Single-cell trajectories

Wouter Saelens

Saeys lab @ VIB - Ghent University, Belgium

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

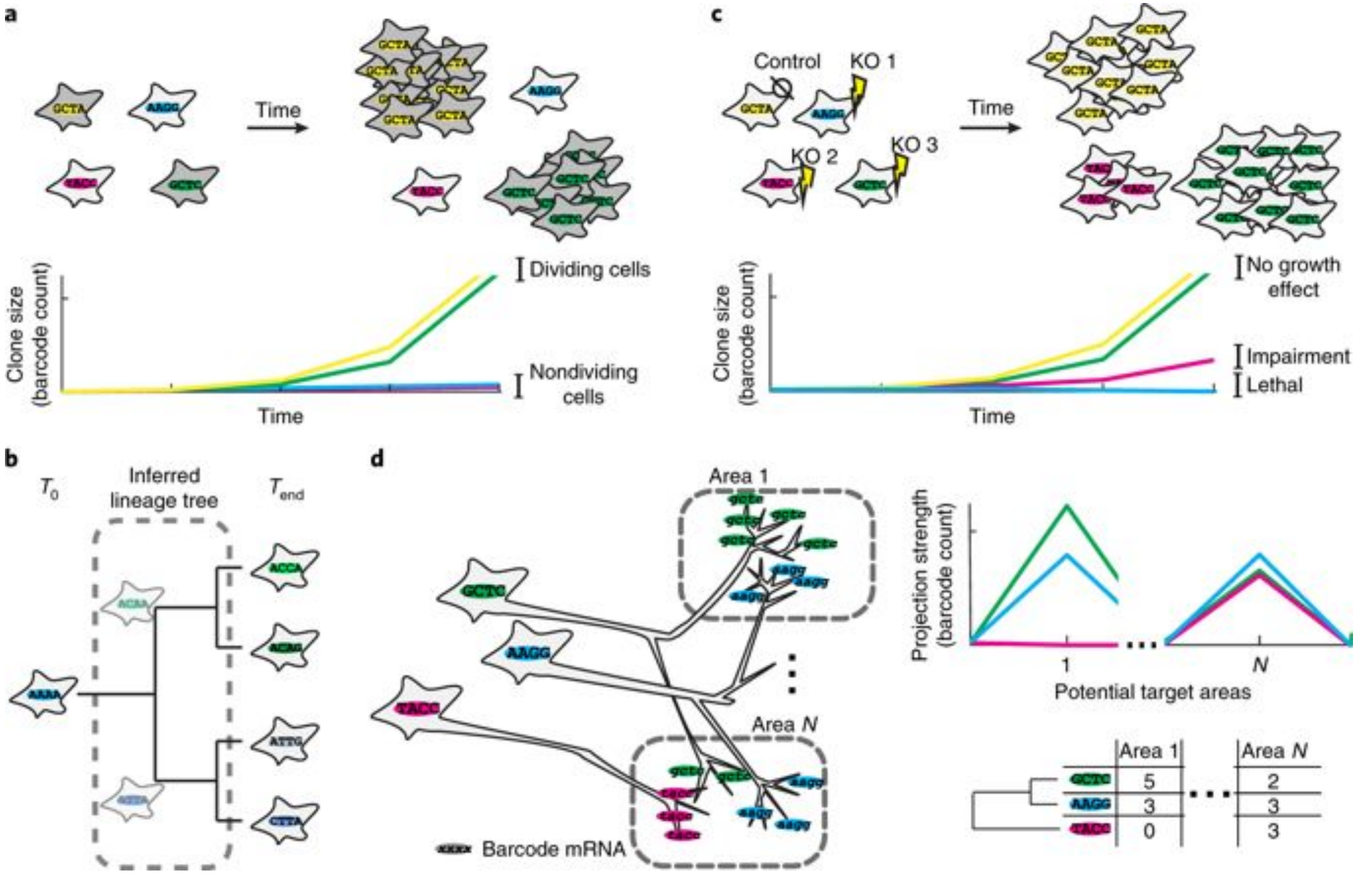
Barcoding

& Lineage tracing

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

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



	Area 1	Area N
GCTC	5	2
AAGG	3	3
TACC	0	3

Cellular barcoding: lineage tracing, screening and beyond. https://doi.org/10.1038/s41592-018-0185-x

3 ways to assess single-cell dynamics

Wet

Dry

Barcoding
& *Lineage tracing*

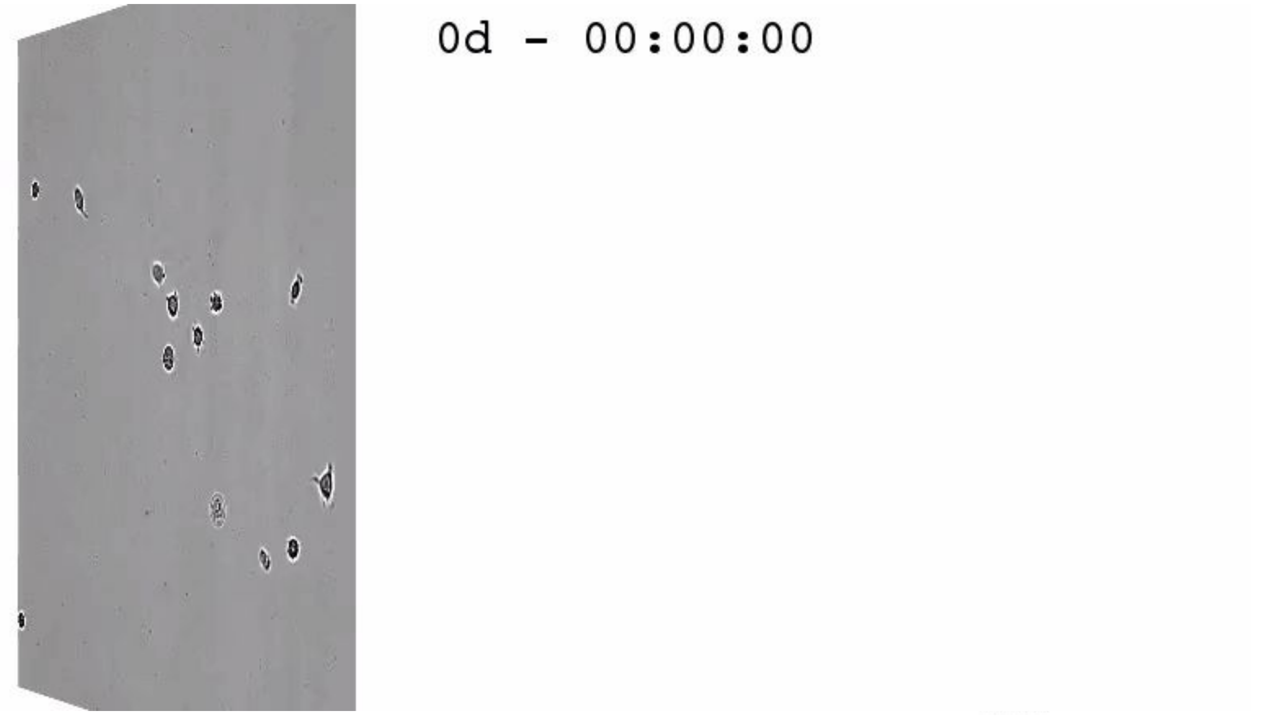
1. Barcode progenitors
2. Wait
3. Analyze barcodes & phenotype of descendants

Live cell imaging
& Cell tracking

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

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



On long term single cell imaging. https://doi.org/10.1038/nmeth.1577

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

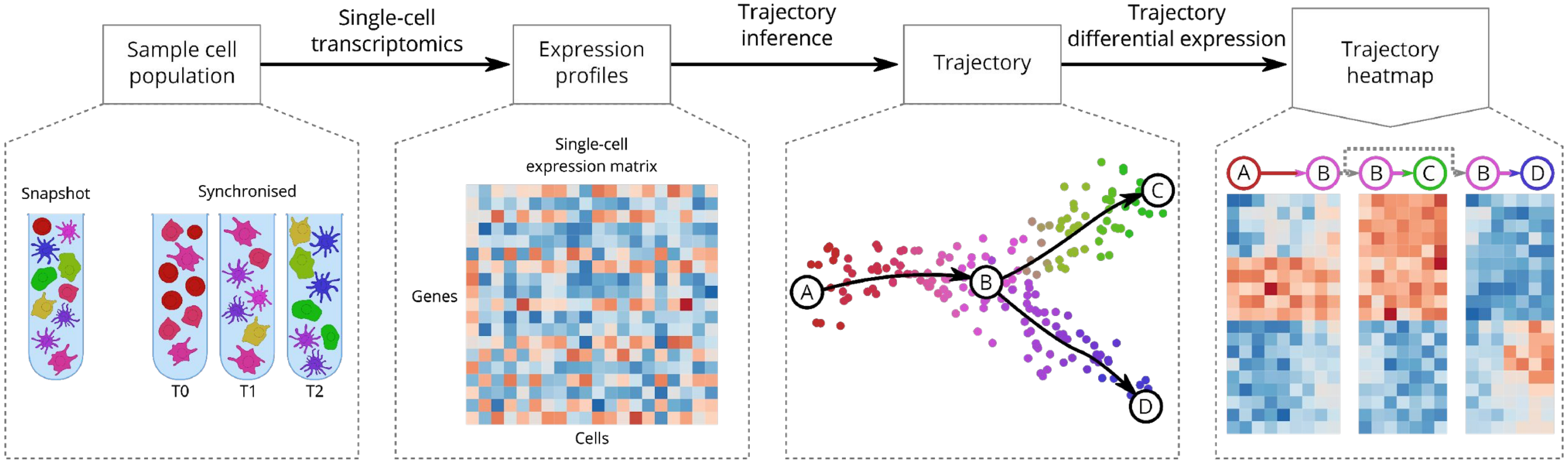
1. Image cells continuously
2. Segment and track individual cells and descendants

Single-cell omics

& Trajectory inference

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

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

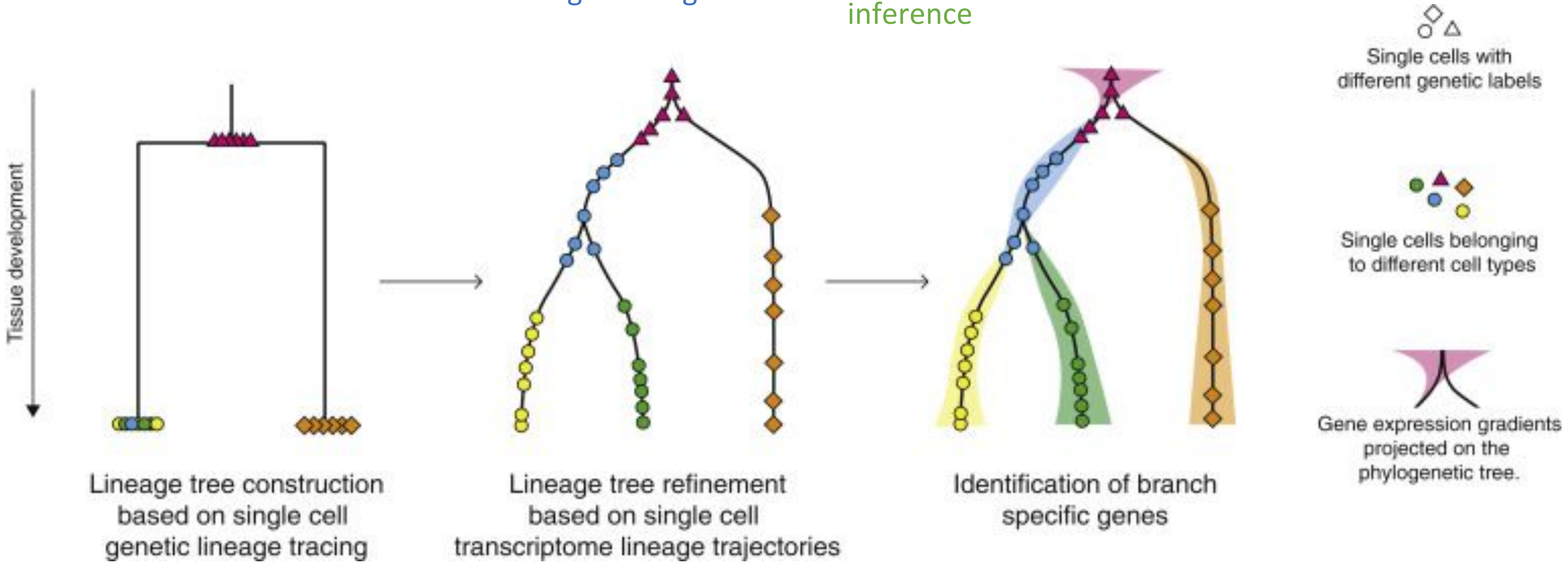
Computational methods for trajectory inference from single-cell transcriptomics. https://doi.org/10.1002/eji.201646347

The future: combinations

Barcoding
& Lineage tracing

\+

Single-cell omics &
Trajectory
inference



How to analyze single-cell dynamics

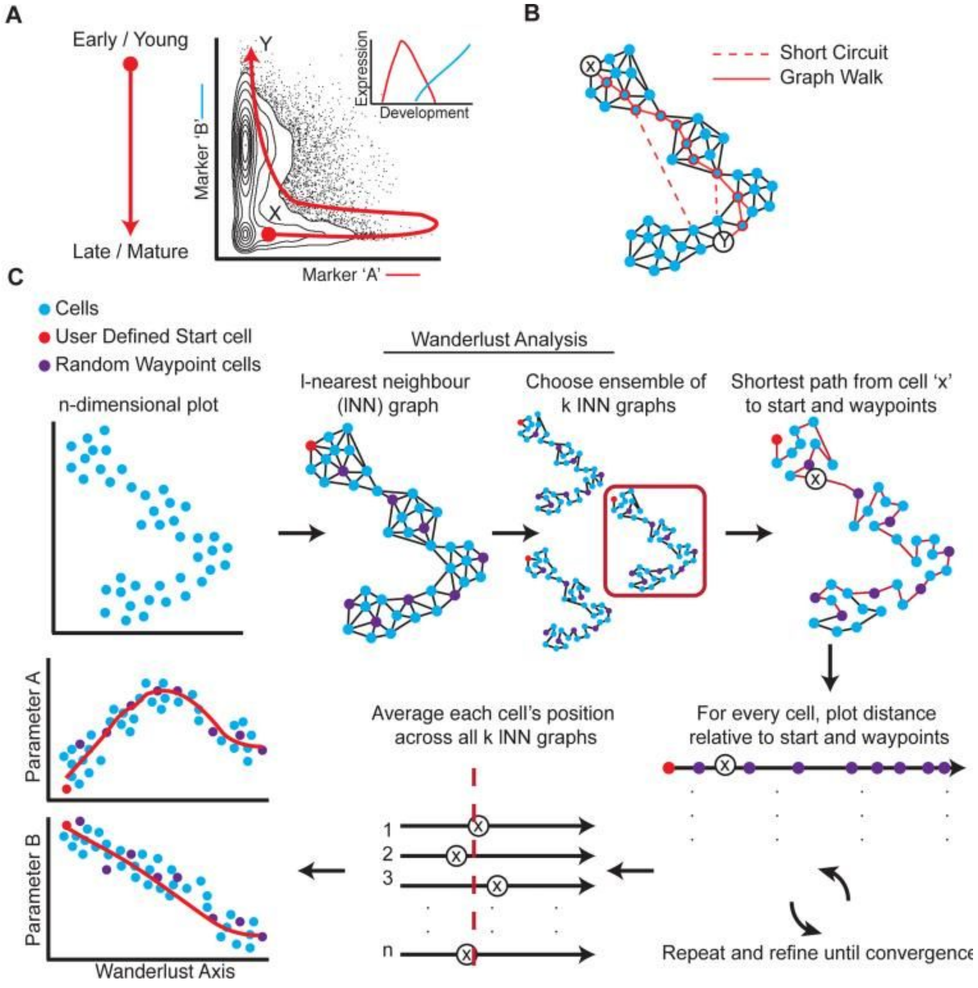
Current state of trajectory inference

Some internal methodology

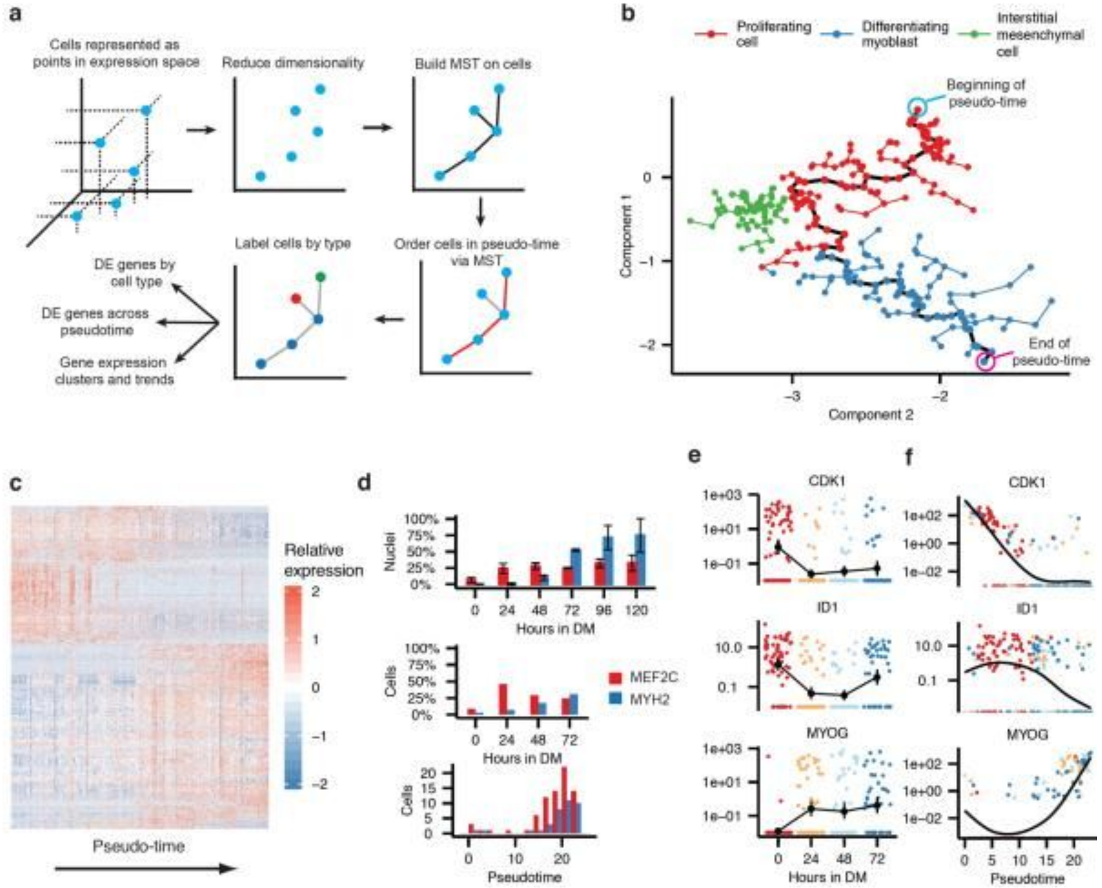
Extensions

2014 2016 2018 2019 2020

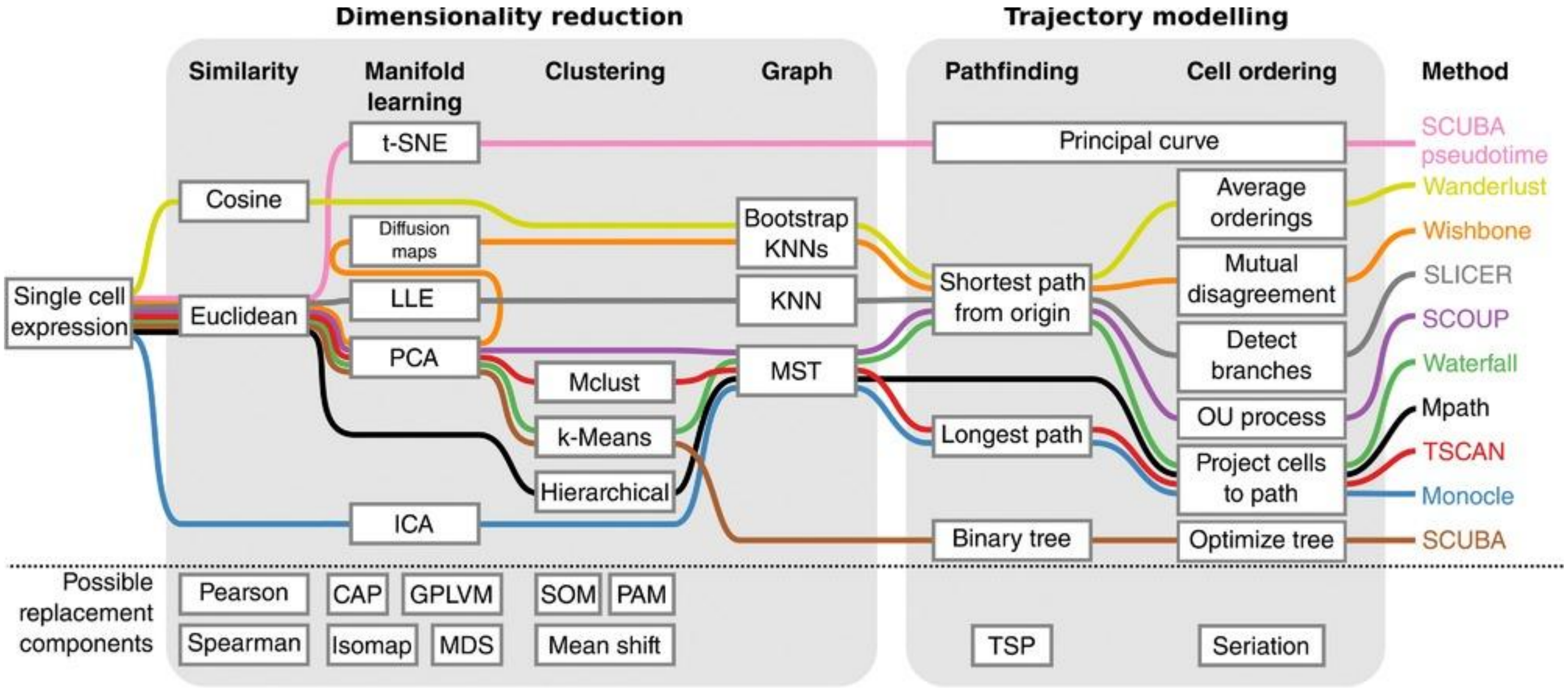
Wanderlust



Monocle (v1)

