Trajectory analysis (BTEP)

Abdalla Abdelmaksoud

Single-Cell Genomics and Trajectory Inference

- During development, in response to stimuli, and throughout life, cells transition from one functional "state" to another.
- Cells in different states express different sets of genes, producing a dynamic repertoire of proteins and metabolites that carry out their work.
- As cells move between states, they undergo a process of transcriptional re-configuration, with some genes being silenced and others newly activated.
- Using Single-cell RNA-Seq trajectory analysis attempts to find these transient states
- This is done mostly by learning the sequence of gene expression changes each cell goes through

Trajectory Analysis Limitations

- You only capture a snapshot
- Differentiation is assumed to be a continuous process
- You cannot track the history

Too many tools?

- Not a one size fits all
- Over 70 methods available
- Luckily a group benchmark 45 of these methods on 110 real and 229 synthetic datasets for cellular ordering, topology, scalability and usability.
- http://guidelines.dynverse.org/

Things to consider

What you know about your data

How to group your cells

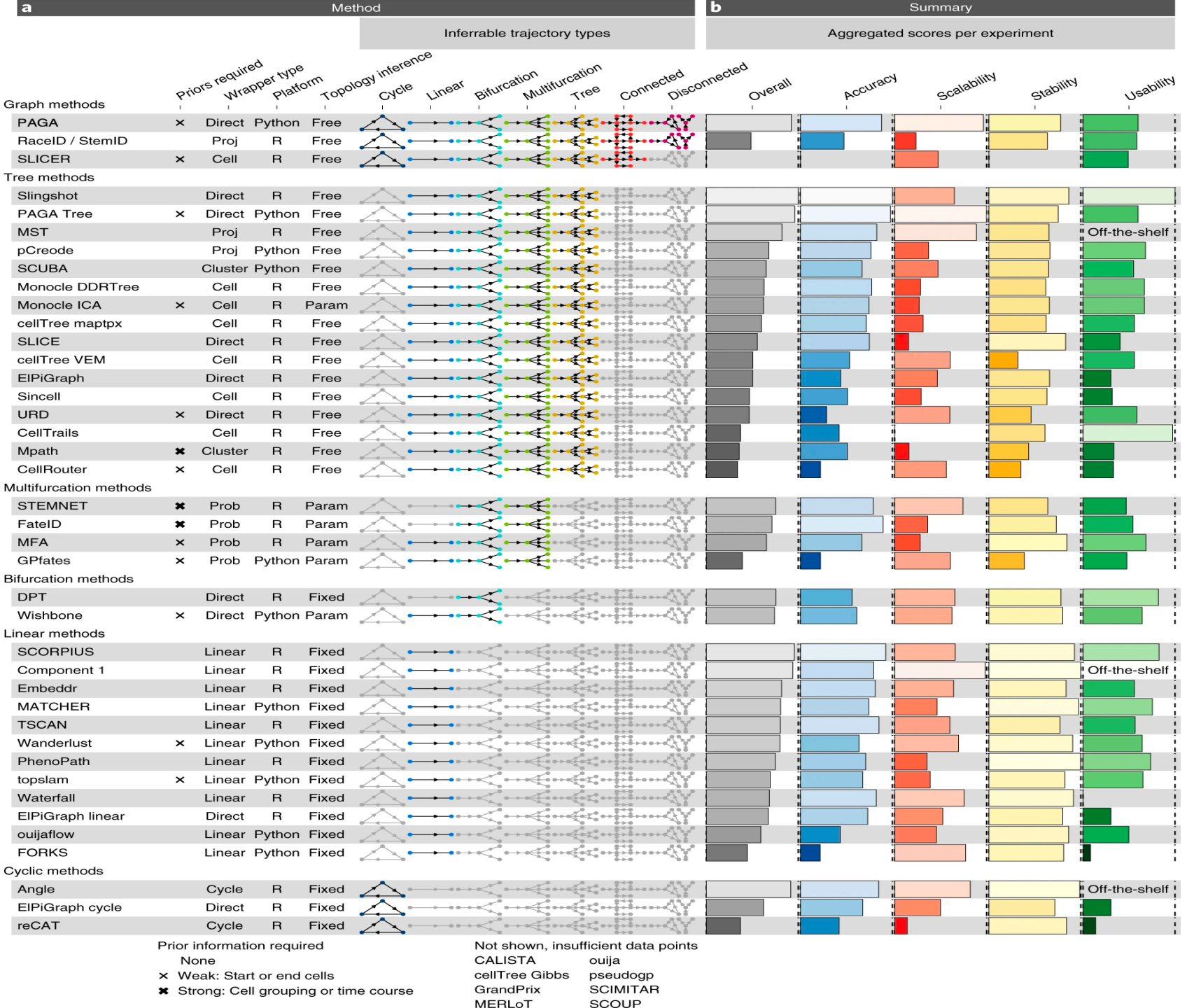
- Machine learning approaches (eg seurat clusters)
- Manual clustering
- Cell type annotations

Comparisons

Compared based on:

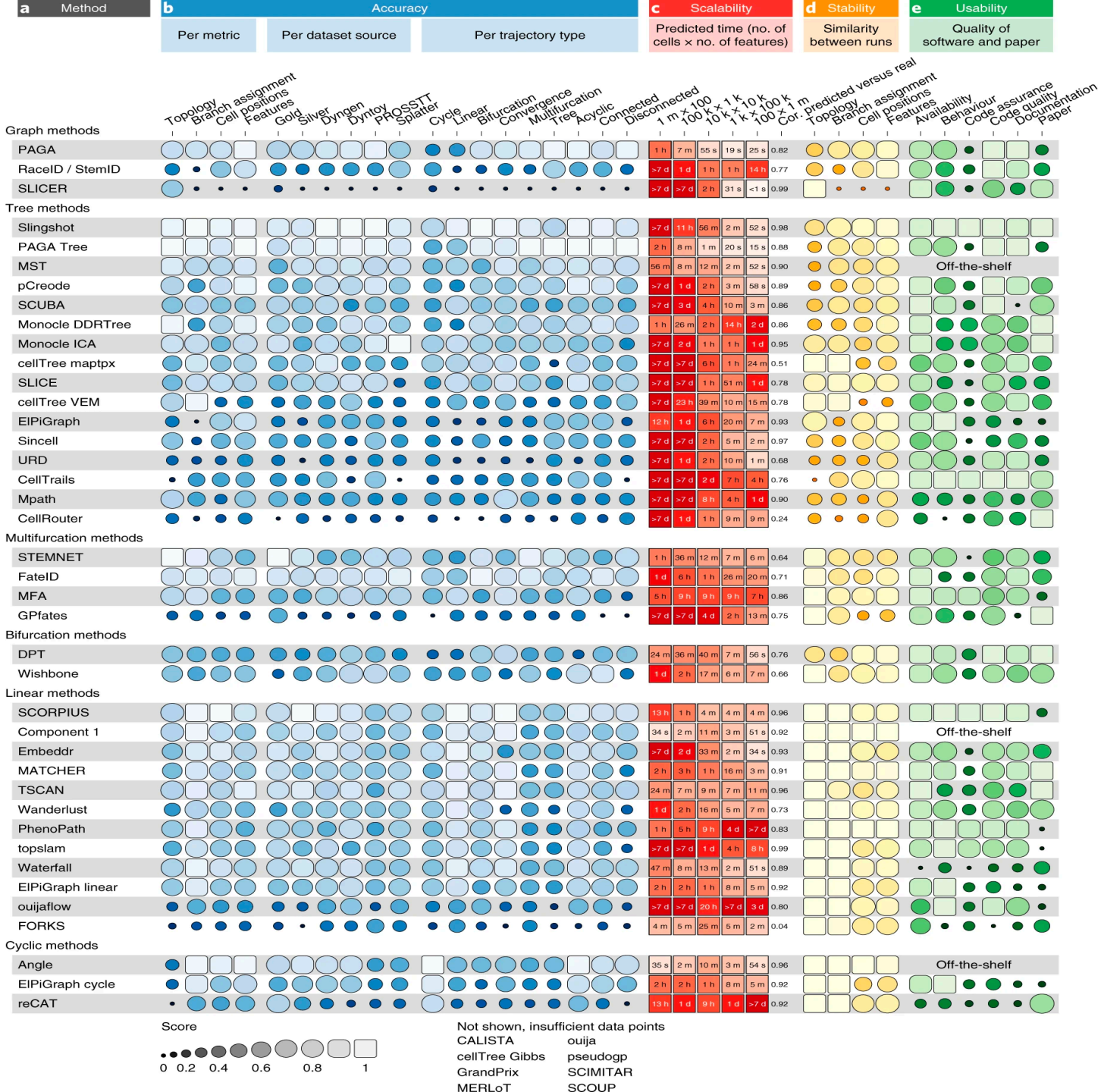
- Accuracy
- Scalability
- Stability
- Usability

Also looked into inferable trajectory types



Comparisons

A more detailed comparison showing metrics for different scenarios



Different topologies defined

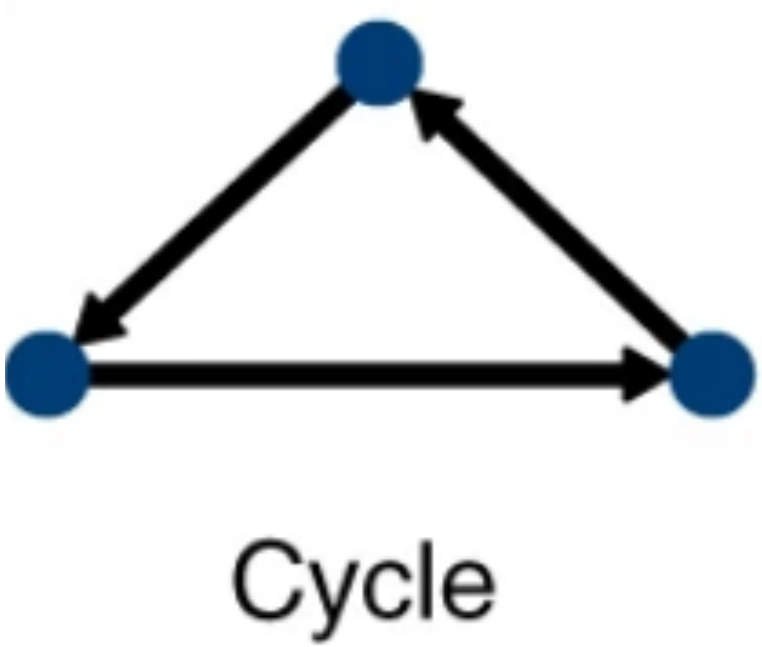
basic topologies: linear, cyclical and bifurcating

