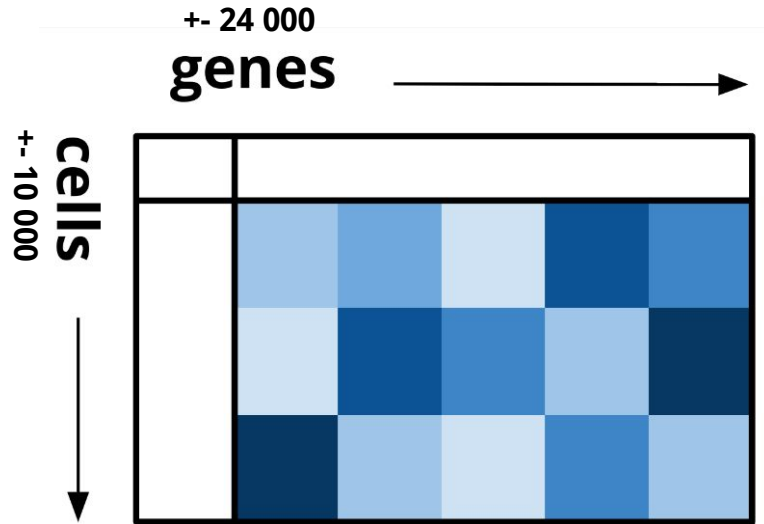


Analysing pseudotemporal single-cell data

Louise Deconinck

Single cell data

Measure gene expression at single-cell level



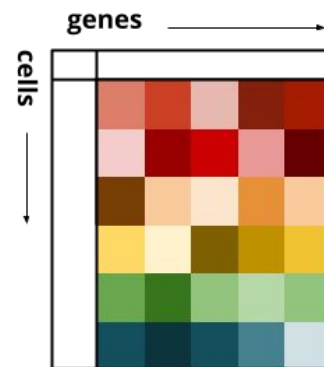
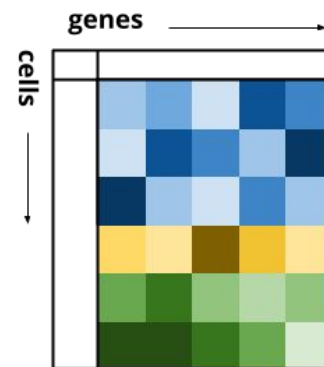
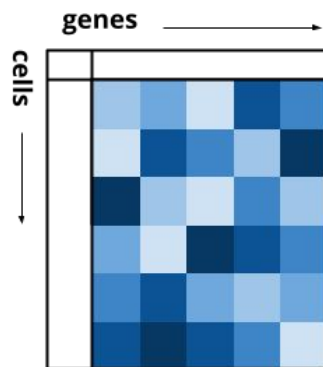
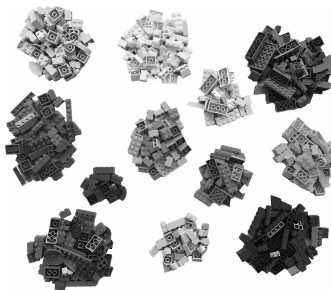
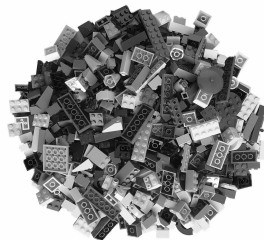
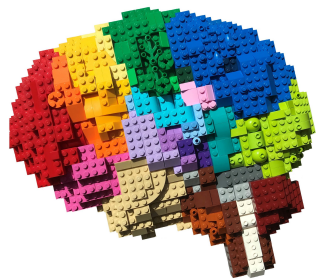
Single cell data

