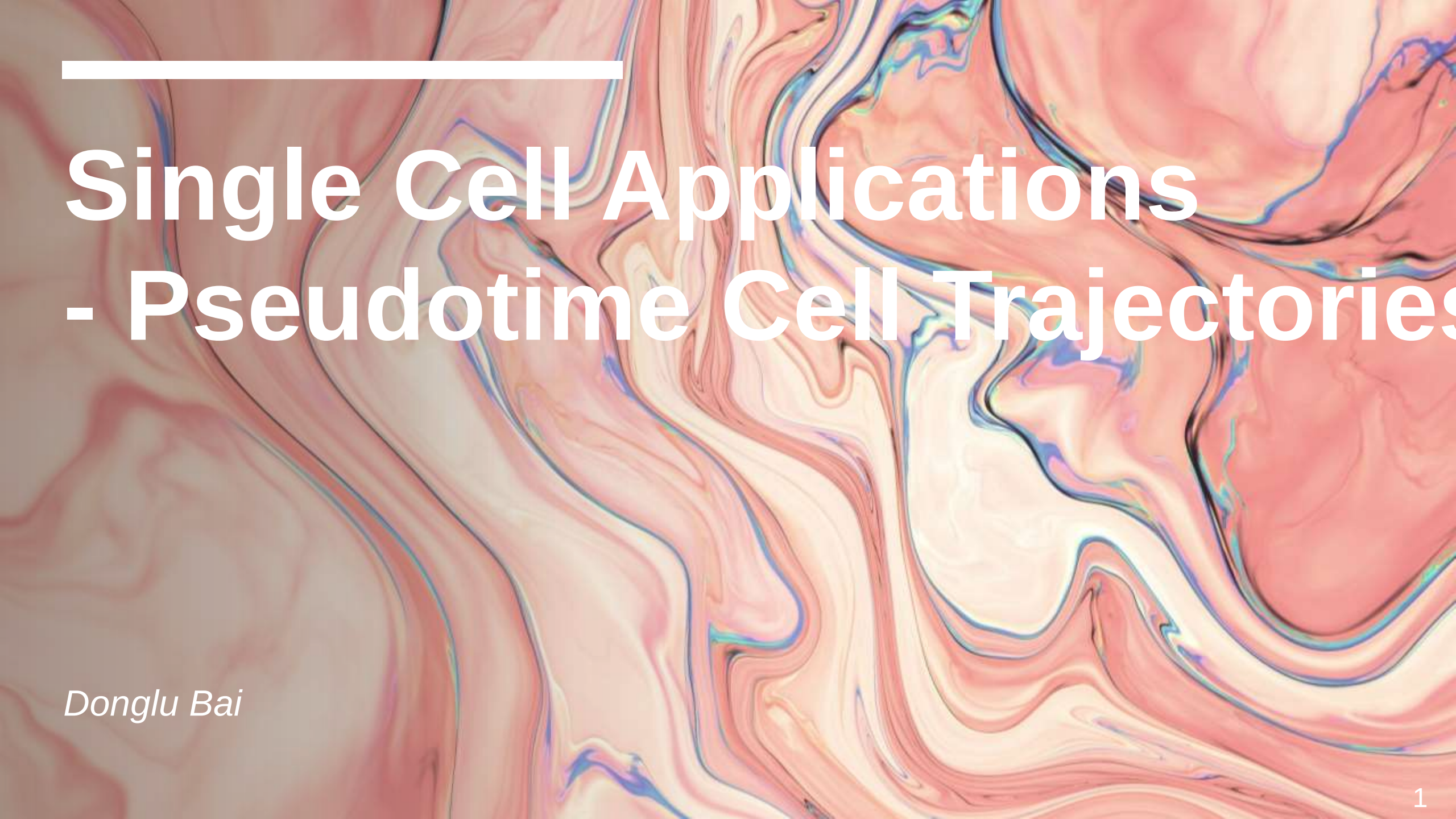


Single Cell Applications - Pseudotime Cell Trajectories

Donglu Bai

Outline

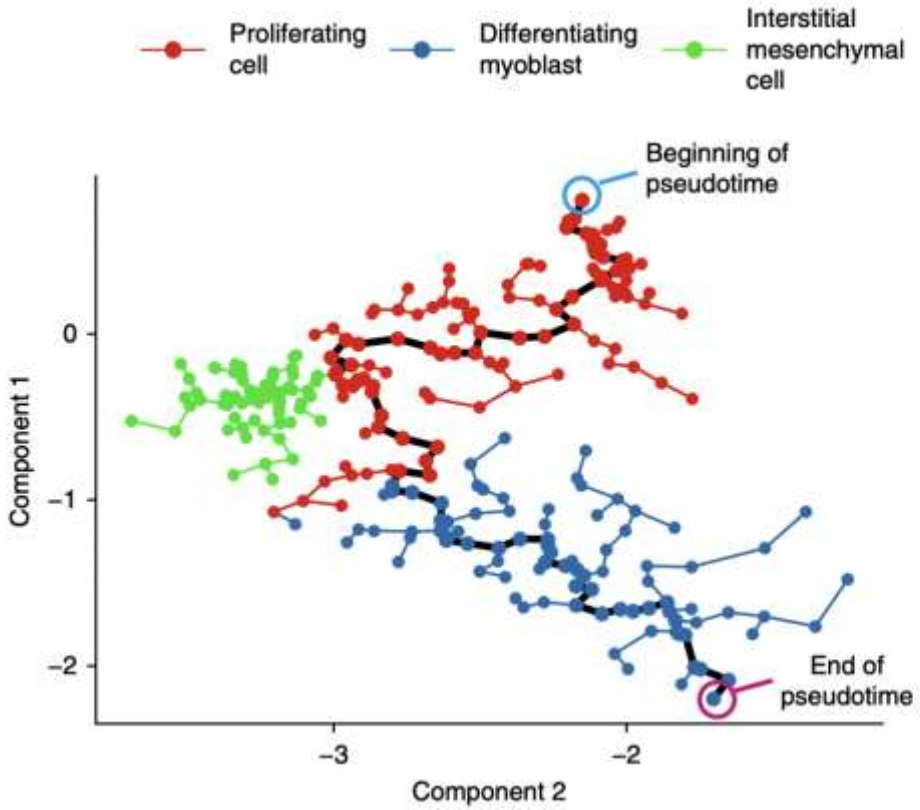
What is pseudotime

Monocle

RNA Velocity

Challenges of Computational Inferences

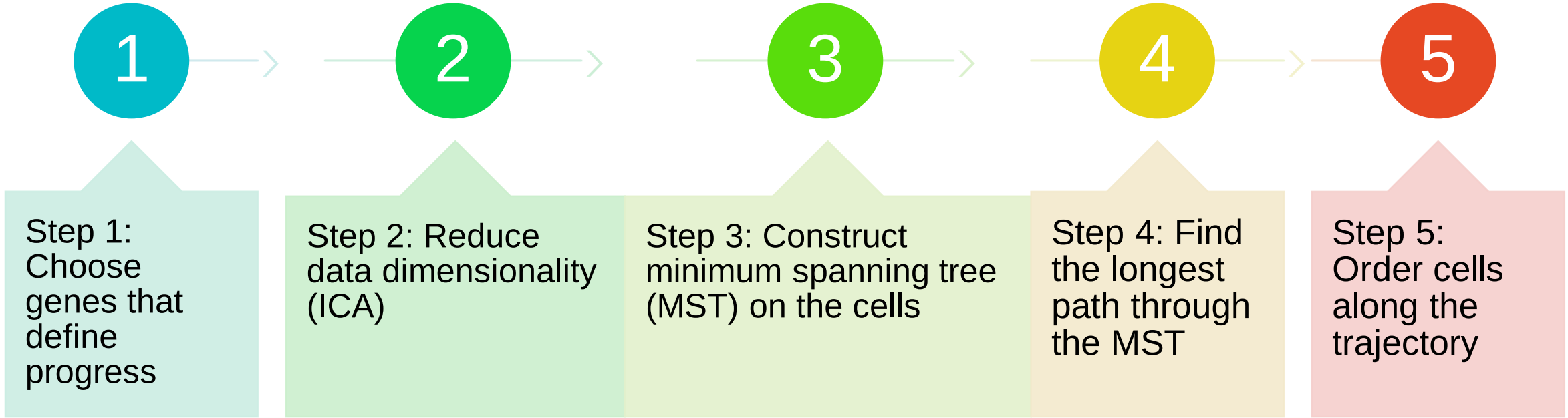
What is Pseudotime/Trajectory



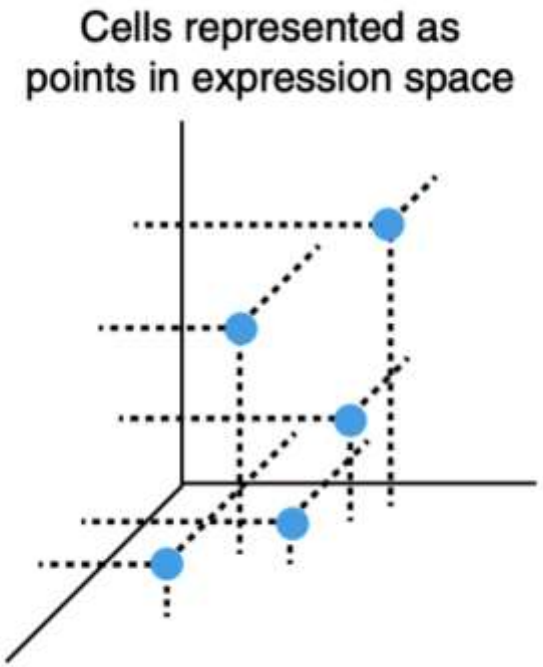
1) Cell differentiation occurs through dynamic developmental processes
2) Pseudotime orders cells along trajectories that represent these process
3) Trajectory inference (TI) reconstructs these cellular transitions

Trapnell, C et al. Nat Biotechnology 2014
Saelens, W et al. Nat Biotechnology 2019

Monocle 2


1
2
3
4
5
Step 1: Choose genes that define progress
Step 2: Reduce data dimensionality (ICA)
Step 3: Construct minimum spanning tree (MST) on the cells
Step 4: Find the longest path through the MST
Step 5: Order cells along the trajectory

