Pseudotime Trajectory Inference

Single Cell Transcriptomics in Python

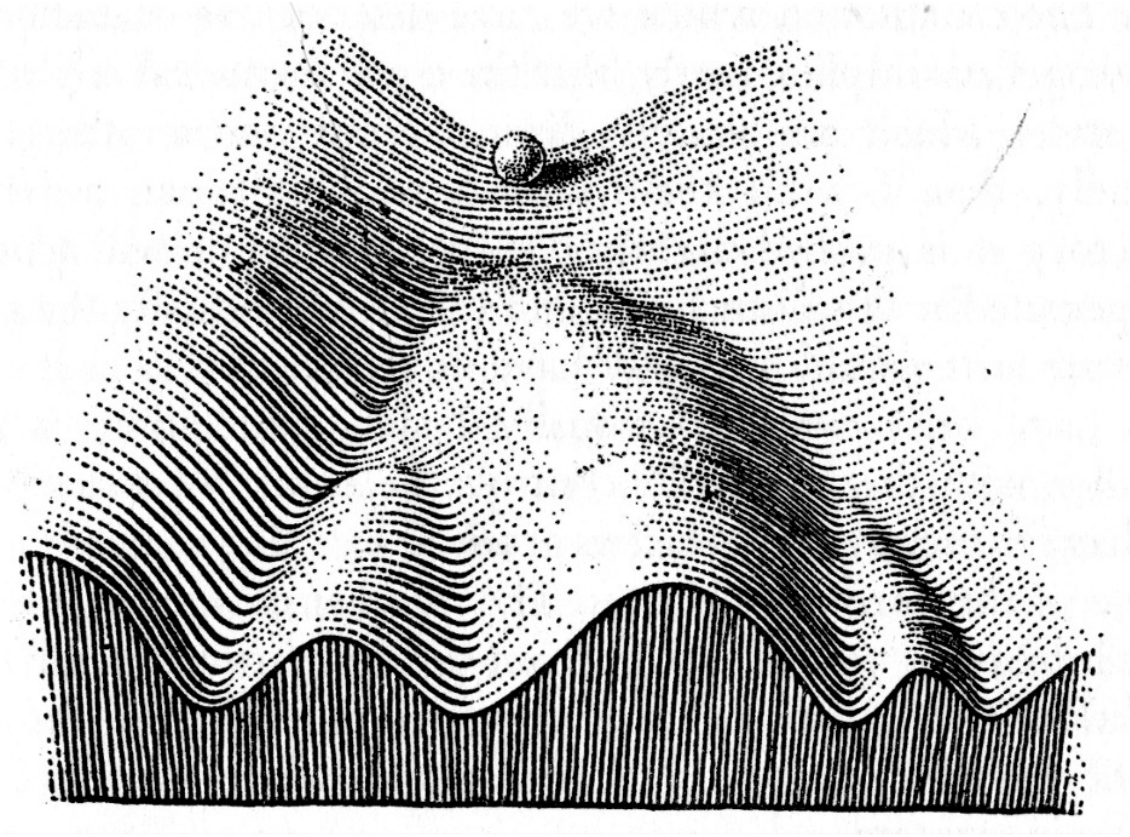
Alex Lederer

What is a cell state transition?

… the process by which cells change in the multi-dimensional feature space over time

What is a cell state transition?

… the process by which cells change in the multi-dimensional feature space over time



“Part of an epigenetic landscape”

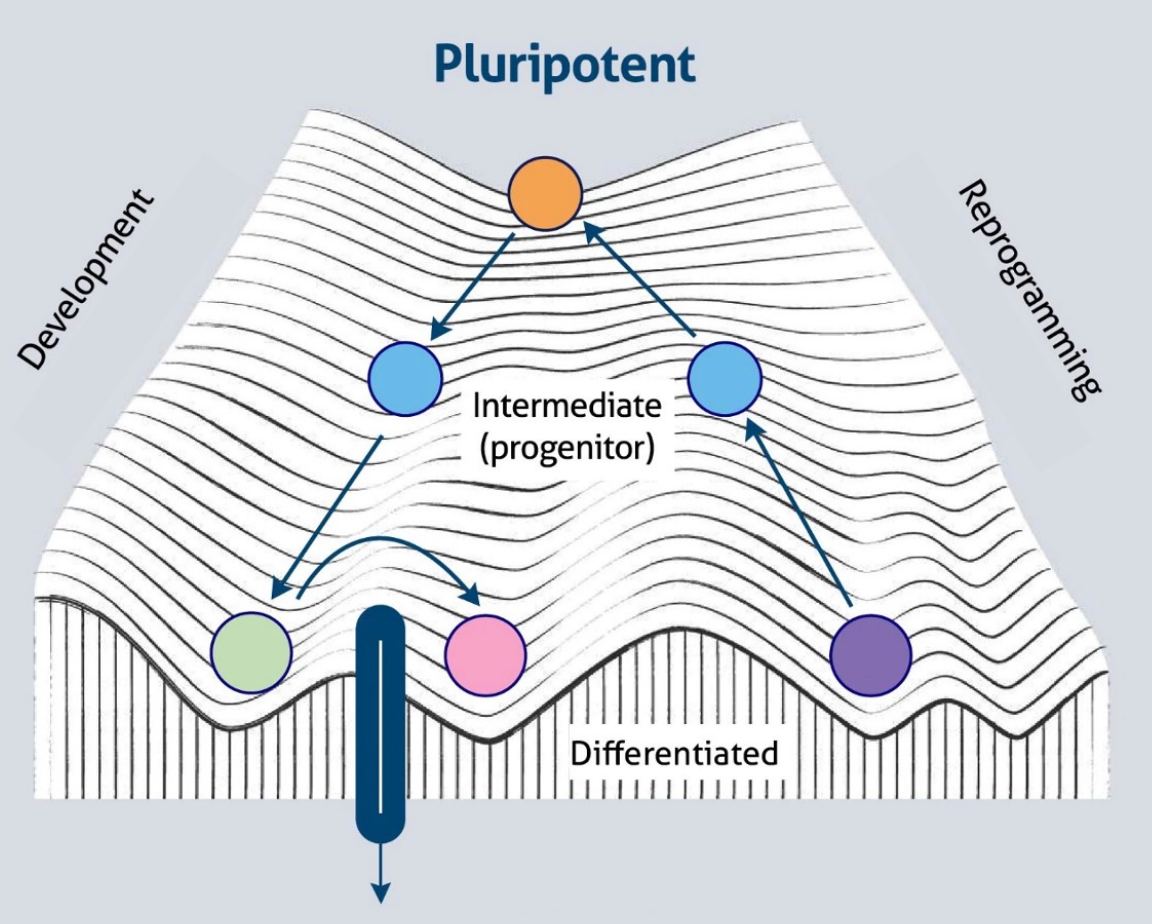
“A system containing many components can be represented by a point in multidimensional space. [...]

In the study of development, we are interested not only in the final state to which the system arrives, but also the course by which it gets there [...]

Conrad H. Waddington
The strategy of genes (1952)

What is a cell state transition?

… the process by which cells change in the multi-dimensional feature space over time



From: Proteintech (https://www.ptglab.com/news/blog/cell-fate-commitment-and-the-waddington-landscape-model/)

"A system containing many components can be represented by a point in multidimensional space. [...]

In the study of development, we are interested not only in the final state to which the system arrives, but also the course by which it gets there [...]

