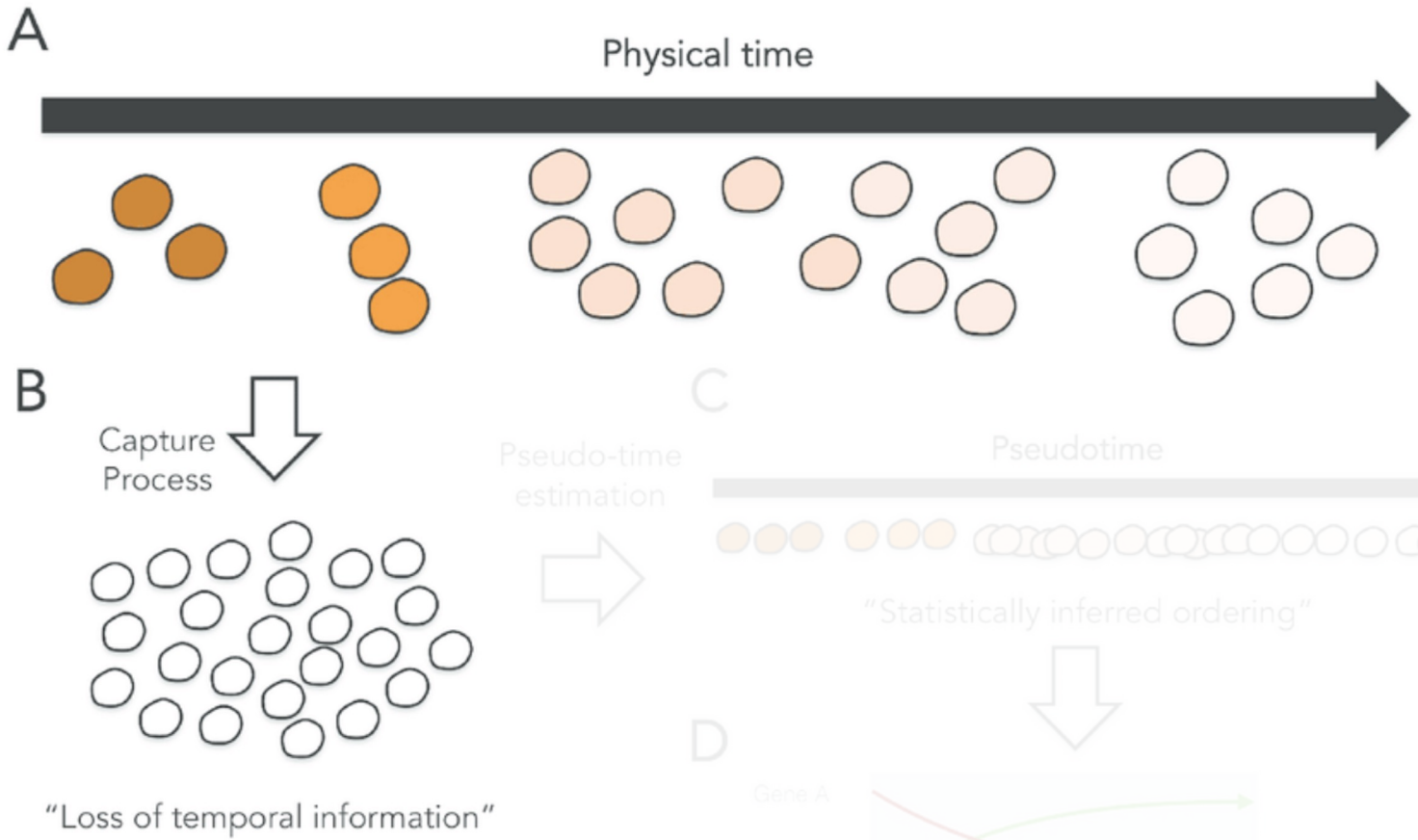
Courses



Outline: Inferring trajectories

- What does it mean to infer a trajectory?
- How to infer trajectories.
- How to infer pseudotime.
- How to identify genes that are differentially expressed along a trajectory.

Reconstructing trajectories from scRNA-Seq measurements

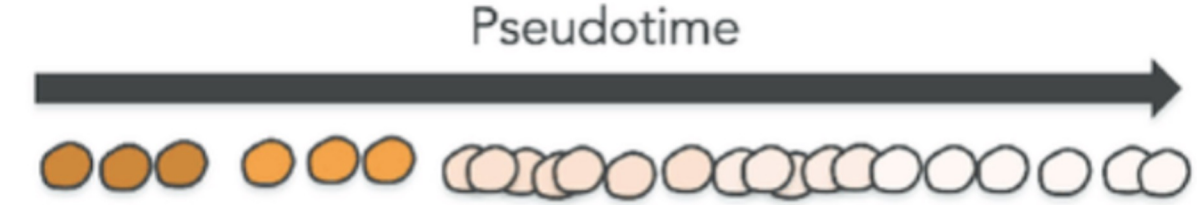


- In single-cell expression studies of processes such as cell differentiation, captured cells may be widely distributed in terms of progress.
- In a population of cells captured at the same time, some cells might be far along, while others might not yet have begun the process.

Defining pseudotime along a trajectory

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

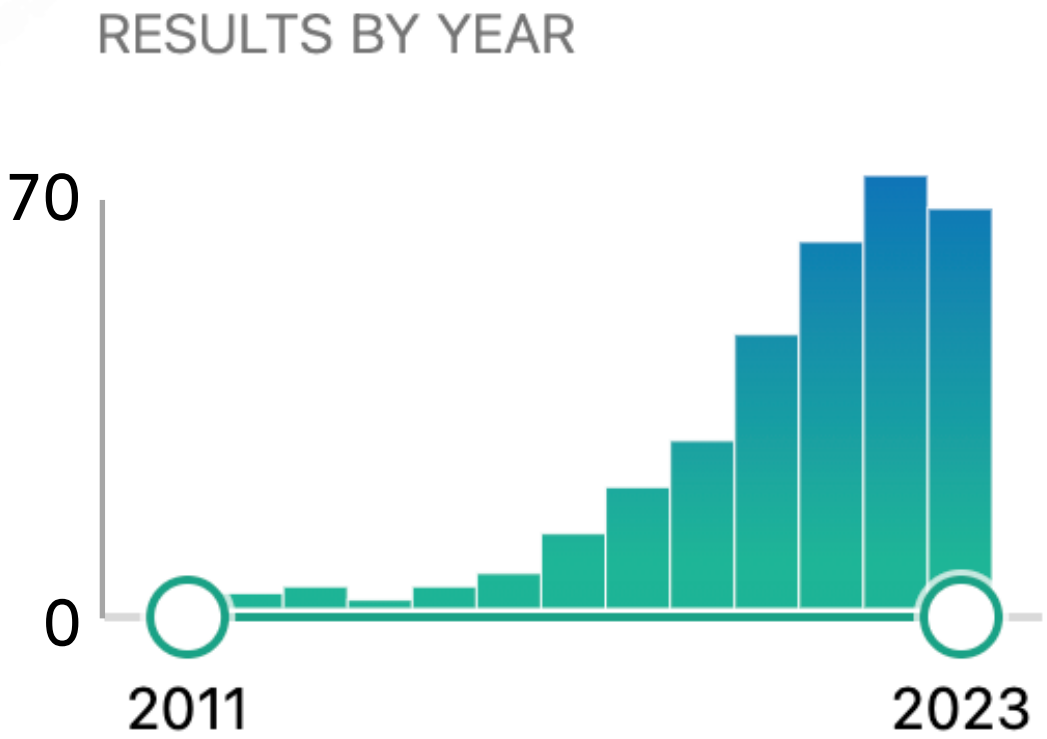
- Pseudotime is an abstract unit of progress: it's simply the distance between a cell and the start of the trajectory, measured along the shortest path
- The trajectory's total length is defined in terms of the total amount of transcriptional change that a cell undergoes as it moves from the starting state to the end state.