Step 1: Identify genes that define the progress



1) Cells exist in a high-dimensional space

- Each cell X_i is a point in d dimensional space

$$x_i = (g_1, g_2, ..., g_d)$$

2) Select high variance genes or based on differential expression patterns

A. High variability across cells

- Compute the variance across cells for each gene

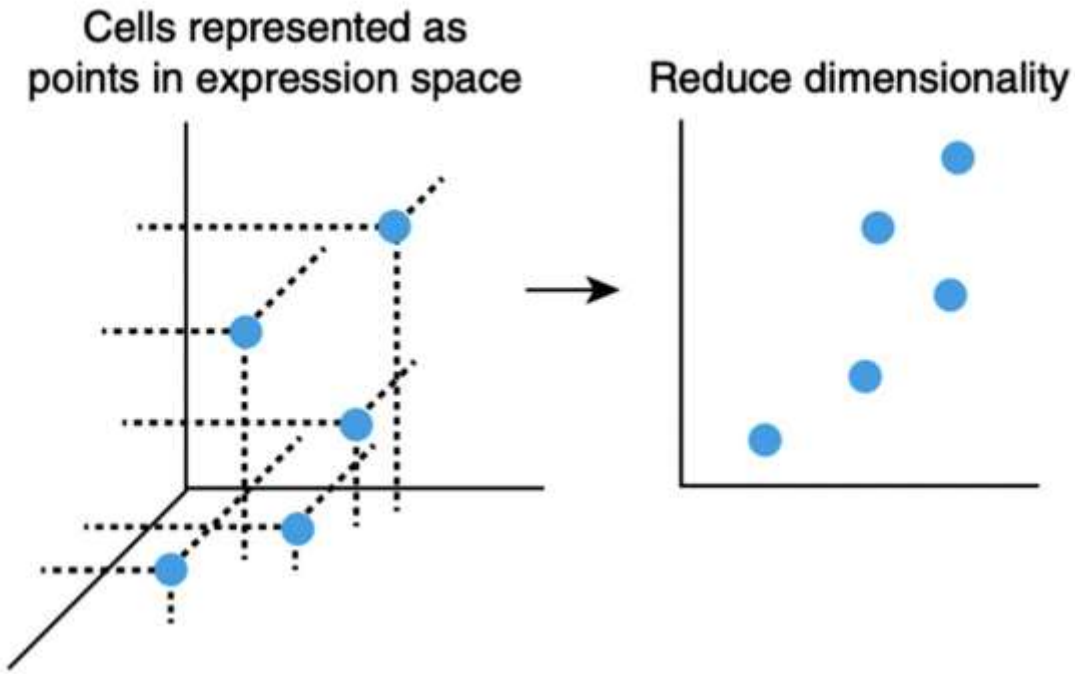
$$\mathrm{Var}(g) = \frac{1}{n}\sum_{i=1}^{n}(X_{i,g} - \bar{X}_g)^2$$

B. Genes that define biologically relevant processes (e.g., stem cell markers)

C. Differer $H_0 : \mu_1 = \mu_2, \quad H_A : \mu_1 \neq \mu_2$ tes

Trapnell, C et al. Nat Biotechnology 2014

Step 2: Reduce data dimensionality (ICA)



1) Mathematical equations:

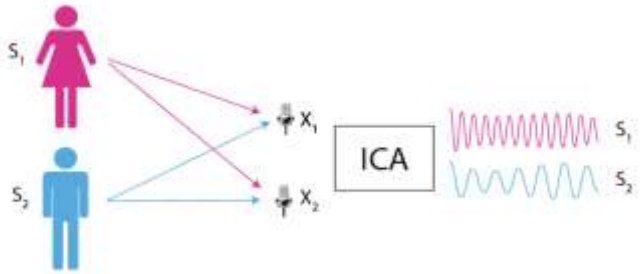
$$x = As$$

- x is the observed mixed signals
- A is the mixing matrix
- s is the original independent source signals