Intermediate: multifurcation, convergence, acyclic and tree

complex topologies: connected and disconnected graphs

Trajectory types

A cycle is a connected graph in which every node has a degree equal to 2



Cycle

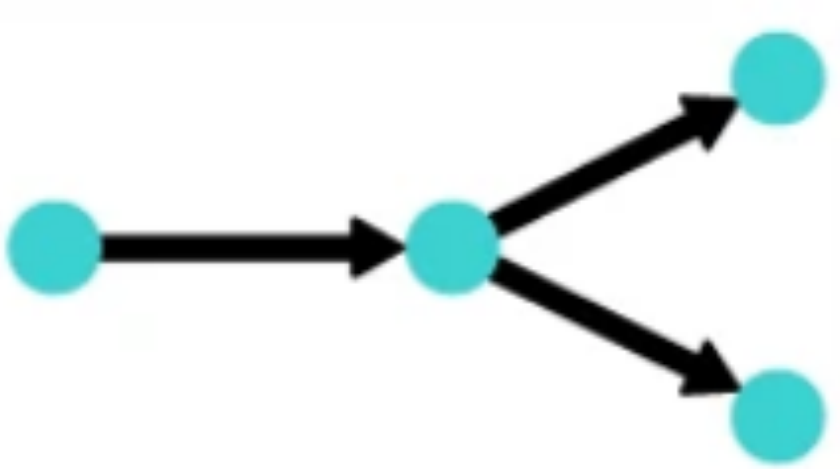
Trajectory types

A linear topology is a graph in which no node has a degree larger than 3

Linear

Trajectory types

A bifurcation is a multifurcation in which only one node has a degree equal to 3



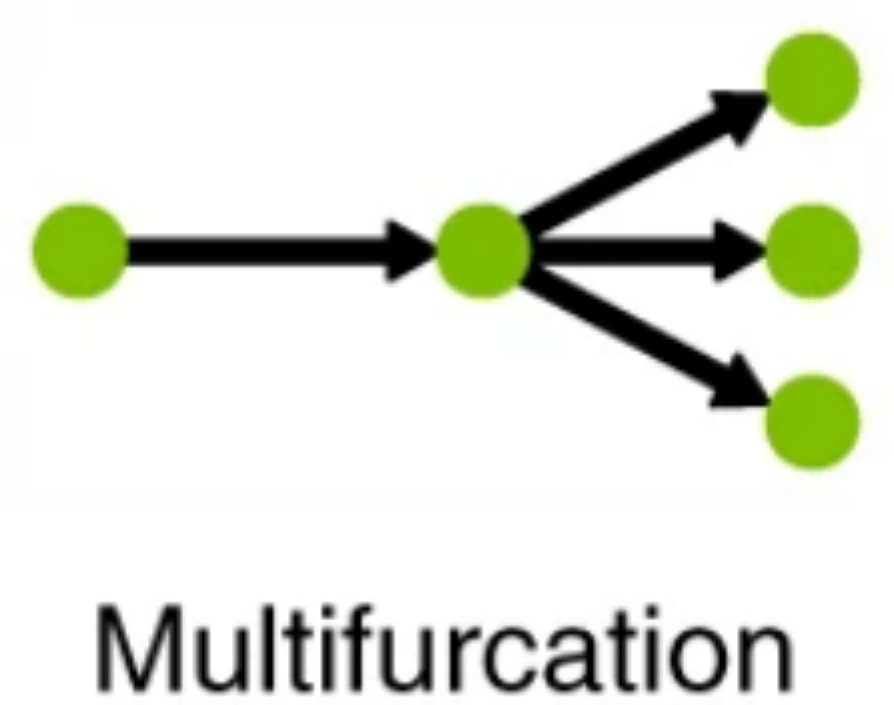
Bifurcation

Trajectory types

A convergence only one node has a degree larger than 1 and this same node has an in-degree of 1
Also like to think of it as the opposite of bifurcation

Trajectory types

A multifurcation is a tree in which only one node has a degree larger than 1

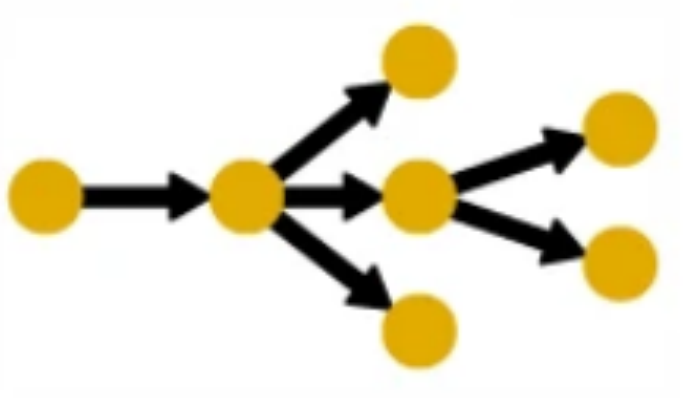


Trajectory types

An acyclic graph is a graph containing no cycles

Trajectory types

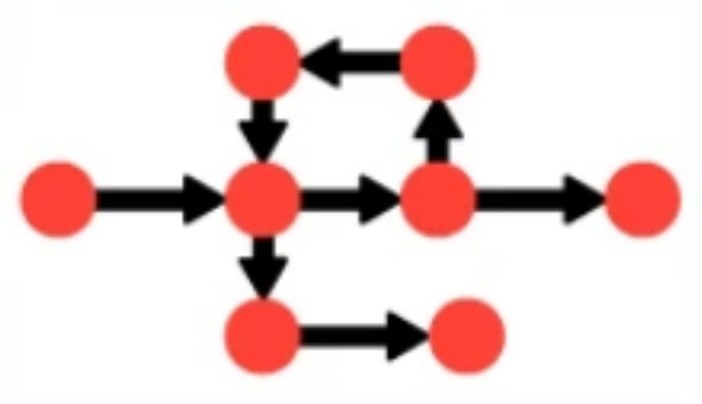
A tree is an acyclic graph containing no convergences



Tree

Trajectory types

Connected (graph): only one edge can exist between two nodes



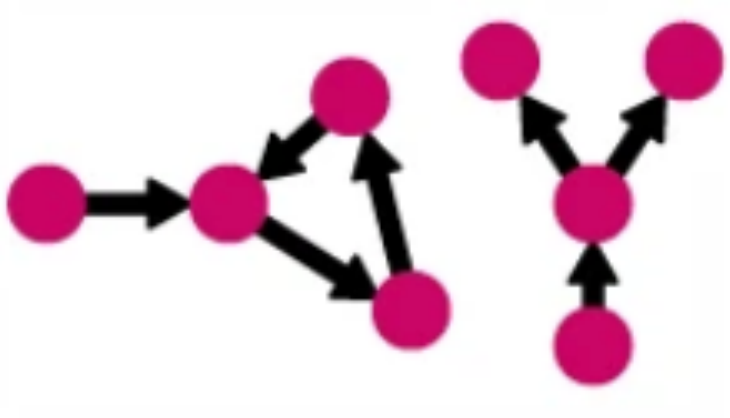
Connected graph

Trajectory types

A disconnected, like the connected is a graph in which only one edge can exist between two nodes

