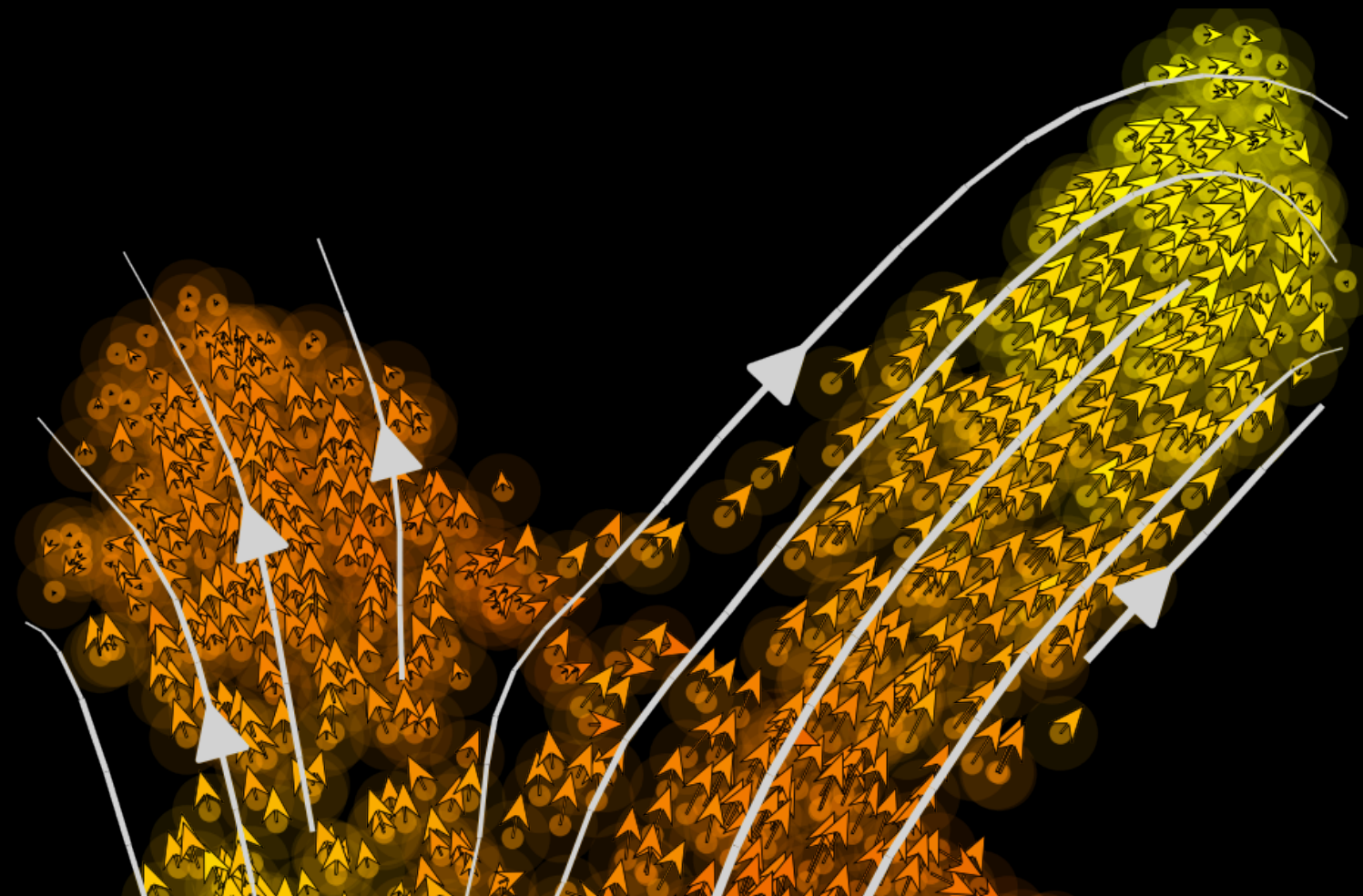


# NBIS/SIB single-cell school RNA velocity

Aug 30, 2021

Volker Bergen  
vbergen@cellarity.com





MSc Mathematics & MBA Finance

(2011-2017)



PhD Computational Biology

(2017-2020)

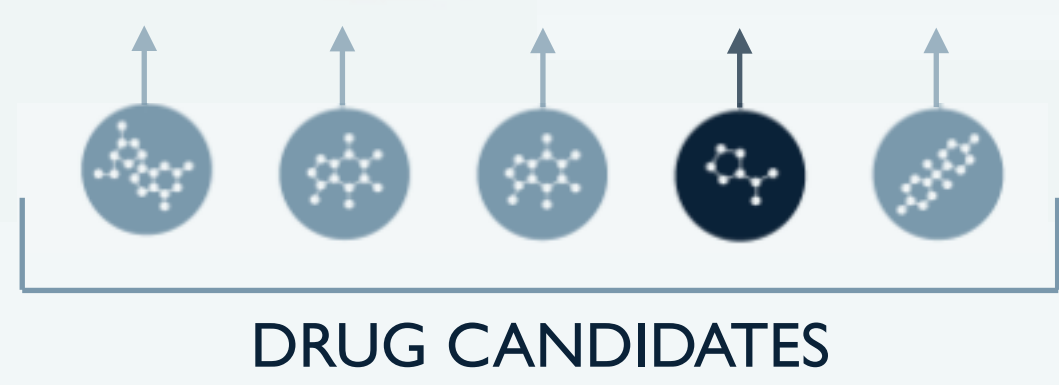
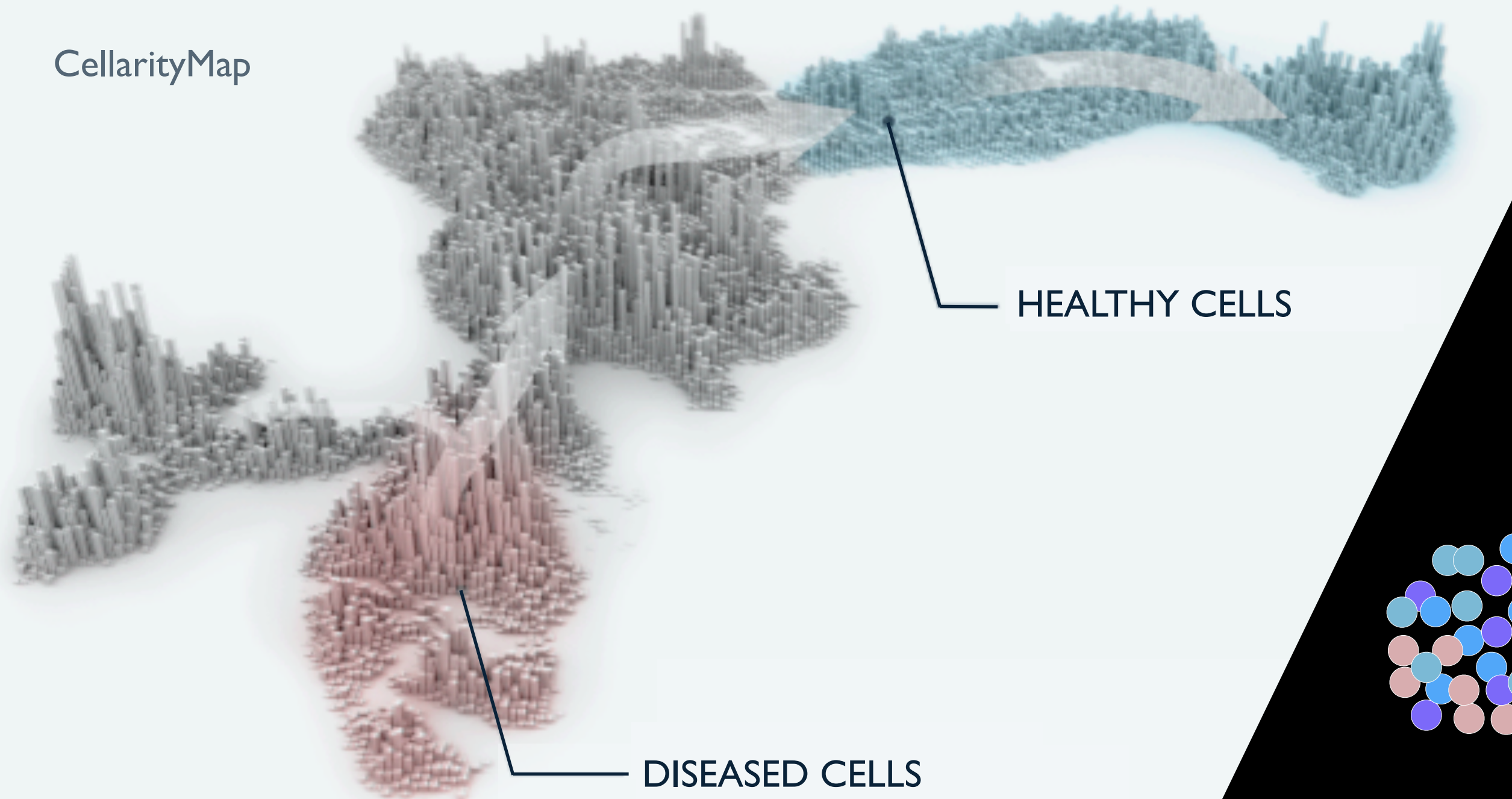


ML Scientist & Strategy

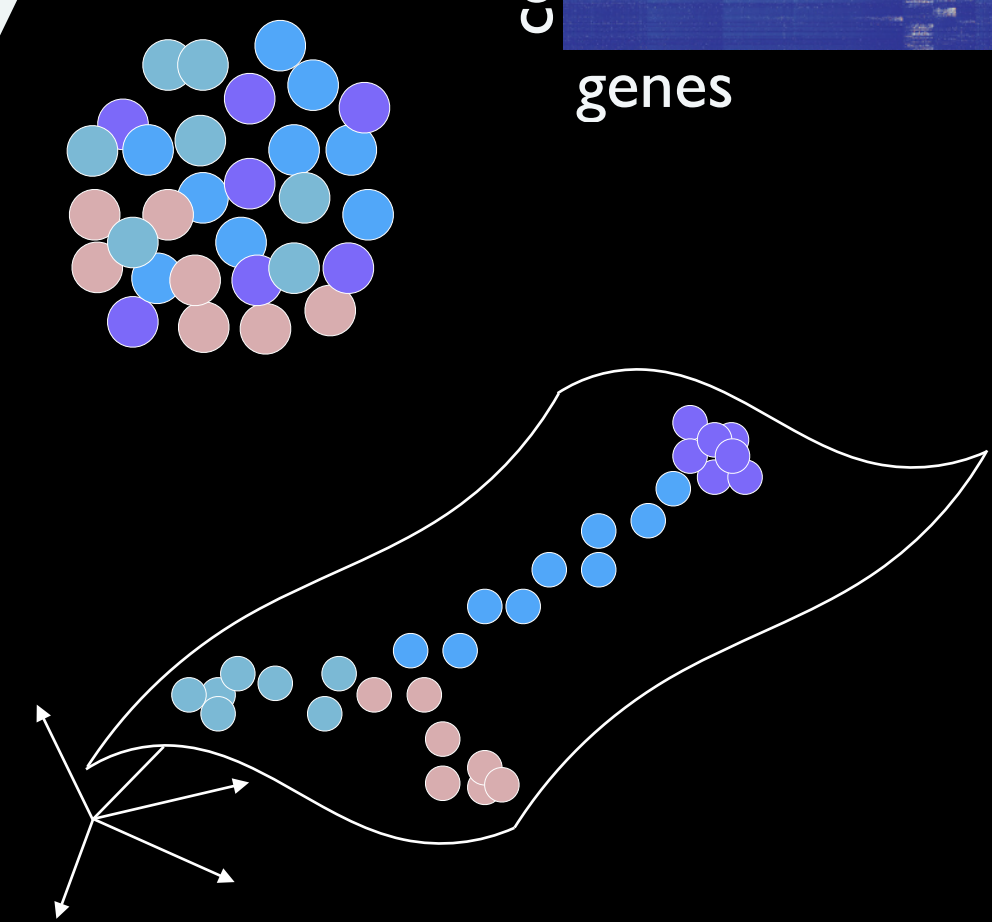
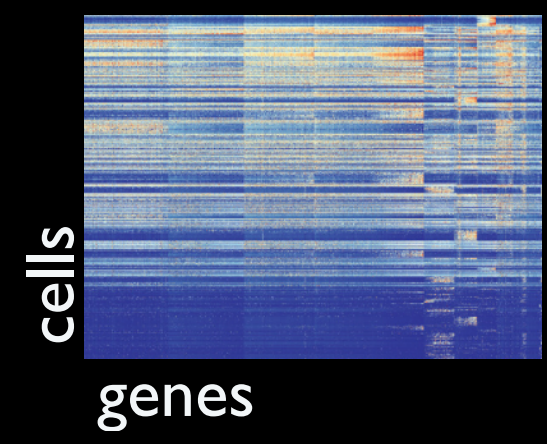
(present)



