

Single-cell trajectories

Wouter Saelens

PhD student

Saeys lab @ VIB - Ghent University, Belgium

<https://github.com/zouter/dyno-workshop-roscoff-2019>
(includes a link to this presentation)

How to analyze single-cell dynamics

Current state of trajectory inference

Some internal methodology

Extensions

How to analyze single-cell dynamics

Current state of trajectory inference

Some internal methodology

Extensions

3 ways to assess single-cell dynamics

Wet

Dry

Barcoding

& Lineage tracing

1. Barcode progenitors
2. Wait
3. Analyze barcodes & phenotype of descendants

Live cell imaging

& Cell tracking

1. Image cells continuously
2. Segment and track individual cells and descendants

Single-cell omics

& Trajectory inference

1. Analyze the (transcript)-ome of many cells and samples
2. Order and infer topology

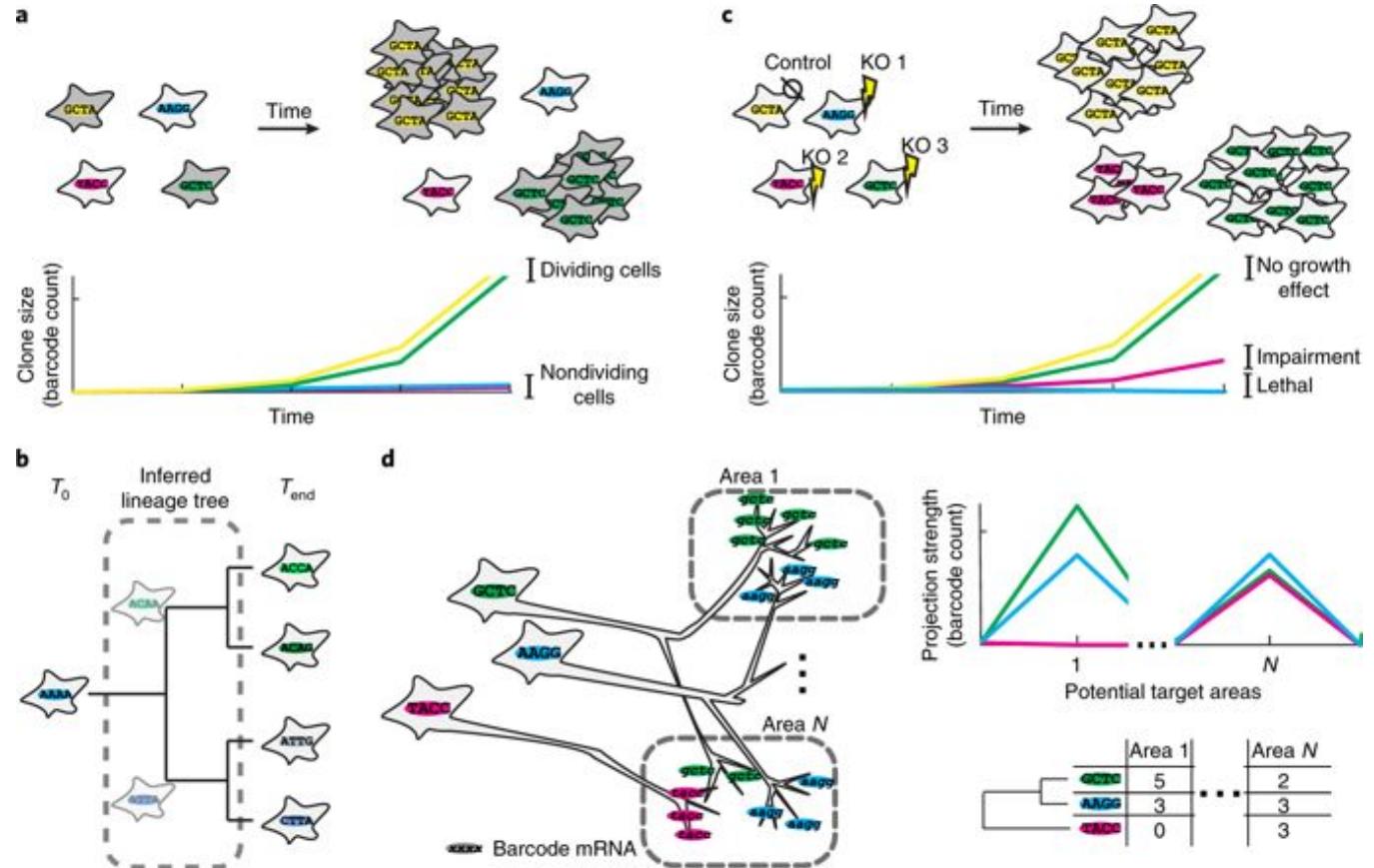
Barcoding & Lineage tracing

- **Transient but stochastic barcoding**

- Technical challenge: a lot of trade-offs
- Harder *in vivo*, nearly impossible in patients

- **Read-out**

- Unlimited possibilities, as long as the barcode can be read
- Only the final state is assessed



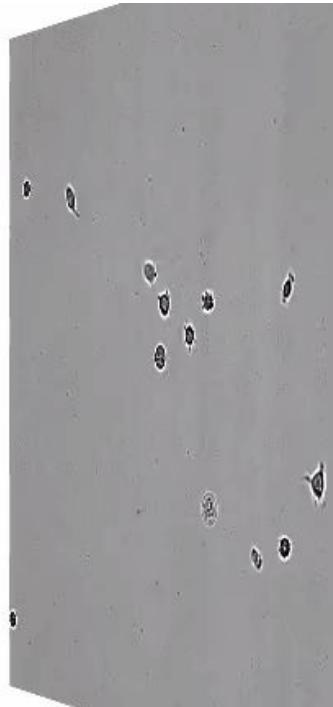
Live cell imaging & Cell tracking

- **Image cells continuously**

- Difficult *in vivo*, nearly impossible *in patients*
- Limitations on additional cellular features

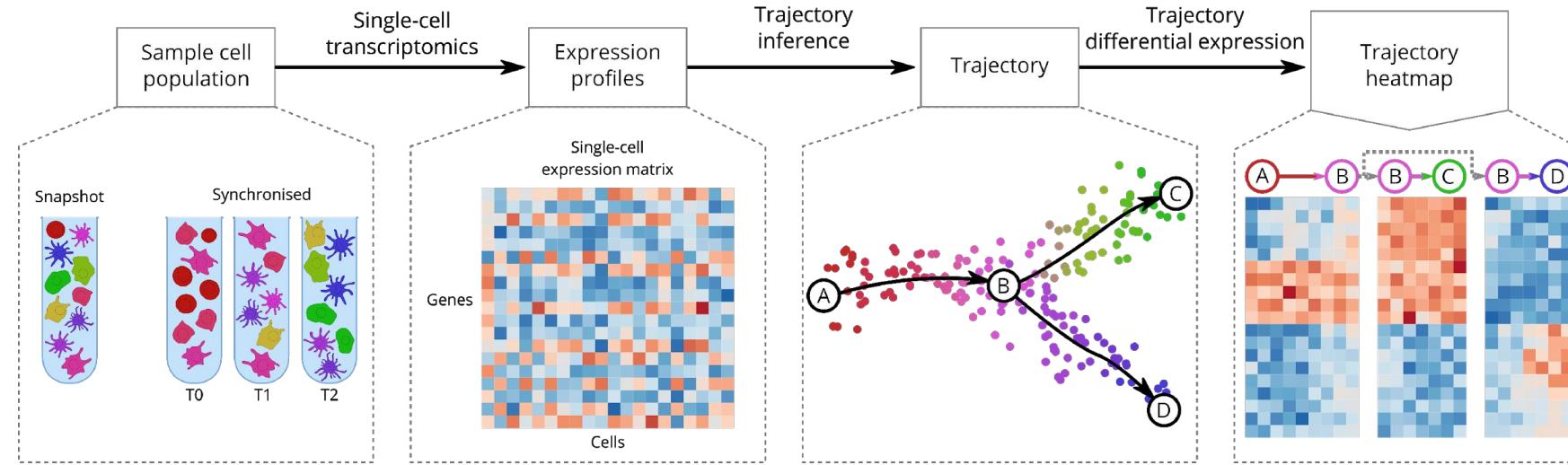
- **Segment and track individual cells and descendants**

- Segmentation errors
- Cells going out-of-bounds



0d - 00:00:00

Single-cell omics & Trajectory inference



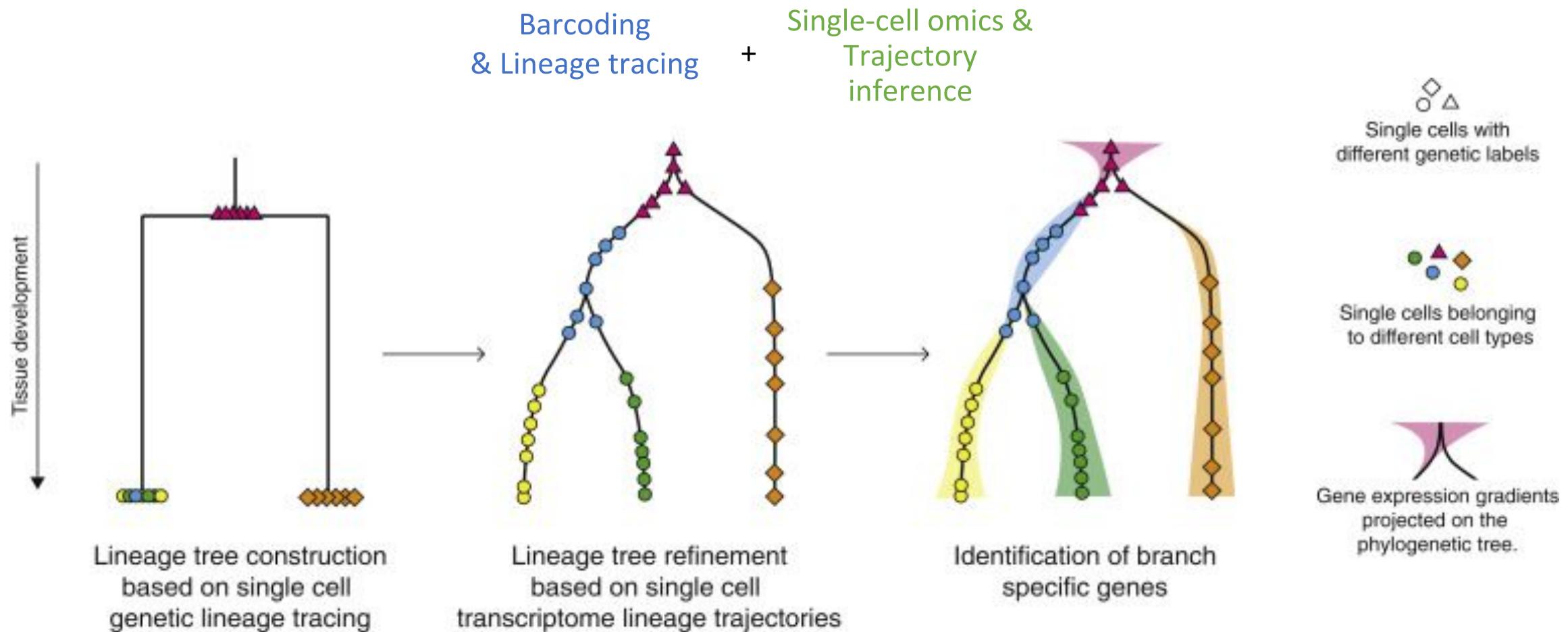
- **Analyze the (transcript)-ome of many cells and samples**

- Destructive
- All intermediates must be present

- **Trajectory inference**

- Trajectories is only a "model", one of many possible ones in the data

The future: combinations



How to analyze single-cell dynamics

Current state of trajectory inference

Some internal methodology

Extensions

2014

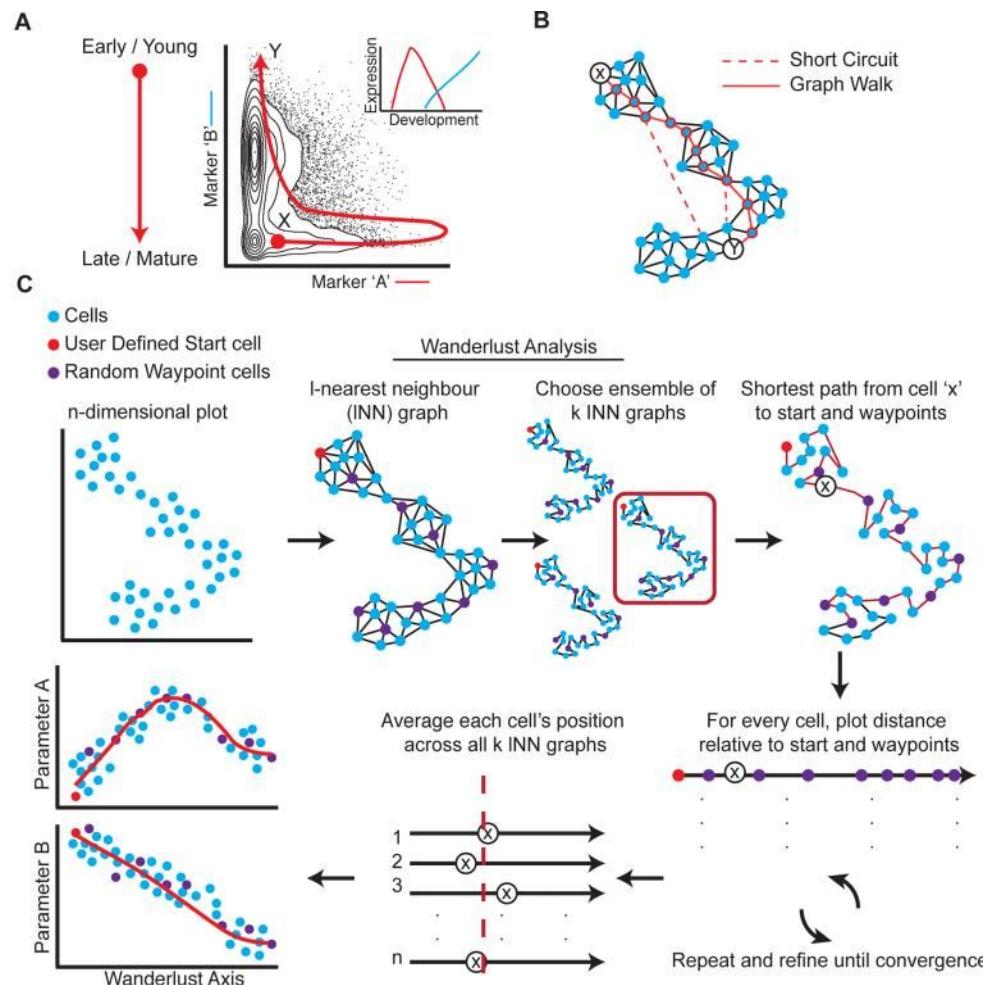
2016

2018

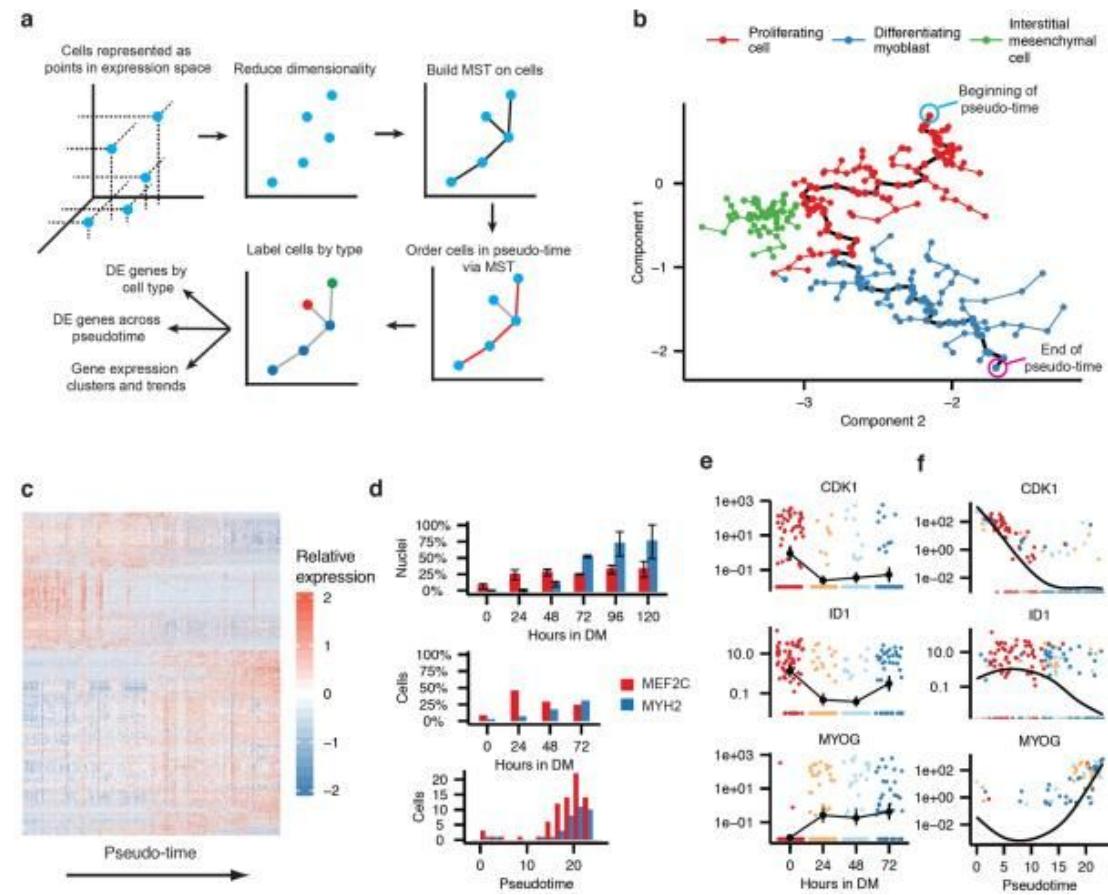
2019

2020

Wanderlust



Monocle (v1)