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

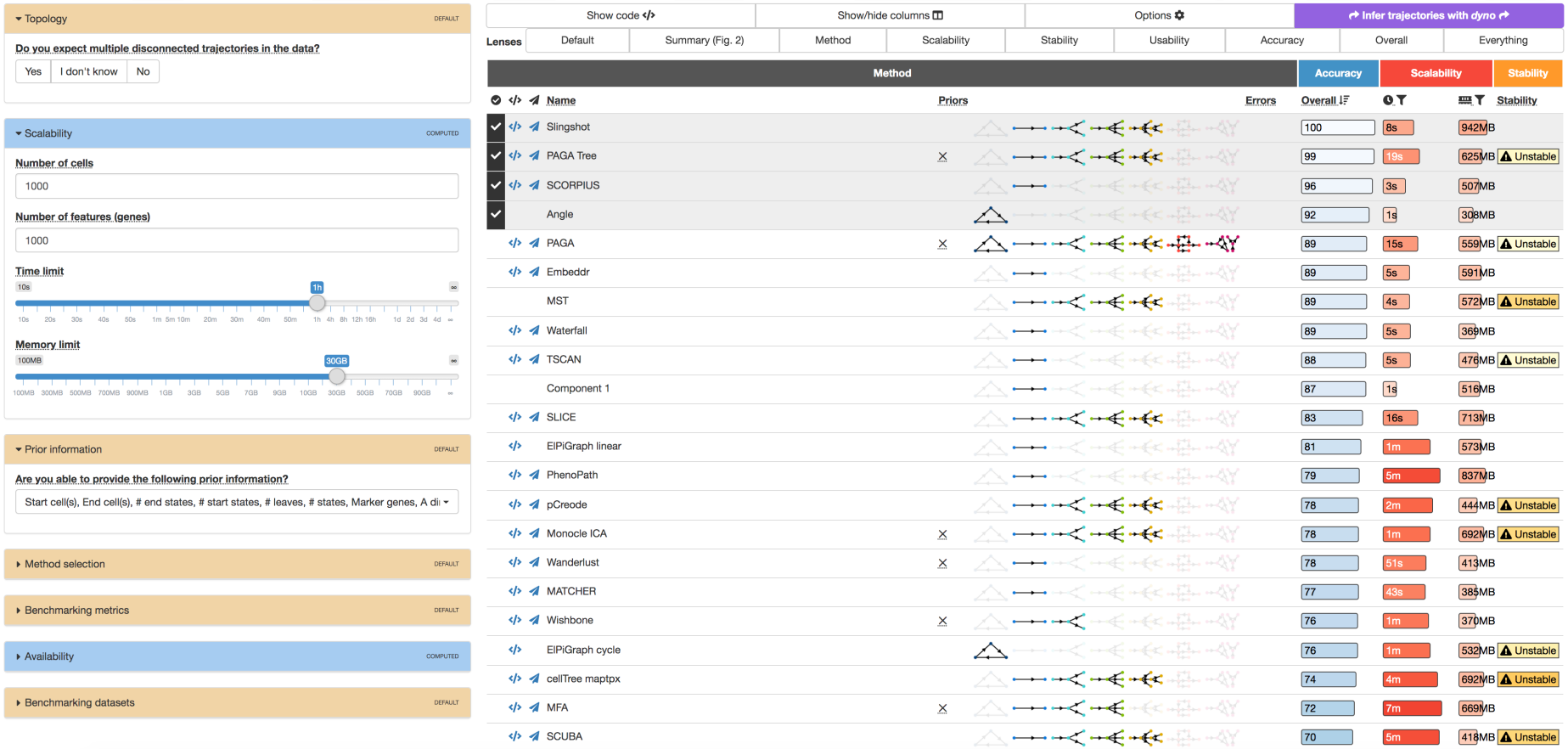
There are many trajectory inference tools!

Pubmed results for "*trajectory inference method single-cell*"



Comparisons of trajectory inference tools

Dynverse lets you infer and compare trajectories using multiple algorithms.

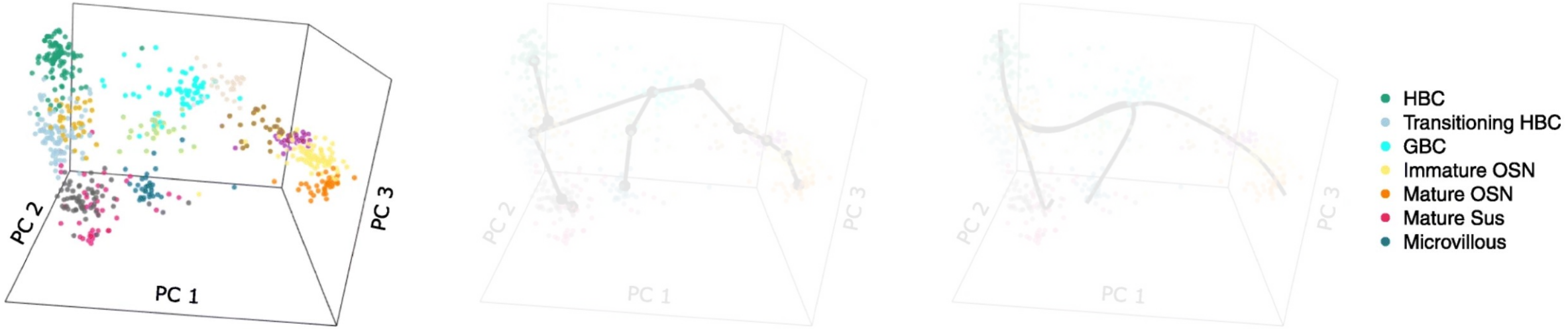


guidelines.dynverse.org

Saelens W. et al. (2019) *Nature Biotechnology.*

Slingshot is one of the most widely used and robust approaches to infer trajectory