Single cell
transcriptomics data

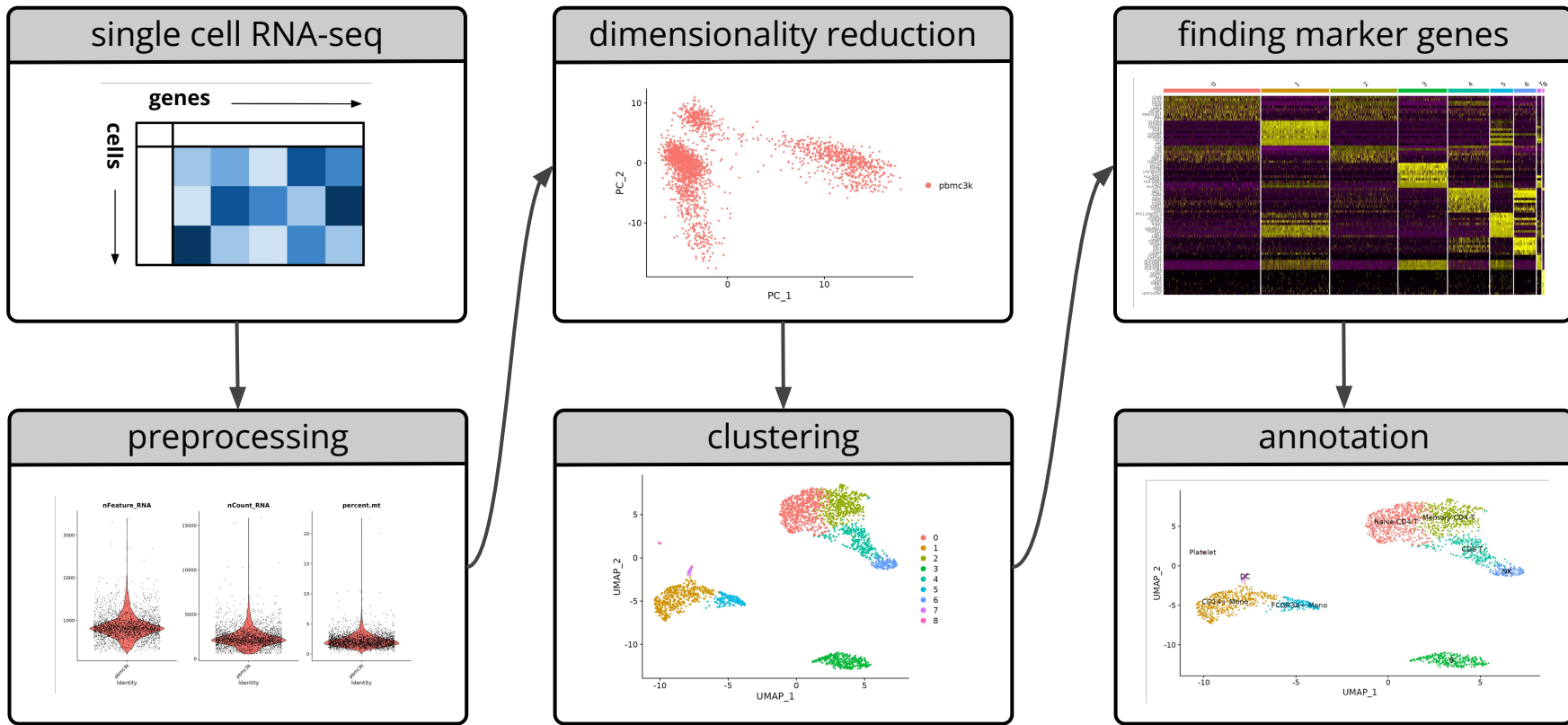
clustering

annotation

trajectory inference

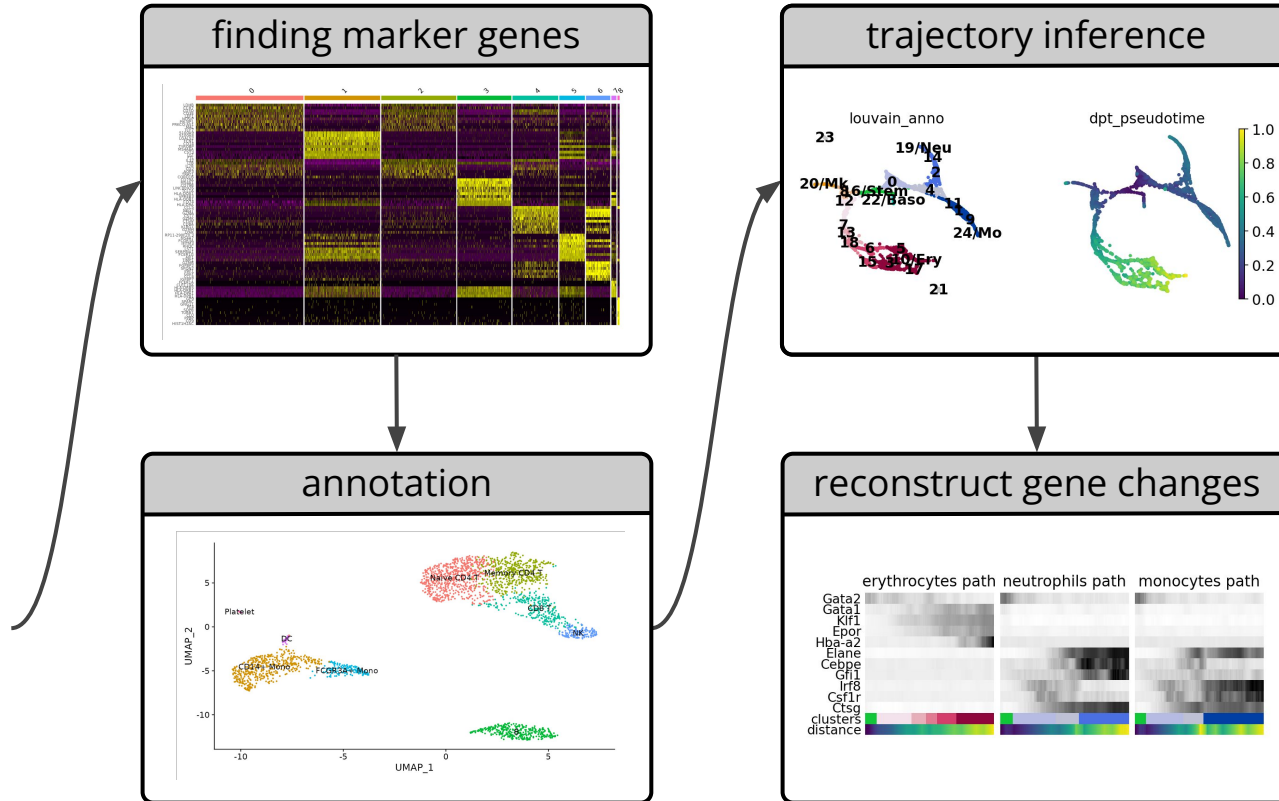


Single cell data

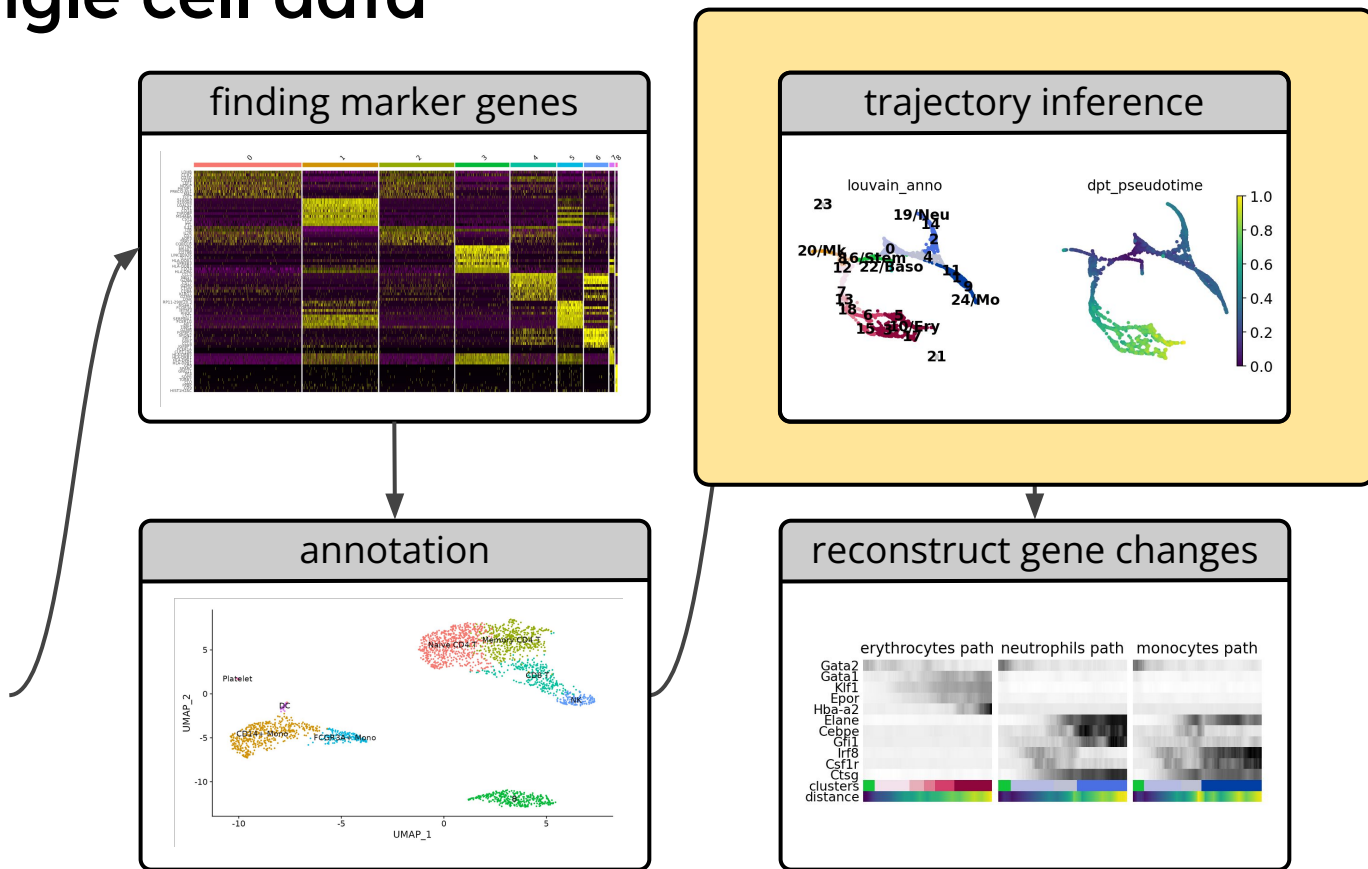


All figures except single cell RNA-seq from:
Seurat guided clustering tutorial: https://satijalab.org/seurat/articles/pbm3k_tutorial.html

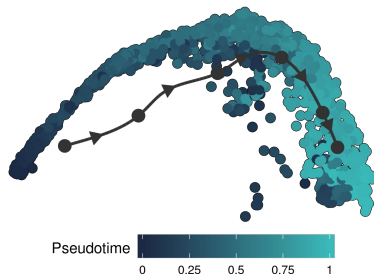
Single cell data



Single cell data



Trajectory inference

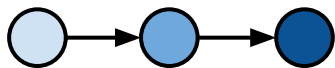


3 assumptions:

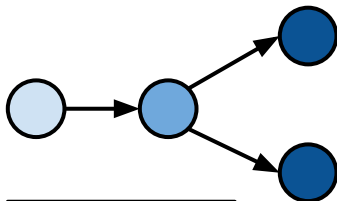
1. Dynamic biological process: cells are **representative** samples
 2. Sampling to sufficient depth to capture very **transient** states
 3. Changes in gene expression is **gradual** during developmental process
- + Method specific assumptions: underlying shape, prior knowledge (starting or ending point)

Trajectory inference

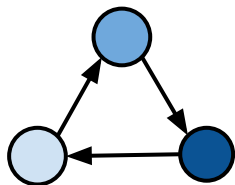
- + Method specific assumptions: underlying shape, prior knowledge (starting point)