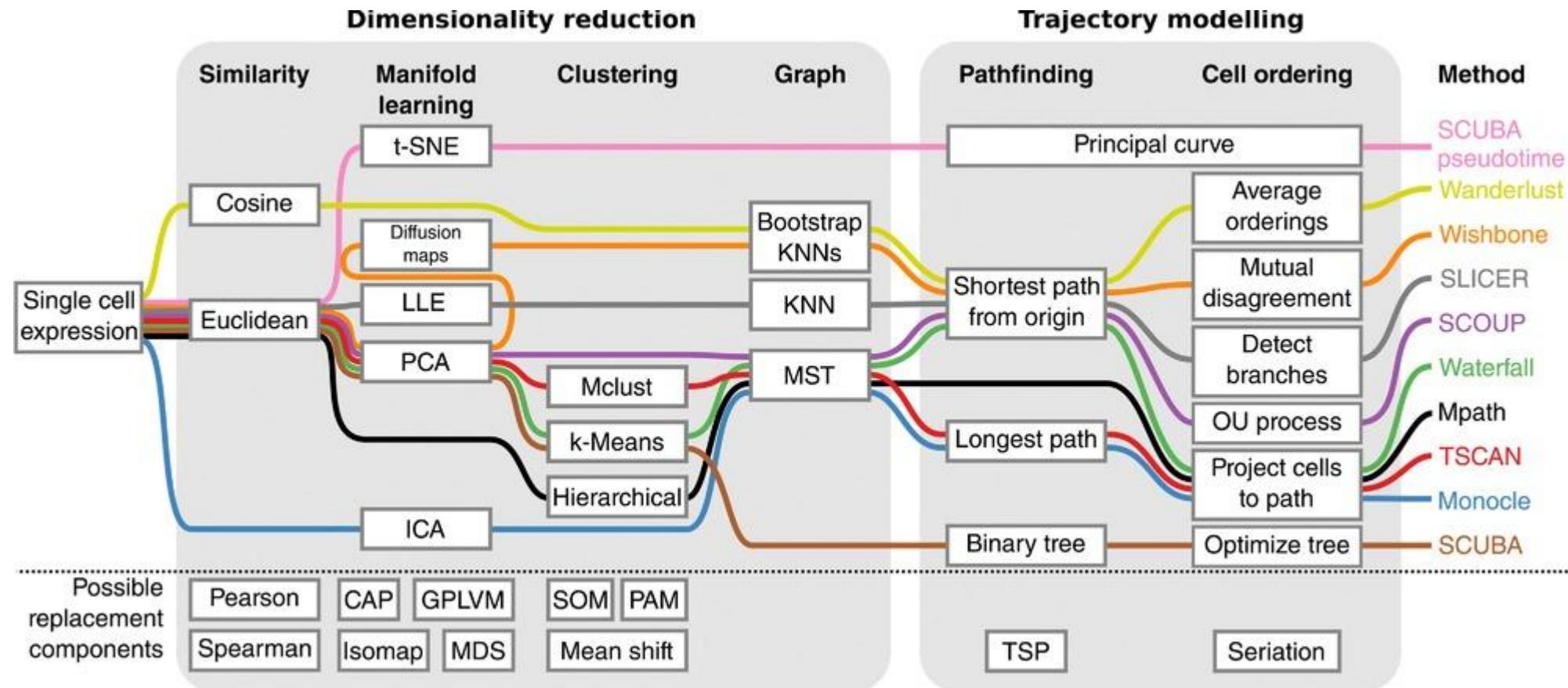
2014

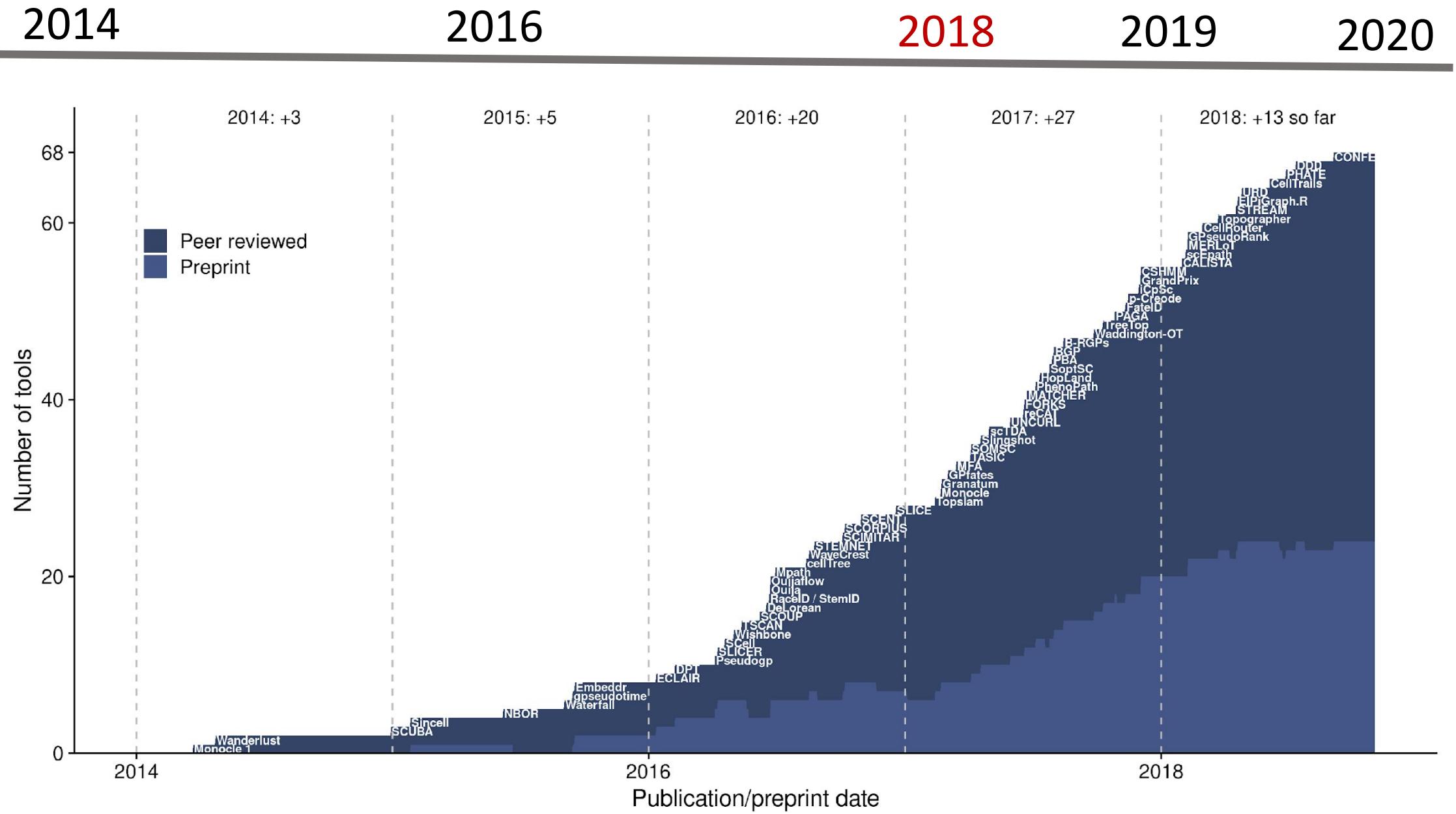
2016

2018

2019

2020





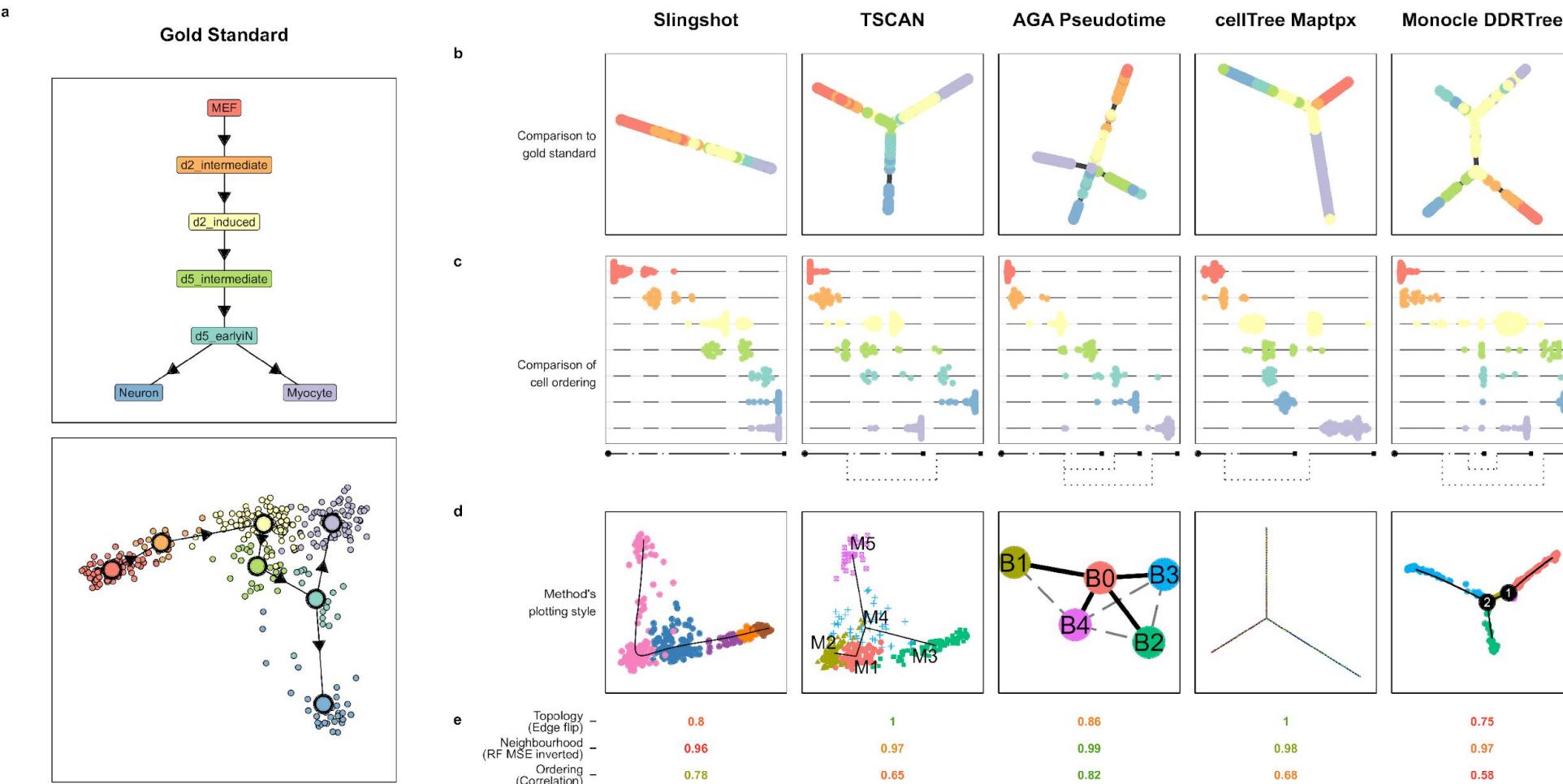
2014

2016

2018

2019

2020



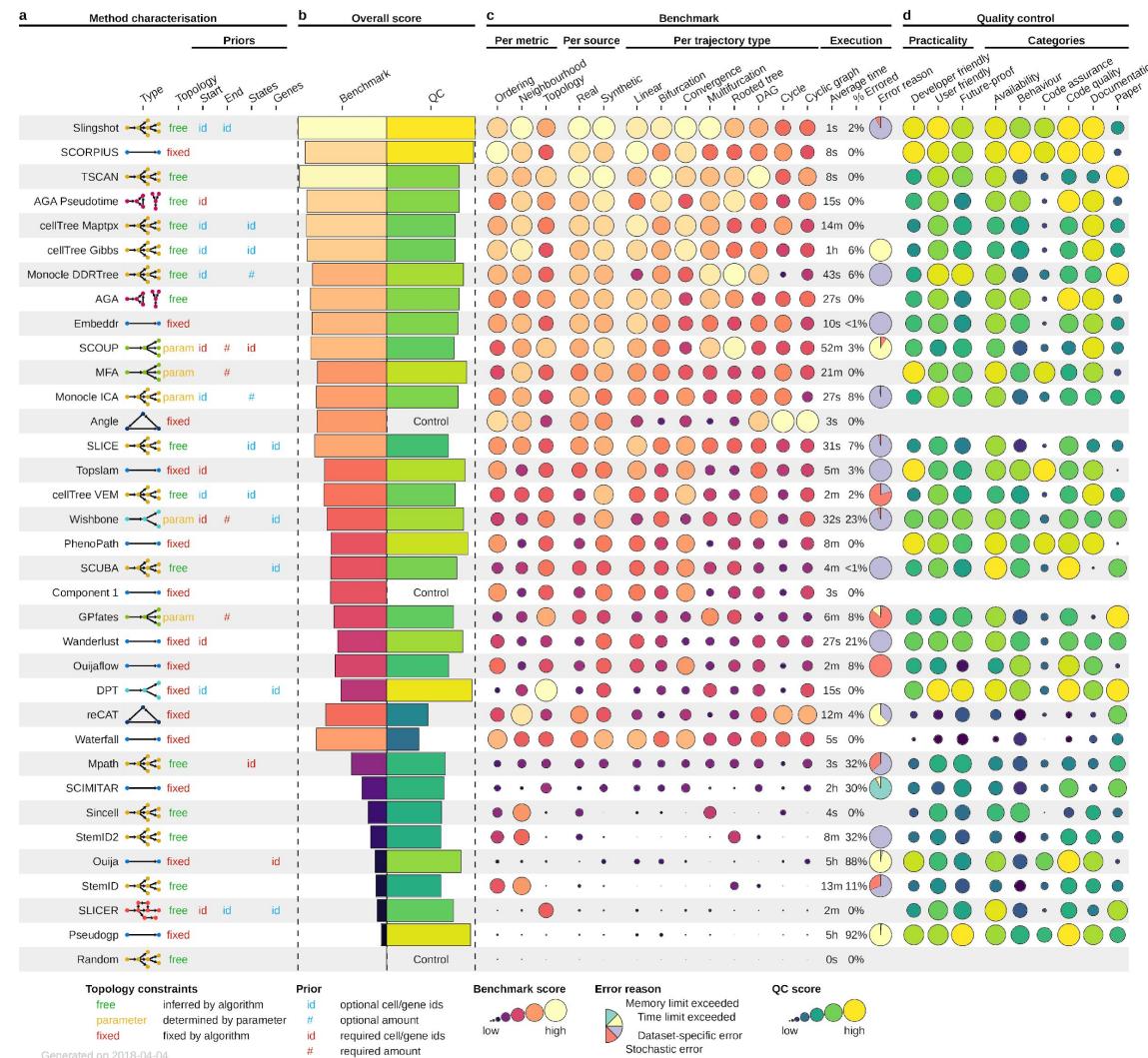
2014

2016

2018

2019

2020



2014

2016

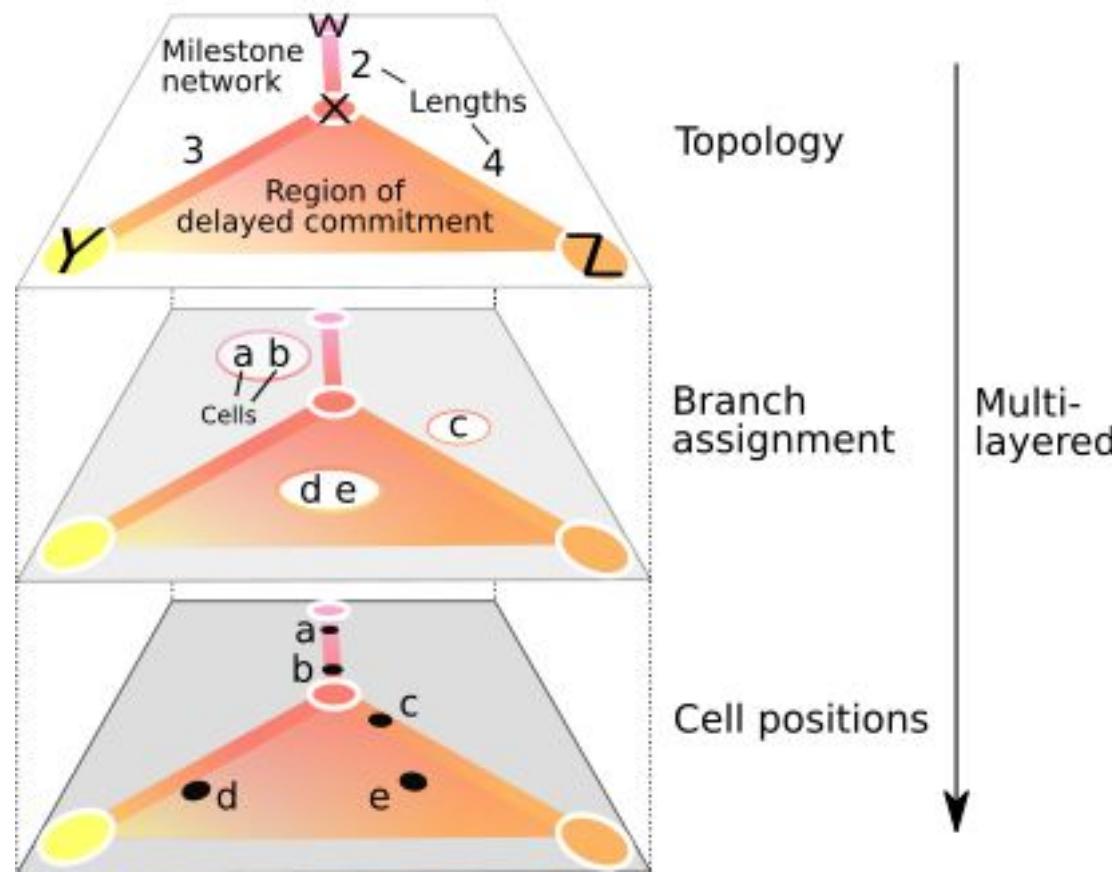
2018

2019

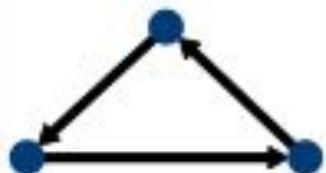
2020

110 real
& 229 synthetic datasets + 45 trajectory inference methods + 4 metrics

Common probabilistic trajectory model



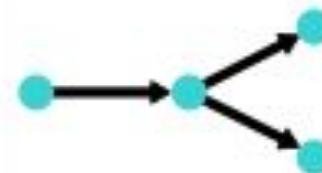
Main difference: trajectory types



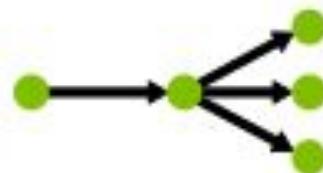
Cycle



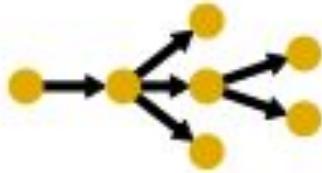
Linear



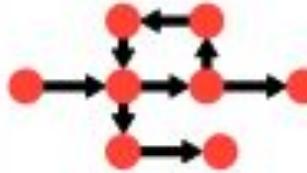
Bifurcation



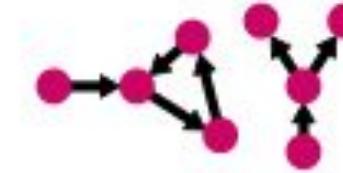
Multifurcation



Tree

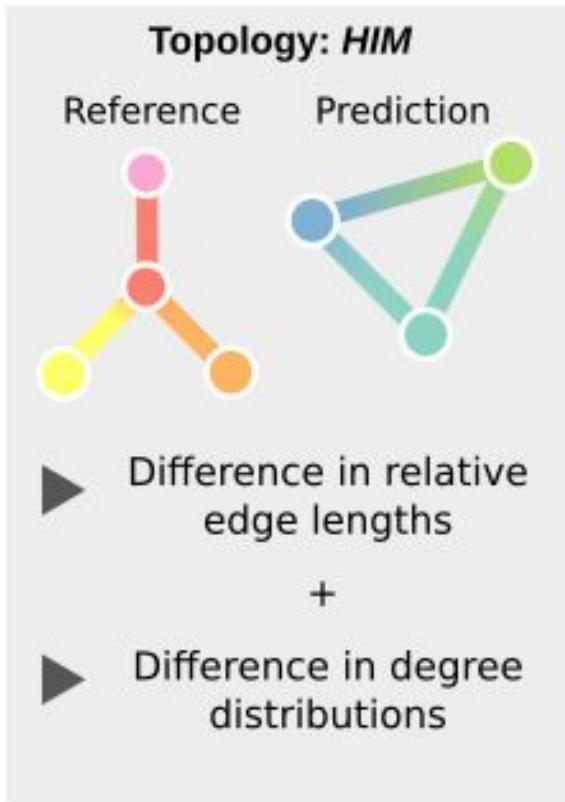


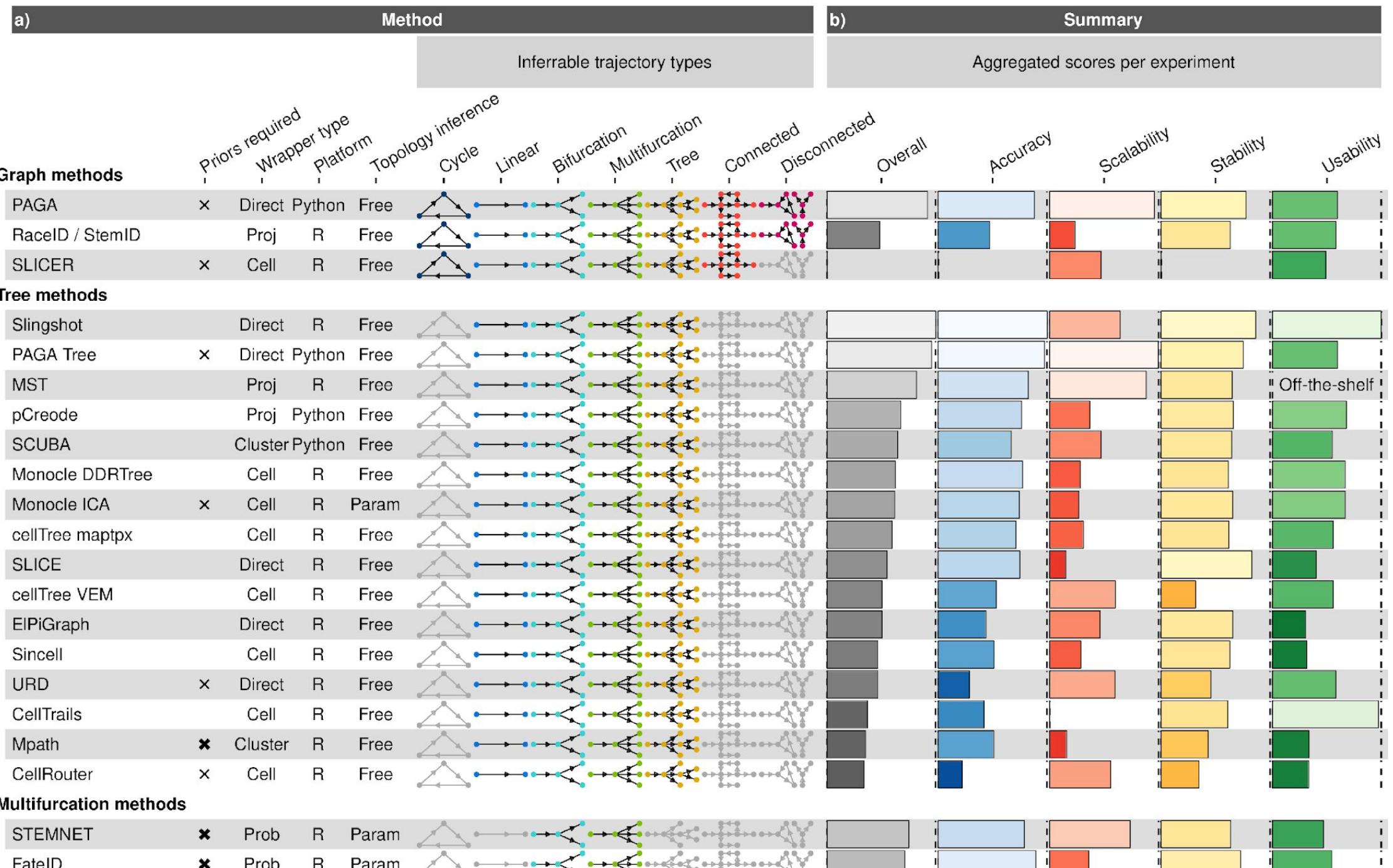
Connected
graph

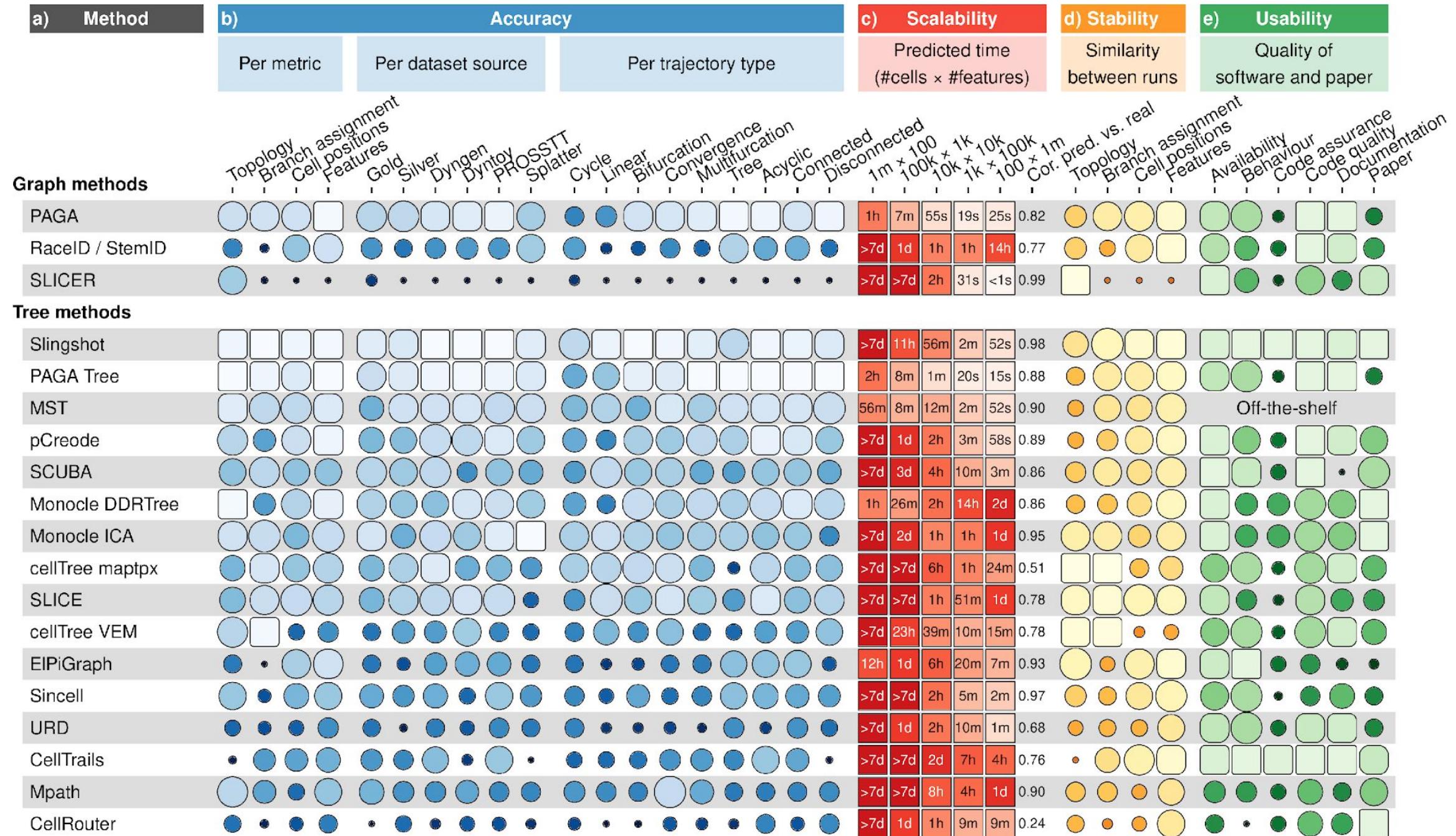


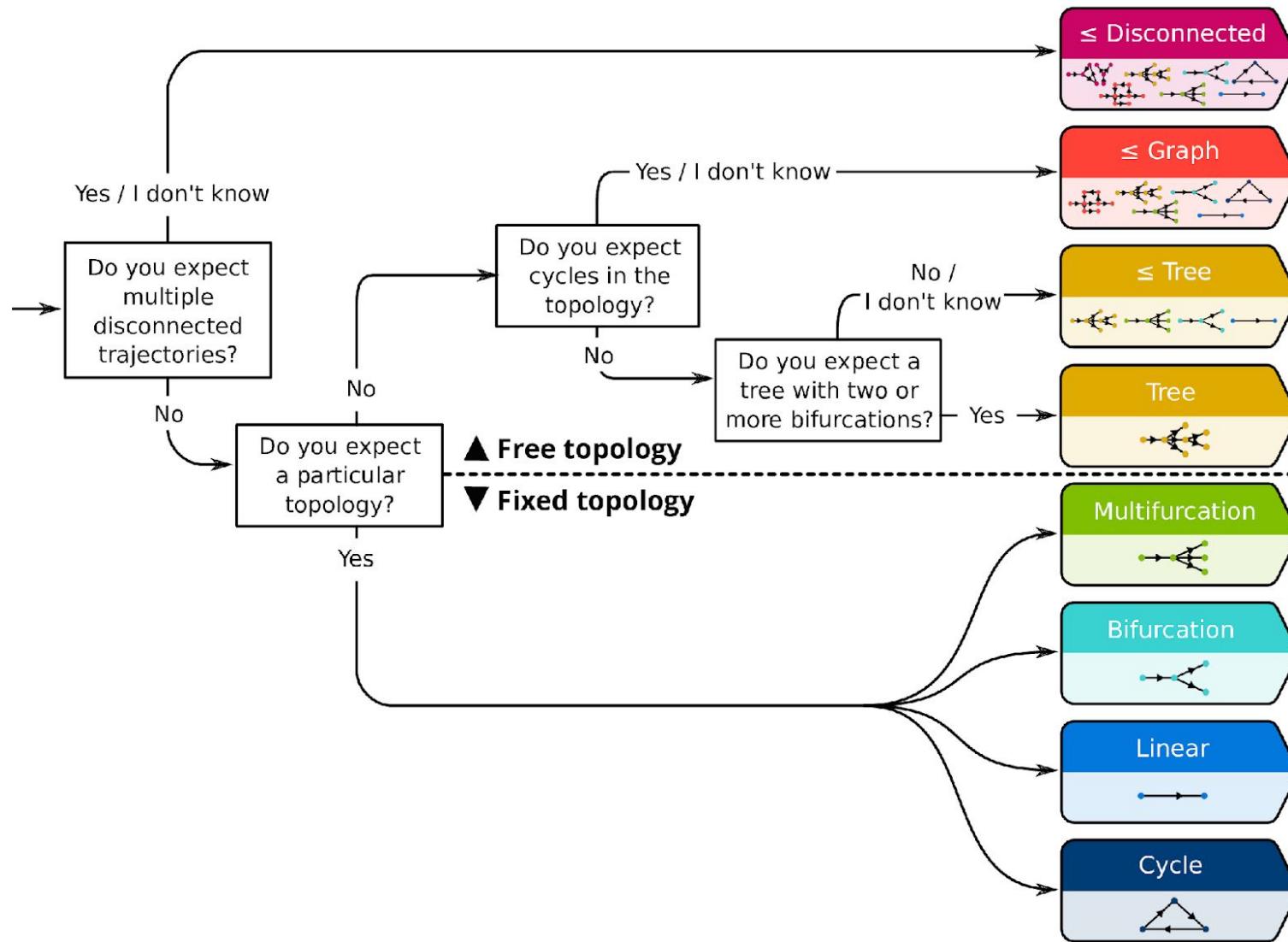
Disconnected
Graph

How to assess accuracy