2) Goals:

$$y = Wx = WAs$$

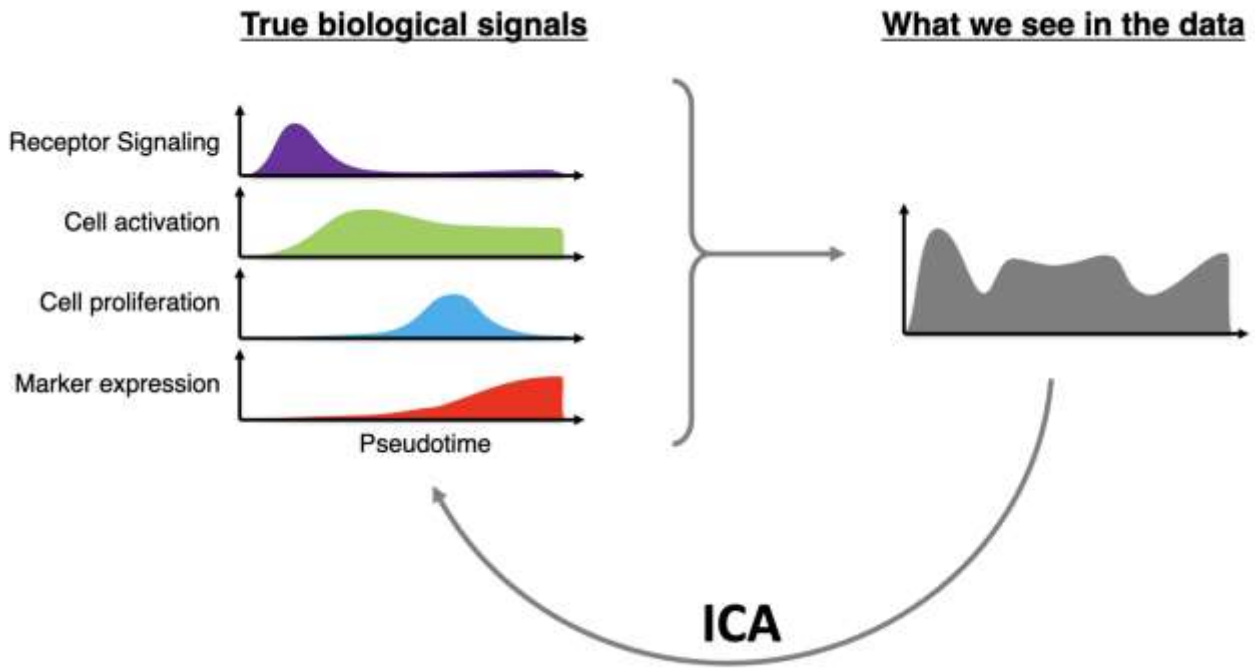
- y is the estimated independent components
- W is the unmixing matrix, computed by ICA



s A $x = As$ $y = Wx$

Trapnell, C et al. Nat Biotechnology 2014

Why ICA

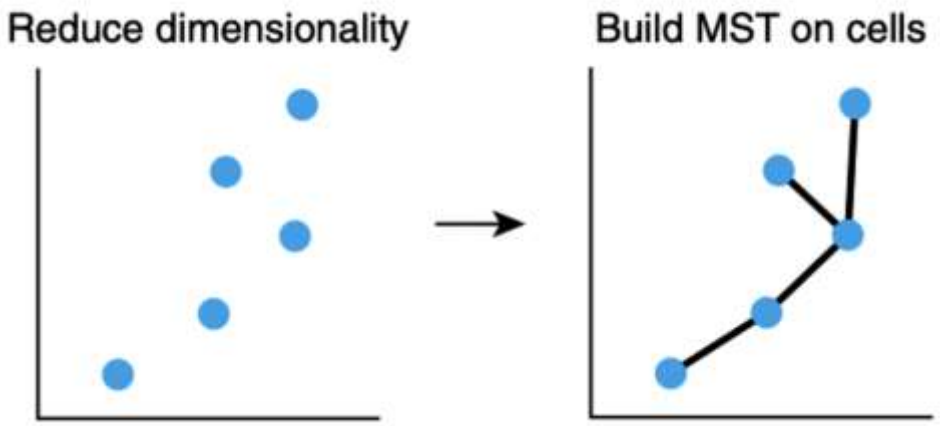


1) PCA : find the directions of maximal variance
- Find the loudest sound in the room (dominant but may be mixed)

2) ICA: find the direction of maximal independence
- Find individual voices in a conversion

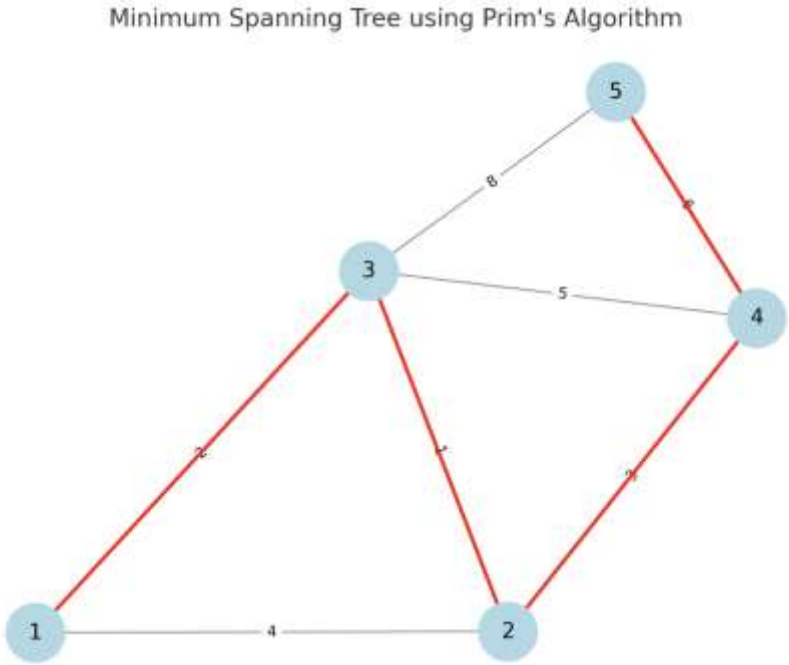
https://github.com/NBISweden/excelerate-scRNAseq/blob/master/session-trajectories/trajectory_inference_analysis.pdf

Step 3: Construct minimum spanning tree (MST) on the cells



- Connect all vertices (cells) in V in $G = (V, E)$
- Has no cycles
- Minimizes the total sum of edge weights

1) Compute the pairwise distance in reduced-dimensional space
2) Prim's algorithm



Trapnell, C et al. Nat Biotechnology 2014

Step 4: Find the longest path through the MST

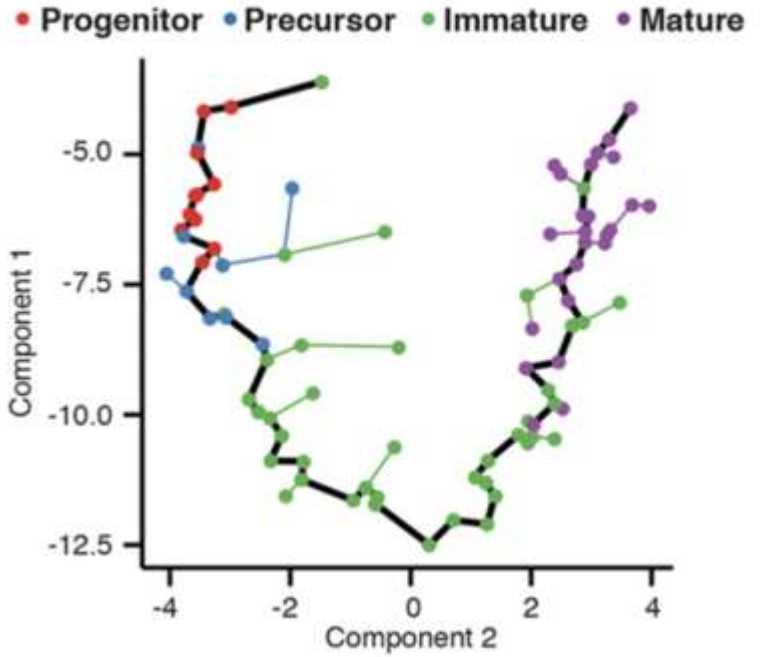
1) The longest continuous path is the best proxy for the differentiation timeline
- Longest path is called the diameter

