

# Unlocking the Power of Continuity in Single Cell RNA-Seq: Differential Gene Expression Along Developmental Trajectories

Talk given at the Statistics  
and Genomics Seminar on  
04/18



Hector Roux de Bézieux  
*Group in Biostatistics*  
*Sandrine Dudoit's lab*

**GitHub:** HectorRDB

**Website:** <http://hectorrdb.github.io>



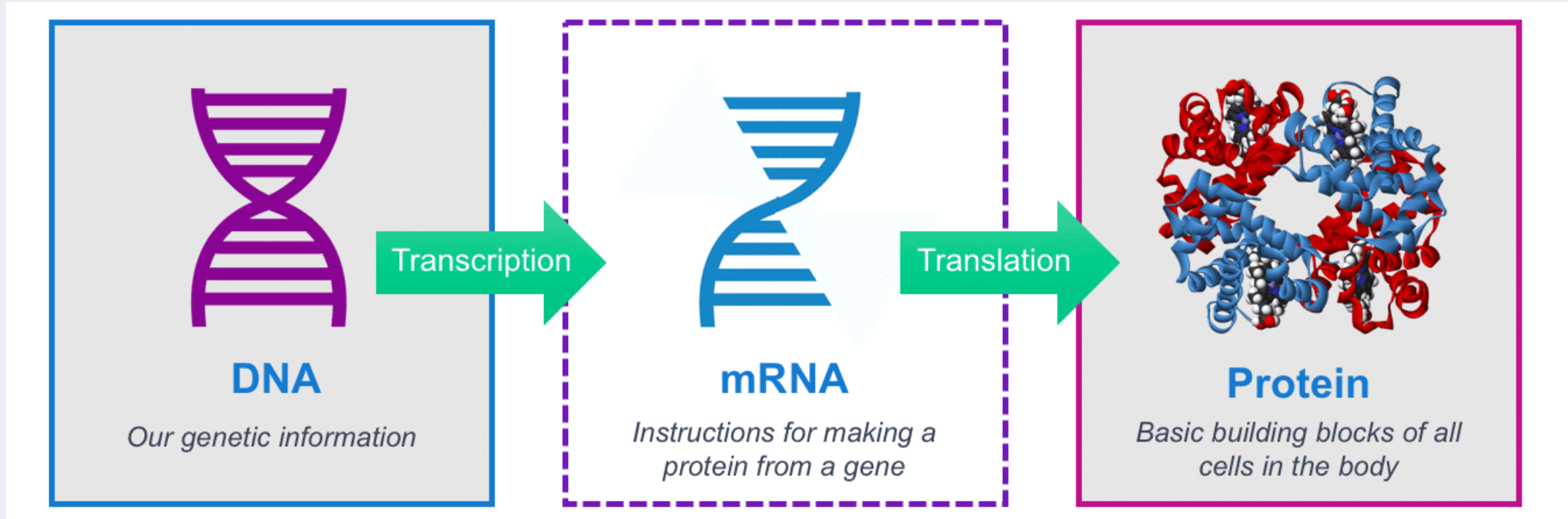
# Overview

- 1) Introduction to scRNA-Seq
- 2) Trajectory Inference with **Slingshot**
- 3) Differential Expression with **tradeSeq**,
- 4) Clustering gene patterns with **RSEC**

# 1. Introduction to scRNA-Seq



# Central Dogma of biology



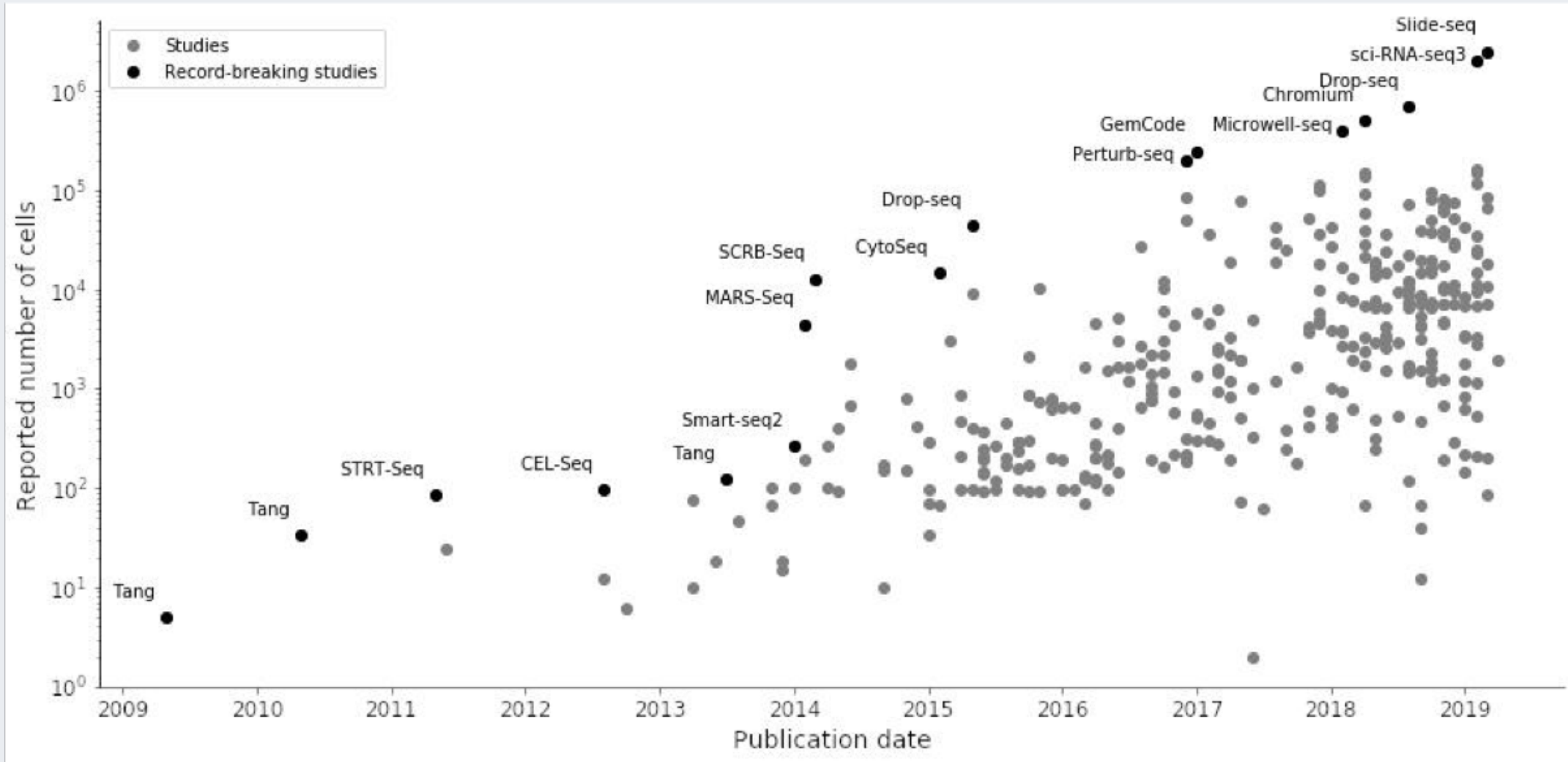
# Single-cell RNA-Seq



Unmixing the smoothie



# Recent explosion in scRNA-Seq



# Data structure



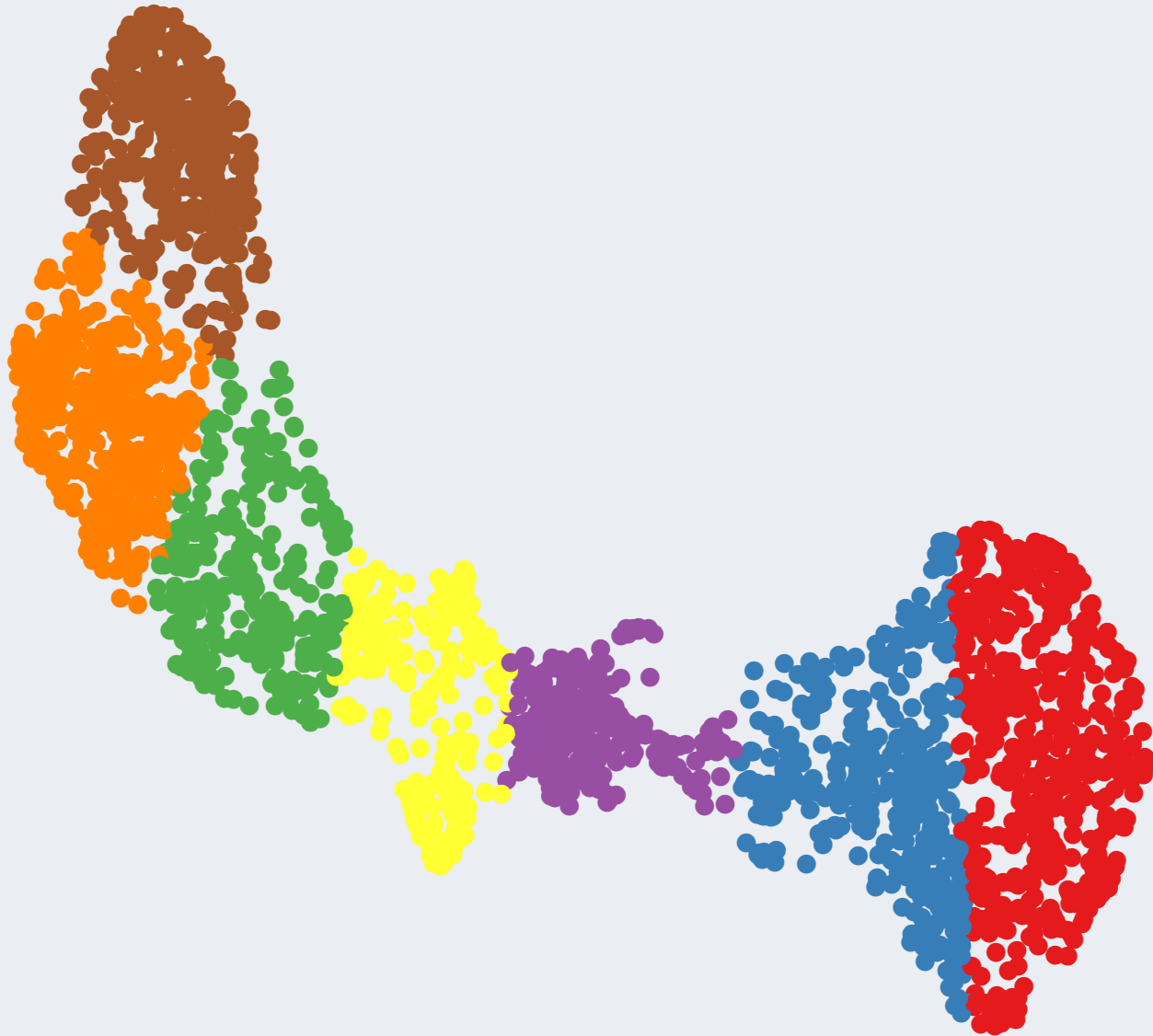
	<i>Cell 1</i>	<i>Cell 2</i>	<i>Cell 3</i>	<i>...</i>	<i>Cell n</i>
<i>Gene 1</i>	0	28	25	...	2
<i>Gene 2</i>	0	3	8	...	36
<i>Gene 3</i>	5	0	0	...	0
<i>...</i>	...	...	...	...	...
<i>Gene G</i>	12	8	0	...	11

