

Models for Dynamical Inference (Diffusion-Drift, Velocities, Optimal Transport)

LALEH HAGHVERDI

Final projects initiation

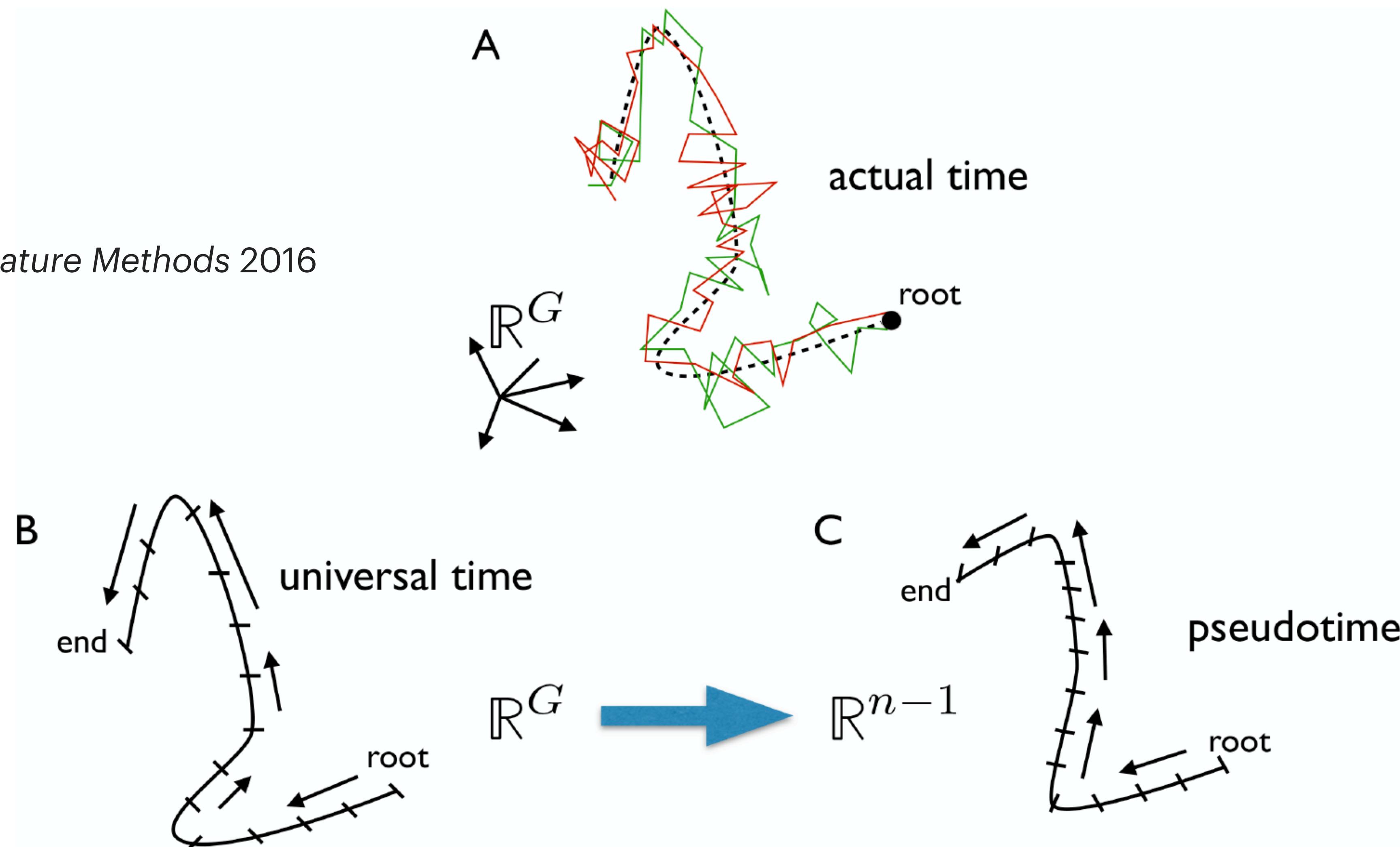
- Projects from topics until next session (25 Farvardin)
- Decide on your topic + group until 31 Farvardin or 2 Ordibehesht (on a google sheet?)
- 2 students (max. 3)
- 2 Additional sessions 3 & 10 Ordibehesht
- Each project ~10 min presentation discussion (& approval)
 - What question you want to address
 - Which data set (or simulation)
 - Theory part
 - Project plans: e.g. what plots you want to make?
 - Any progress so far?
- Submission deadline 1 Tir.

Grading system

- 10 points: (~5) Exercises and seminars
- 10 points: final project
 - Groups of 2-3 people
 - 5-6 pages report including Abstract, Intro, Methods, Results, Discussion
 - In the Abstract specify what problem you are addressing and based on which lecture(s)/exercise(s) it is
 - Share coding scripts and data (on GitHub) for reproducing your plots

How pseudotime is defined

Haghverdi et al. Nature Methods 2016
(Supplement)



Supplementary Figure N7: A) Two (red and green) actual time single cell trajectories in gene expression space(\mathbb{R}^G). Each jump on a trajectory happens in an (equidistant) unit of actual time. B) Universal time is defined as *arc length* on the data manifold starting from the root. This manifold $C \subset \mathbb{R}^G$ remains the same for several single cell trajectories, as well as for snapshot samples of single cells. C) Pseudotime (in