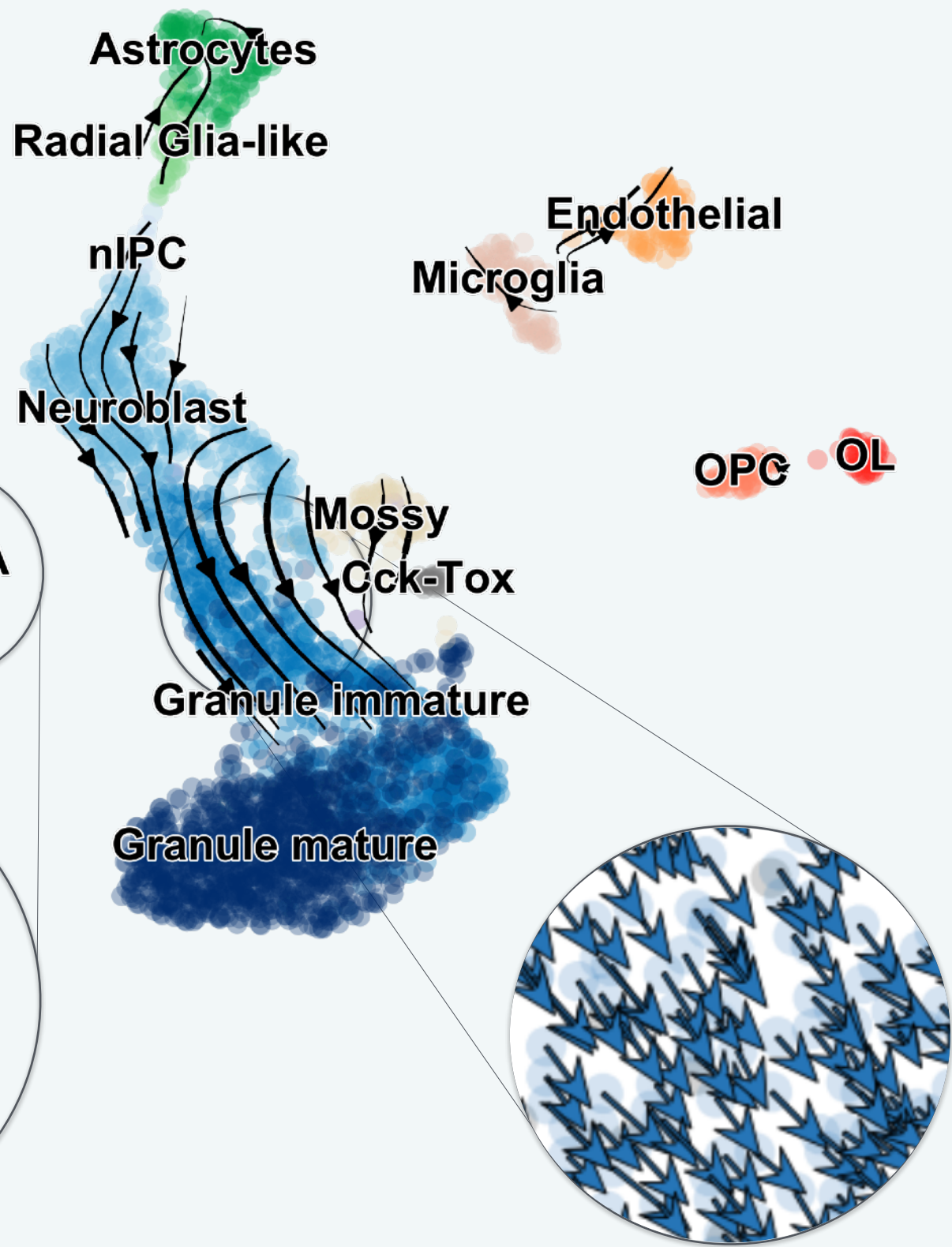
CellarityMap



  
scanpy  
Wolf et al. (2018)







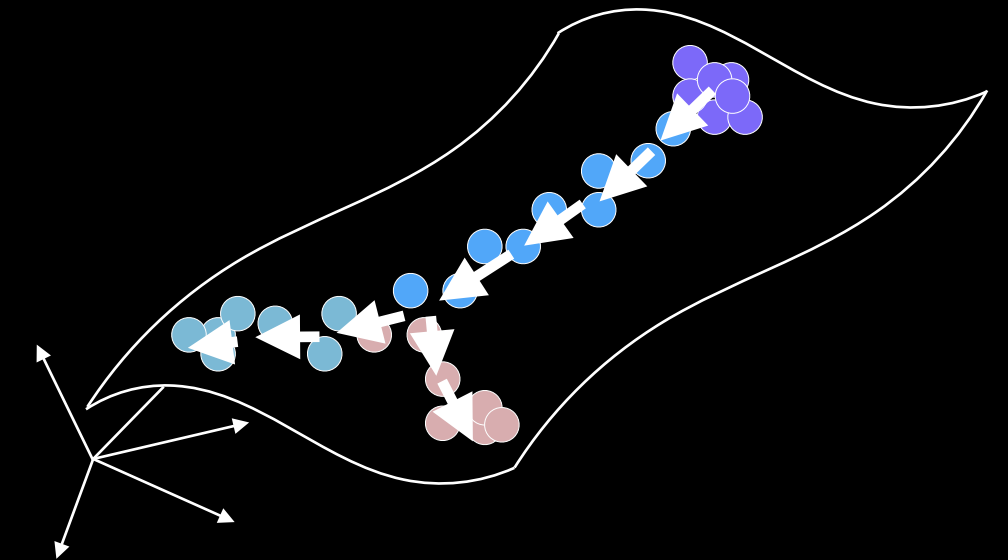
## RNA velocity

La Manno *et al.* (Nature 2018)  
 Bergen *et al.* (Nature Biotech 2020)