## 2. Trajectory Inference with Slingshot



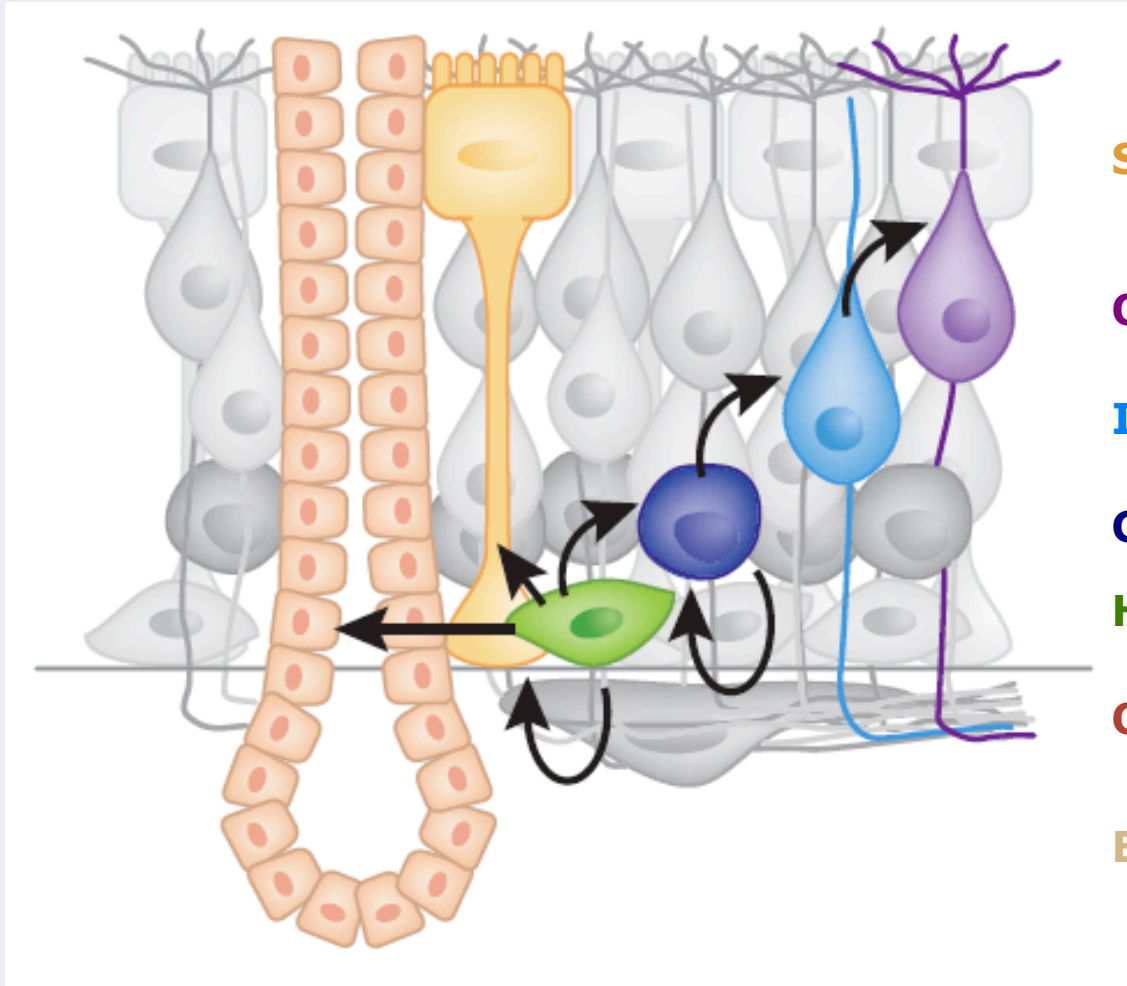


# Dimensionality reduction



- Bone-marrow stem cells from the monocle 3 vignette
- 2660 cells and 3004 genes
- Now 2660 cells in two dimensions using UMAP

# Olfactory Epithelium



**Sustentacular cell (Sus)**

**Olfactory receptor neuron (ORN)**

**Immature olfactory neuron**

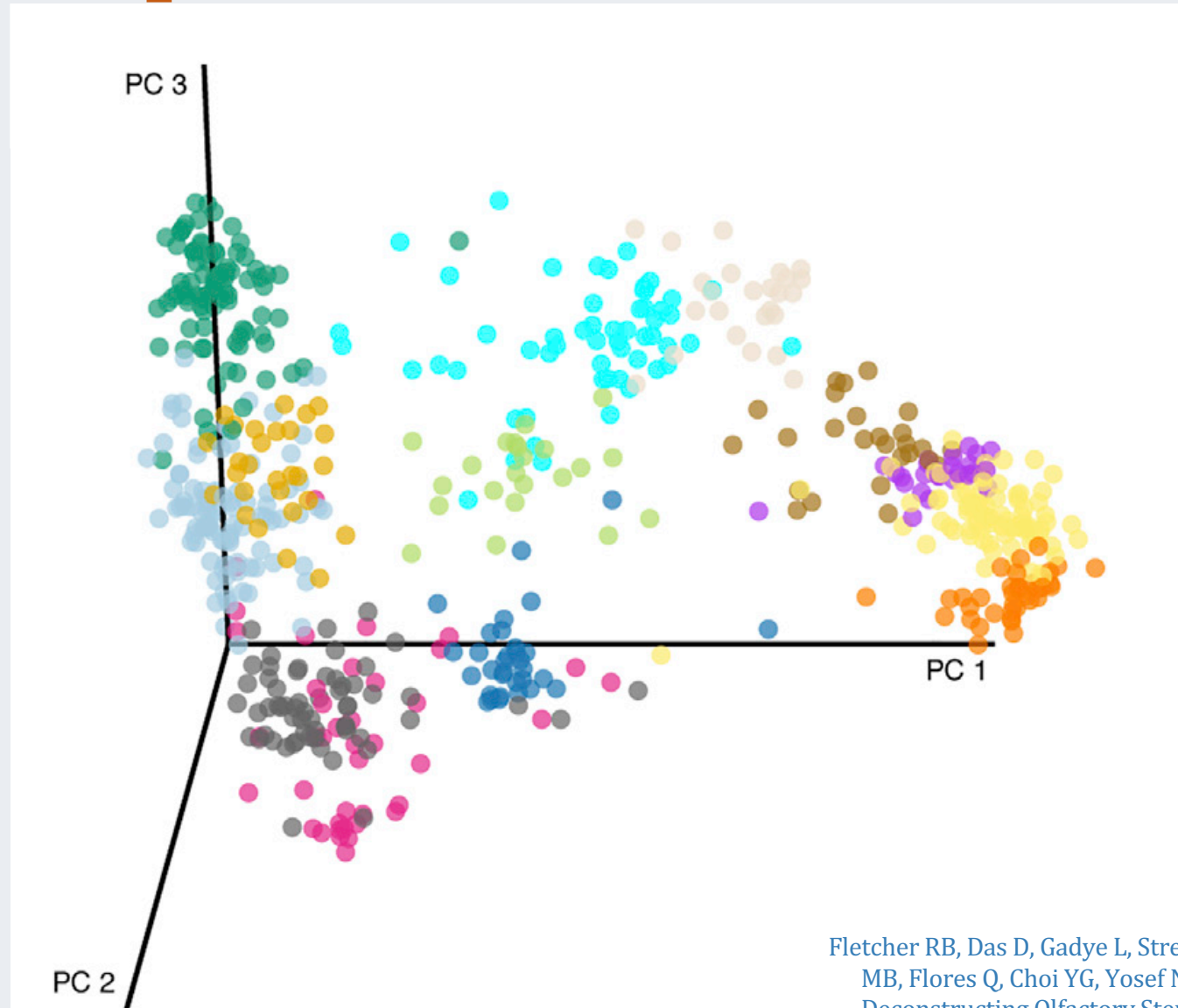
**Globose basal cell (GBC)**

**Horizontal basal cell (HBC)**

**Olfactory ensheathing glia**

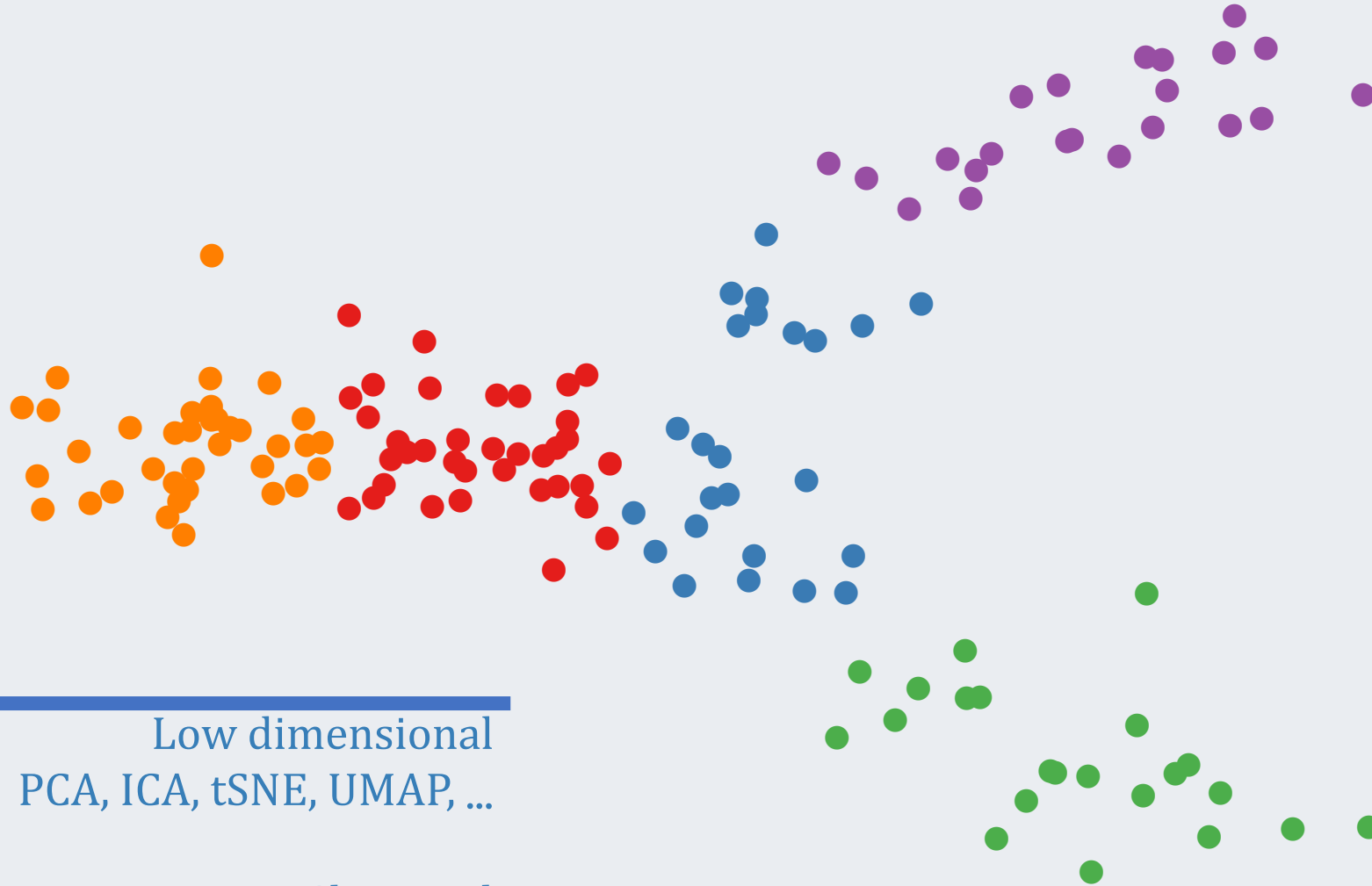
**Bowman's gland**

# Olfactory Epithelium



Fletcher RB, Das D, Gadye L, Street K, Baudhuin A, Risso D, Wagner A, Cole MB, Flores Q, Choi YG, Yosef N, Purdom E, Dudoit S, Ngai J. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell*. 2017; 20(6): 817–30.

# Input Data

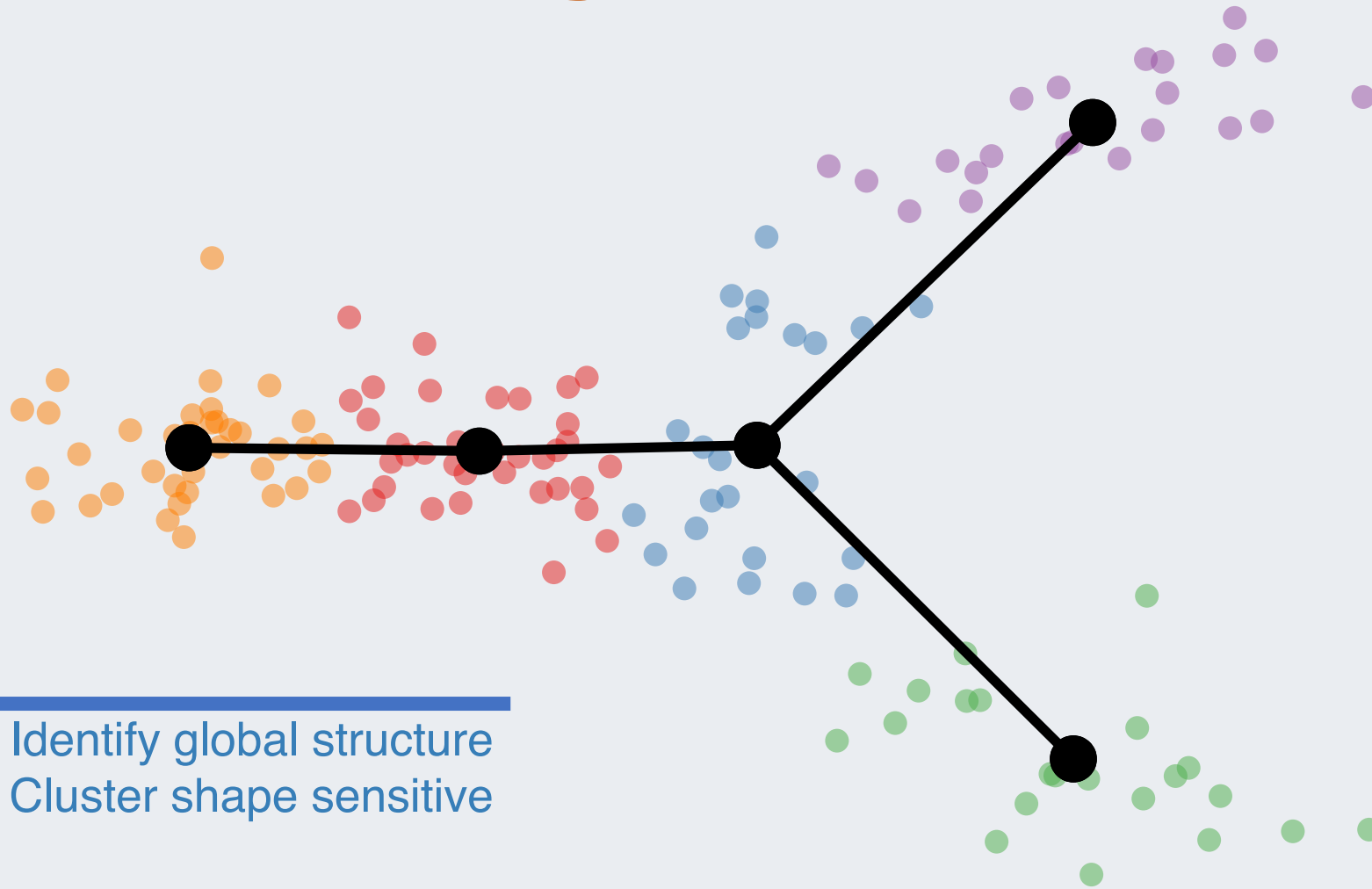


---

Low dimensional  
PCA, ICA, tSNE, UMAP, ...