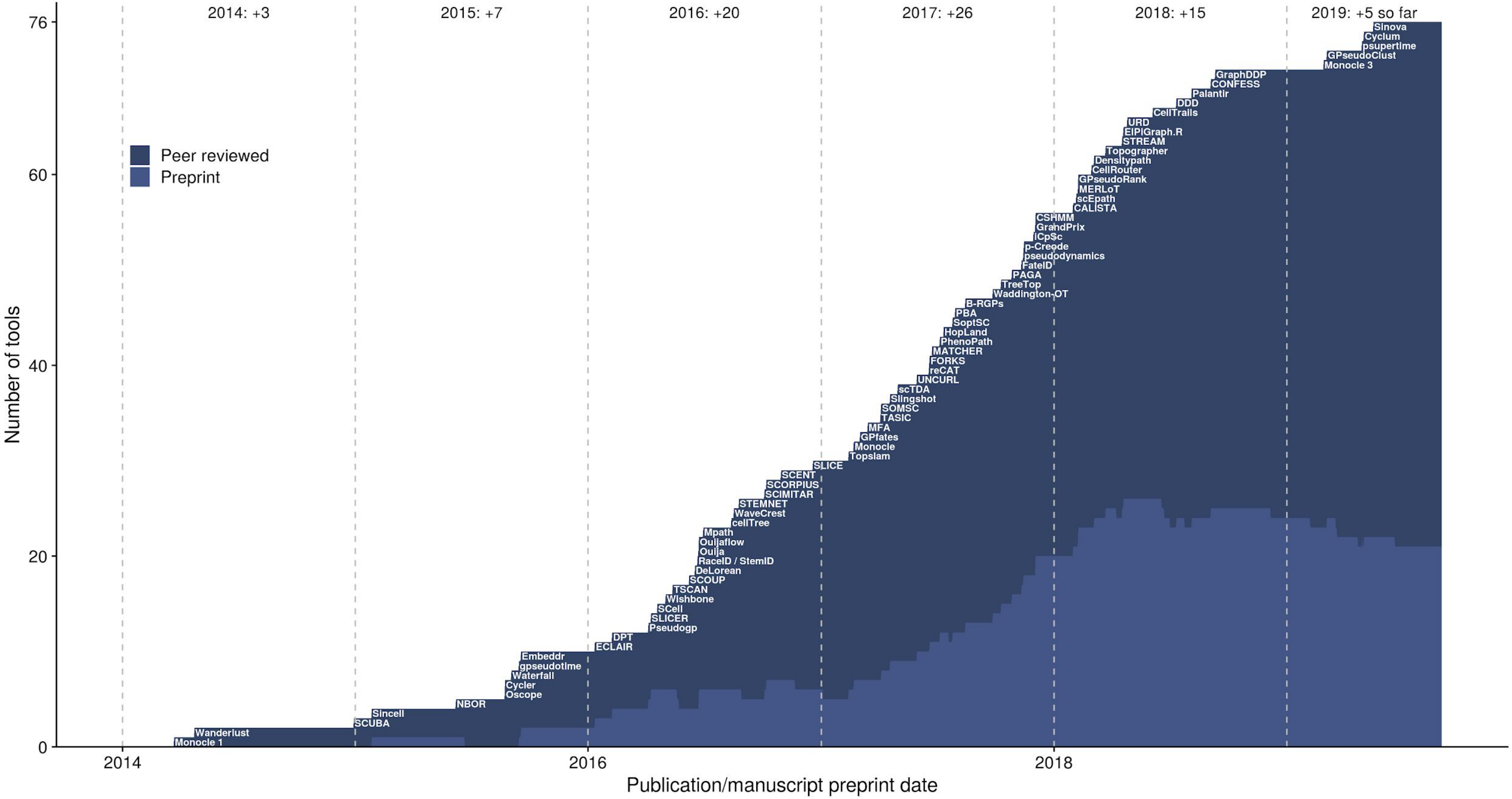


2014 2016 2018 2019 2020

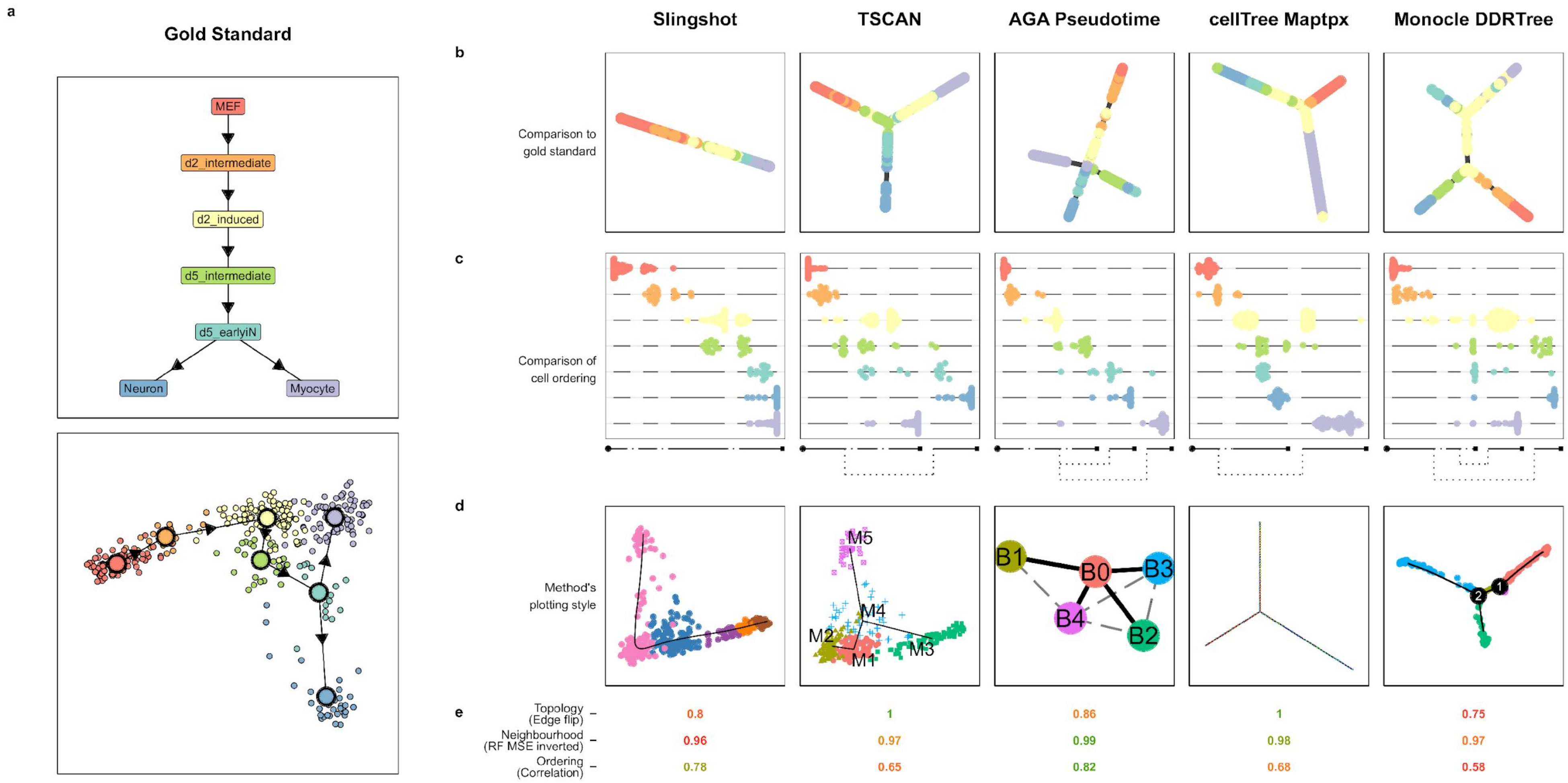


Computational methods for trajectory inference from single-cell transcriptomics. https://doi.org/10.1002/eji.201646347

2014 2016 2018 2019 2020

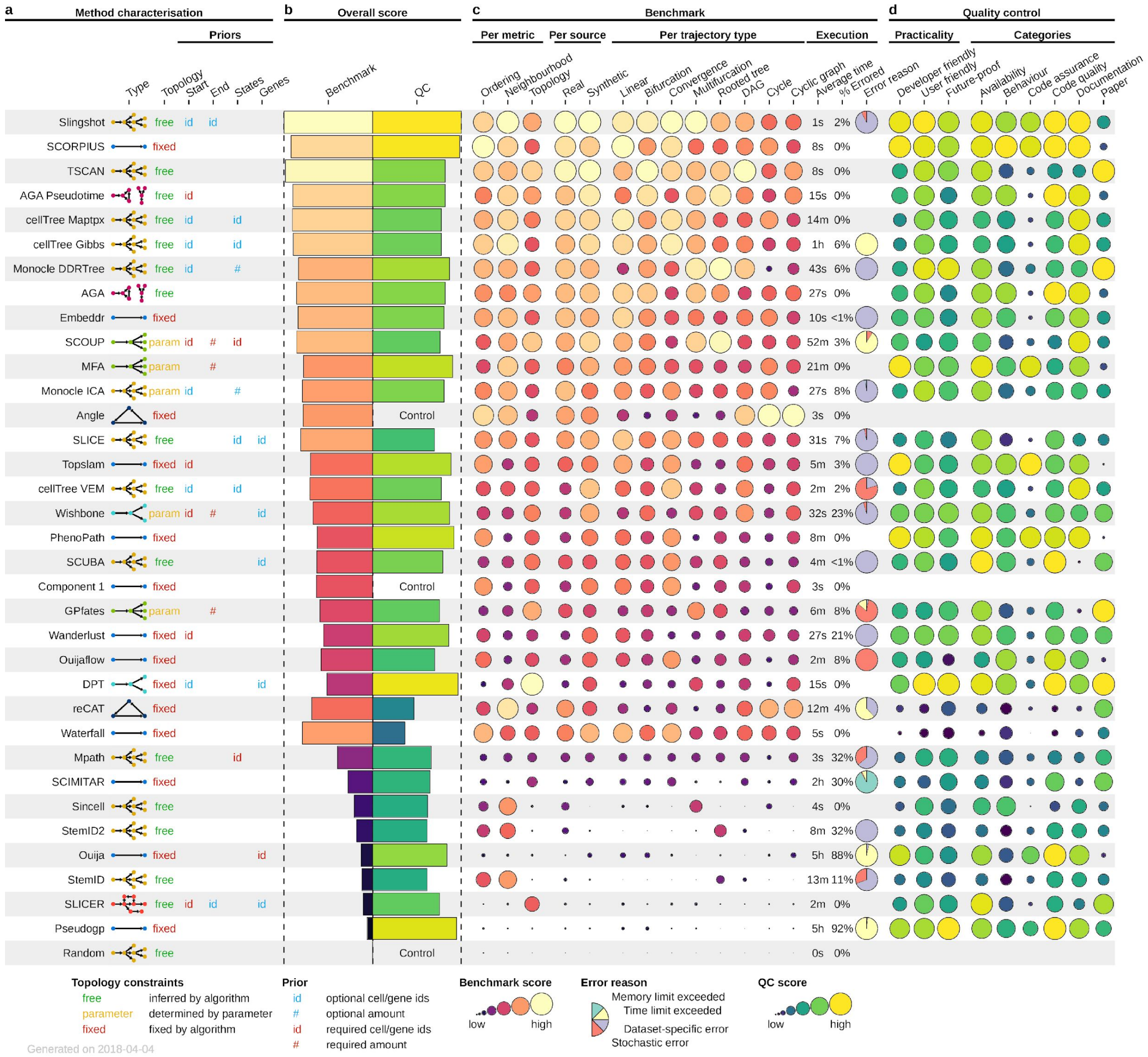


2014 2016 2018 2019 2020



A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. https://doi.org/10.1101/276907

2014 2016 2018 2019 2020



A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. https://doi.org/10.1101/276907

2014 2016 2018 2019 2020

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

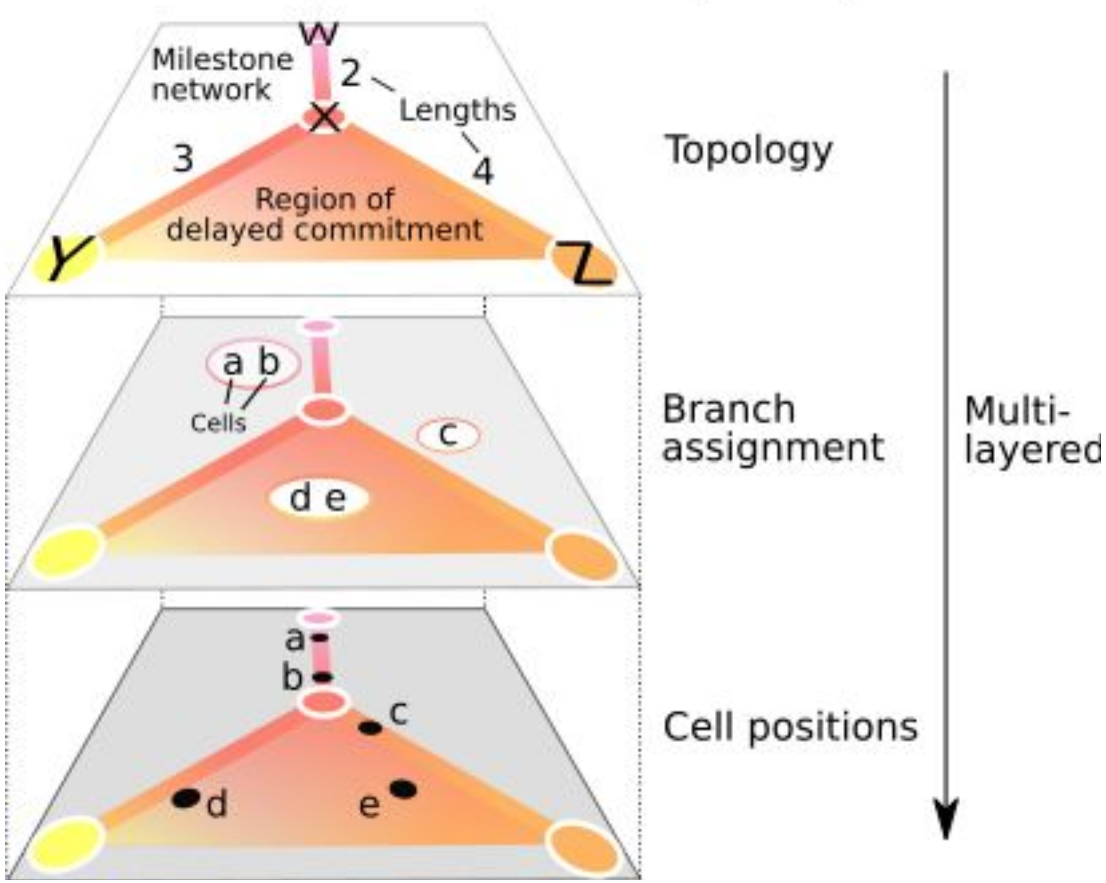
Accuracy + Scalability + Stability + Usability

User guidelines + New possibilities for developers

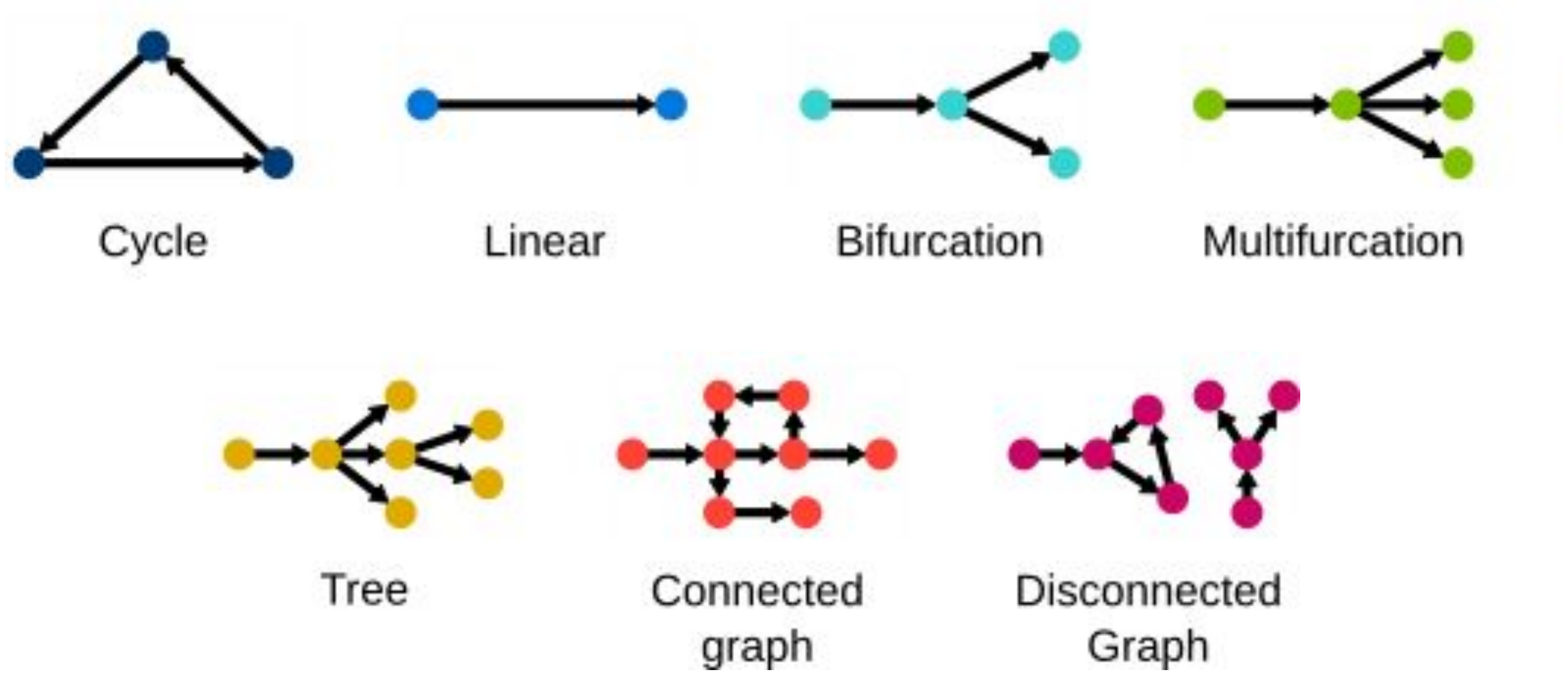
Method wrappers methods.dynverse.org
Guidelines app guidelines.dynverse.org
Benchmarking pipeline benchmark.dynverse.org

A comparison of single-cell trajectory inference methods. https://doi.org/10.1038/s41587-019-0071-9