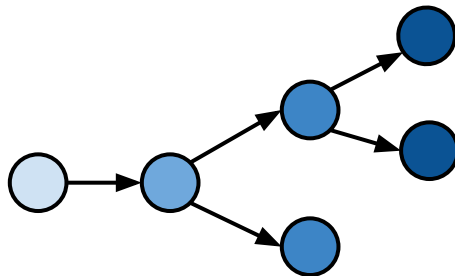
Linear



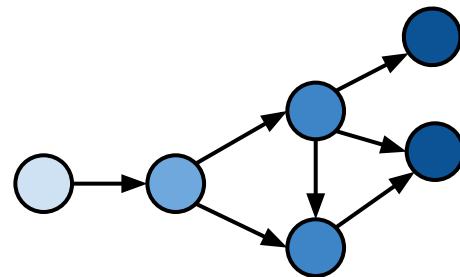
Branching



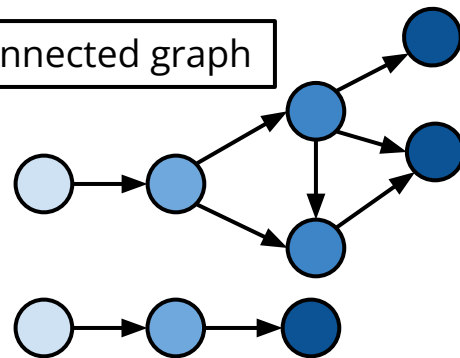
Cycle



Tree

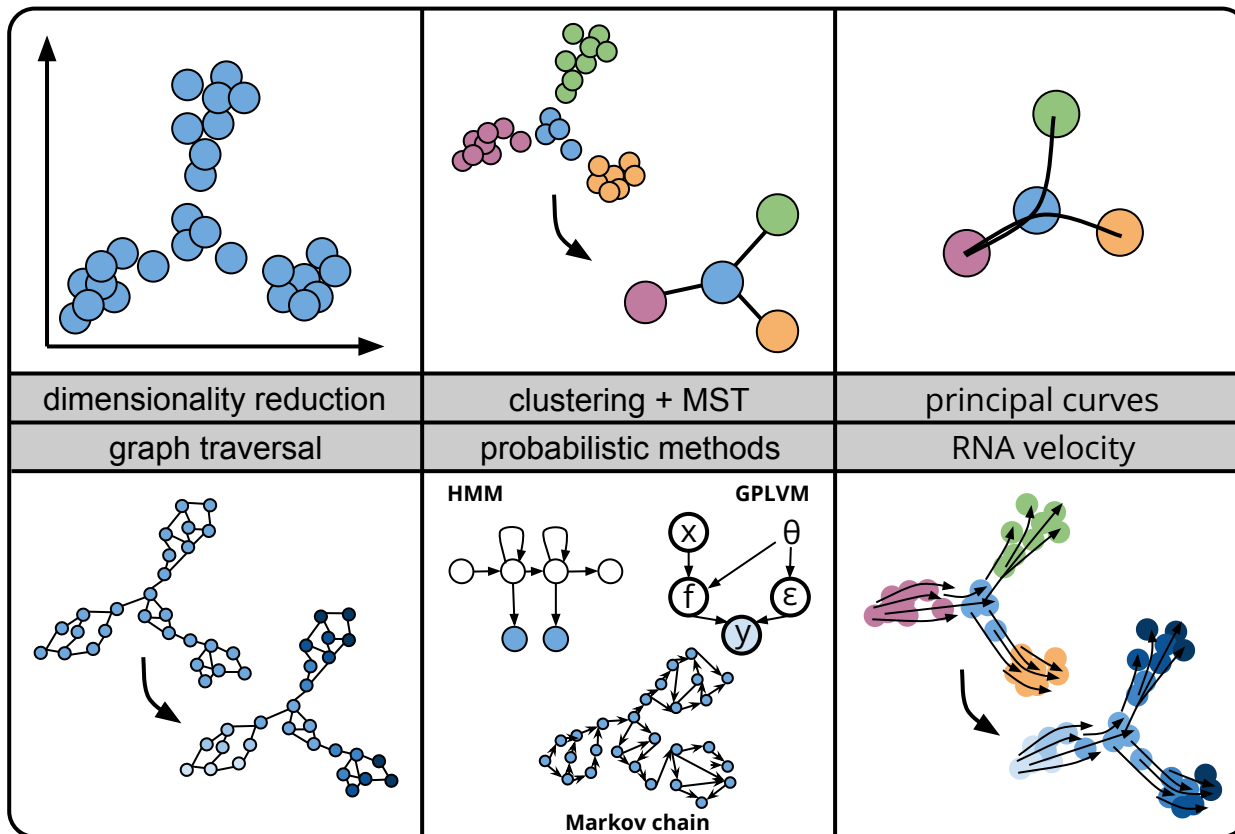


Connected graph

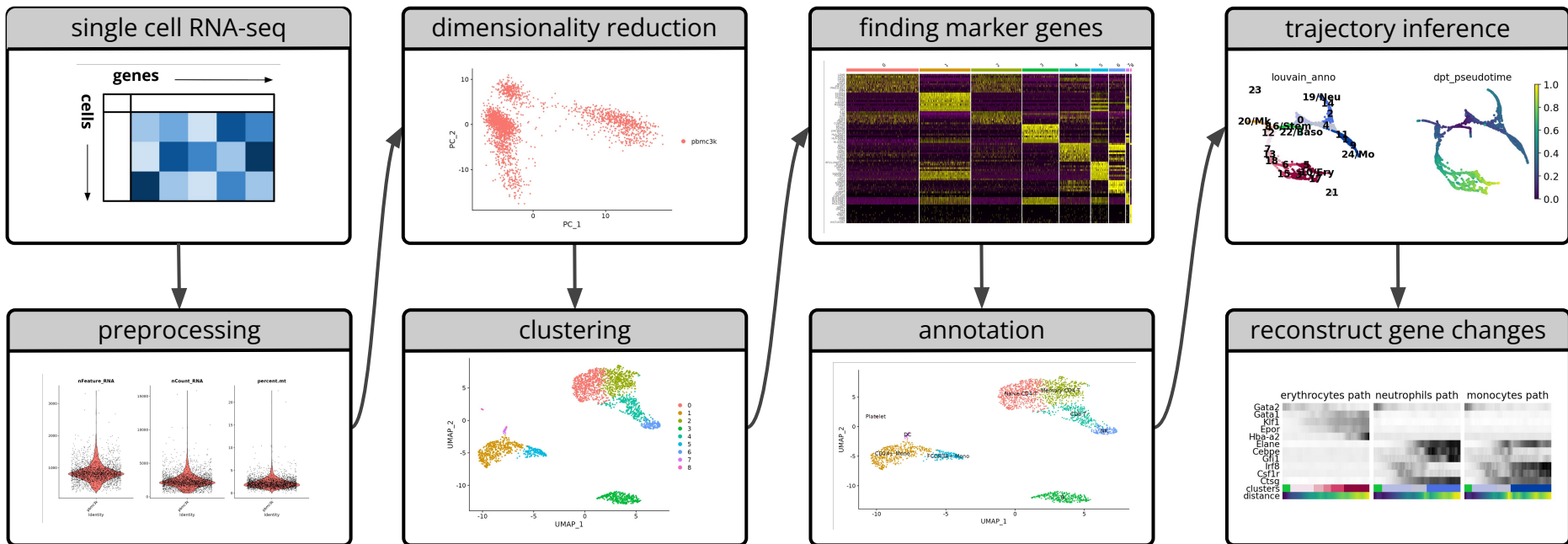


Disconnected graph

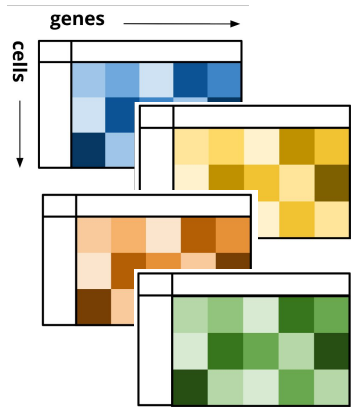
Trajectory inference



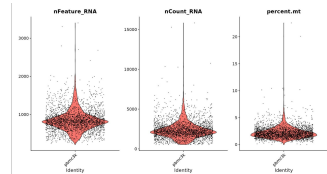