Clustered  
SC3, Seurat, RSEC, ...

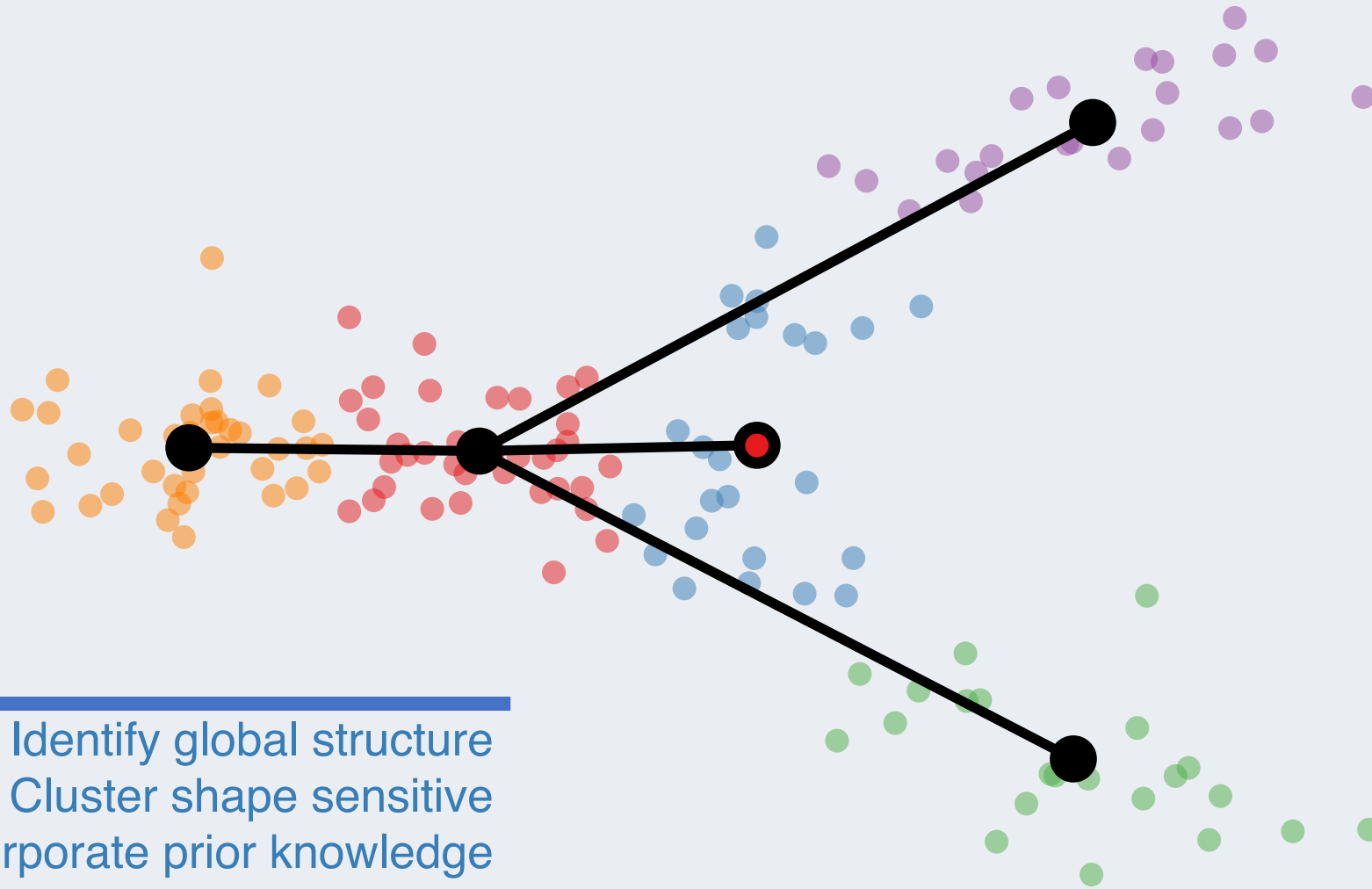
# Minimal Spanning Tree



---

Identify global structure  
Cluster shape sensitive

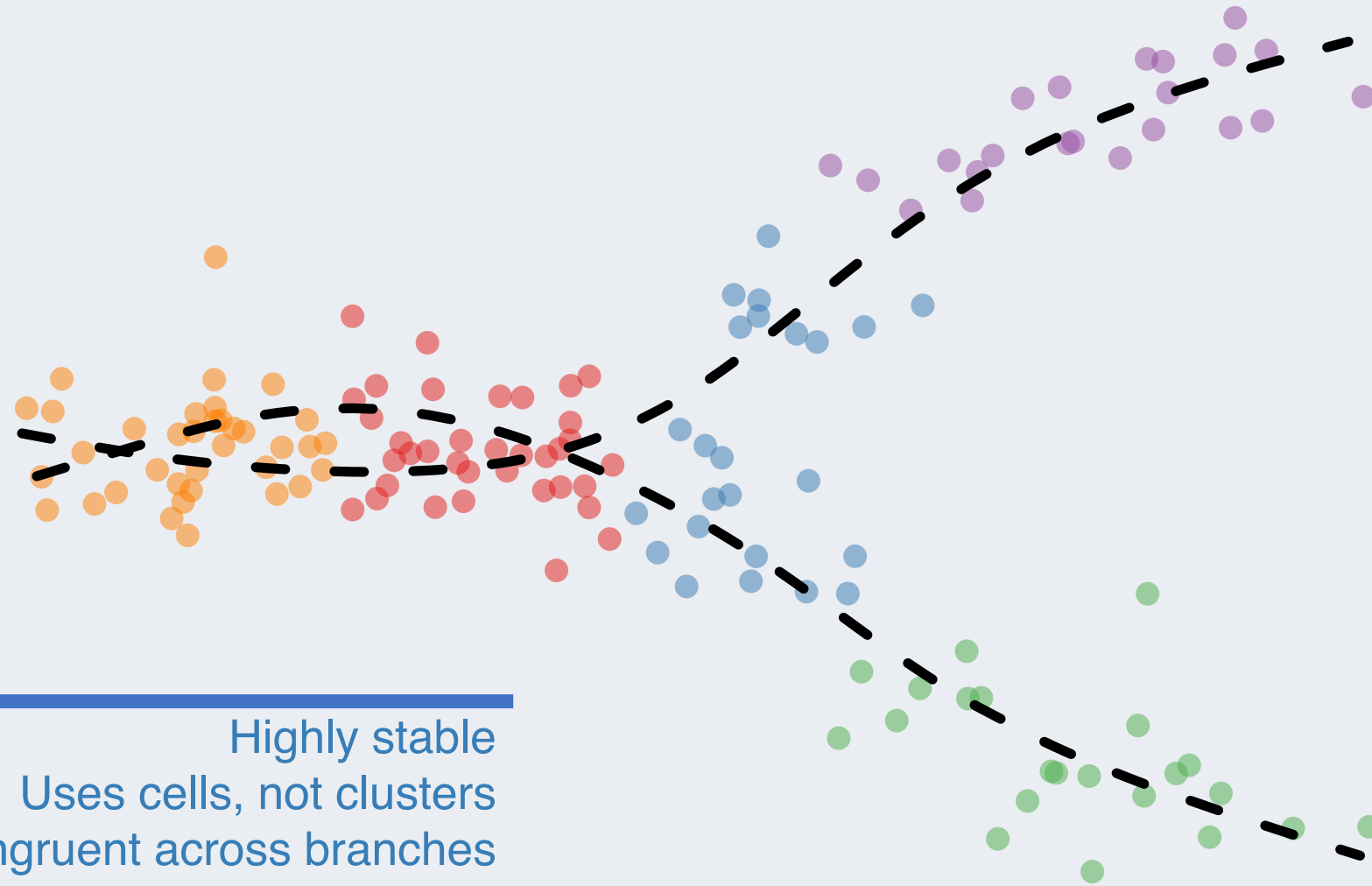
# Constrained MST



---

Identify global structure  
Cluster shape sensitive  
Incorporate prior knowledge