Diffusion maps

Coifman *et al.* PNAS 2005

row normalised transition matrix P (from the cells' pairwise Gaussian kernel matrix W)

$$P_{n \times n} = D^{-1}W = D^{-1/2}(D^{-1/2}WD^{-1/2})D^{1/2}$$

$$D_{ii} = \sum_j W_{ij}$$

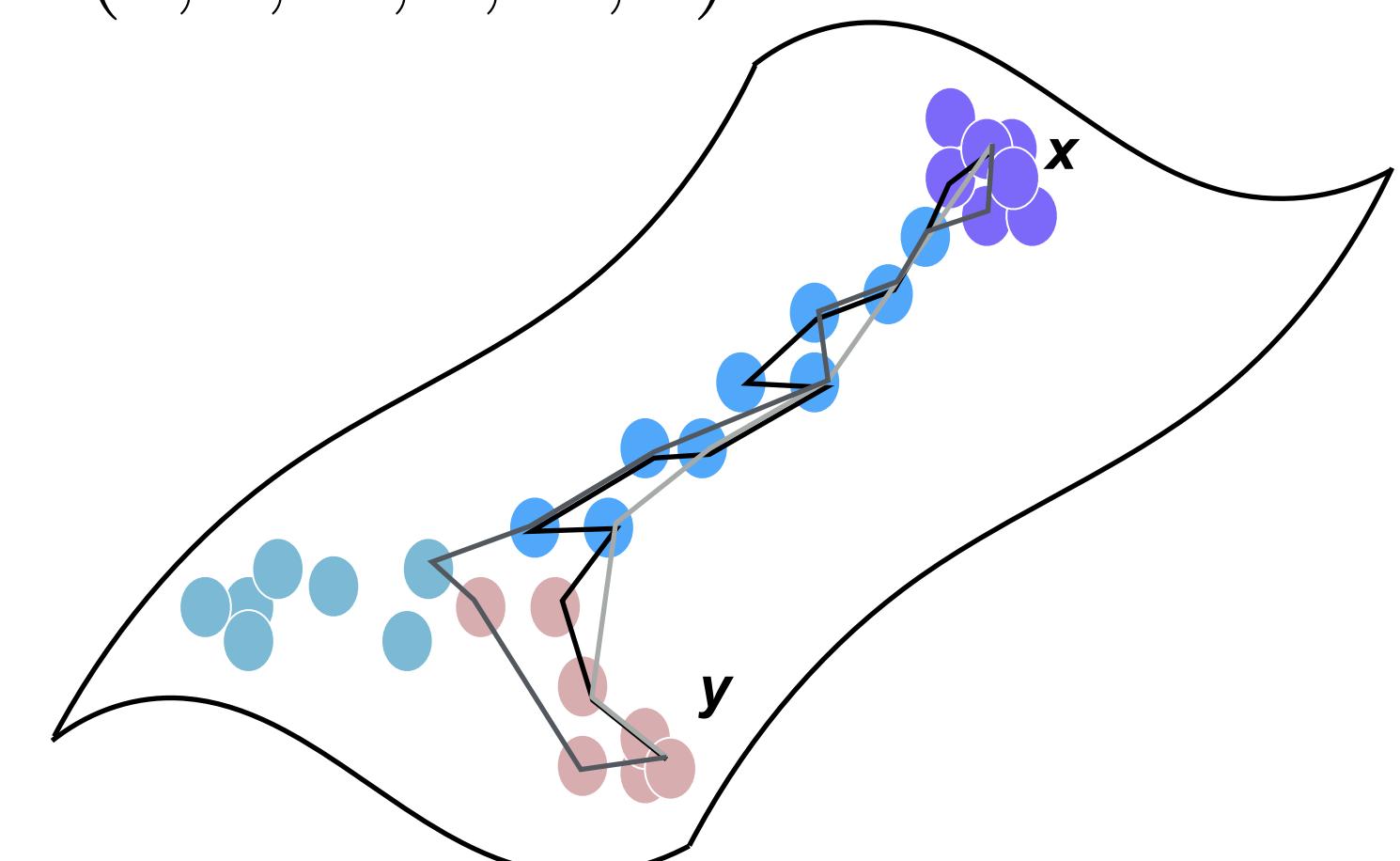
P^t transition probabilities in t steps

$$\phi_i P = \lambda_i \phi_i, \quad P \psi_i = \lambda_i \psi_i, \quad \lambda_i \leq 1$$

The state vector X can represent one cell (with only one nonzero value at position x) but can a distribution of cells (with more 1 enters)

$$\begin{aligned} D_t^2(x, y) &= \|P^t(x, \cdot) - P^t(y, \cdot)\|_{1/\phi_0}^2 \\ &= \sum_{i=1}^{n-1} \lambda_i^{2t} (\psi_i(x) - \psi_i(y))^2 \end{aligned}$$