Issues with samples & batches



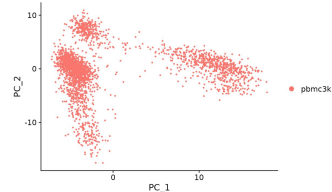
single cell RNA-seq



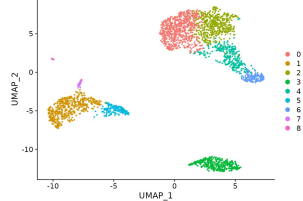
preprocessing



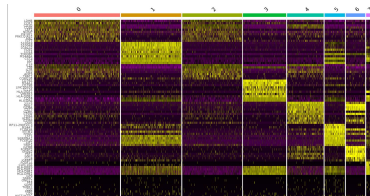
dimensionality reduction



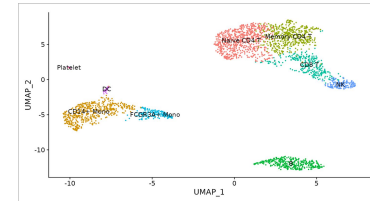
clustering



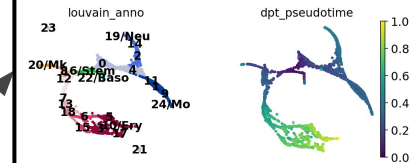
finding marker genes



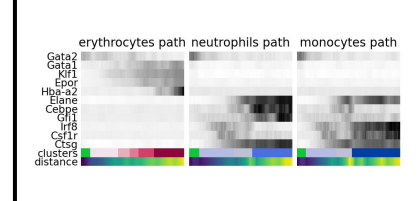