Trapnell, C et al. Nat Biotechnology 2014

Step 5: Order cells along the trajectory

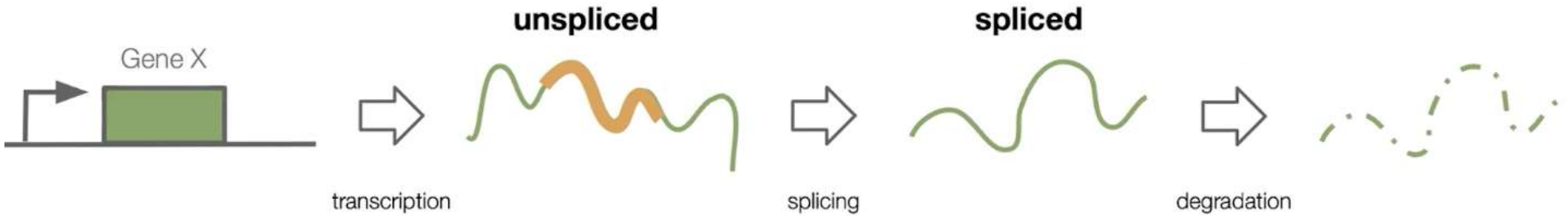
1) Assign a pseudotime value to each cell
- Cells early in the path: undifferentiated states
- Cells later in the path: fully differentiated states

Developmental trajectory of olfactory neurons in mice



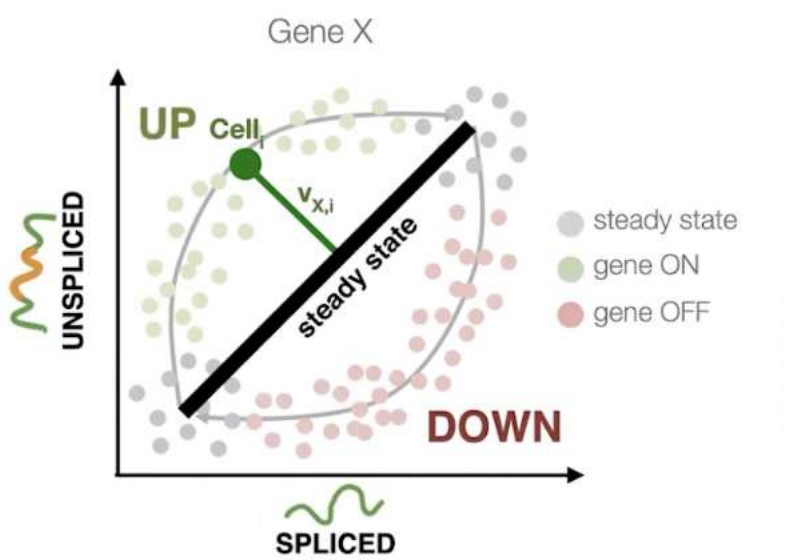
https://cole-trapnell-lab.github.io/projects/sc-trajectories/
Trapnell, C et al. Nat Biotechnology 2014

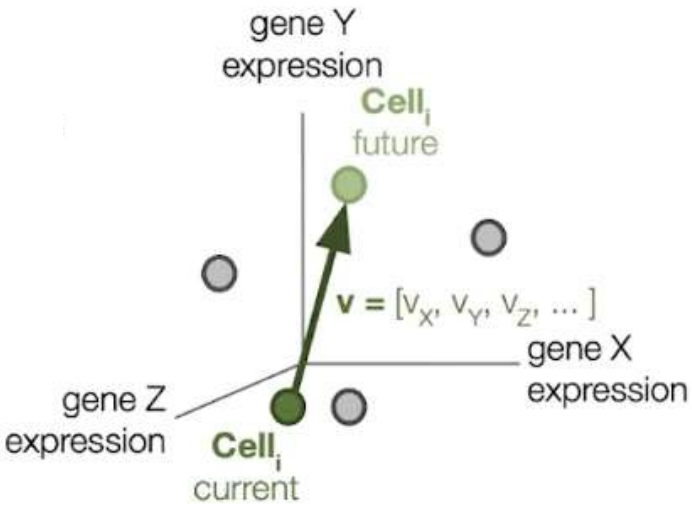
RNA Velocity: spliced and unspliced RNA levels indicate changing gene expression

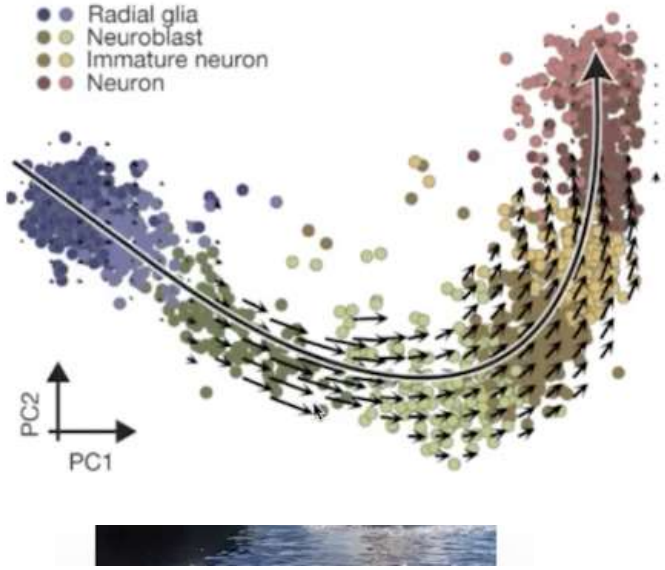


Bioinformatics, 2021: tinyurl.com/veloviz
Sofrware +tutorials : jef.works/veloviz