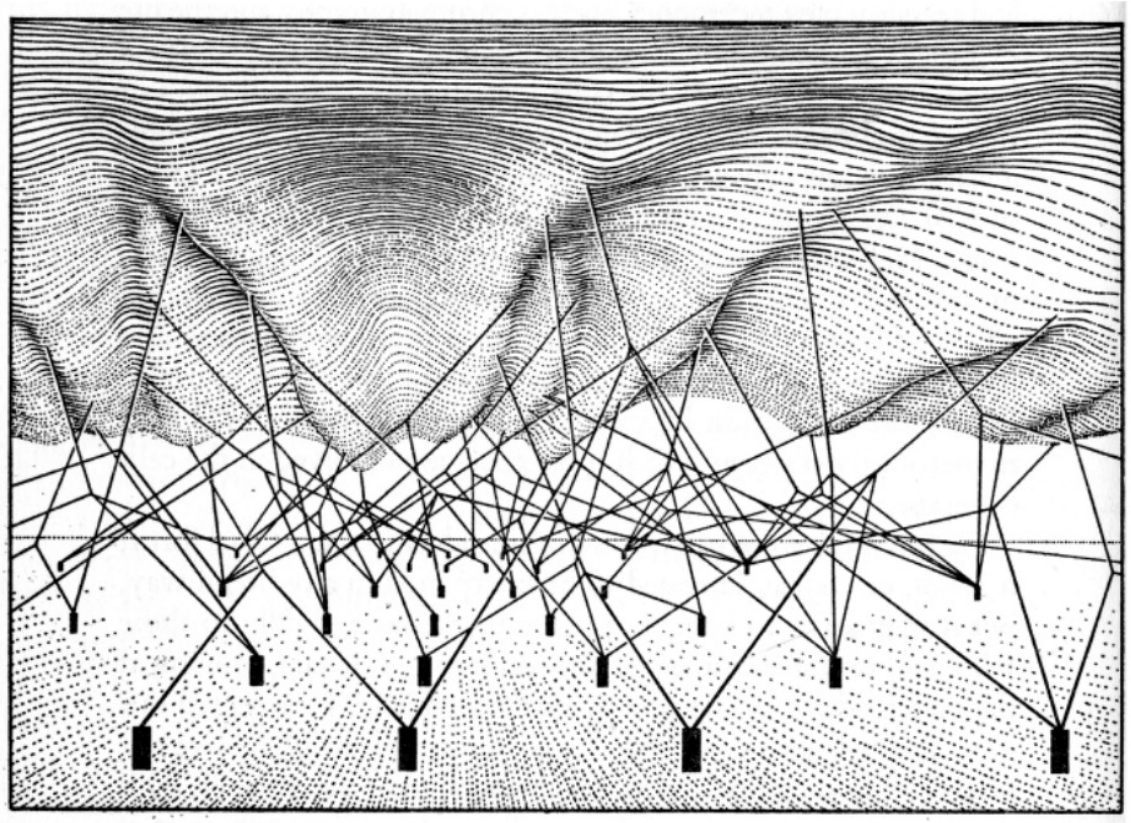
Conrad H. Waddington
The strategy of genes (1952)

What defines the features of the multi-dimensional space?

… usually, the genes!

What defines the features of the multi-dimensional space?

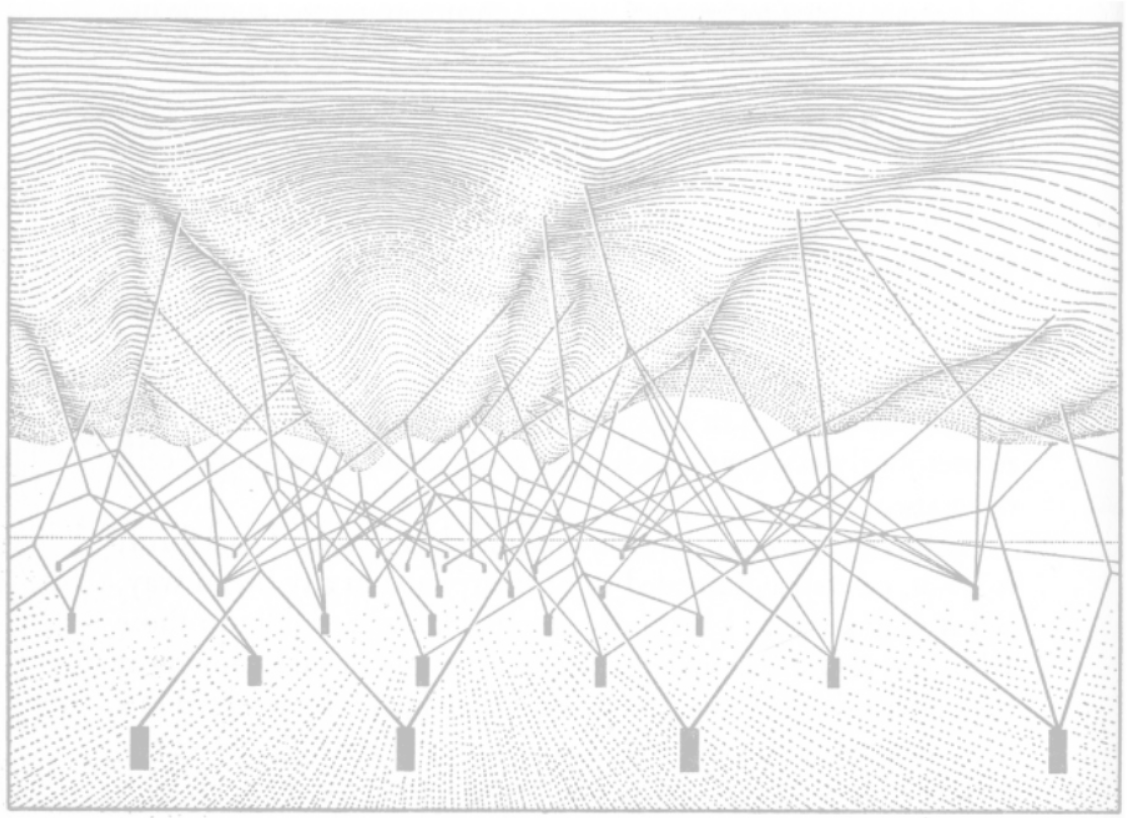
… usually, the genes!



“The complex system of interactions underlying the epigenetic landscape”

Changes in gene abundances define the trajectories taken by cells during cell state transitions

What defines the features of the multi-dimensional space?



"The complex system of interactions underlying the epigenetic landscape"