annotation

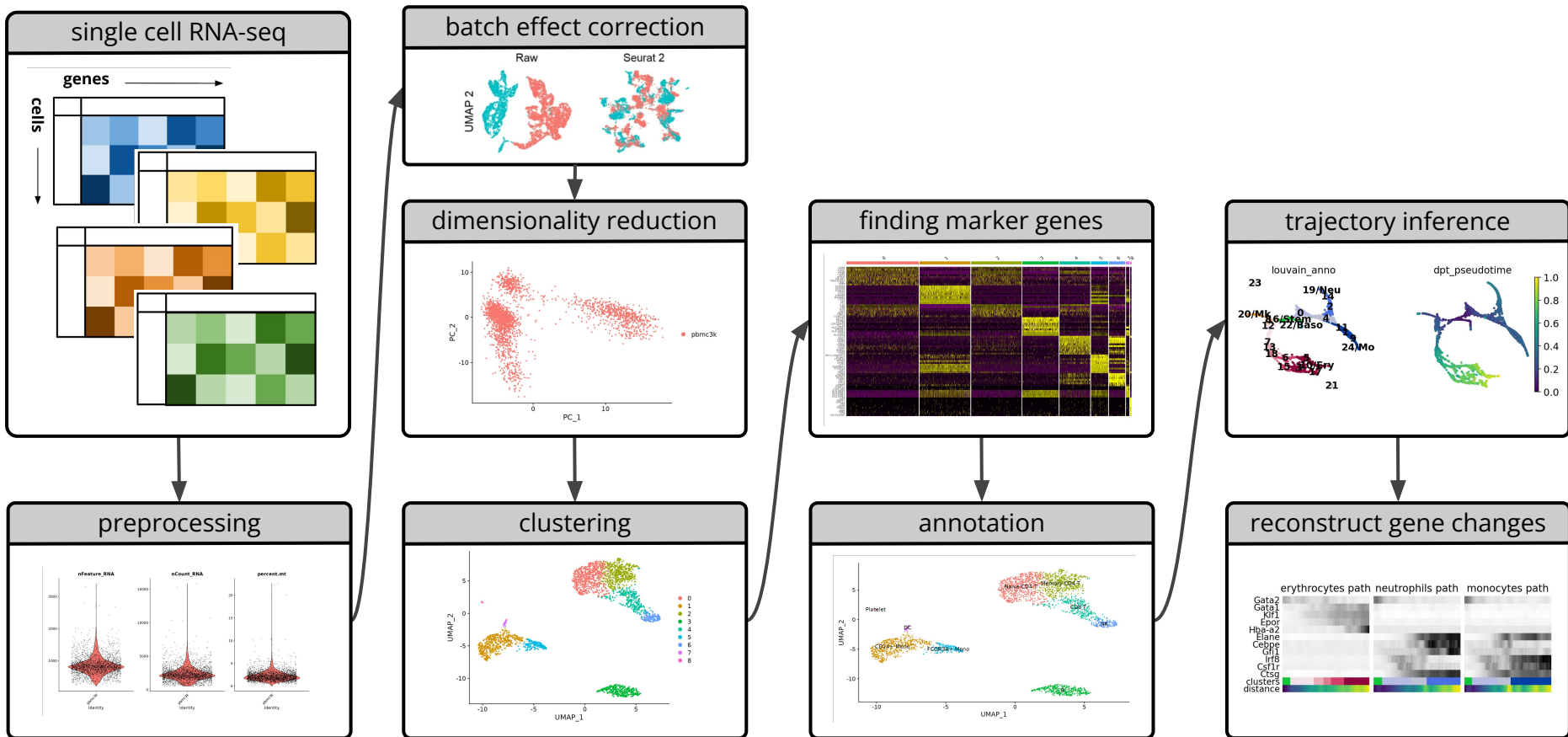


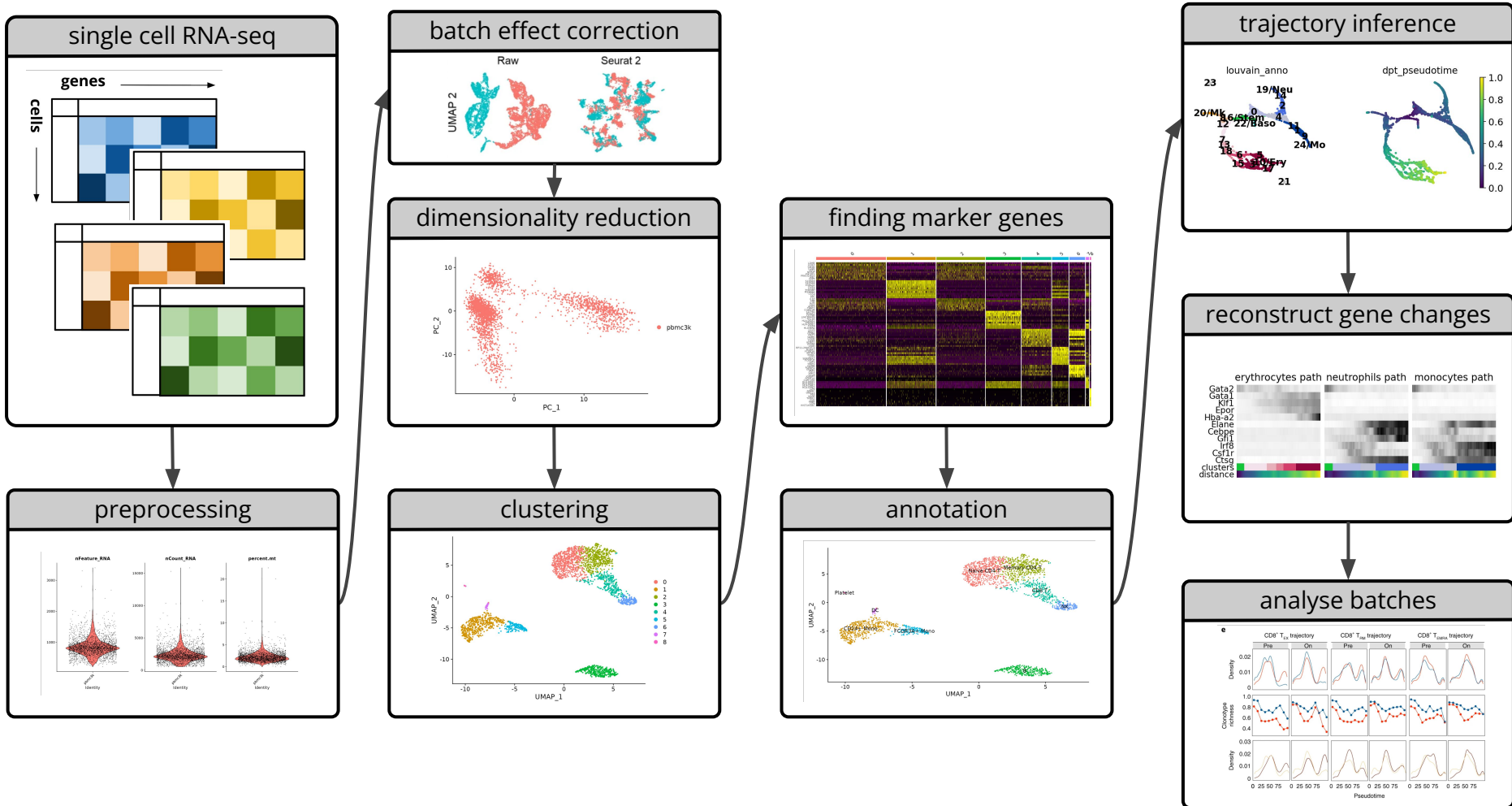
trajectory inference



reconstruct gene changes

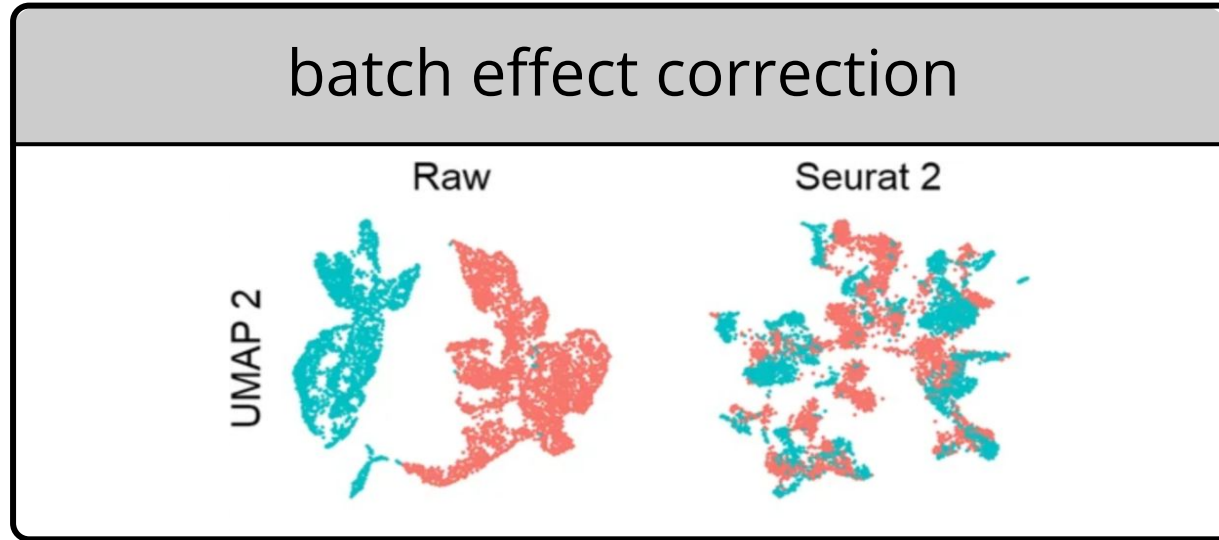






What is the problem?

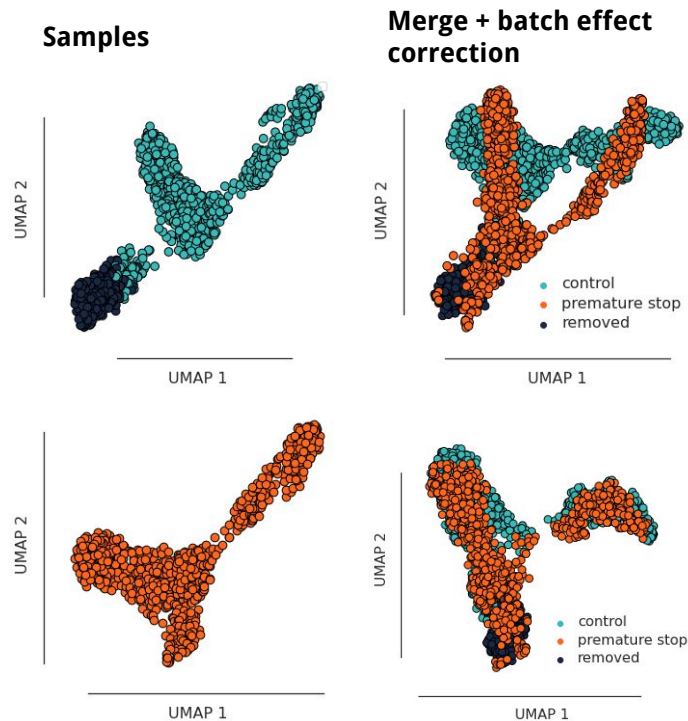
Batch effect correction is really hard & can affect the downstream analysis



What is the problem?