## Reviews

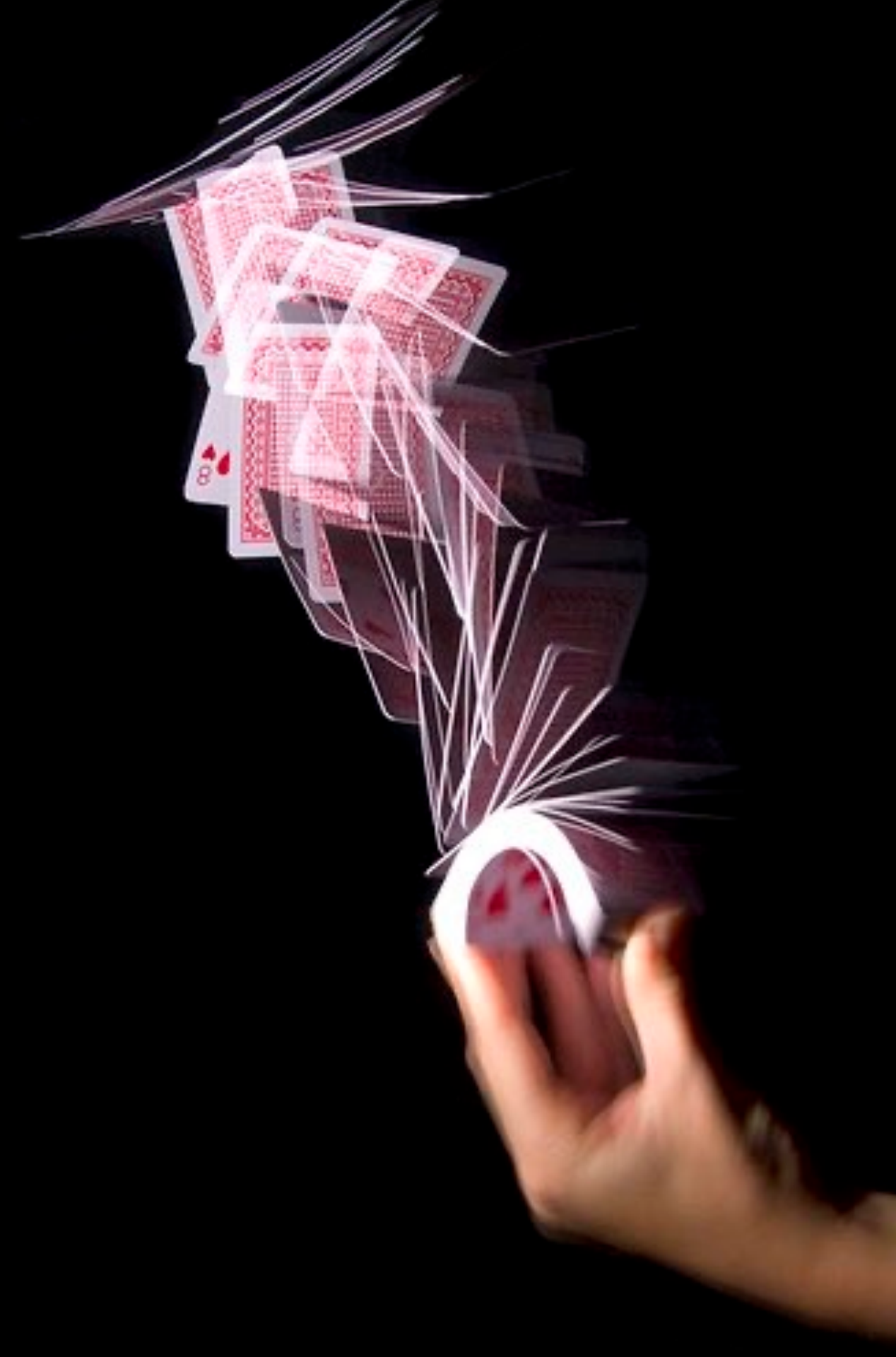
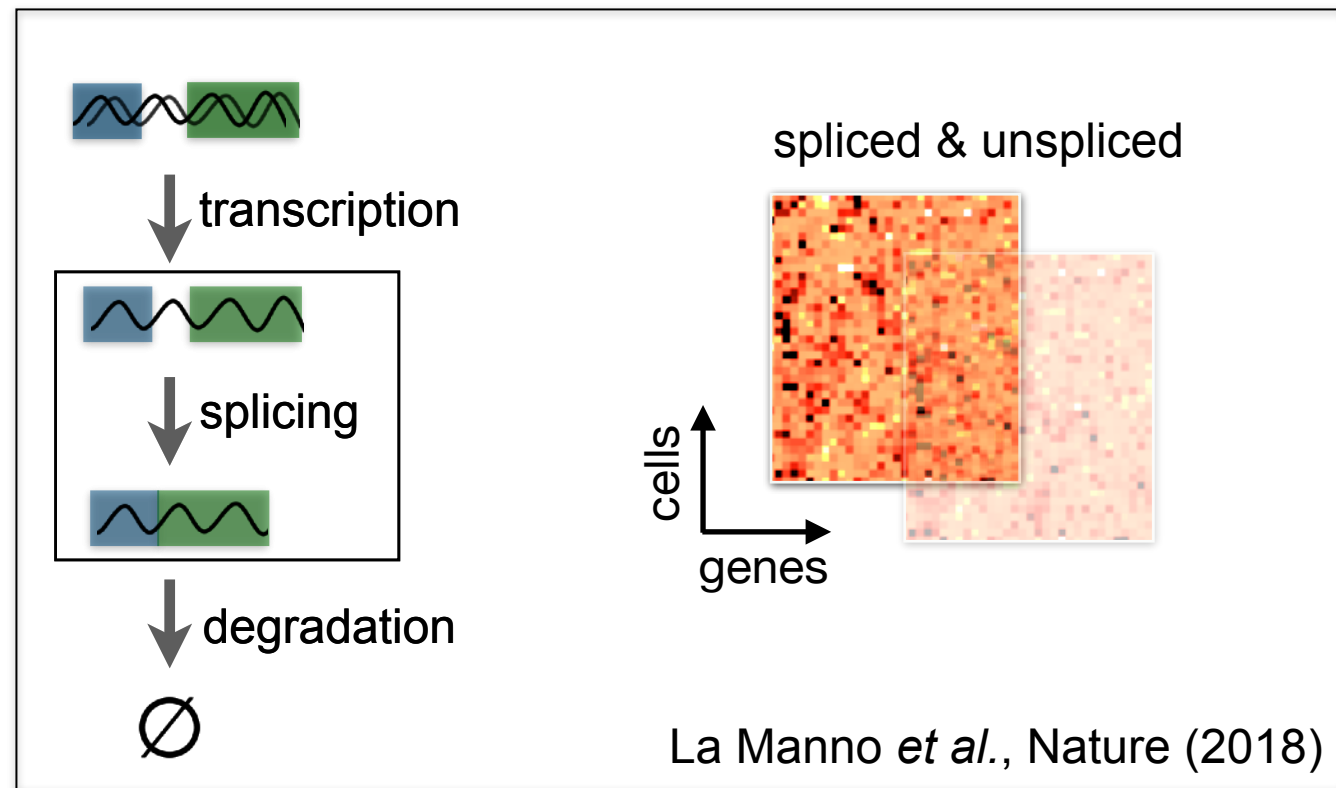
Lederer & La Manno (Nat. Biotech 2020)  
 Bergen *et al.* (MSB, 2021)

Modeling cellular dynamics with  
**RNA Velocity**

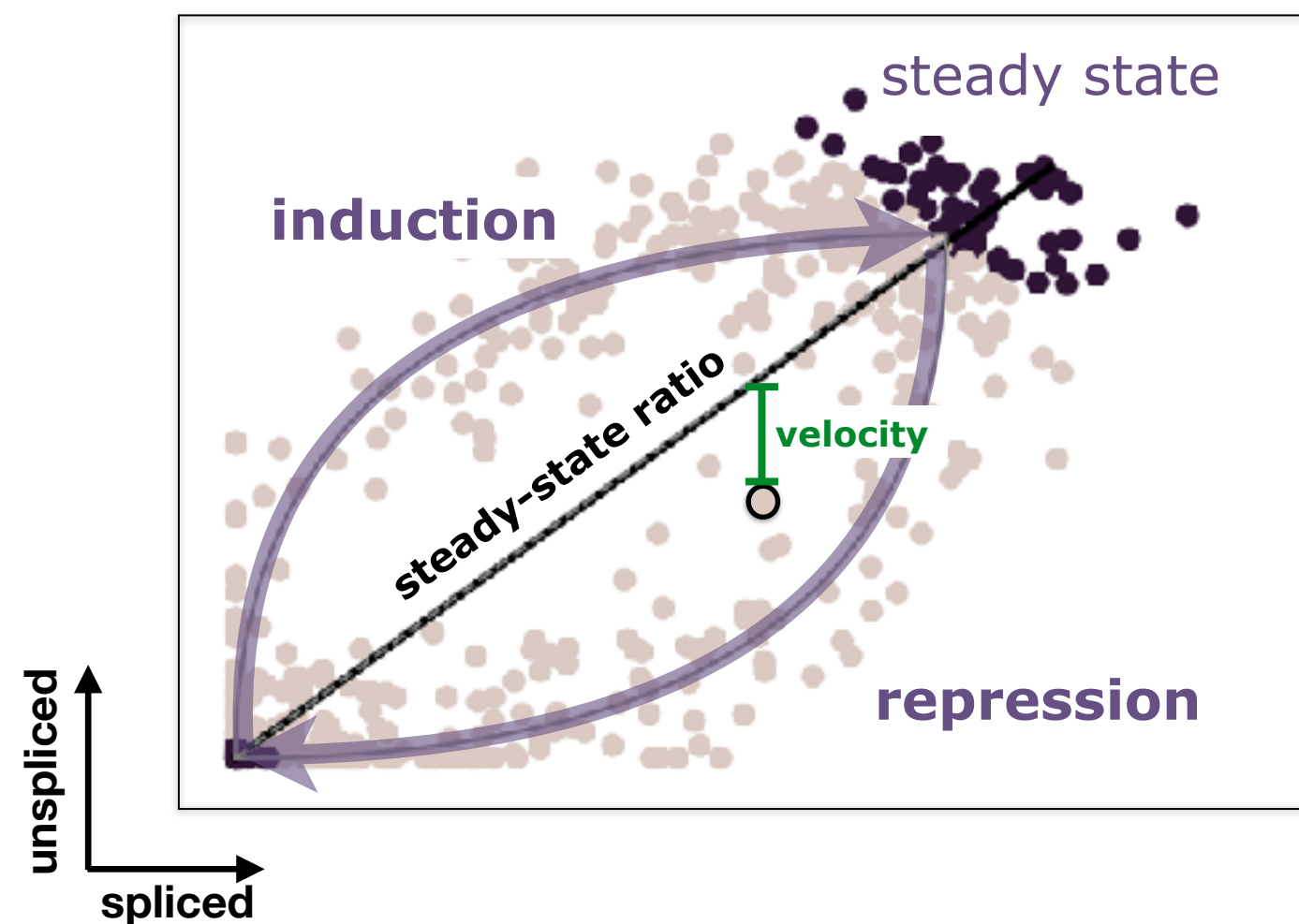
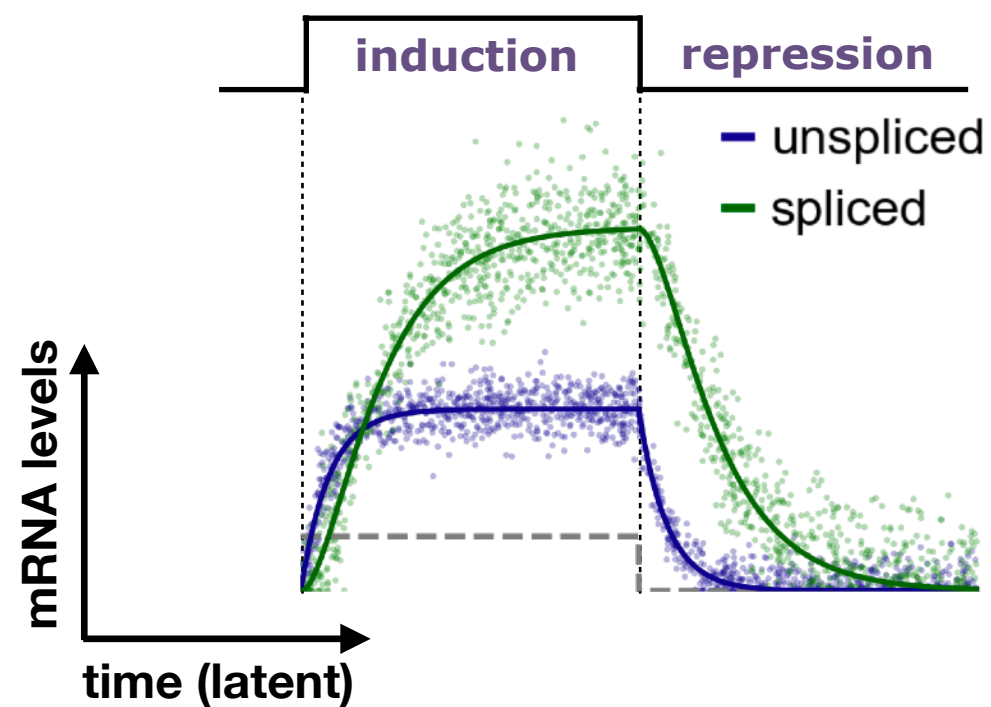
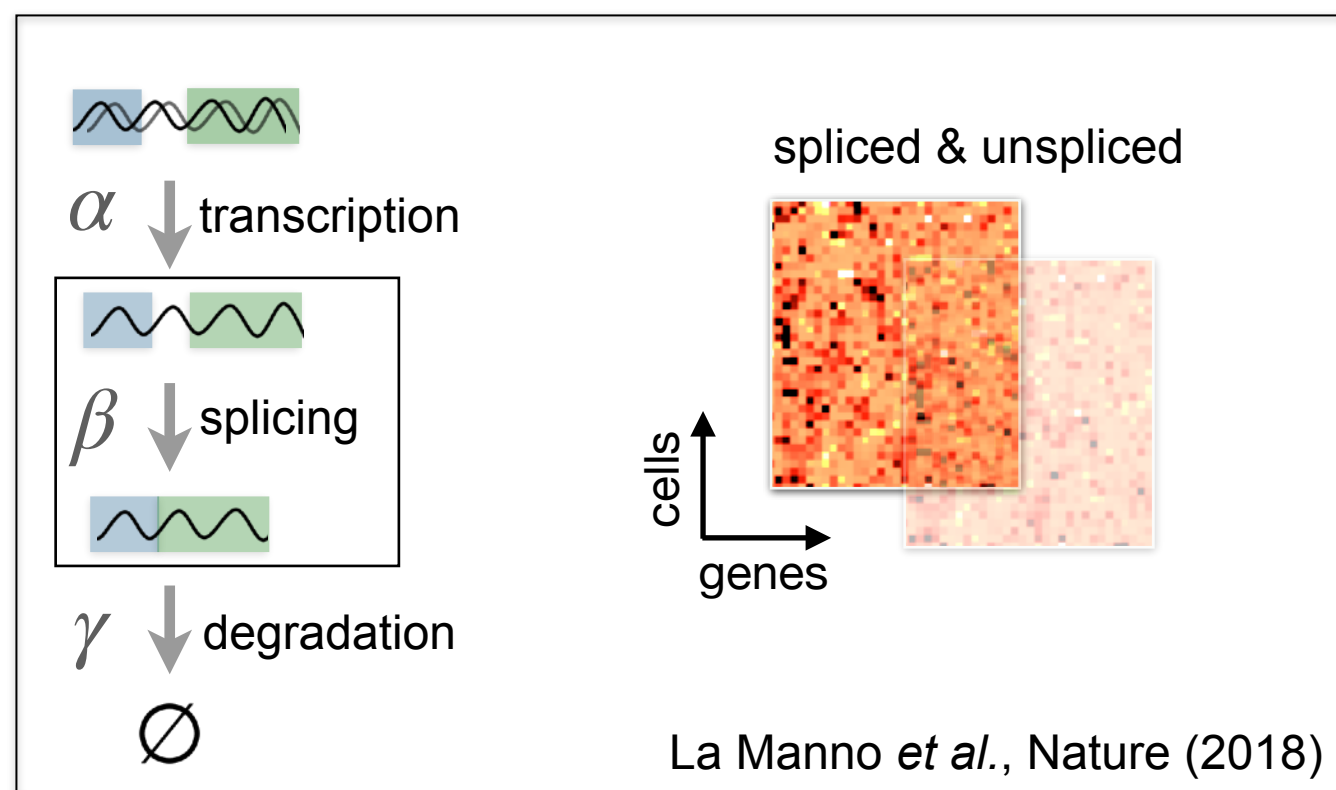