We need technologies to quantify the abundance of molecular features

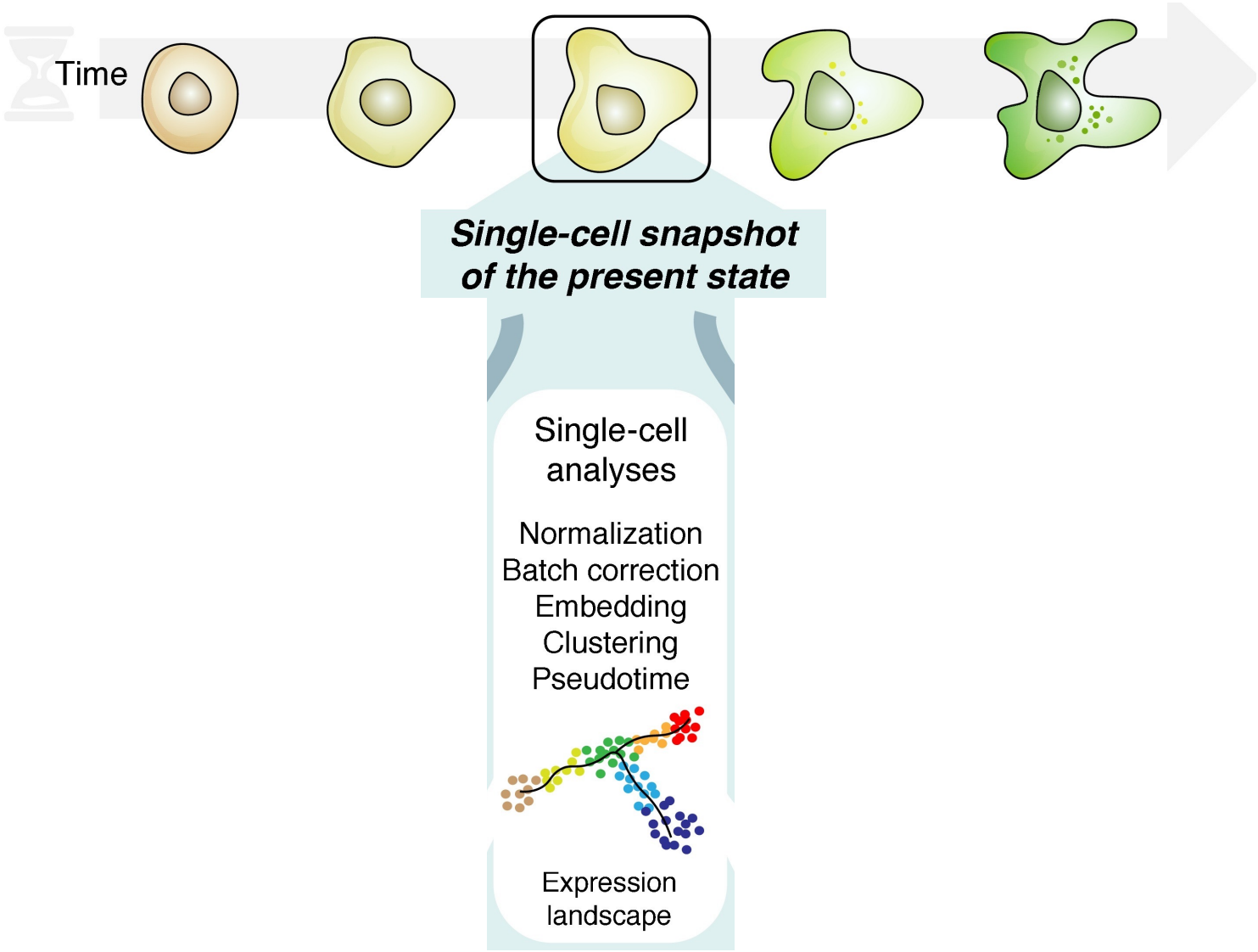
Single cell RNA sequencing!

10x GENOMICS®

Single-cell RNA sequencing is a **destructive** technology

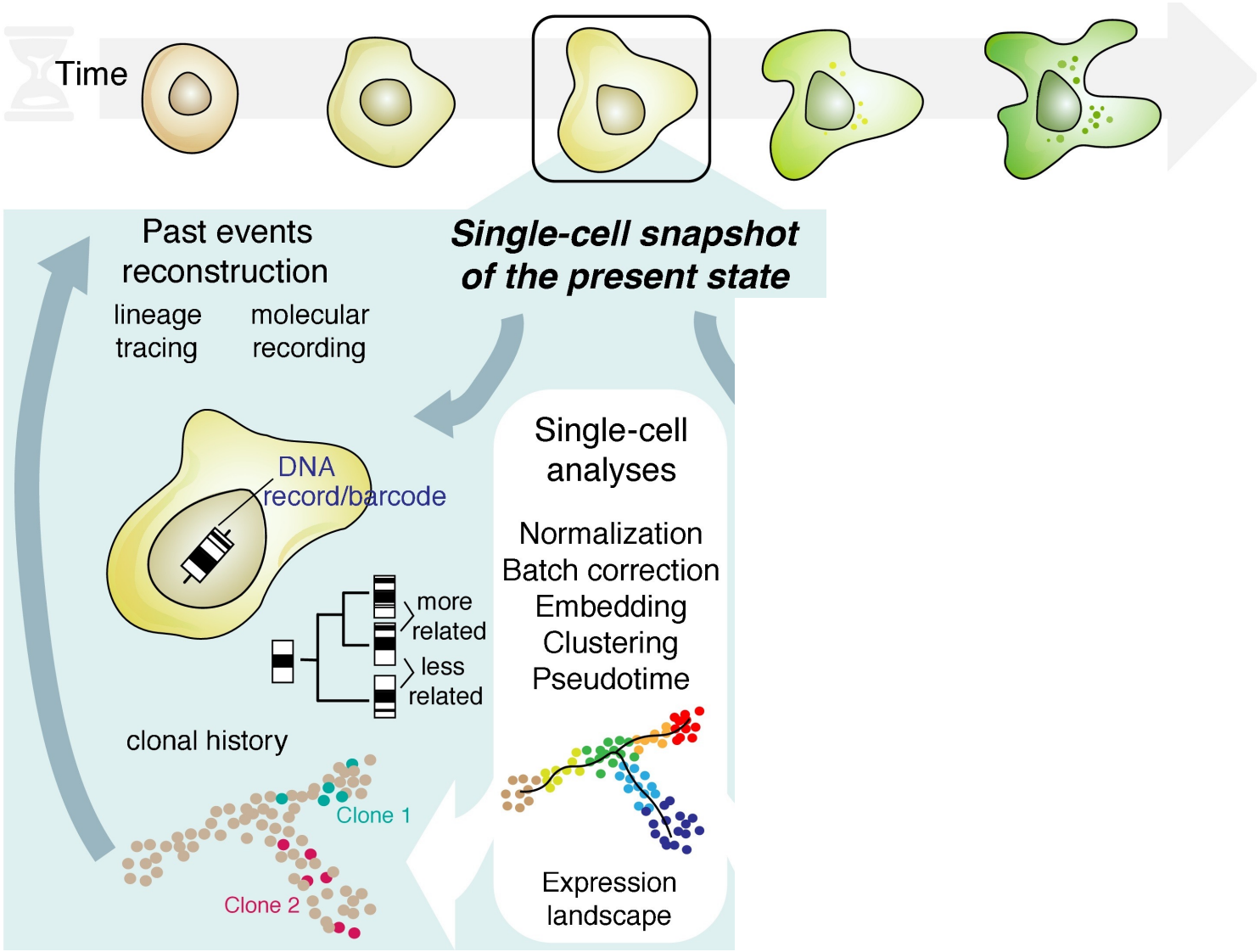
A cell can be profiled only one time,
providing a static **snapshot**