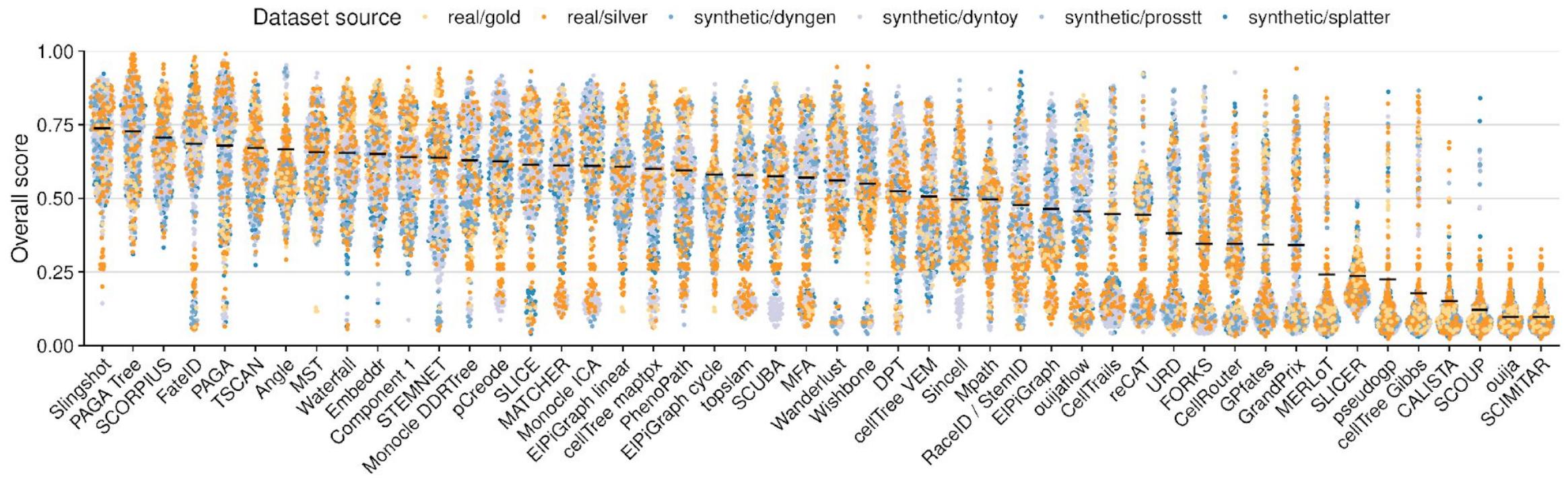




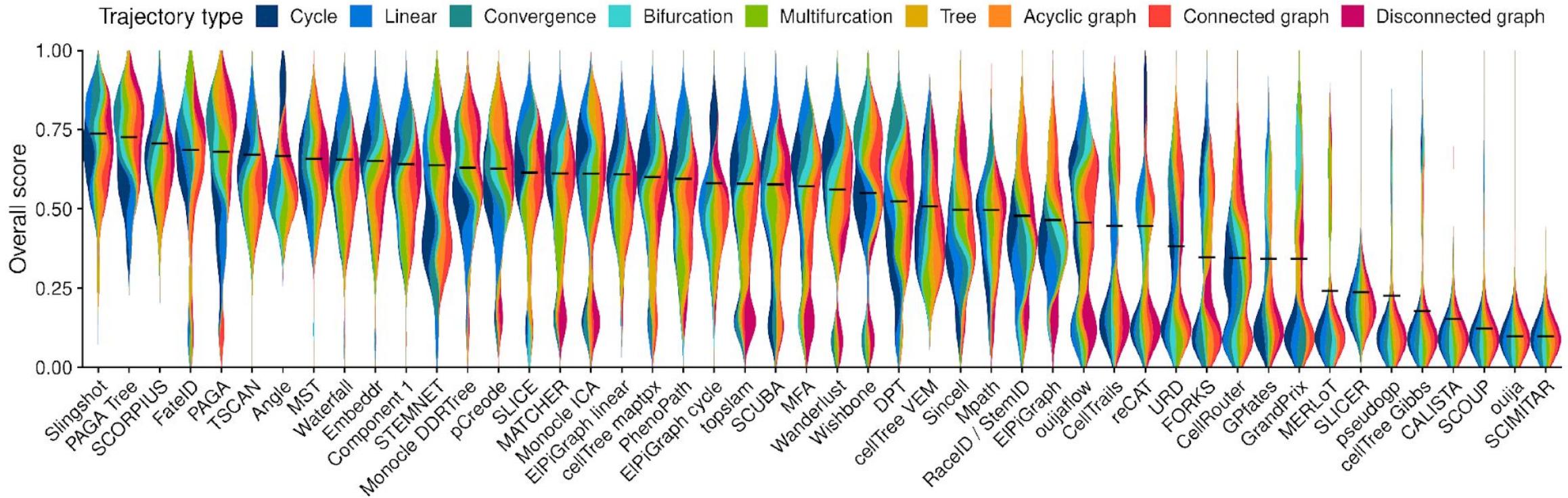




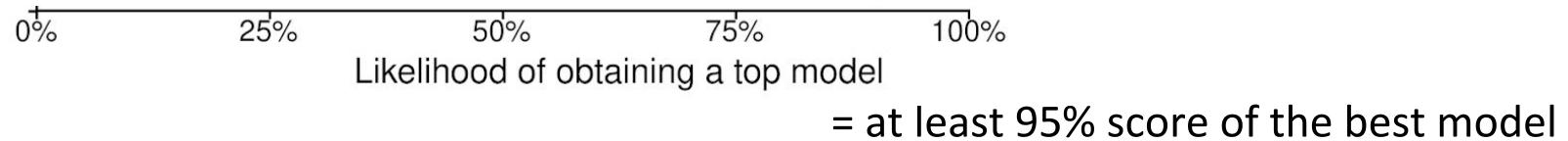
Large variability in scores

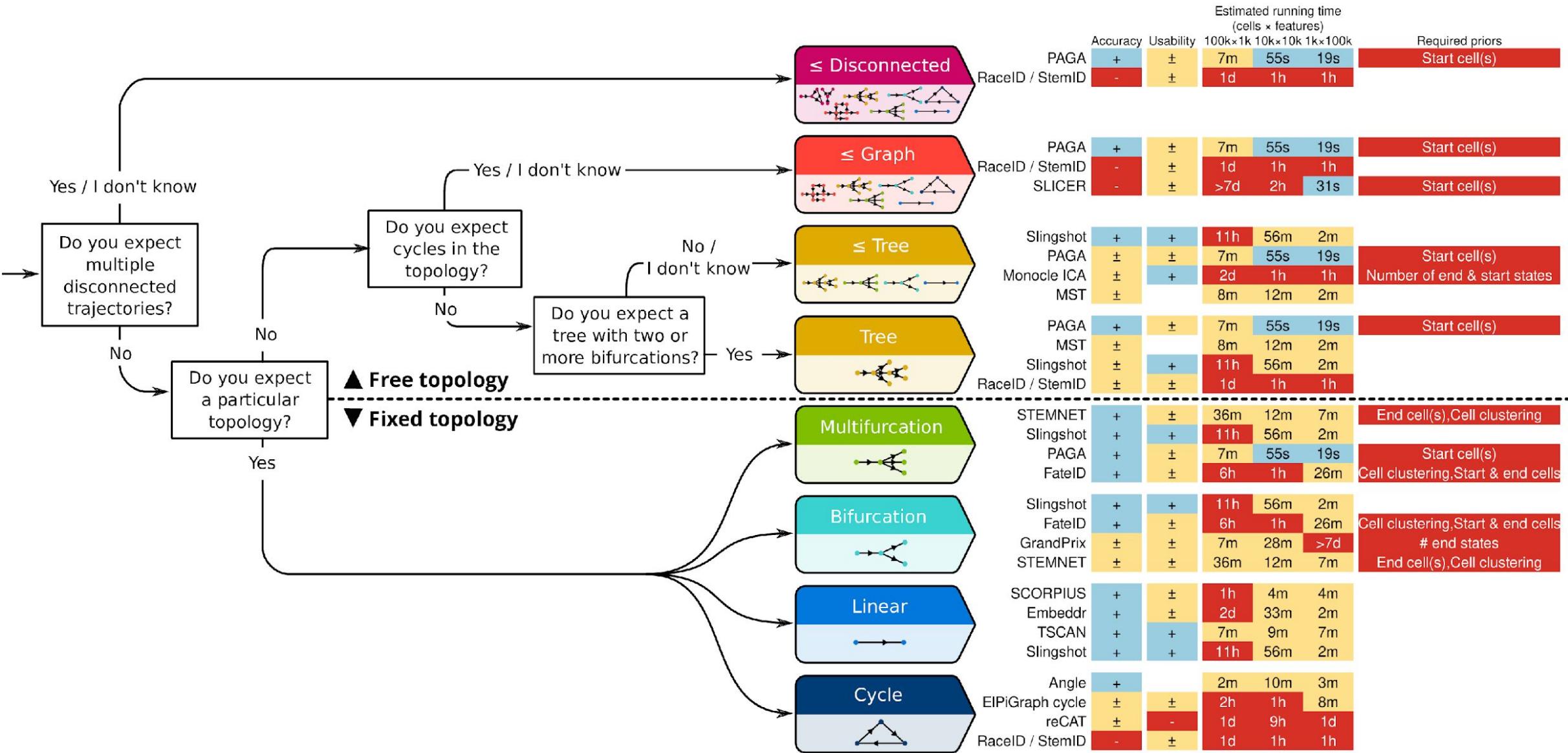


Large variability in scores

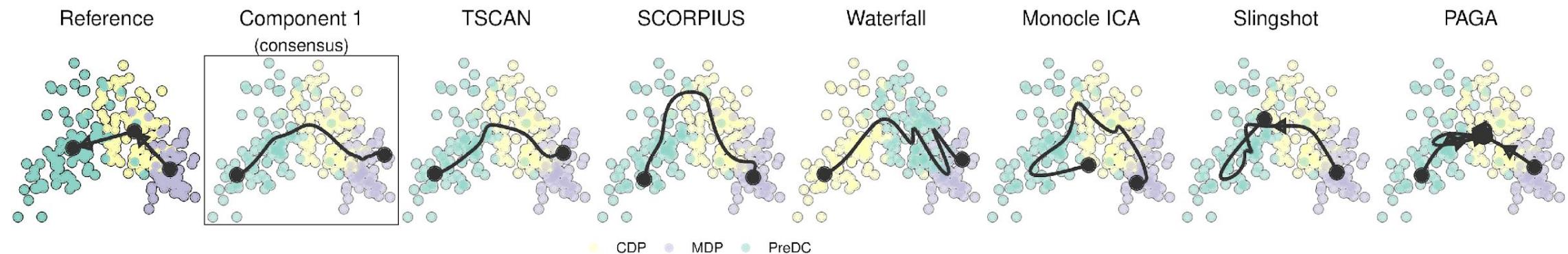


Complementarity

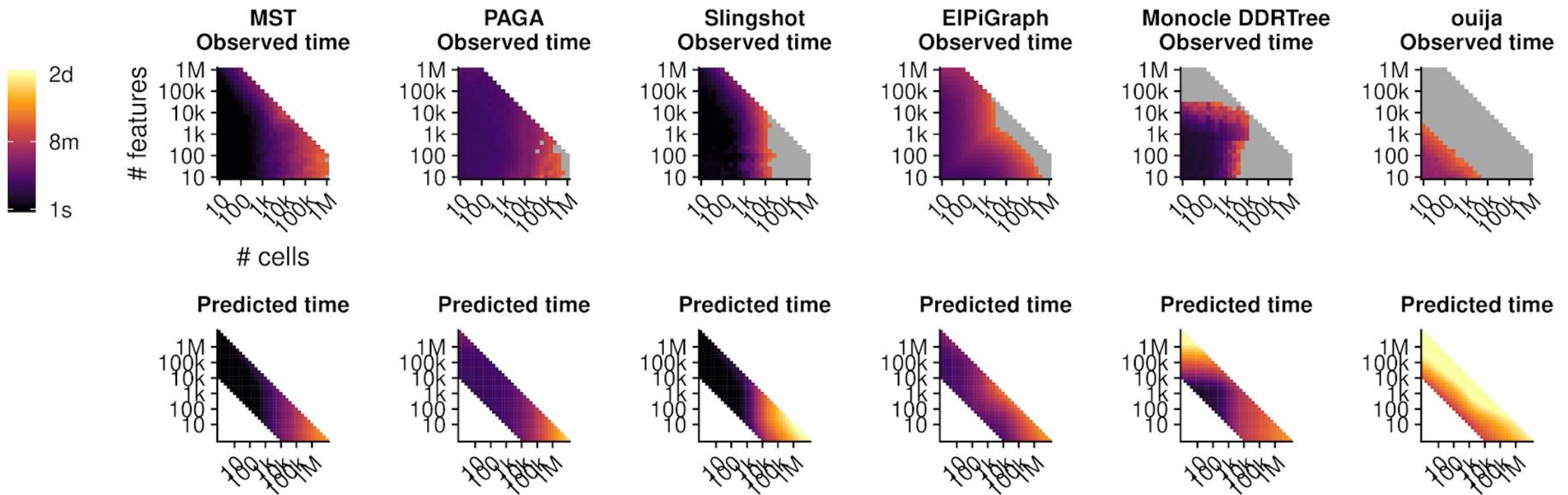


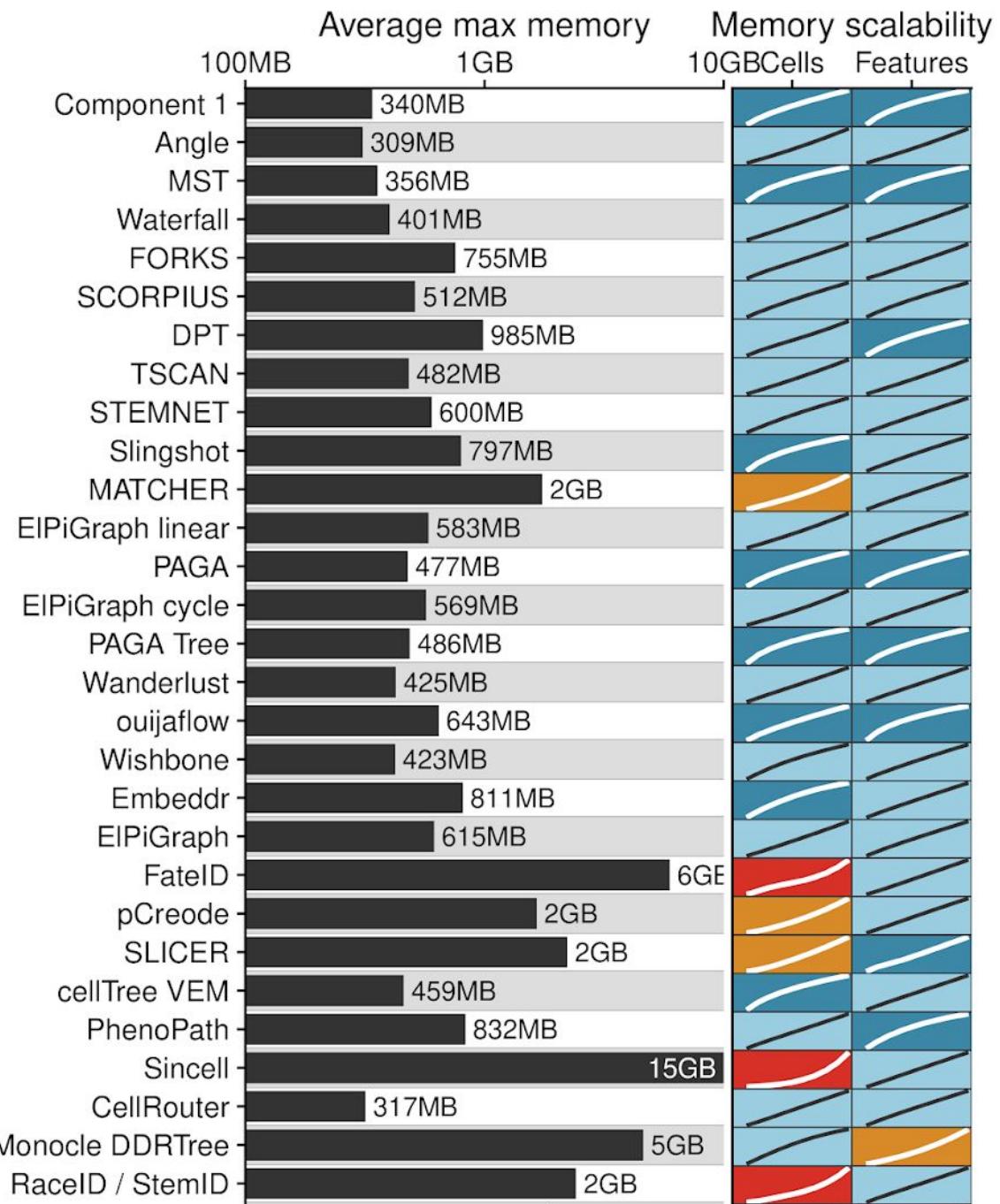
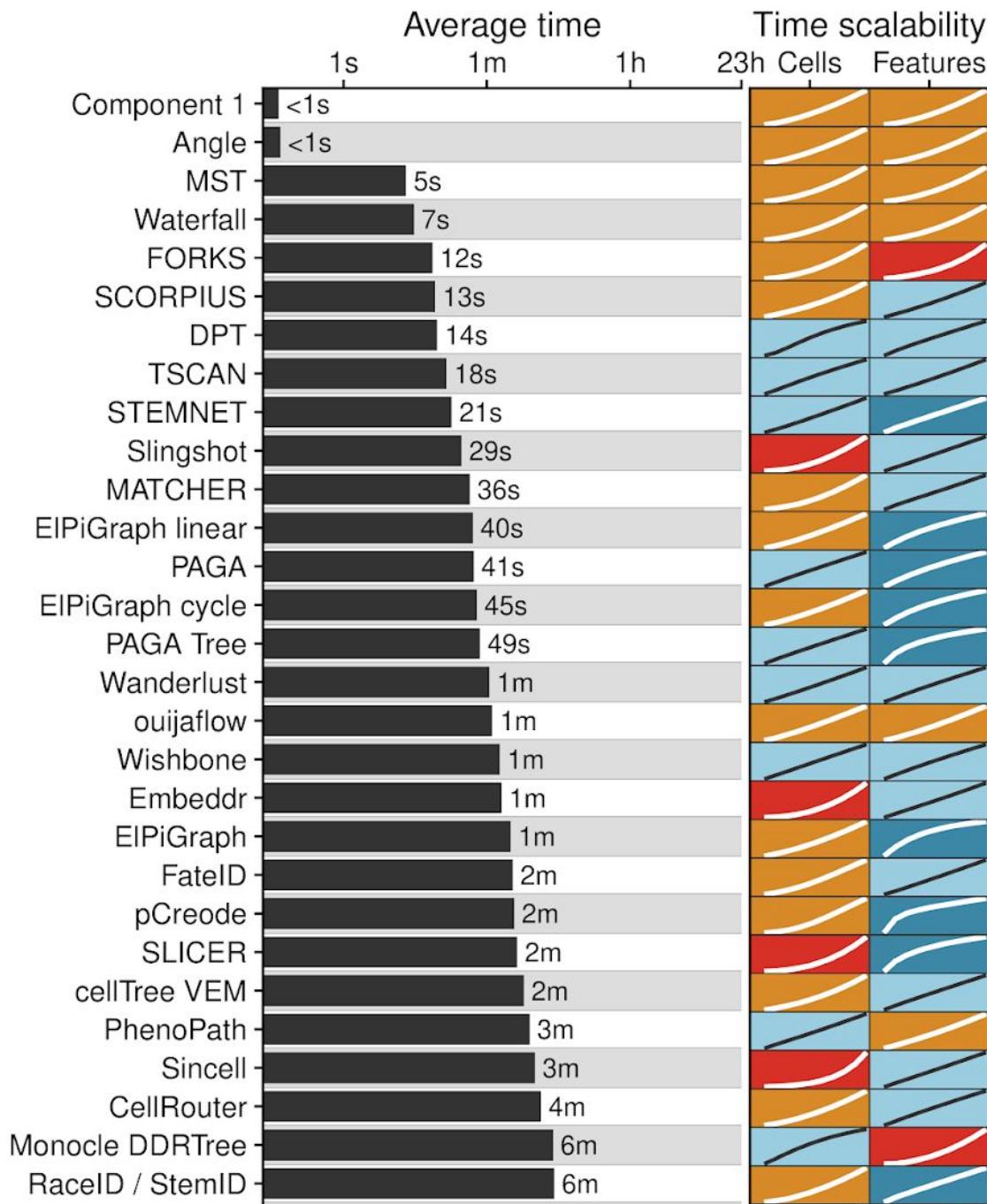


a



Scalability



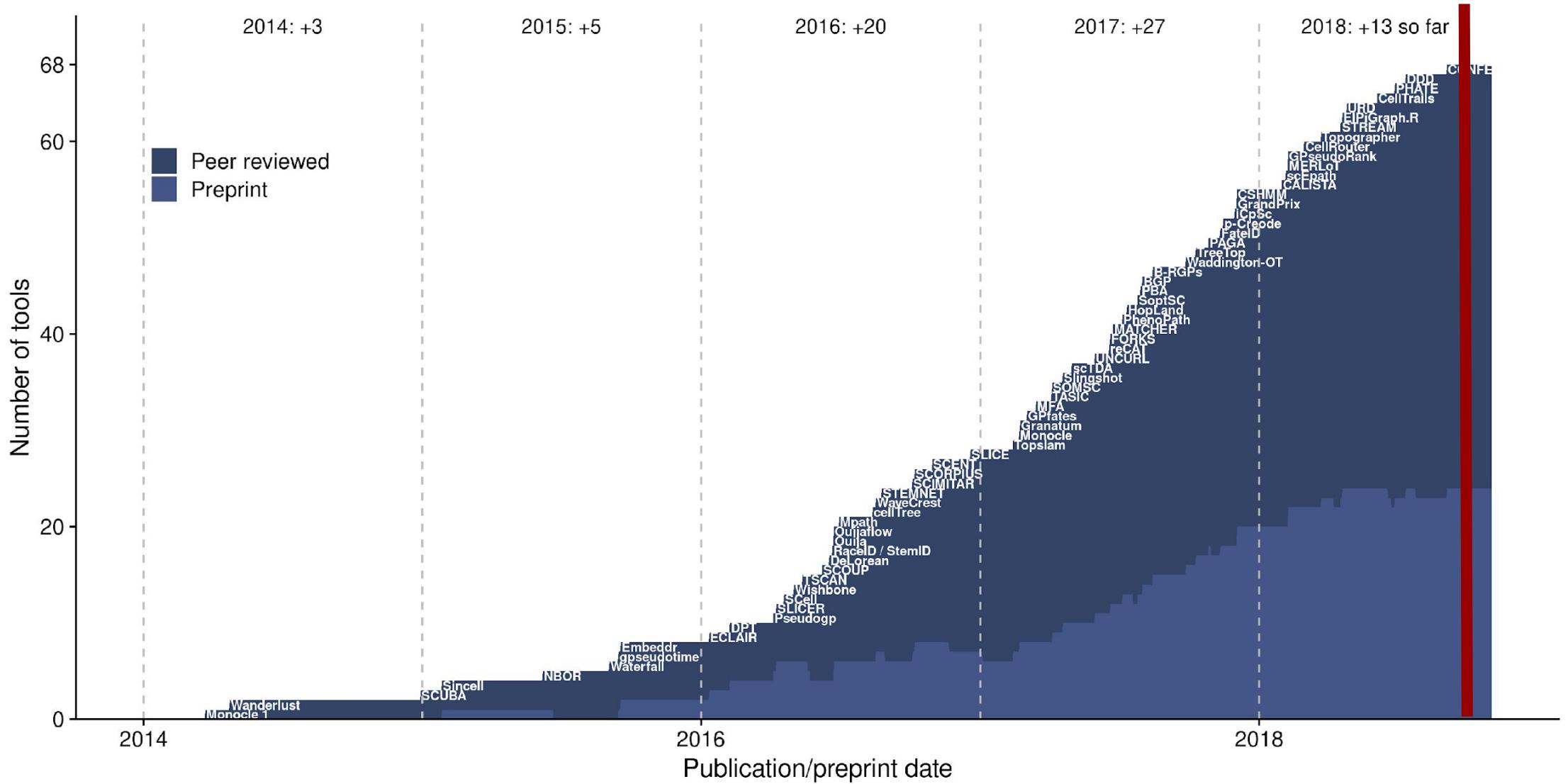


* Advertisement break *

- Clustering?
 - A systematic performance evaluation of clustering methods for single-cell RNA-seq data. <https://doi.org/10.12688/f1000research.15666.1>
 - Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. <https://doi.org/10.12688/f1000research.15809.1>
- Normalization?
 - scRNA-seq mixology: towards better benchmarking of single cell RNA-seq protocols and analysis methods. <https://doi.org/10.1101/433102>
- Differential expression?
 - Bias, robustness and scalability in single-cell differential expression analysis. <https://doi.org/10.1038/nmeth.4612>