Batch effect correction is really hard & can affect the downstream analysis

- Simulate batch effect data using dyngen:
 - Flexible simulator for dynamic cellular processes



What is the problem?

Batch effect correction is really hard & can affect the downstream analysis

Benchmarking atlas-level data integration in single-cell genomics

[Malte D. Luecken](#), [M. Büttner](#), [K. Chaichoompu](#), [A. Danese](#), [M. Interlandi](#), [M. F. Mueller](#), [D. C. Strobl](#), [L. Zappia](#), [M. Dugas](#), [M. Colomé-Tatché](#) ✉ & [Fabian J. Theis](#) ✉

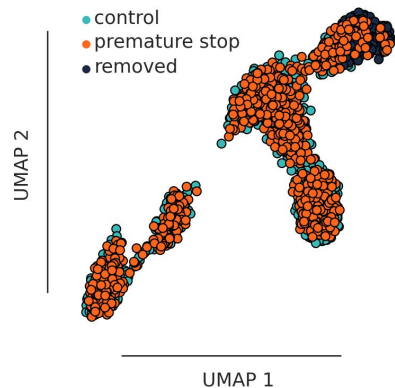
[Nature Methods](#) **19**, 41–50 (2022) | [Cite this article](#)

43k Accesses | **56** Citations | **363** Altmetric | [Metrics](#)

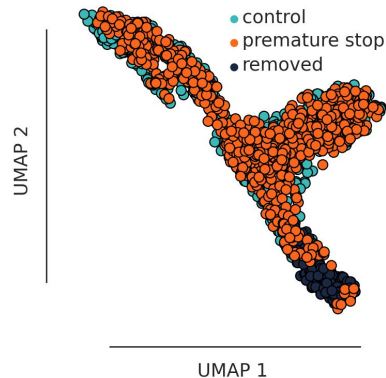
a

	Considerations	scANVI	Scanorama embed	scVI	FastMNN embed	scGen	Harmony	FastMNN gene	Seurat v3 RPCA	BBKNN	Scanorama gene	ComBat	MNN	Seurat v3 CCA	trVAE	Conos	DESC	LIGER	SAUCIE embed	SAUCIE gene
Input	Programming language																			
	Method runs without additional information	✗				✗														
Scib results	Consistent top performer	✓	✓	✓		✓														
	Top method on small/simple tasks		✓		✓	✓	✓													
	Top method on large/complex tasks	✓	✓	✓		✓														
	Top method on ATAC data	—		—			✓											✓		
Task details	Integrates strong batch effects	✓	—	—		✓			—	—				—						
	Top method for recovery cell states or modules	✓	✓								✓	✓	✓							
	Confounding of bio and batch variance	✓	—			✓														
	Top method for trajectories	—	✓	—	✓	✓														
Speed	Method deals with varying compositions											✗								
	Fast method for quick results									✓		✓								
	Scales well to large datasets on CPU	✓	—	✓						✓	—								✓	✓
	Method has GPU support	✓		✓		✓									✓		✓		✓	✓
	Scales well to feature spaces beyond genes														✓	✓				
Output	Method shows corrected expression					✓		✓	✓		✓	✓	✓	✓						✓
	Method gives relative cell embeddings									✗						✗				

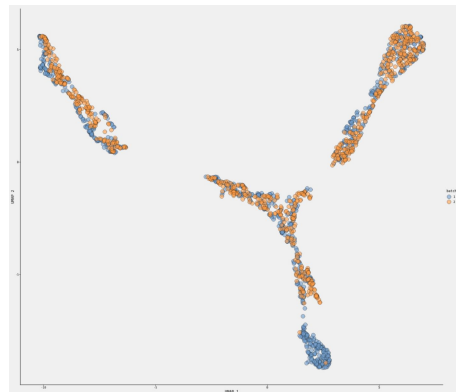
What is the problem?



scanorama



scGen



fastMNN

These methods also overcorrect on this simple dataset

Existing work: cellAlign

Brief Communication | Published: 12 March 2018

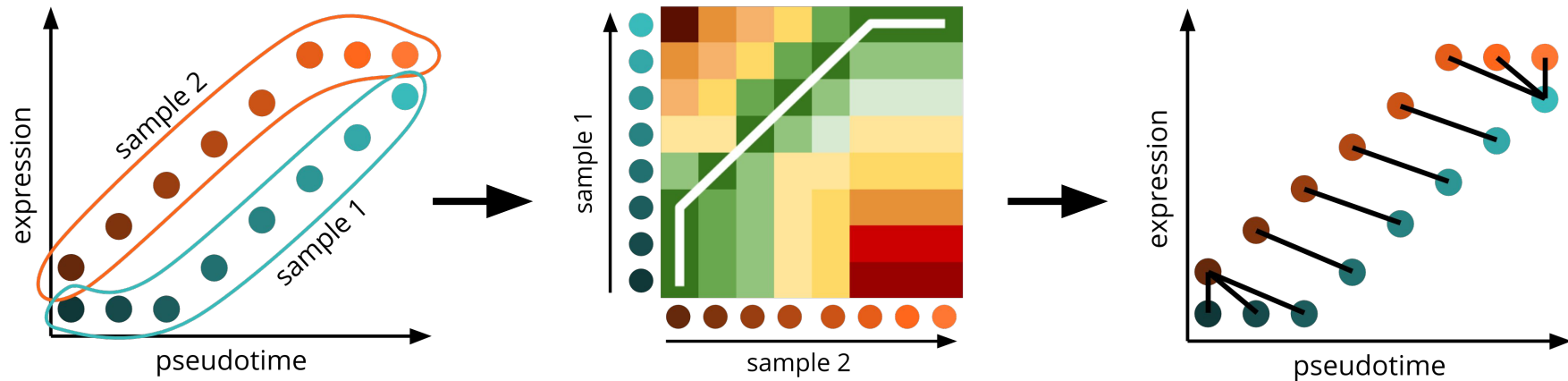
Alignment of single-cell trajectories to compare cellular expression dynamics