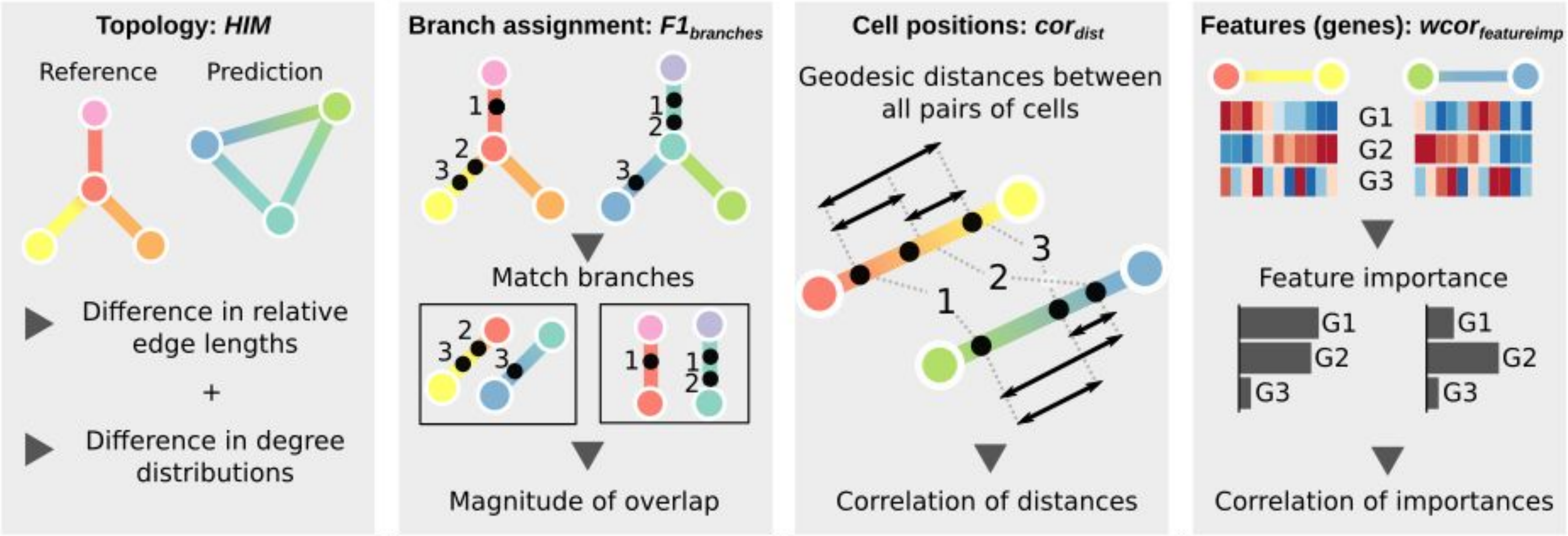
Common probabilistic trajectory model

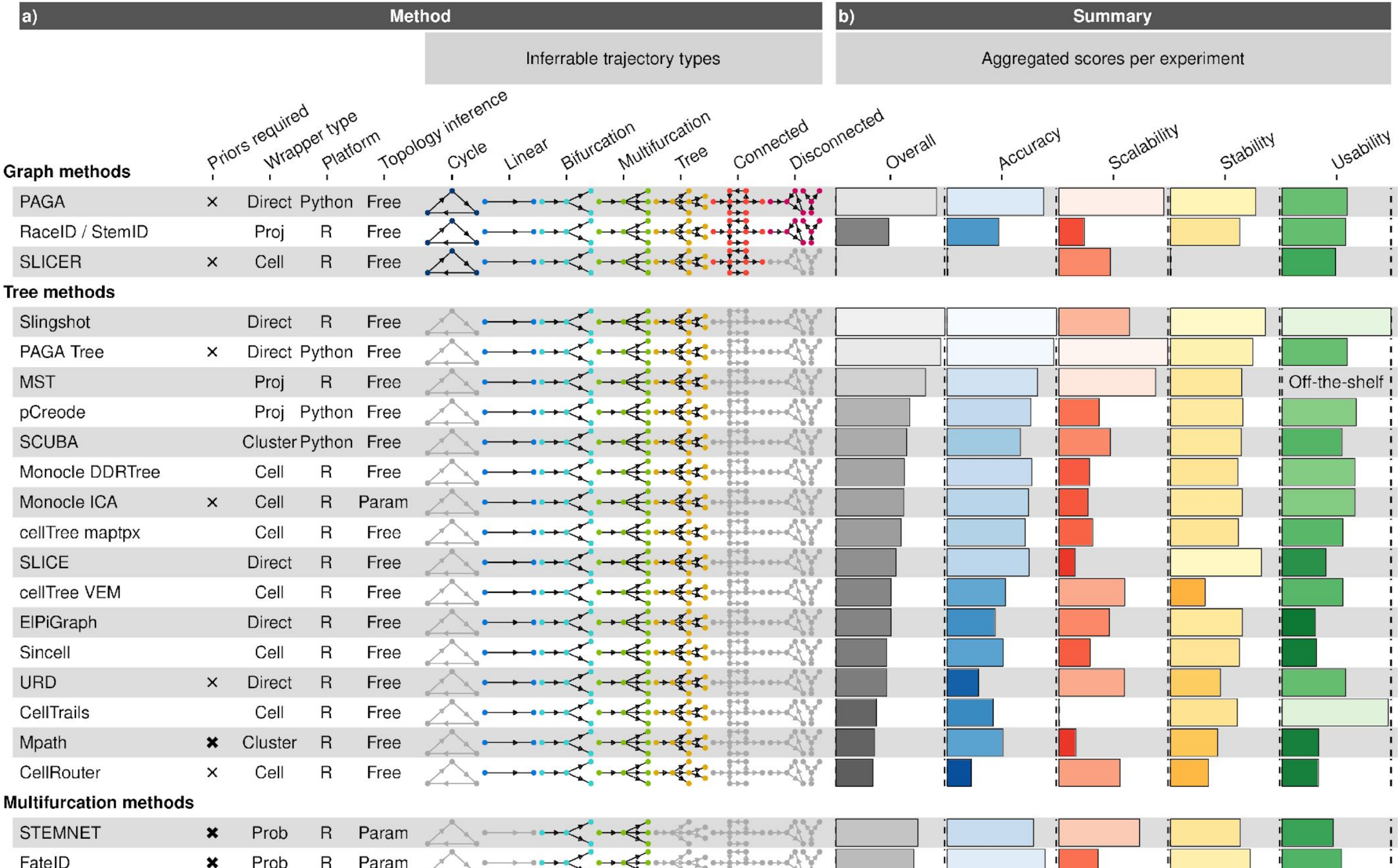


Main difference: trajectory types



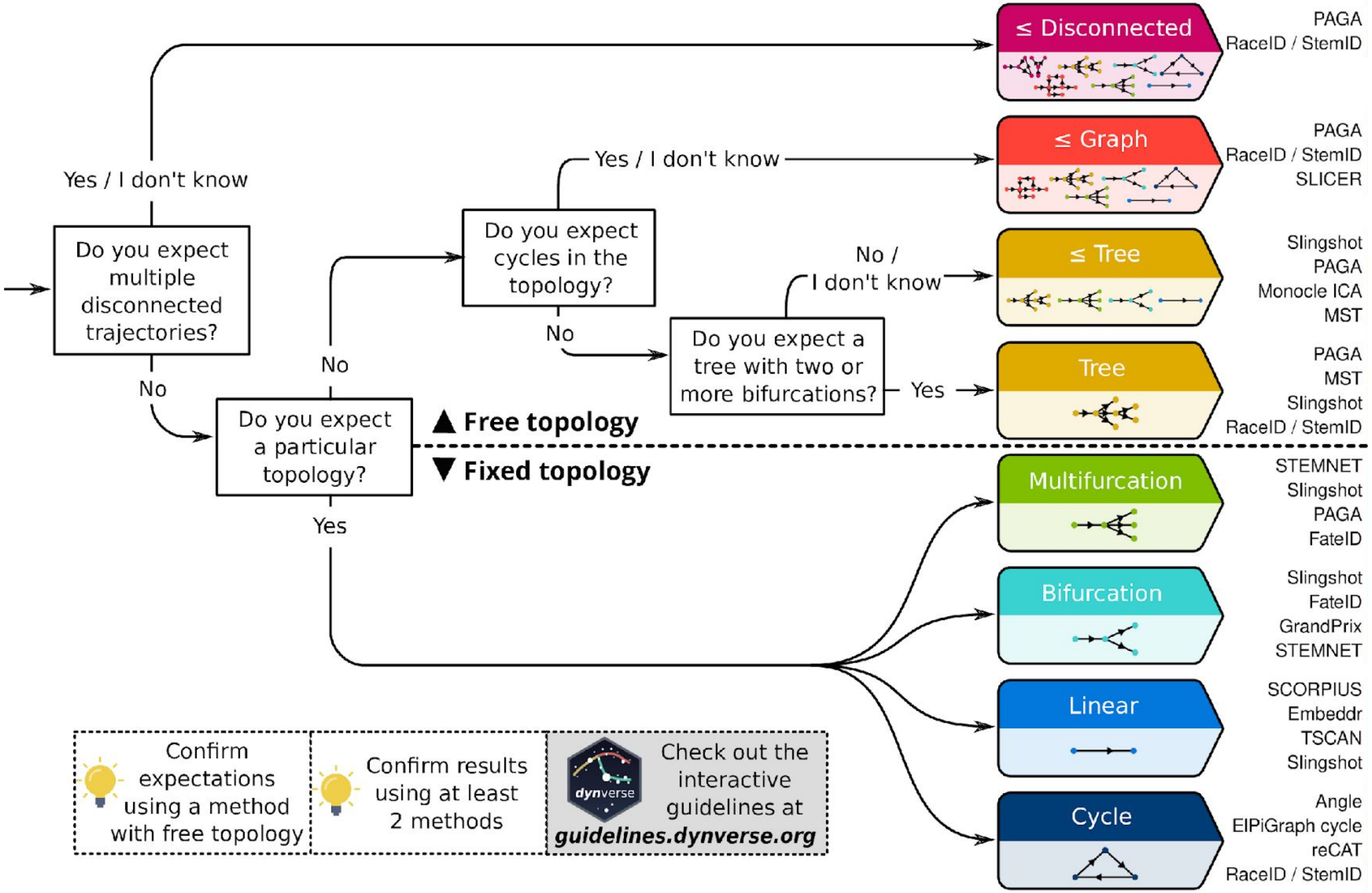
How to assess accuracy



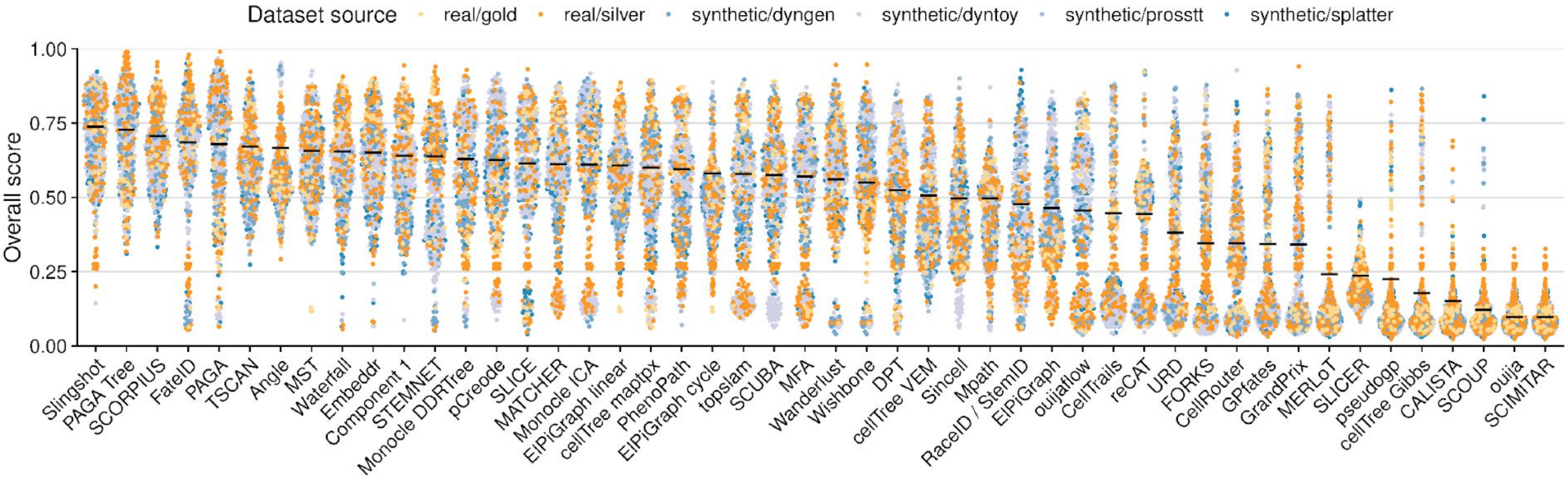


	Priors required	Wrapper type	Platform	Topology inference
Graph methods				
PAGA	×	Direct	Python	Free
RaceID / StemID		Proj	R	Free
SLICER	×	Cell	R	Free
Tree methods				
Slingshot		Direct	R	Free
PAGA Tree	×	Direct	Python	Free
MST		Proj	R	Free
pCreode		Proj	Python	Free
SCUBA		Cluster	Python	Free
Monocle DDRTree		Cell	R	Free
Monocle ICA	×	Cell	R	Param
cellTree maptpx		Cell	R	Free
SLICE		Direct	R	Free
cellTree VEM		Cell	R	Free
ElPiGraph		Direct	R	Free
Sincell		Cell	R	Free
URD	×	Direct	R	Free
CellTrails		Cell	R	Free
Mpath	✖	Cluster	R	Free
CellRouter	×	Cell	R	Free
Multifurcation methods				
STEMNET	✖	Prob	R	Param
FateID	✖	Prob	R	Param

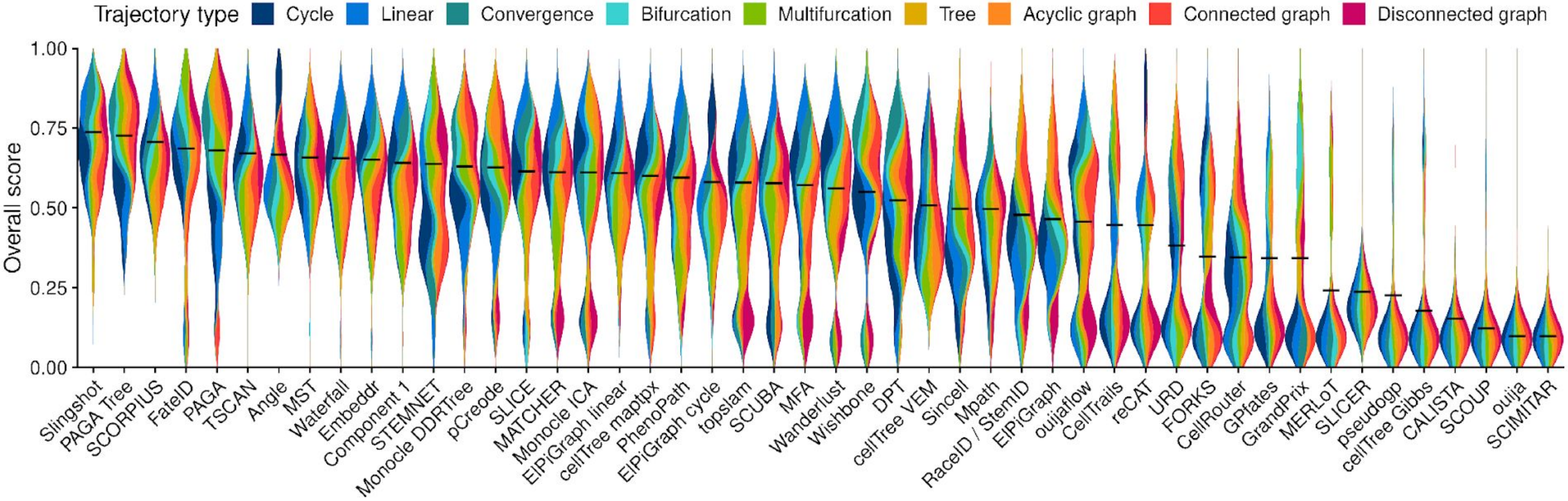

a) Method
b) Accuracy
Per metric
Per dataset source
Per trajectory type
c) Scalability
Predicted time (#cells × #features)
d) Stability
Similarity between runs
e) Usability
Quality of software and paper
Topology
Branch assignment
Cell positions
Features
Gold
Silver
Dyngen
Dyntoy
PROSSTT
Splatter
Cycle
Linear
Bifurcation
Convergence
Multifurcation
Tree
Acyclic
Connected
Disconnected
1m × 100
100k × 1k
10k × 10k
1k × 100k
100 × 1m
Cor. pred. vs. real
Topology
Branch assignment
Cell positions
Features
Availability
Behaviour
Code assurance
Code quality
Documentation
Paper
Graph methods
PAGA 1h 7m 55s 19s 25s 0.82
RaceID / StemID >7d 1d 1h 1h 14h 0.77
SLICER >7d >7d 2h 31s <1s 0.99
Tree methods
Slingshot >7d 11h 56m 2m 52s 0.98
PAGA Tree 2h 8m 1m 20s 15s 0.88
MST 56m 8m 12m 2m 52s 0.90 Off-the-shelf
pCreode >7d 1d 2h 3m 58s 0.89
SCUBA >7d 3d 4h 10m 3m 0.86
Monocle DDRTree 1h 26m 2h 14h 2d 0.86
Monocle ICA >7d 2d 1h 1h 1d 0.95
cellTree maptpx >7d >7d 6h 1h 24m 0.51
SLICE >7d >7d 1h 51m 1d 0.78
cellTree VEM >7d 23h 39m 10m 15m 0.78
ElPiGraph 12h 1d 6h 20m 7m 0.93
Sincell >7d >7d 2h 5m 2m 0.97
URD >7d 1d 2h 10m 1m 0.68
CellTrails >7d >7d 2d 7h 4h 0.76
Mpath >7d >7d 8h 4h 1d 0.90
CellRouter >7d 1d 1h 9m 9m 0.24


Do you expect multiple disconnected trajectories?
Yes / I don't know
No
Do you expect a particular topology?
No
Yes
Do you expect cycles in the topology?
Yes / I don't know
No
Do you expect a tree with two or more bifurcations?
No / I don't know
Yes
▲ Free topology
▼ Fixed topology
≤ Disconnected
PAGA
RaceID / StemID
≤ Graph
PAGA
RaceID / StemID
SLICER
≤ Tree
Slingshot
PAGA
Monocle ICA
MST
Tree
PAGA
MST
Slingshot
RaceID / StemID
Multifurcation
STEMNET
Slingshot
PAGA
FateID
Bifurcation
Slingshot
FateID
GrandPrix
STEMNET
Linear
SCORPIUS
Embeddr
TSCAN
Slingshot
Cycle
Angle
ElPiGraph cycle
reCAT
RaceID / StemID
Confirm expectations using a method with free topology
Confirm results using at least 2 methods
dynverse
Check out the interactive guidelines at guidelines.dynverse.org