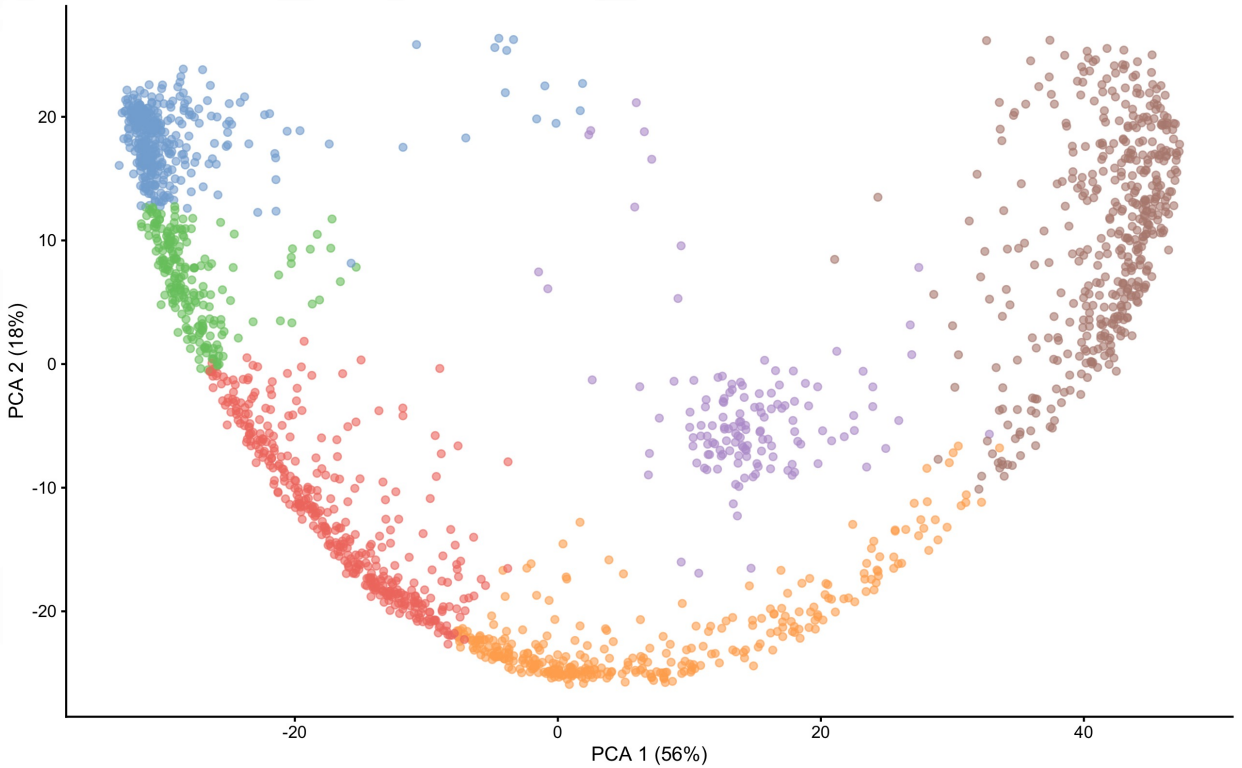
1. Cluster scRNA-Seq data in a low-dimensional space.
2. Building a minimum spanning tree (MST) on the clusters.
3. Fit principal curves through the MST.



Street K. et al. (2018) *BMC Genomics.*

Building a minimum spanning tree

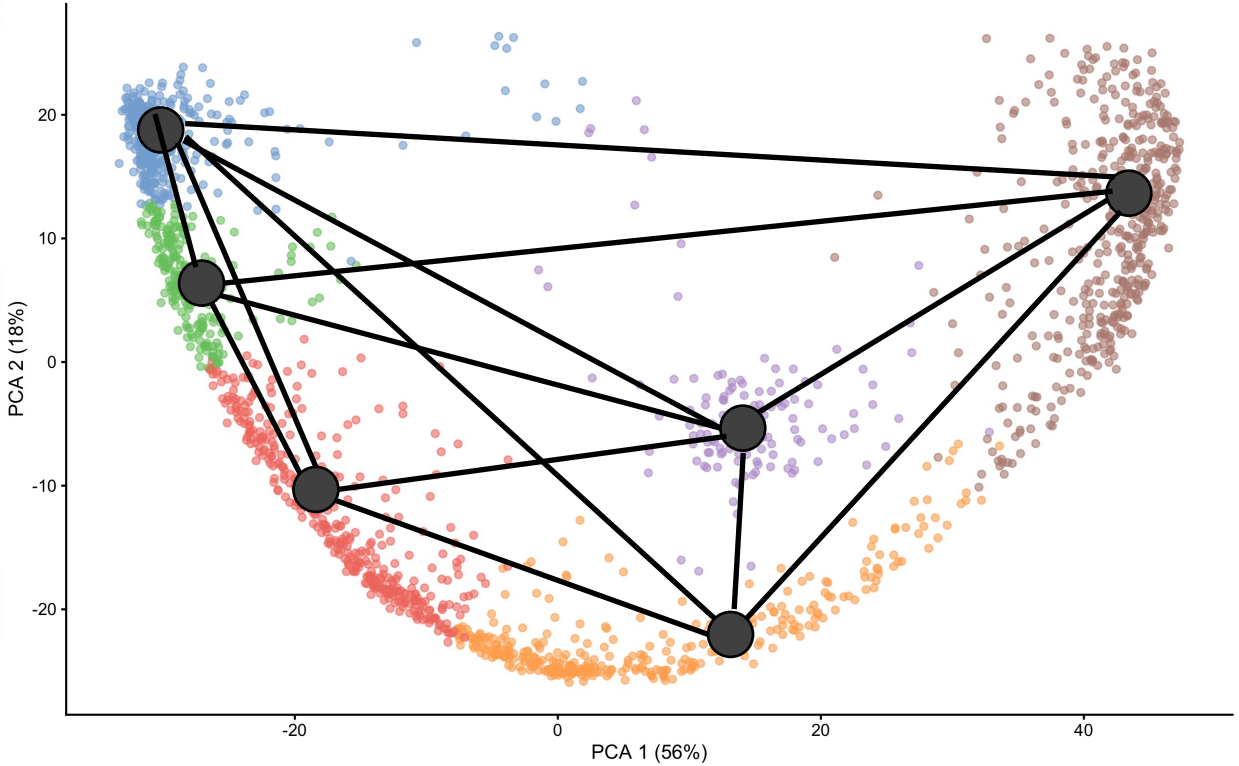
1. Build a graph with edges between each pair of clusters.