It is also disconnected

Based on their testing only PAGA can reliably call this type



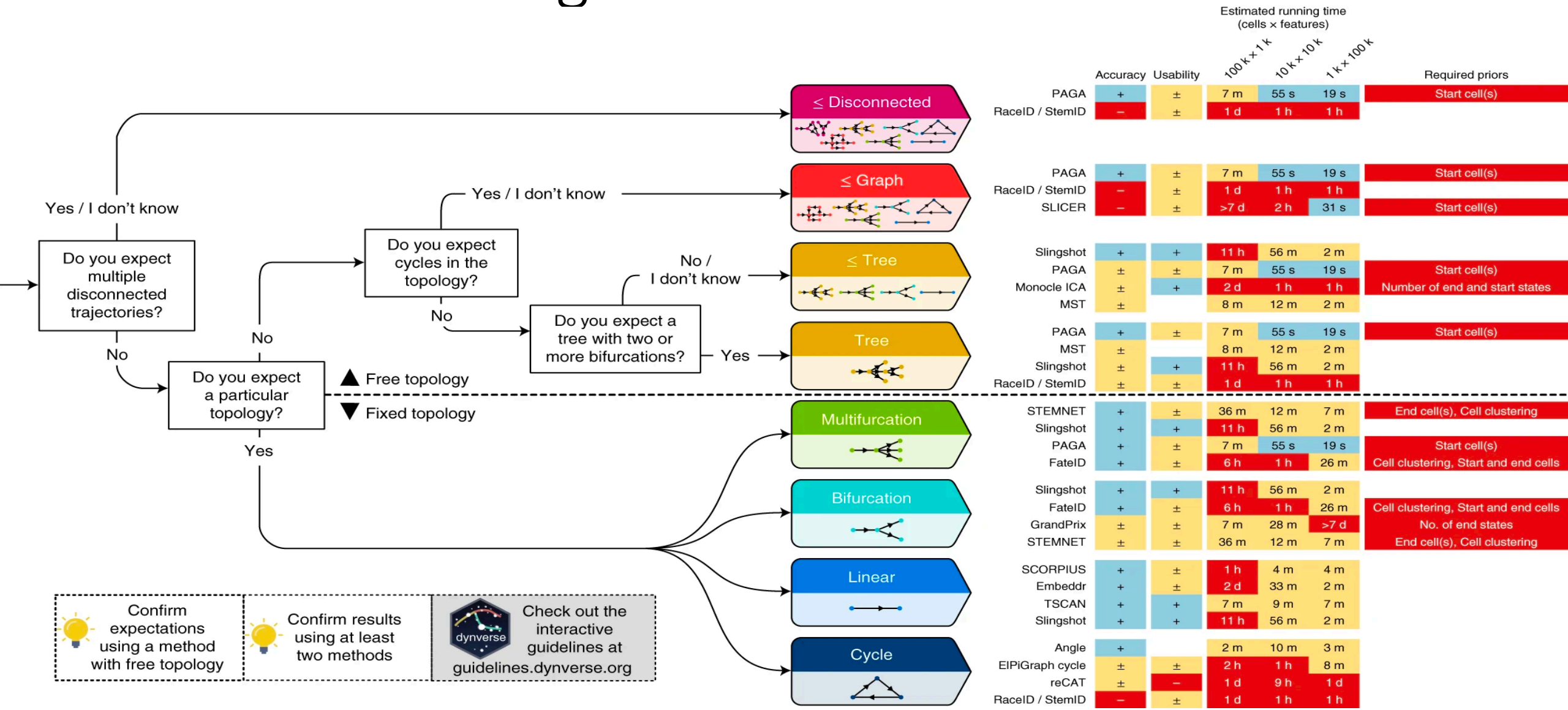
Disconnected graph

Decision making

Obviously this is complicated and almost no way of knowing exactly what trajectory types before actually running that's why they simplified to:

- Do you expect multiple disconnected trajectories
- Do you expect a particular topology
- Do you expect cycles in the topology
- Do you expect a tree with two or more bifurcation

Decision making

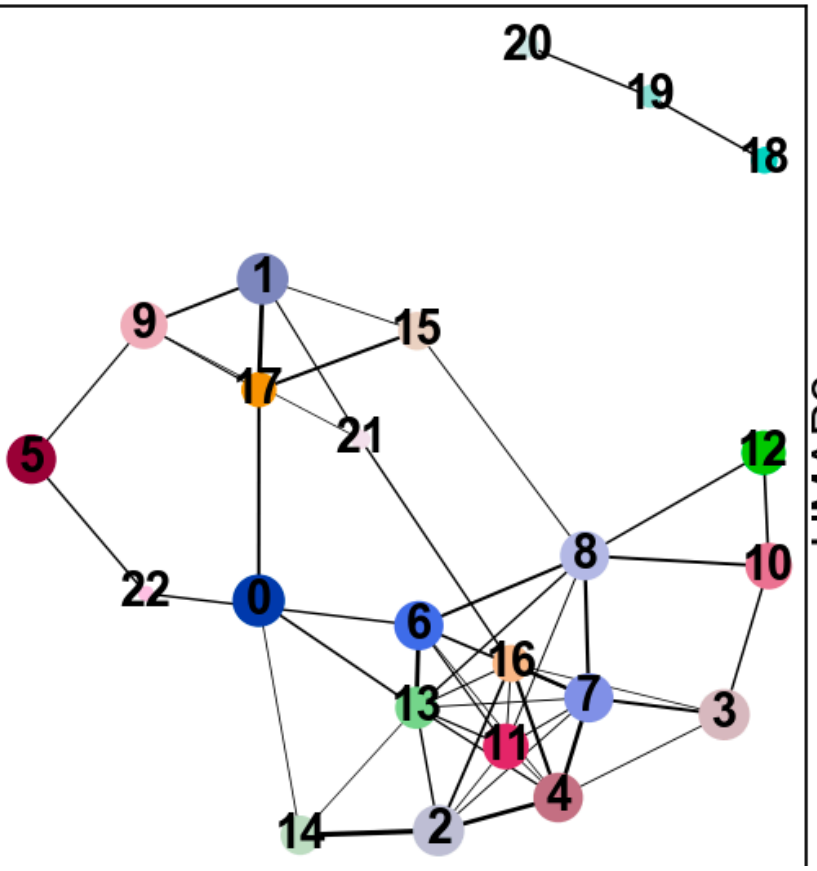


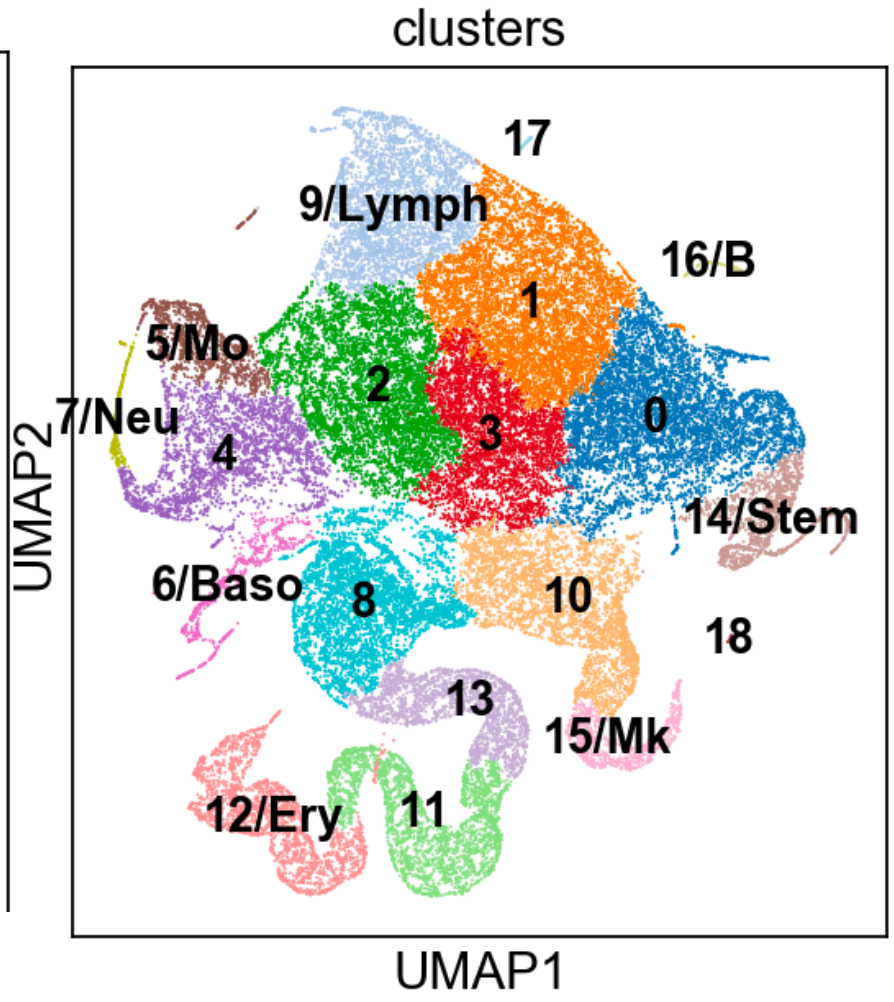
- http://guidelines.dynverse.org/

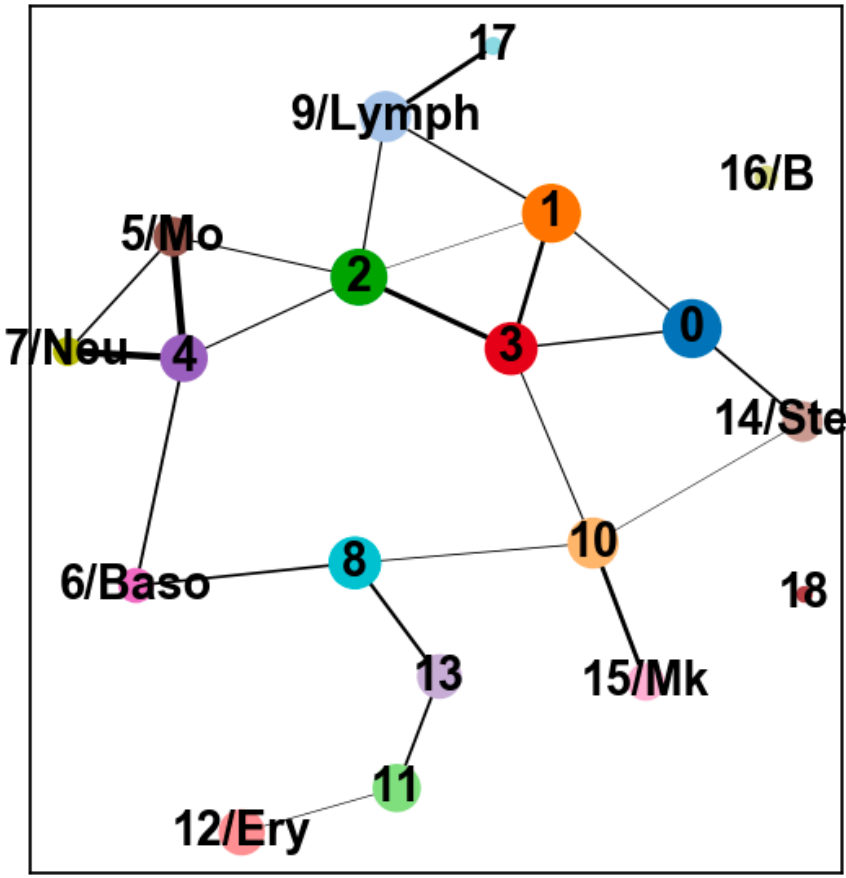
Partition-based Graph Abstraction (PAGA)

- Gives graph-like map of data manifold, based on estimating connectivity of manifold partitions (e.g. cell clusters)
- Preserves global topology of data, allowing analysis at different resolutions
- Unifies both the clustering and continuous change approaches