Emergence of single-cell temporal-omics approaches



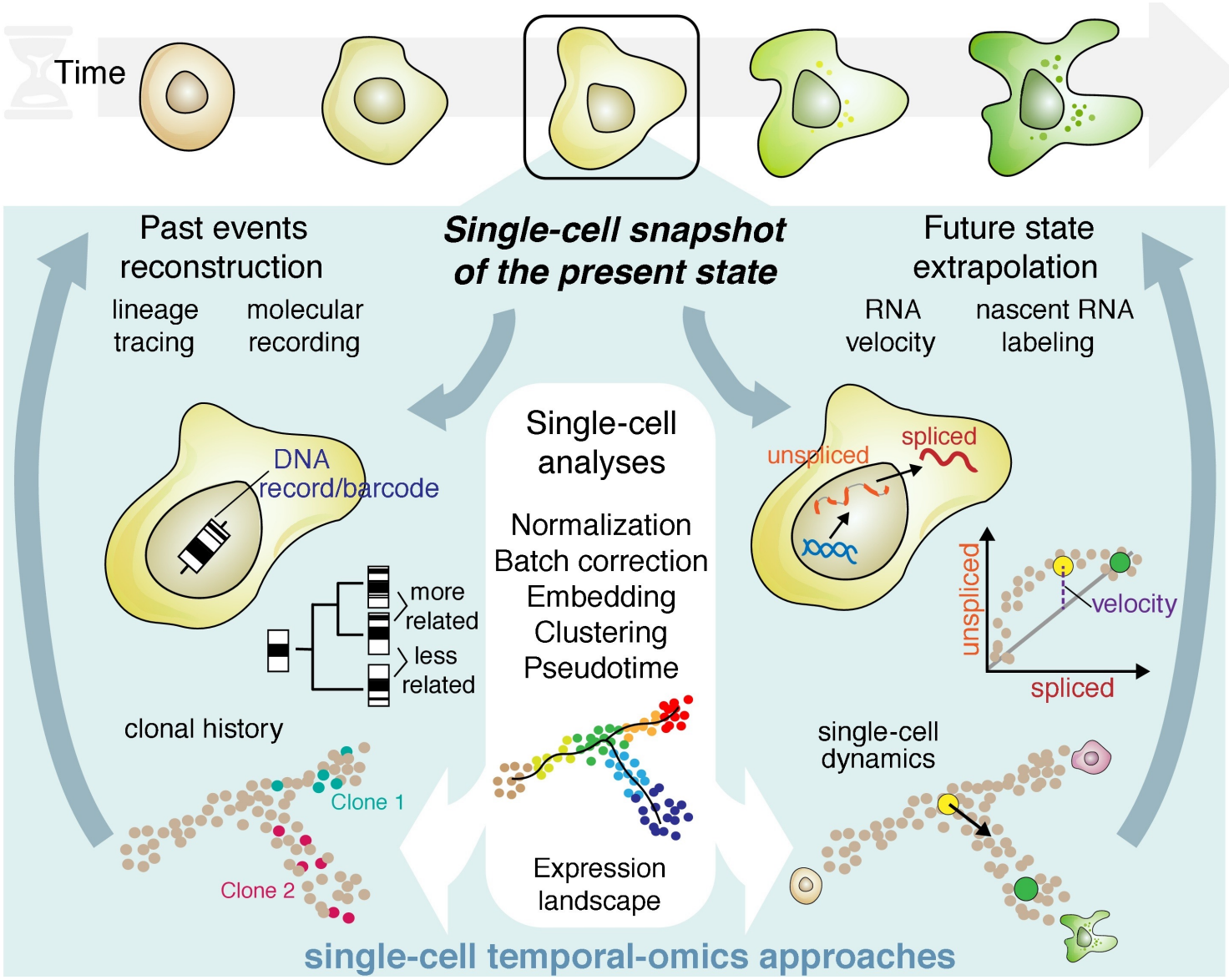
Lederer and La Manno 2020

Emergence of single-cell temporal-omics approaches



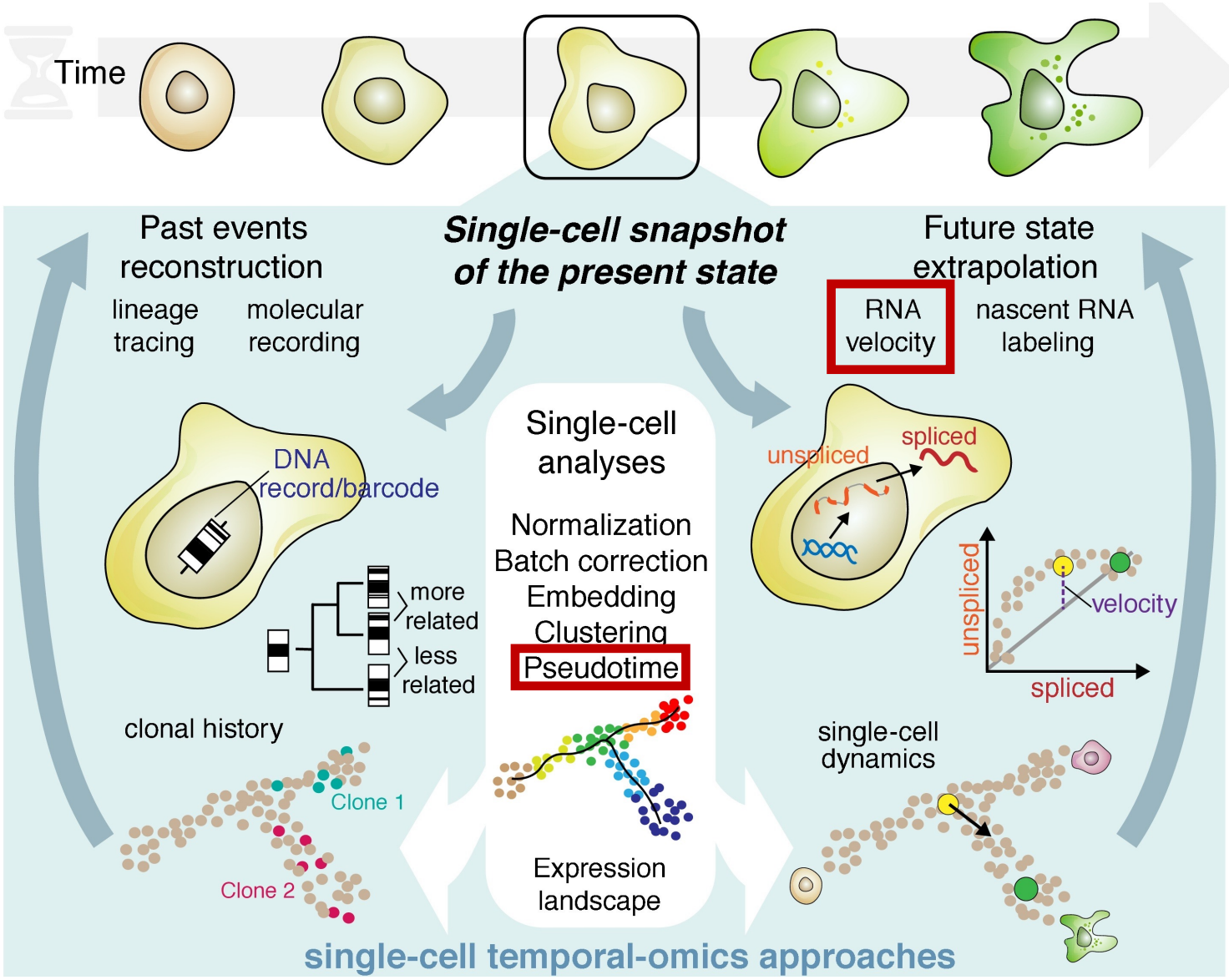
Lederer and La Manno 2020

Emergence of single-cell temporal-omics approaches



Lederer and La Manno 2020

Emergence of single-cell temporal-omics approaches



Lederer and La Manno 2020

Pseudotime trajectory inference

Pseudotime trajectory inference

- Differences in gene expression between cells might be attributed to **dynamic** processes:
  - Cell cycle
  - Development or differentiation
  - Response to a stimuli (environmental change, drug treatment)

- **Trajectory inference** orders a population of individual cells along a path or lineage