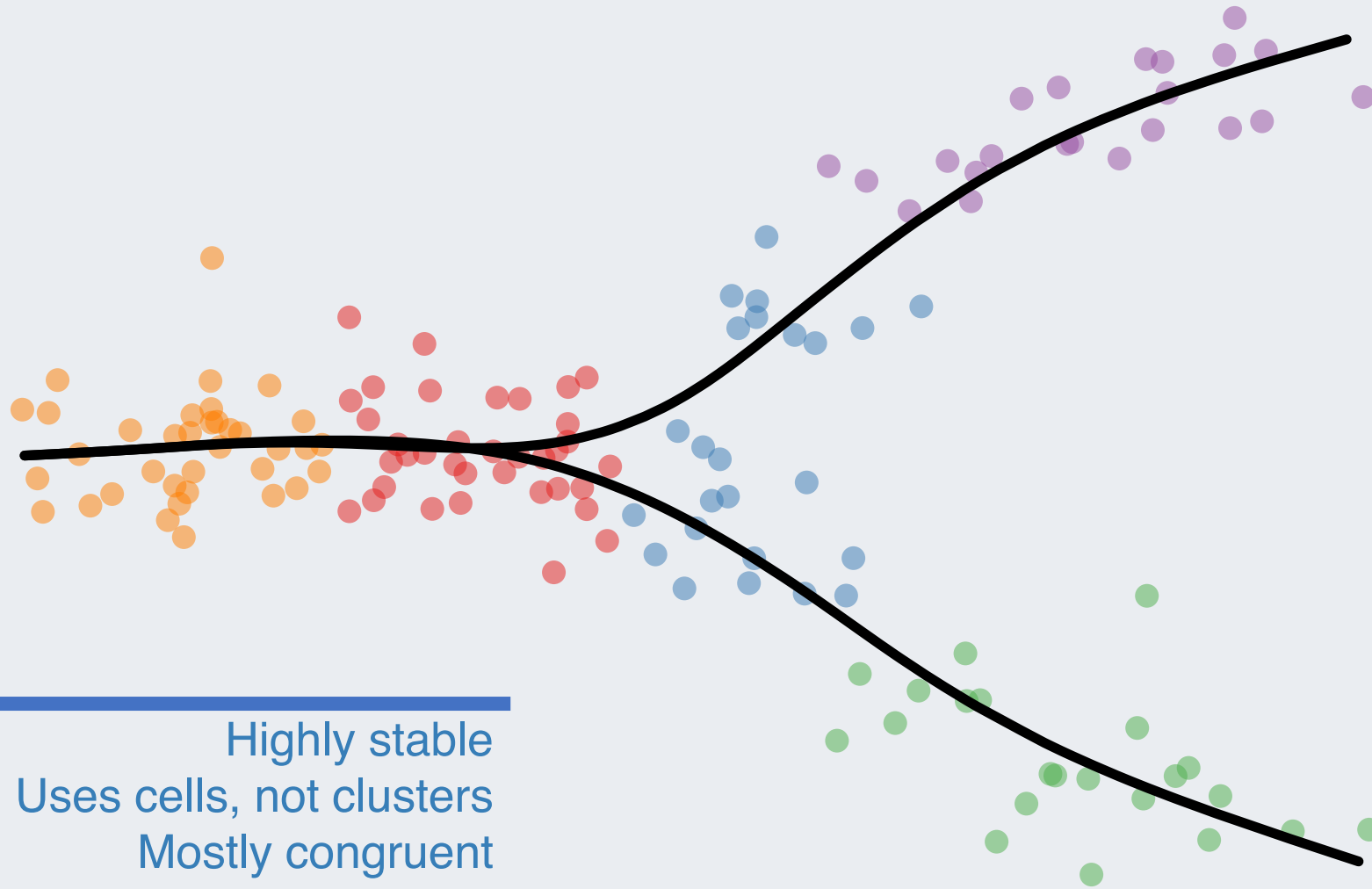
# Principal Curves



---

Highly stable  
Uses cells, not clusters  
Incongruent across branches

# Simultaneous Principal Curves

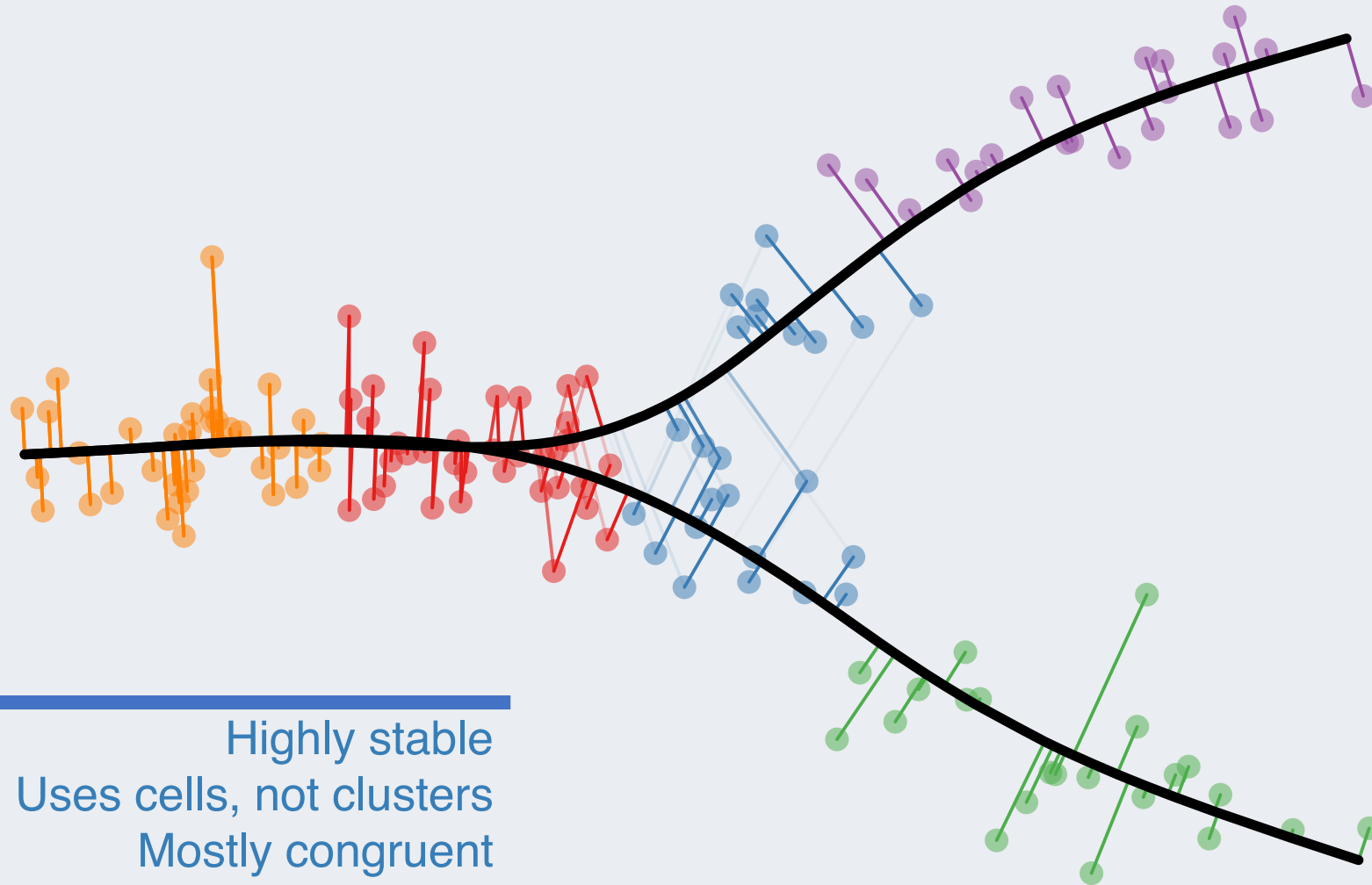


---

Highly stable  
Uses cells, not clusters  
Mostly congruent  
across branches



# Computing Pseudotime

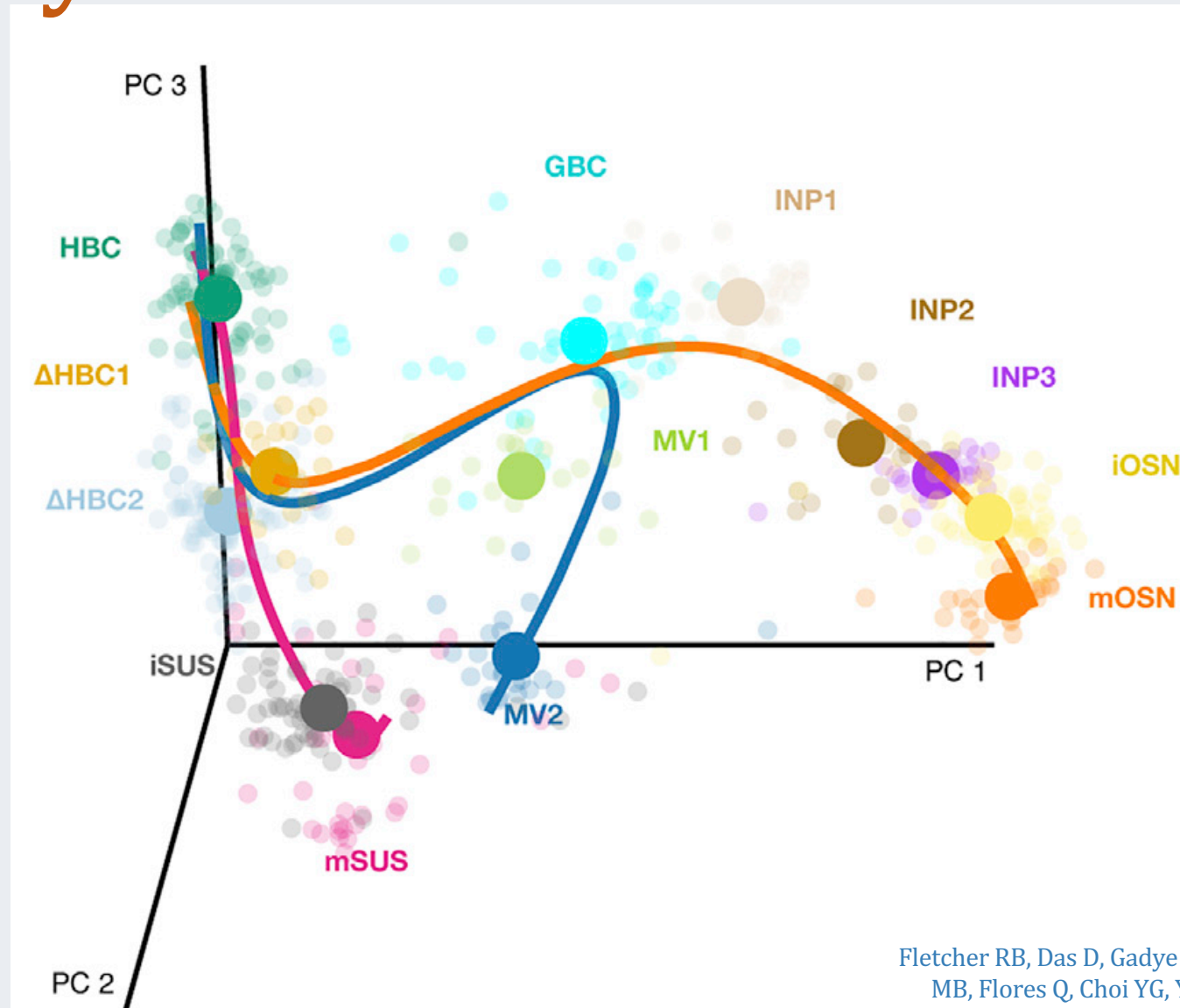


---

Highly stable  
Uses cells, not clusters  
Mostly congruent  
across branches

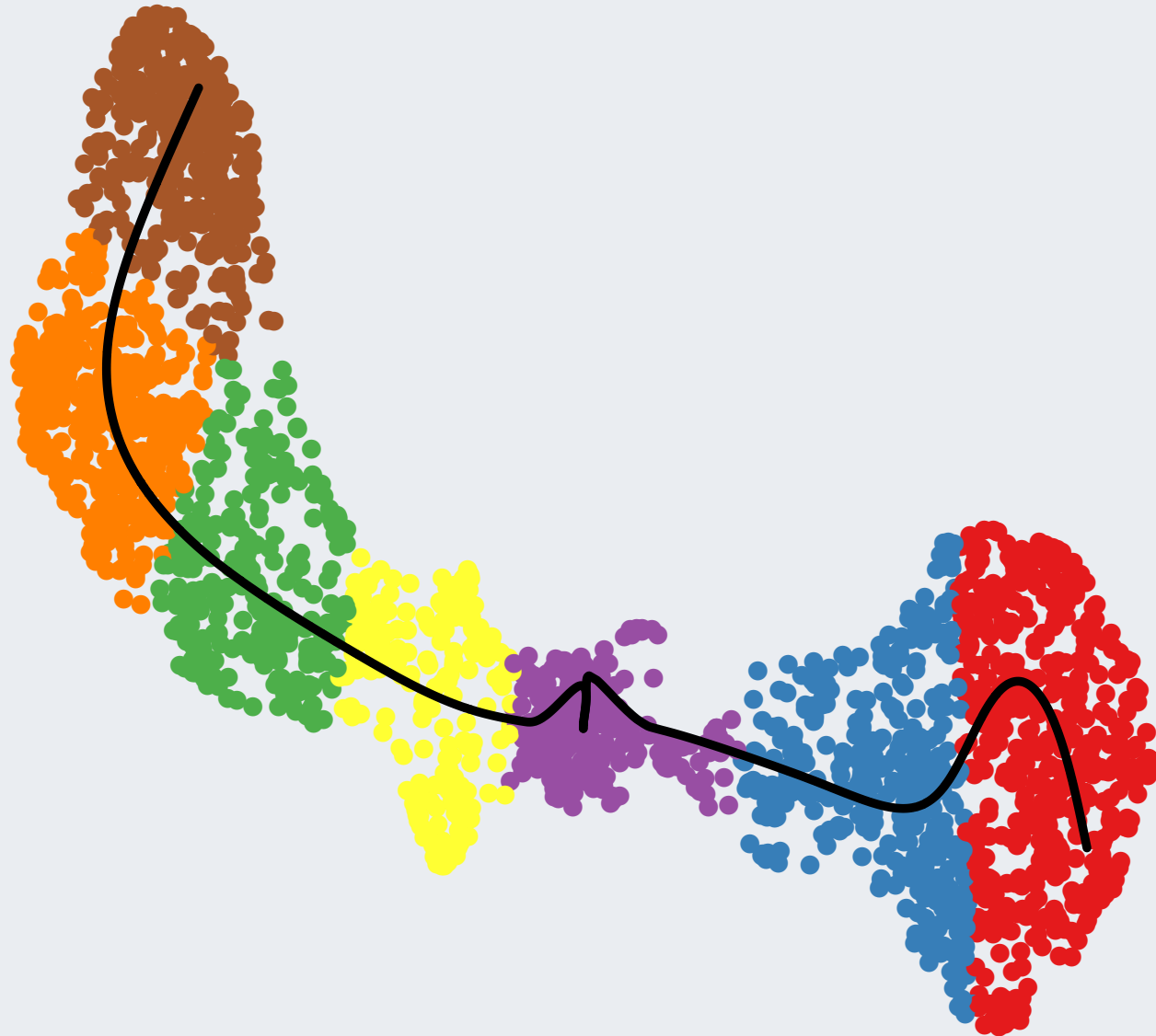
Kelly Street, Davide Risso, Russell B. Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics, 19(1): 477, 12 2018. ISSN 1471-2164. doi: 10.1186/s12864-018-4772-0. URL <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4772-0>

# Trajectory inference



Fletcher RB, Das D, Gadye L, Street K, Baudhuin A, Risso D, Wagner A, Cole MB, Flores Q, Choi YG, Yosef N, Purdom E, Dudoit S, Ngai J. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell*. 2017; 20(6): 817–30.