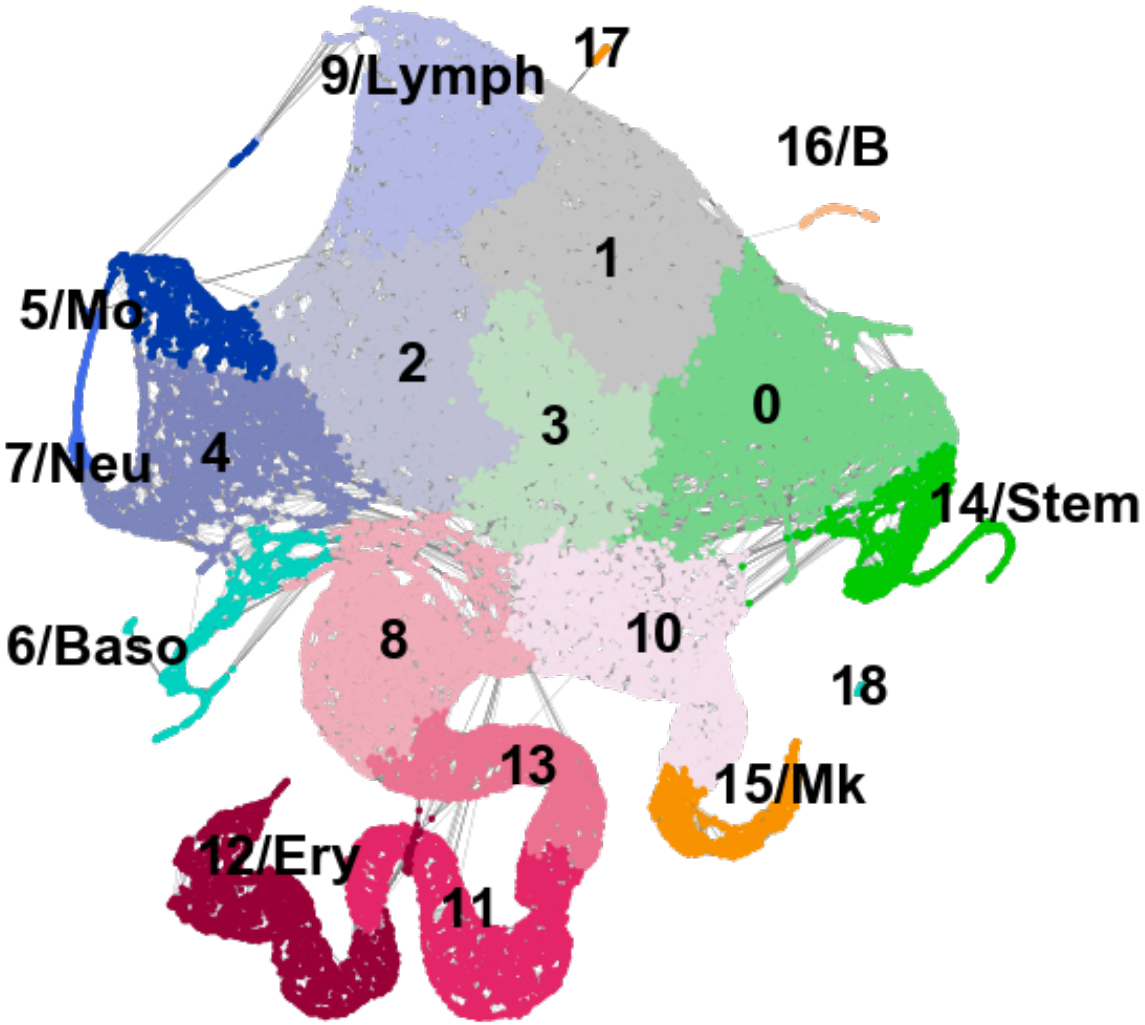
PAGA

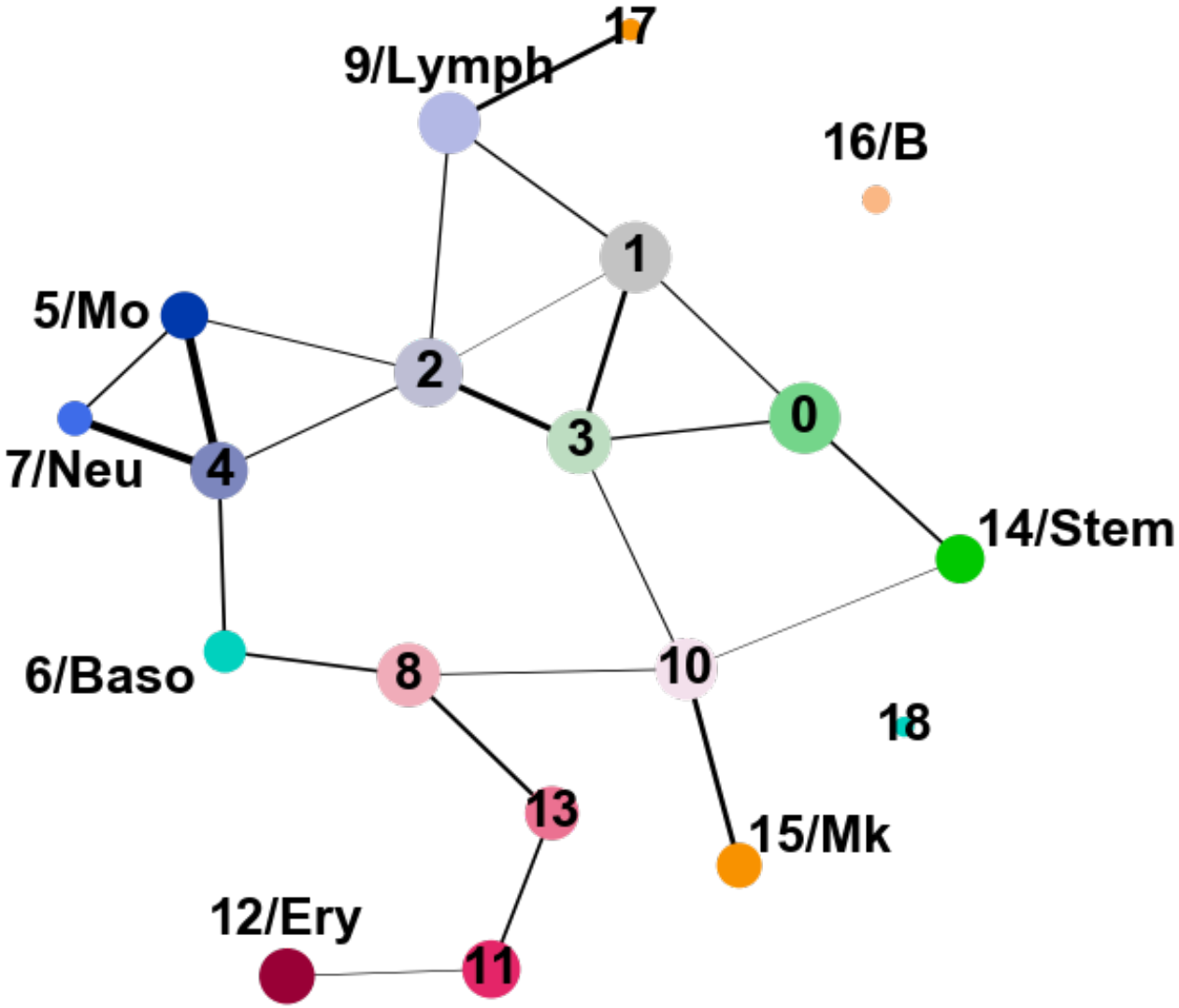

20
19
18
1
9
15
17
21
5
12
22
0
8
10
6
16
13
7
3
11
4
14
2


clusters
17
9/Lymph
16/B
1
5/Mo
2
7/Neu
4
3
0
14/Stem
6/Baso
8
10
18
13
15/Mk
12/Ery
11
UMAP2
UMAP1