# Concept of RNA velocity

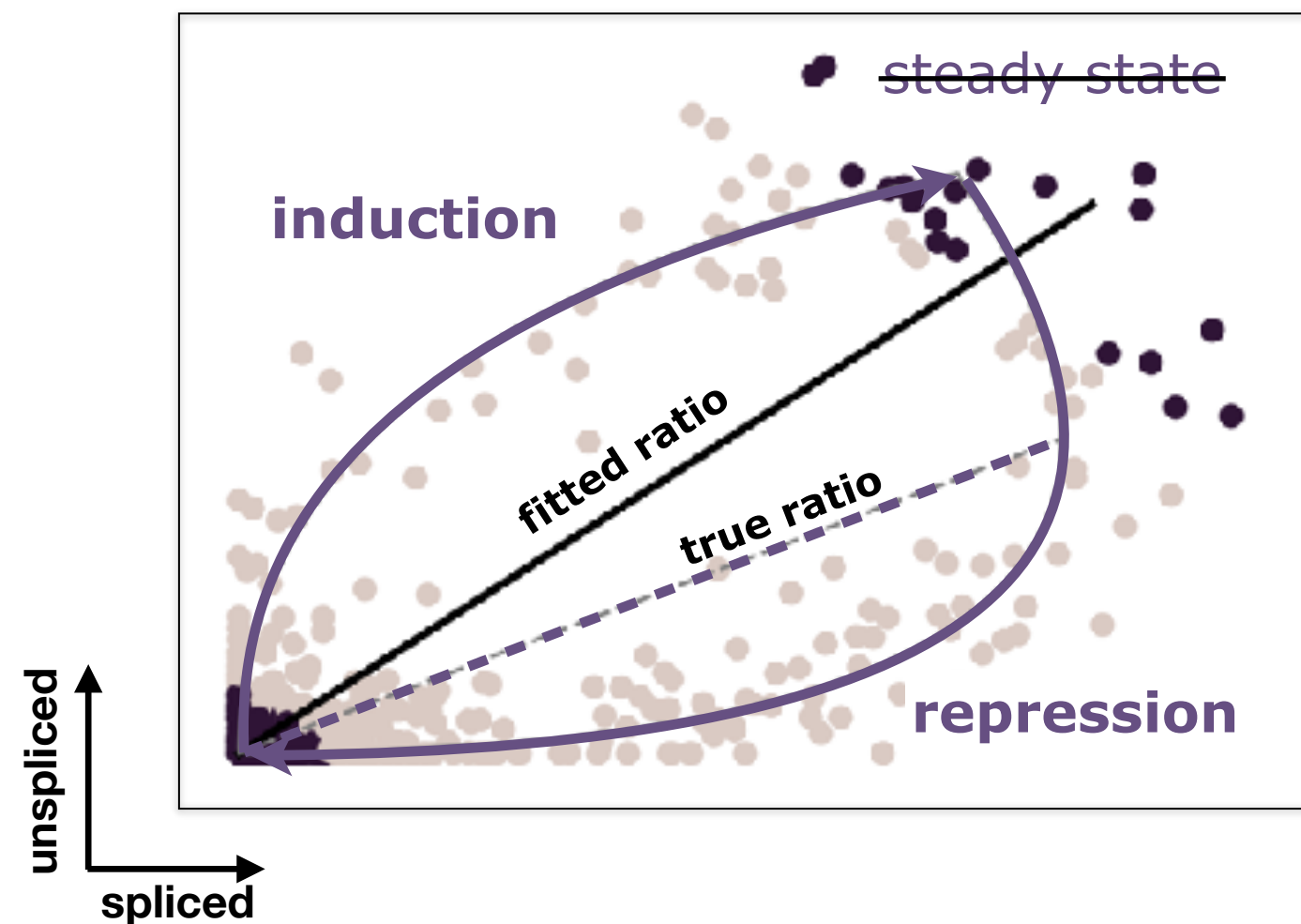
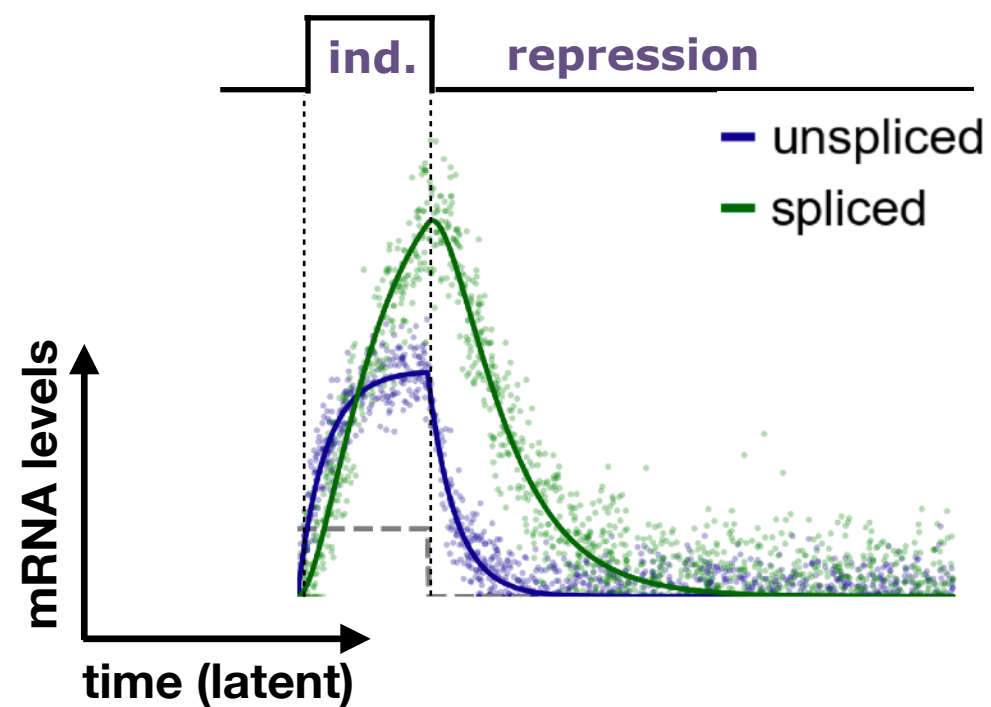
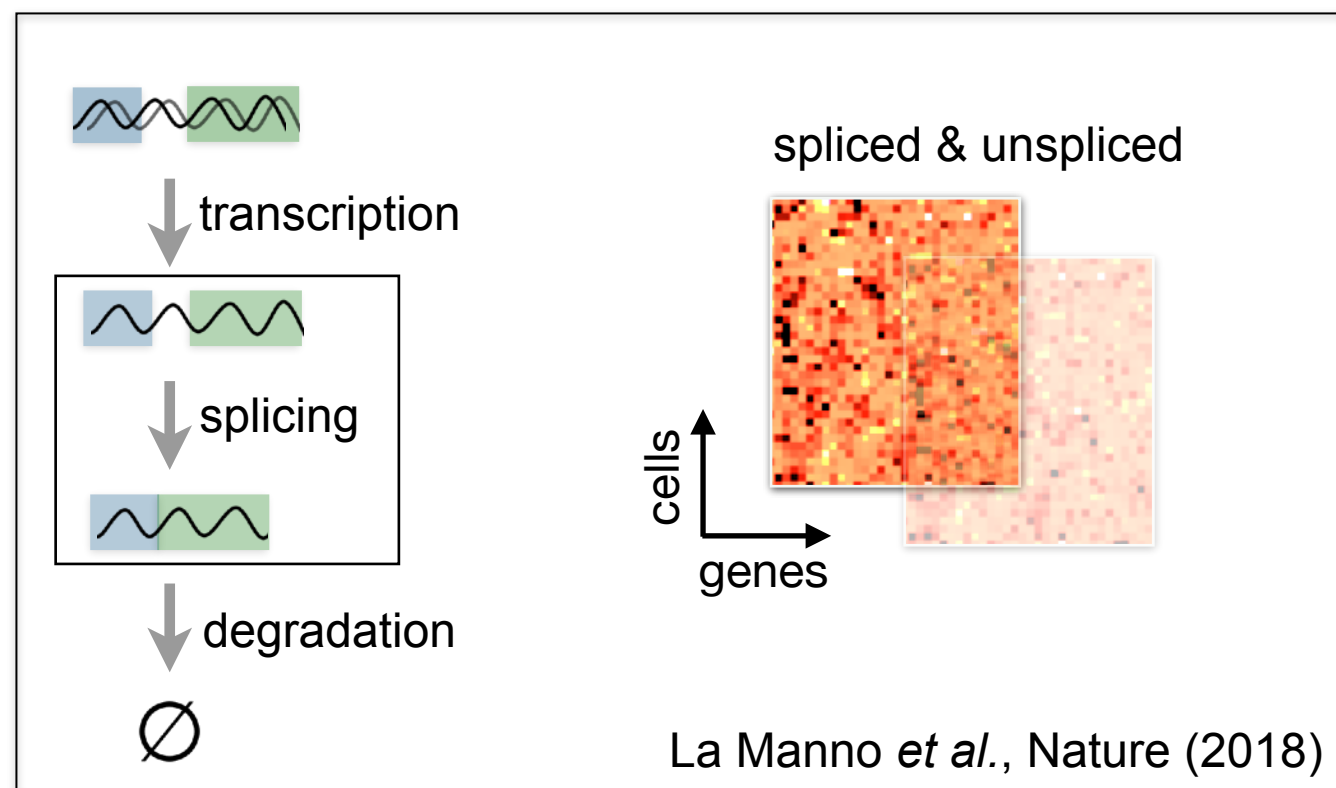