Street K. et al. (2018) *BMC Genomics.*

Building a minimum spanning tree

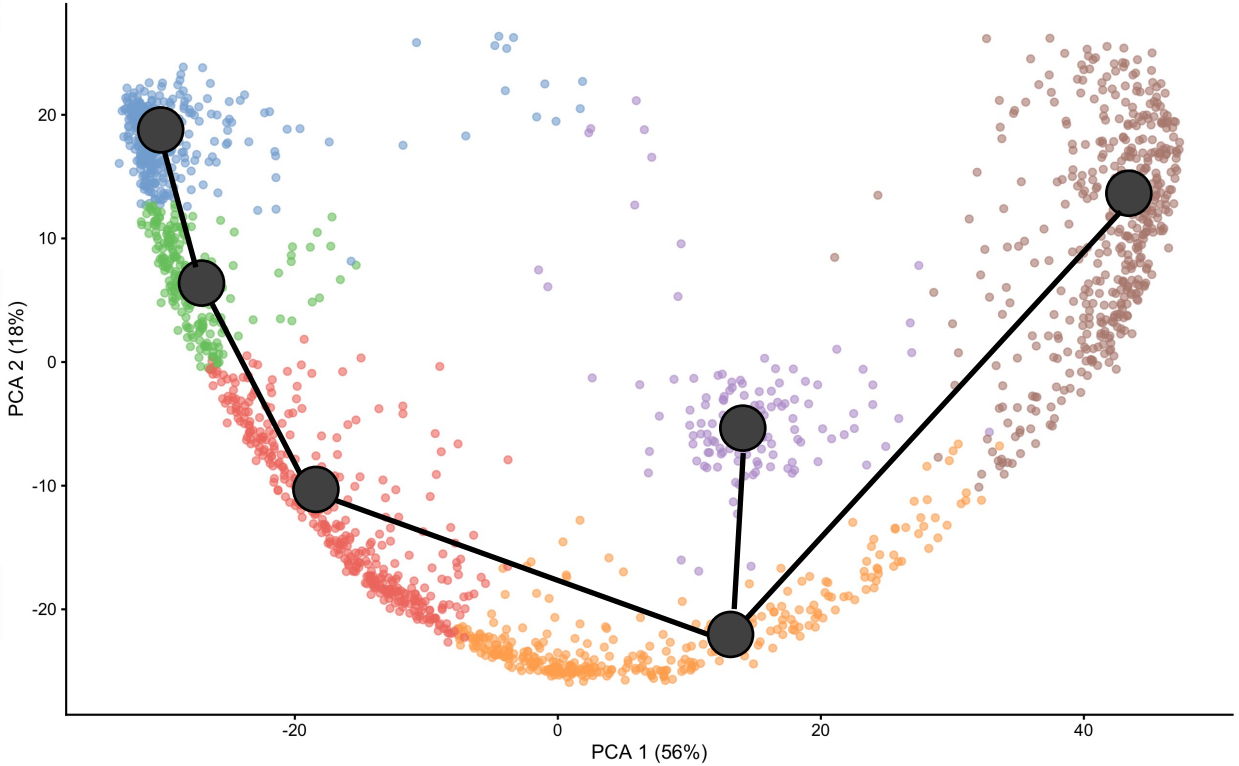
1. Build a graph with edges between each pair of clusters.
2. Add weights to each edge according to proximity of the two clusters.



Street K. et al. (2018) *BMC Genomics.*

Building a minimum spanning tree

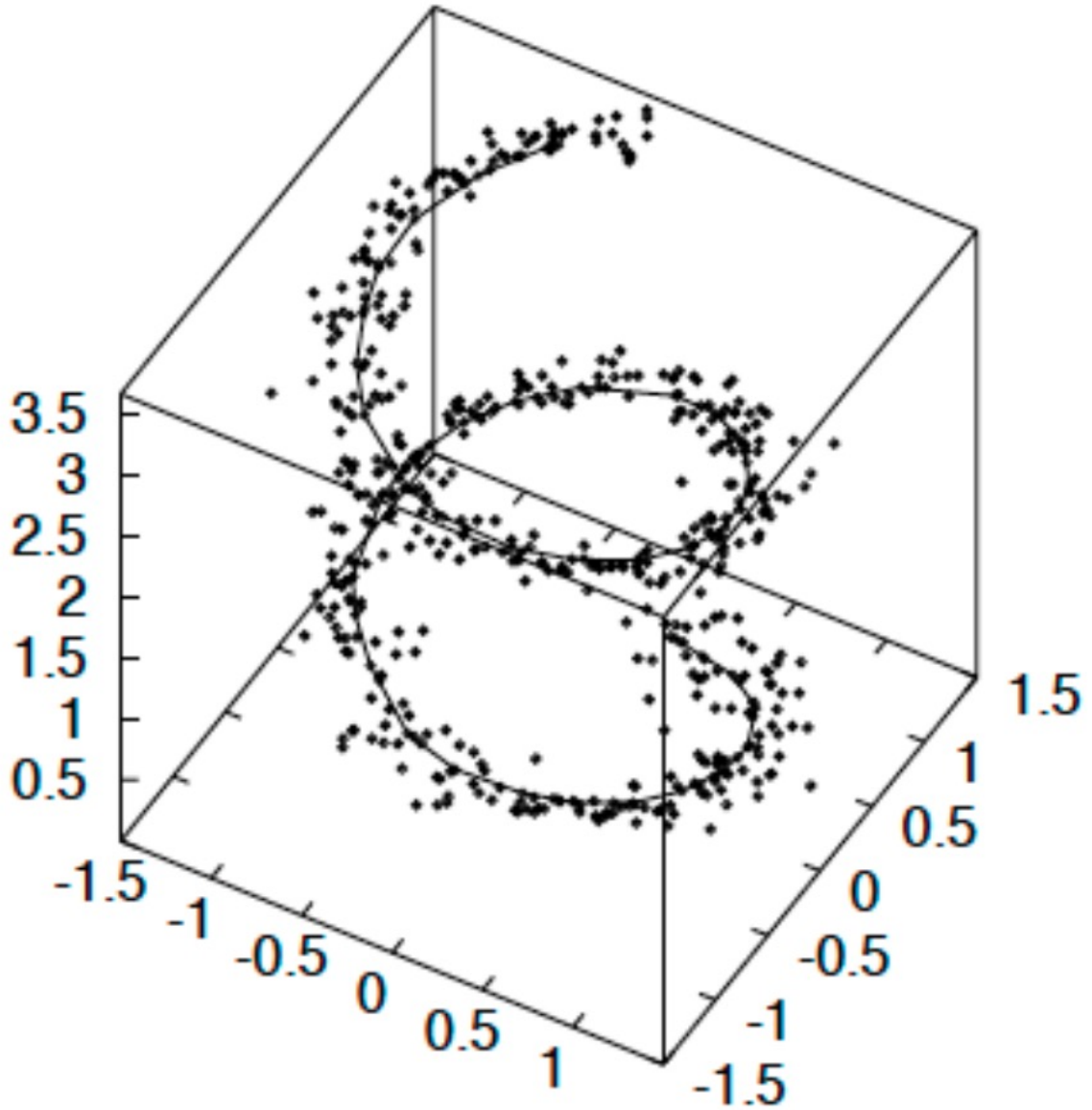
1. Build a graph with edges between each pair of clusters.
2. Add weights to each edge according to proximity of the two clusters.
3. Find the shortest path between clusters.