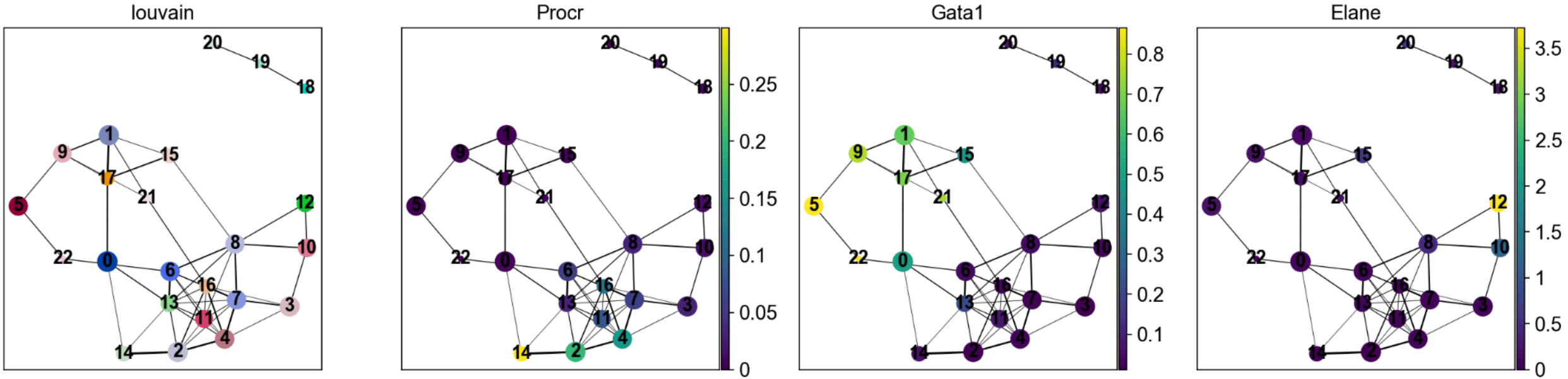

17
9/Lymph
16/B
1
5/Mo
2
7/Neu
4
3
0
14/Ste
10
6/Baso
8
18
13
15/Mk
11
12/Ery

PAGA

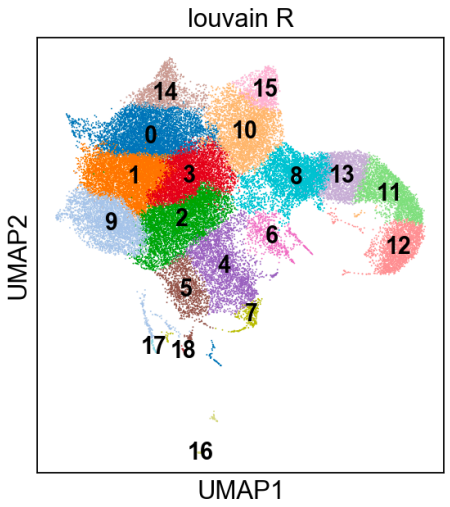

9/Lymph
17
16/B
1
5/Mo
2
3
0
7/Neu
4
14/Stem
6/Baso
8
10
18
13
15/Mk
12/Ery
11

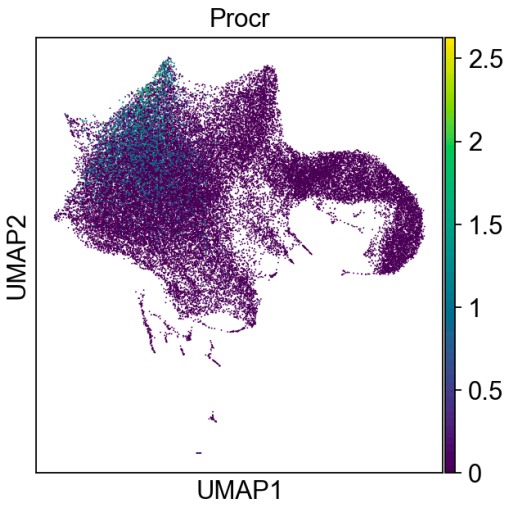

9/Lymph
17
16/B
1
5/Mo
2
3
0
7/Neu
4
14/Stem
6/Baso
8
10
18
13
15/Mk
12/Ery
11

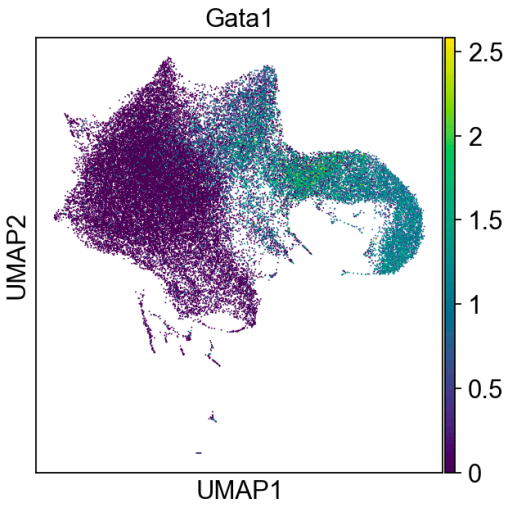
PAGA

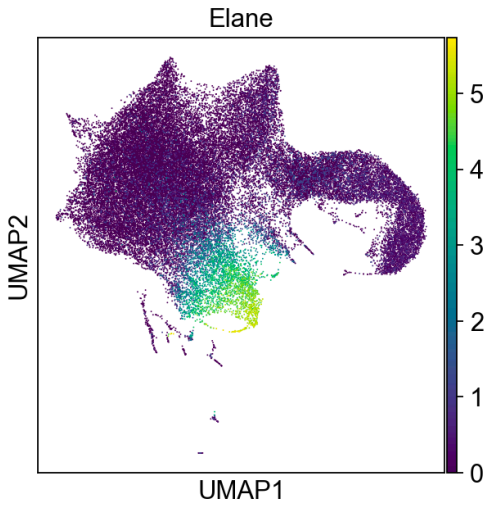

louvain
Procr
Gata1
Elane
0.25
0.2
0.15
0.1
0.05
0
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
3.5
3
2.5
2
1.5
1
0.5
0

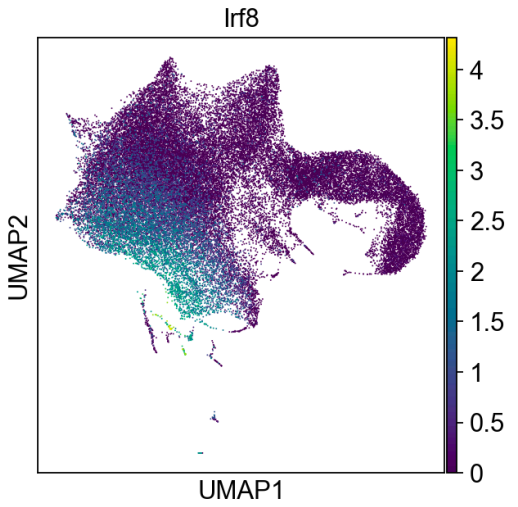
PAGA


louvain R
14
15
0
10
1
3
8
13
11
9
2
6
12
4
5
7
17
18
16
UMAP2
UMAP1


Procr
2.5
2
1.5
1
0.5
0
UMAP2
UMAP1


Gata1
2.5
2
1.5
1
0.5
0
UMAP2
UMAP1


Elane
5
4
3
2
1
0
UMAP2
UMAP1


Irf8
4
3.5
3
2.5
2
1.5
1
0.5
0
UMAP2
UMAP1

PAGA

How to run

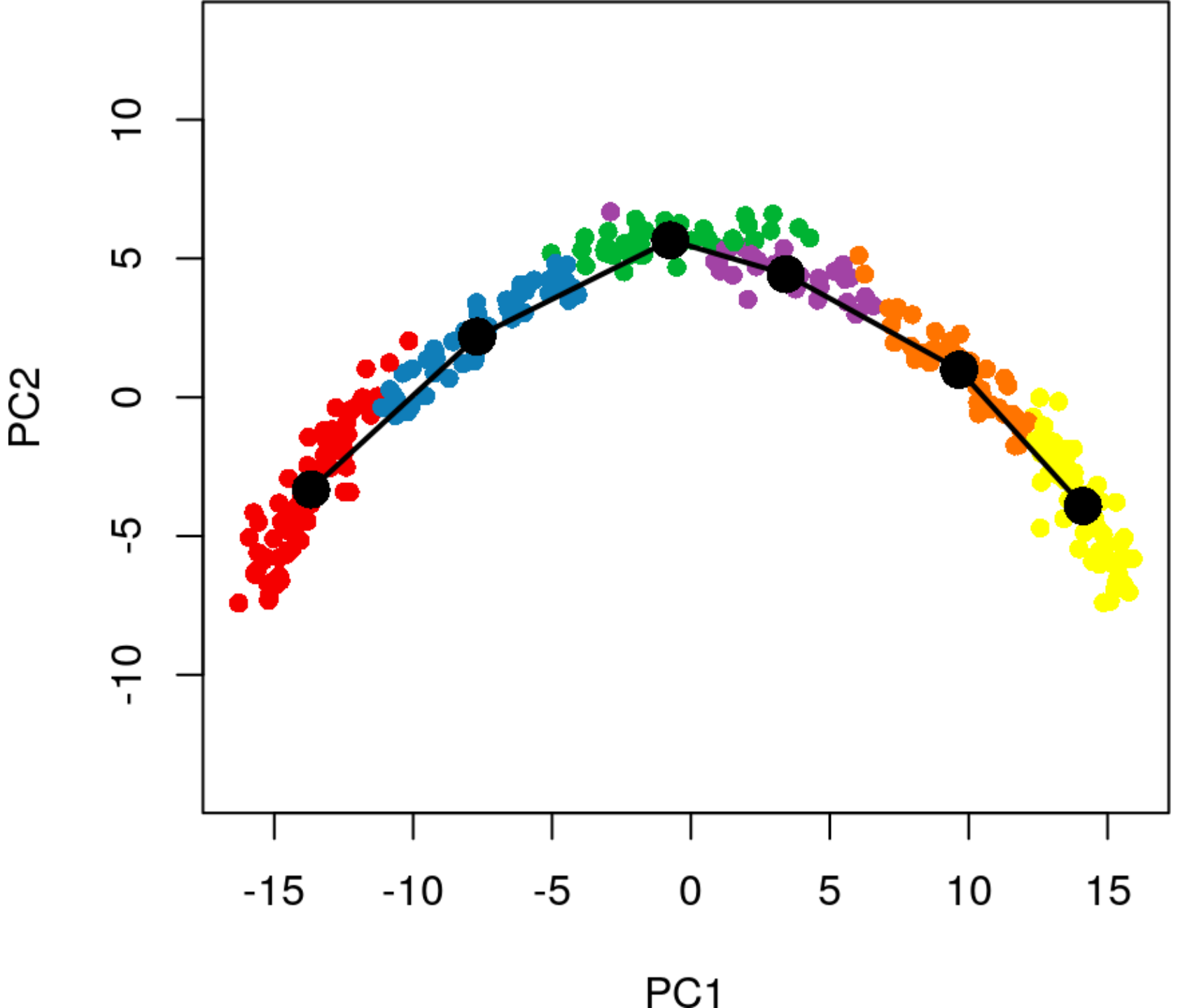
- PAGA is python based
- Save Seurat object as loom and import into anndata (doesn’t always work smoothly)
- Can run from scratch using the scanpy manual
- Can add matrix, UMAP coordinates and meta data separately