# Concept of RNA velocity



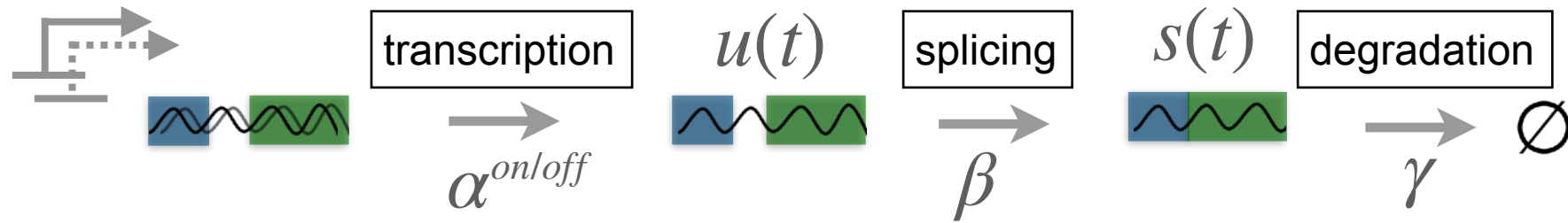
$$\frac{du(t)}{dt} = \alpha - \beta u(t), \quad \frac{ds(t)}{dt} = \beta u(t) - \gamma s(t)$$

# Concept of RNA velocity





# RNA velocity generalized through dynamical modeling

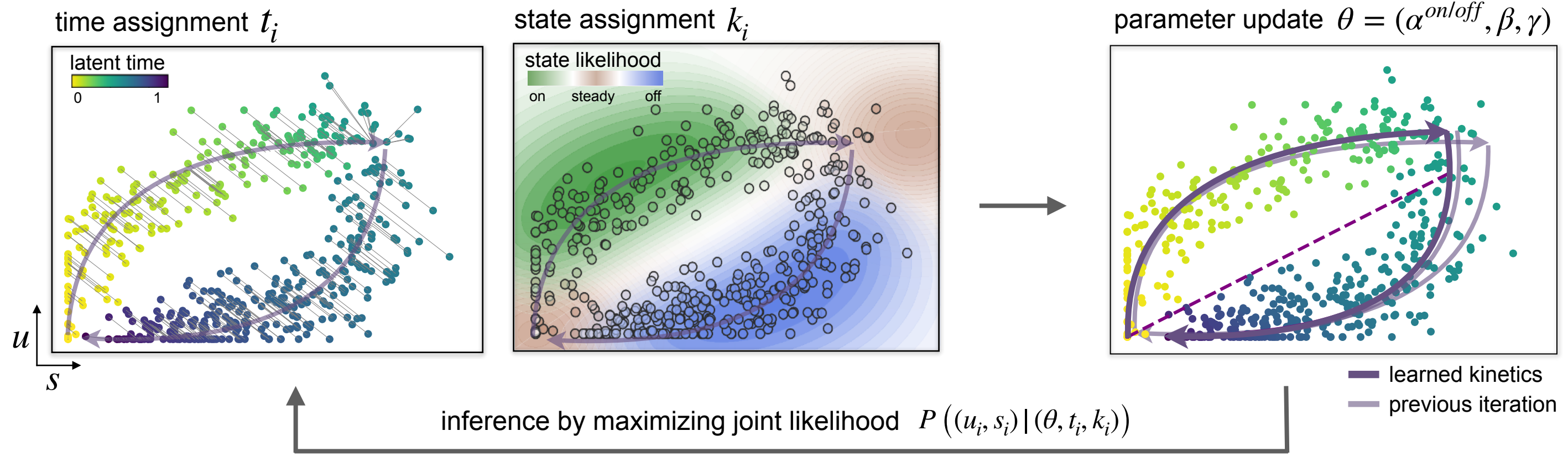


$$u(t) = u_0 e^{-\beta\tau} + \frac{\alpha}{\beta} (1 - e^{-\beta\tau})$$

$$s(t) = s_0 e^{-\gamma\tau} + \frac{\alpha}{\gamma} (1 - e^{-\gamma\tau}) + \frac{\alpha - \beta u_0}{\gamma - \beta} (e^{-\gamma\tau} - e^{-\beta\tau}) \quad \tau = t - t_0$$

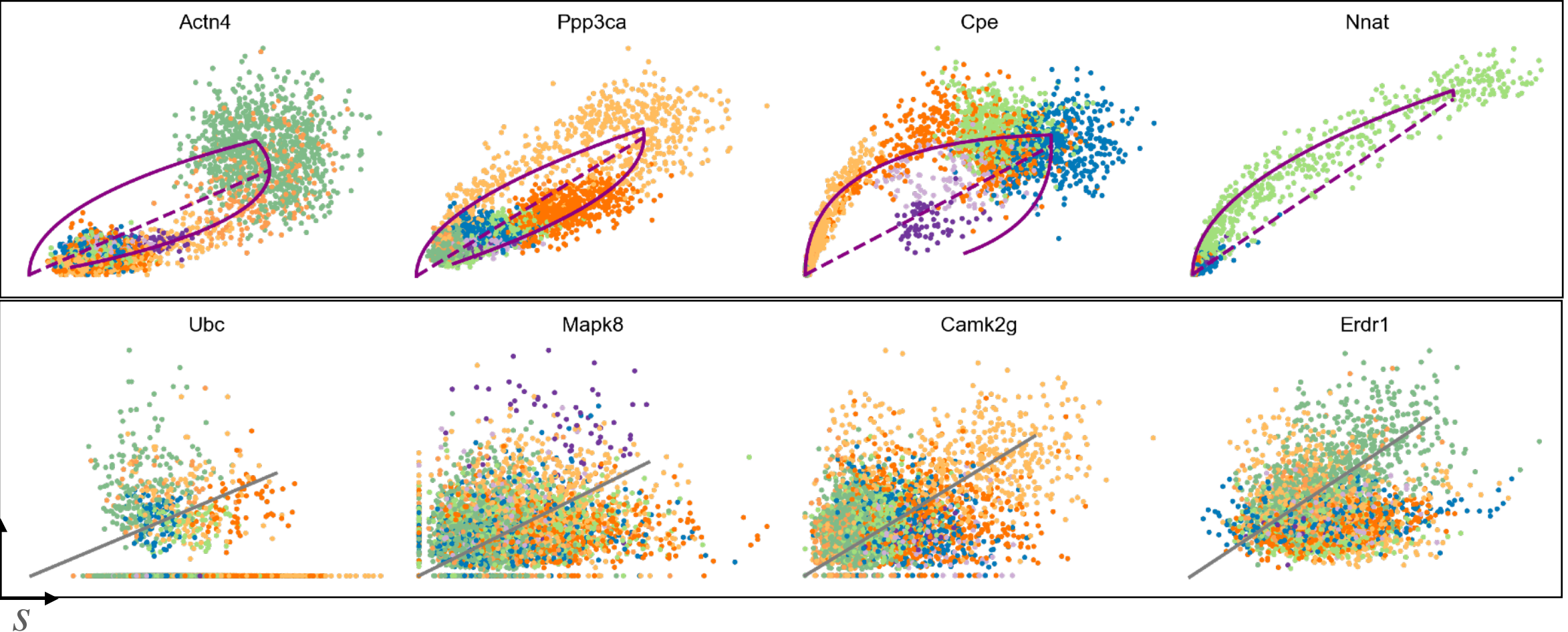
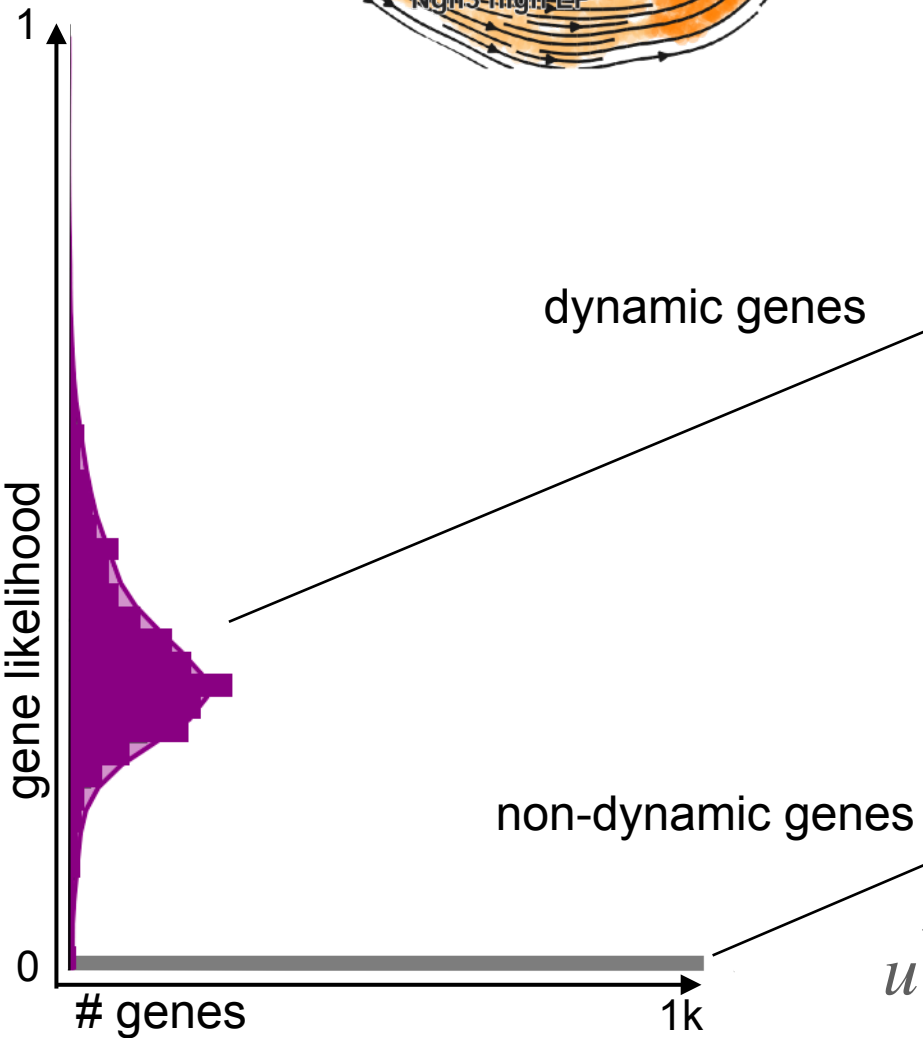
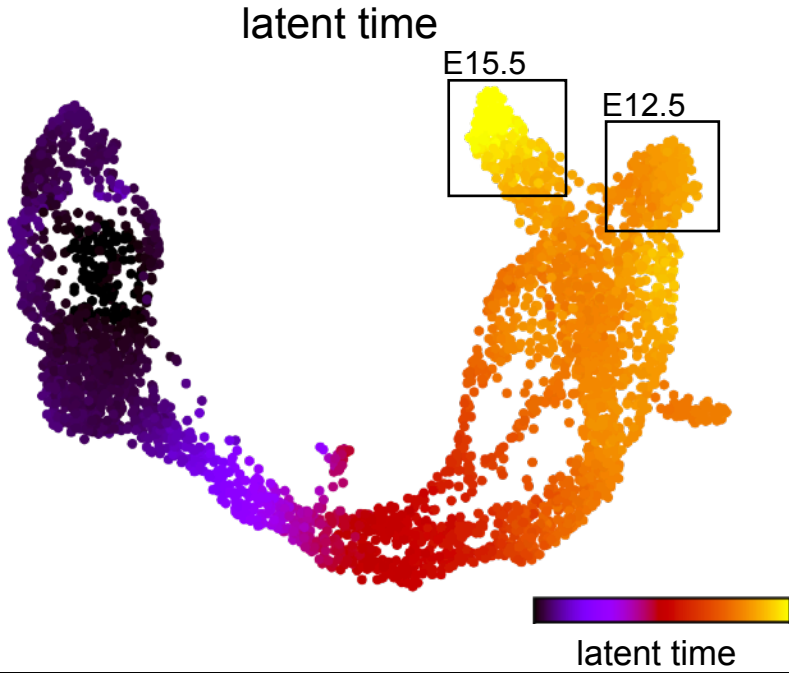
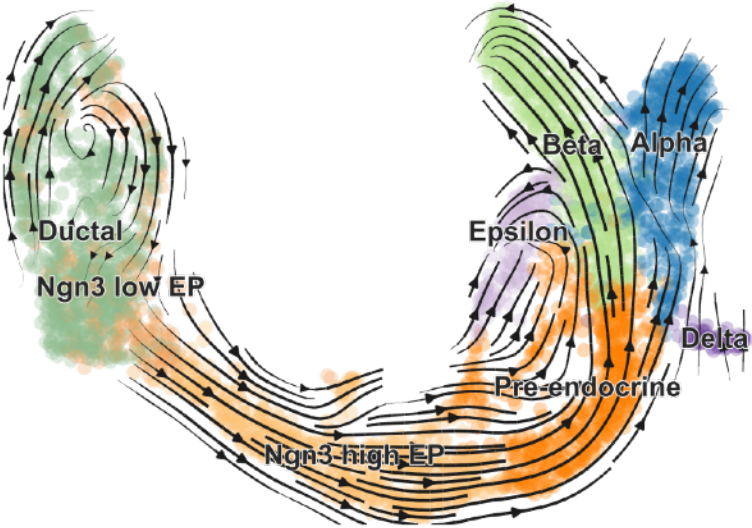
parameters of reaction rates  $\theta = (\alpha^{off}, \alpha^{on}, \beta, \gamma)$