For large t we can cut the sum at $i=K$ for low-dimensional approximation



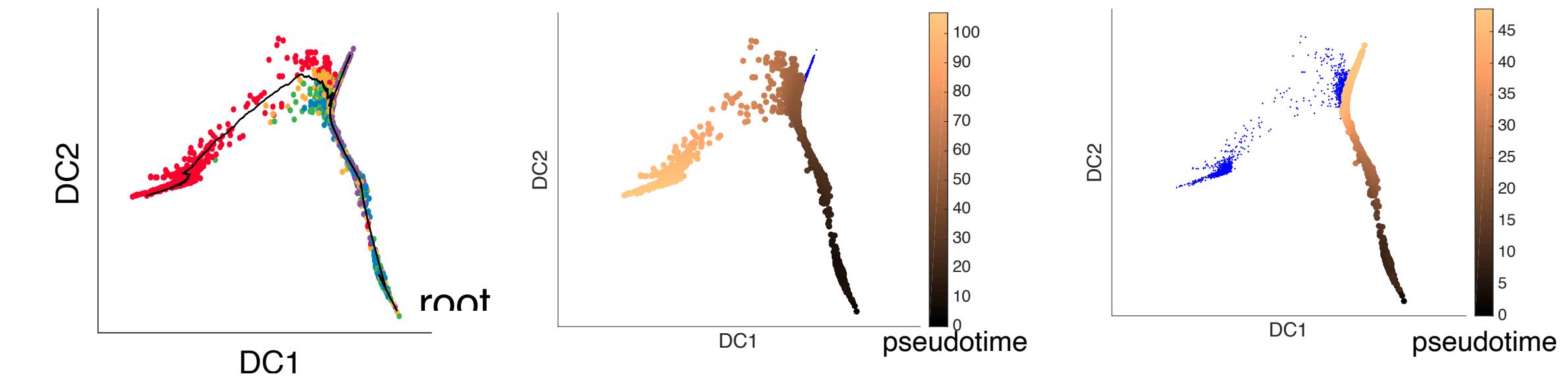
Diffusion pseudotime

$$T_{n \times n} = D^{-1/2} W D^{-1/2}$$

- "Diffusion pseudotime robustly reconstructs lineage branching"
Haghverdi et al. Nature methods 2016
- "Geometric diffusions for reconstruction of cell differentiation dynamics". Haghverdi, L., Diss. Technische Universität München, 2016.

Symmetric transition matrix

T^t Transition probabilities in t steps



$$\phi_i T = \lambda_i \phi_i, \quad T \phi_i = \lambda_i \phi_i, \quad \lambda_i \leq 1$$

$$\tilde{T} = T - \phi_0 \phi_0^T$$

Remove the stationary state

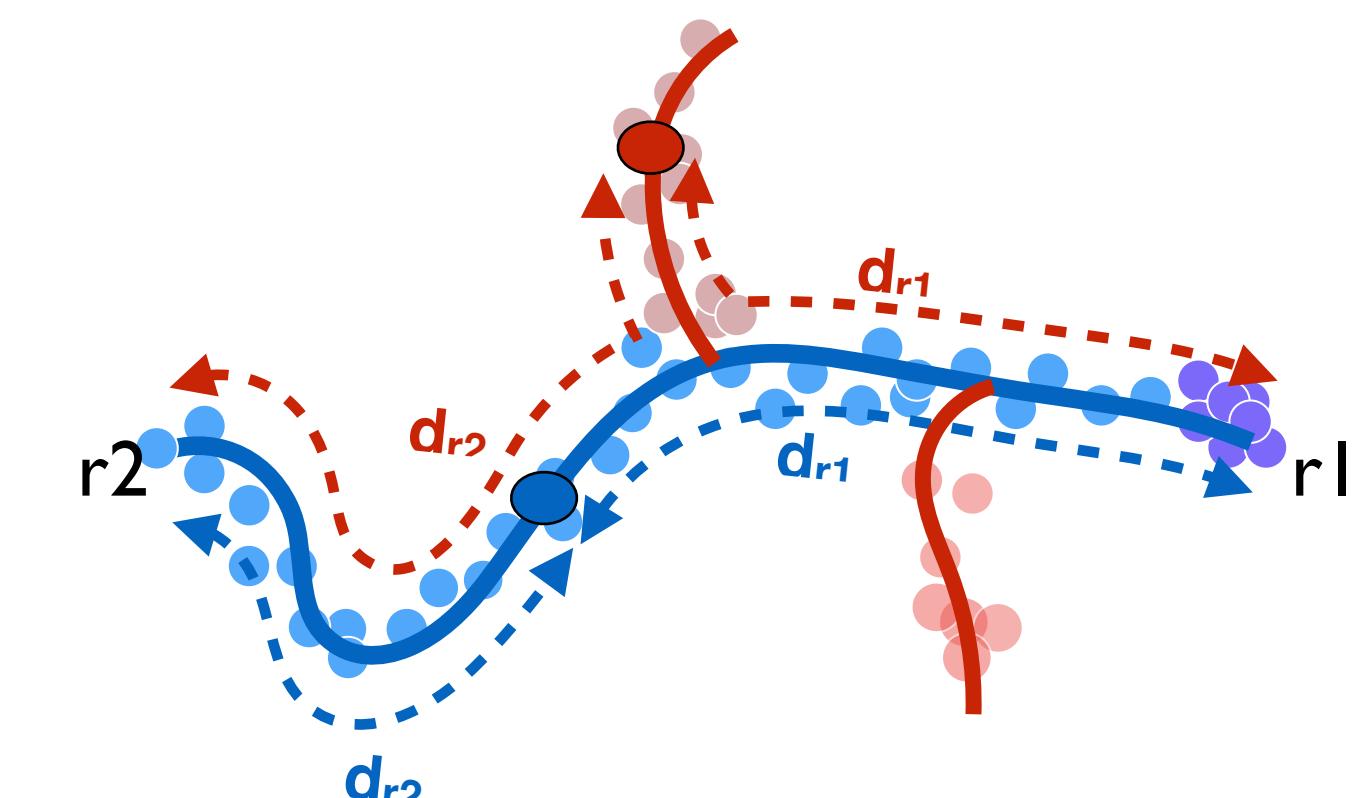
$$\begin{aligned} M &= \sum_{t=1}^{\infty} \tilde{T}^t \\ &= (I - T + \phi_0 \phi_0^T)^{-1} - I \end{aligned}$$

$$dpt_t^2(x, y) = \|M(x, \cdot) - M(y, \cdot)\|^2 = XMX^T + YMY^T - XMY^T - YMX^T$$