* Advertisement break *

First independent benchmark



* Advertisement break *

- **Rapidly outdated**

What if a new method comes along next week which *outperforms everything*?

- **Hard to adapt and extend**

What if you want to add just one metric, would I need to *delve into this huge codebase* and start *rerunning everything*?

- **Authoritative**

What with *alternative* but *sensible* interpretations?

- **Published too late**

It should have been published *before 70 methods were created* (each with their own small benchmark)



komparo: Continuous and collaborative benchmarking in computational biology using modern software technologies



How to analyze single-cell dynamics

Current state of trajectory inference

Some internal methodology

Extensions

75+ methods, but appearances are deceptive

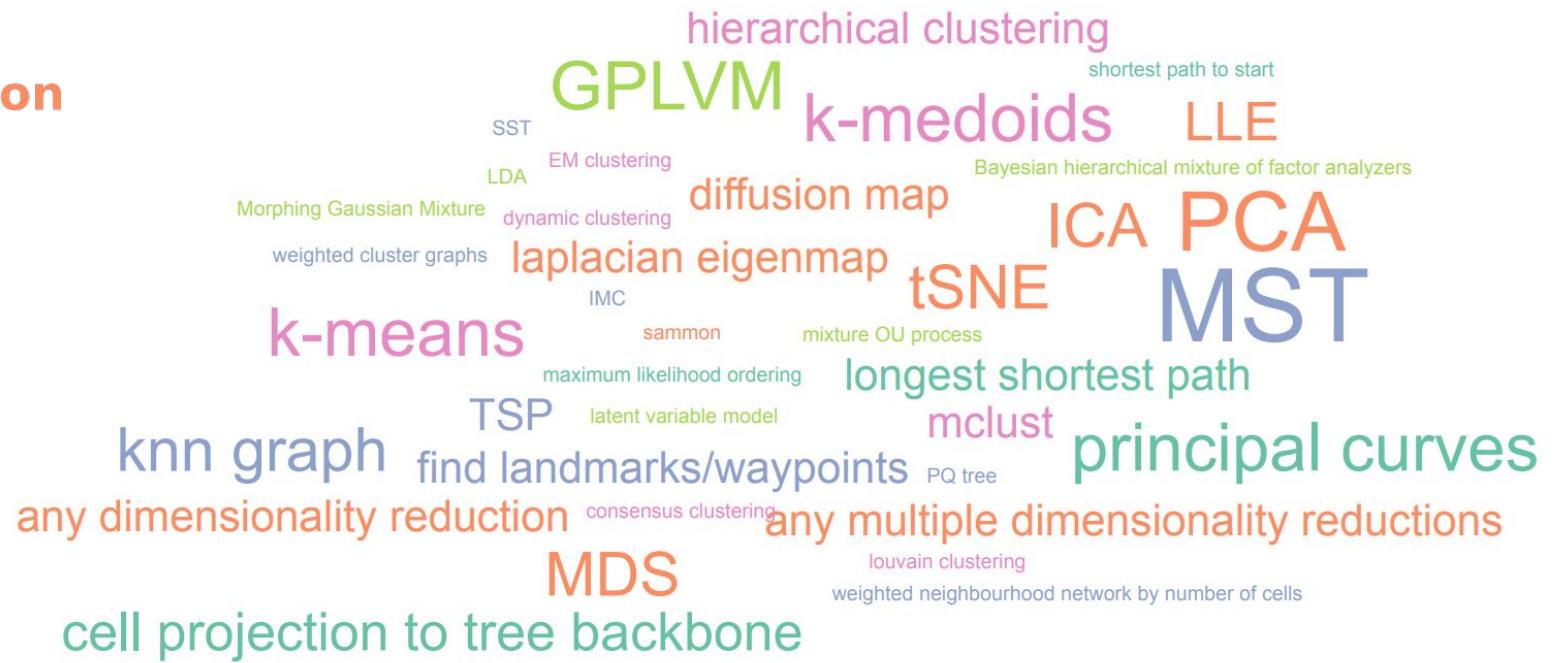
Dimensionality reduction

Clustering

Graph operation

Generative model

Ordering



Principal curves

Graph diffusion

KNN

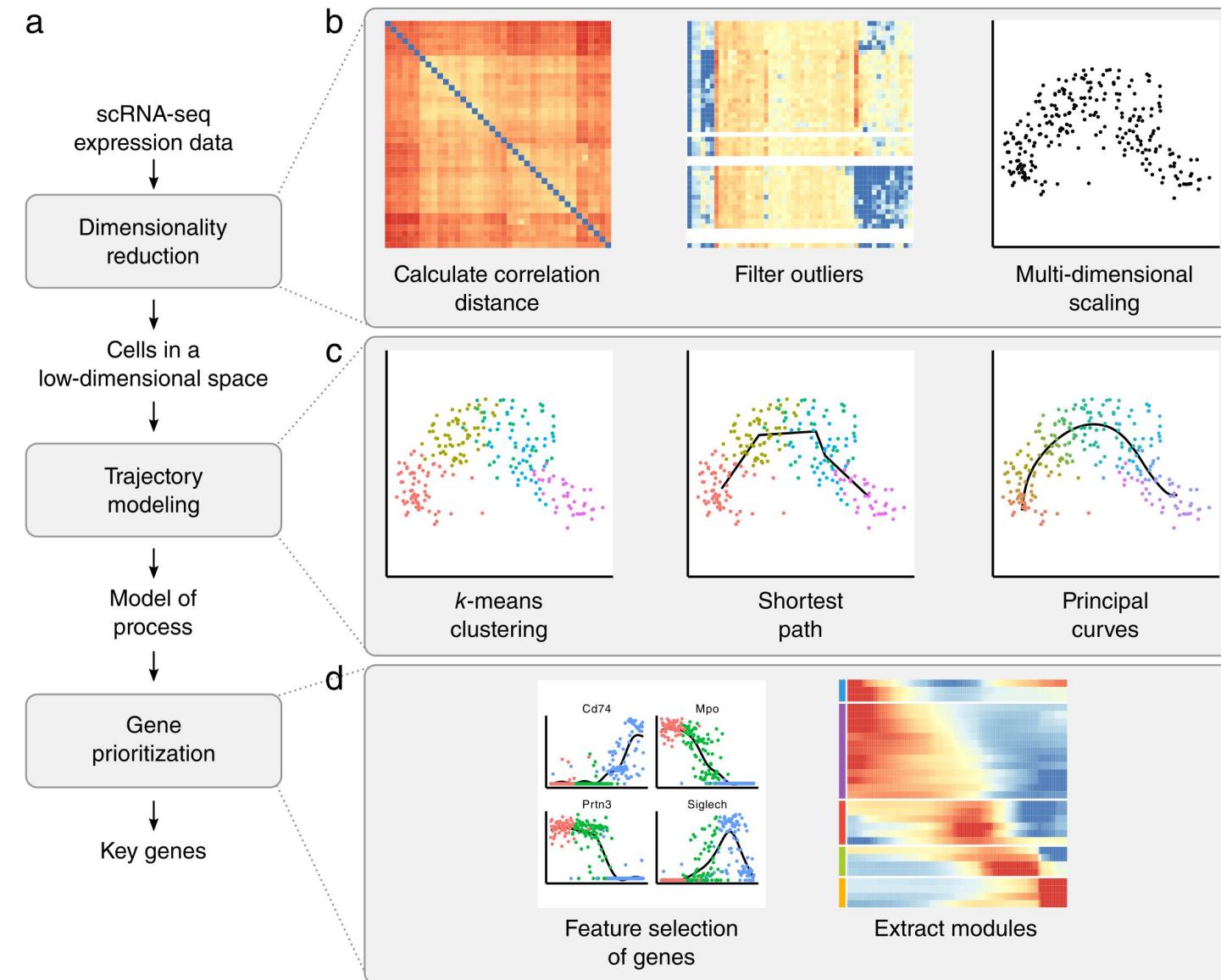
GPLVM

...