cell-specific latent variables (switch, time, state)  $\eta_i = (t_0^{(i)}, t_i, k_i)$



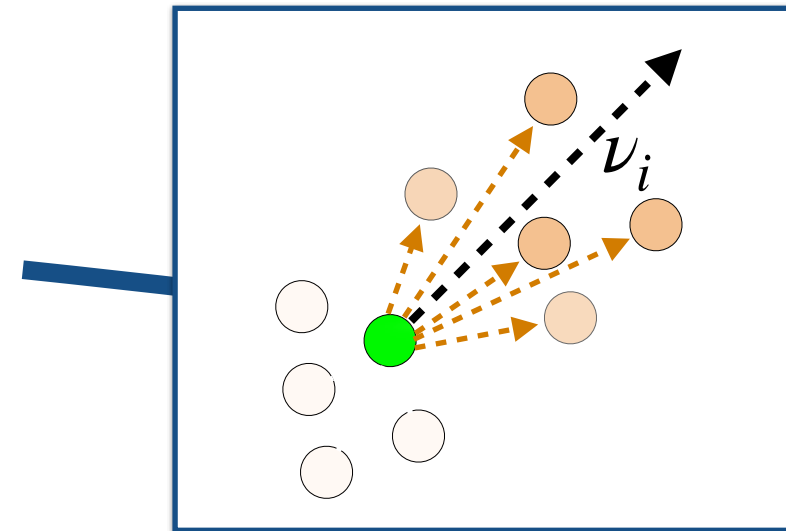
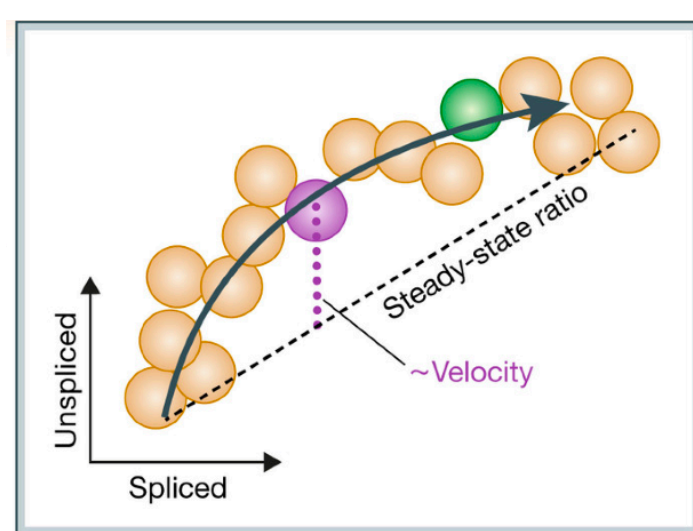
— learned kinetics  
- - previous iteration

# Applications of RNA velocity



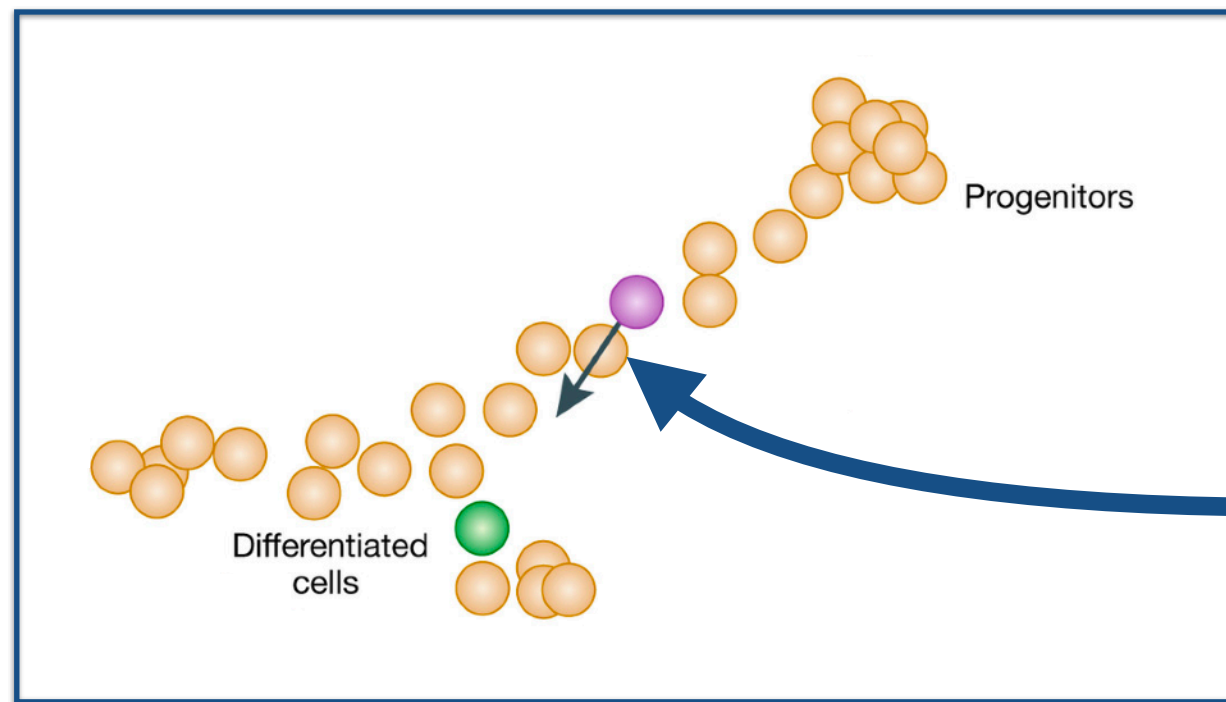
# RNA velocity in two dimensions

Compute cell-to-cell transition probabilities by how much the transition correlates with the velocity vector (high-dim)

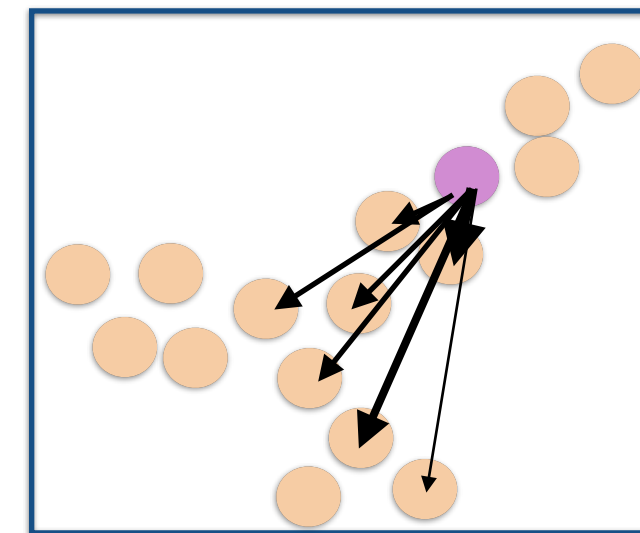


$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,j} & \dots & P_{1,S} \\ P_{2,1} & P_{2,2} & \dots & P_{2,j} & \dots & P_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \dots & P_{i,j} & \dots & P_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{S,1} & P_{S,2} & \dots & P_{S,j} & \dots & P_{S,S} \end{bmatrix}$$