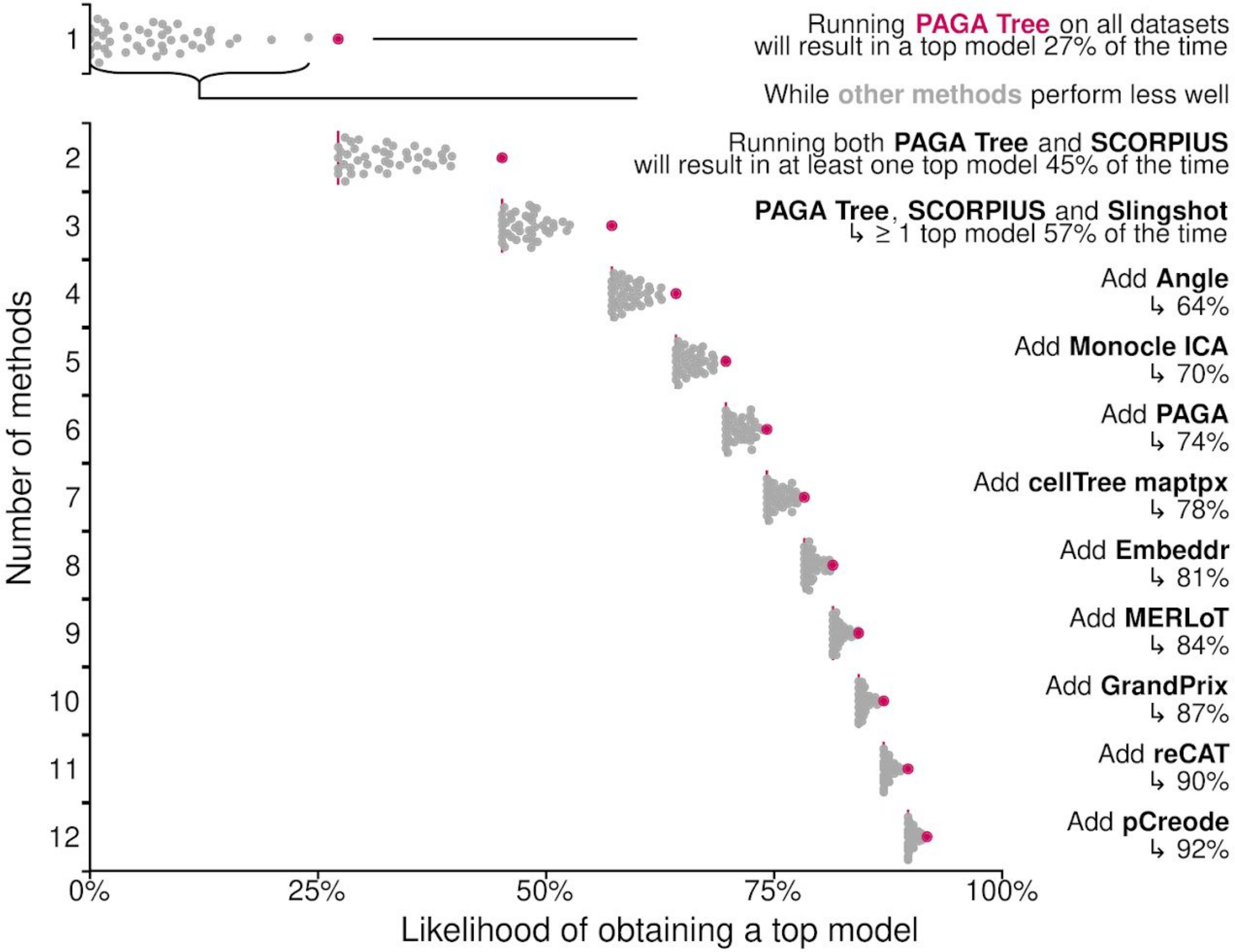
Large variability in scores



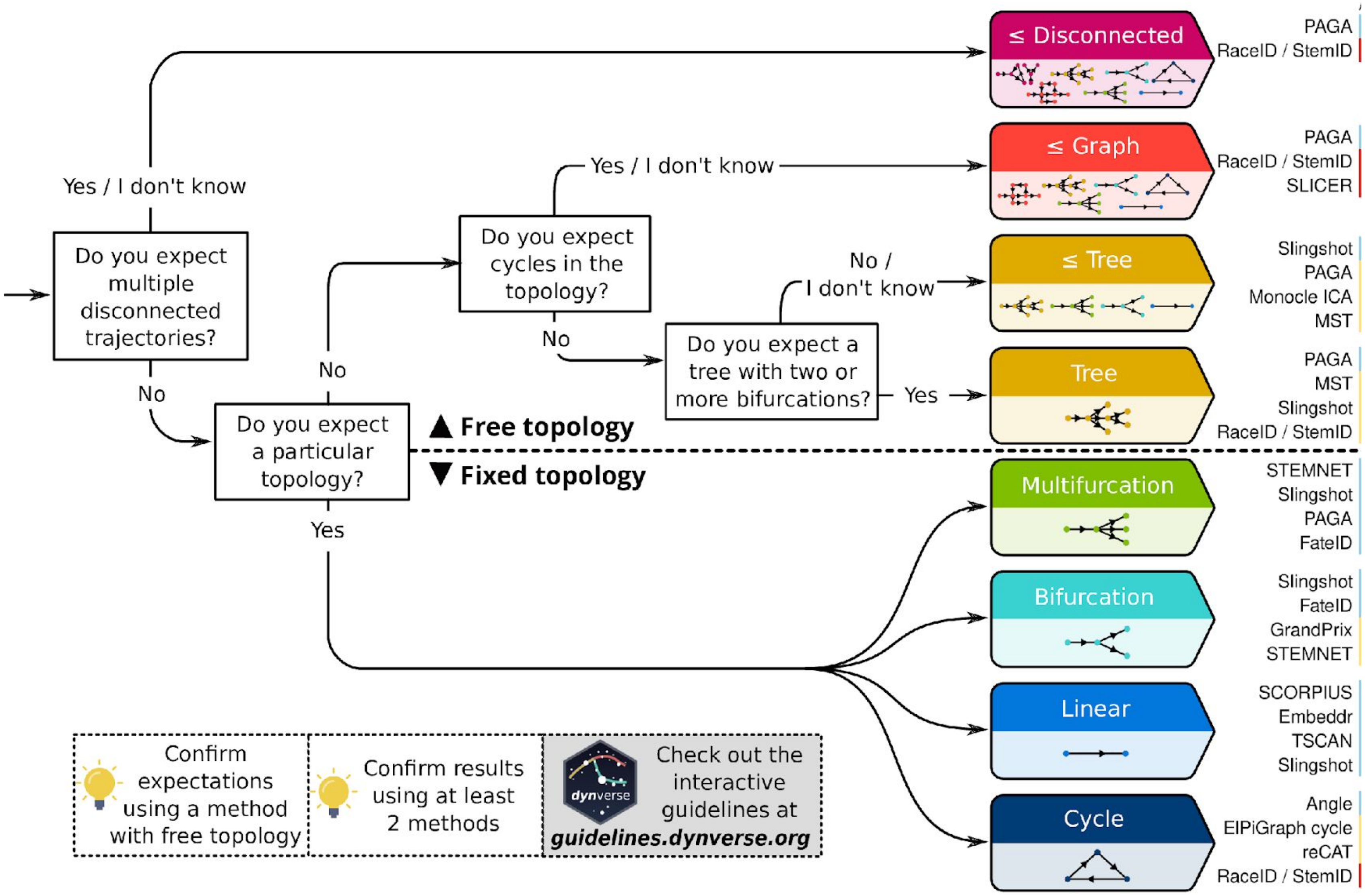
Large variability in scores



Complementarity



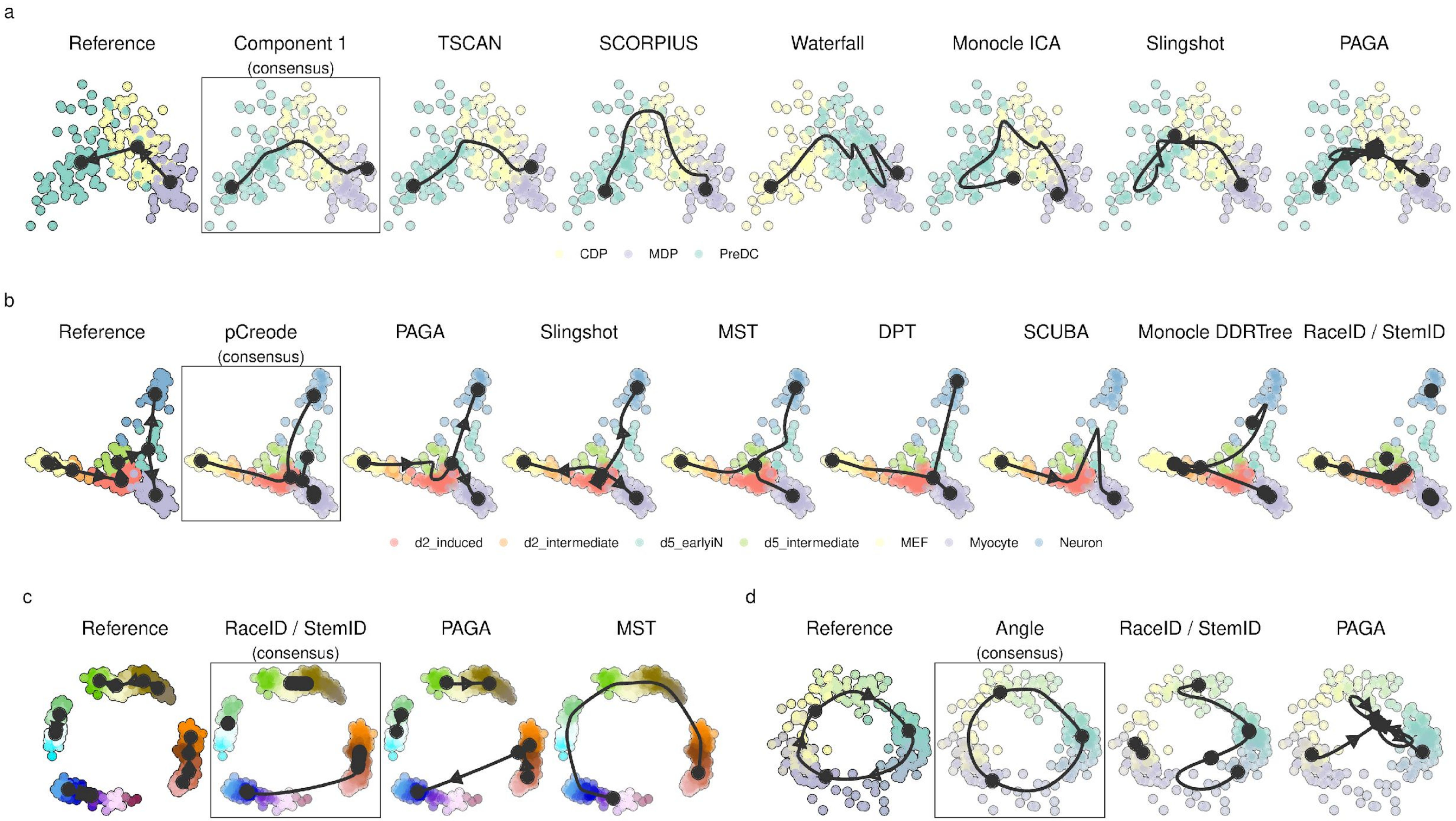
= at least 95% score of the best model



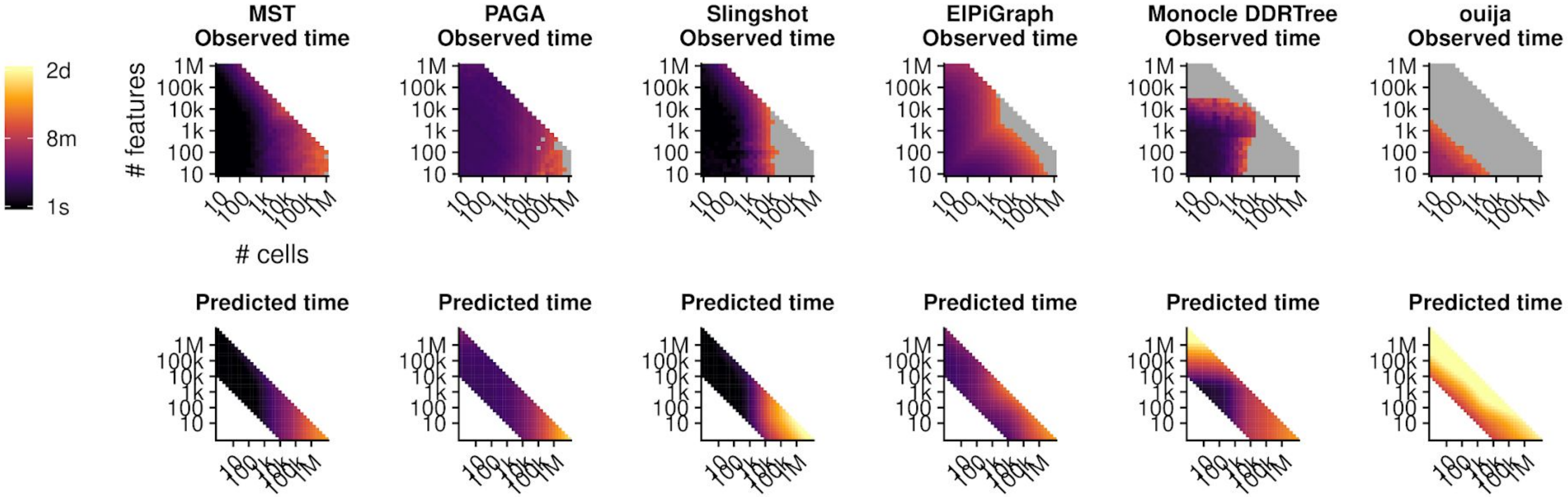
Confirm expectations using a method with free topology

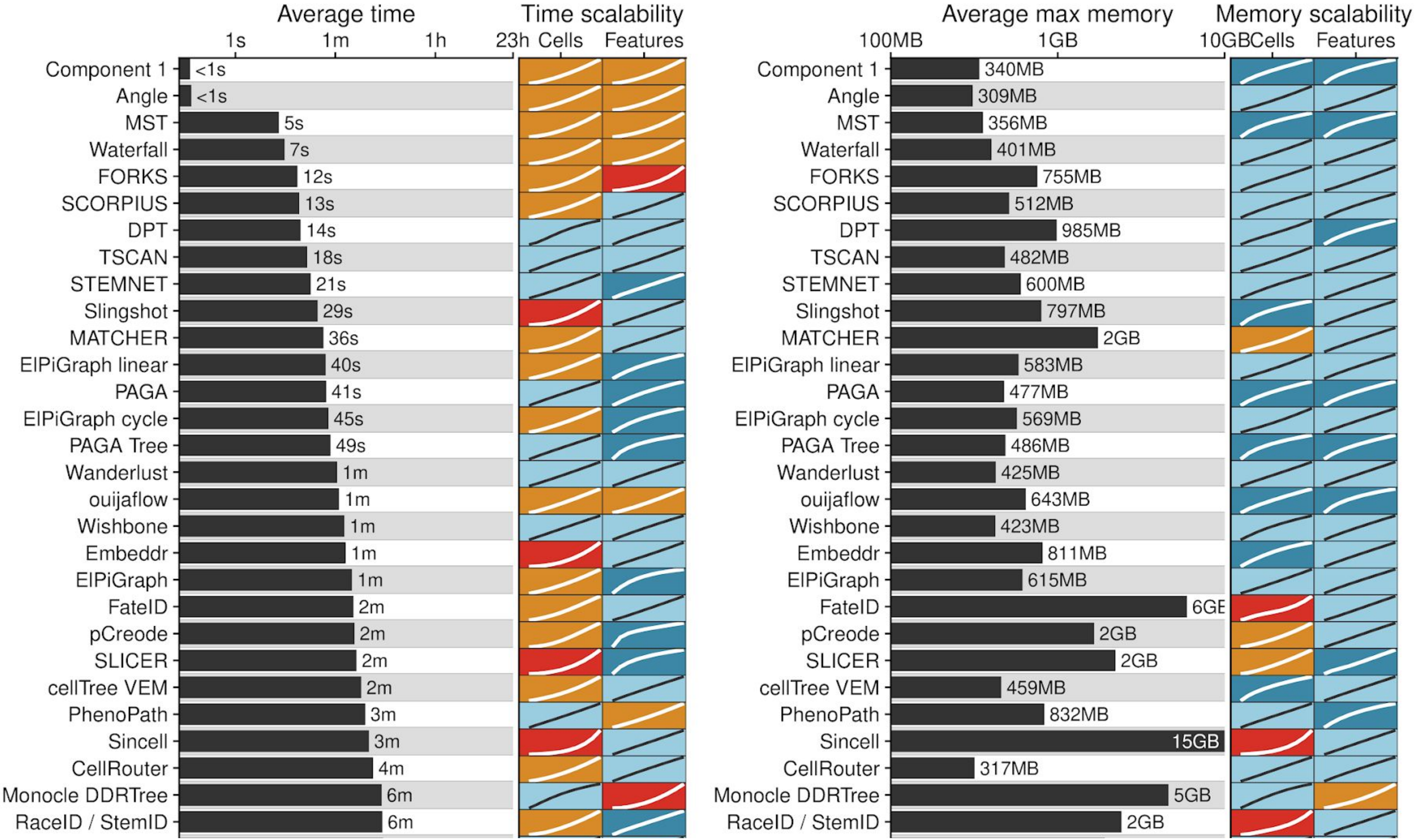
Confirm results using at least 2 methods

Check out the interactive guidelines at ***guidelines.dynverse.org***


a
Reference
Component 1
(consensus)
TSCAN
SCORPIUS
Waterfall
Monocle ICA
Slingshot
PAGA
CDP
MDP
PreDC
b
Reference
pCreode
(consensus)
PAGA
Slingshot
MST
DPT
SCUBA
Monocle DDRTree
RaceID / StemID
d2_induced
d2_intermediate
d5_earlyiN
d5_intermediate
MEF
Myocyte
Neuron
c
Reference
RaceID / StemID
(consensus)
PAGA
MST
d
Reference
Angle
(consensus)
RaceID / StemID
PAGA

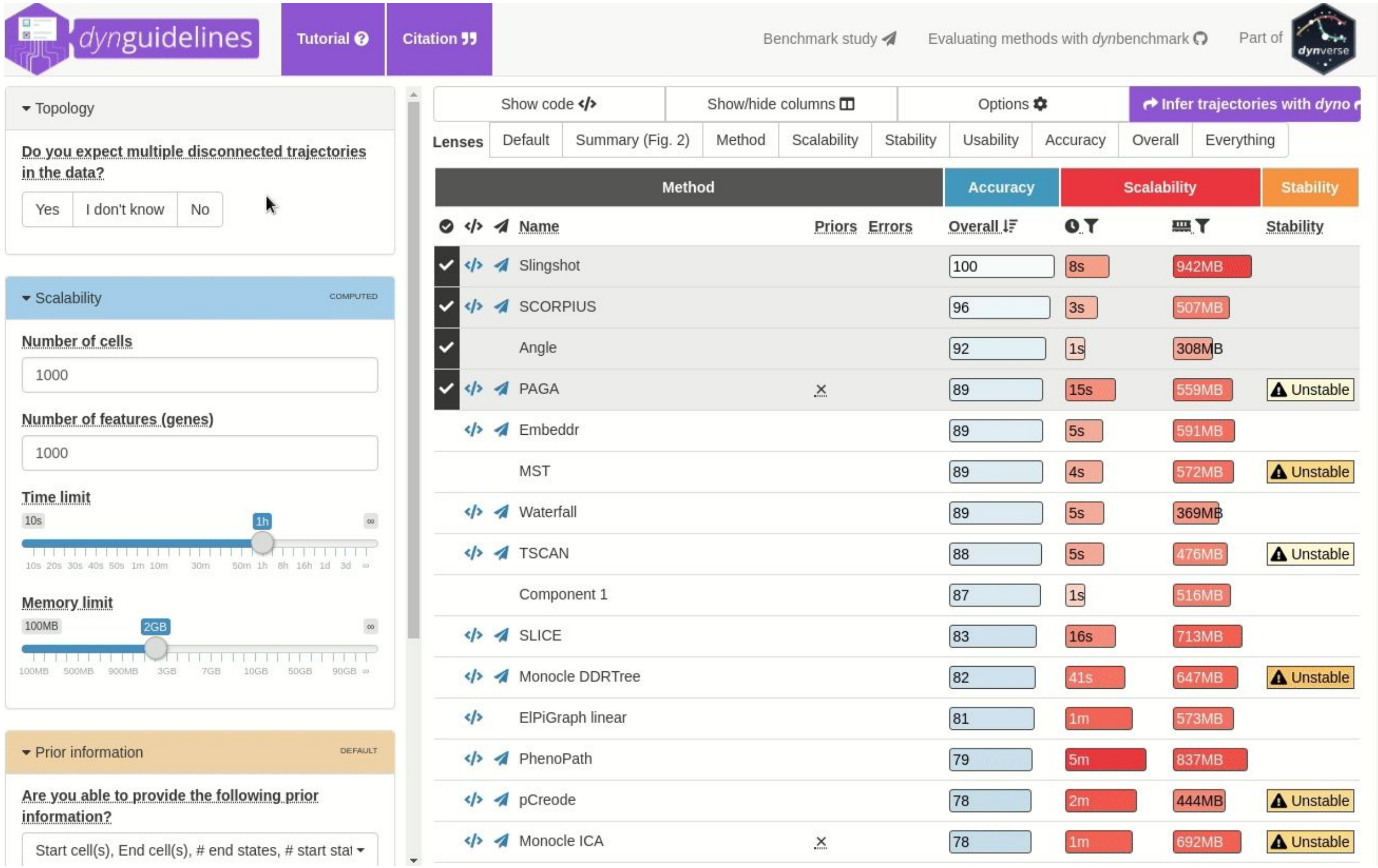
Scalability




Average time
Time scalability
1s
1m
1h
23h
Cells
Features
Average max memory
Memory scalability
100MB
1GB
10GB
Cells
Features
Component 1
<1s
340MB
Angle
<1s
309MB
MST
5s
356MB
Waterfall
7s
401MB
FORKS
12s
755MB
SCORPIUS
13s
512MB
DPT
14s
985MB
TSCAN
18s
482MB
STEMNET
21s
600MB
Slingshot
29s
797MB
MATCHER
36s
2GB
ElPiGraph linear
40s
583MB
PAGA
41s
477MB
ElPiGraph cycle
45s
569MB
PAGA Tree
49s
486MB
Wanderlust
1m
425MB
ouijaflow
1m
643MB
Wishbone
1m
423MB
Embeddr
1m
811MB
ElPiGraph
1m
615MB
FateID
2m
6GE
pCreode
2m
2GB
SLICER
2m
2GB
cellTree VEM
2m
459MB
PhenoPath
3m
832MB
Sincell
3m
15GB
CellRouter
4m
317MB
Monocle DDRTree
6m
5GB
RaceID / StemID
6m
2GB

1. Selecting the most optimal method(s)

`guidelines.dynverse.org`



Topology

Do you expect multiple disconnected trajectories in the data?

Yes | I don't know | No

Scalability COMPUTED

Number of cells: 1000

Number of features (genes): 1000

Time limit: 1h (10s – ∞)

Memory limit: 2GB (100MB – ∞)

Prior information DEFAULT

Are you able to provide the following prior information?

Start cell(s), End cell(s), # end states, # start stat

Method					Accuracy	Scalability		Stability
✓	</>	Name	Priors	Errors	Overall	🕓	▦	Stability
✓	</>	Slingshot			100	8s	942MB	
✓	</>	SCORPIUS			96	3s	507MB	
✓		Angle			92	1s	308MB	
✓	</>	PAGA	×		89	15s	559MB	⚠ Unstable
	</>	Embeddr			89	5s	591MB	
		MST			89	4s	572MB	⚠ Unstable
	</>	Waterfall			89	5s	369MB	
	</>	TSCAN			88	5s	476MB	⚠ Unstable
		Component 1			87	1s	516MB	
	</>	SLICE			83	16s	713MB	
	</>	Monocle DDRTree			82	41s	647MB	⚠ Unstable
	</>	ElPiGraph linear			81	1m	573MB	
	</>	PhenoPath			79	5m	837MB	
	</>	pCreode			78	2m	444MB	⚠ Unstable
	</>	Monocle ICA	×		78	1m	692MB	⚠ Unstable

2. Inferring trajectories

- A common input and output interface for 55 methods

Input ▷ **Method** ▷ **Output**

Input	Output
Raw counts	Topology
Normalised	Cell positions
Start cell(s)	
End cell(s)	
Cell groups	

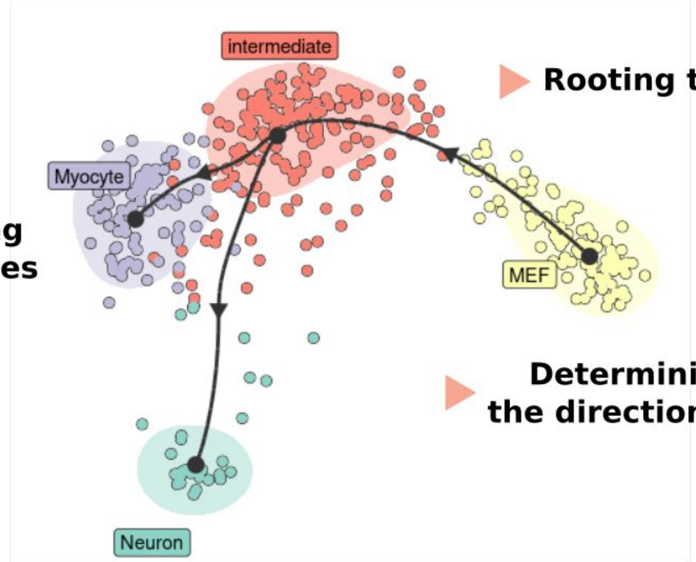
- One line and run any method:

```
infer_trajectory(dataset, "paga")
infer_trajectory(dataset, "slingshot")
infer_trajectory(dataset, "my_favorite_ti_method")
```

- Easy to include a new method

`methods.dynverse.org`

3. Annotating the trajectory



Rooting the trajectory

Labelling milestones

Rooting the trajectory

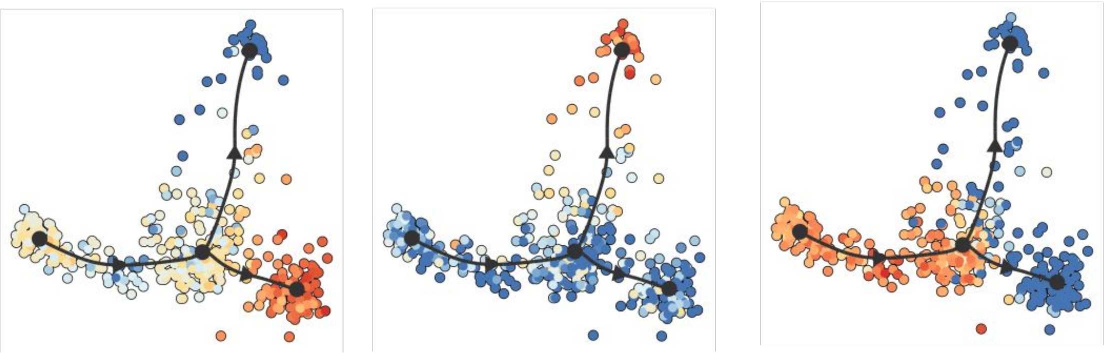
Determining the directionality

4. Detecting differential expression

Overall

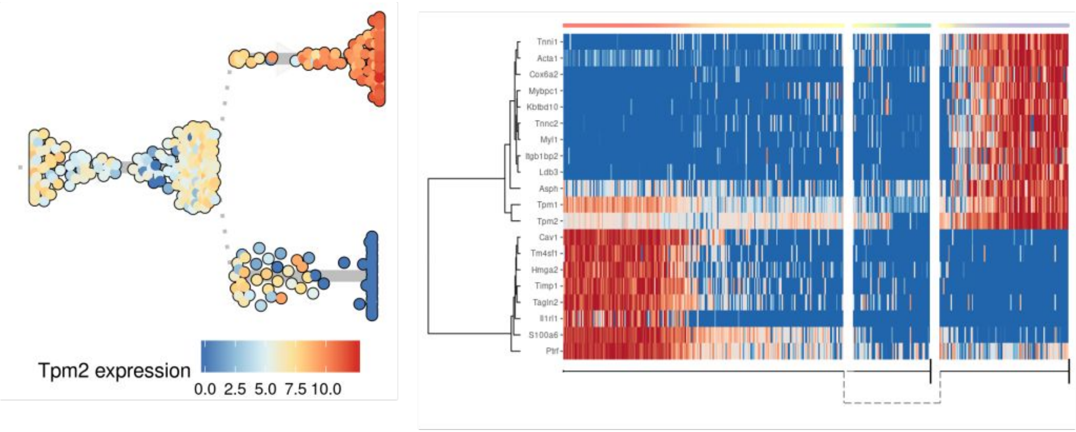
Branch

Branch point

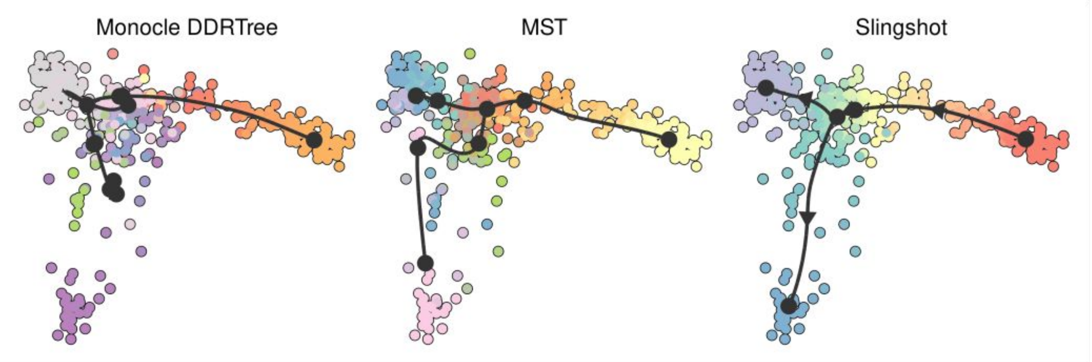


5. Visualizing the trajectory

Visualize a model in **multiple ways**



Compare models on a **common embedding**



Why use dyno?