SCORPIUS

<https://doi.org/10.1101/079509>

Principal curves



Slingshot

<https://dx.doi.org/10.1186/s12864-018-4772-0>

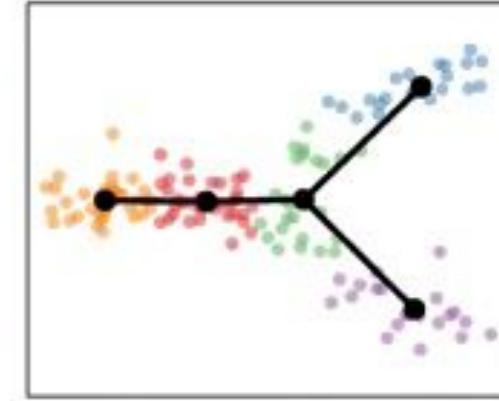
Principal curves

Dimensionality reduction & clustering

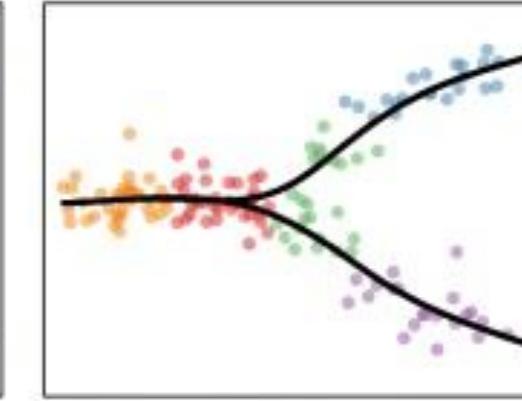
a



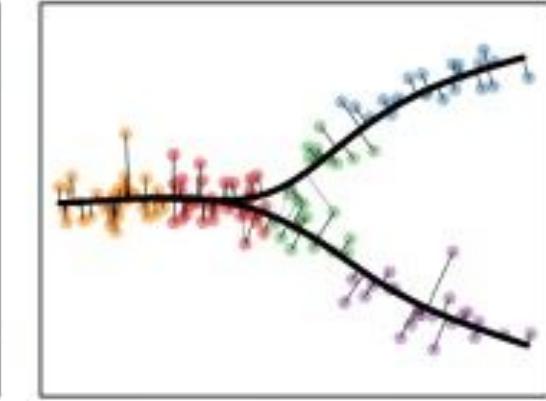
MST



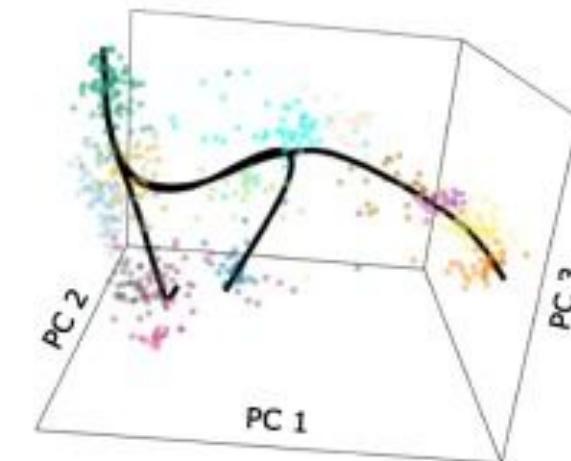
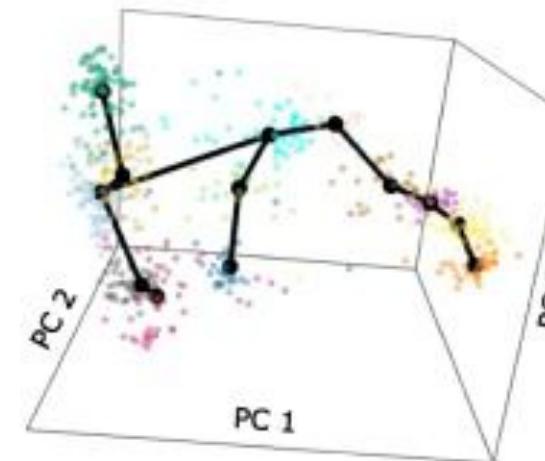
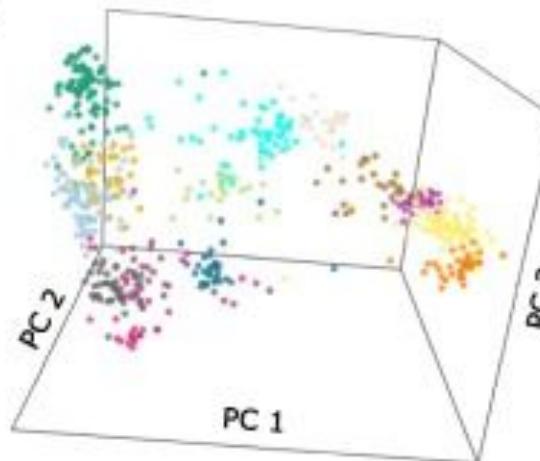
Principal curves



Projection onto curve

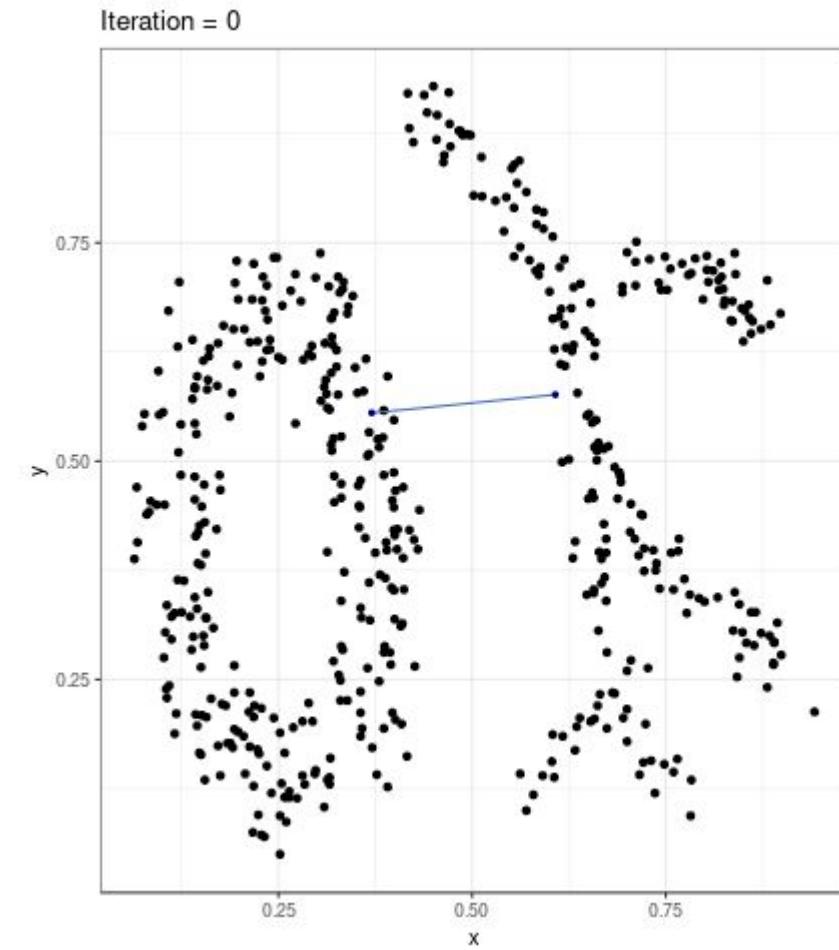
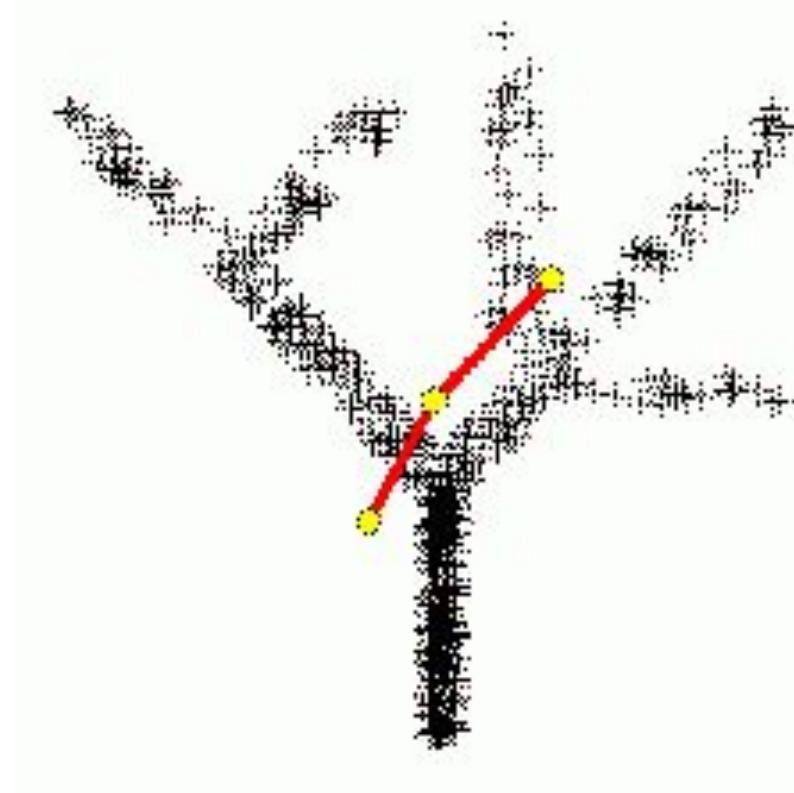


b



- HBC
- Transitioning HBC
- GBC
- Immature OSN
- Mature OSN
- Mature Sus
- Microvillous

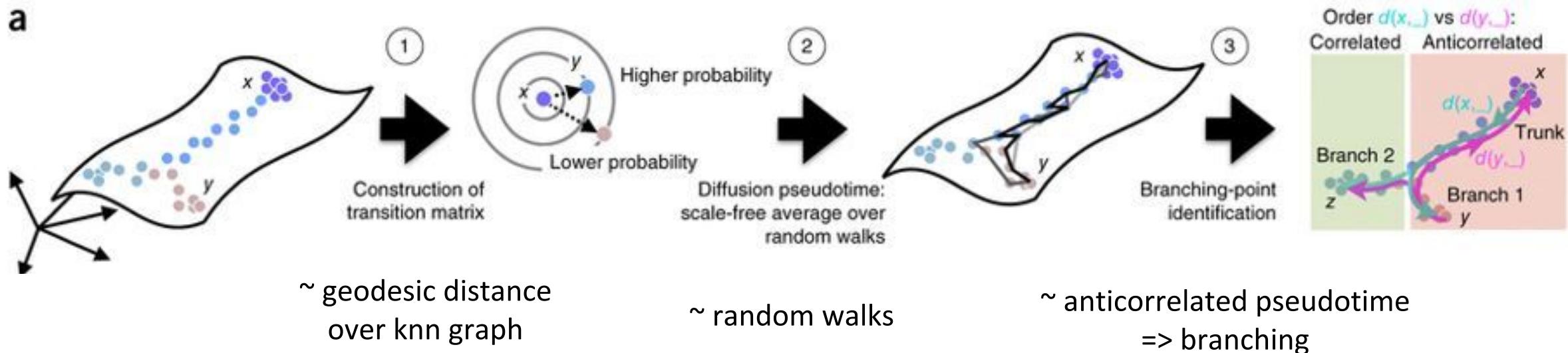
Principal graphs and neural gas



Diffusion pseudotime (DPT)

Graph diffusion

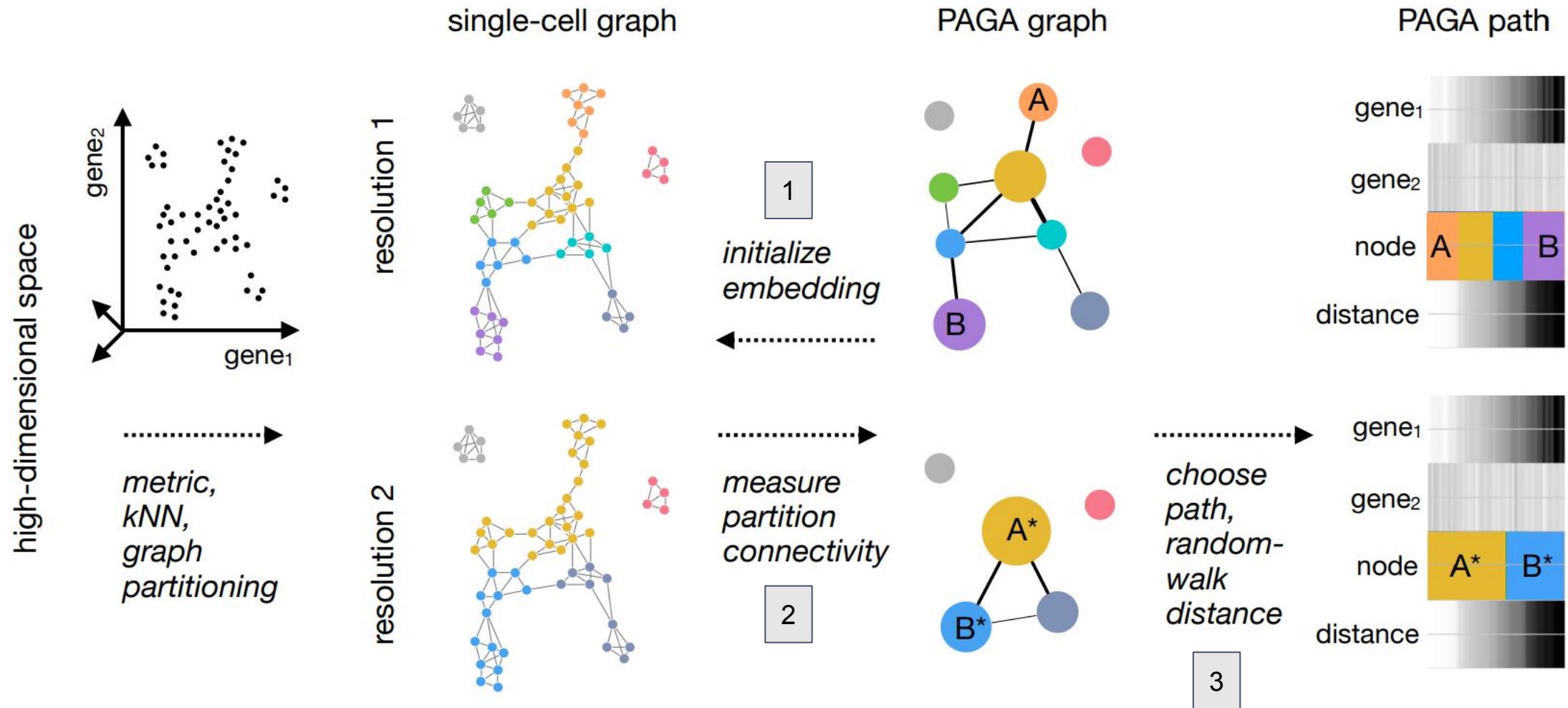
<https://doi.org/10.1038/nmeth.3971>



PAGA

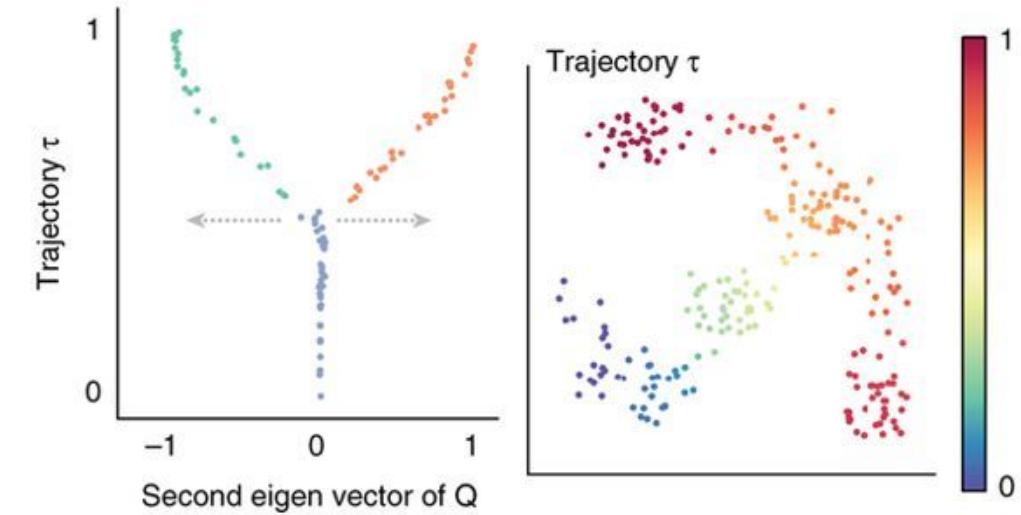
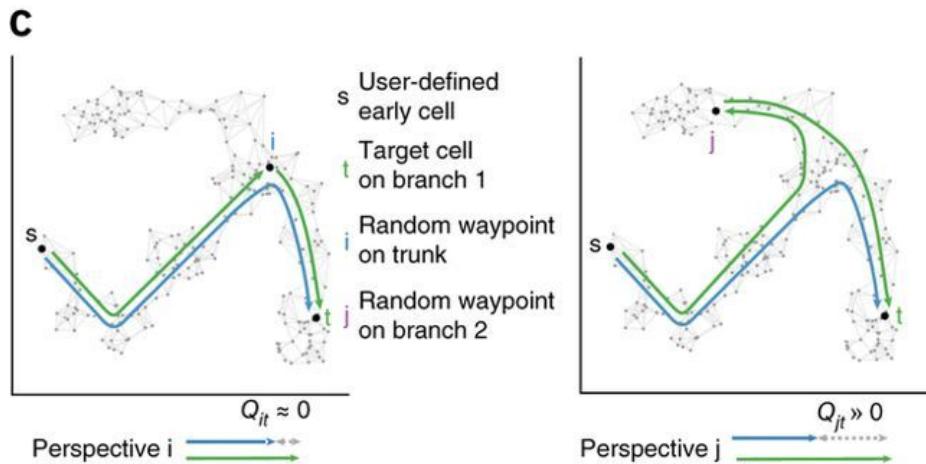
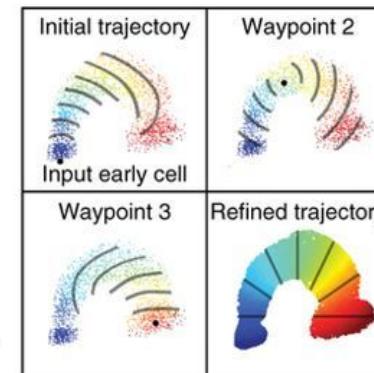
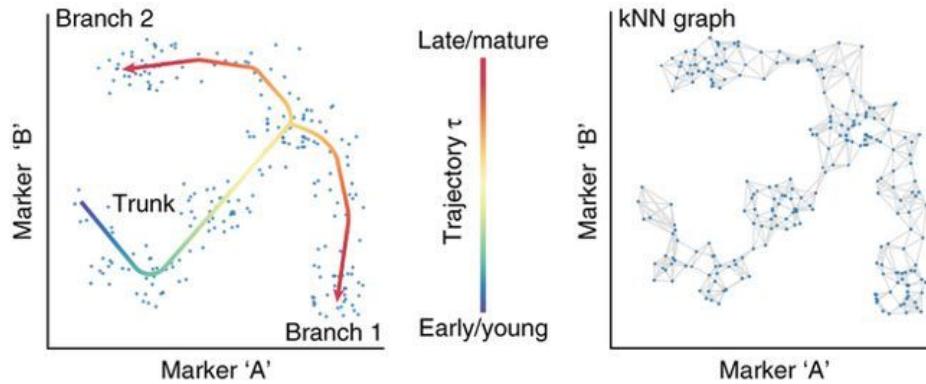
Graph diffusion

<https://doi.org/10.1101/208819>



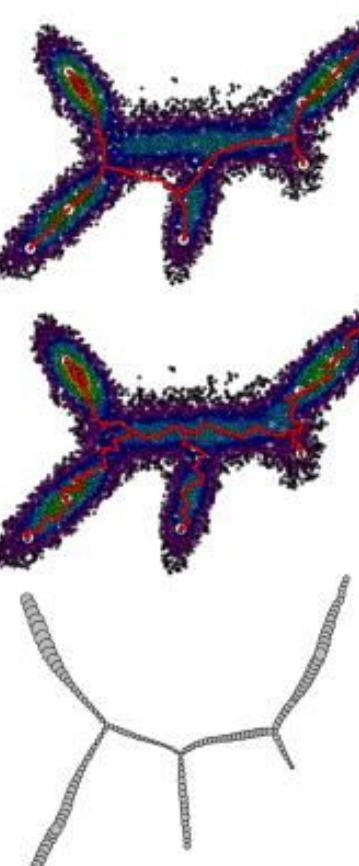
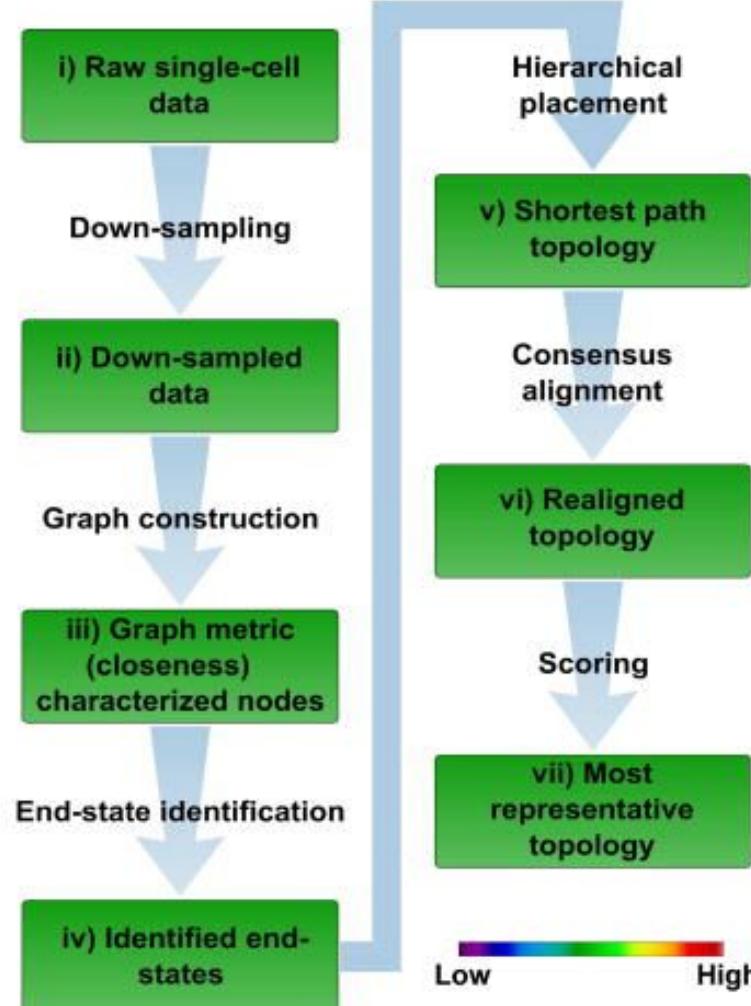
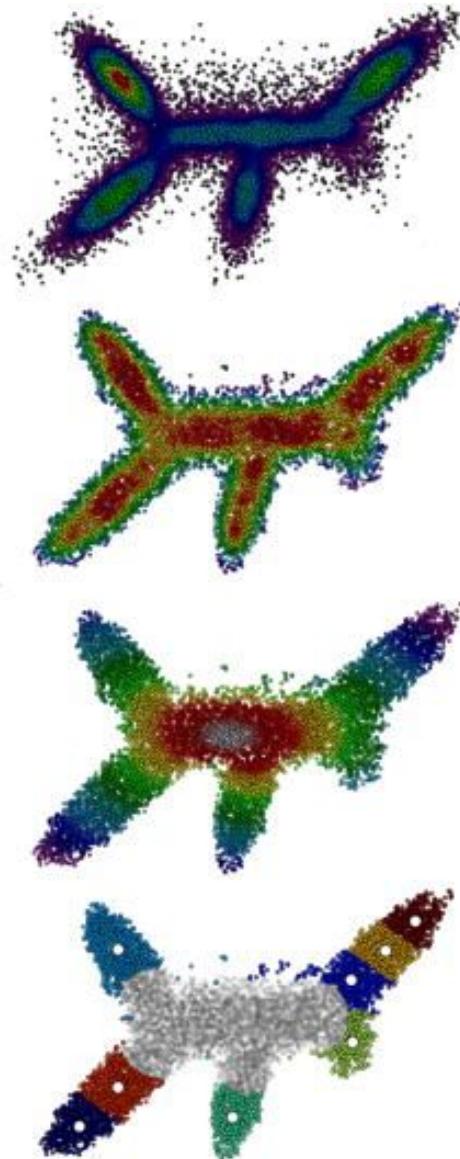
Wanderlust/Wishbone

<https://doi.org/10.1038/nbt.3569>



p-Creode

KNN



<https://doi.org/10.1016/j.cels.2017.10.012>

GrandPrix, GPfates, ...