$$P_{ij} = e^{\rho(\delta_{ij}, \nu_i)} / \sigma_i^2$$



UMAP / TSNE



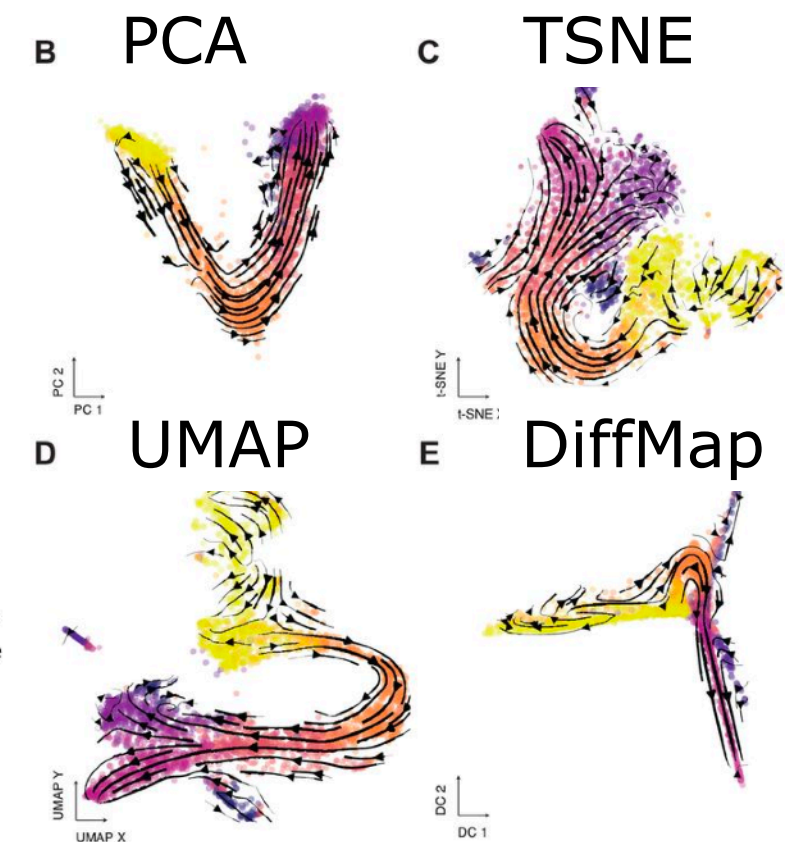
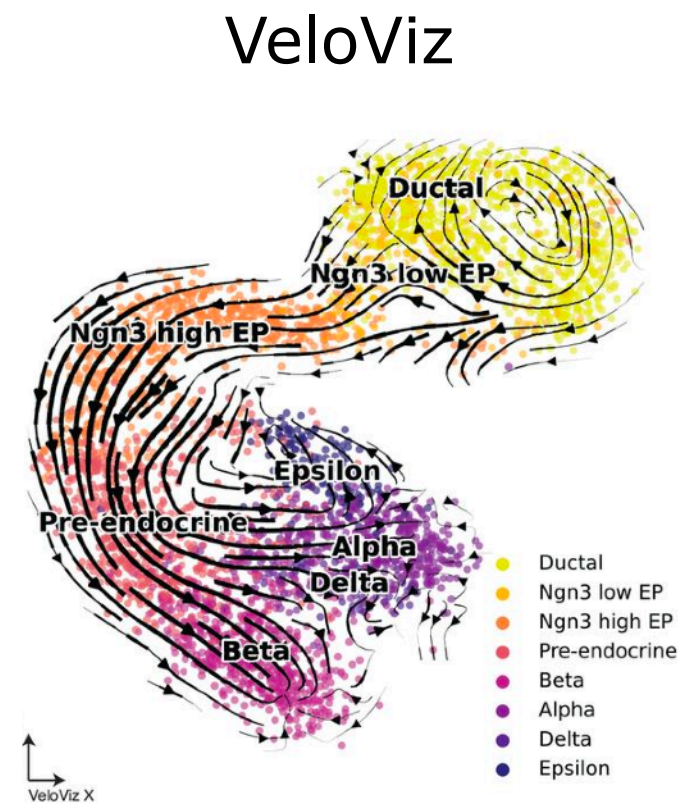
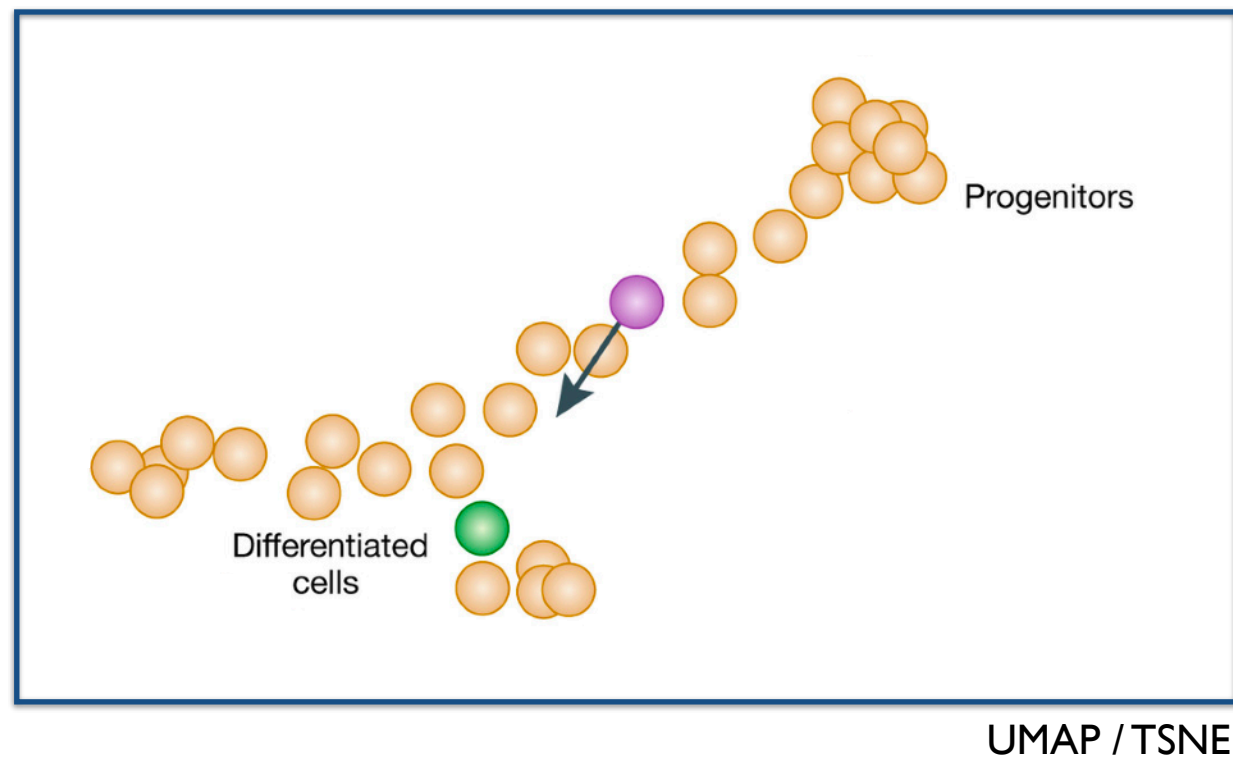
Given the transitions probabilities, compute the expected transition in the lower-dim embedding



# RNA velocity - finding a good representation

## Topic 1

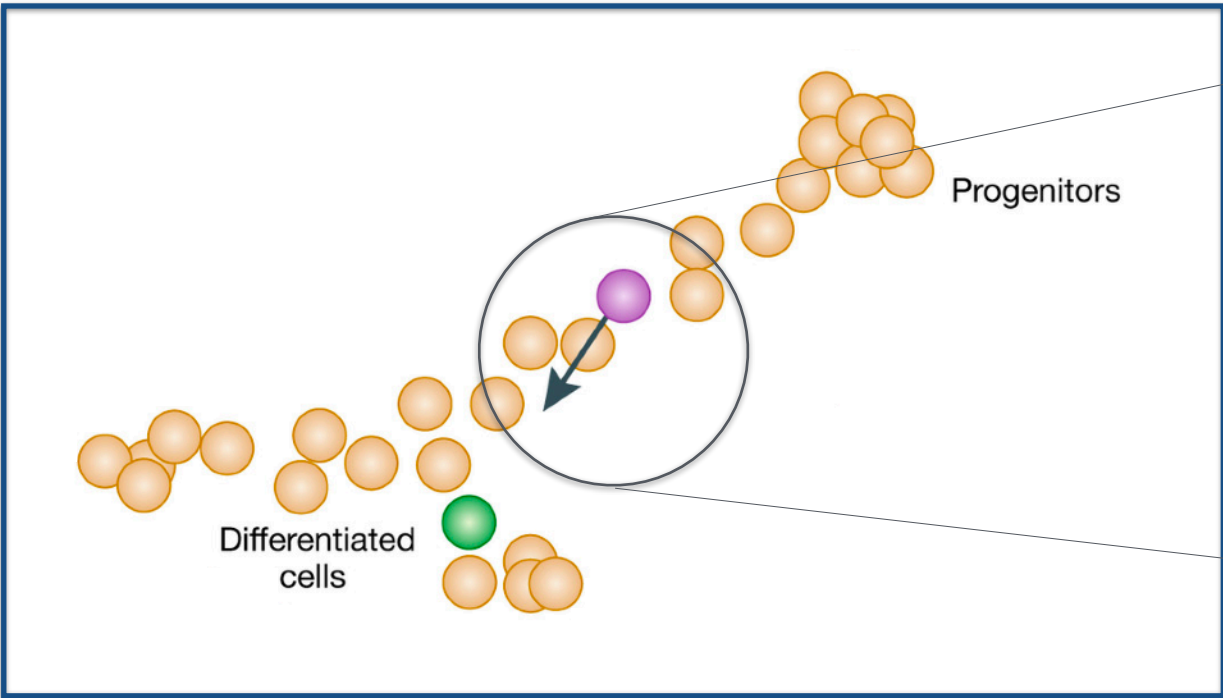
- What embedding specification best represents the high-dim vector field?
- Can we find other ways to project the data (e.g., parametric UMAP)?



# RNA velocity - identifying relevant genes

## Topic 2

- What genes are driving the projected arrows in the low-dim manifold?
- Can we systematically identify genes that are important in a particular compartment?