Slingshot

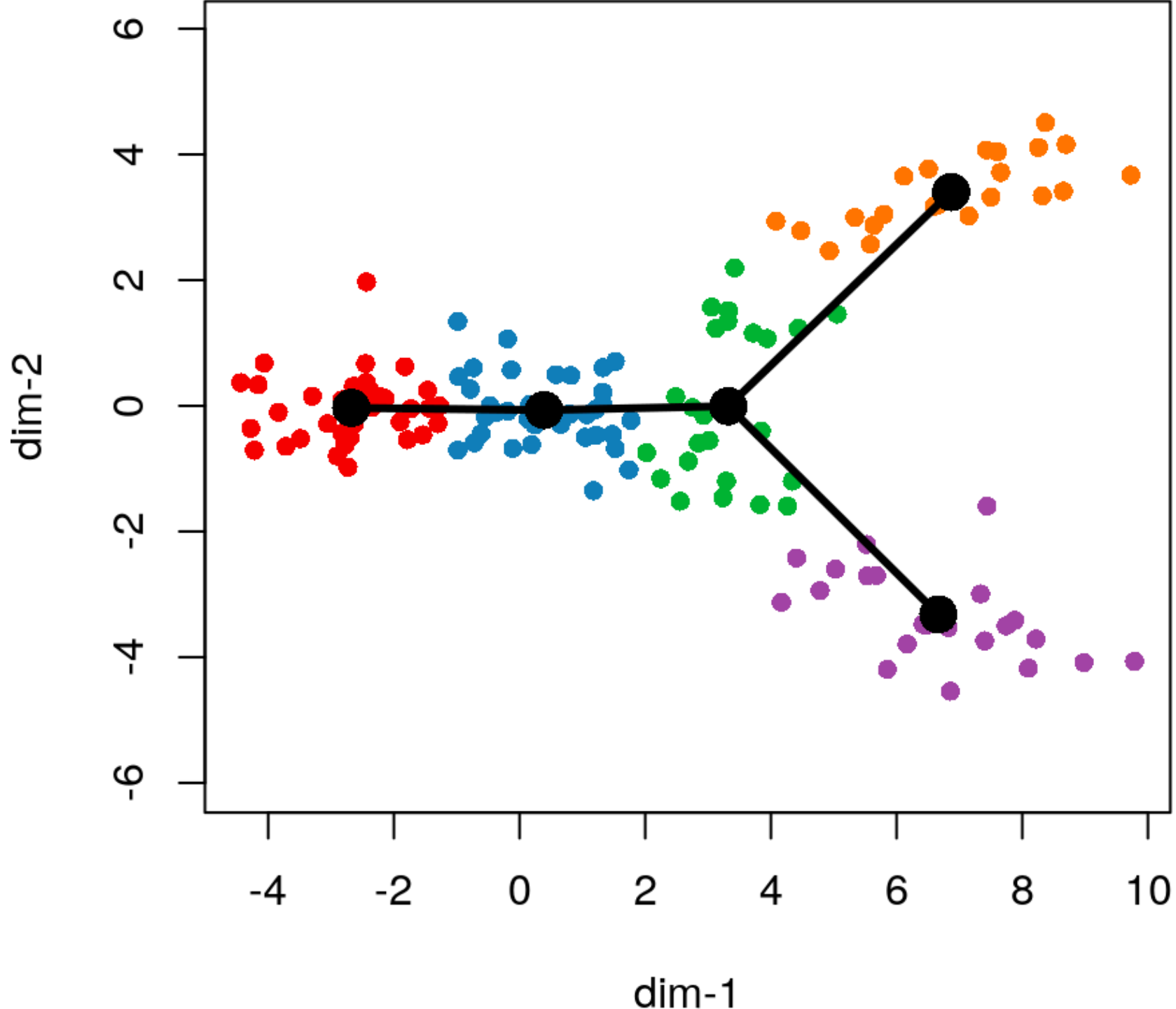
Two-step process:

1. identifying the global lineage structure with a cluster-based minimum spanning tree (MST)
2. Fitting simultaneous principal curves representing each cell's transcriptional progression toward the terminal state to describe each lineage.