Pseudotime trajectory inference

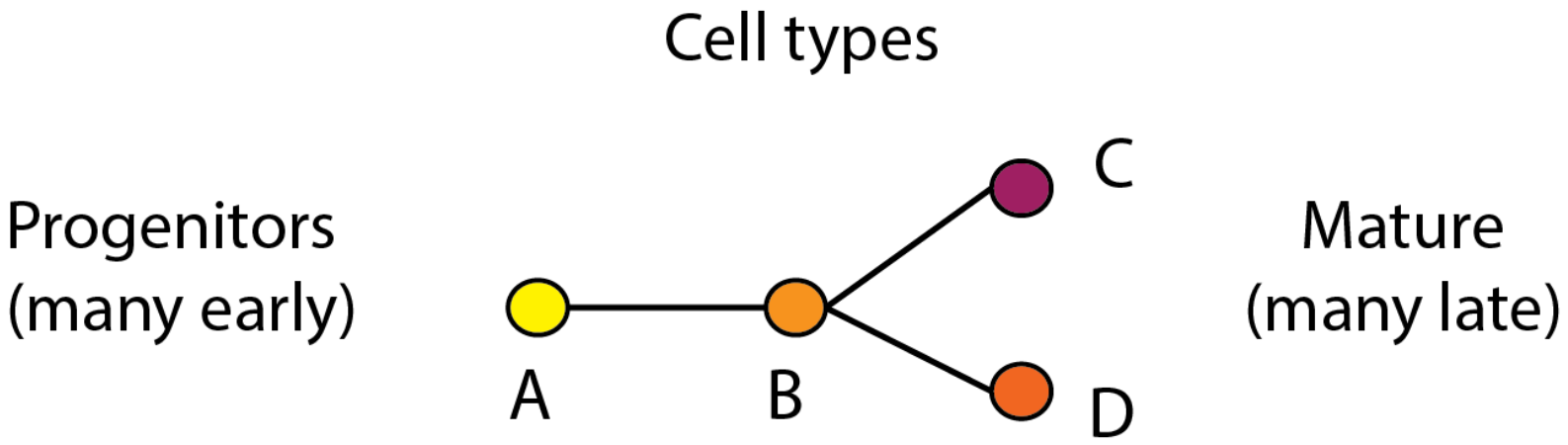
- Differences in gene expression between cells might be attributed to **dynamic** processes:
  - Cell cycle
  - Development or differentiation
  - Response to a stimuli (environmental change, drug treatment)

- **Trajectory inference** orders a population of individual cells along a path or lineage

- Cells on the learned path can be assigned a "**pseudotime**", which is a measure of their amount progress along the path.

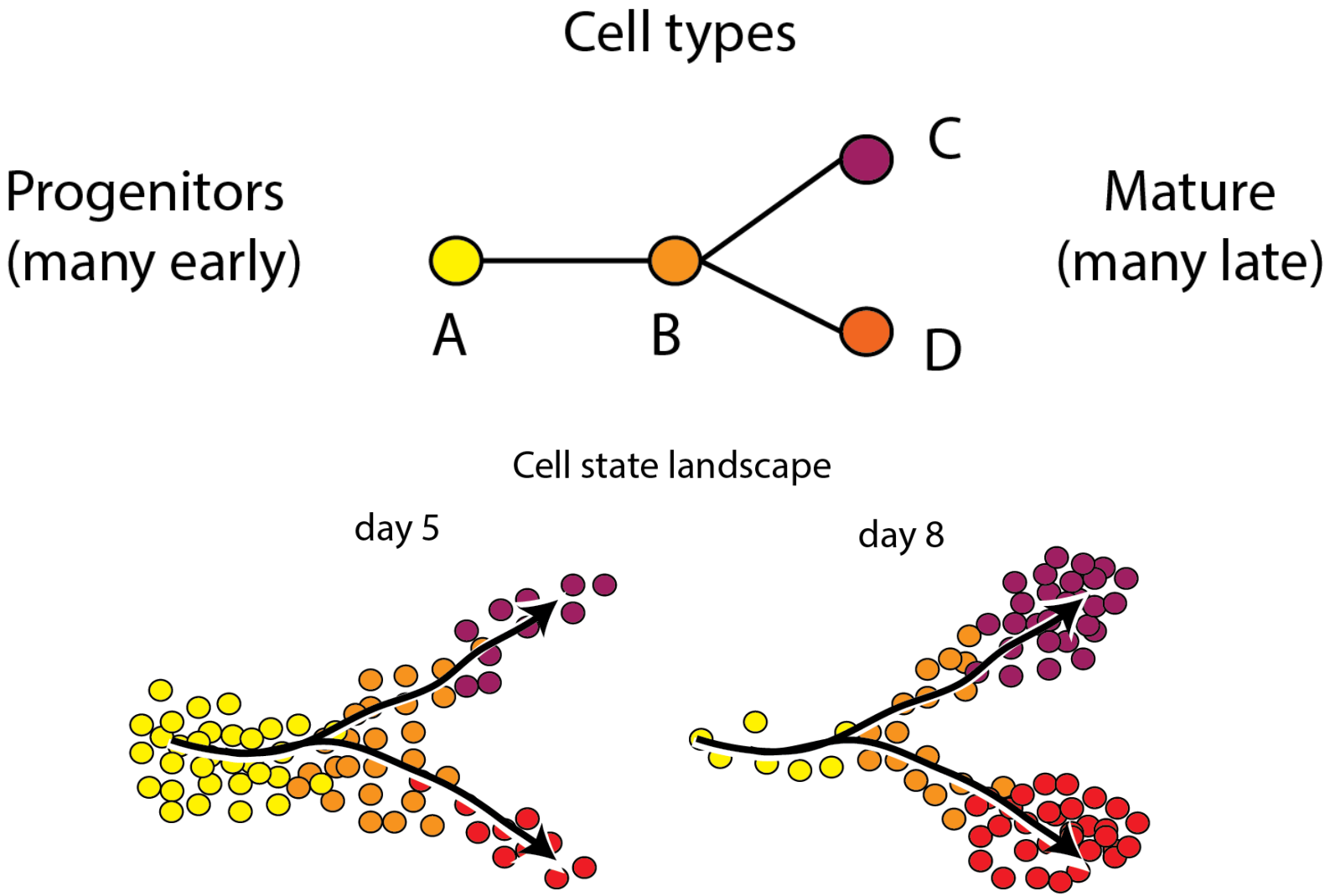
- Can be a good starting point for further analysis:
  - Determine gene expression programs driving changes in conditions that drive cells towards more or less differentiated states or phenotypes (i.e., number of cells in the beginning vs end of of the pseudotime axis).