Street K. et al. (2018) *BMC Genomics.*

Fitting principal curves to tree and measuring pseudotime

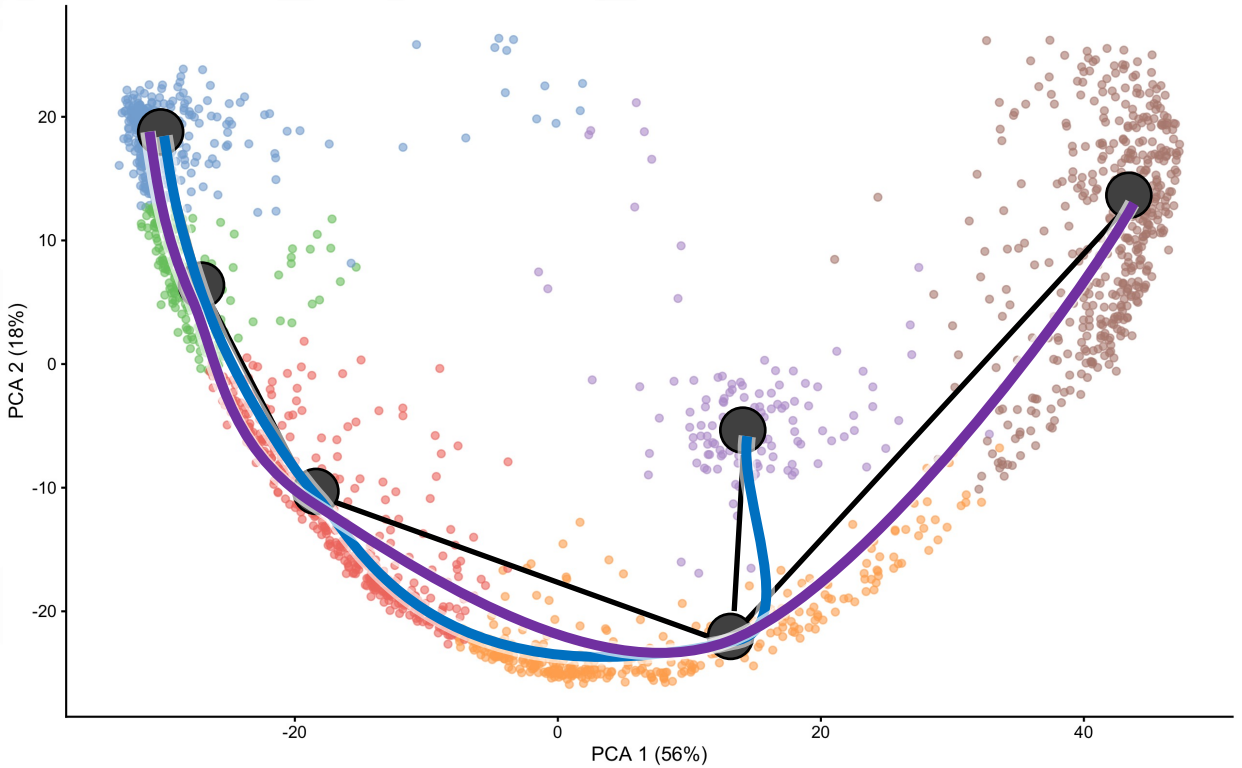
A principal curve is a **smooth, one-dimensional curve** that passes through the middle of a high-dimensional data set, providing a **nonlinear summary of the data**.



Hastie T. et al. (1989) *Journal of the American Statistical Association.*

Fitting principal curves to tree and measuring pseudotime

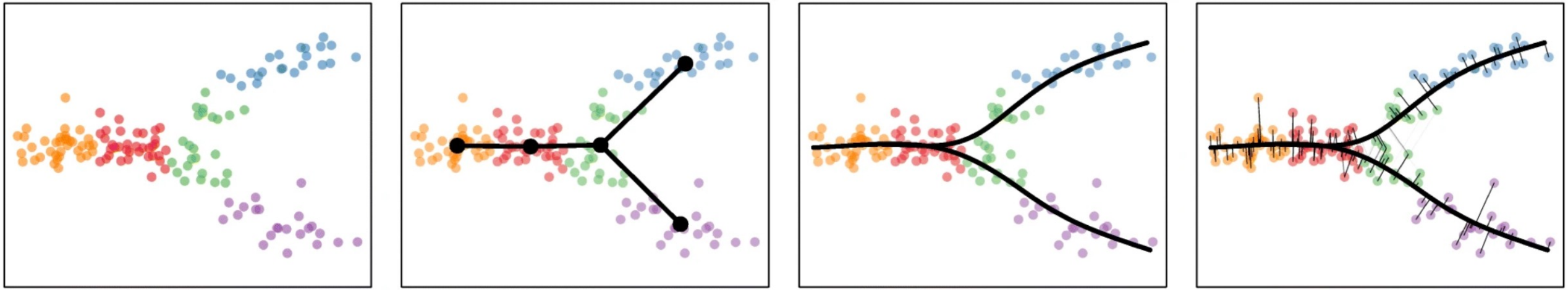
A principal curve is a **smooth, one-dimensional curve** that passes through the middle of a high-dimensional data set, providing a **nonlinear summary of the data**.



Street K. et al. (2018) *BMC Genomics*.

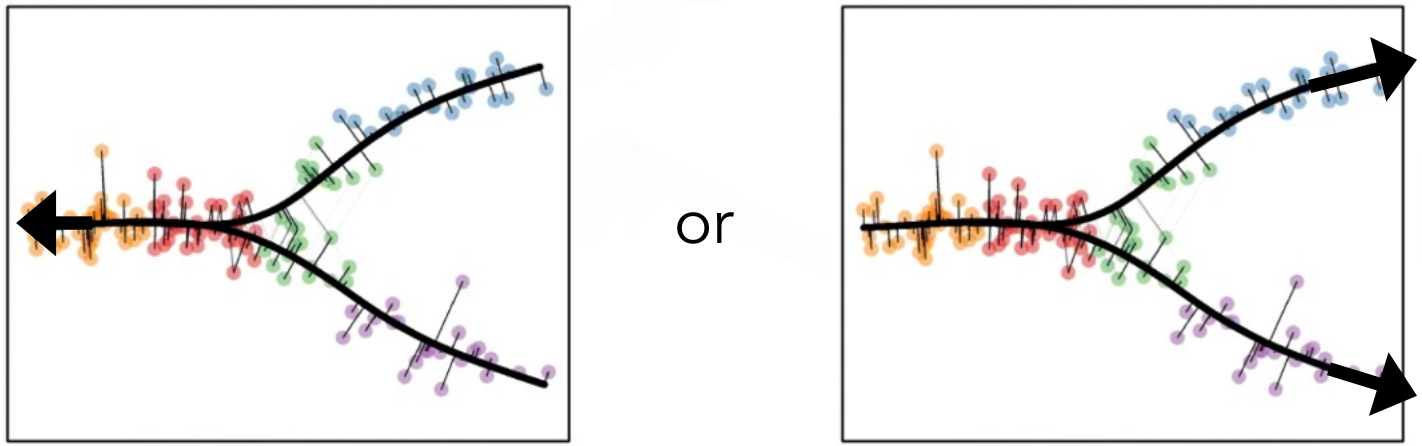
Fitting principal curves to tree and measuring pseudotime

In Slingshot, the pseudotime 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).