# Trajectory inference



➤ Finding developmental paths

Each cell has a pseudotime, which measure how far along it is in the developmental process



# Challenges of Slingshot

**Slingshot** can only tackle tree structures.

It can not handle connected (including cyclic) trajectories, nor non connected trajectories.

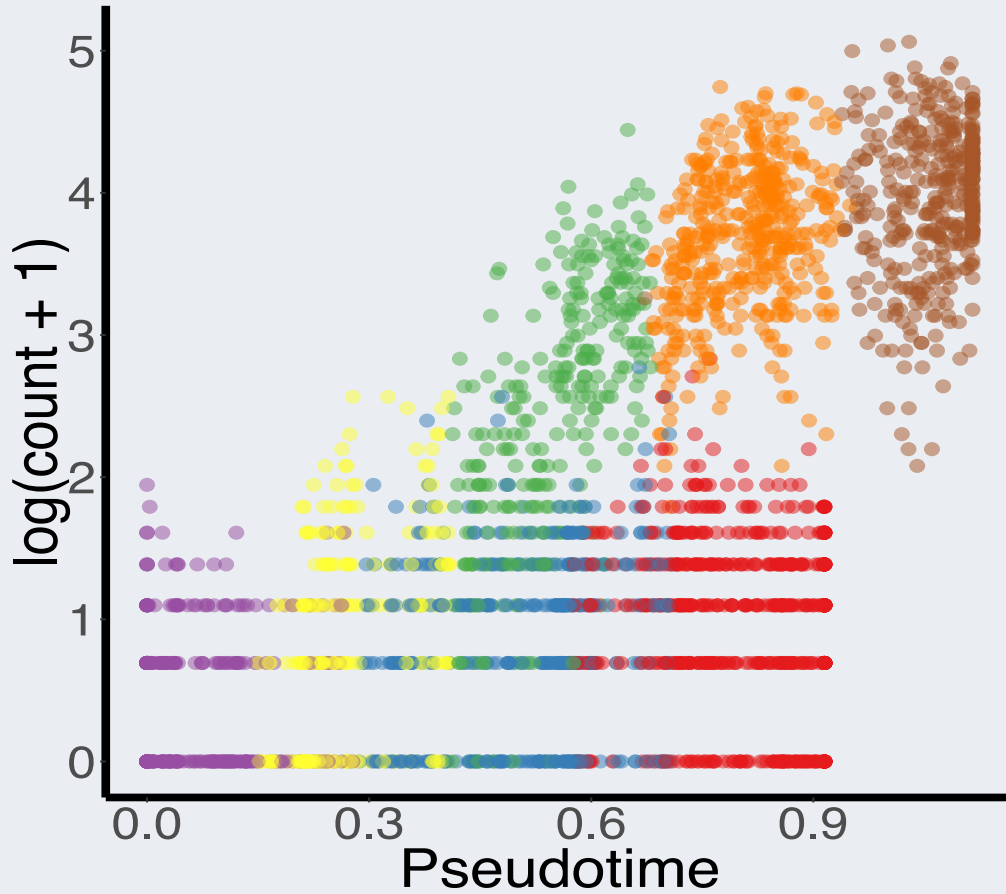


Can be handled by slingshot

### 3. Differential Expression with *tradeSeq*

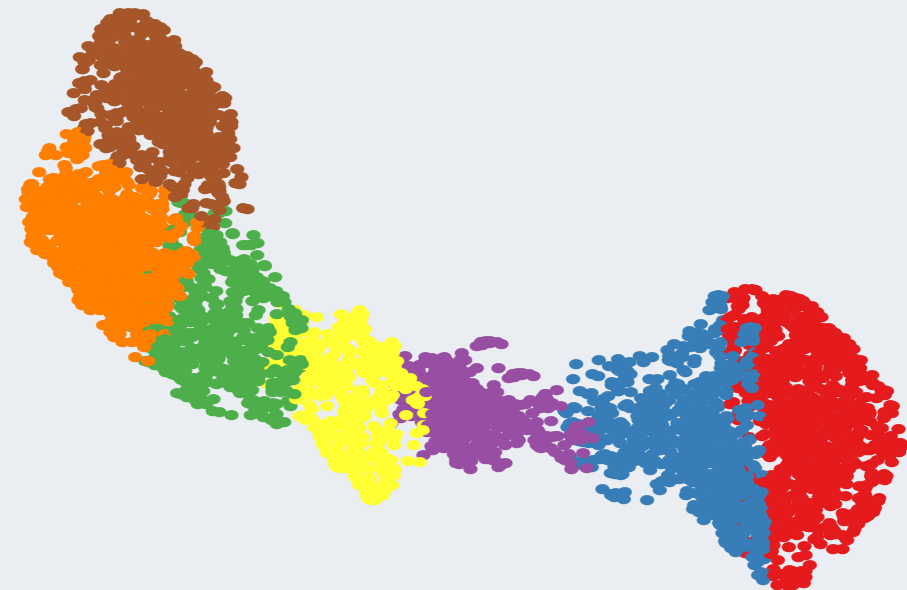


# cluster-based DE is artificial



Genes are now expressed in a continuous manner (since 2014)

Differential Expression is still cluster-based, i. e. discrete.





# Trajectory-based DE

We developed *tradeSeq*, an algorithm that leverages the continuous nature of scRNA-Seq.

- Available as an **R** package on Github ([statOmics/tradeSeq](https://github.com/statOmics/tradeSeq)). Soon on Bioconductor.
- Modular tool that work with any dimensionality reduction and trajectory inference method.



# Statistical model

$$Y_{gi} = NB(\mu_{gi}, \phi_g)$$

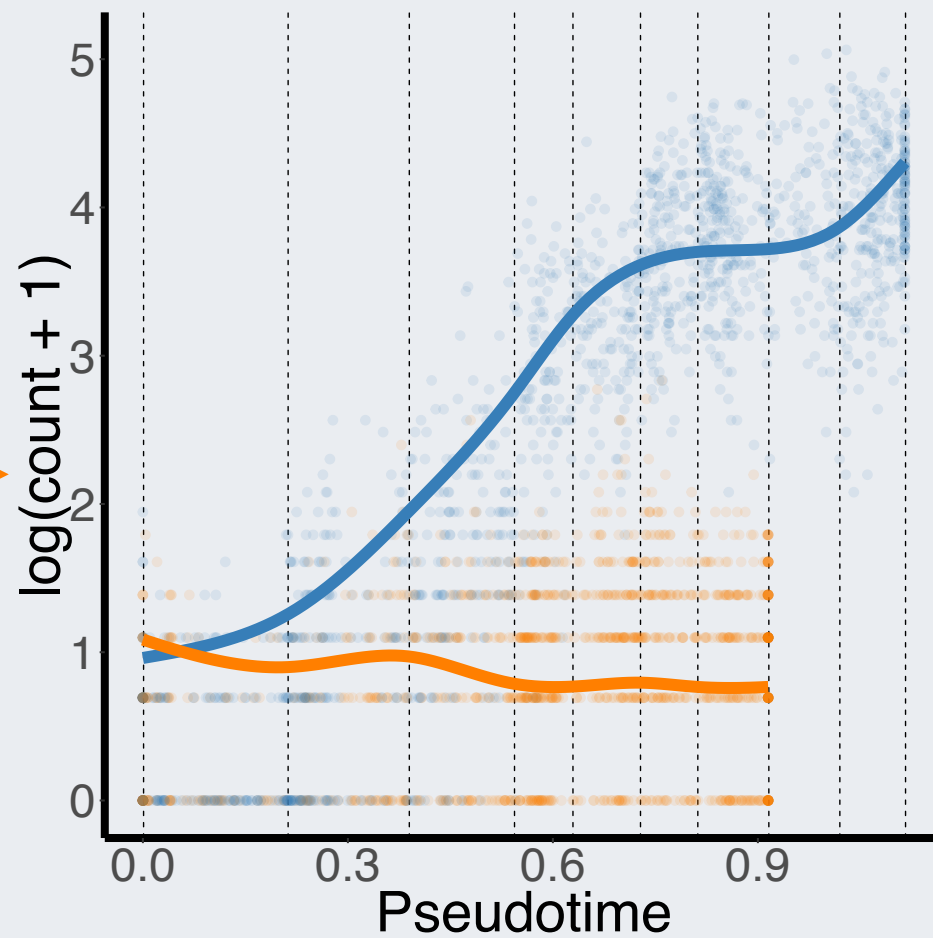
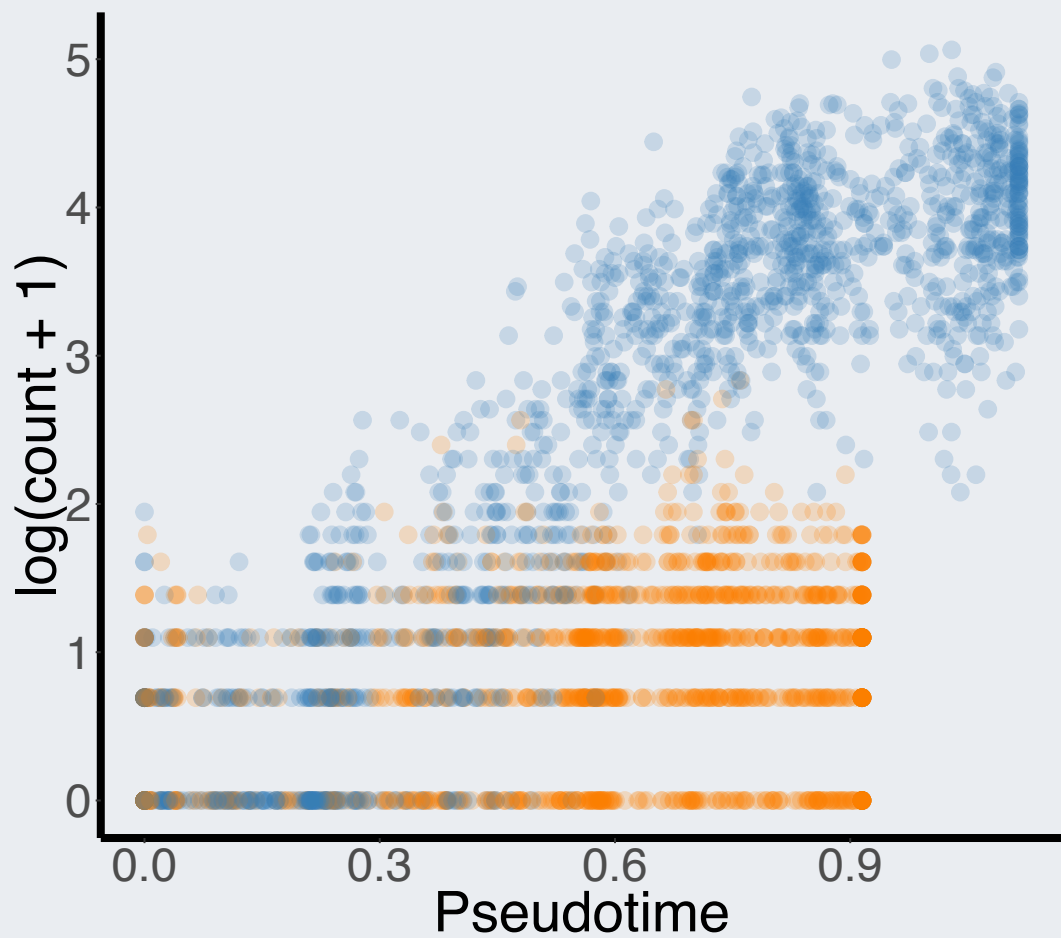
$$\mu_{gi} = \sum_{l=1}^n s_{gl}(T_i) Z_{li} + \mathbf{U}_i \alpha_g + \log(N_i)$$

Can accommodate

- Design matrix
- Different sequencing depth
- Weights



# Statistical model



# An investigation tool



$$s_{gl}(T_i) = \sum_{k=1}^K b_k(t) \beta_{g lk}$$

Testing null hypotheses of the form:

$$H_0: \mathbf{C}^T \beta_g = 0$$

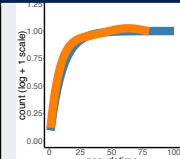
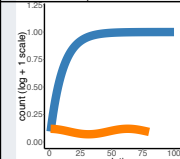
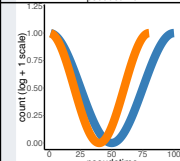
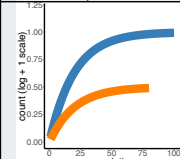
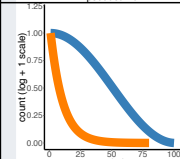
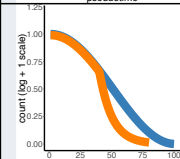
Using Wald Statistics of the form:

$$W_g = \hat{\beta}_g^T \mathbf{C} (\mathbf{C}^T \hat{\Sigma} \mathbf{C})^{-1} \mathbf{C}^T \hat{\beta}_g$$