UMAP / TSNE

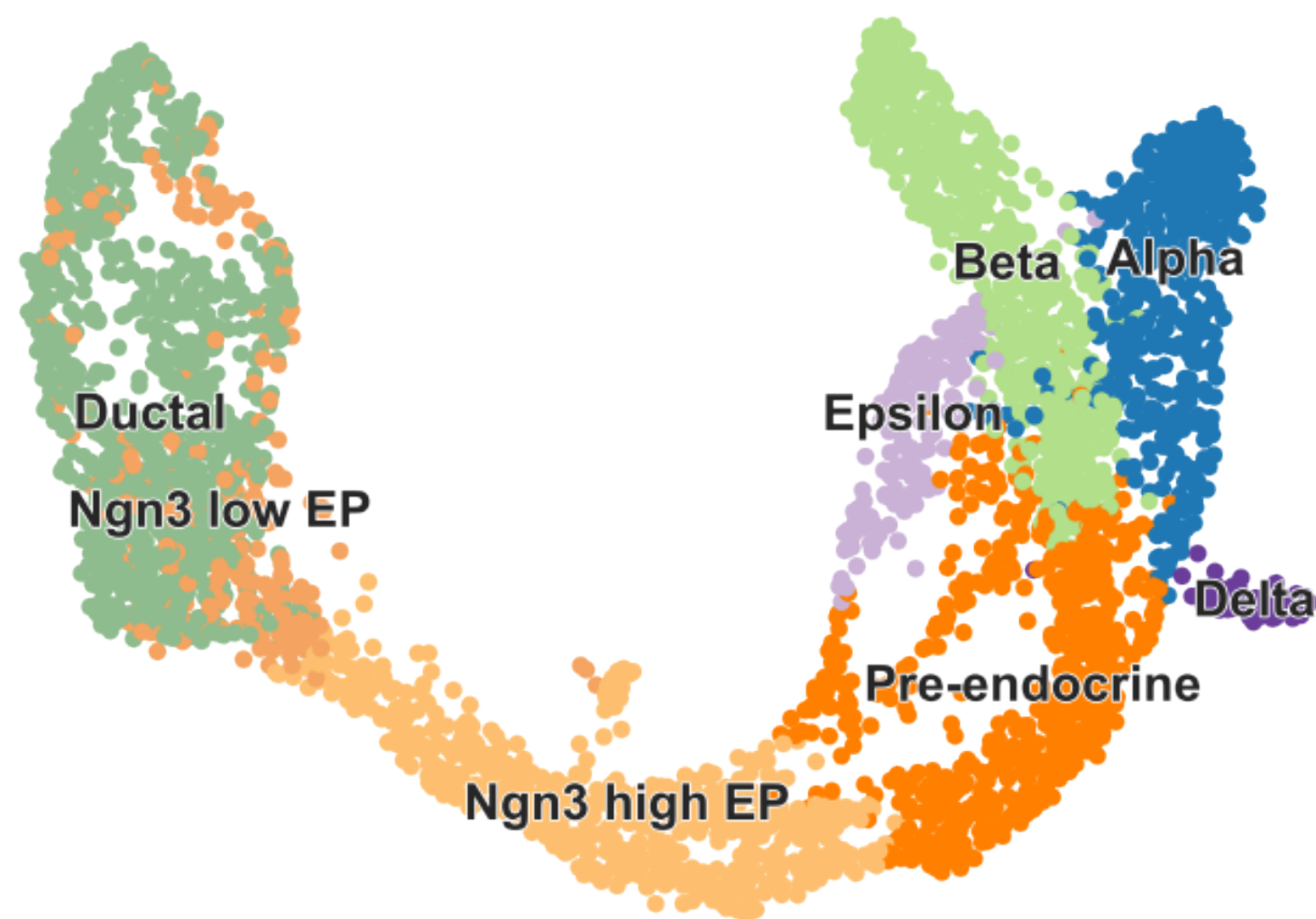
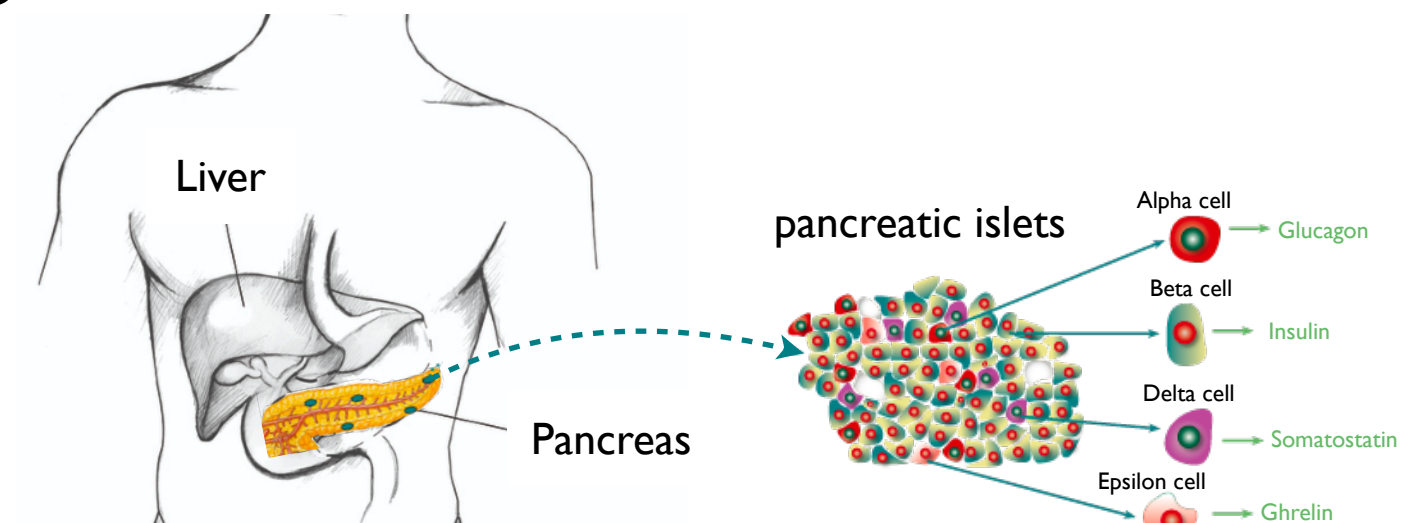
### Cell type A

Genes ranked	some metric
Gene XX	0.85
Gene XY	0.72
....	
Gene XZ	0.02

### Cell type B

Genes ranked	some metric
Gene AA	0.91
Gene AB	0.83
....	
Gene AZ	0.05

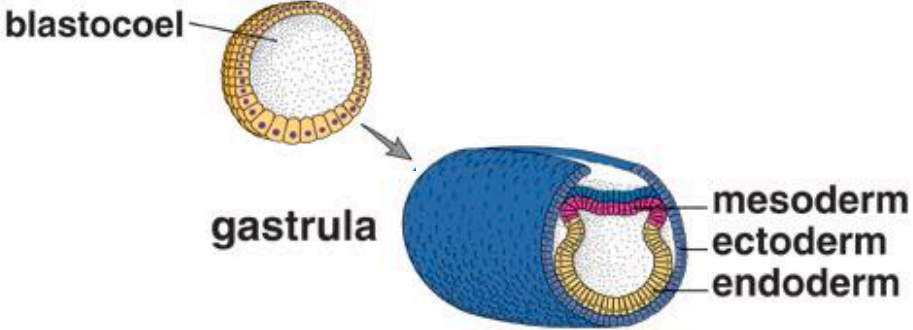
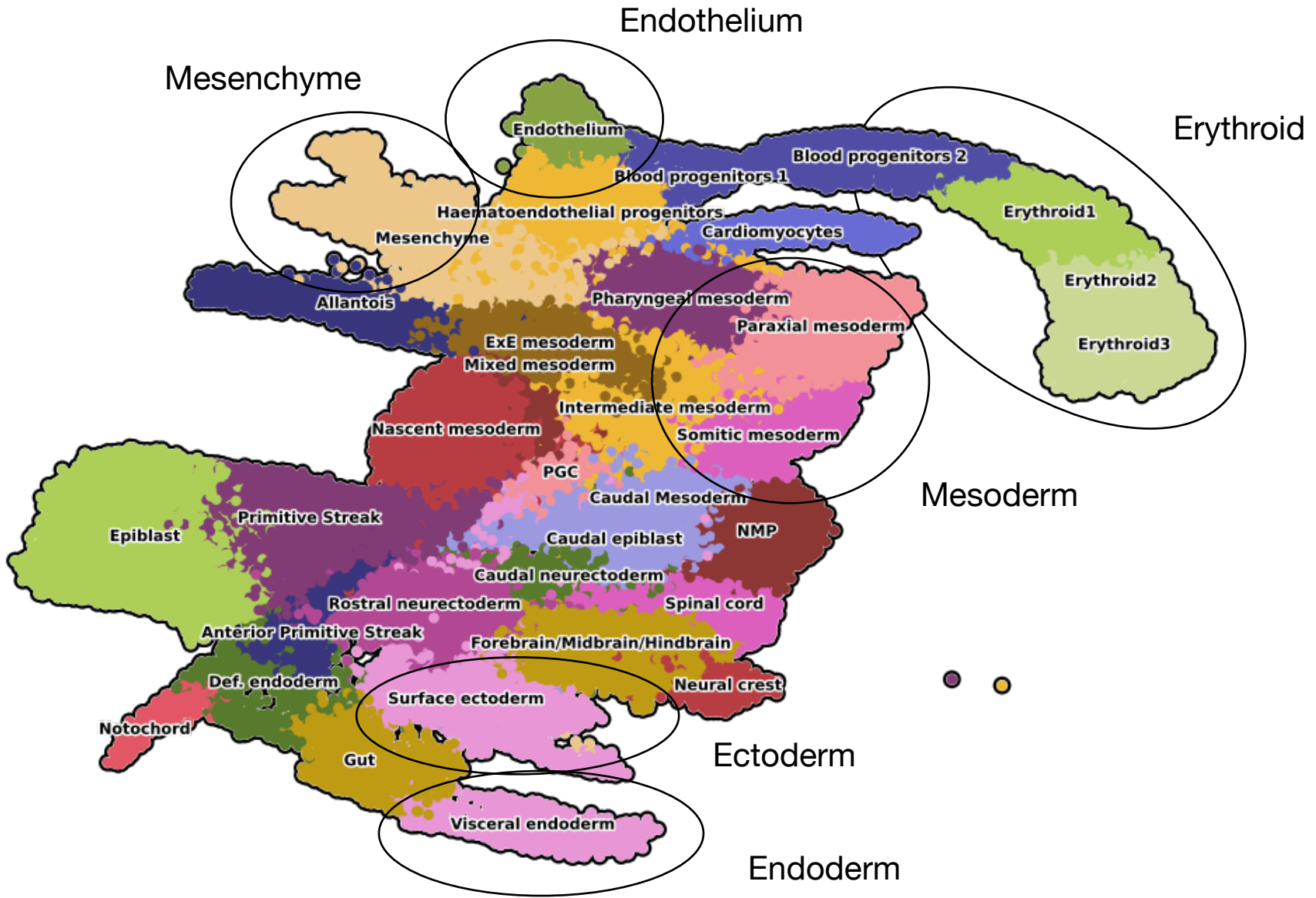
# Datasets - pancreatic endocrinogenesis



Bastidas-Ponce *et al.*, Development (2019)  
 Bergen *et al.*, Nat. Biotech (2020)



# Datasets - gastrulation and early organogenesis



# The scvelo API

```
import scvelo as scv
```

## Read the data

```
adata = scv.datasets.pancreas()
```

## Preprocessing

```
scv.pp.filter_and_normalize(adata, n_top_genes=2000)
```

```
scv.pp.moments(adata)
```

## Velocity estimation

```
scv.tl.velocity(adata)
```

## Velocity projection

```
scv.tl.velocity_embedding(adata, basis='umap')
```

## Visualization

```
scv.pl.velocity_embedding_stream(adata, basis='umap')
```

The screenshot shows the scvelo.org API documentation website. The page is titled "Utils" and lists various utility functions. The functions are organized into categories: "Get data by key", "Get gene info", "Data preparation", "Getters", "Converters", and "Least squares and correlation".

The "Get gene info" section contains the following functions:

Function	Description
<code>utils.gene_info</code> (name[, fields])	Retrieve gene information from biothings

The "Getters" section contains the following functions:

Function	Description
<code>utils.get_moments</code> (adata[, layer, ...])	Computes moments for
<code>utils.get_transition_matrix</code> (adata[, vkey, ...])	Computes cell-to-cell t
<code>utils.get_cell_transitions</code> (adata[, ...])	Simulate cell transition
<code>utils.get_extrapolated_state</code> (adata[, vkey, ...])	Get extrapolated cell st

The "Converters" section contains the following functions:

Function	Description
<code>utils.convert_to_ensembl</code> ([gene_names])	Retrieve ensembl IDs
<code>utils.convert_to_gene_names</code> ([ensembl_names])	Retrieve gene names f

The "Least squares and correlation" section contains the following functions:

Function	Description
<code>utils.leastsq</code> (x, y[, fit_offset, perc, ...])	Solves least squares $X*b=Y$
<code>utils.vcorrcoef</code> (X, y[, mode, axis])	Pearsons/Spearman's correla
<code>utils.test_bimodality</code> (x[, bins, kde, plot])	Test for bimodal distribution