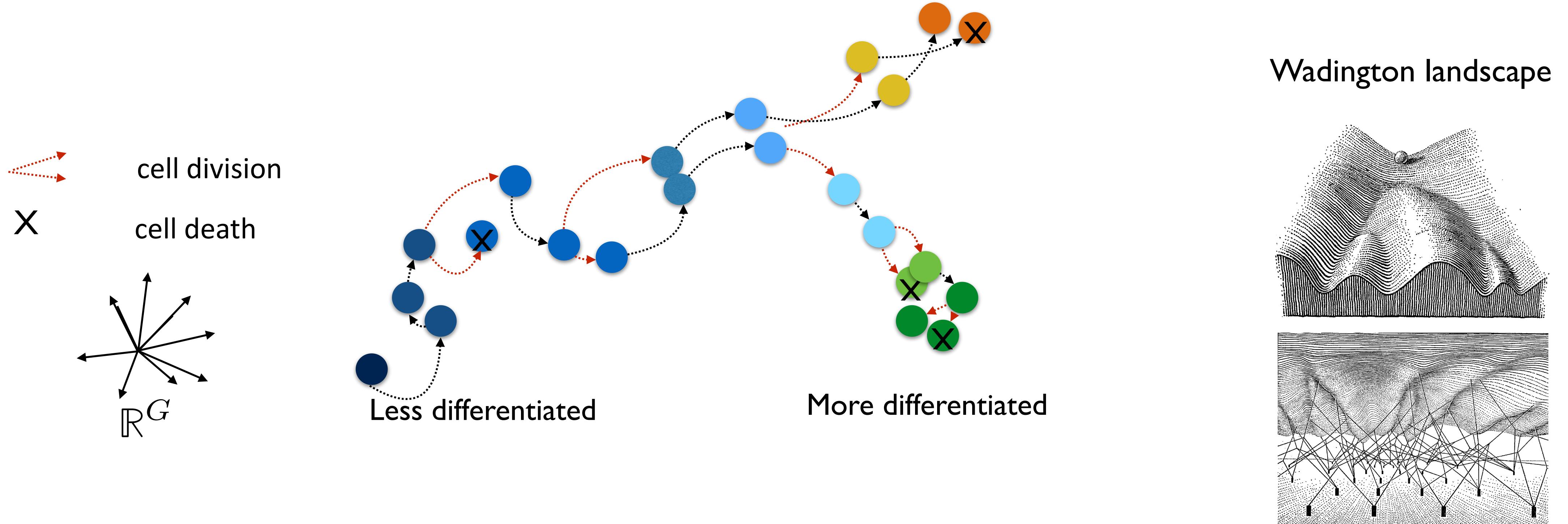
$$= \sum_{i=1}^{n-1} \left(\frac{\lambda_i}{1 - \lambda_i} \right)^2 (\phi_i(x) - \phi_i(y))^2$$

Exercises:

- a) Can we terminate the summation in the previous page for an approximation and calculation of dpt only based on the first eigenvectors of T? Would this have any advantages? Any disadvantages?
- b) Haghverdi et al. in Nature methods 2016 used a symmetric version of the transition matrix for ease of calculations. Also because they assumed they should not worry much about exact conservation of probabilities in cell differentiation process with cell birth and death events. Is this a good argument for abandoning the row-normalised (but assymmetric) P matrix? Does this introduce any disadvantages? Which matrix (P or T) produces better data visualisations?
- c) Show that both P and T have real positive eigenvalues.
- d) What would be the problem with a data mapping (DR) based on eigenvectors with imaginary or negative eigenvalues?
- e) Is dpt really a geodesic distance? Could goedesics calculation on a diffusion map of data result in better pseudotime and branch identification?



Diffusion-drift model



- Describes the temporal propagation of the probability density

$$\frac{\partial}{\partial t} \underline{p(s,t)} = \nabla \cdot \left(\underline{\nabla D(s)p(s,t)} + \underline{p(s,t)\nabla U(s)} + \underline{\nabla S(s,t)p(s,t)} \right)$$

Probability density at
position s , time t

Diffusion coefficient
(Position dependent)

Drift force (Position
dependent)

Birth/death
(Position dependent)

Diffusion-drift relation with cell state velocities

- "Diffusion pseudotime robustly reconstructs lineage branching"
Haghverdi et al. Stem cell Reports 2023 (Supplemental Note 1)

$$\frac{\partial}{\partial t} p(s, t) = \nabla \cdot \left(\nabla D(s)p(s, t) + p(s, t)\nabla U(s) + \nabla S(s, t)p(s, t) \right) \quad (1)$$

We can rewrite equation 1 as:

- J: probability flux $\frac{\partial}{\partial t} p(s, t) = \nabla \cdot \vec{J}(s, t) \quad (10)$

$$J(s, t) = \nabla D(s)p(s, t) + p(s, t)\nabla U(s) + \nabla S(s, t)p(s, t)$$

- V: Cell state velocity $= \vec{V}(s)p(s, t) + \nabla S(s, t)p(s, t) \quad (11)$