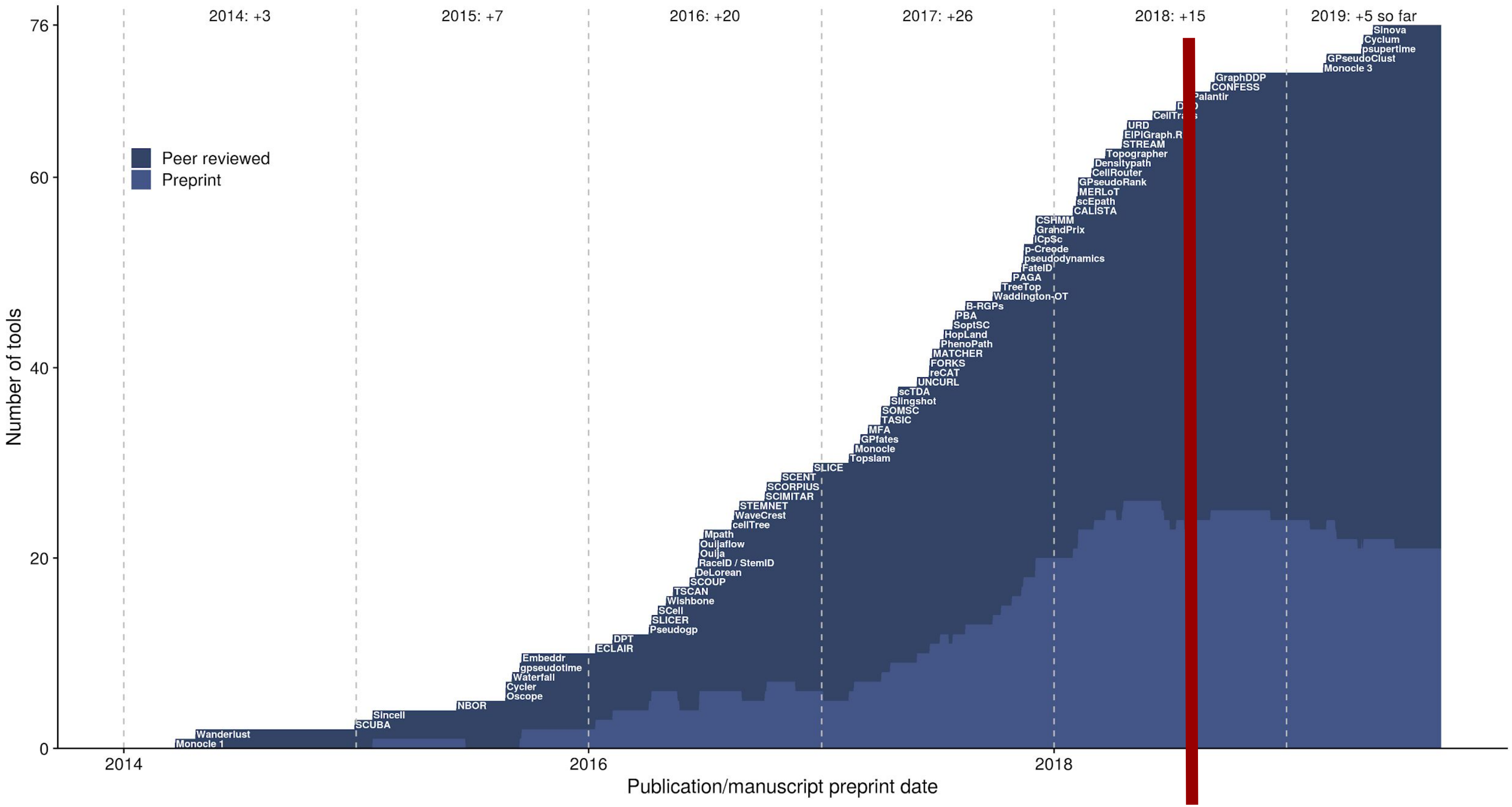
Now	**In the near future**
Tight integration with benchmarking, use the state-of-the-art for your dataset	Benchmark new methods
Modular and consistent interface to each method	Include additional inputs (e.g. RNA velocity)
Annotate a trajectory, using your own insights into the data	Add more ways to annotate (suggestions welcome!)
Detect genes that change at any step in the trajectory	Other powerful analyses (e.g. trajectory alignment)
Look at the trajectory in multiple ways and easily compare different models	Interactive plotting of a trajectory

* 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

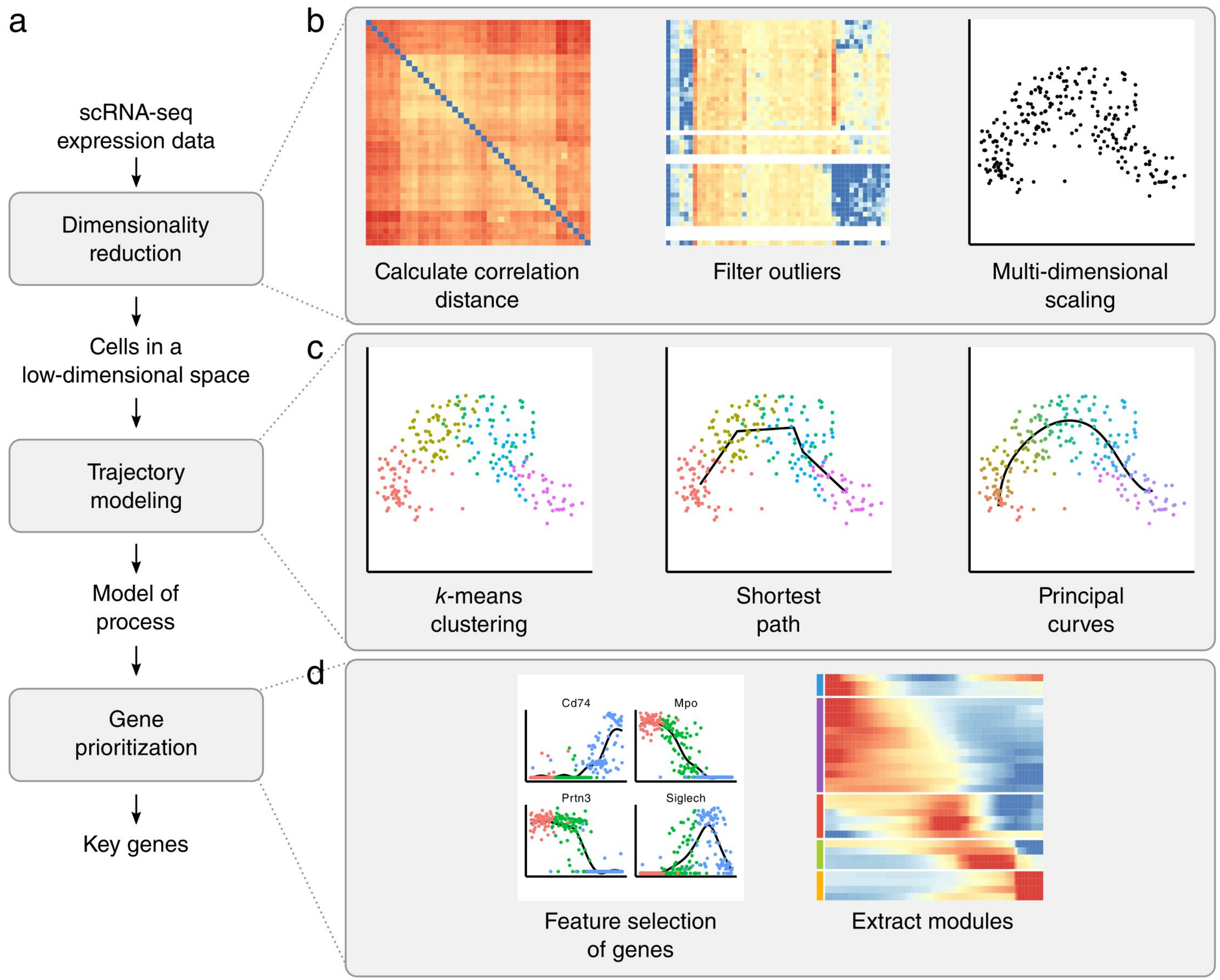
GPLVM

...

SCORPIUS

https://doi.org/10.1101/079509

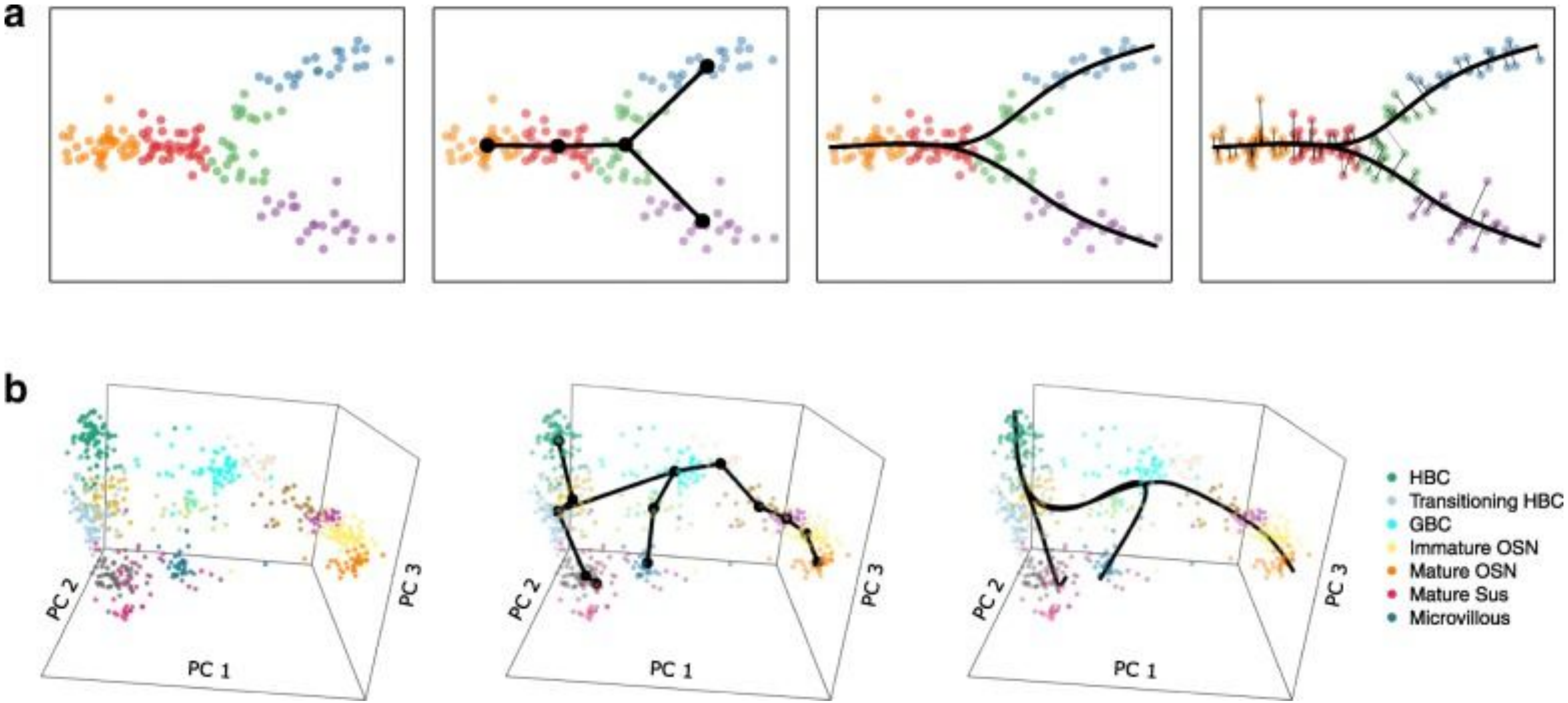
Principal curves



Slingshot

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

Principal curves



Dimensionality reduction & clustering

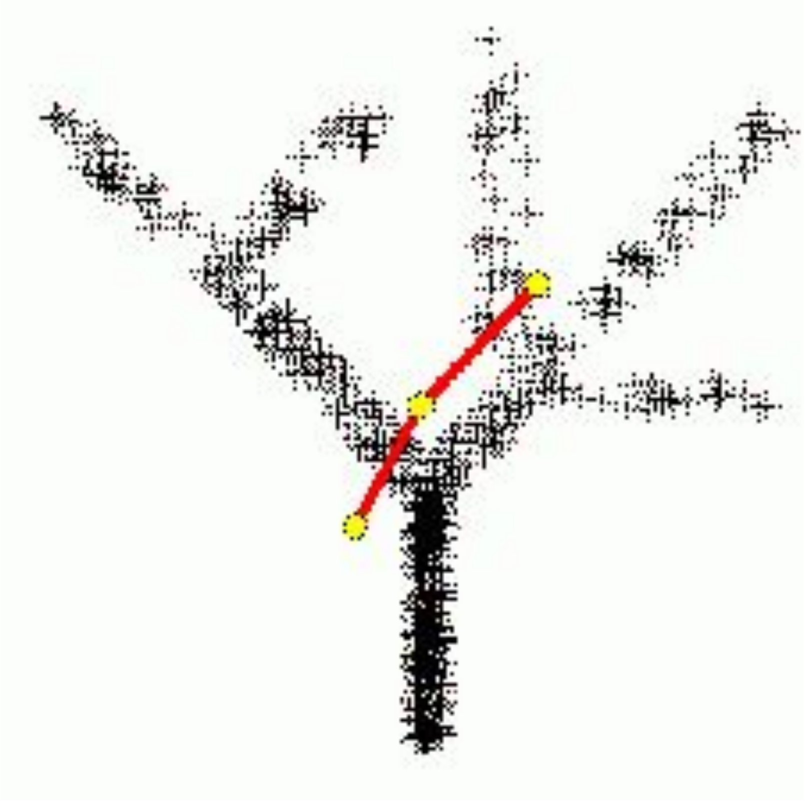
MST

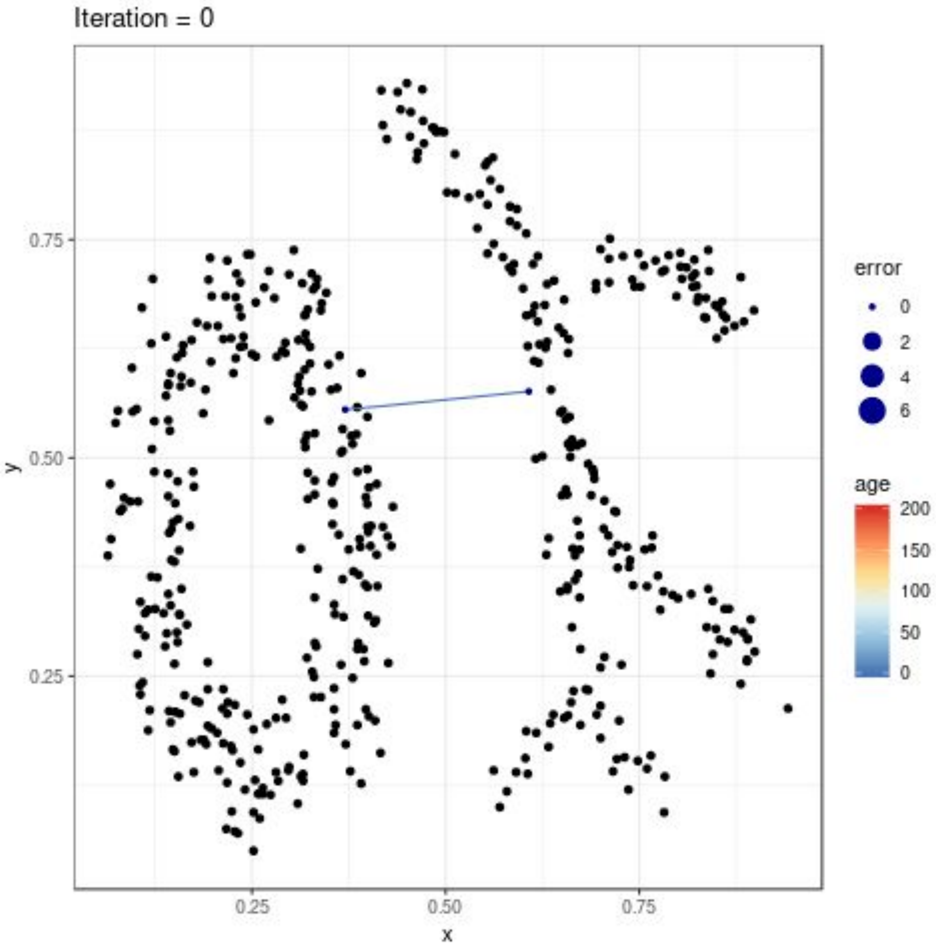
Principal curves

Projection onto curve

Principal graphs and neural gas

e.g. Monocle DDRTree

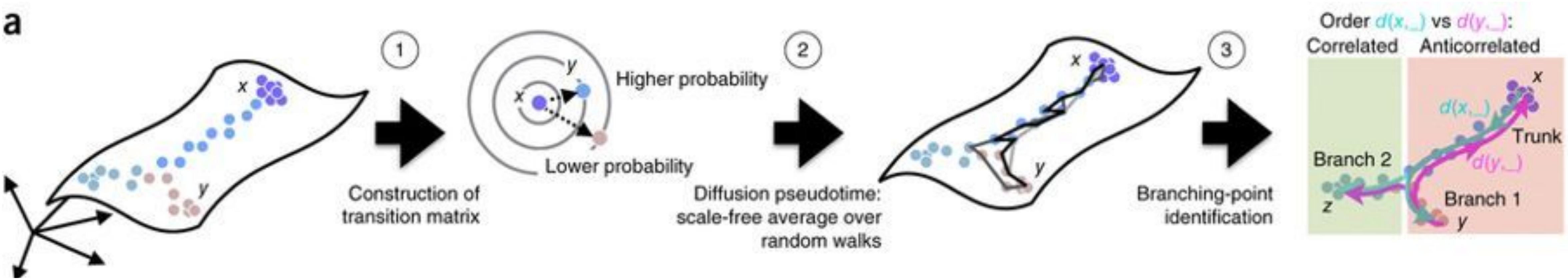




Diffusion pseudotime (DPT)