Which cell came first? How is a trajectory oriented?

In most trajectory inference algorithms, the trajectory is not a vector as it lacks an orientation!



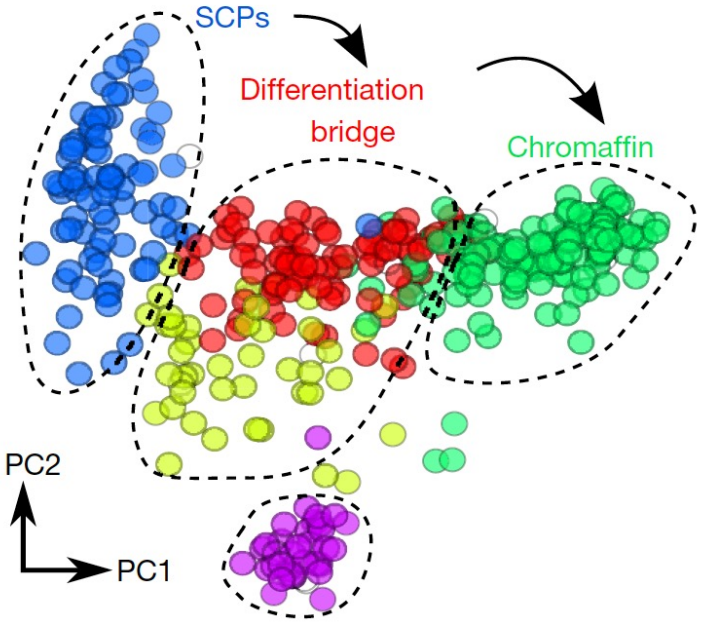
or

We typically need prior biological knowledge (e.g. stem cell…) to provide a root to the inferred trajectory.

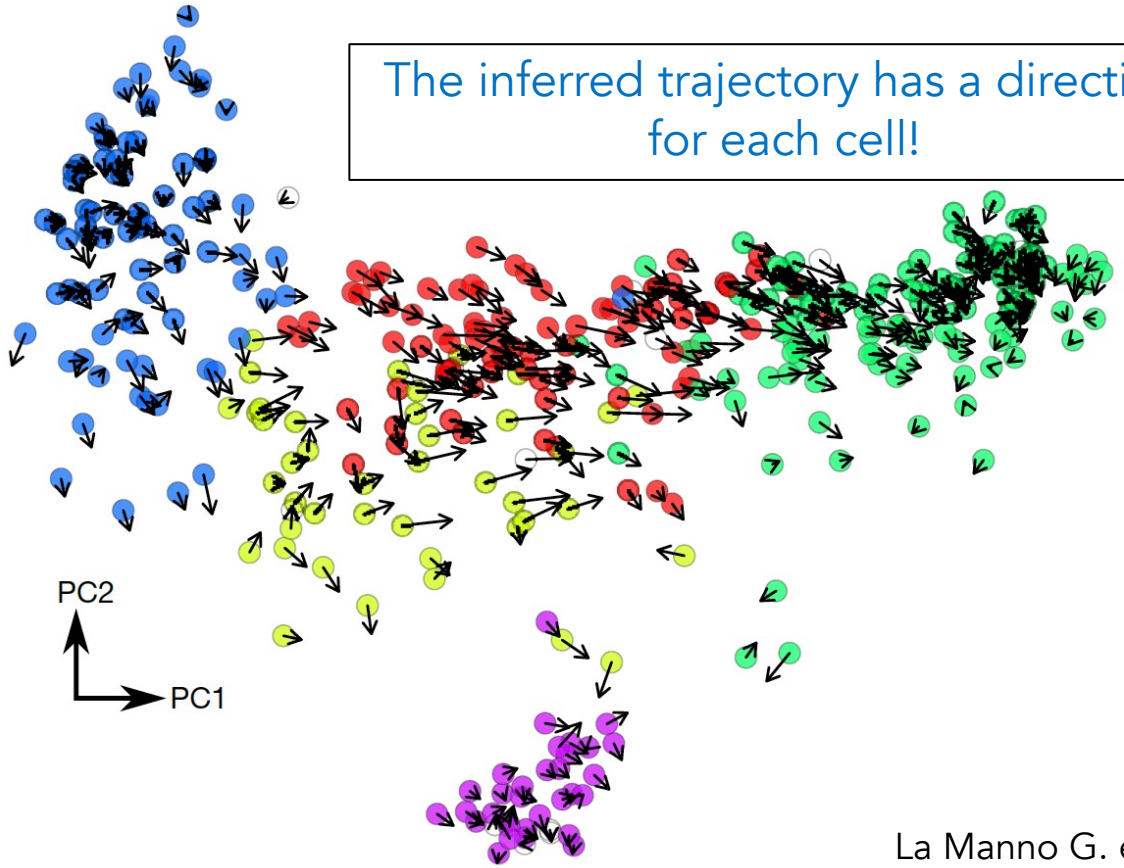
RNA velocity: predicting the future state of individual cells

RNA velocity

- Time derivative of the gene expression state.
- High-dimensional vector that predicts the future state of individual cells on a timescale of hours.



Schwann cell precursors (SCPs) differentiating into chromaffin cells in embryonic day (E)12.5 mice