- Model gene expression as a function of pseudotime: (B)GPLVM
 - Gaussian process: "Regression", but non-parametric.
I.e. do not assume a function but test all possible ones within constraints
 - Latent variable: The values of the variables you use are not yet known.
I.e. We do not know the pseudotime, even though we use it as input for regression
 - Bayesian: Include some prior information
I.e. Known experimental time points
- Branching?
 - Mixture GPLVM, with multiple functions over time

<https://doi.org/10.1126/sciimmunol.aal2192>

<http://dx.doi.org/10.1093/bioinformatics/bty533>

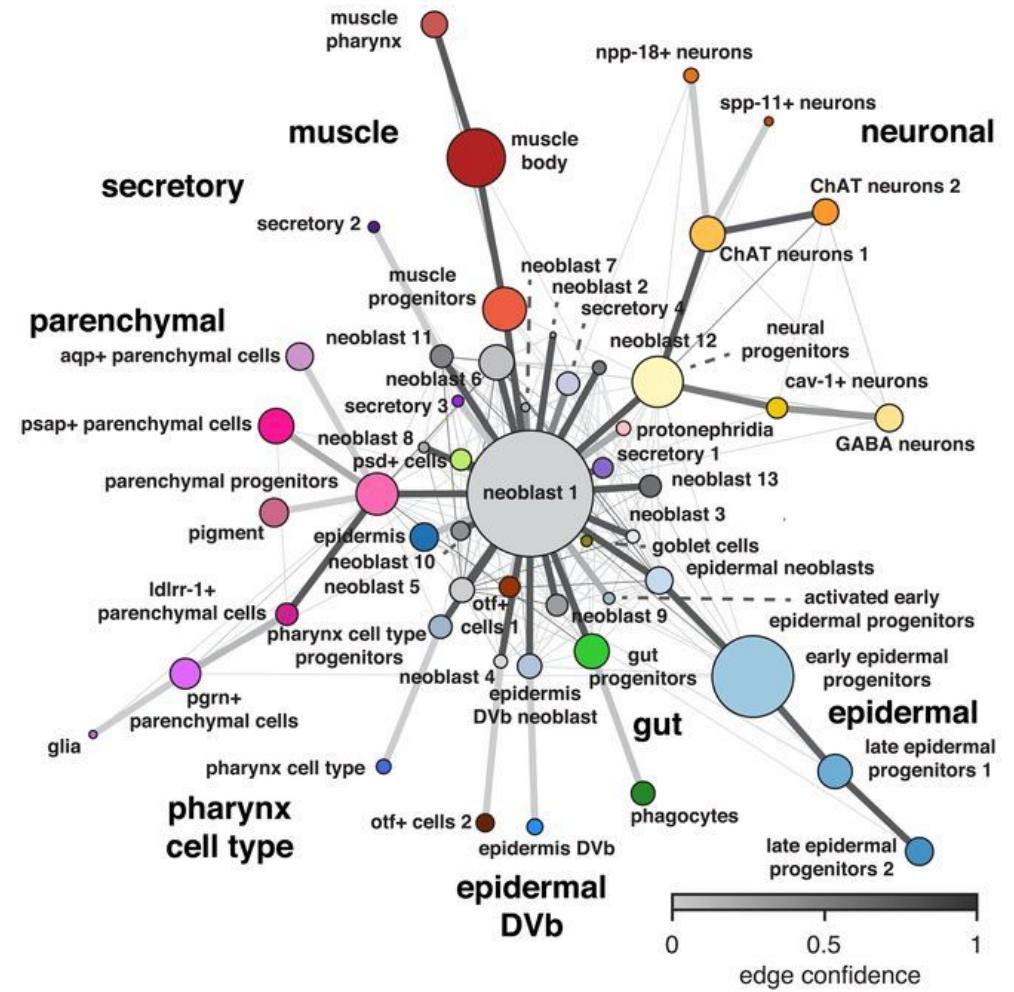
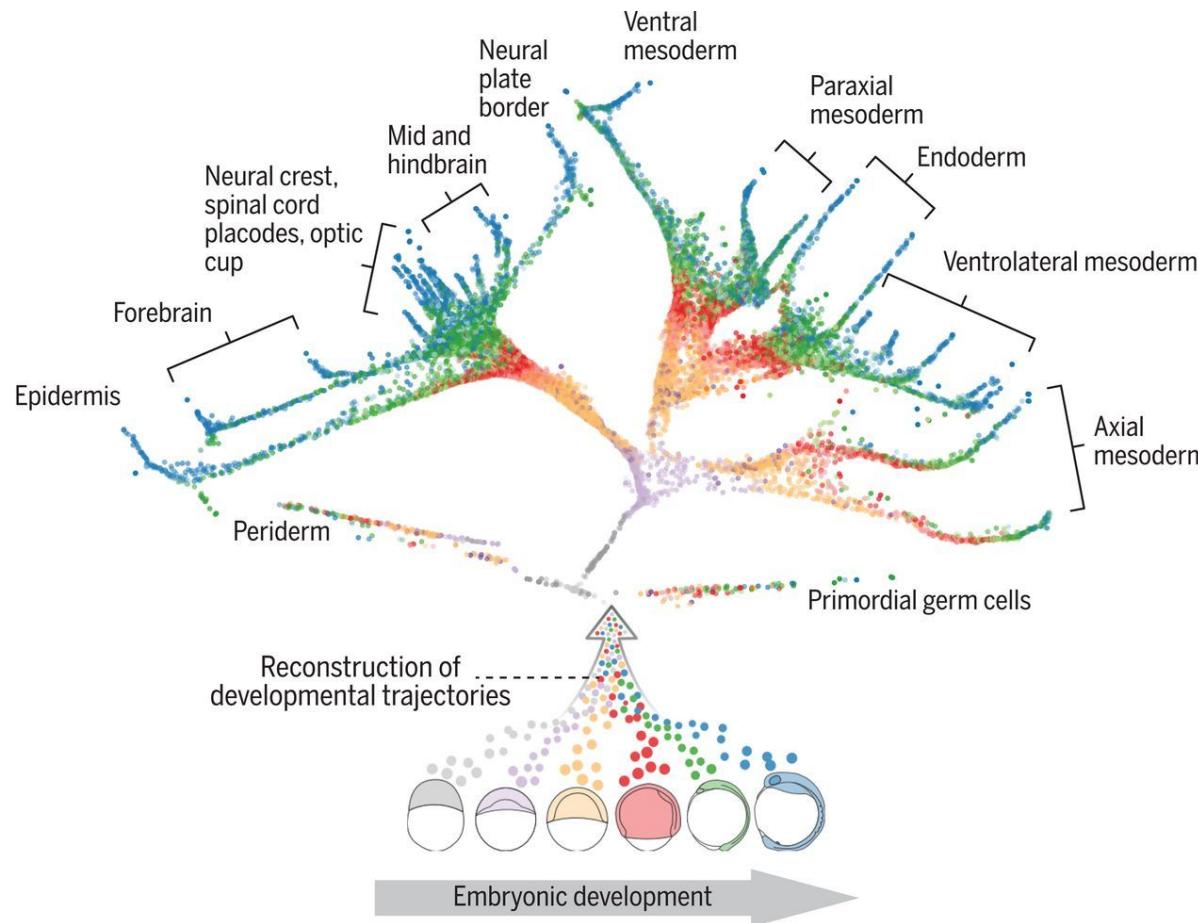
How to analyze single-cell dynamics