Graph diffusion

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



~ geodesic distance
over knn graph

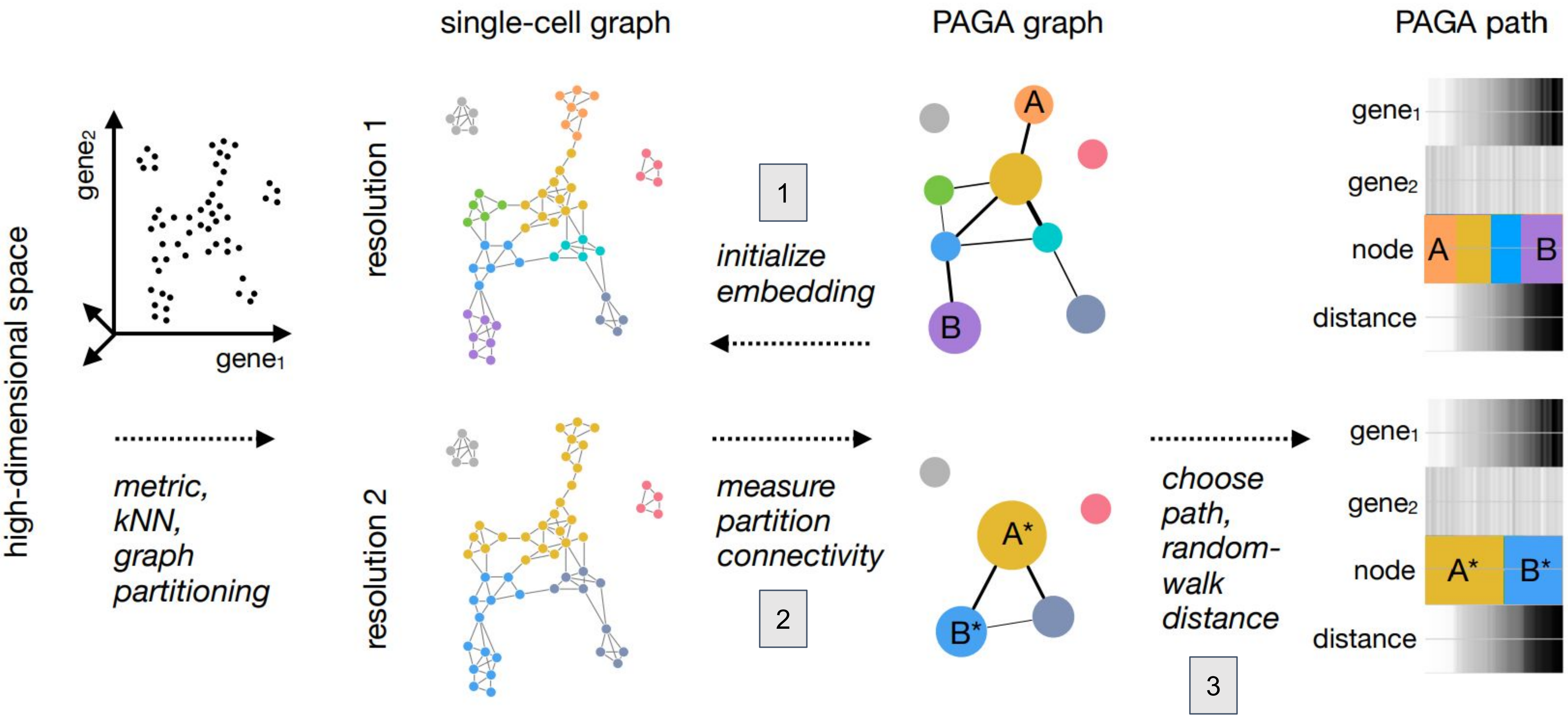
~ random walks

~ anticorrelated pseudotime
=> branching

PAGA

Graph diffusion

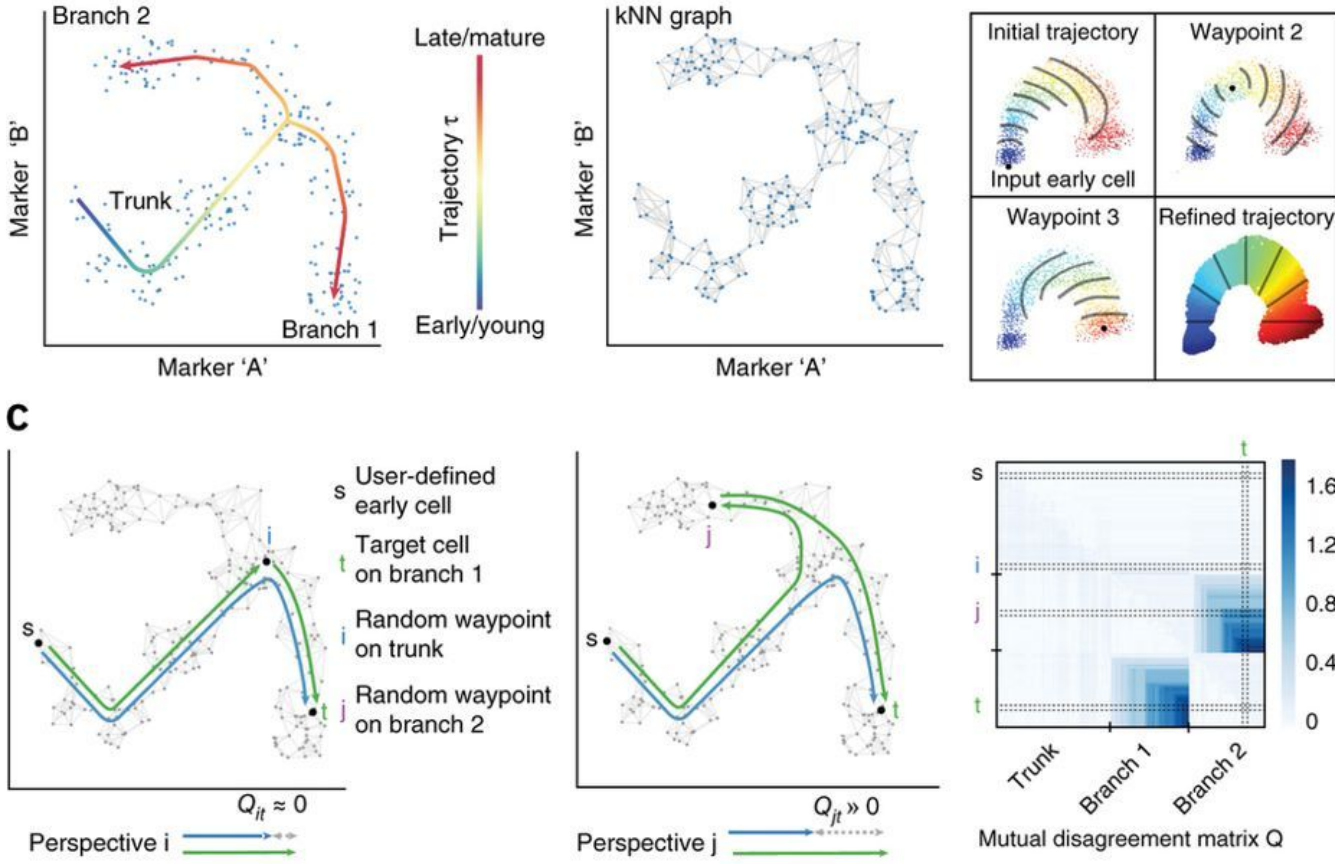
https://doi.org/10.1101/208819

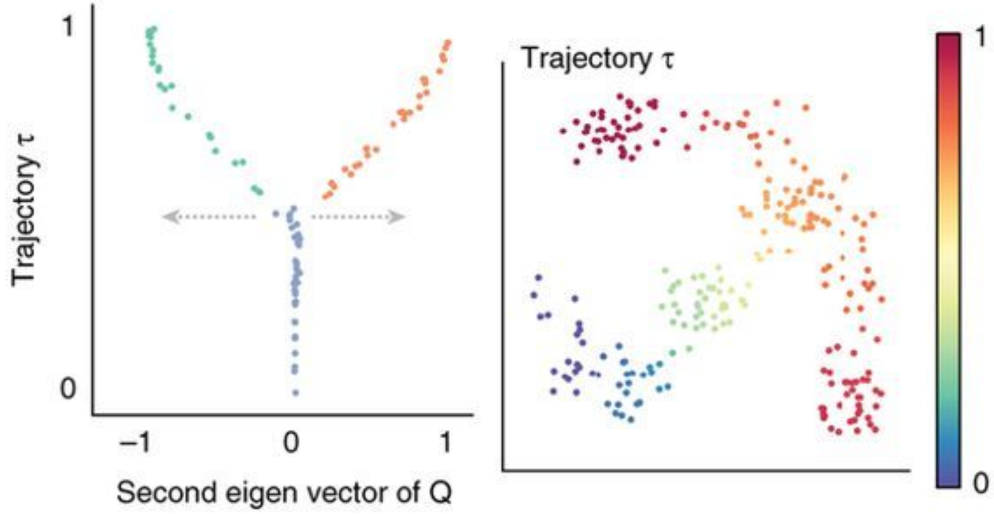


Wanderlust/Wishbone

KNN

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

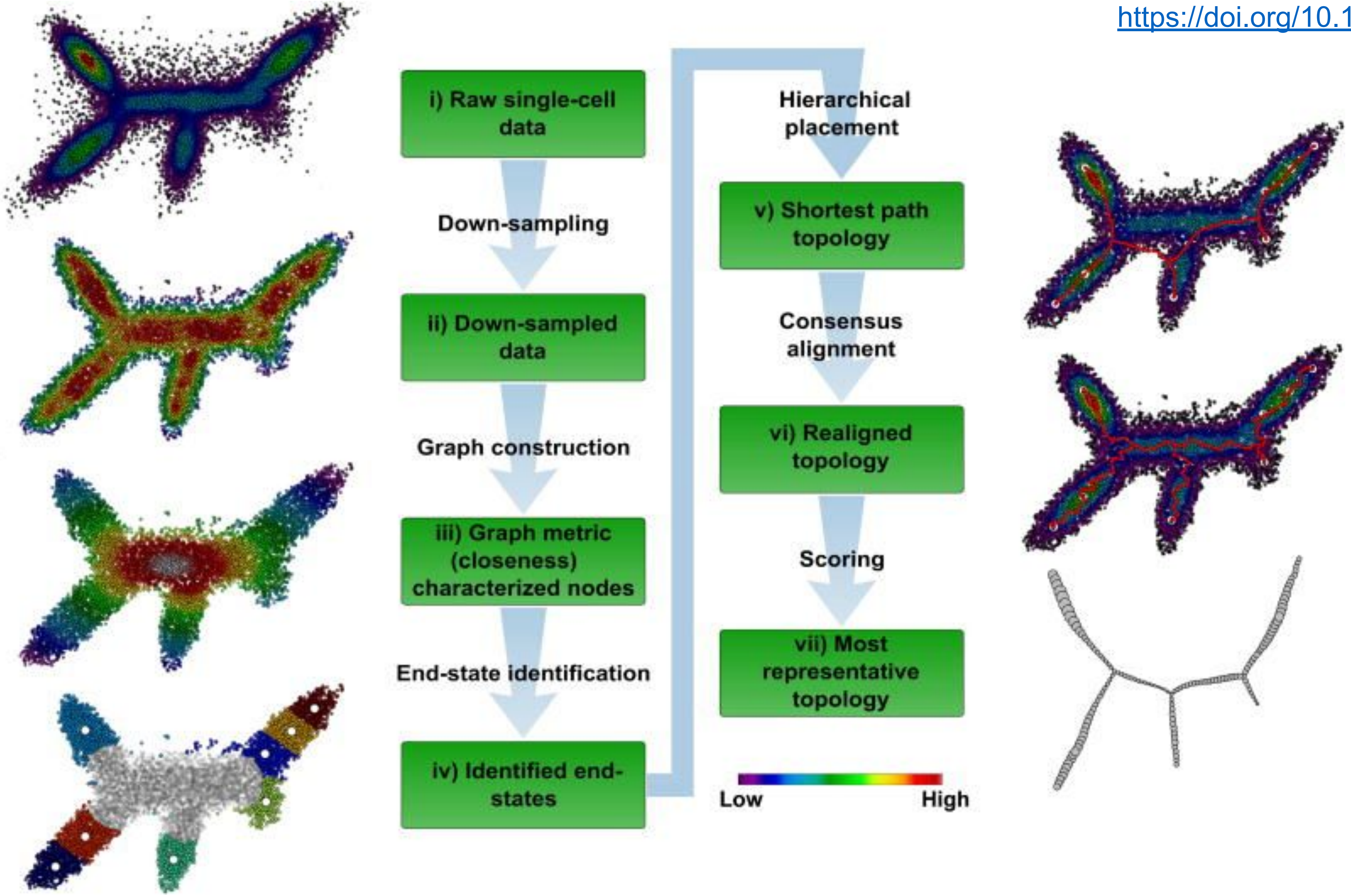




p-Creode

KNN

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



GrandPrix, GPfates, ...

- Model gene expression as a function of pseudotime: (B)GPLVM
  - **G**aussian **p**rocess: "Regression", but non-parametric.
    I.e. do not assume a function but test all possible ones within constraints
  - **L**atent **v**ariable: 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
  - **B**ayesian: Include some prior information
    I.e. Known experimental time points