Slingshot

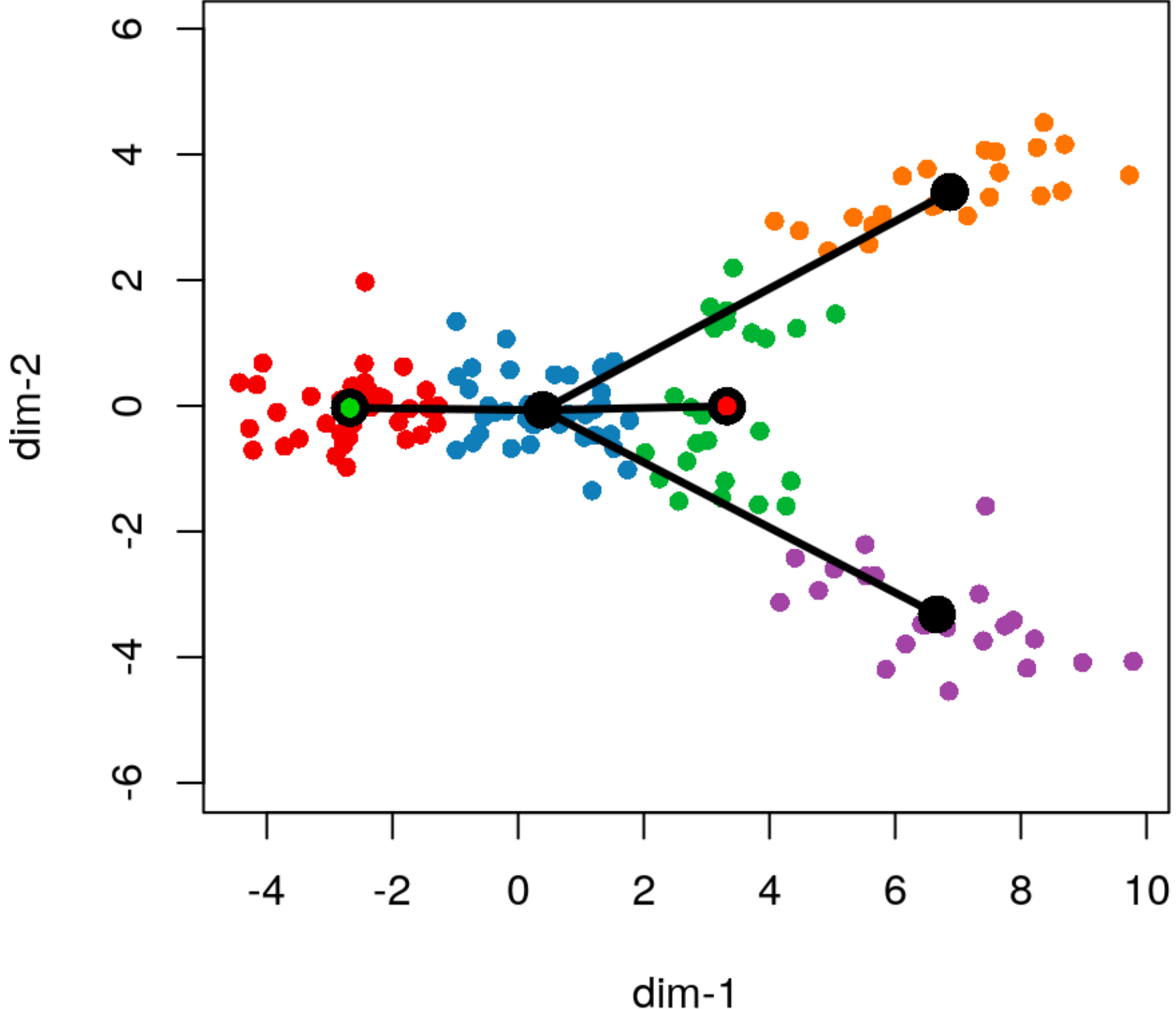


Slingshot



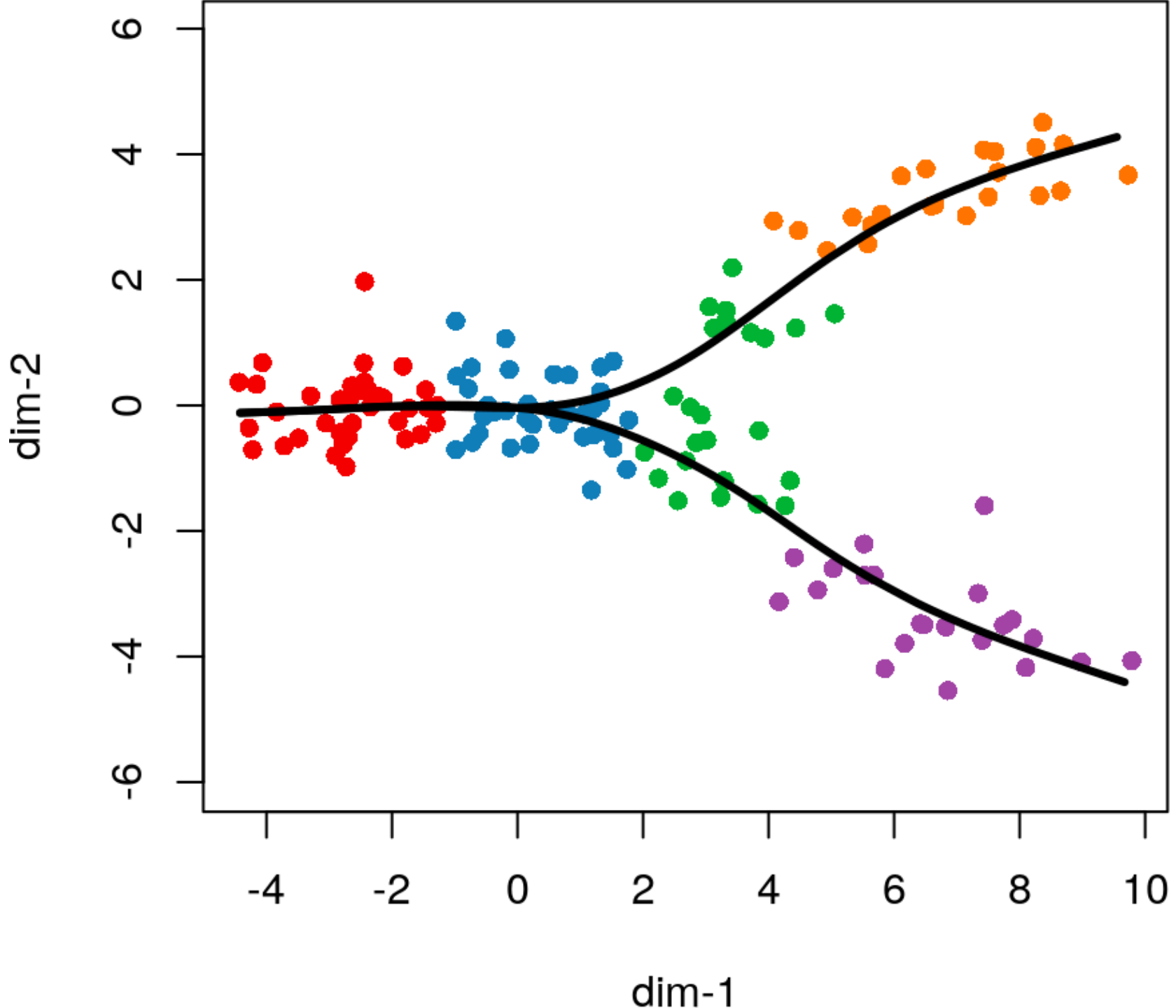
Slingshot

Slingshot also allows you to select start and end clusters



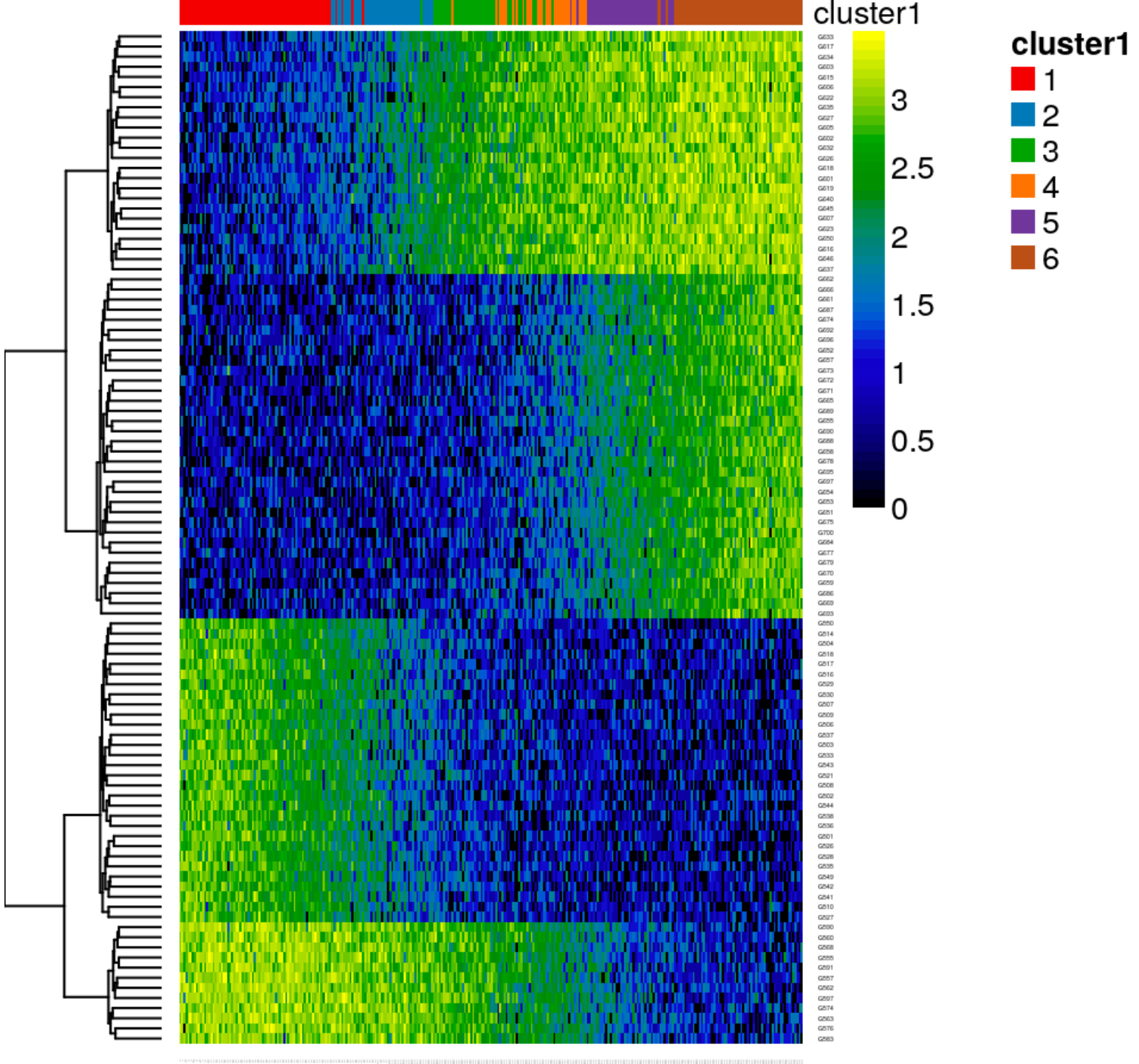
Slingshot

More robust to noise



Slingshot

Using top genes based on p-value and visualize their expression over developmental time with a heatmap.



Slingshot

How to run

From Scratch

- Using the instructions in their manual on how to process using SingleCellExperiment object

From Seurat:

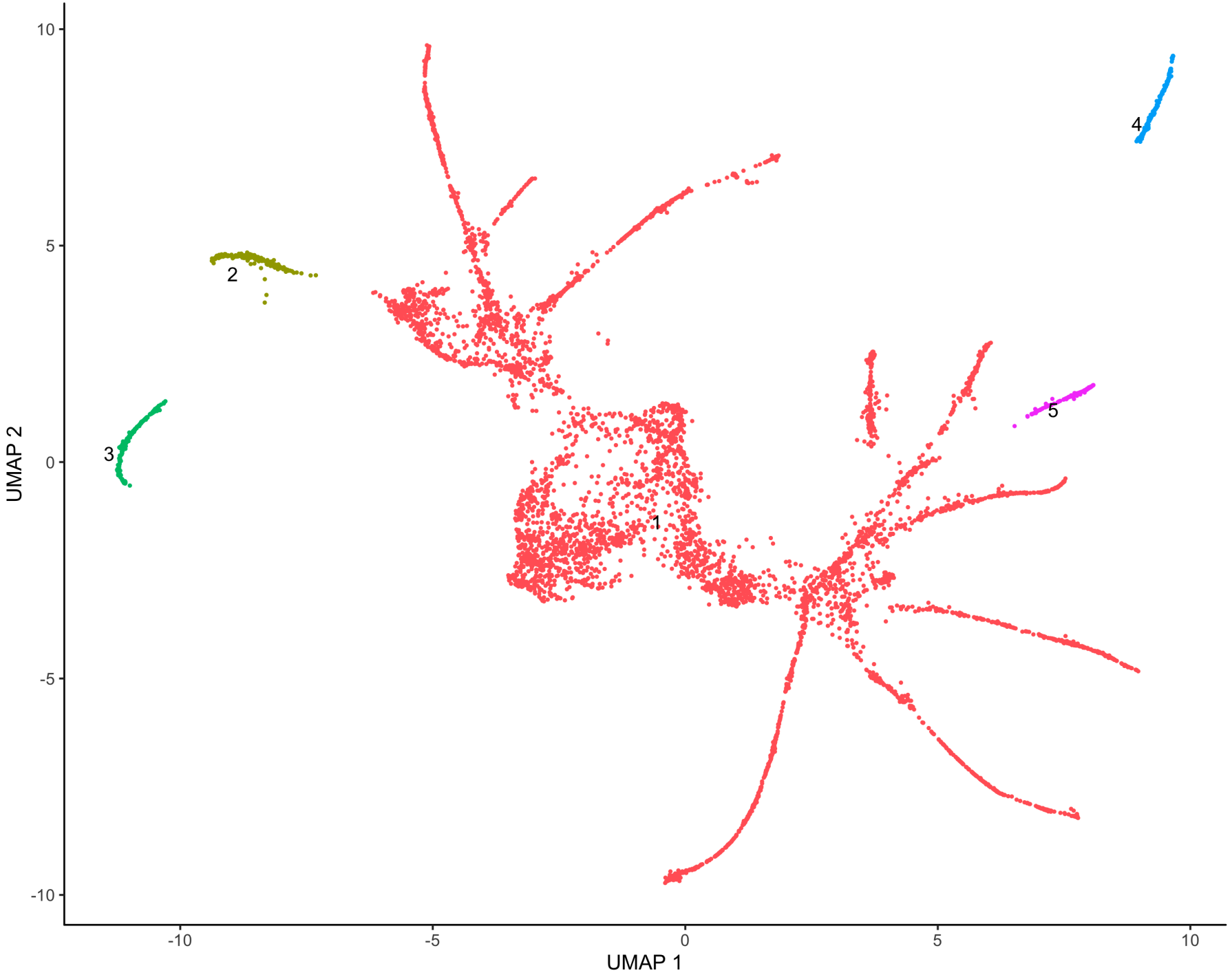
- Using the as.SingleCellExperiment() from Seurat
- This should import all of your meta-data and UMAP coordinates