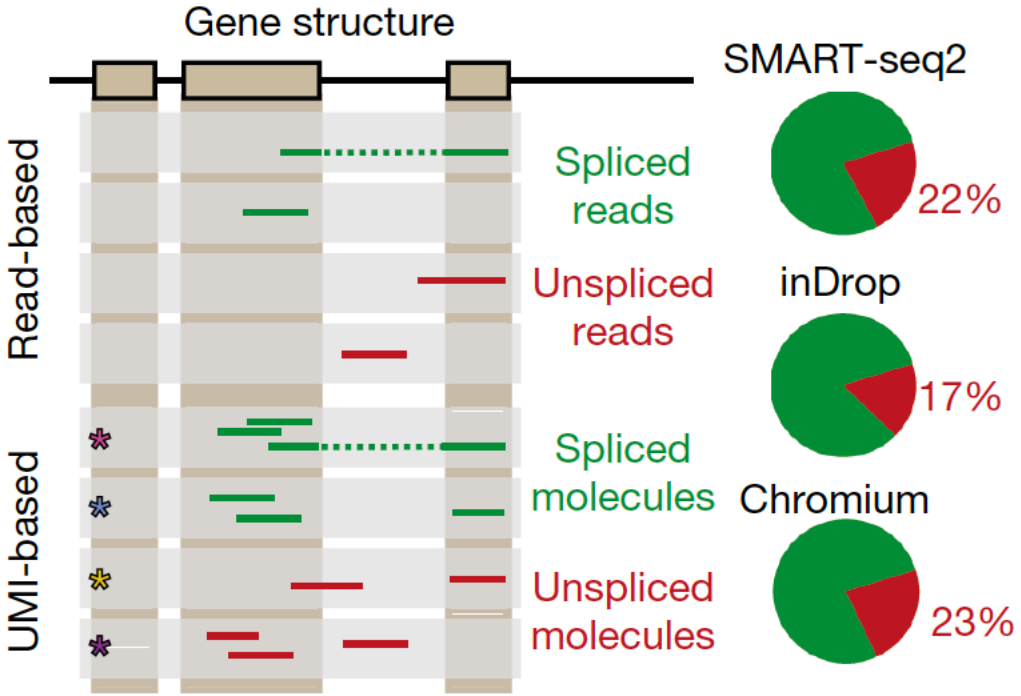


La Manno G. et al. (2018) *Nature*.

RNA velocity: predicting the future state of individual cells

RNA velocity

- Time derivative of the gene expression state.
- High-dimensional vector that predicts the future state of individual cells on a timescale of hours.
- Directly estimated by distinguishing between unspliced and spliced mRNAs in scRNA-Seq sequencing reads.

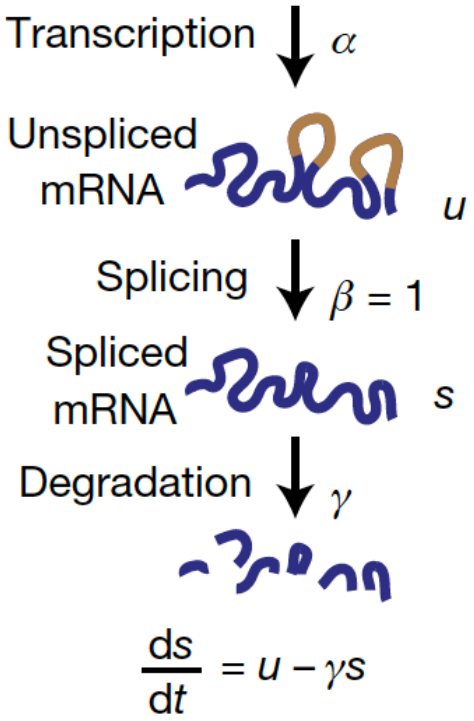


La Manno G. et al. (2018) *Nature*.

RNA velocity: modeling transcriptional dynamics

The time derivative of the spliced mRNA abundance (RNA velocity).

- Balance between production of spliced mRNA from unspliced mRNA and mRNA degradation.

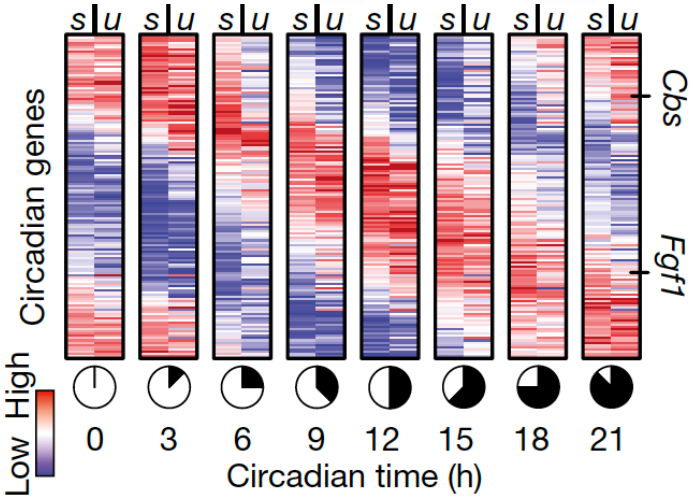


Central assumption is that transcription is at steady state (violated in many cases, e.g. differentiation).

La Manno G. et al. (2018) *Nature*.

RNA velocity: example with circadian-associated genes

- Circadian-associated genes in mouse liver measured over 24 hours.
- *Fgf1* and *Cbs* are two circadian-driven genes.



Bulk RNA-Seq of circadian cycle in mouse liver

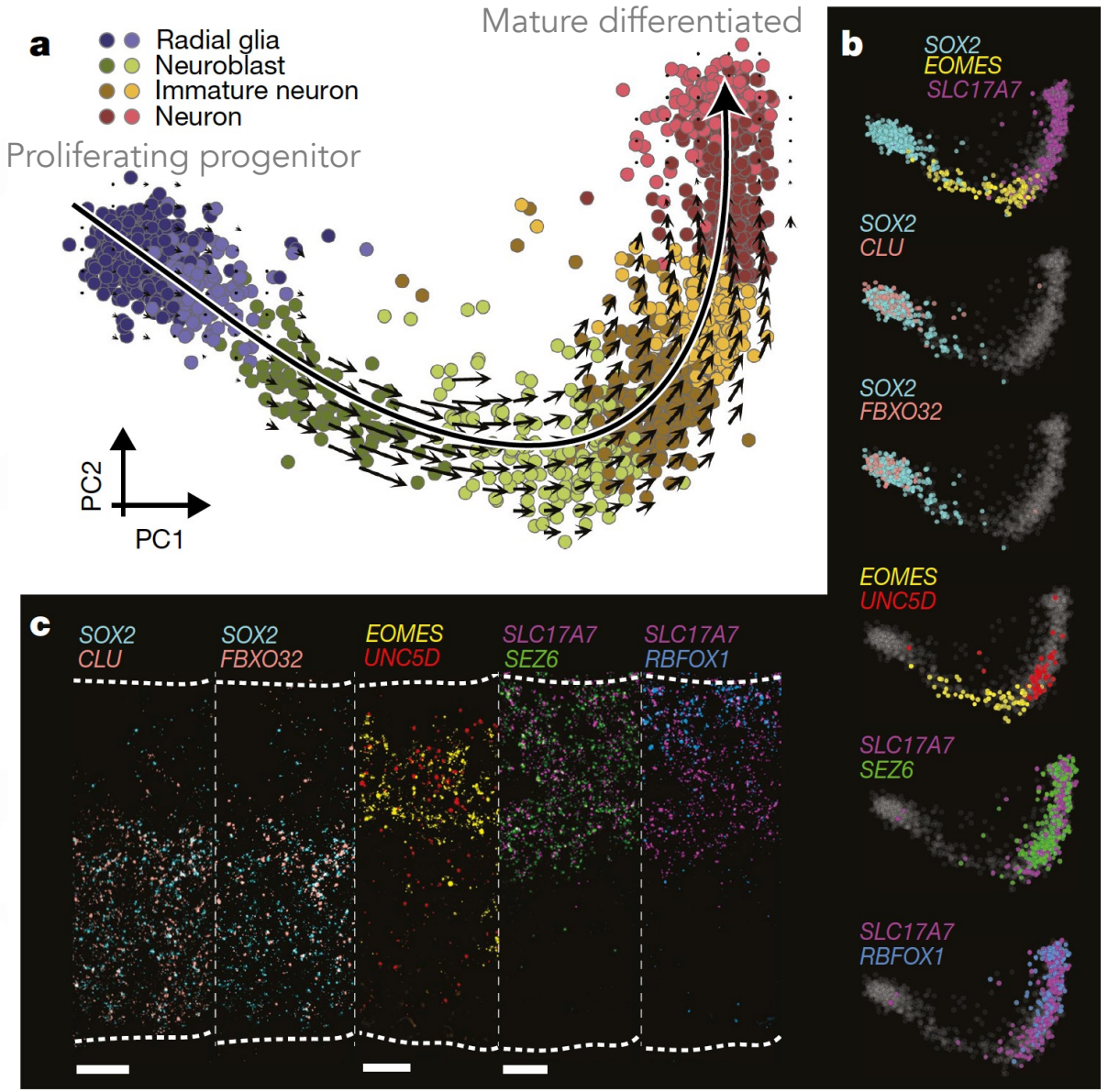
Transcriptional dynamics of individual genes