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

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

- Branching?
  - Mixture GPLVM, with multiple functions over time

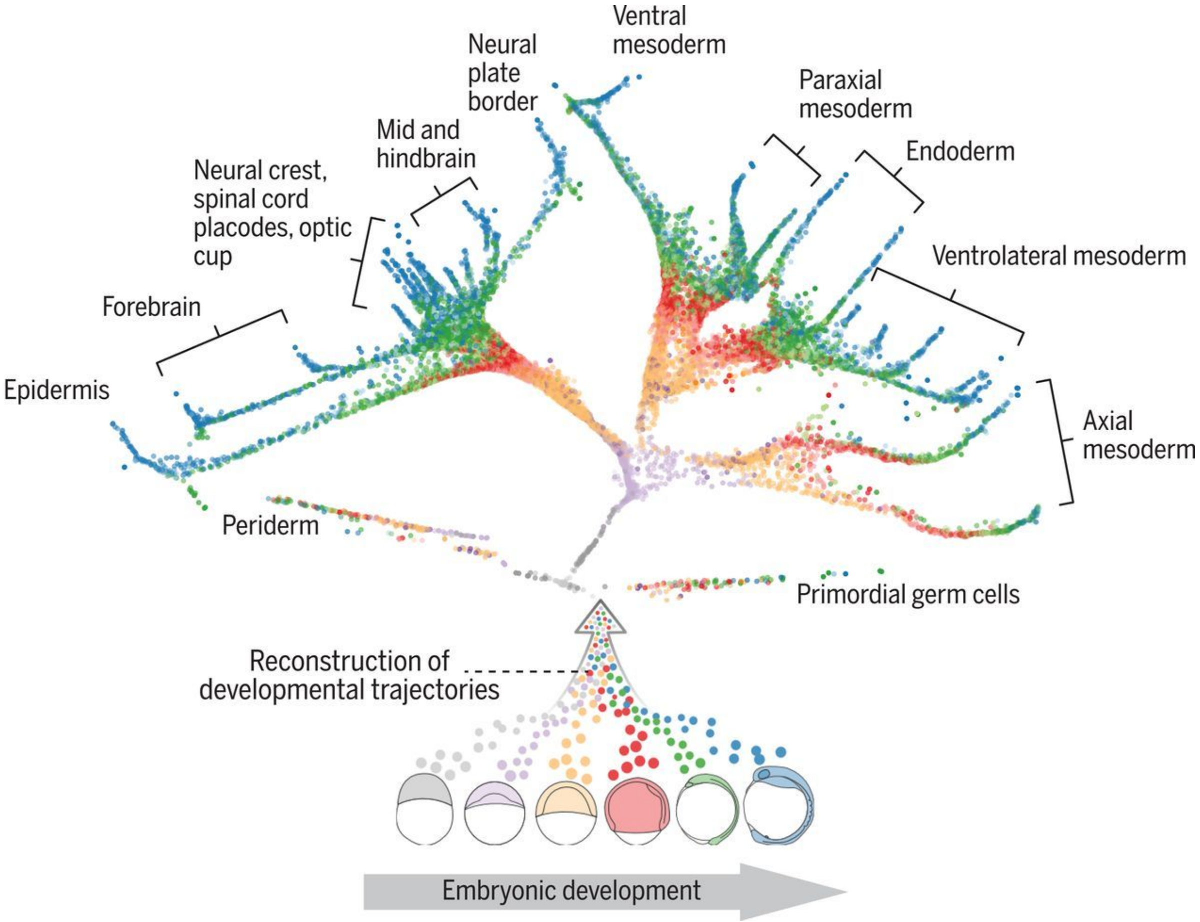
How to analyze single-cell dynamics

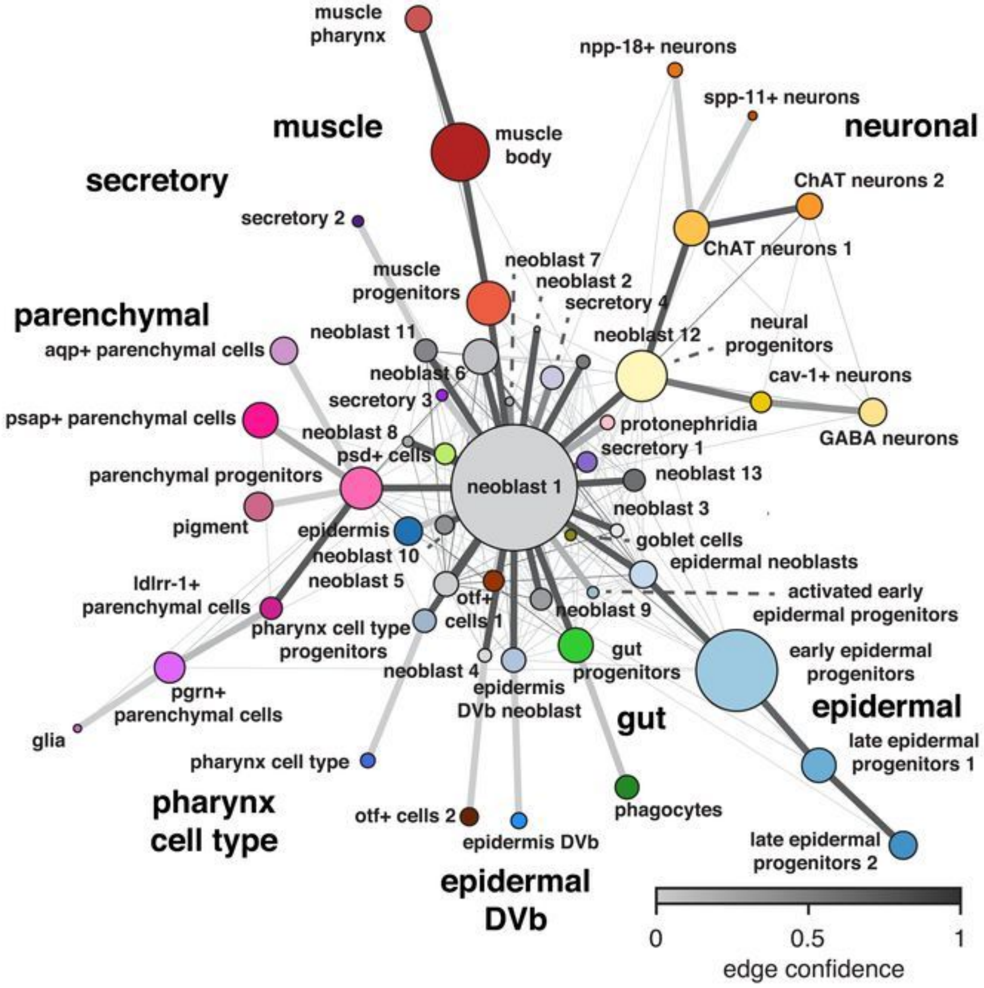
Current state of trajectory inference

Some internal methodology

Extensions

Cautionary note

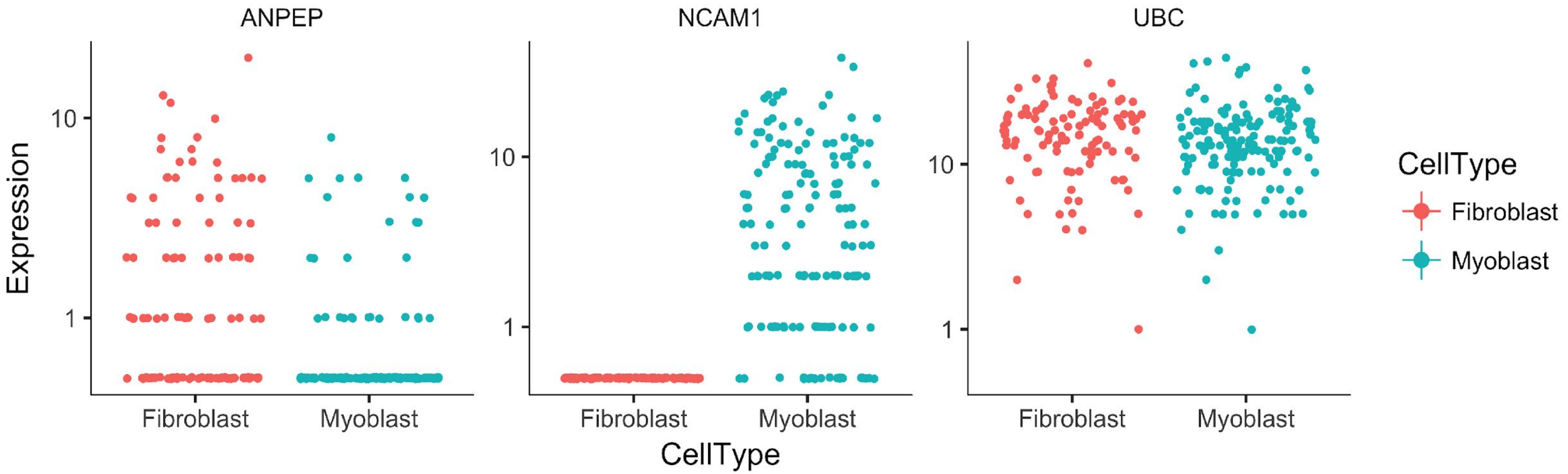




Clean and iteratively filtered data + A LOT of parameter tuning

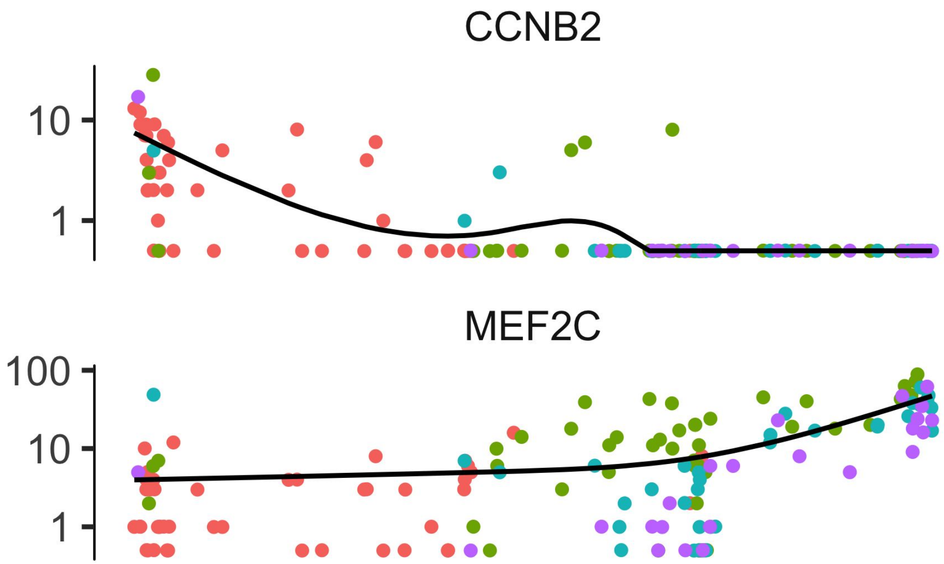
Trajectory differential expression

Between states

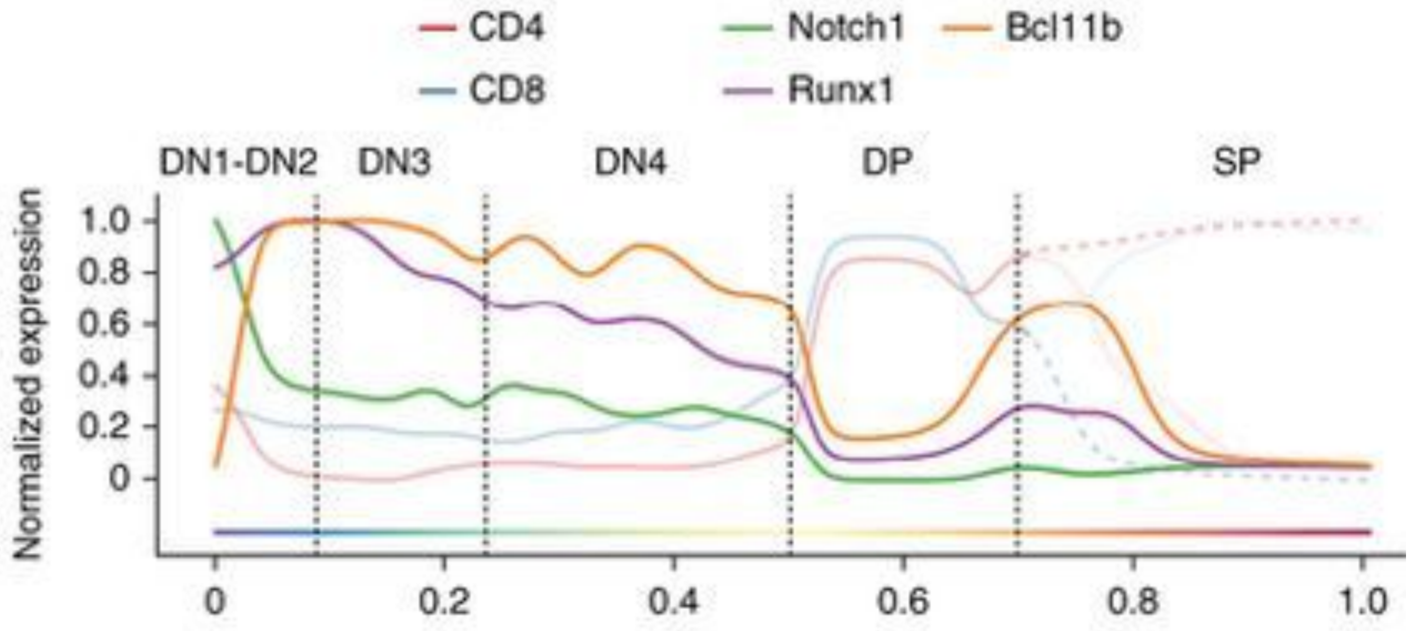


Trajectory differential expression

Monocle v2

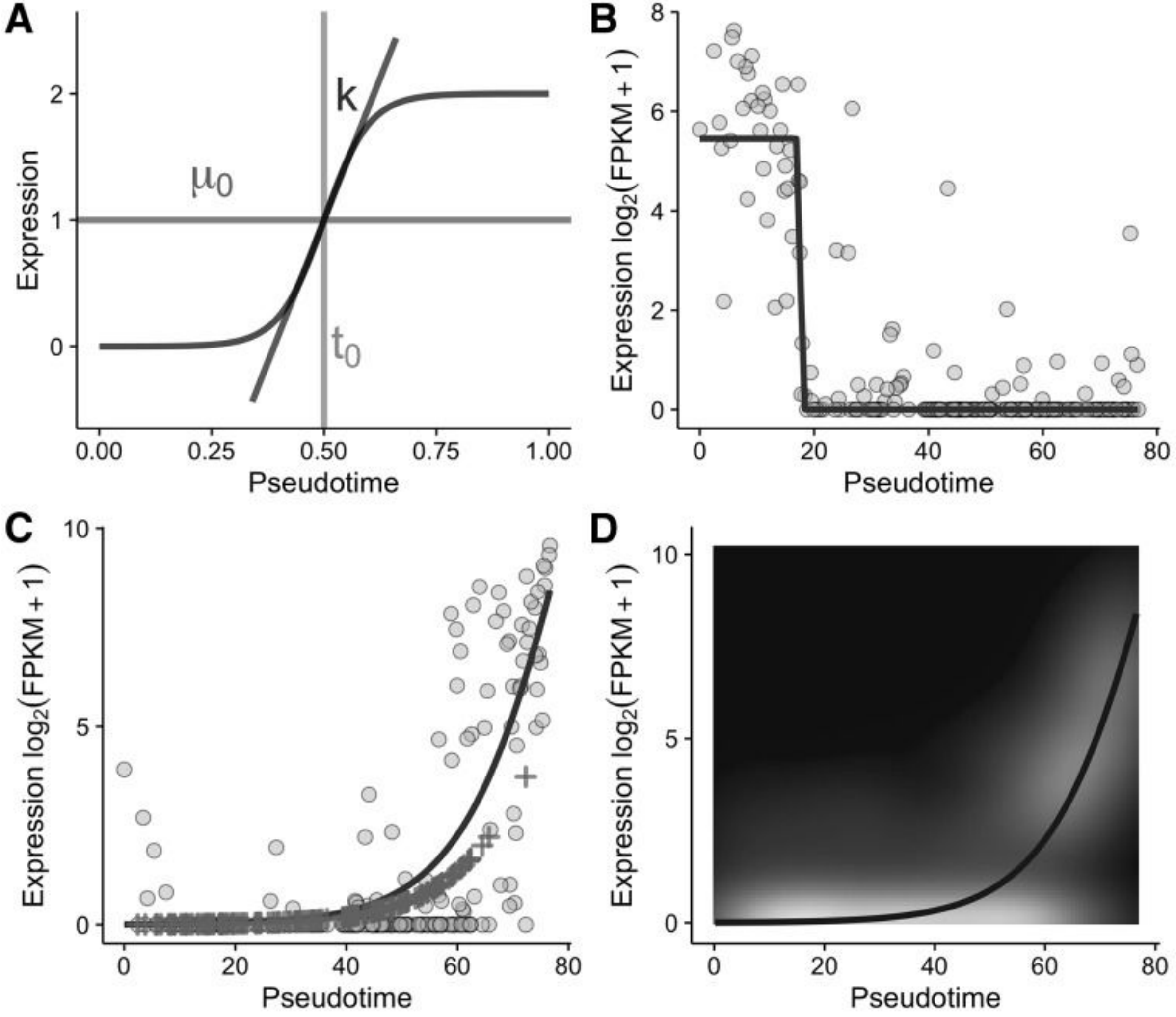


Wishbone

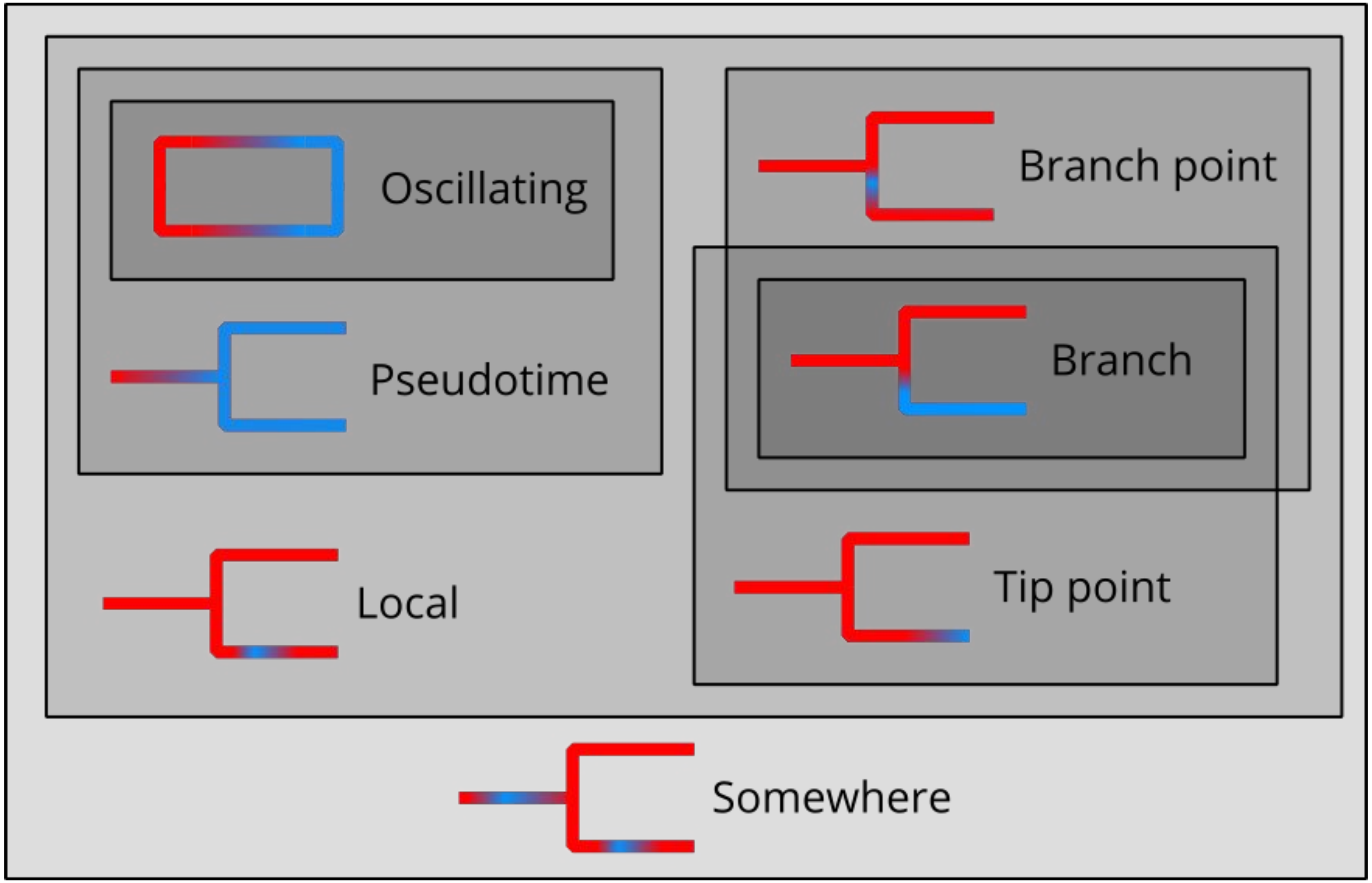


Trajectory differential expression

switchDE



Trajectory differential expression

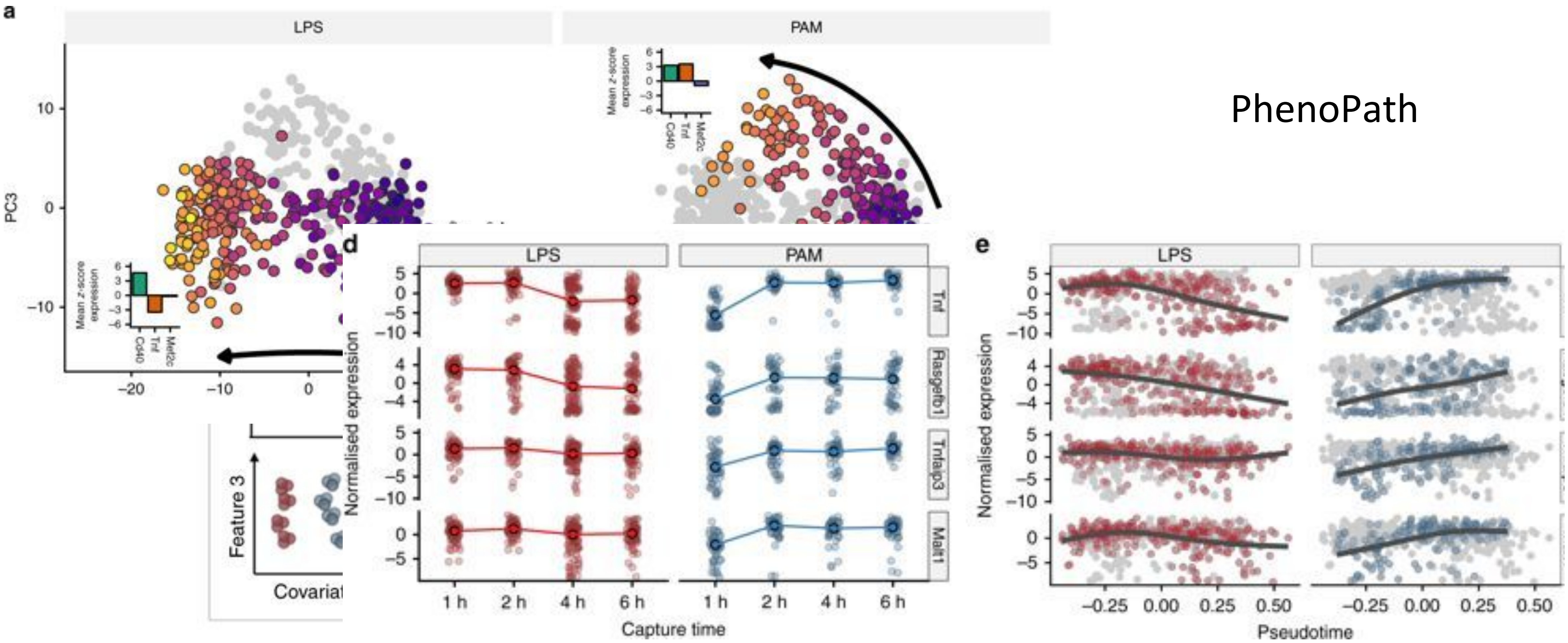


Trajectory differential expression

Current limitations

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

Alignment

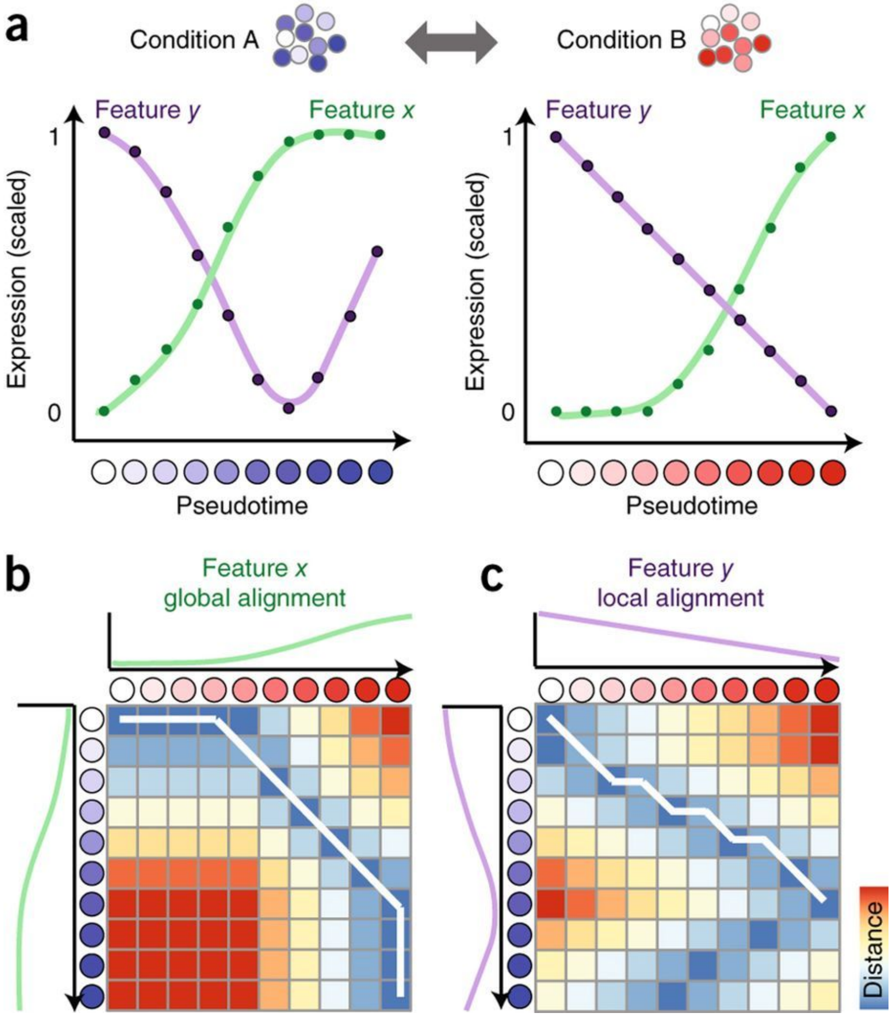


PhenoPath

Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data https://doi.org/10.1038/s41467-018-04696-6

Alignment

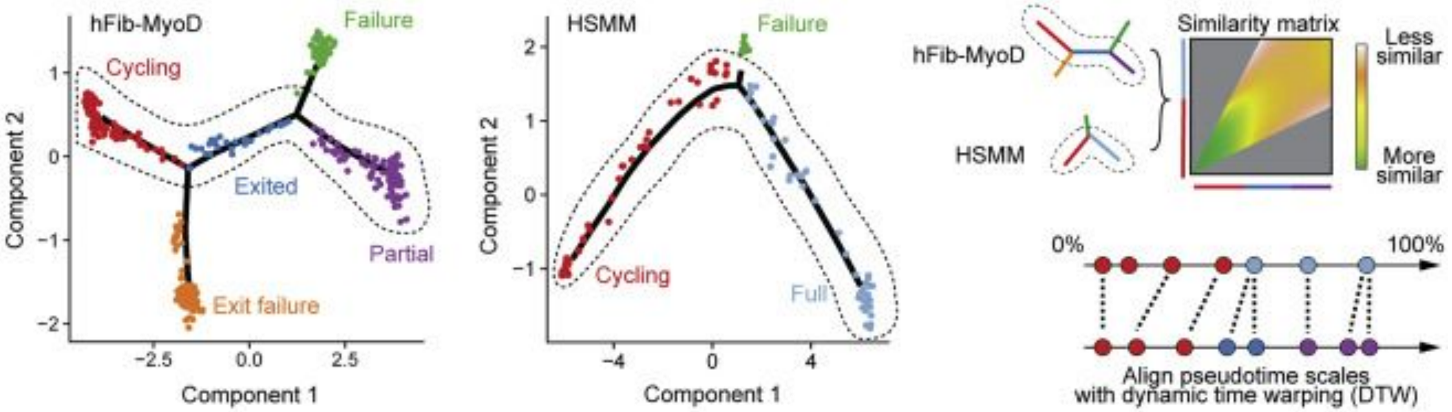
CellAlign



Alignment of single-cell trajectories to compare cellular expression dynamics https://doi.org/10.1038/nmeth.4628

Alignment

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



Current limitations

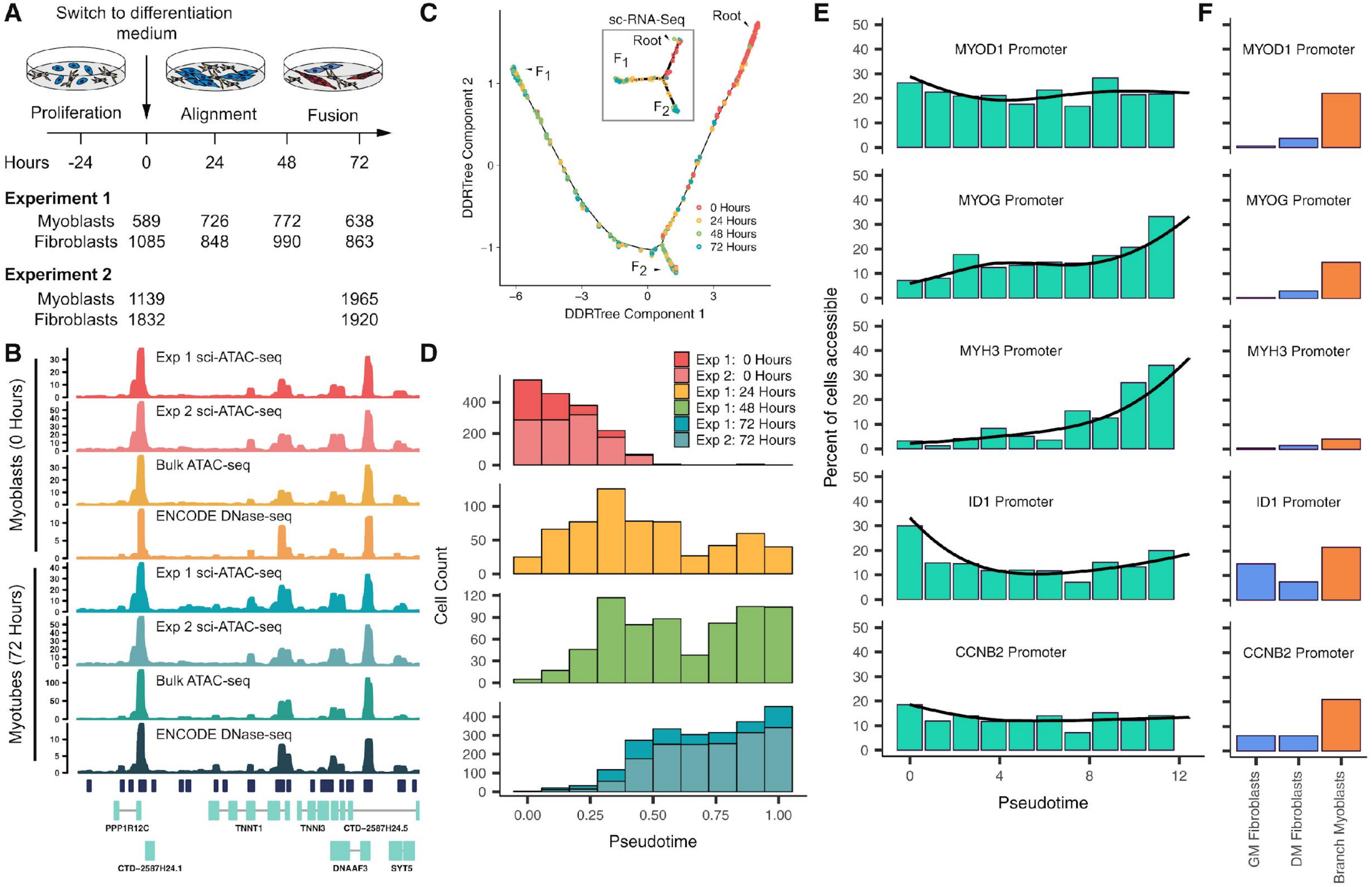
- No alignment of branches
- Assumes approx. same start and end state
- "Batch" sensitivity unclear