Current state of trajectory inference

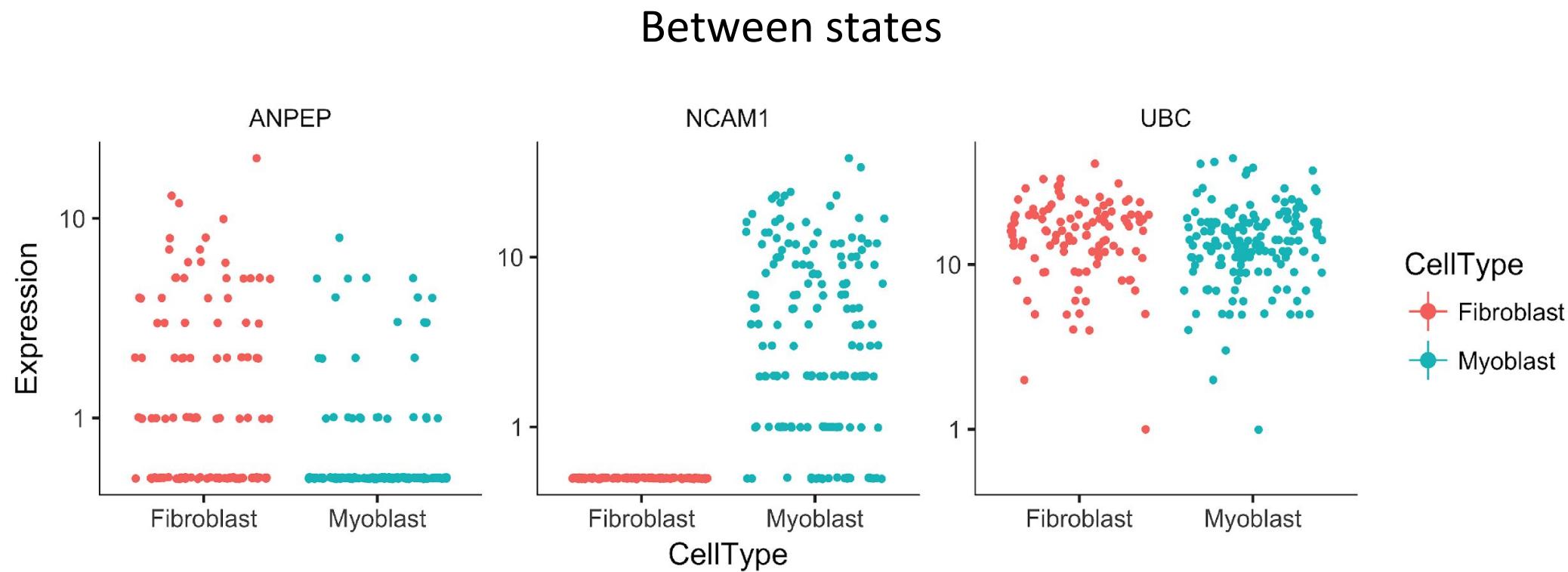
Some internal methodology

Extensions

Cautionary note

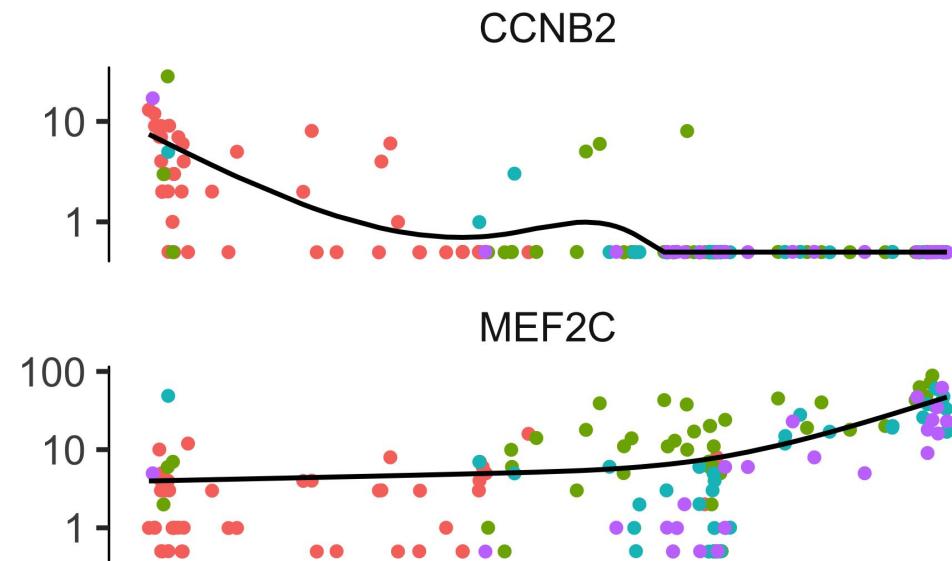


Trajectory differential expression

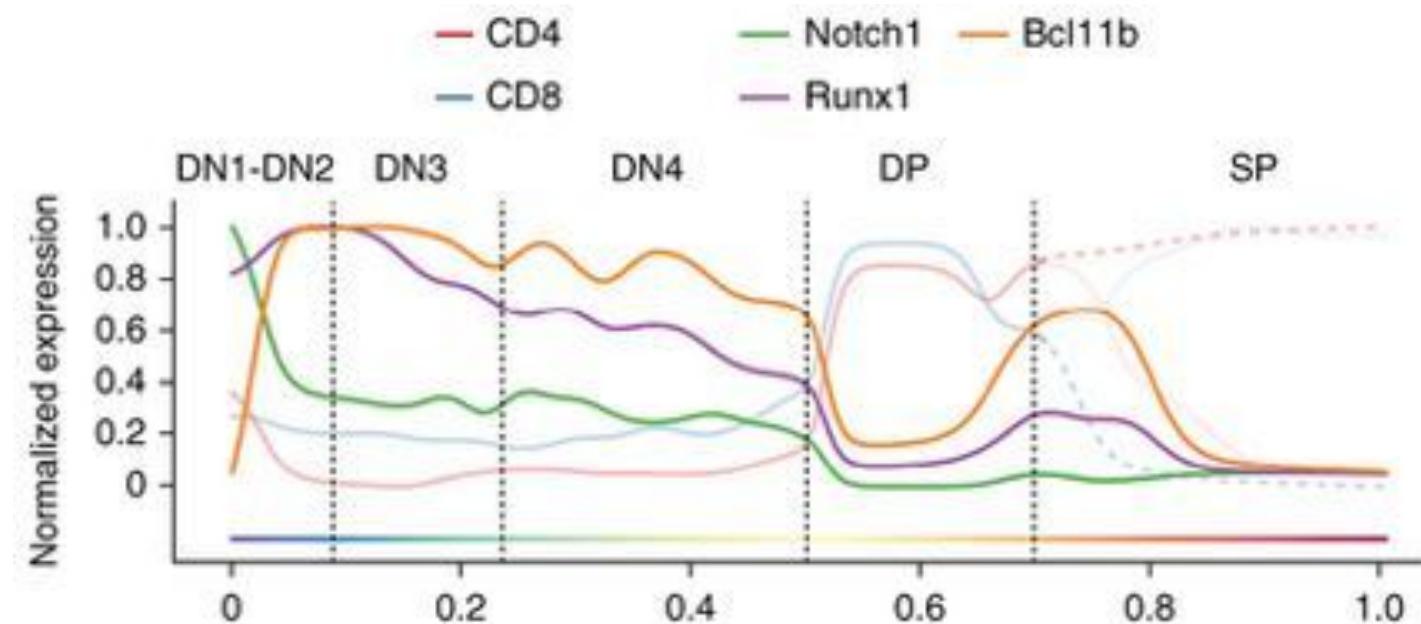


Trajectory differential expression

Monocle v2

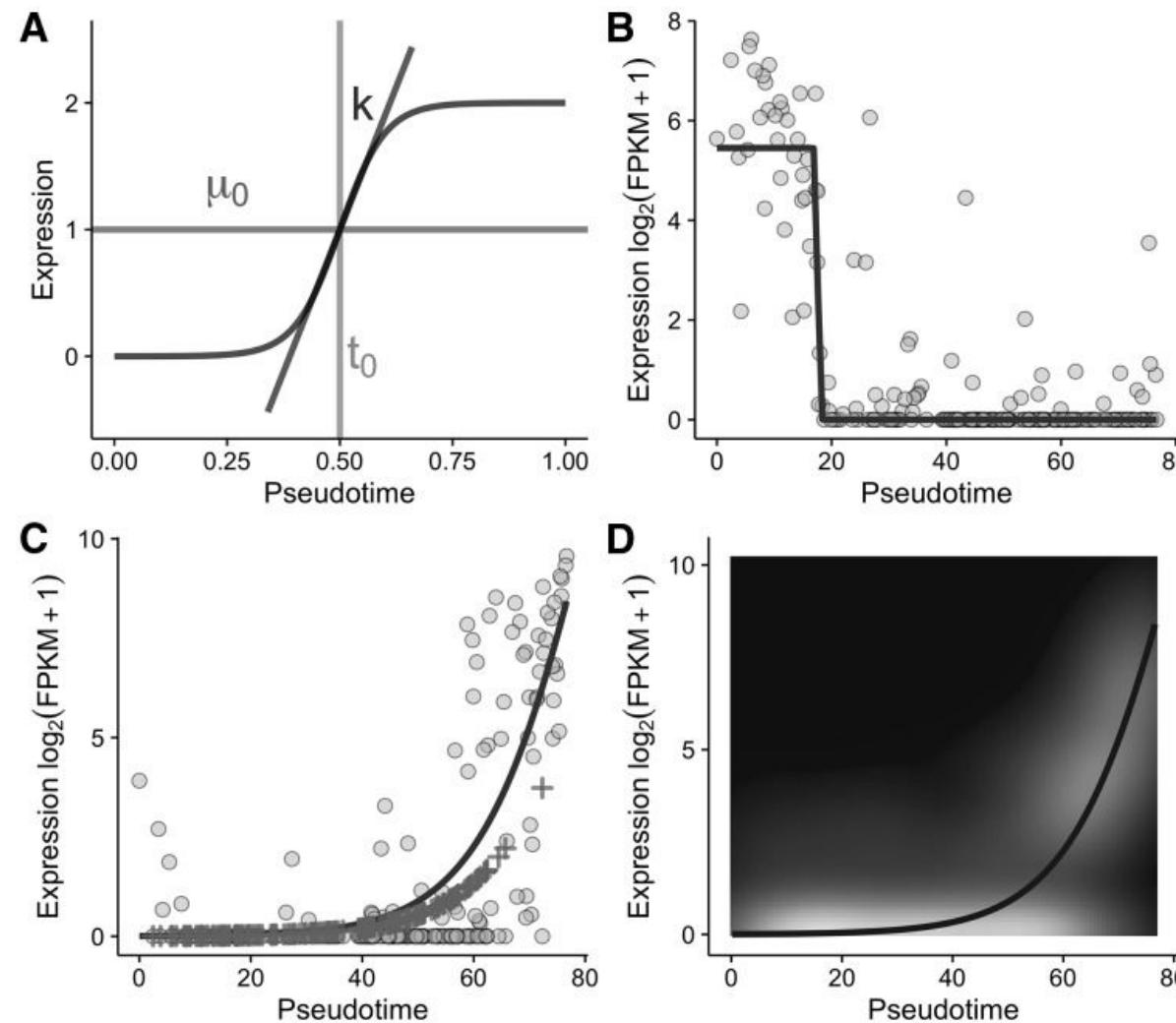


Wishbone

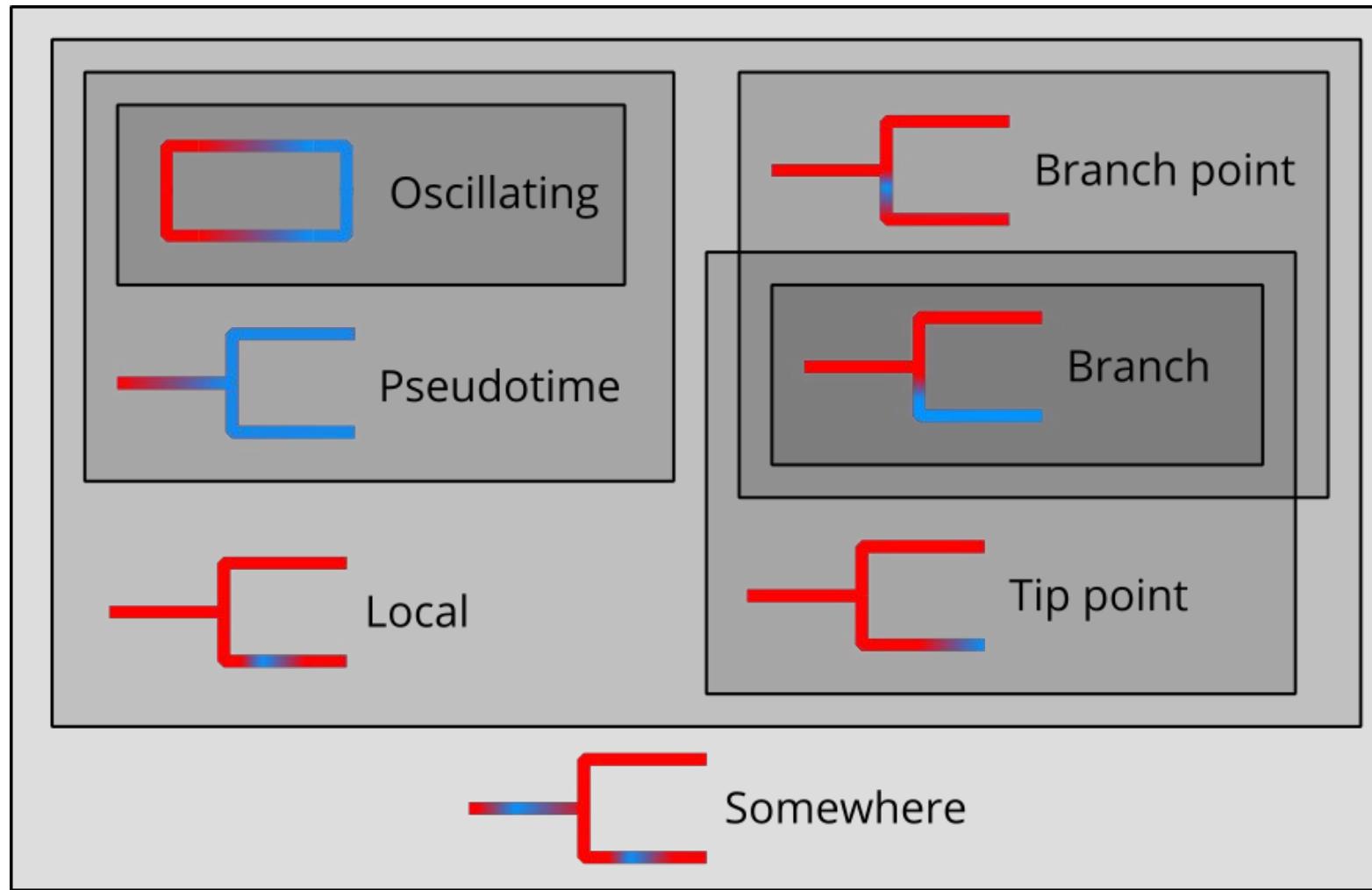


Trajectory differential expression

switchDE



Trajectory differential expression



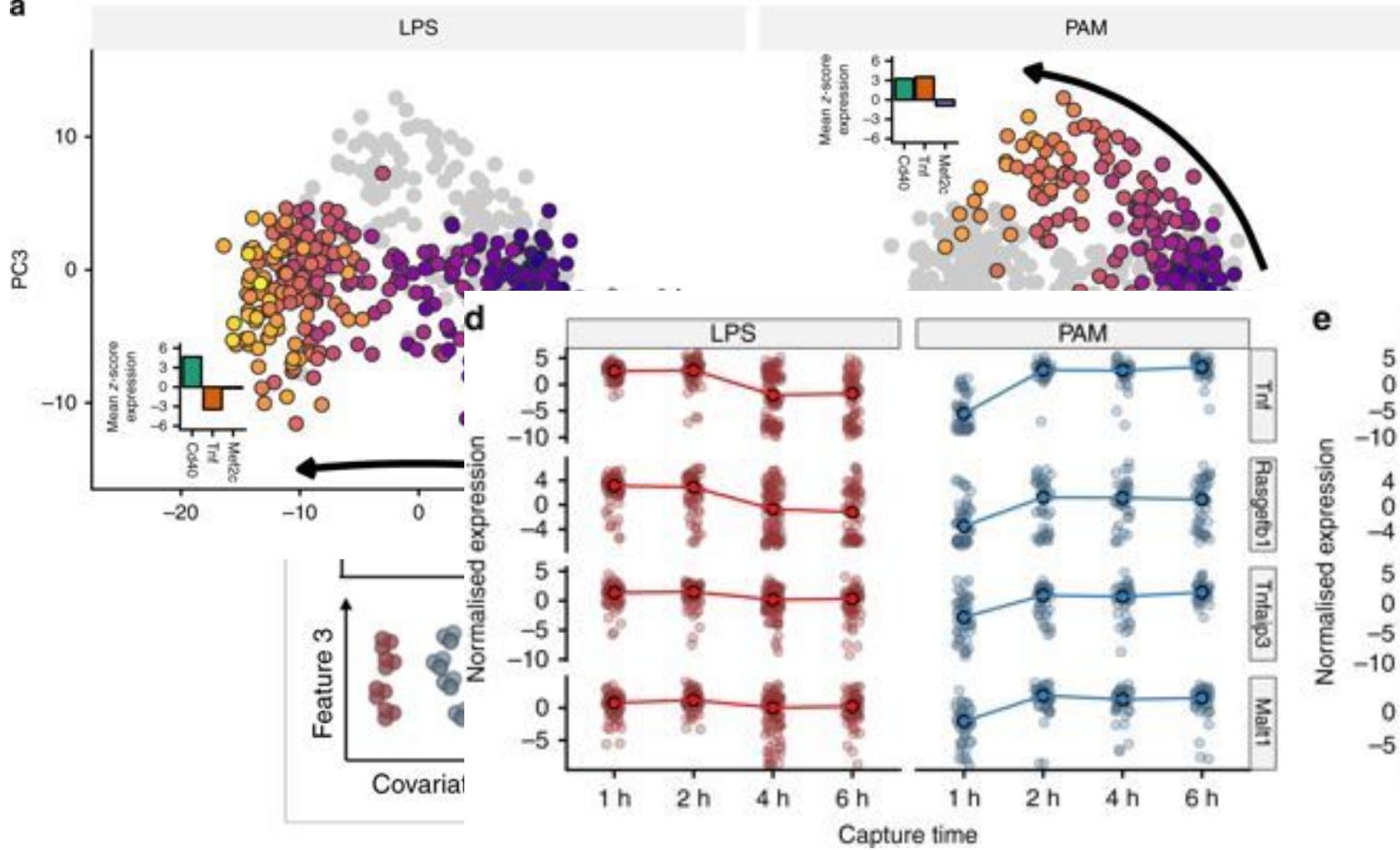
Trajectory differential expression

Current limitations

- Easy for linear – hard with branching
- No consensus on terminology
- Dubious statistics

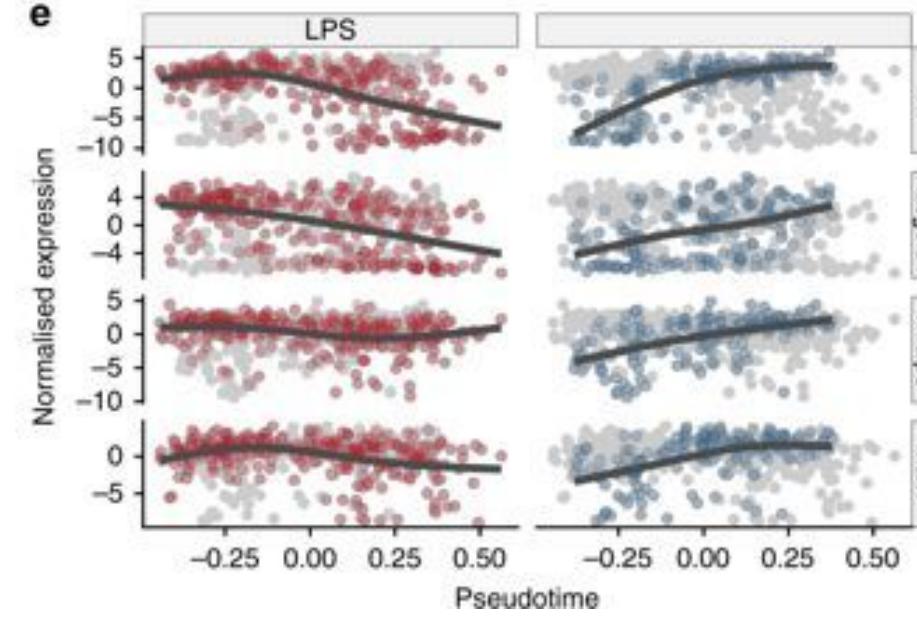
Alignment

a



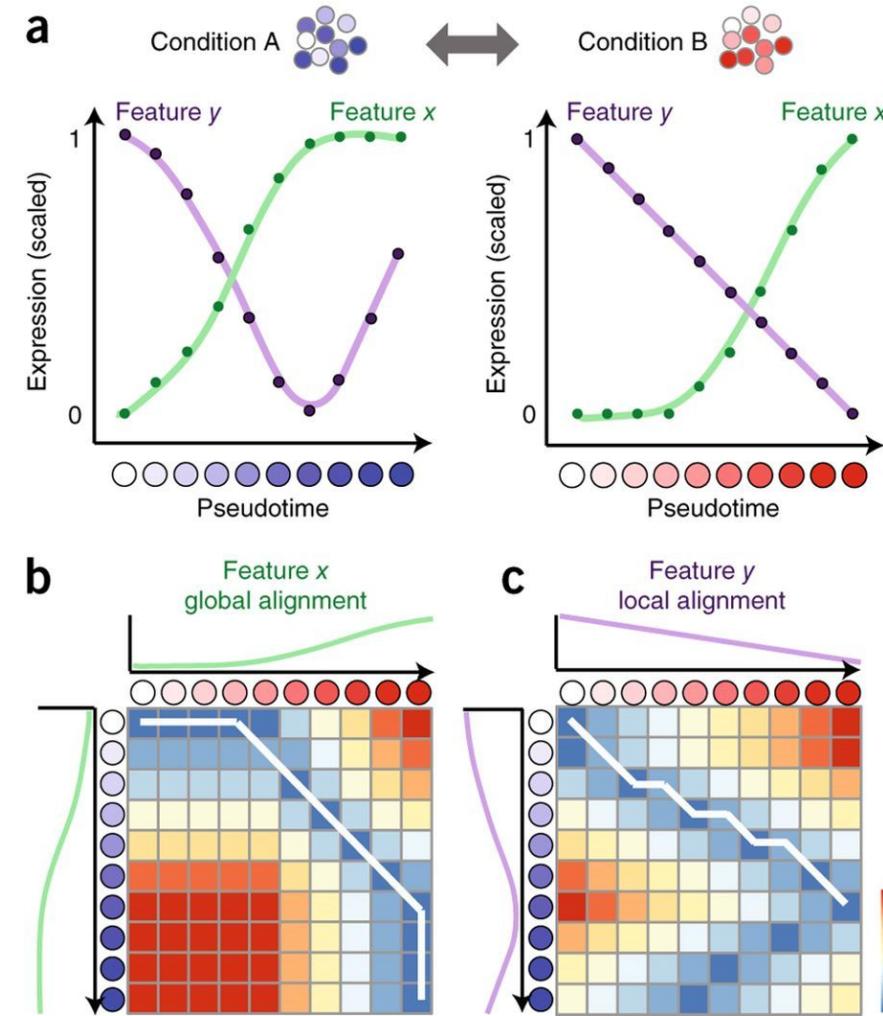
PhenoPath

e



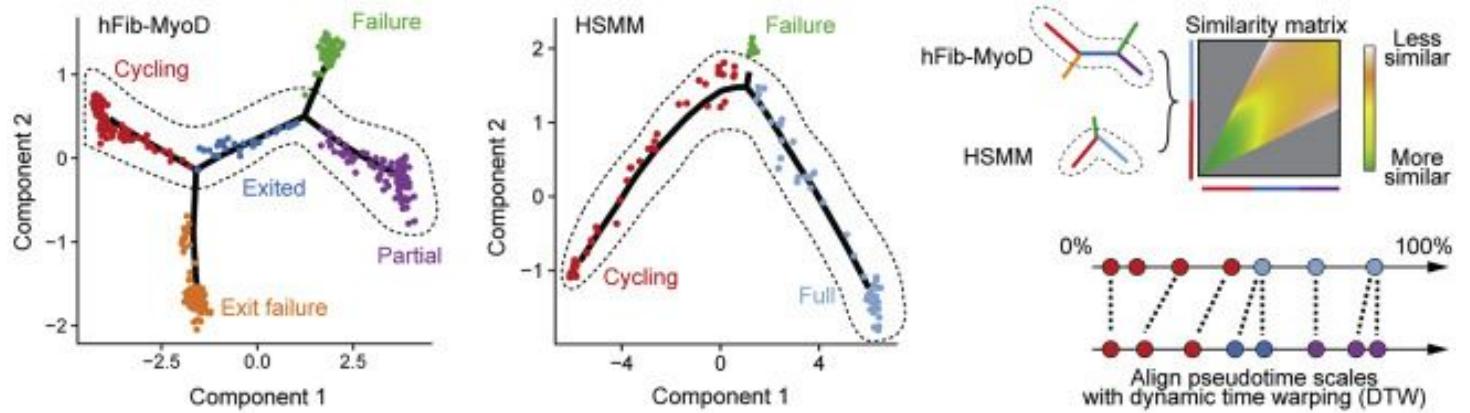
Alignment

CellAlign



Alignment

Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of [...]
<https://doi.org/10.1016/j.cels.2018.07.006>



Current limitations

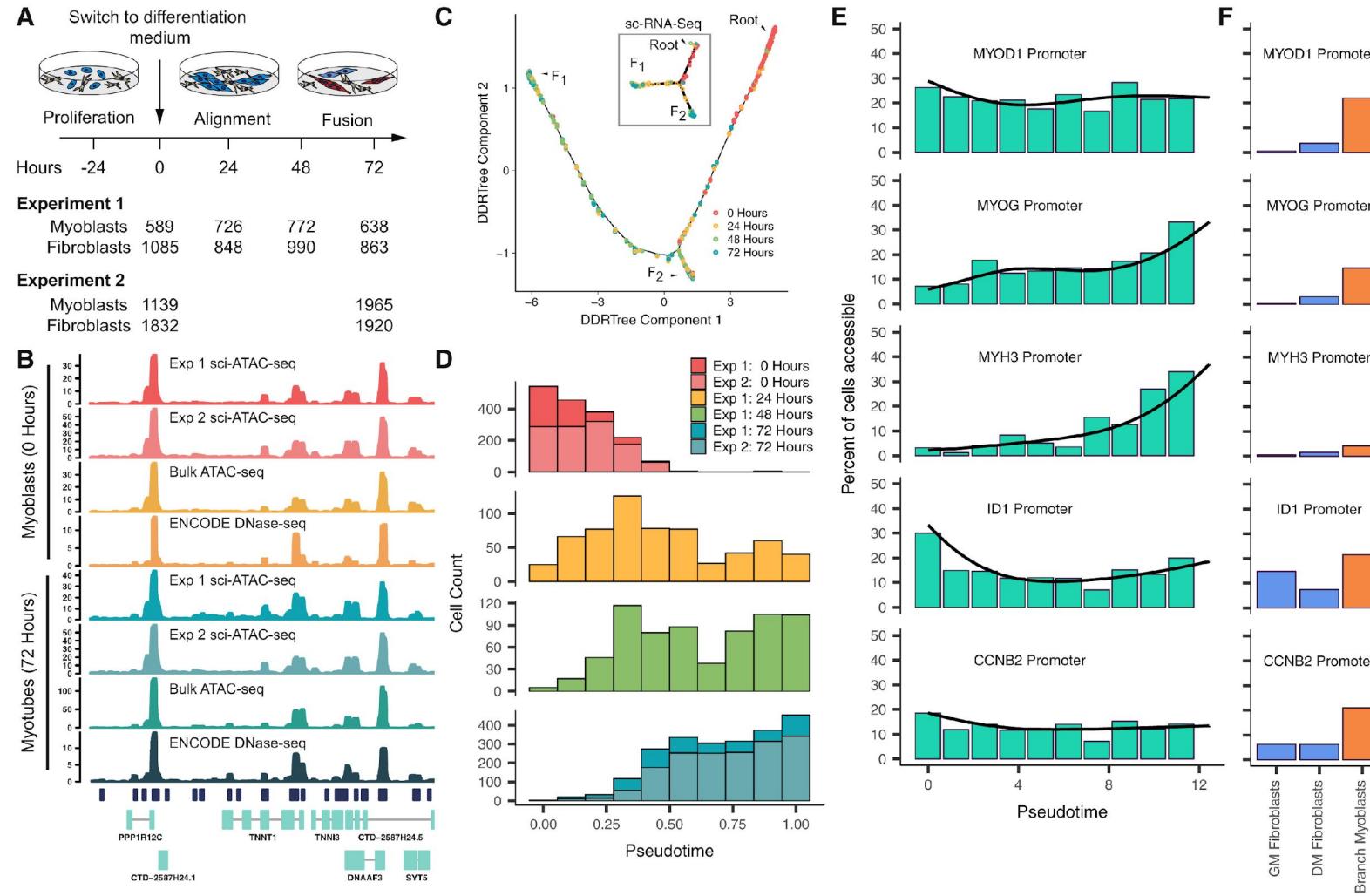
- No alignment of branches
- Assumes approx. same start and end state
- "Batch" sensitivity unclear