Monocle

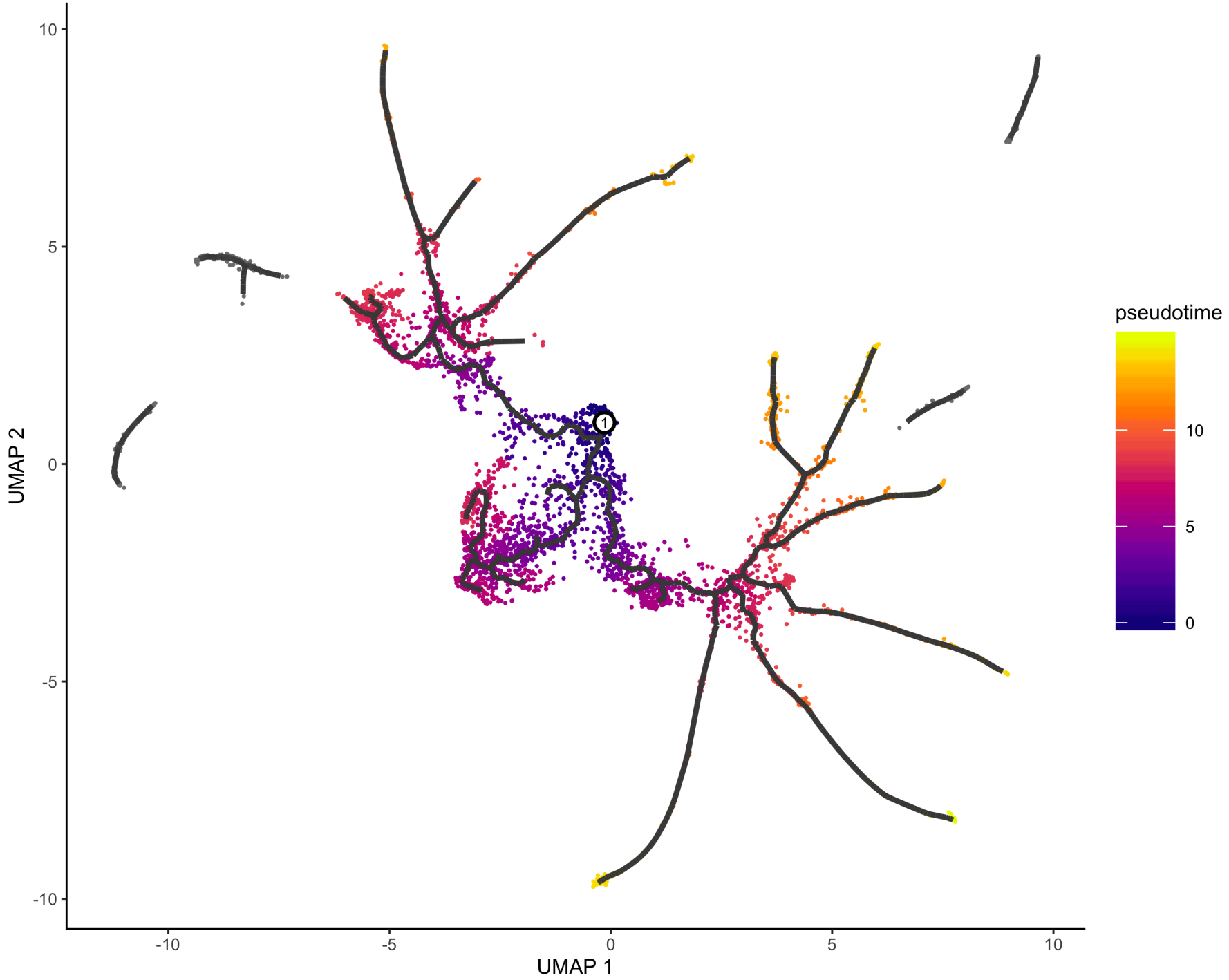
Monocle3 claims

- Support for trajectories with multiple roots.
- Ways to learn trajectories that have loops or points of convergence.
- Algorithms that automatically partition cells to learn disjoint or parallel trajectories.
- A 3D interface to visualize trajectories and gene expression.

Monocle



Monocle



Monocle

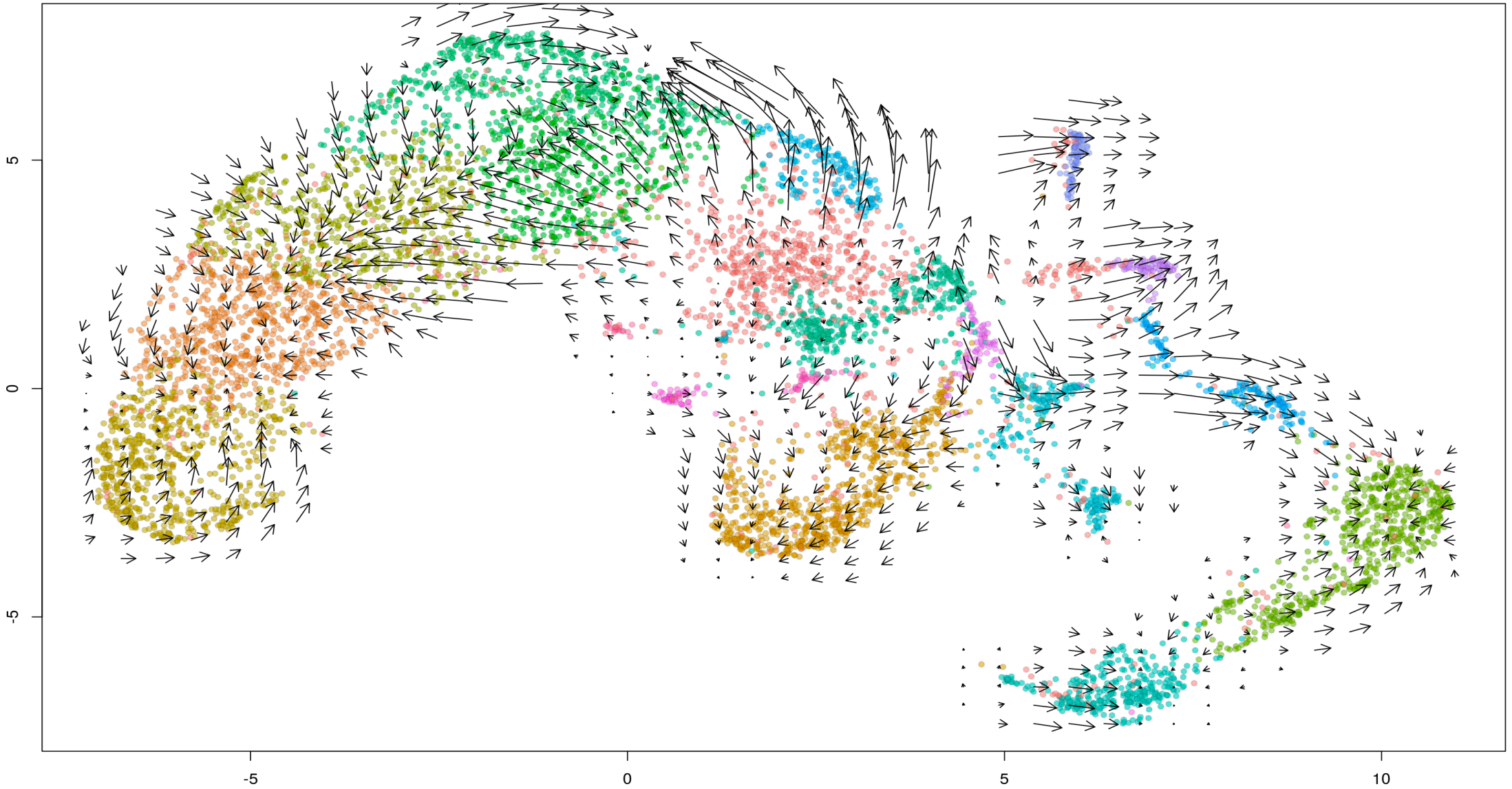
How to run

- Monocle2 use to have a function to convert from Seurat2 but Monocle 3 has been in beta mode for almost a year now and this function does not exist yet
- Can run from scratch using the manual
- Can probably add matrix, UMAP coordinates and meta data separately like paga

Velocyto

- RNA abundance is a powerful indicator of the state of individual cells. Single-cell RNA sequencing can reveal RNA abundance with high quantitative accuracy, sensitivity and throughput[1].
- However, this approach captures only a static snapshot at a point in time, posing a challenge for the analysis of time-resolved phenomena such as embryogenesis or tissue regeneration.
- Built under the notion that unspliced mRNAs consistently preceded spliced mRNAs during both up- and downregulation, claims to predicts the future state of individual cells on a timescale of hours
- *RNASEH2B* exhibited fast kinetics, with little difference between unspliced and spliced RNAs.
- By contrast, genes such as *DCX*, *ELAVL4* and *STMN2* showed evidence of spliced transcripts following a noticeably delayed trajectory.

Velocyto



Package	Vignette	Reference	Source
Conos	Integration of datasets using Conos	Barkas et al, Nature Methods 2019	https://github.com/hms-dbmi/conos
LIGER	Integrating Seurat objects using LIGER	Welch et al, Cell 2019	https://github.com/MacoskoLab/liger
fastMNN	Running fastMNN on Seurat Objects	Nature Biotechnology 2018	https://bioconductor.org/packages/release/bioc/html/batchelor.html
Harmony	Integration of datasets using Harmony	Korsunsky et al, bioRxiv 2018	https://github.com/immunogenomics/harmony
ALRA	Zero-preserving imputation with ALRA	Linderman et al, bioRxiv 2018	https://github.com/KlugerLab/ALRA
Velocity	Estimating RNA Velocity using Seurat	La Manno et al, Nature 2018	https://velocyto.org
schex	Using schex with Seurat	Freytag, R package 2019	https://github.com/SaskiaFreytag/schex

Velocyto

How to run

- velocyto run10x
- Run using Seurat wrappers

Final notes

- Understand limitations of the trajectory analysis
- Choose clusters carefully (make sure they are biologically meaningful)
- PAGA might be a good first choice

References

1. Saelens, Wouter, et al. "A comparison of single-cell trajectory inference methods." *Nature biotechnology* 37.5 (2019): 547.
2. Wolf, F. Alexander, et al. "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells." *Genome biology* 20.1 (2019): 59.
3. Street, Kelly, et al. "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics." *BMC genomics* 19.1 (2018): 477.
4. Trapnell, Cole, Davide Cacchiarelli, and Xiaojie Qiu. "Monocle: Cell counting, differential expression, and trajectory analysis for single-cell RNA-Seq experiments." (2019): 10.
5. La Manno, Gioele, et al. "RNA velocity of single cells." *Nature* 560.7719 (2018): 494.