Non-transcriptome trajectory inference

ATAC-Seq

CITE-seq

…

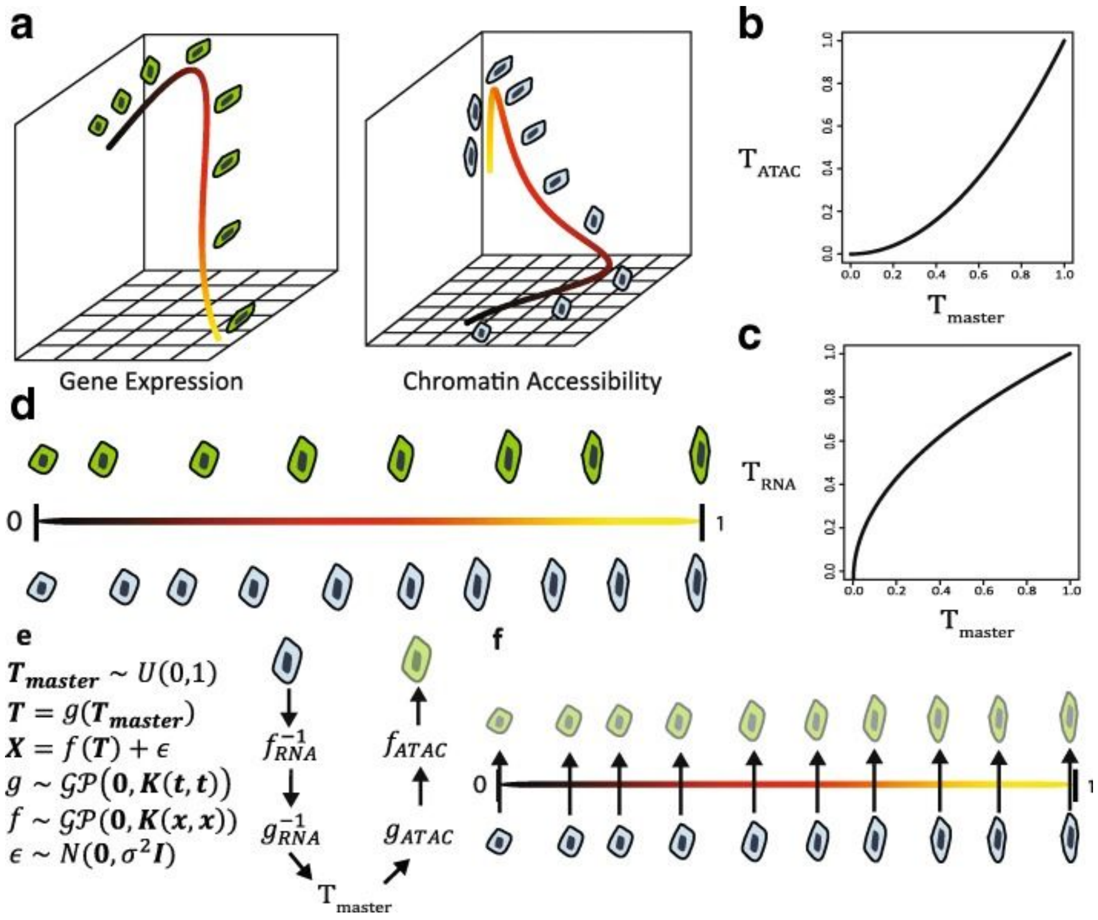


Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data.

https://doi.org/10.1016/j.molcel.2018.06.044

Multi-omics trajectory inference

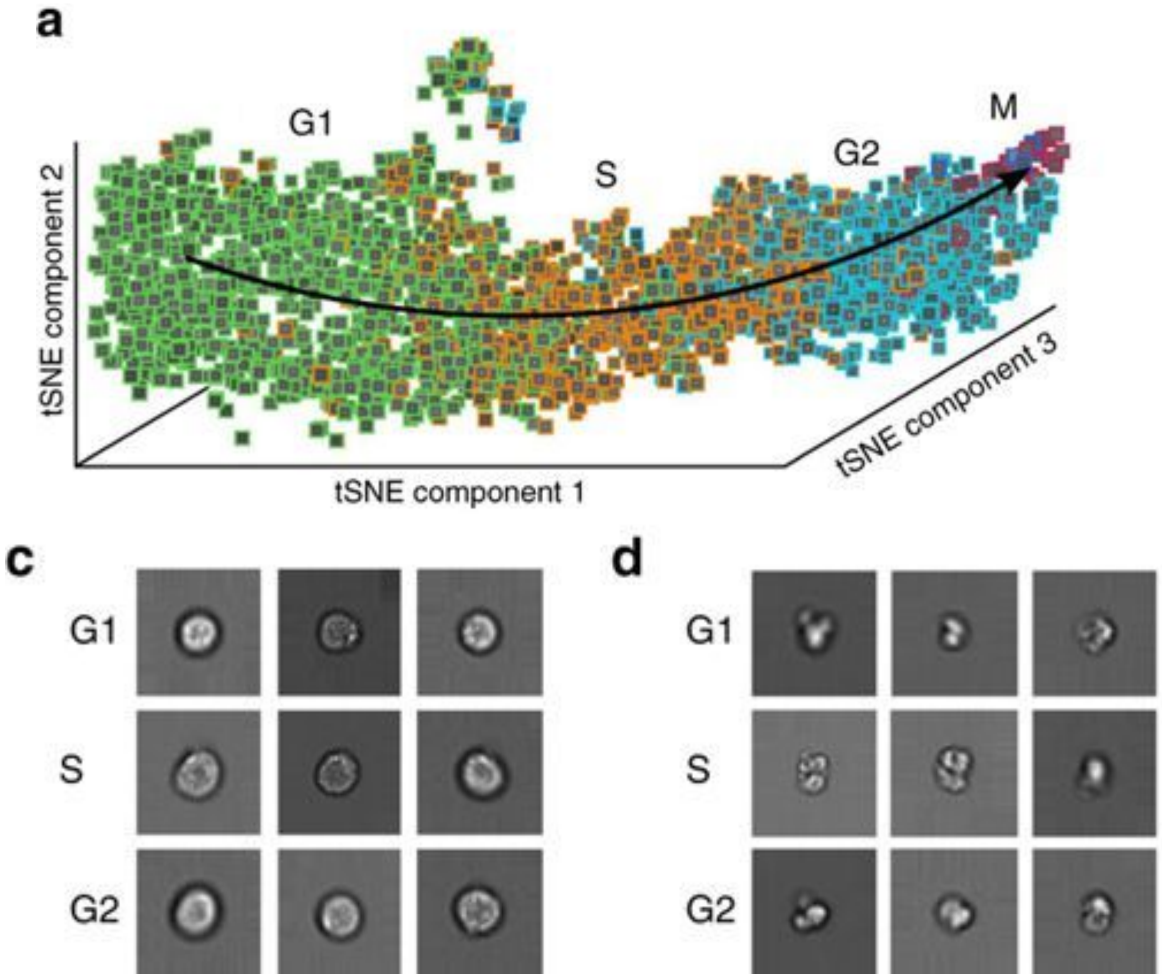
MATCHER

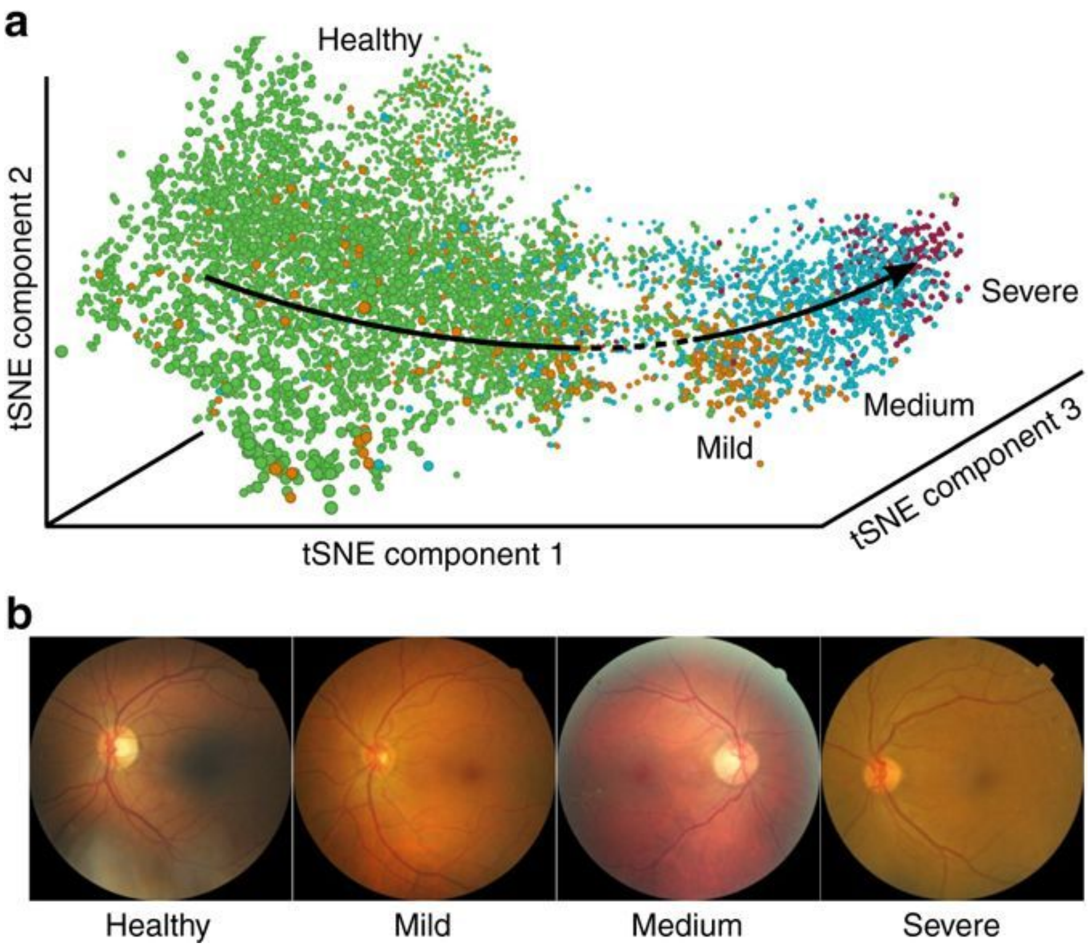


Current limitations

- No branching
- Limited accuracy & scalability

Trajectory inference on images

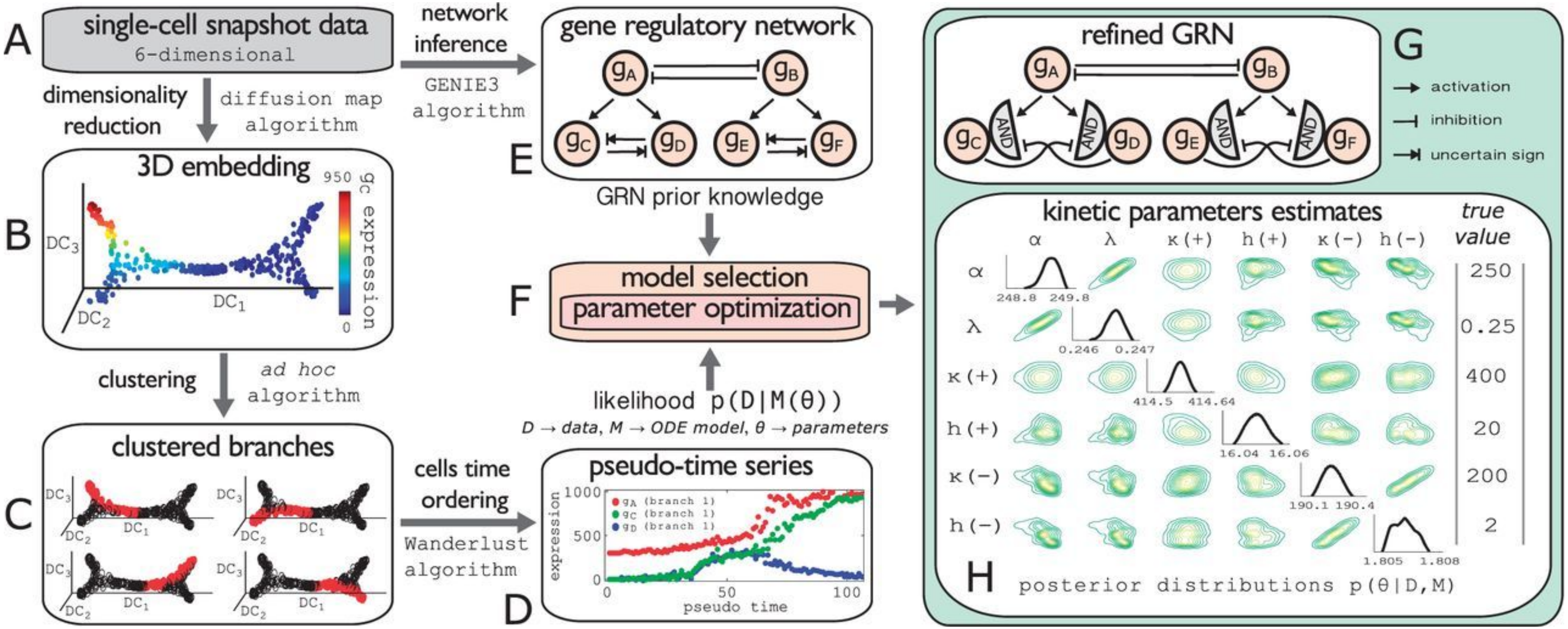




Reconstructing cell cycle and disease progression using deep learning. https://doi.org/10.1038/s41467-017-00623-3

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

Dynamic network inference



Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data.
https://doi.org/10.1093/bioinformatics/btv257

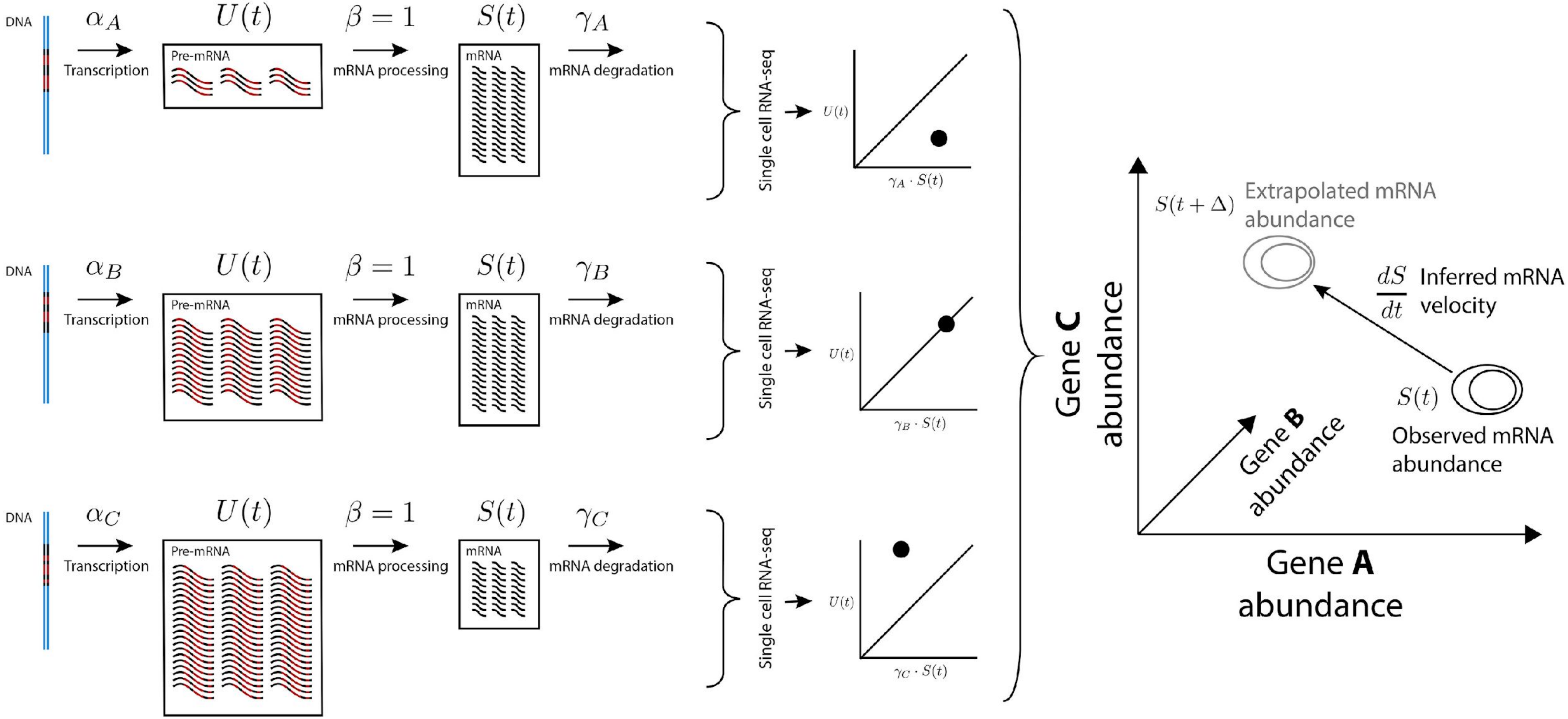
Dynamic network inference

Current limitations

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

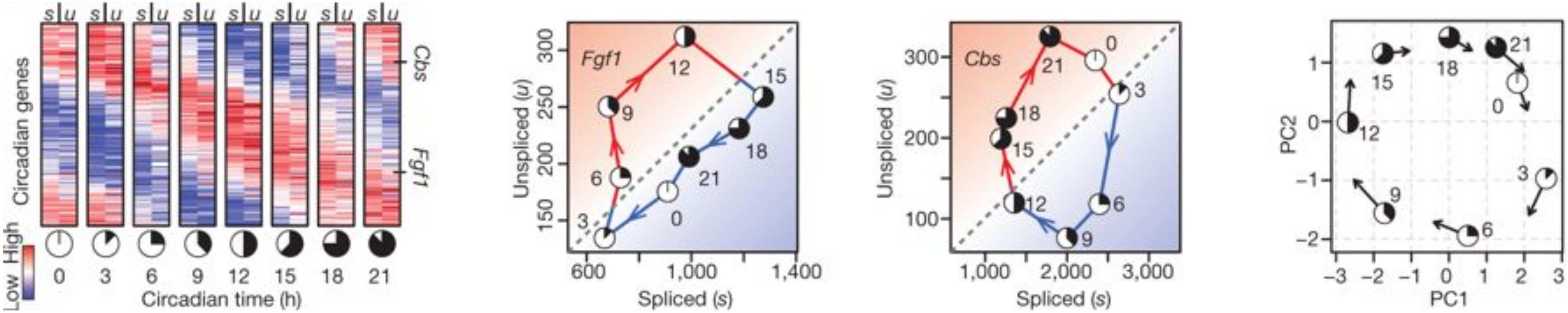
RNA velocity of single cells. https://doi.org/10.1038/s41586-018-0414-6

RNA velocity



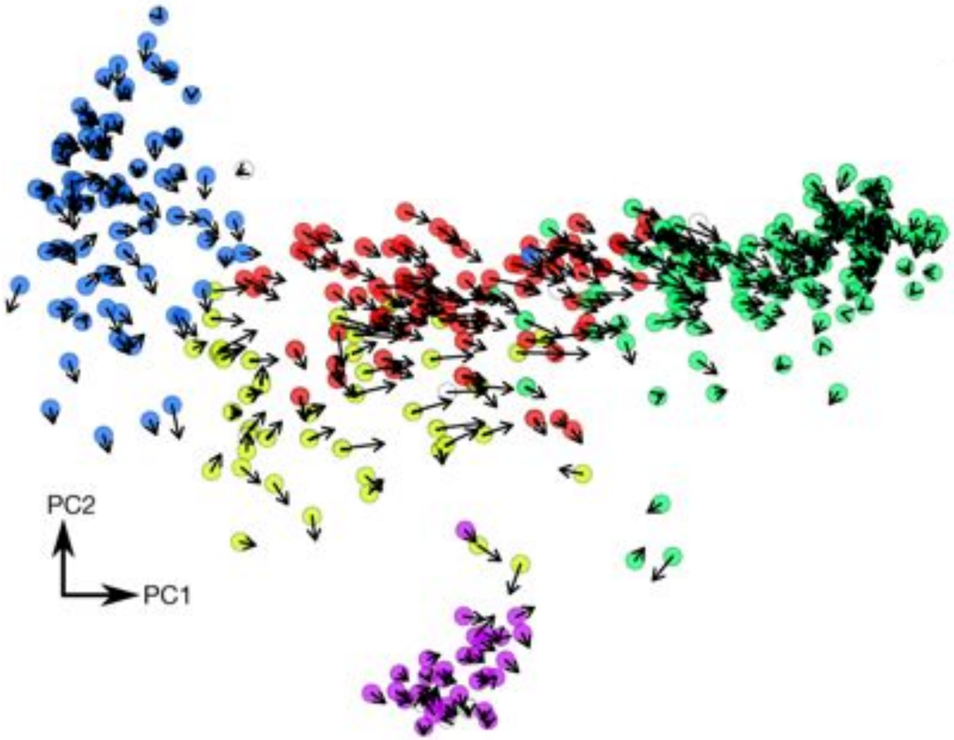
RNA velocity of single cells. https://doi.org/10.1038/s41586-018-0414-6

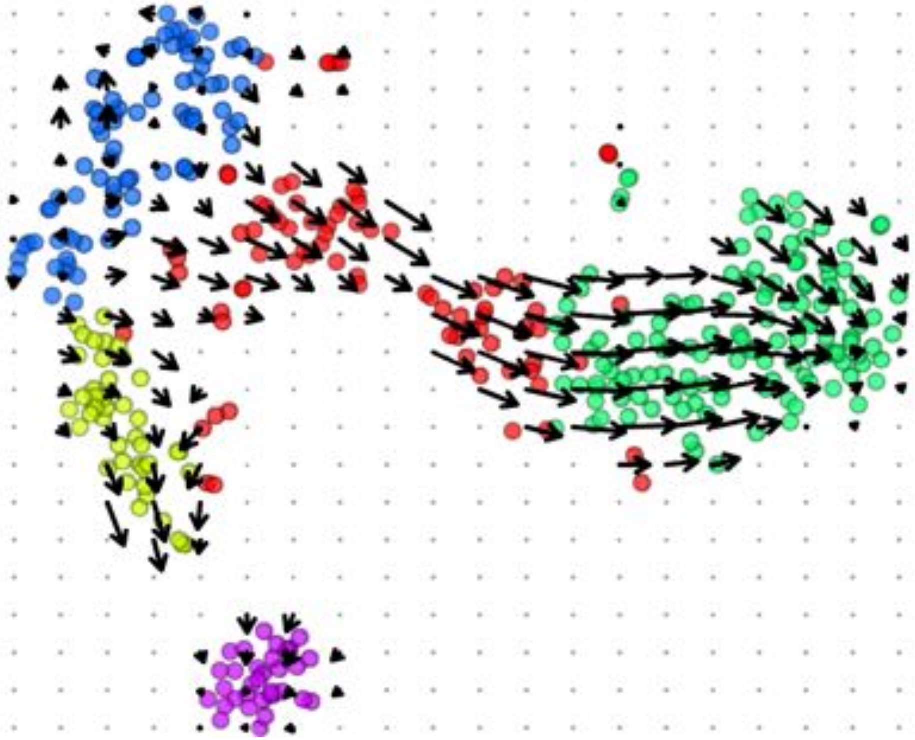
RNA velocity



RNA velocity of single cells. https://doi.org/10.1038/s41586-018-0414-6

RNA velocity





2014 2016 2018 2019 2020

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