Pseudotime trajectory inference

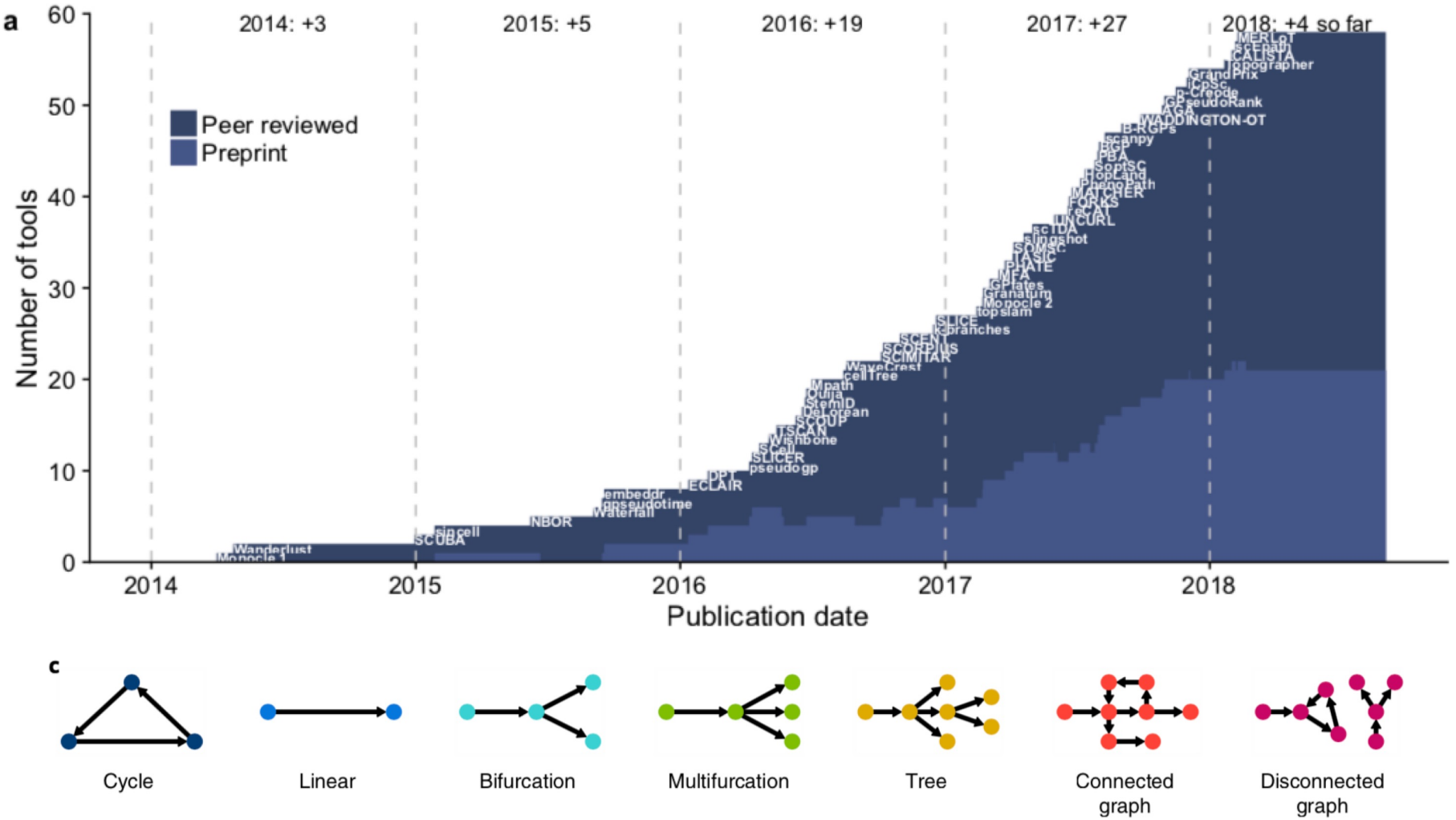


let's say this process takes 7 days to unfold…

Pseudotime trajectory inference



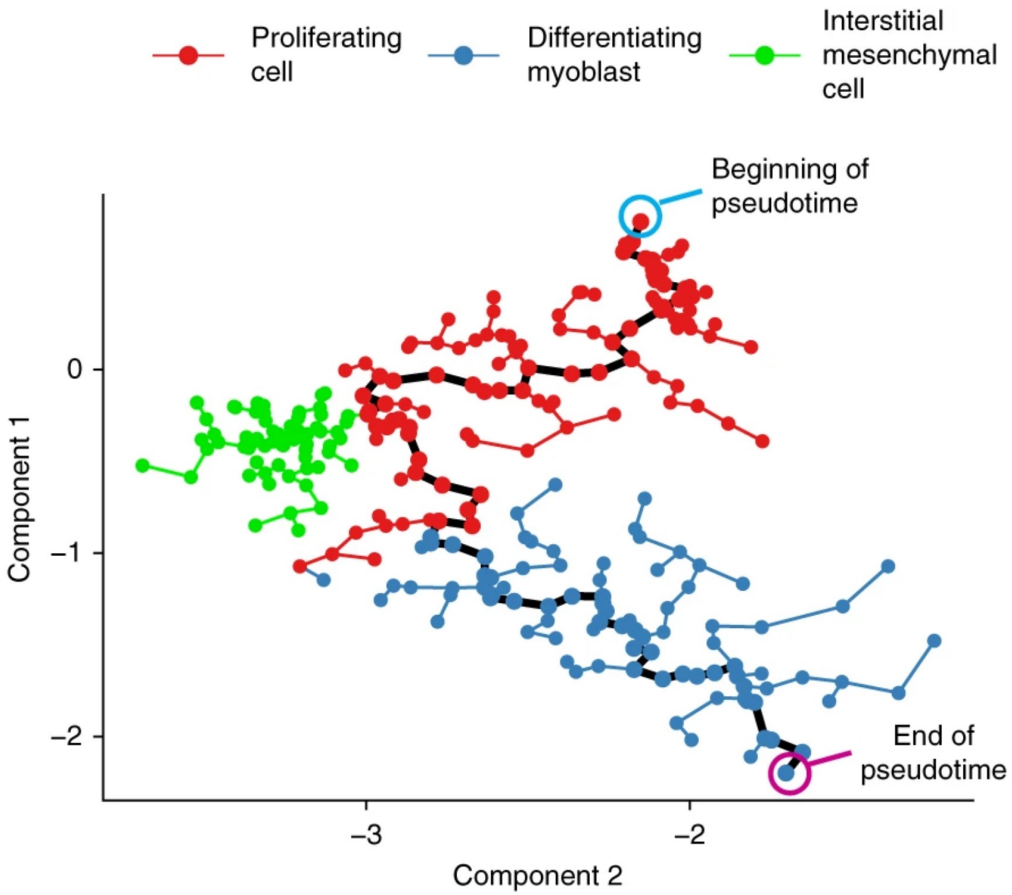
There are many trajectory inference methods to choose from!

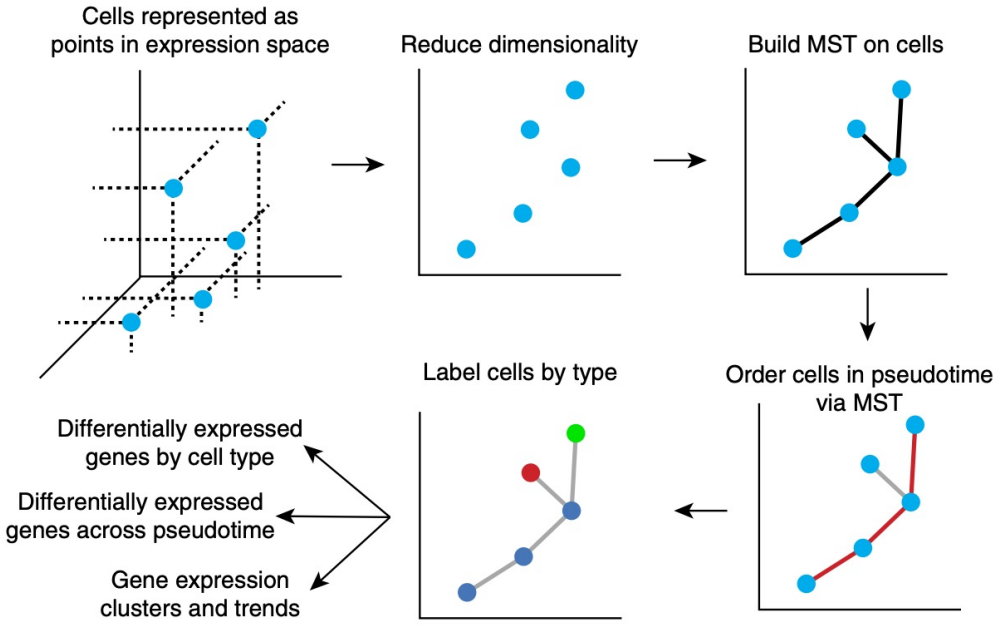


Saelens et al. 2019

The first pseudotime algorithm: Monocle

Monocle

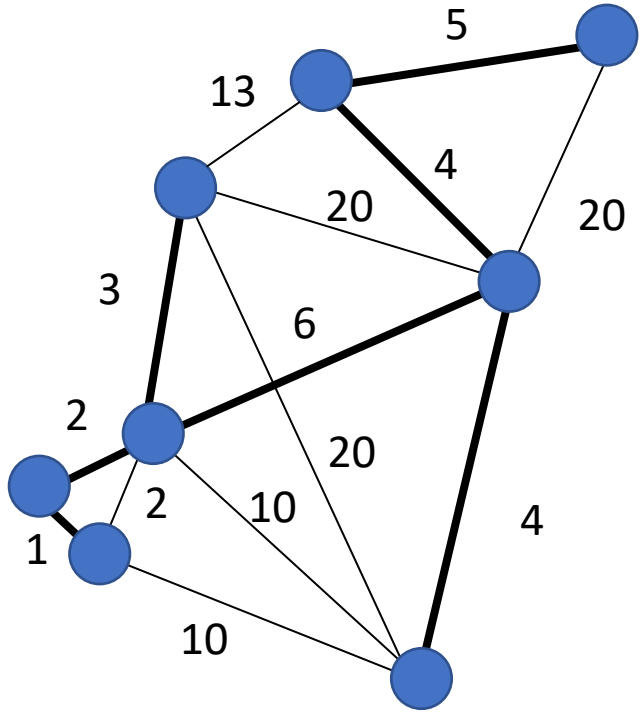




Trapnell et al. 2014

Minimum spanning tree

- Sum of all distances in the tree (graph) among single cells is at its minimum
- Having more intermediate cells improves the definition of the tree
- The weights are usually a distance in the dimensionality reduction space (PCA, UMAP)
- MST has no cycles, cell cycles will not work in here