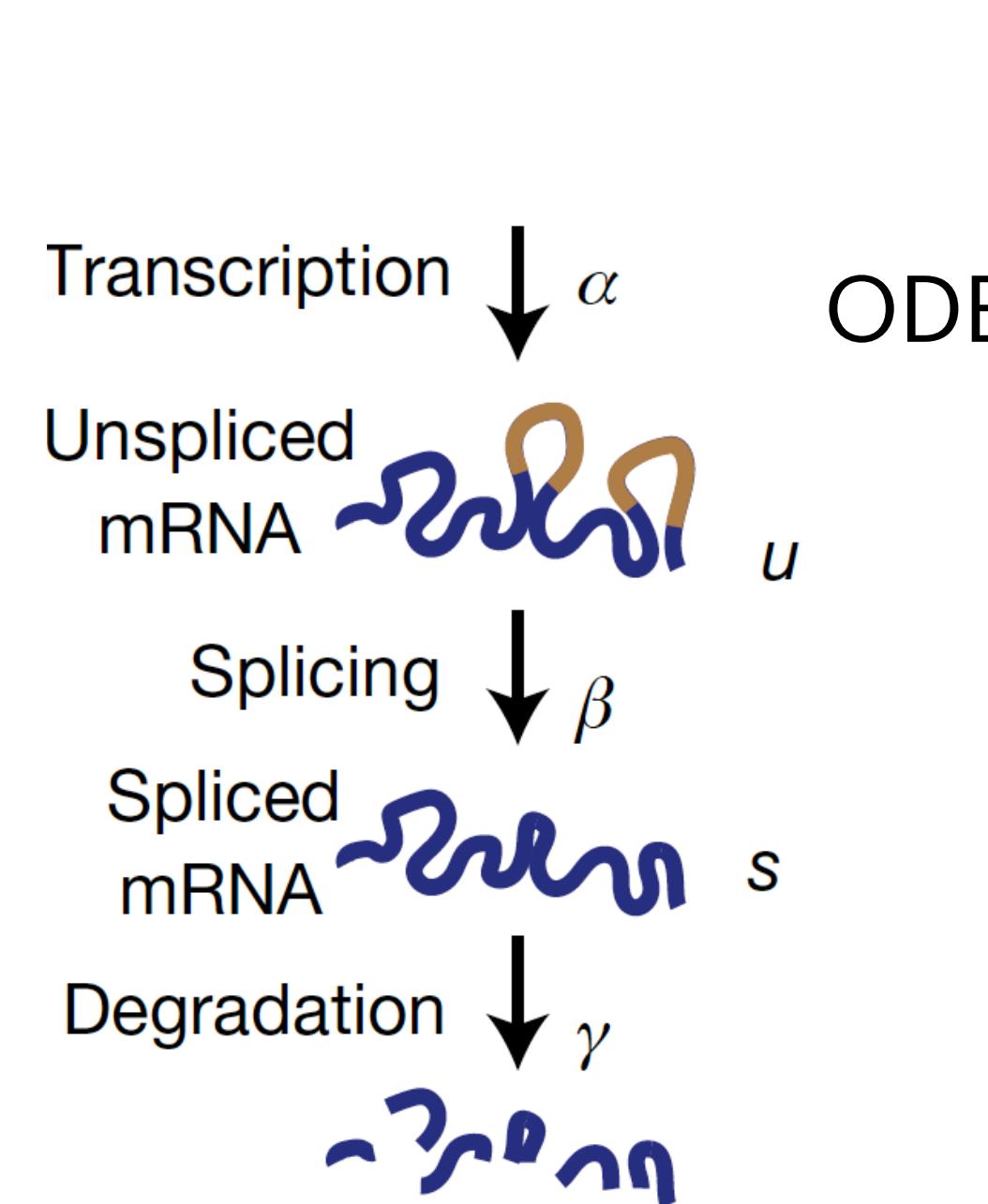
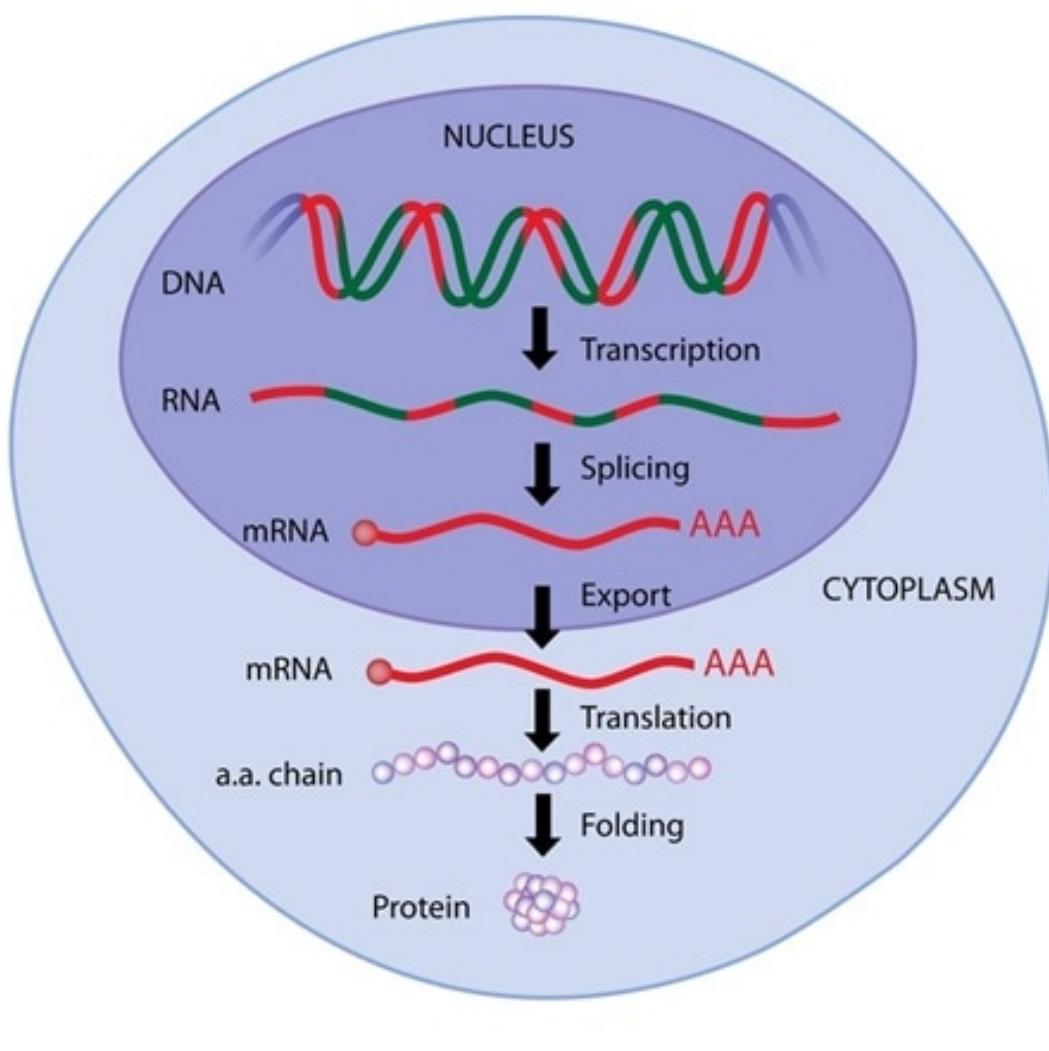
$$\vec{V}(s) = \nabla D(s) + \nabla U(s) \quad (12)$$

Equation 12 is also known as "Langevin equation" in statistical physics literature for Brownian motion.