Future expression state predicted by RNA velocity

La Manno G. et al. (2018) *Nature*.

RNA velocity: example with human embryonic glutamergic neurogenesis (10 weeks p.c.)

RNA velocity and pseudotime from fitting a principal curve



Gene expression of known markers

RNAscope (fluorescence in situ hybridization)

La Manno G. et al. (2018) *Nature*.

Important follow-up studies in trajectory inference and pseudotime

Follow-up studies to RNA velocity:

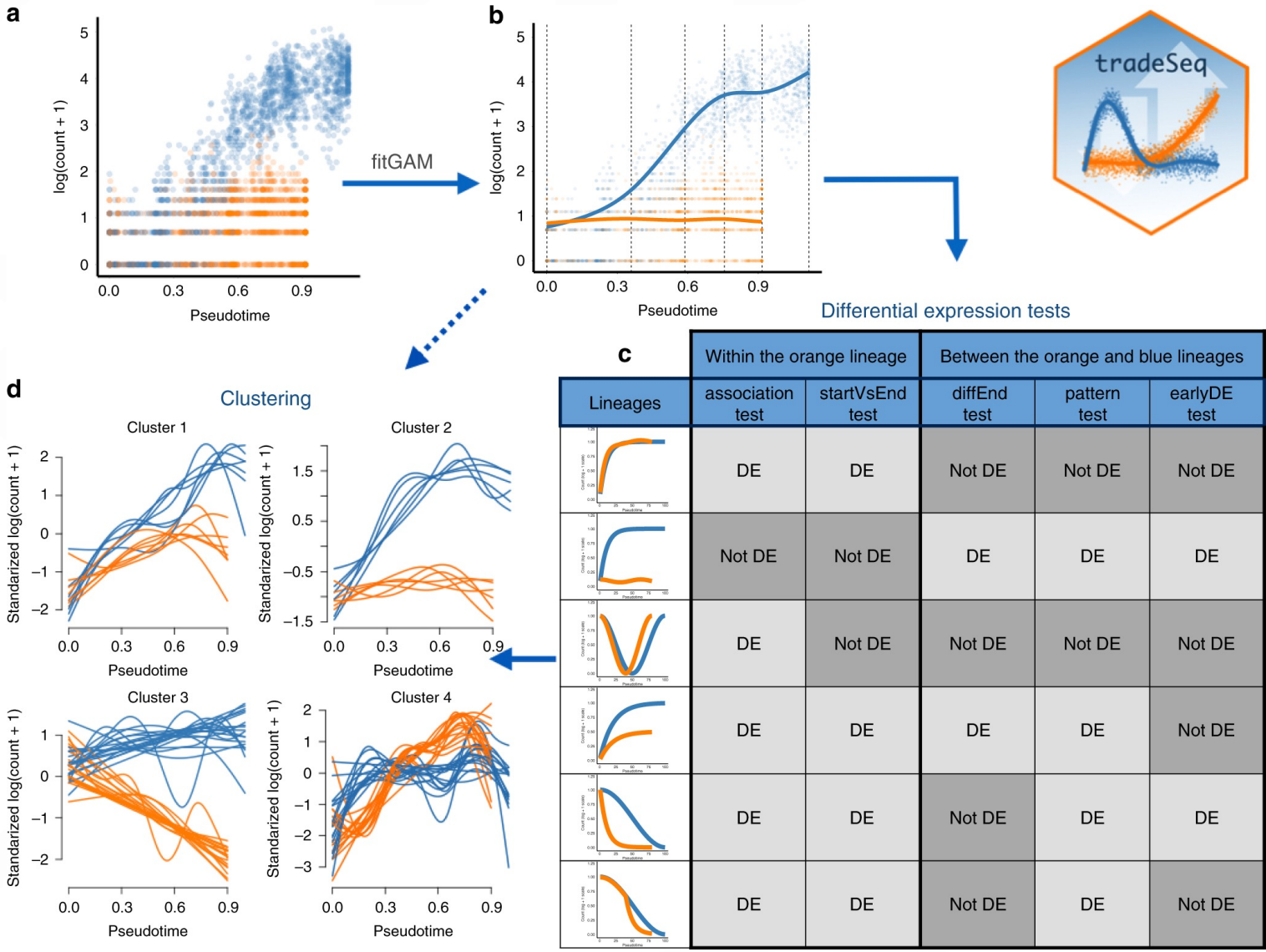
- Bergen, V. et al. "Generalizing RNA velocity to transient cell states through dynamical modeling." *Nature biotechnology* (2020).
- Li, C. et al. "Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction." *Nature biotechnology* (2022).
- Qiu, X. et al. "Mapping transcriptomic vector fields of single cells." *Cell* (2022).
- Lange, M. et al. "CellRank for directed single-cell fate mapping." *Nature Methods* (2022).
- Gayoso, A. et al. "Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells." *Nature biotechnology* (2023).
- Gorin, G. et al. "RNA velocity unraveled." *PLoS Computational Biology* (2022).

Follow-up studies to Slingshot:

- Cao, J. et al. "The single-cell transcriptional landscape of mammalian organogenesis." *Nature* (2019).

Discovering genes associated with the lineages in the trajectory, or differentially expressed between lineages

- For each gene, plot a cell's position in pseudotime (x-axis) vs the expression of the gene in the cell.
- Model time-dependent expression by fitting a negative binomial generalized additive model (GAM) to gene expression~pseudotime.
- DE gene analysis **within** a lineage and **between** lineages.



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, K. et al. (2020) *Nature Communications*.