RNA Velocity: spliced dynamics predict future cell state

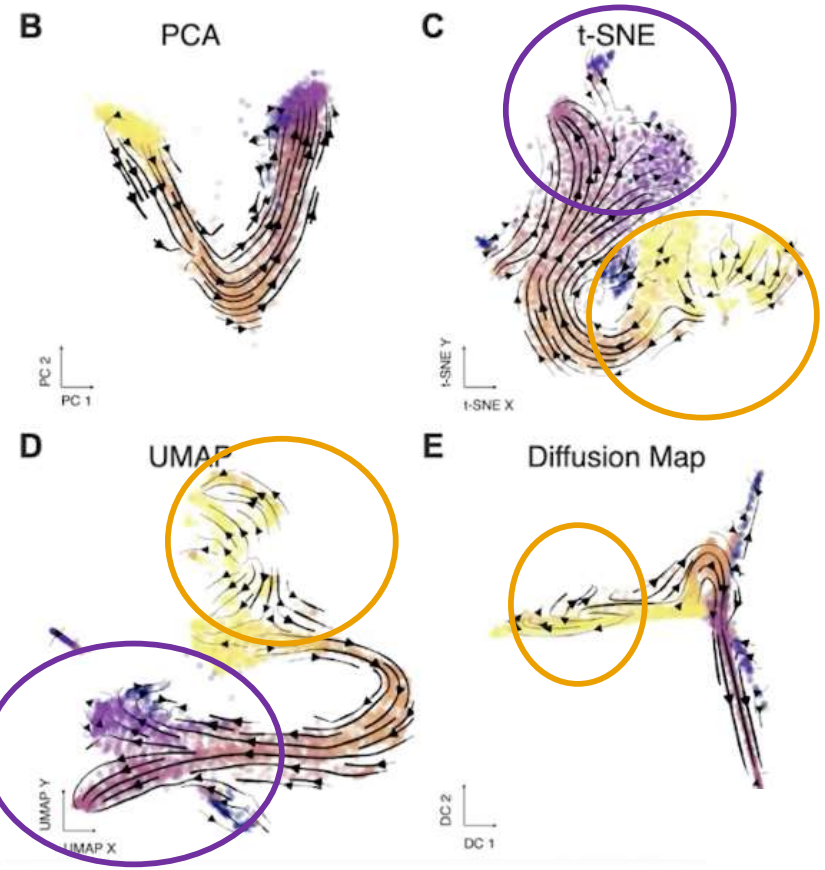


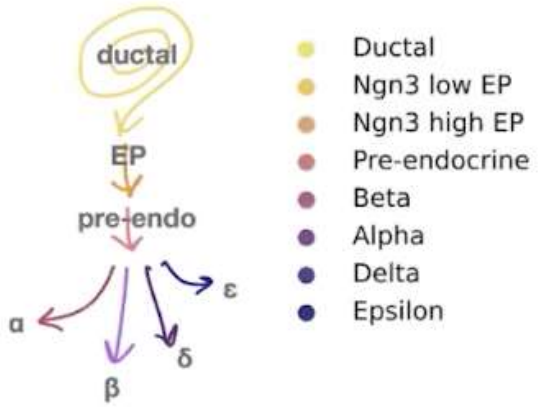




Bioinformatics, 2021: tinyurl.com/veloviz
Sofrware +tutorials : jef.works/veloviz

Visualizing RNA velocity trends: projecting onto existing 2D-embeddings

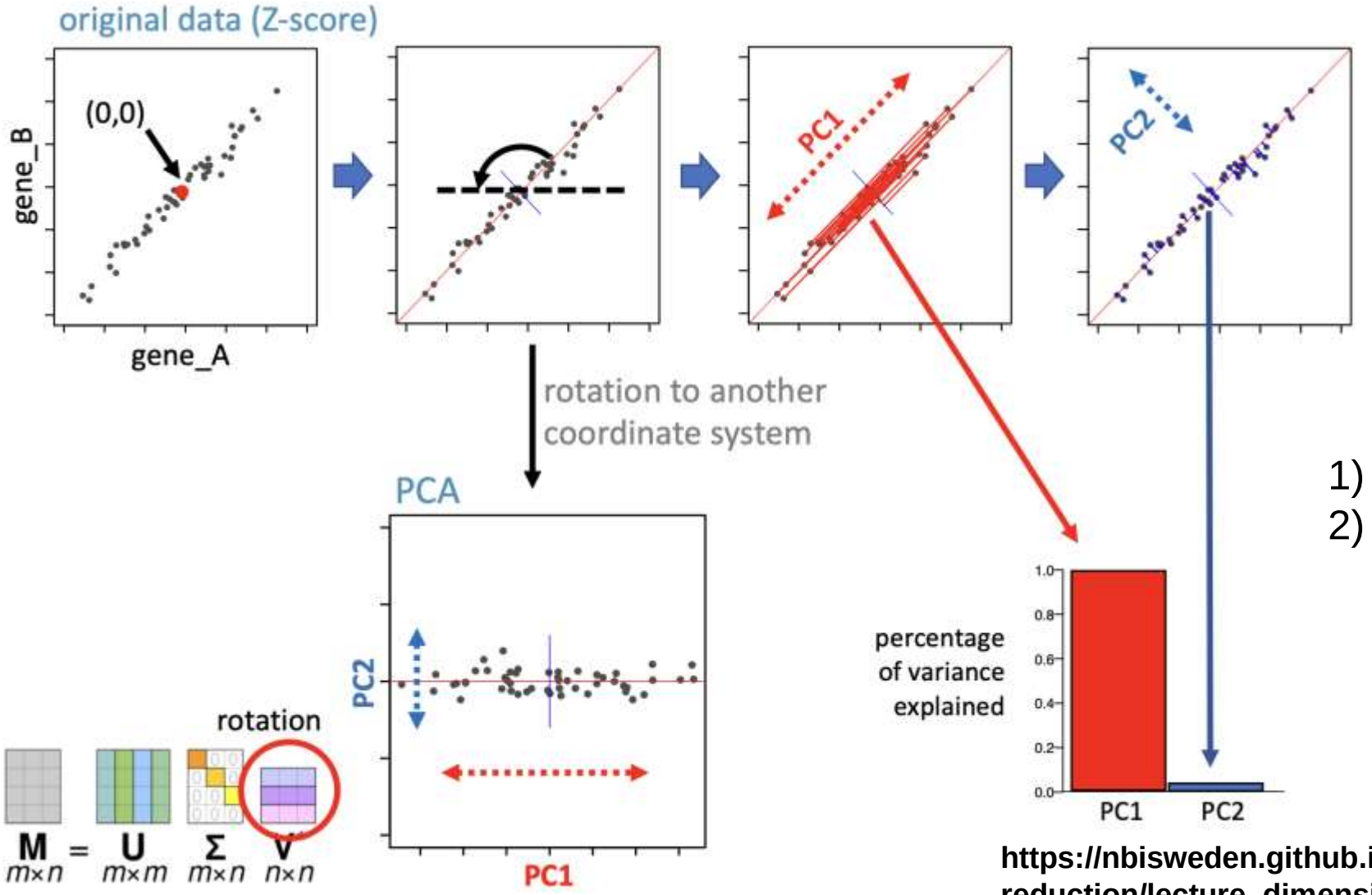




Ductal
Ngn3 low EP
Ngn3 high EP
Pre-endocrine
Beta
Alpha
Delta
Epsilon

Bioinformatics, 2021: tinyurl.com/veloviz
Sofrware +tutorials : jef.works/veloviz