RNA velocity

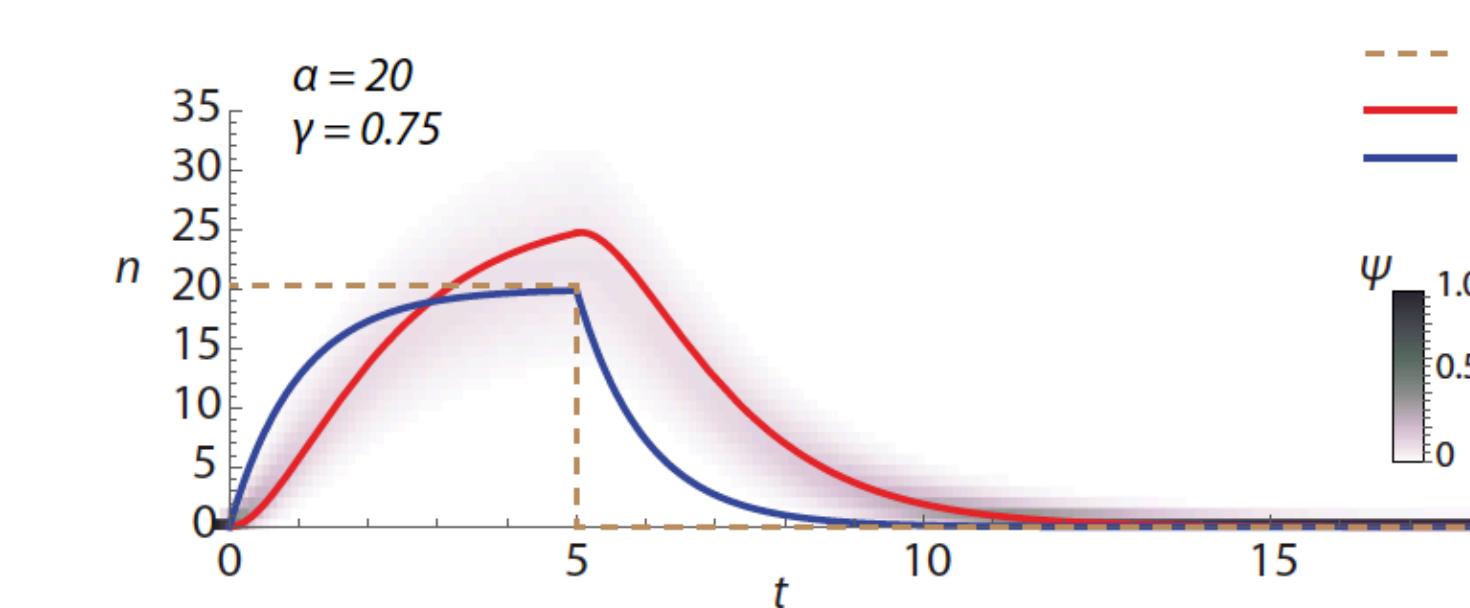
Cell state velocities

La Manno et al,
Nature 2018