Ayelet Alpert, Lindsay S Moore, Tania Dubovik & Shai S Shen-Orr 

Nature Methods **15**, 267–270(2018) | [Cite this article](#)

Existing work: cellAlign

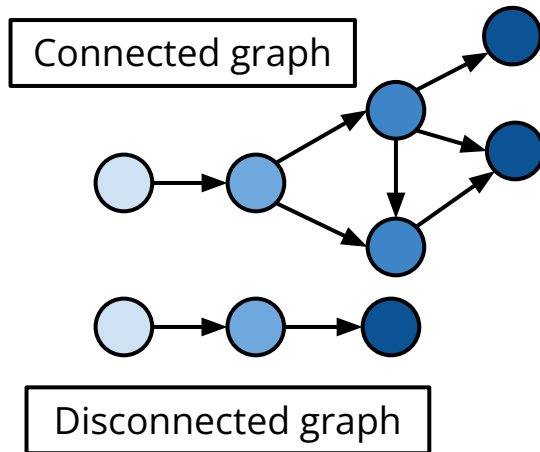
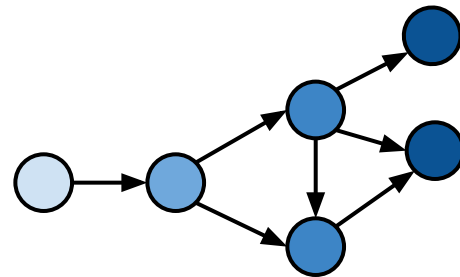
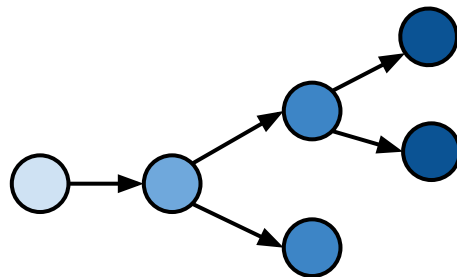
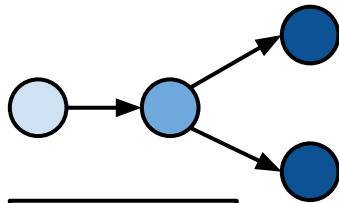
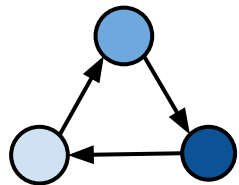
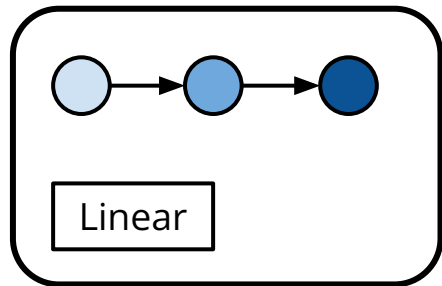
Dynamic Time Warping



→ Needleman-Wunsch

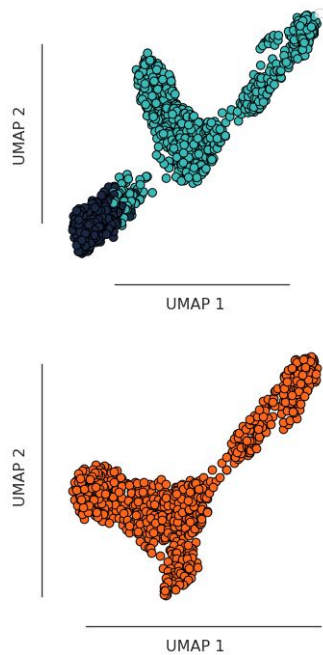
→ Approximate string matching

cellAlign limitations

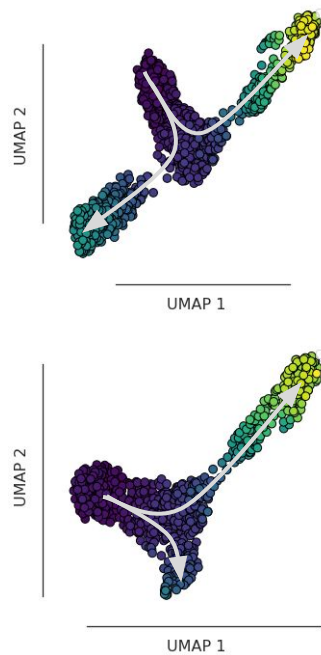


dynchro

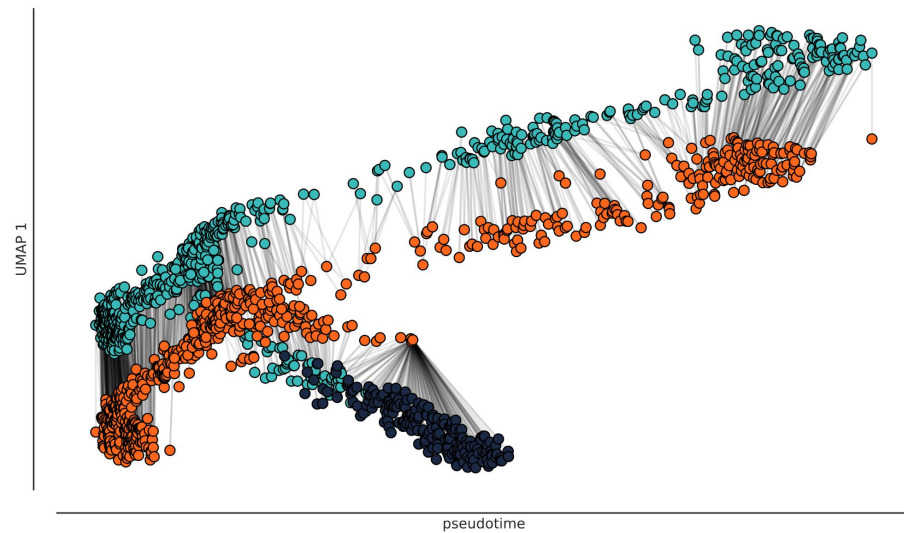
Samples



Pseudotime inference



Pseudotime analysis



dynchro: multiple trajectories

Iterative sample alignment, using a hierarchical clustering approach
(`scikit.cluster.AgglomerativeClustering`)