# An investigation tool



## Differential Expression Tests

Lineages	Within the orange lineage		Between the orange and blue lineages		
	associationTest	startVsEndTest	diffEndTest	patternTest	earlyDETest
	DE	DE	Not DE	Not DE	Not DE
	Not DE	Not DE	DE	DE	DE
	DE	Not DE	Not DE	Not DE	Not DE
	DE	DE	DE	DE	Not DE
	DE	DE	Not DE	DE	DE
	DE	DE	Not DE	DE	Not DE

# Association test

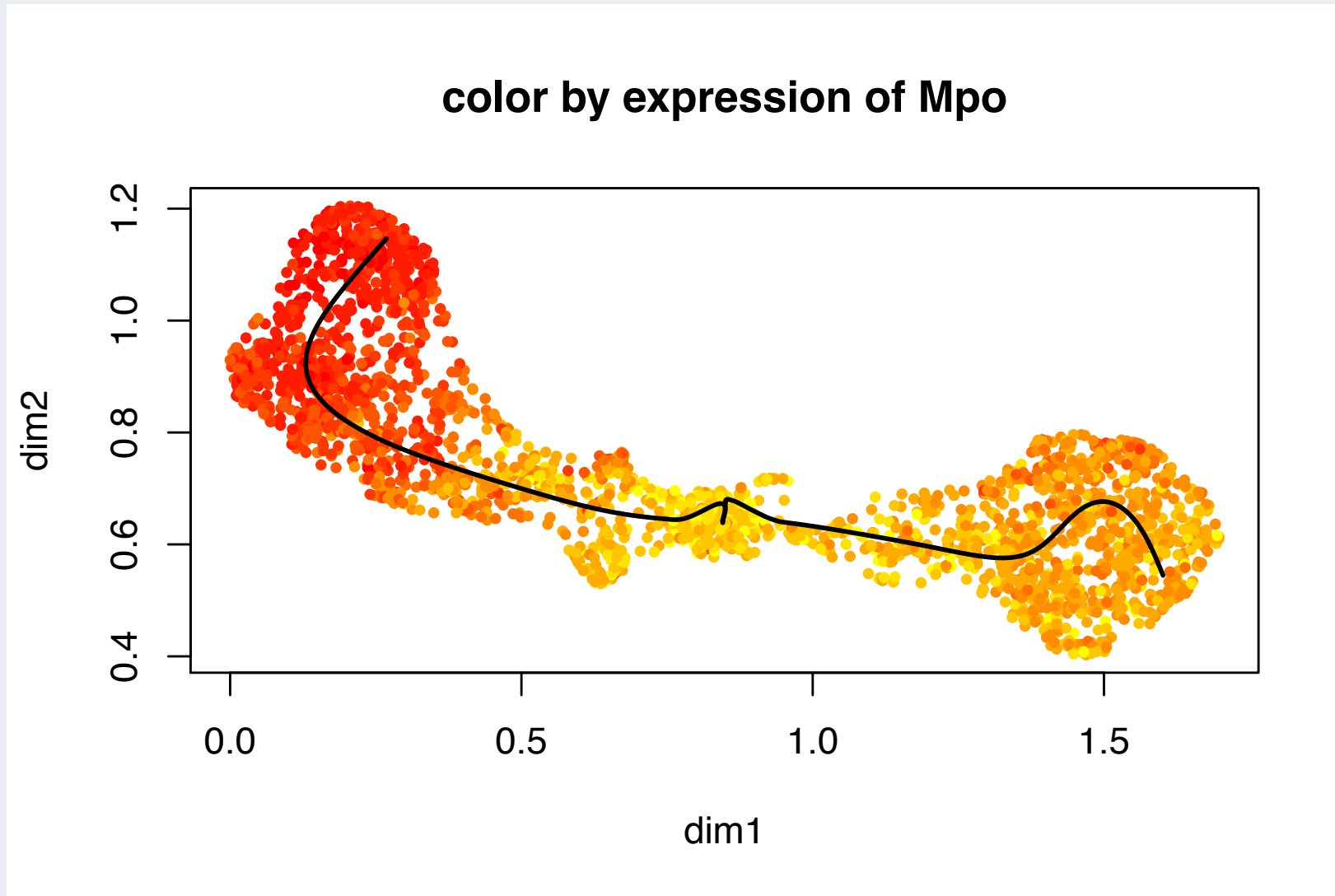


$$H_0: \beta_{lk_g} = \beta_{lk'_g} \text{ for all } k \neq k'$$

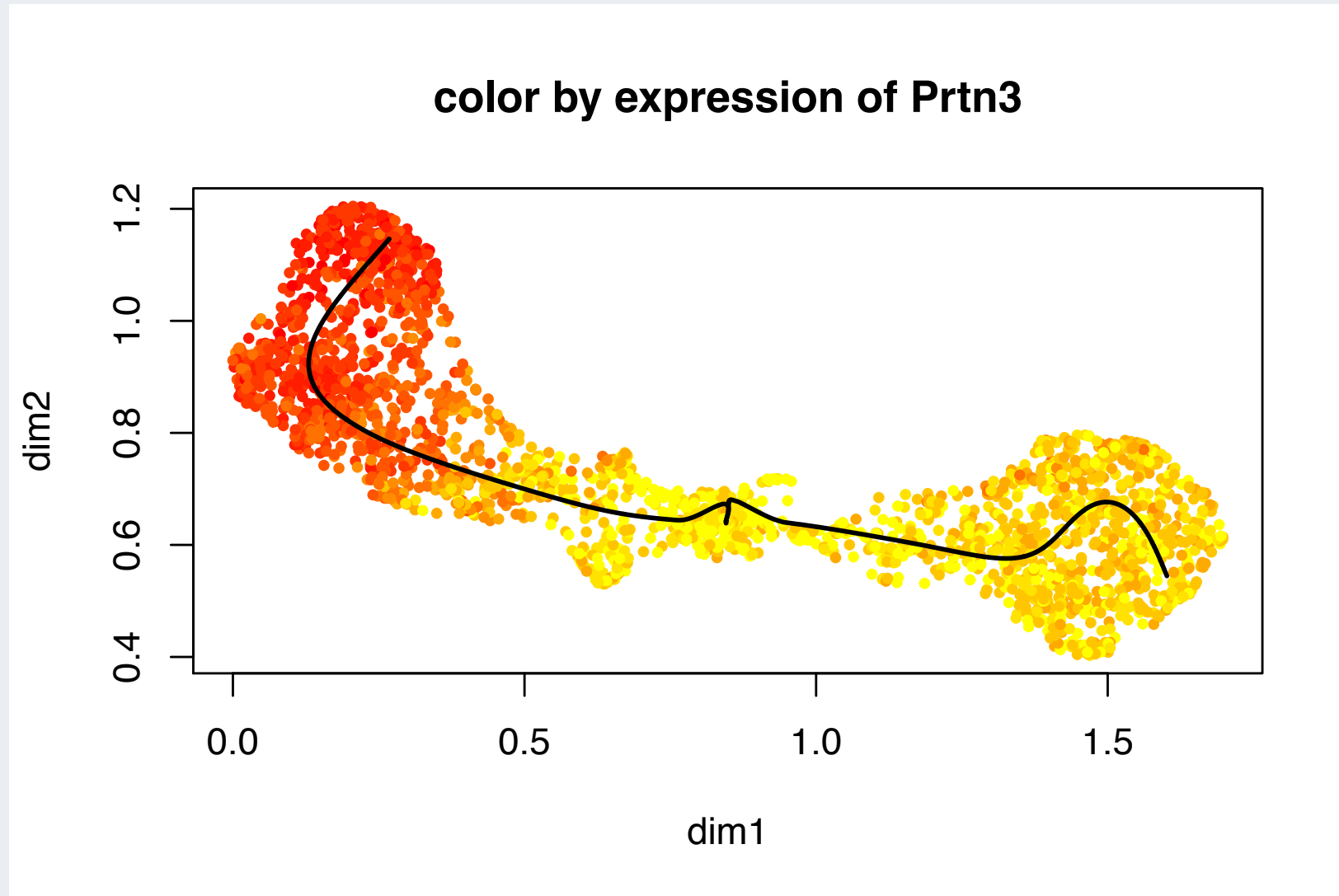
Contrast matrix

$\beta_{l1g}$	$\beta_{l2g}$	$\beta_{l3g}$	...	$\beta_{lk_g}$
1	-1	0	...	0
0	1	-1	...	0
0	0	1	...	0
...	...	...	...	...
-1	0	0	...	1