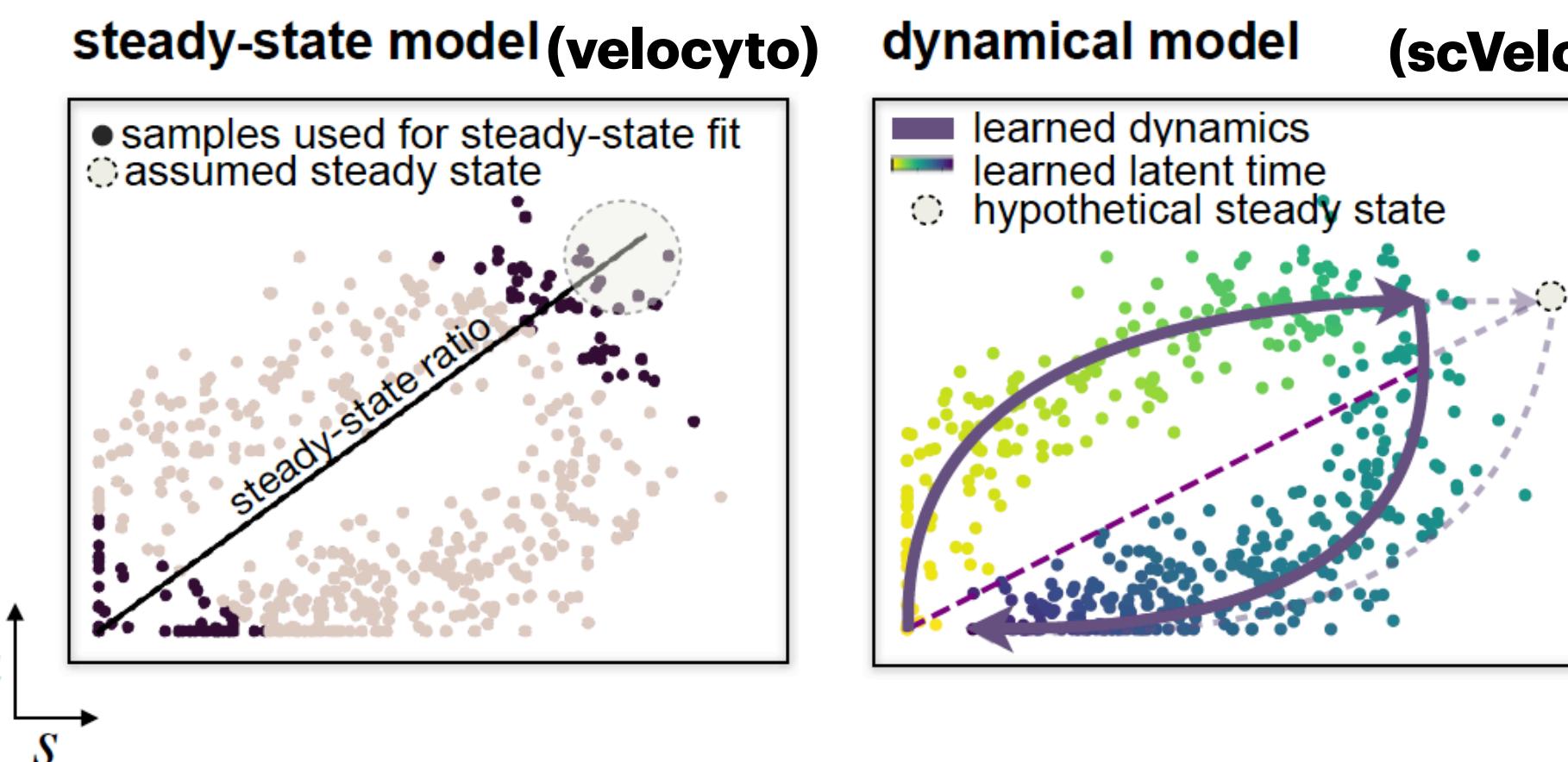
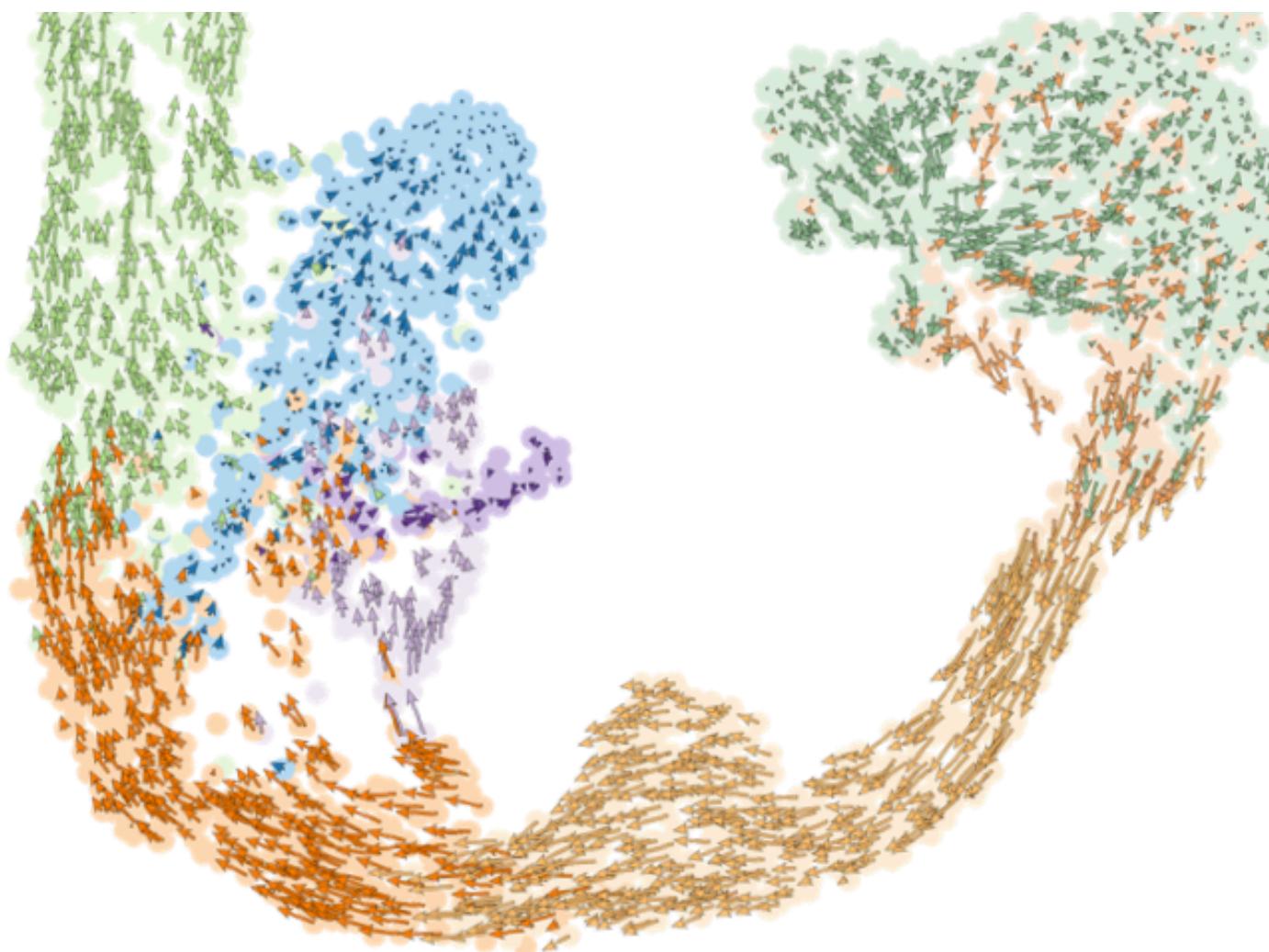
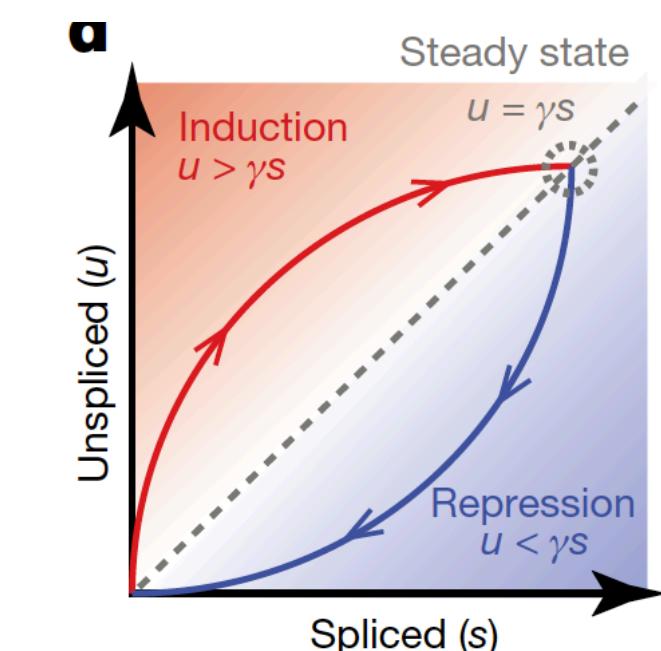