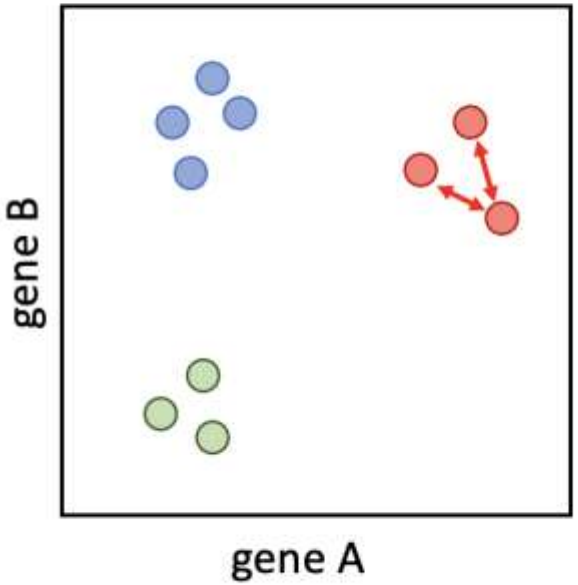
Limitations of PCA for RNA velocity

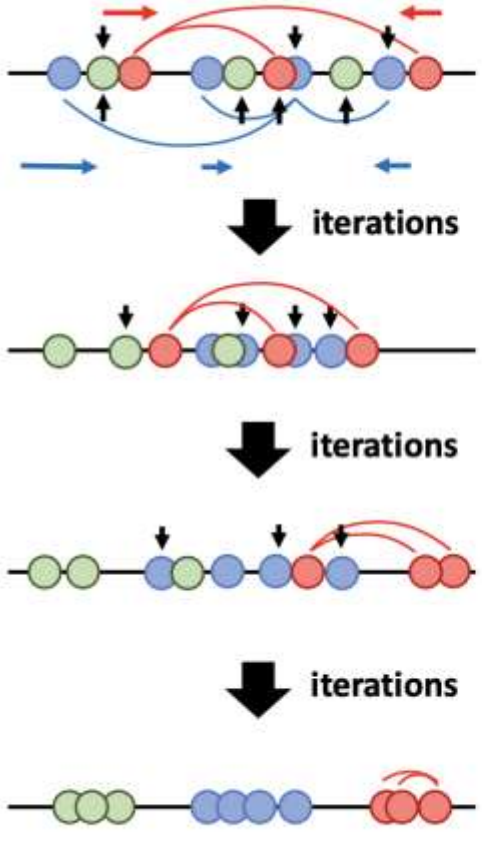


1) Only capture linear trends
2) PCA prioritizes global varianc and ignores local trajectories

https://nbisweden.github.io/excelerate-scRNAseq/session-dim-reduction/lecture_dimensionality_reduction.pdf

Limitations of t-SNE for RNA velocity

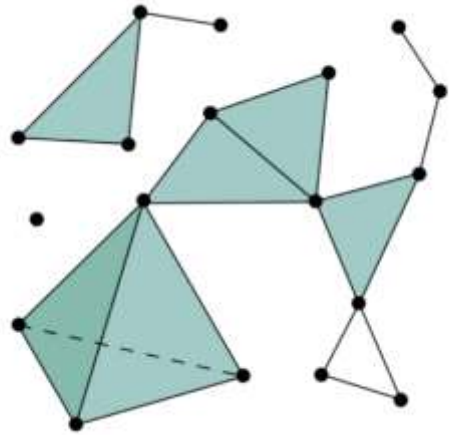




1) Focus on clustering rather than continuity (cells should be connected may end up in different clusters)
2) Stochastic and no fixed geometric structure
3) No global structure

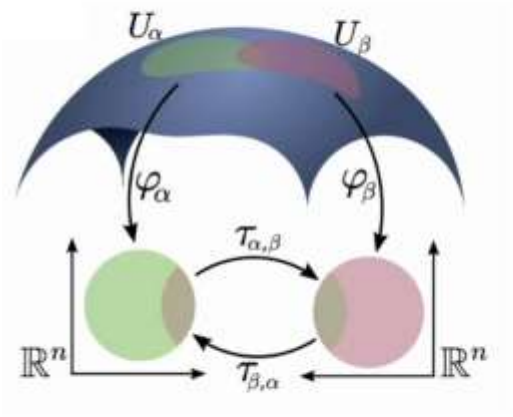
https://nbisweden.github.io/excelerate-scRNAseq/session-dim-reduction/lecture_dimensionality_reduction.pdf

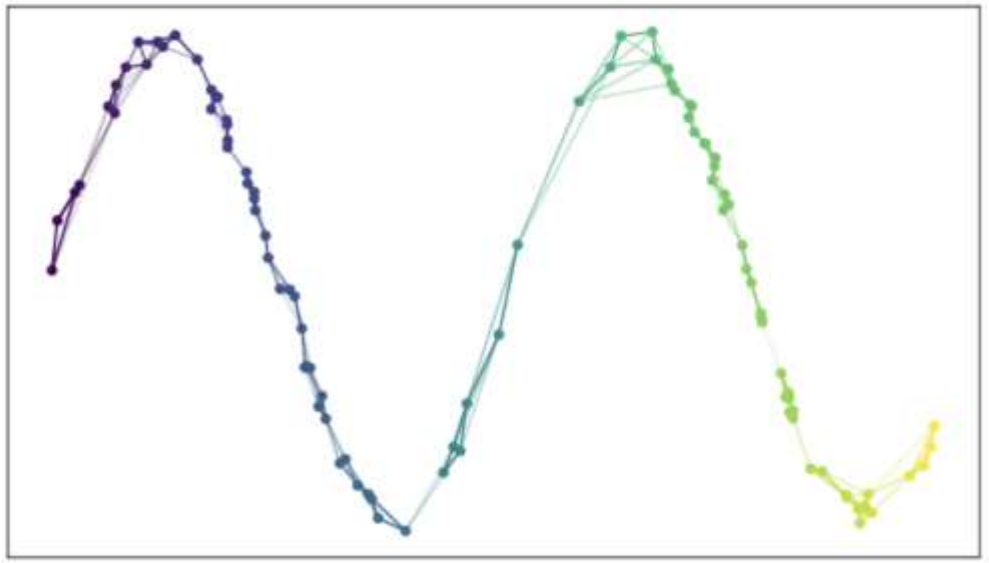
Limitations of UMAP for RNA velocity



- Based on topological structures
- Points are connected if the distance is below a threshold
- Manifold alignment while preserving topology

1) Distance-based but not descriptive of directional process and transcriptional dynamics

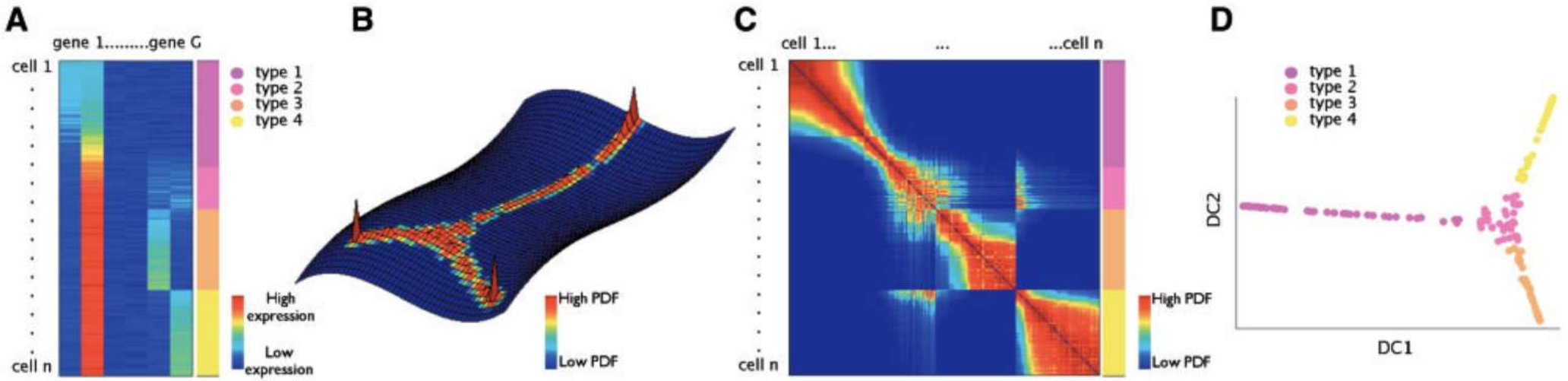




https://nbisweden.github.io/excelerate-scRNAseq/session-dim-reduction/lecture_dimensionality_reduction.pdf