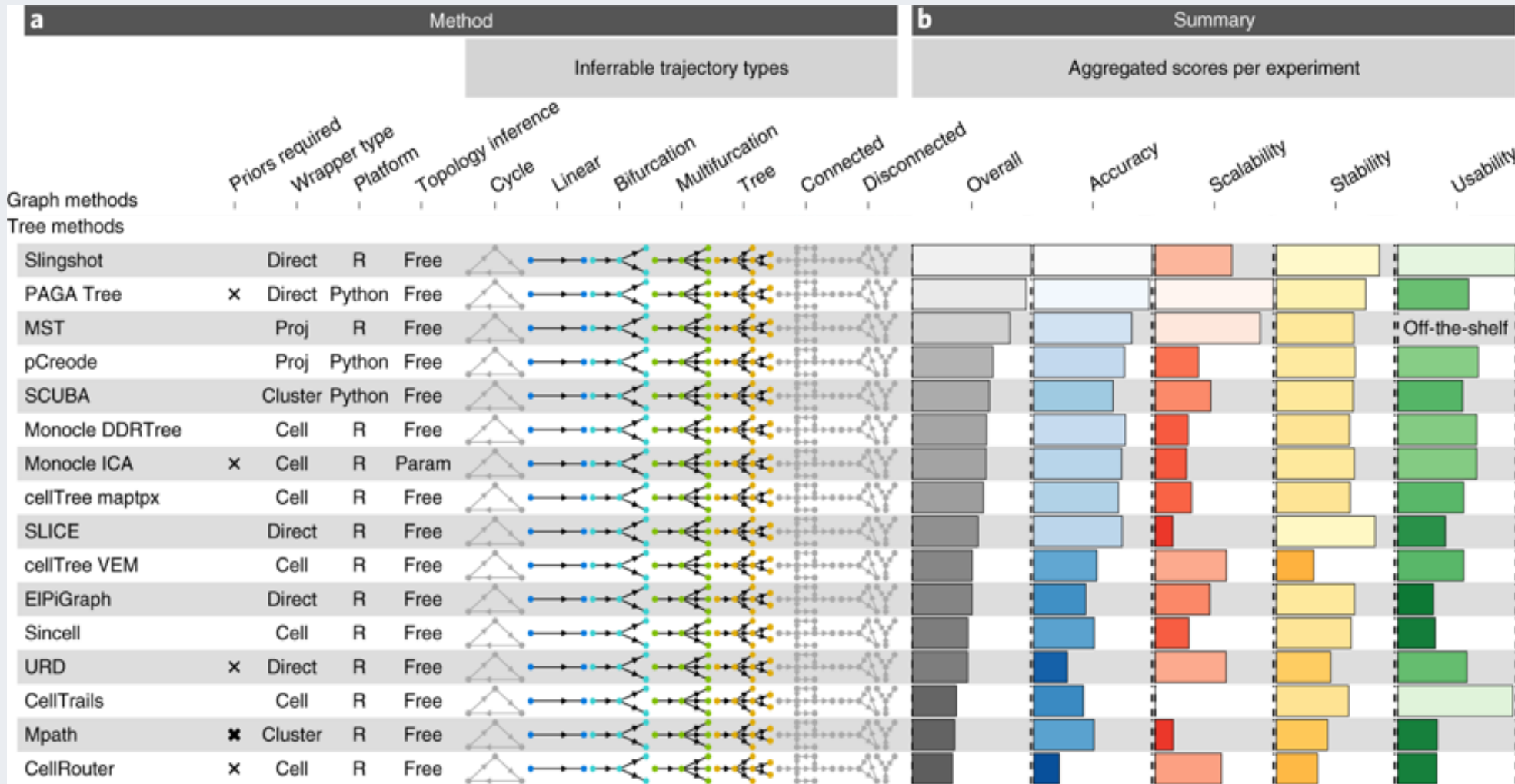
# StartVsEndTest



# DiffEndTest

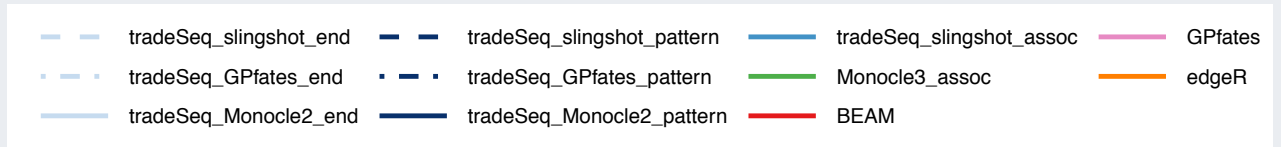
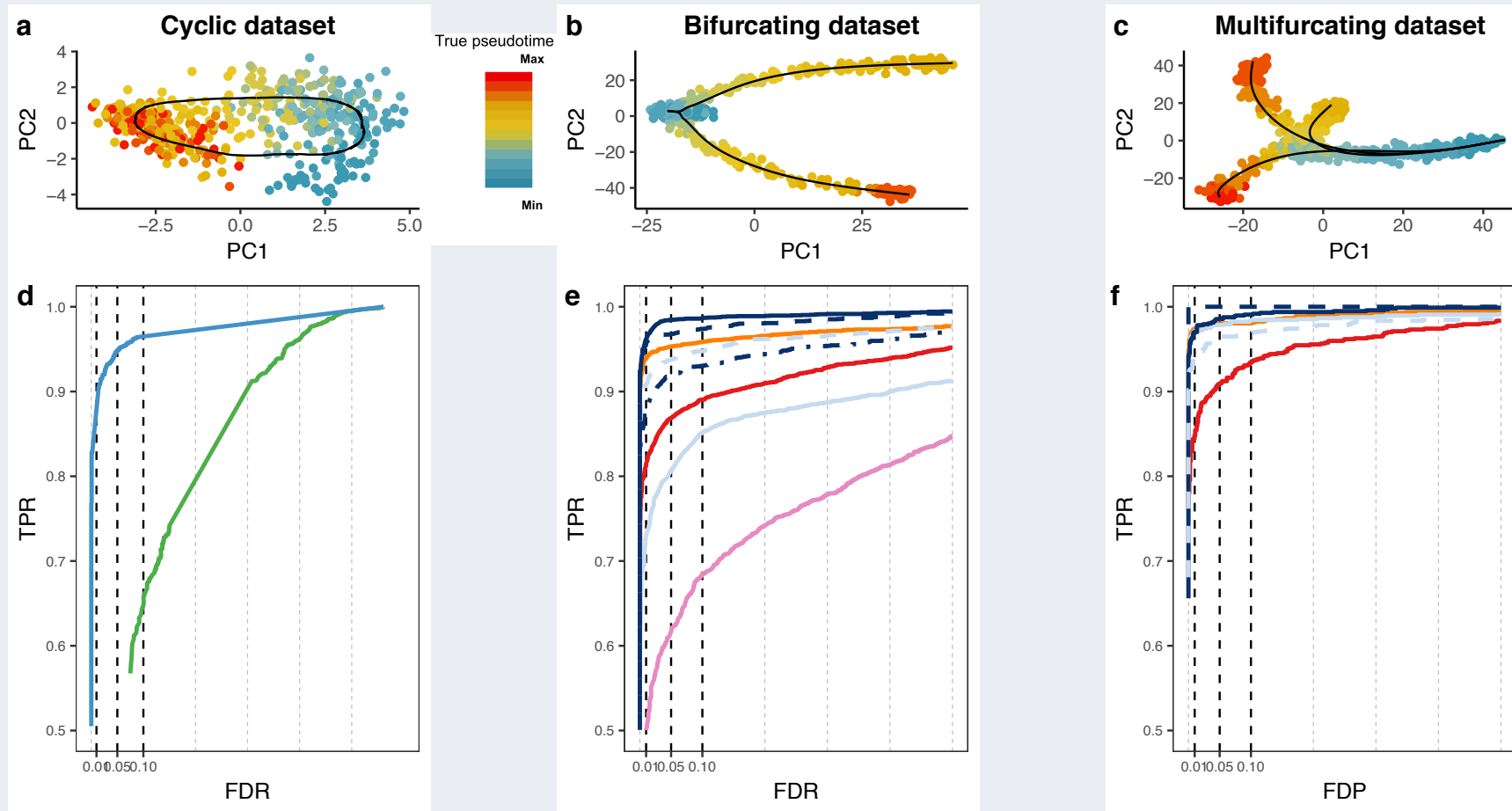


# Simulation framework: dynverse



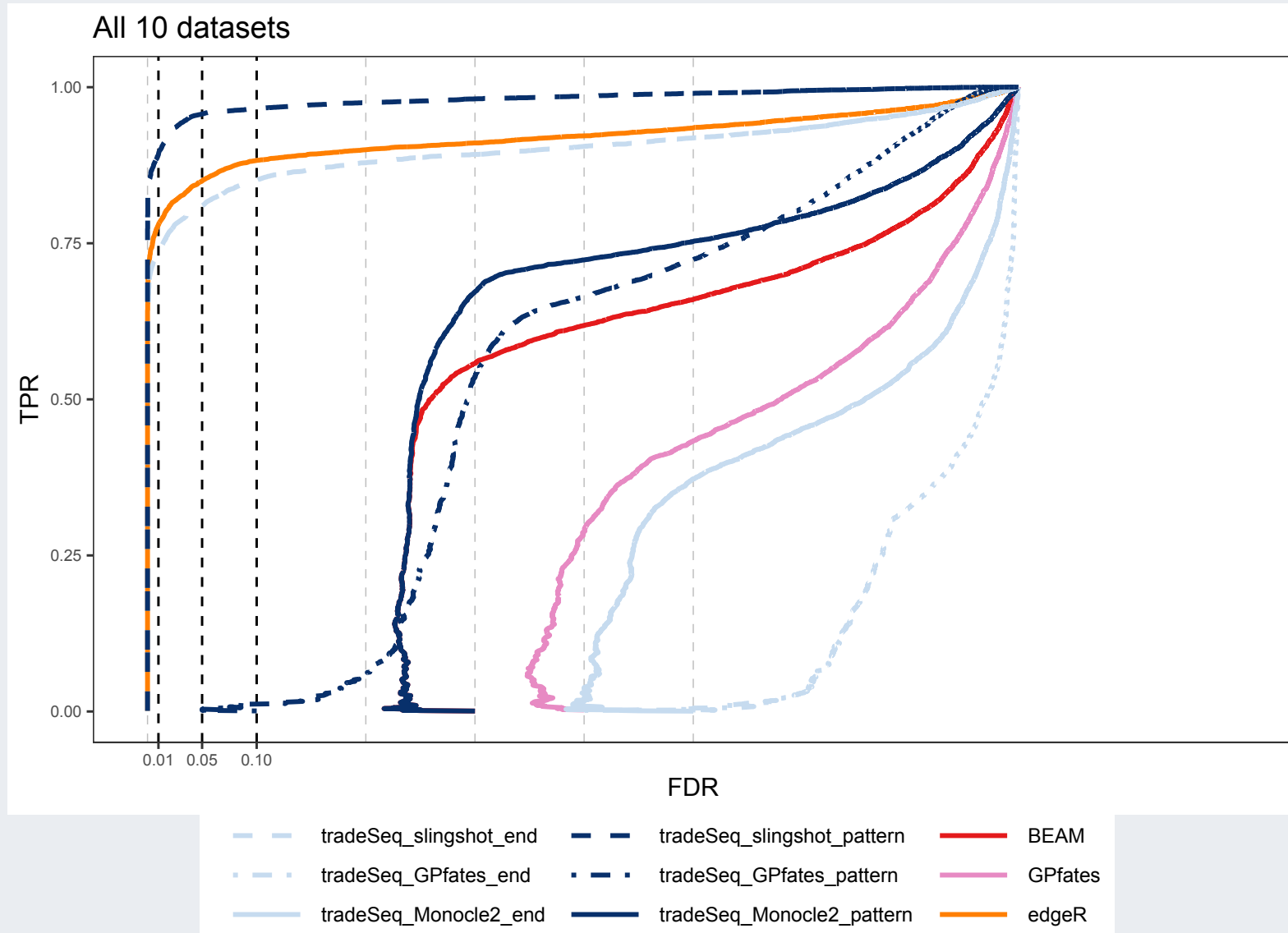
Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeyns. A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, page 1, 4 2019. ISSN 1087-0156. doi: 10.1038/s41587-019-0071-9. URL <http://www.nature.com/articles/s41587-019-0071-9>