Dr. Rachel Marcone

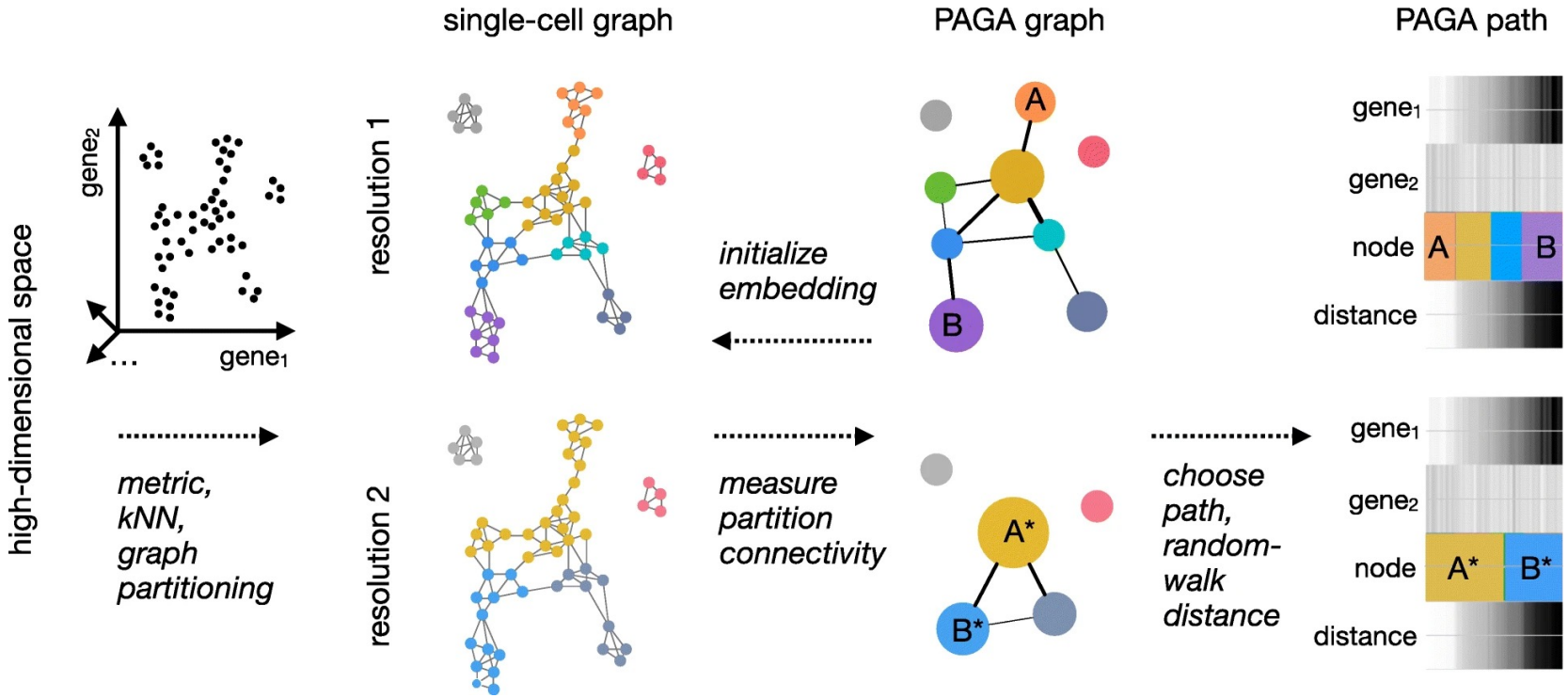
General types of pseudotemporal ordering

1. **Clustering graph-based**: cells are clustered using k-means or Leiden clustering, and then ordered connections between the clusters are constructed based on similarity or a MST (MST, PAGA)
2. **Manifold-learning based**: connections between cells are defined using principal curves, which find a one-dimensional curve connecting cellular observations (Slingshot).
3. **Probabilistic frameworks**: assign transition probabilities to cell-cell pairs (diffusion pseudotime)

Comparison of trajectory inference methods: https://www.nature.com/articles/s41587-019-0071-9

PAGA: Partition-based graph abstraction

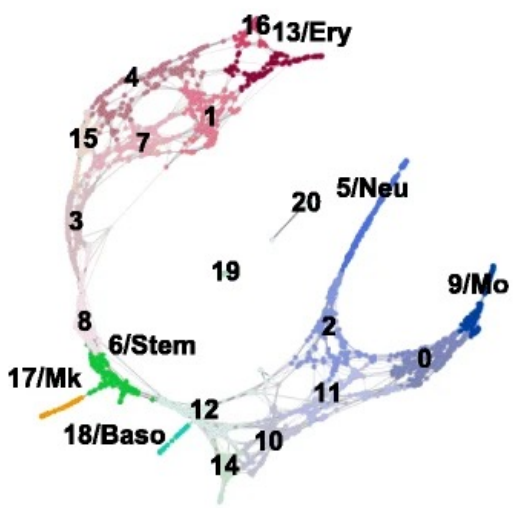
- A graph connecting clusters/partitions (at various resolutions) of single cells is constructed
- Connective measure for each partition: do nodes in a cluster connect more to cells within the cluster, or outside of it?
- Random-walk between cells to identify the most probable path: based on connectivity of different partitions

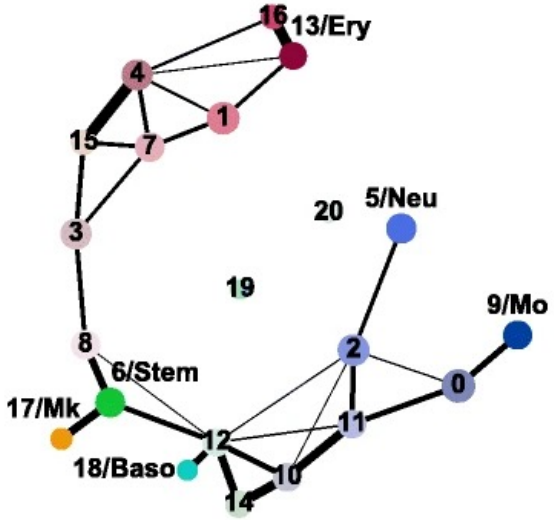


Wolf et al. Genome Biology (2019)
https://doi.org/10.1186/s13059-019-1663-x

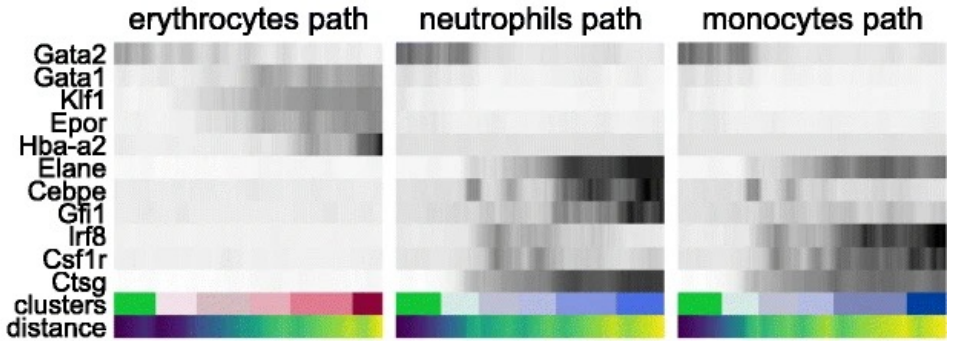
PAGA: Partition-based graph abstraction

- A graph connecting clusters/partitions (at various resolutions) of single cells is constructed
- Connective measure for each partition: do nodes in a cluster connect more to cells within the cluster, or outside of it?
- Random-walk between cells to identify the most probable path: based on connectivity of different partitions