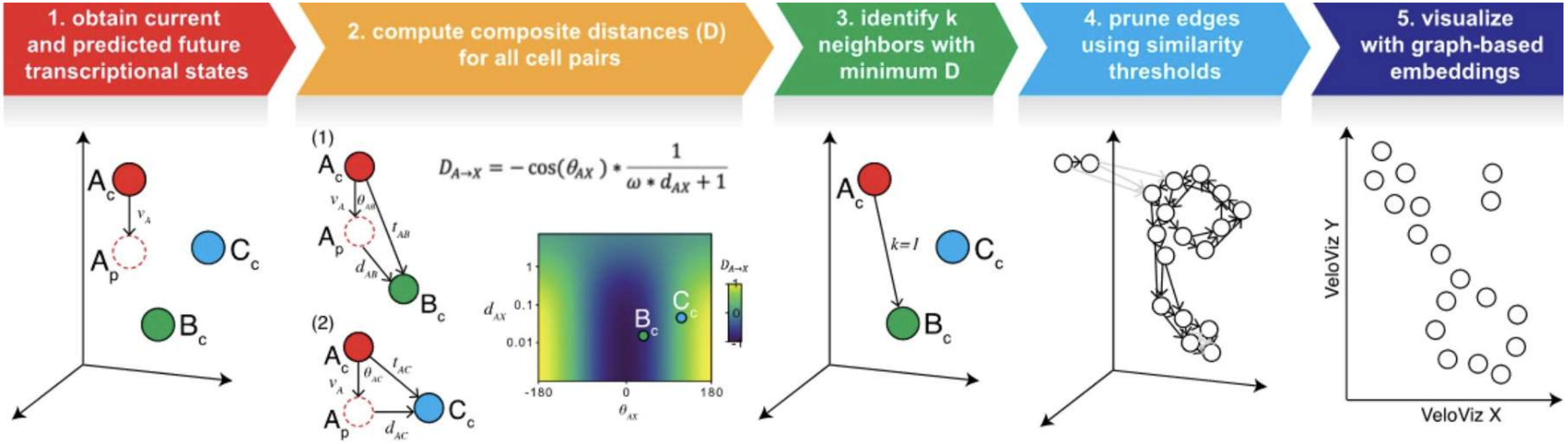
Limitations of diffusion maps for RNA velocity

1) Diffusion maps assume Markovian diffusion (reversible transitions) while RNA velocity is irreversible
2) Global manifold structure and they tend to over smooth local velocity variations



Haghverdi et al 2015

Visualizing RNA velocity: RNA-velocity informed 2D-embeddings using VeloViz



A_c the observed transcriptomic profile
A_p the predicted future state

transition vectors A_c *to* B_c
Cells with small composite distances are more likely transitions

https://nbisweden.github.io/excelerate-scRNAseq/session-dim-reduction/lecture_dimensionality_reduction.pdf

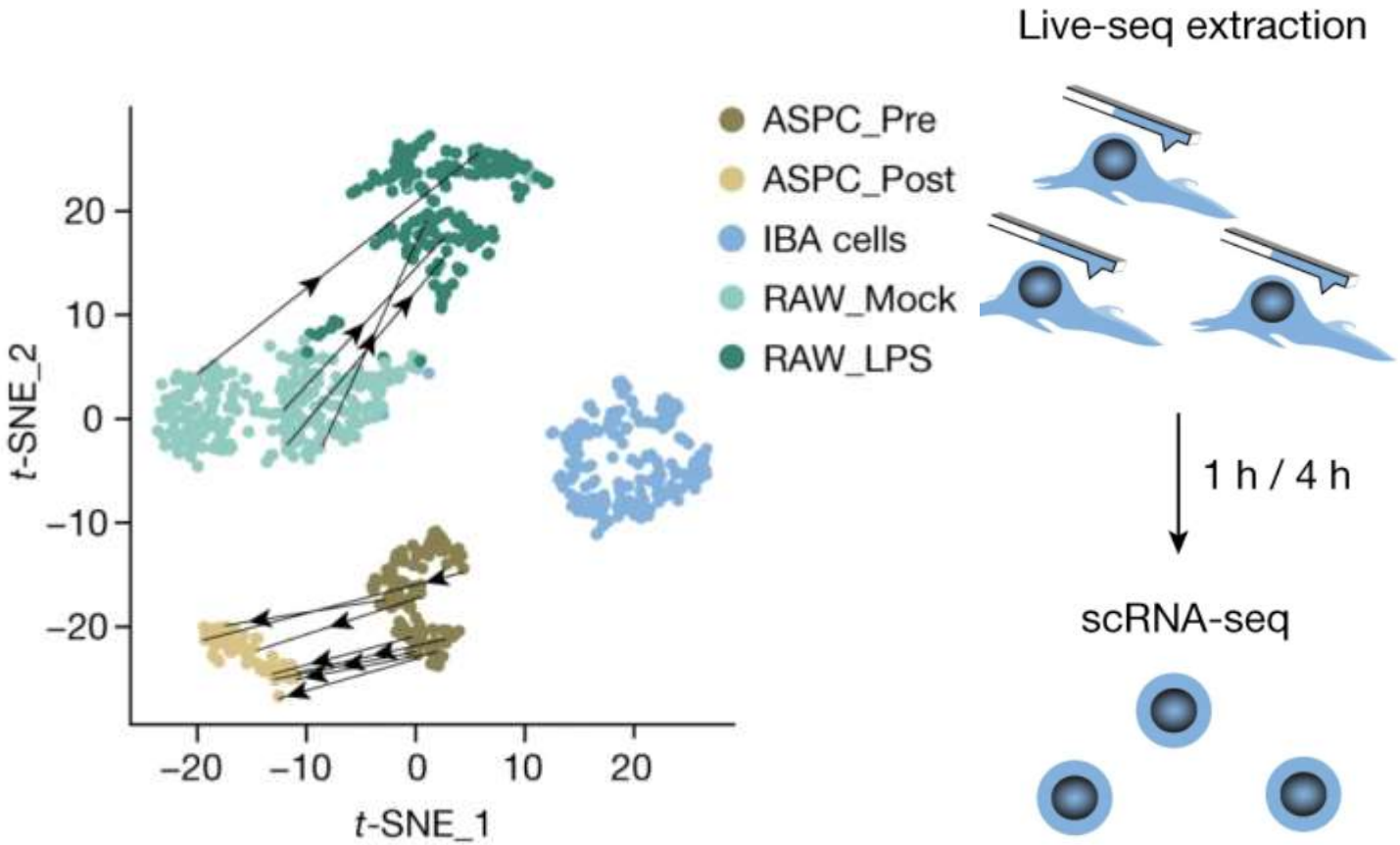
Challenges in Inferring cellular time

1) **Learning ‘pseudo time’ trajectories**

Assume cells take smooth paths

Does the learnt trajectory reflect real biological path?