# Outperforms existing methods





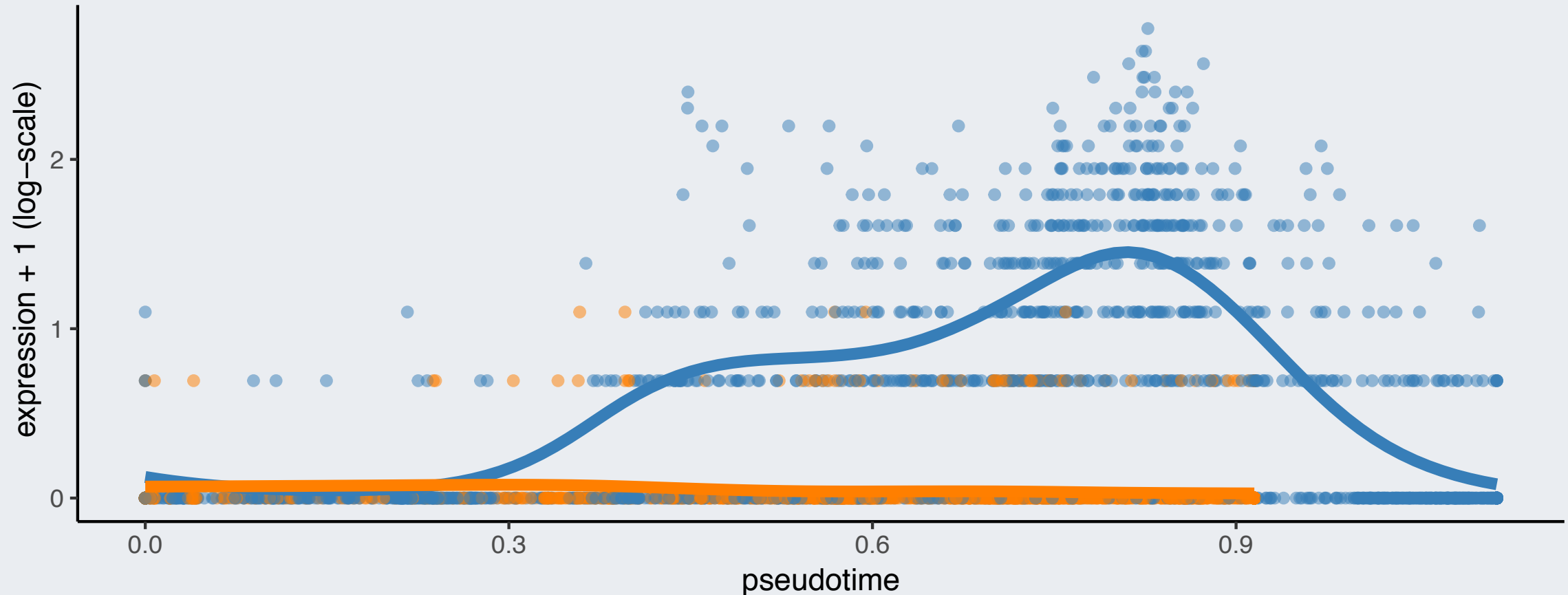
# Outperforms existing methods



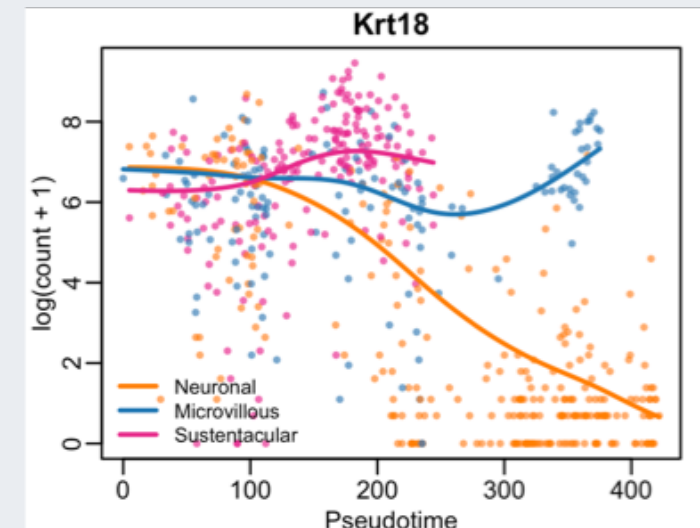
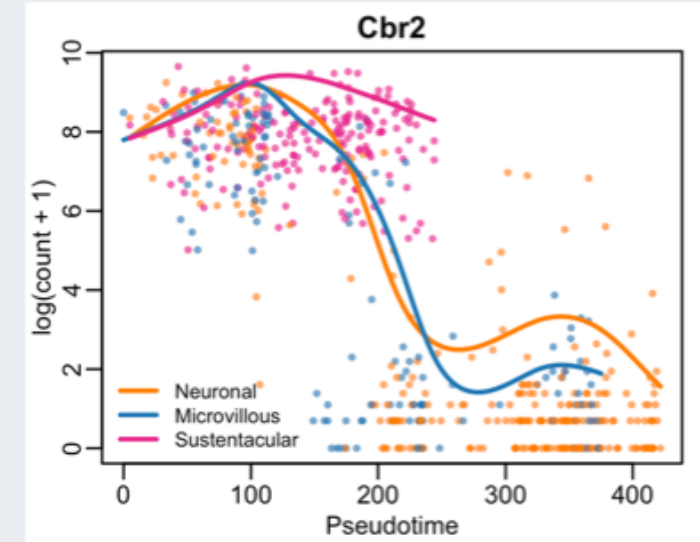
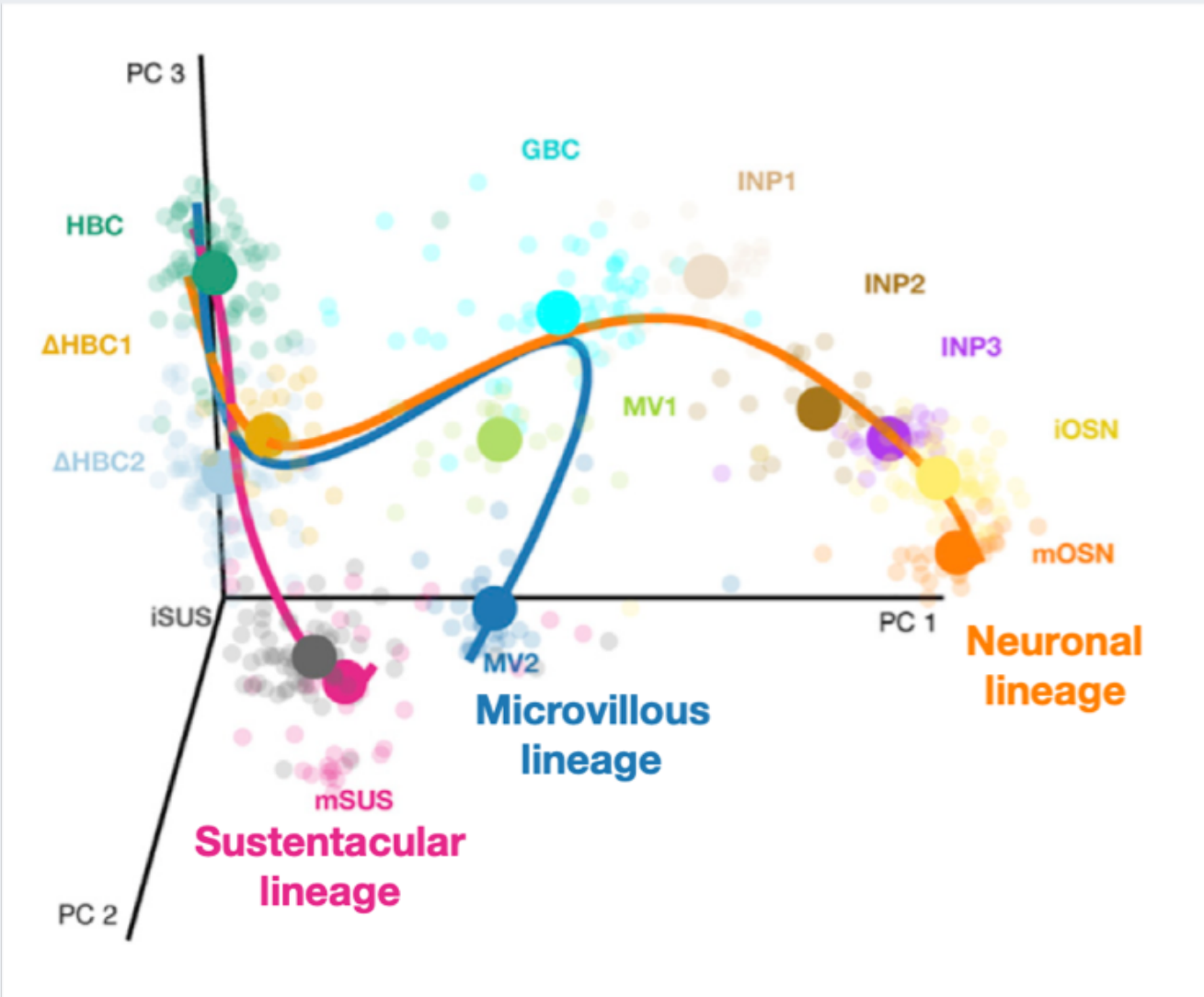
# Provides unique insights



Gene Irf8 in the bone marrow dataset



# Also works with multiple lineages





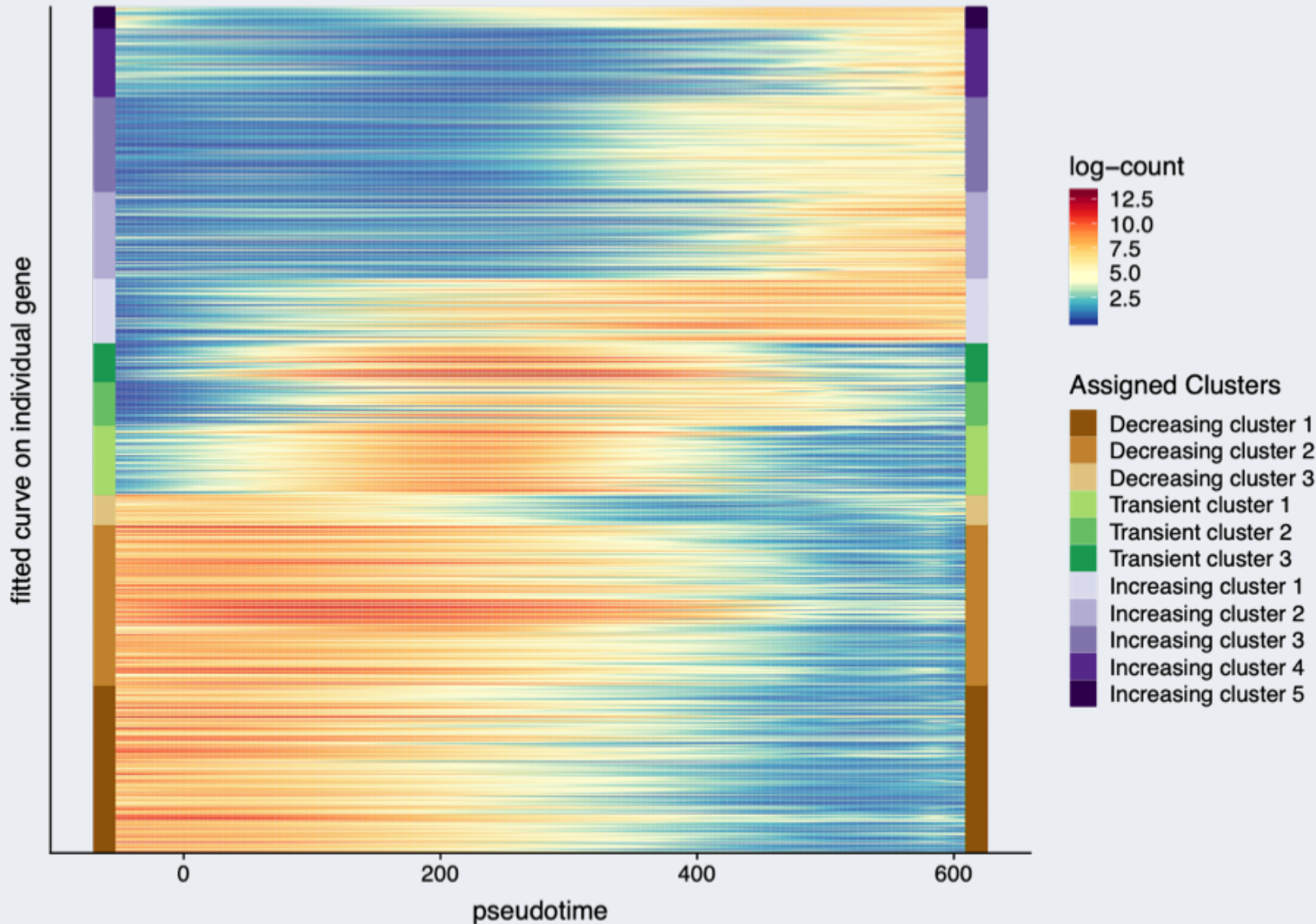
# Perspectives for tradeSeq

- Possible to develop new tests, especially to look at speed or acceleration of gene changes.
- Zero-inflation weights are estimated before the smoothers. Future improvements could focus on joint-improvements.
- Publish the paper and software

# 3. Clustering gene patterns with **RSEC**



# Single lineage clustering

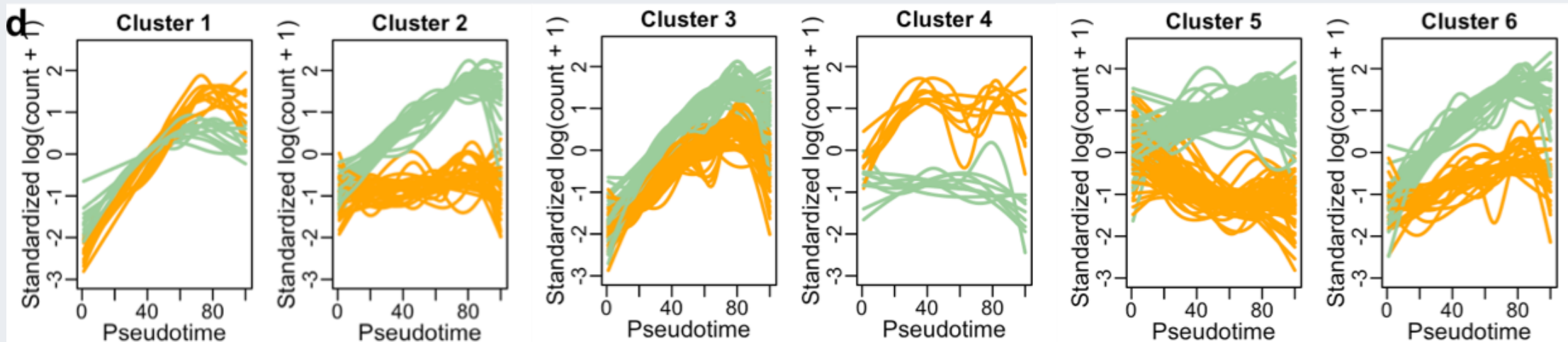


Embryogenesis datasets

- Cluster with RSEC on the genes, using the  $\beta_{lkg}$  as features

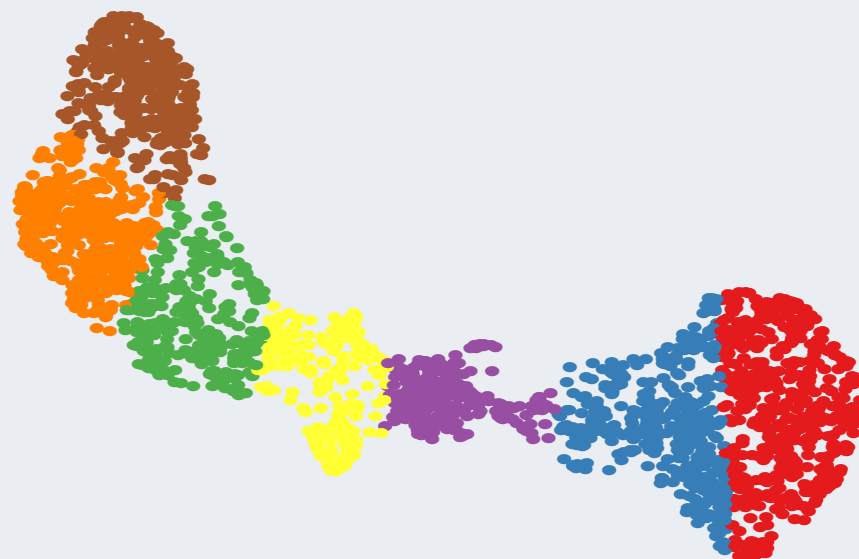


# Multiple lineages clustering



Bone marrow dataset:

Clusters for the top 500 genes





# Limitations

- Clustering with more than one lineage is hard to interpret (and sometimes leads to adherent results).
- Most filtering or merging criteria used in **RSEC** are not applicable here. Filtering only based on cluster size might miss small but very strong signals.
- Current work by Stephanie DeGraaf might be more promising.





Sandrine Dudoit



Lieven Clement



Kelly Street



Koen Van den Berge



Thank you for listening

Any questions?