Non-transcriptome trajectory inference

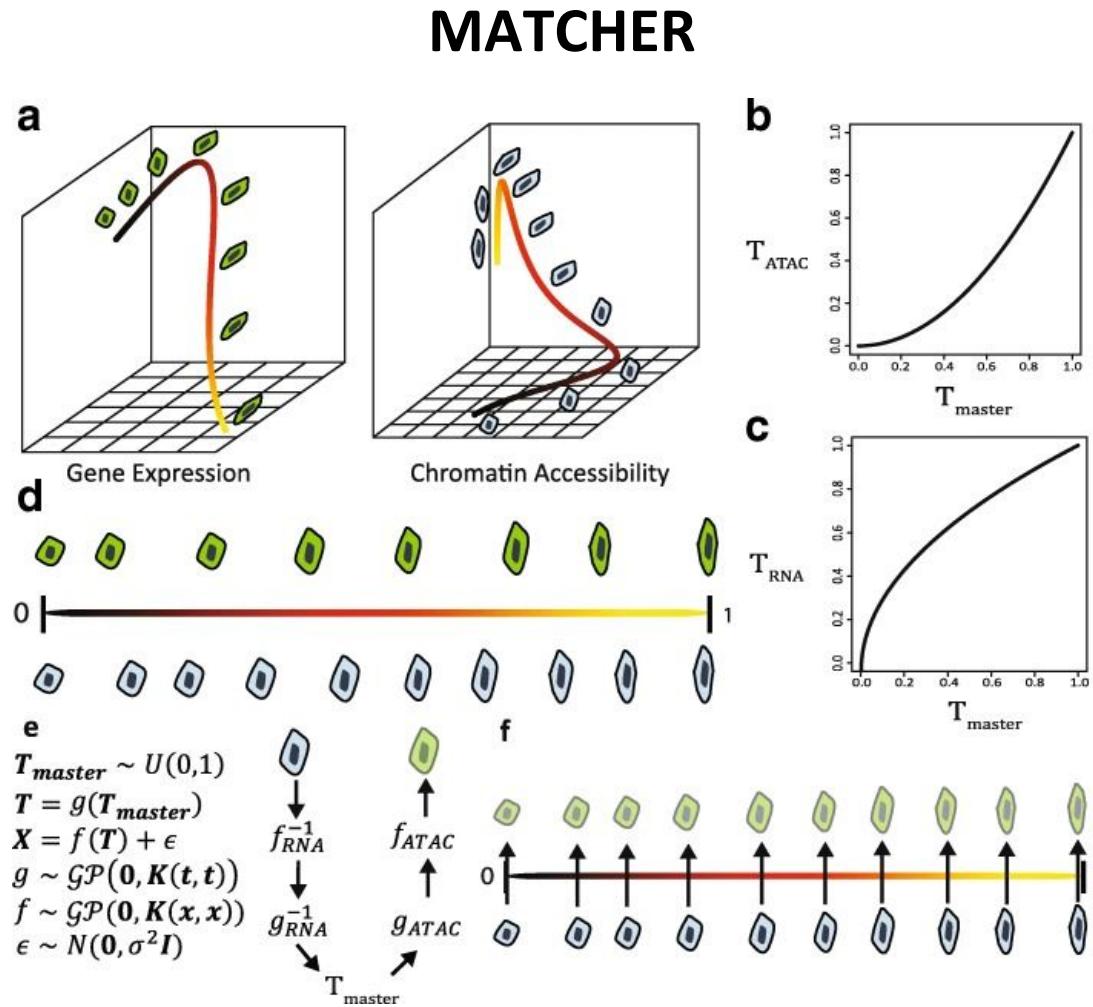
ATAC-Seq

CITE-seq

• • •



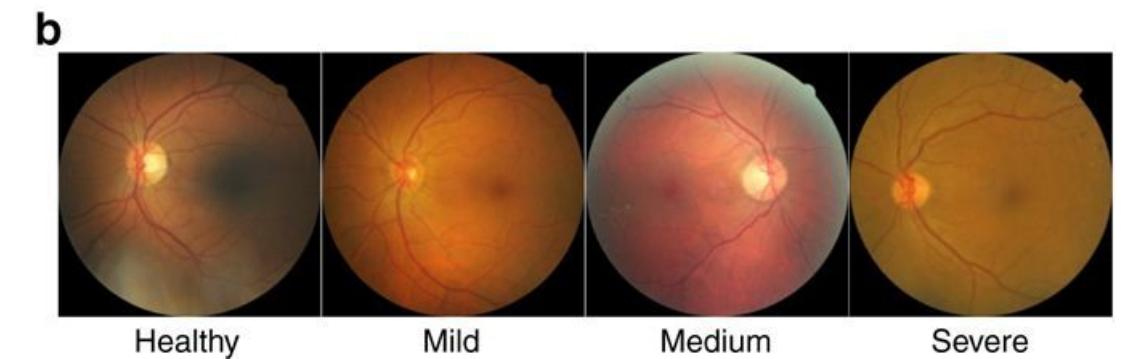
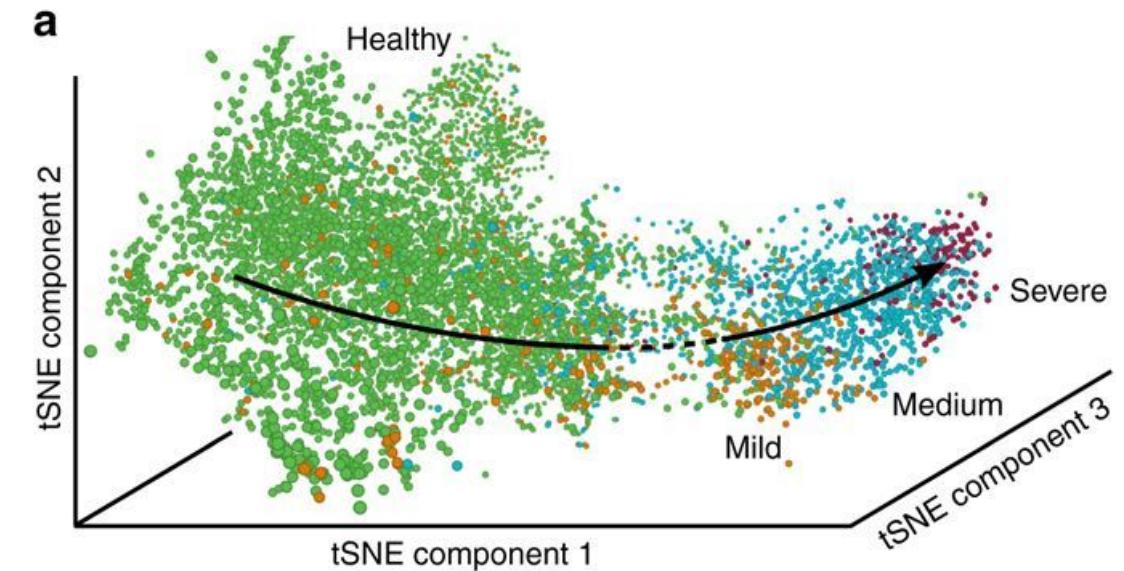
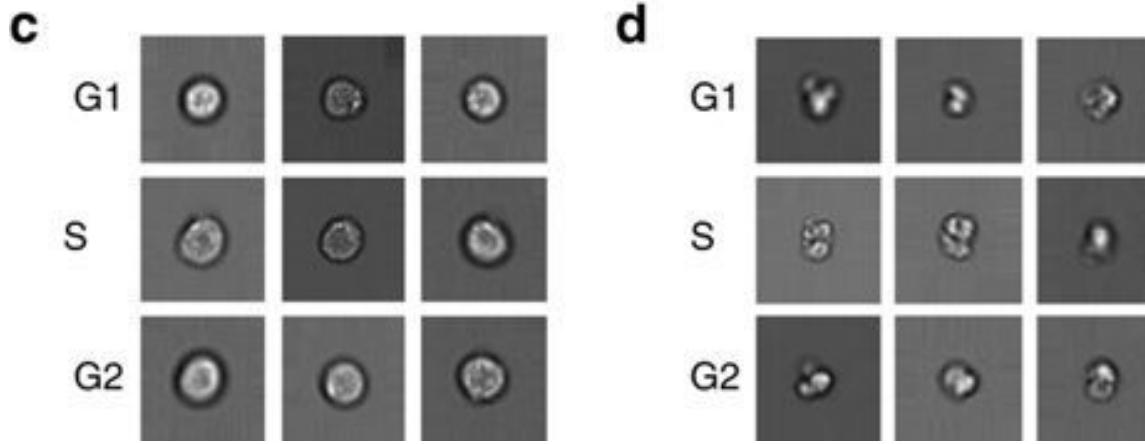
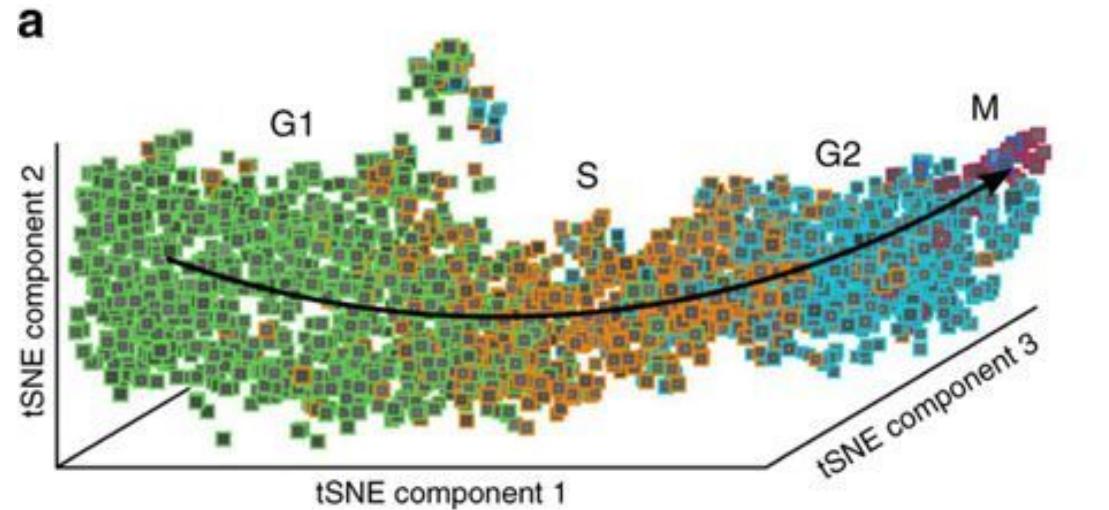
Multi-omics trajectory inference



Current limitations

- No branching
- Limited accuracy & scalability

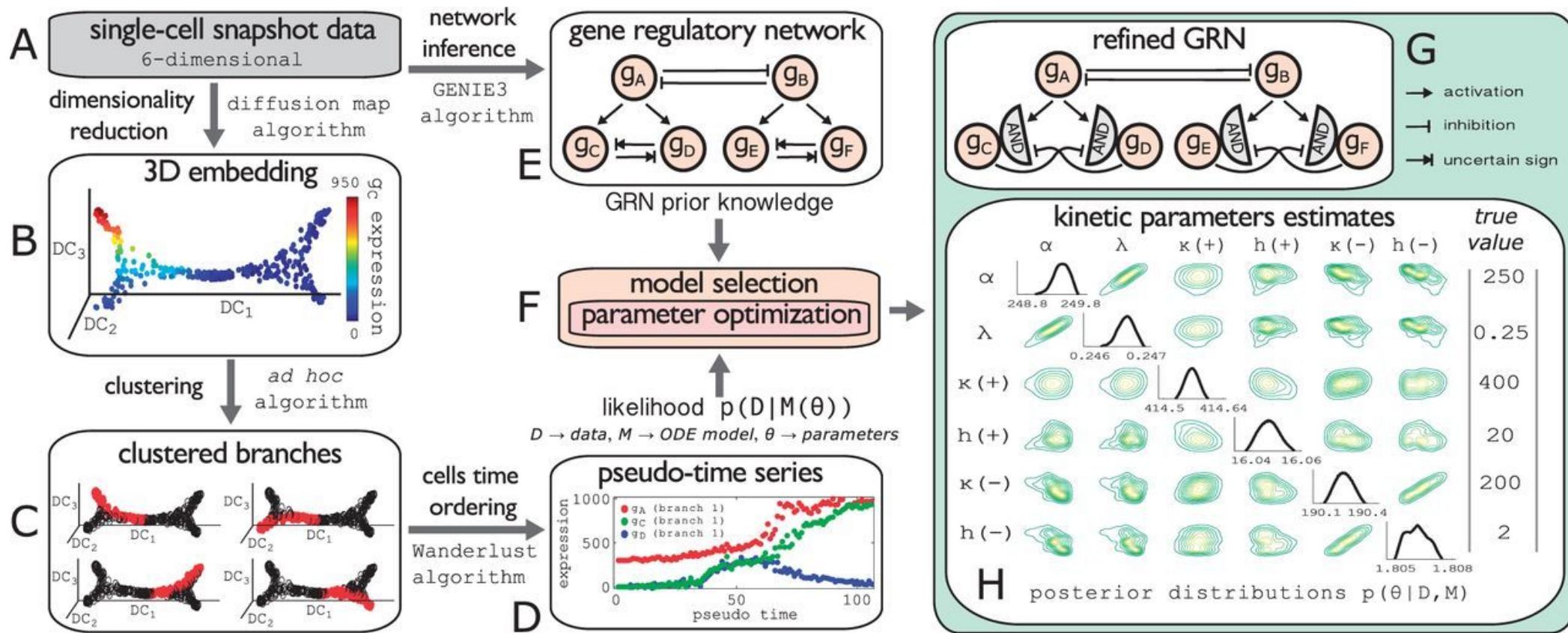
Trajectory inference on images



Reconstructing cell cycle and disease progression using deep learning. <https://doi.org/10.1038/s41467-017-00623>

Single-cell RNA-seq denoising using a deep count autoencoder. <https://doi.org/10.1038/s41467-018-07931-2>

Dynamic network inference



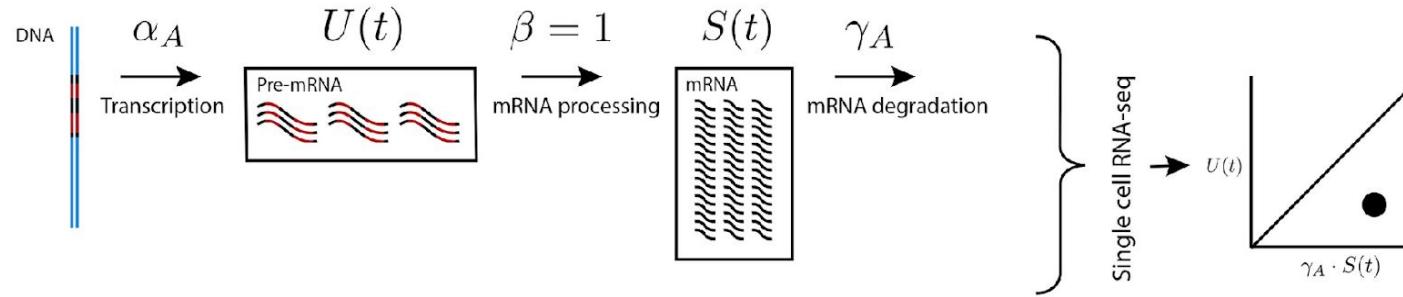
Dynamic network inference

Current limitations

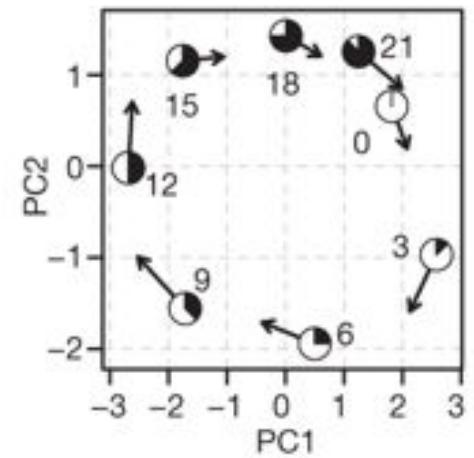
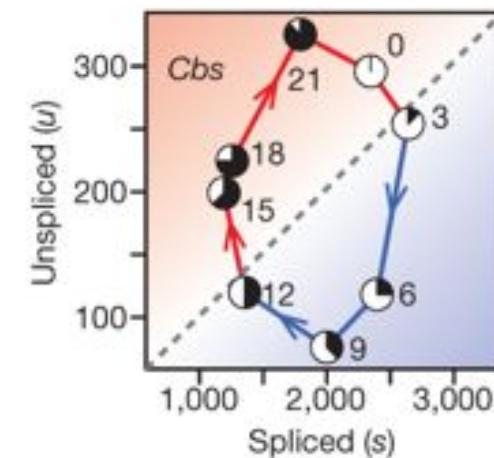
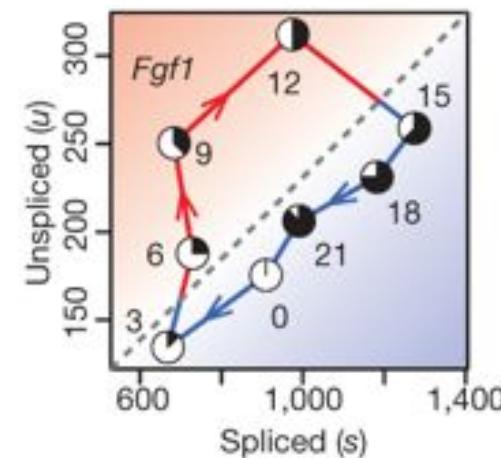
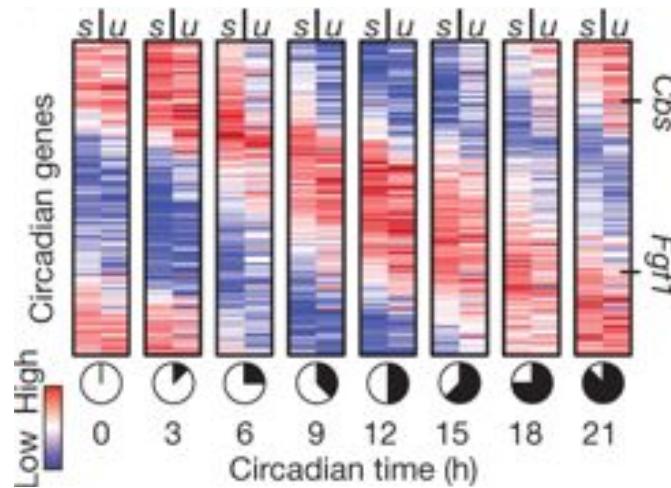
- Only co-expression based
- No robust methods for delay effects
- No multi-omics methods (but huge potential)

RNA velocity

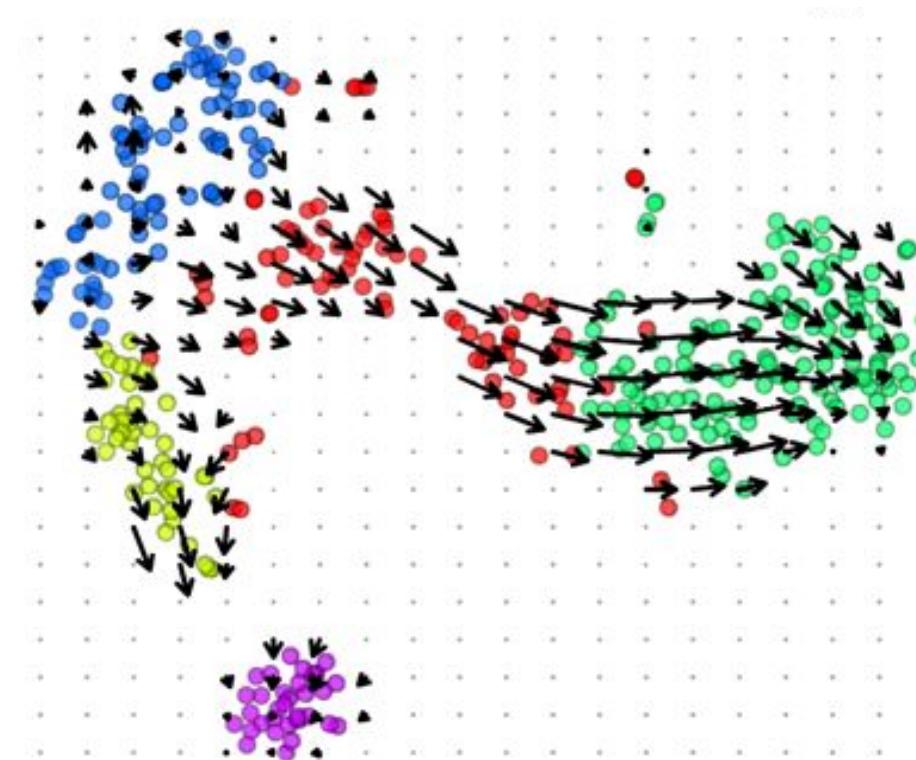
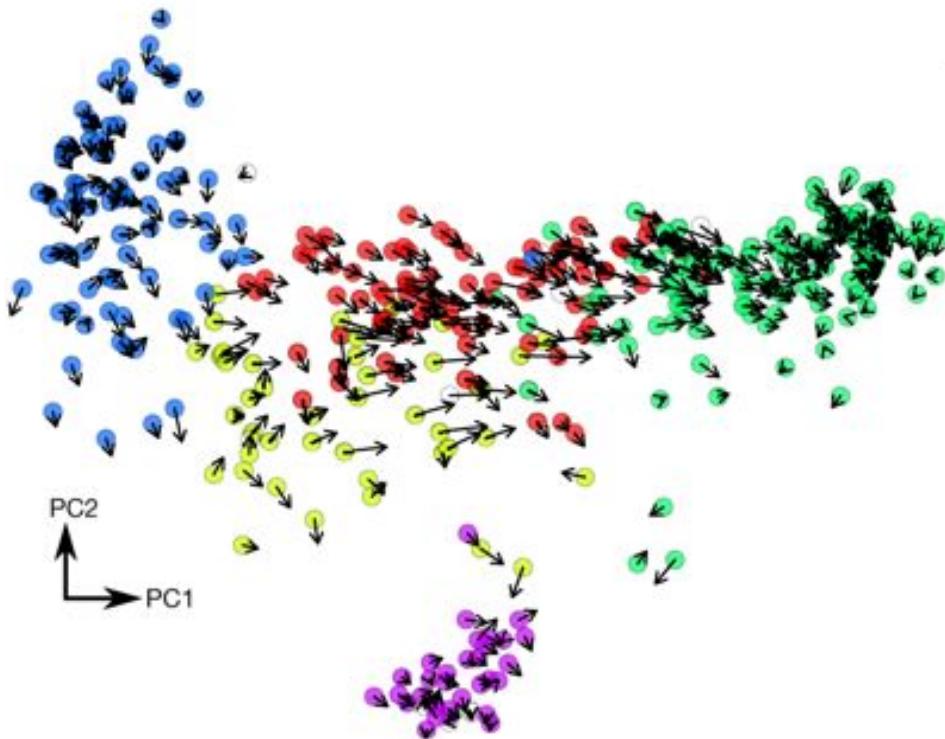
Gene **A** Downregulation, pre-mRNA pool refills slower than mRNA degradation.



RNA velocity



RNA velocity



RNA velocity – comments

Current limitations

- No accurate estimates per gene with 3' protocols
- Doesn't always make sense - parameter dependent!

Possibilities

- RNA acceleration when combining with CITE-Seq or others

2014

2016

2018

2019

2020

- Trajectories after single-cell perturbations? (CRISPR-Cas?)
- Multiple (interacting?) trajectories in the same cells?
- Trajectory inference which uses RNA velocity?



dynverse.org



zouters
rcannood



Only one border away ;-)
We have fries