LIVE-seq



1) Cytoplasmic cellular biopsies

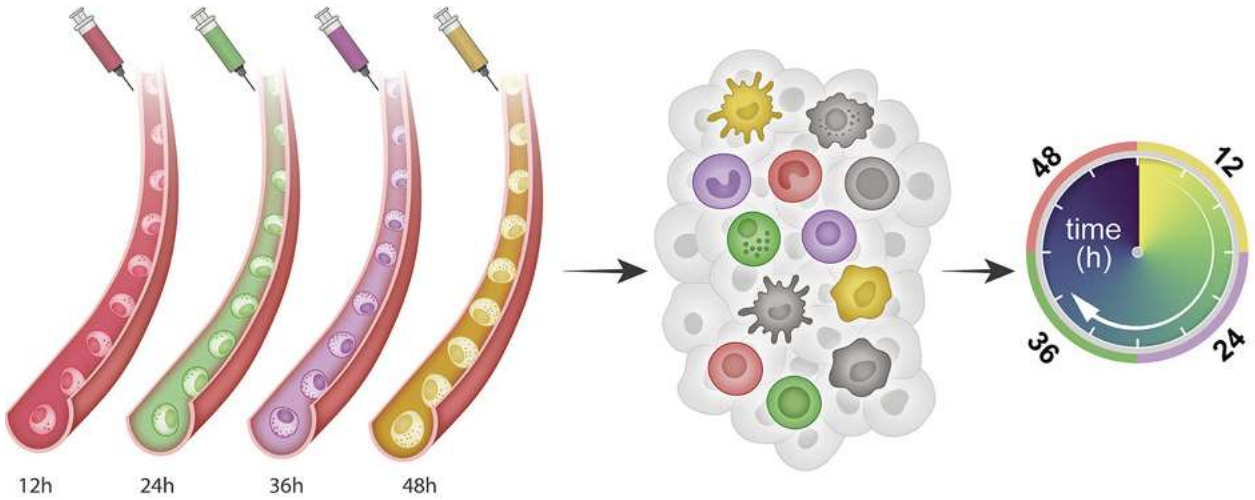
- miniaturized RNA-seq protocols for profiling gene expression

2) Repeated cellular sampling

- time-lapse microscopy enables sequential extractions and profiling
- challenging to scale protocol to large numbers of cells

Chen et al, Nature 2022

Zman-seq

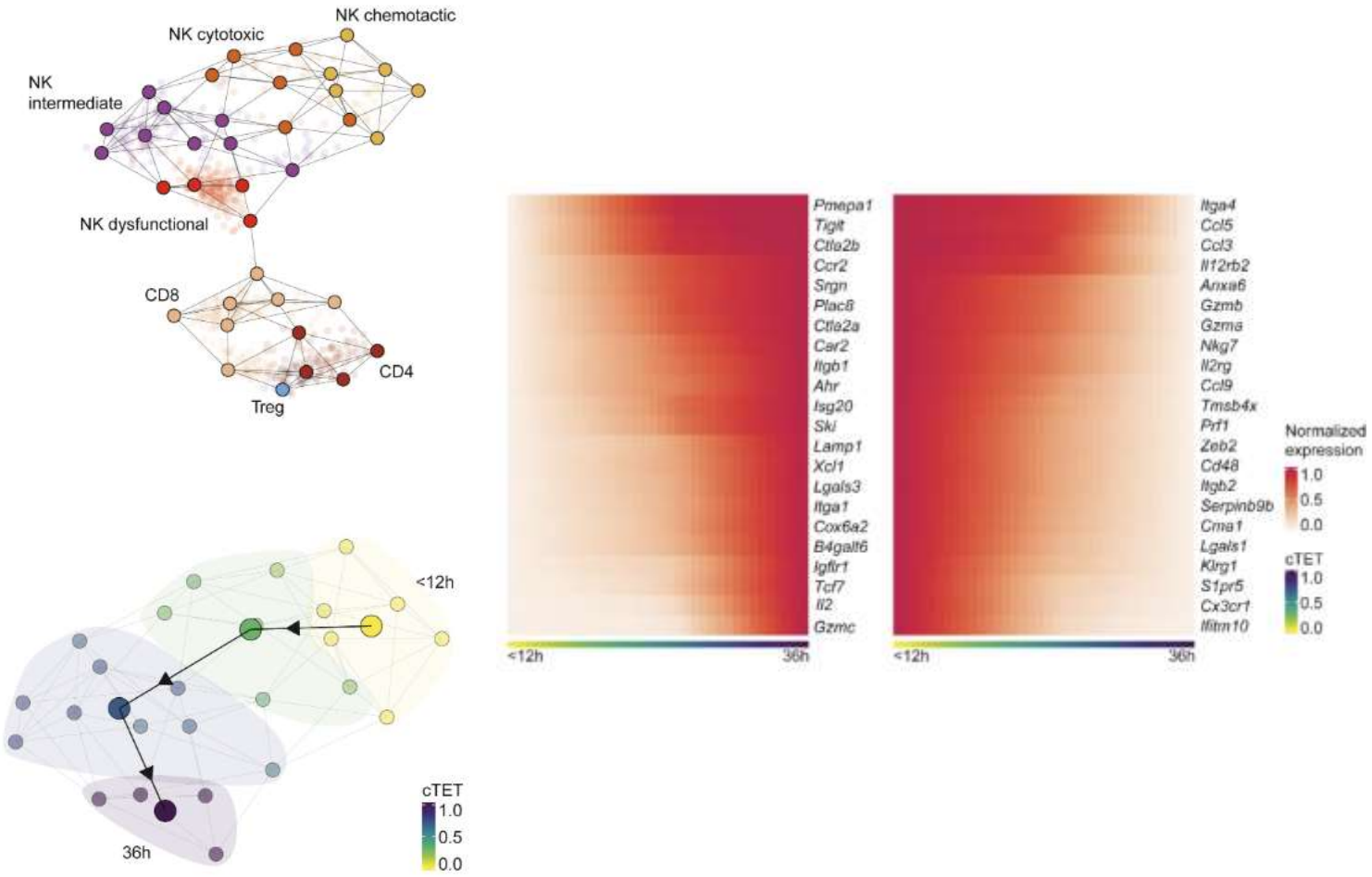


1) Barcoding cellular time in-vivo
- use fluorescent pulse labels

2) Time-stamped cellular dynamics
- cells retain their fluorescent label as a temporal barcode
- can build cellular trajectories with ground truth time-stamps

Kirschenbaum et al, Cell 2024

Zman-seq



1) Barcoding cellular time in-vivo
- use fluorescent pulse labels

2) Time-stamped cellular dynamics
- cells retain their fluorescent label as a temporal barcode
- can build cellular trajectories with ground truth time-stamps

Kirschenbaum et al, Cell 2024

Thank you!