Paul et al. (2015)



Wolf et al. Genome Biology (2019)
https://doi.org/10.1186/s13059-019-1663-x

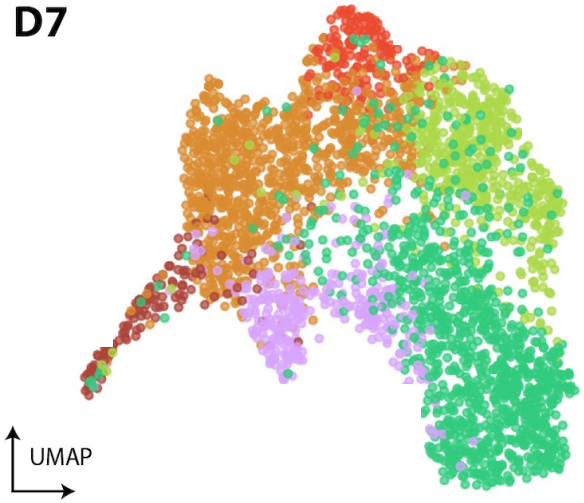
Should you run trajectory inference?

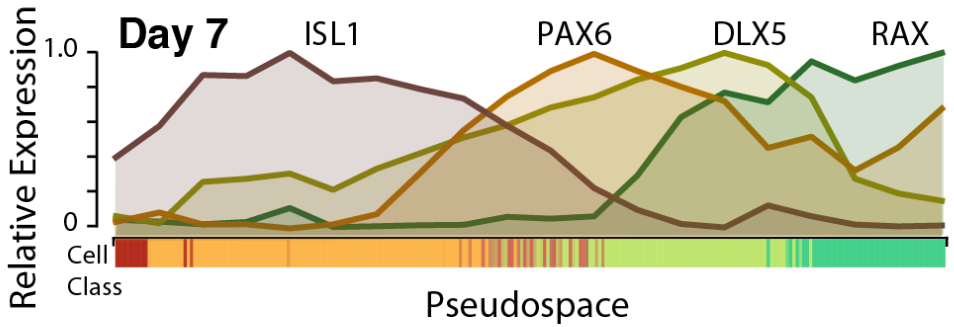
Questions to ask:

- Are you sure that you expect a trajectory?
- Do you have intermediate states?
- Do you think you have branching in your trajectory?
- Do you have a time scale on your cells?
- Do you know your start or end state?

Be aware, any dataset can be forced into a trajectory without any biological meaning!

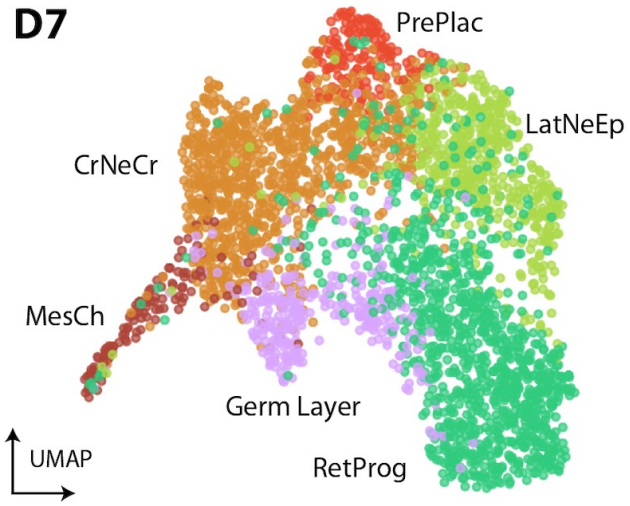
An example where pseudotime can be misinterpreted

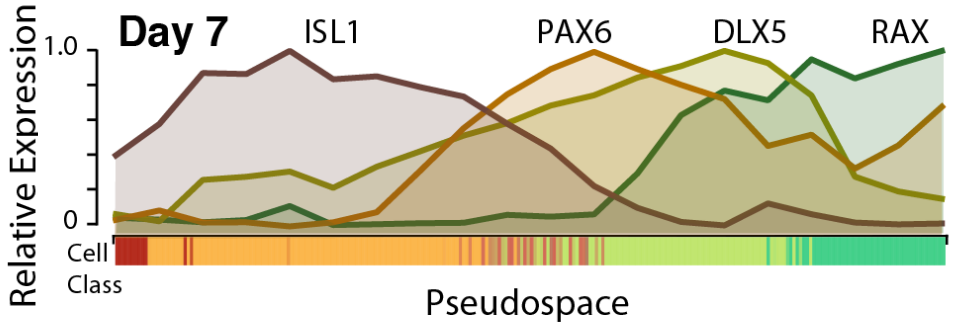


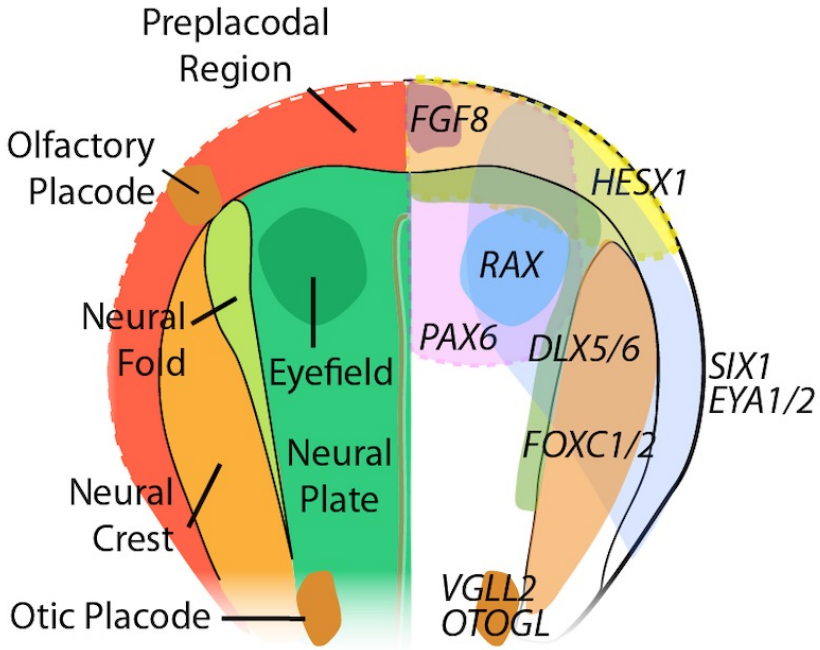


Petrus-Reurer, Lederer et al 2022

An example where pseudotime can be misinterpreted







- Cell types are similar to rostral embryonic tissues
- Spatial patterning (rather than a temporal axis of variation)

Petrus-Reurer, Lederer et al 2022