$$\frac{du_g}{dt} = \alpha_g - \beta_g u_g$$

$$\frac{ds_g}{dt} = \beta_g u_g - \gamma_g s_g = v_g$$



$$\vec{V} = \sum_{g=1}^G \vec{v}_g$$



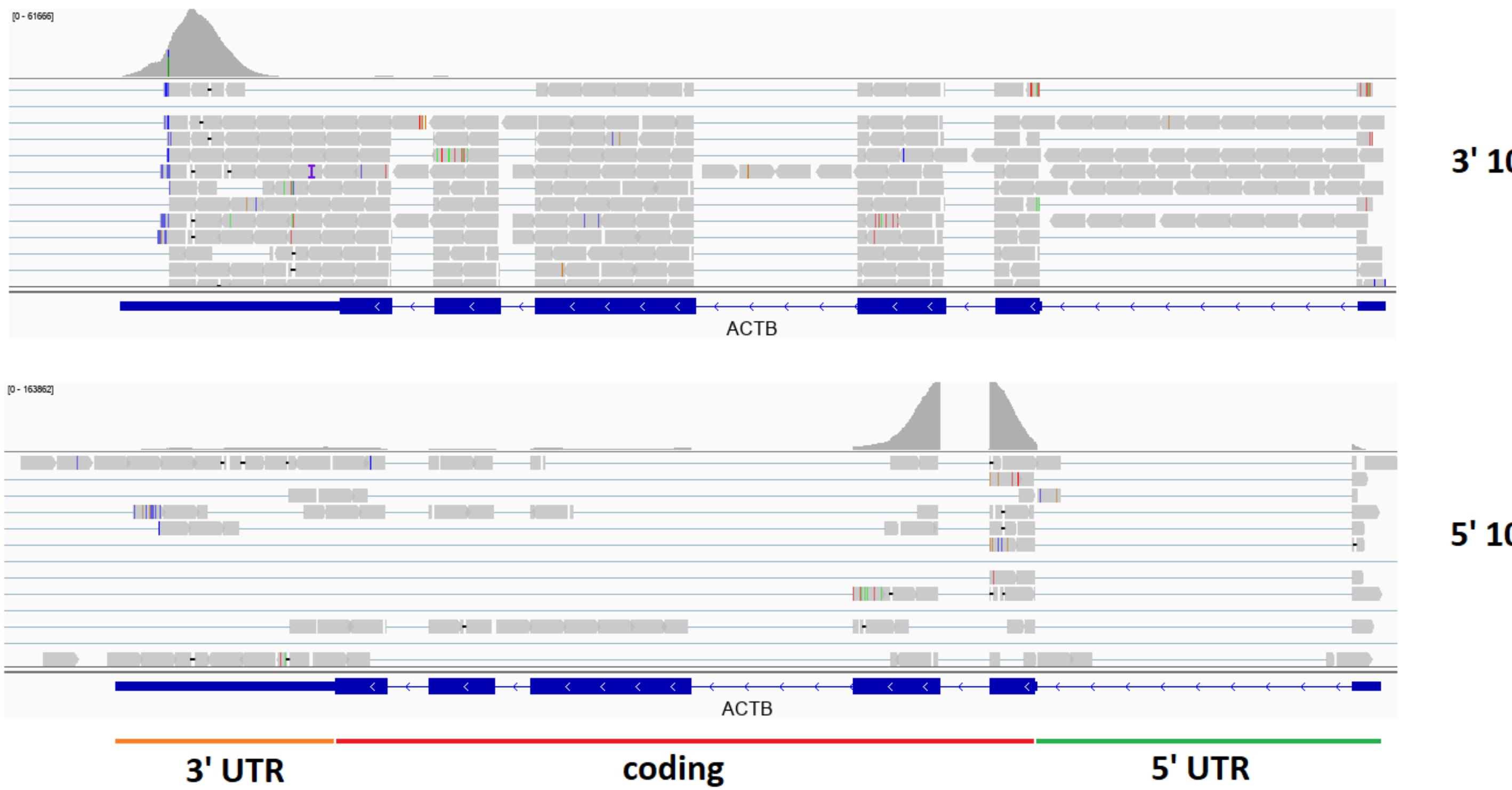
Bergen et al, Nature Biotechnology 2020

Exercises:

a) Derive the time independent relation between u and s from the splicing ODEs.

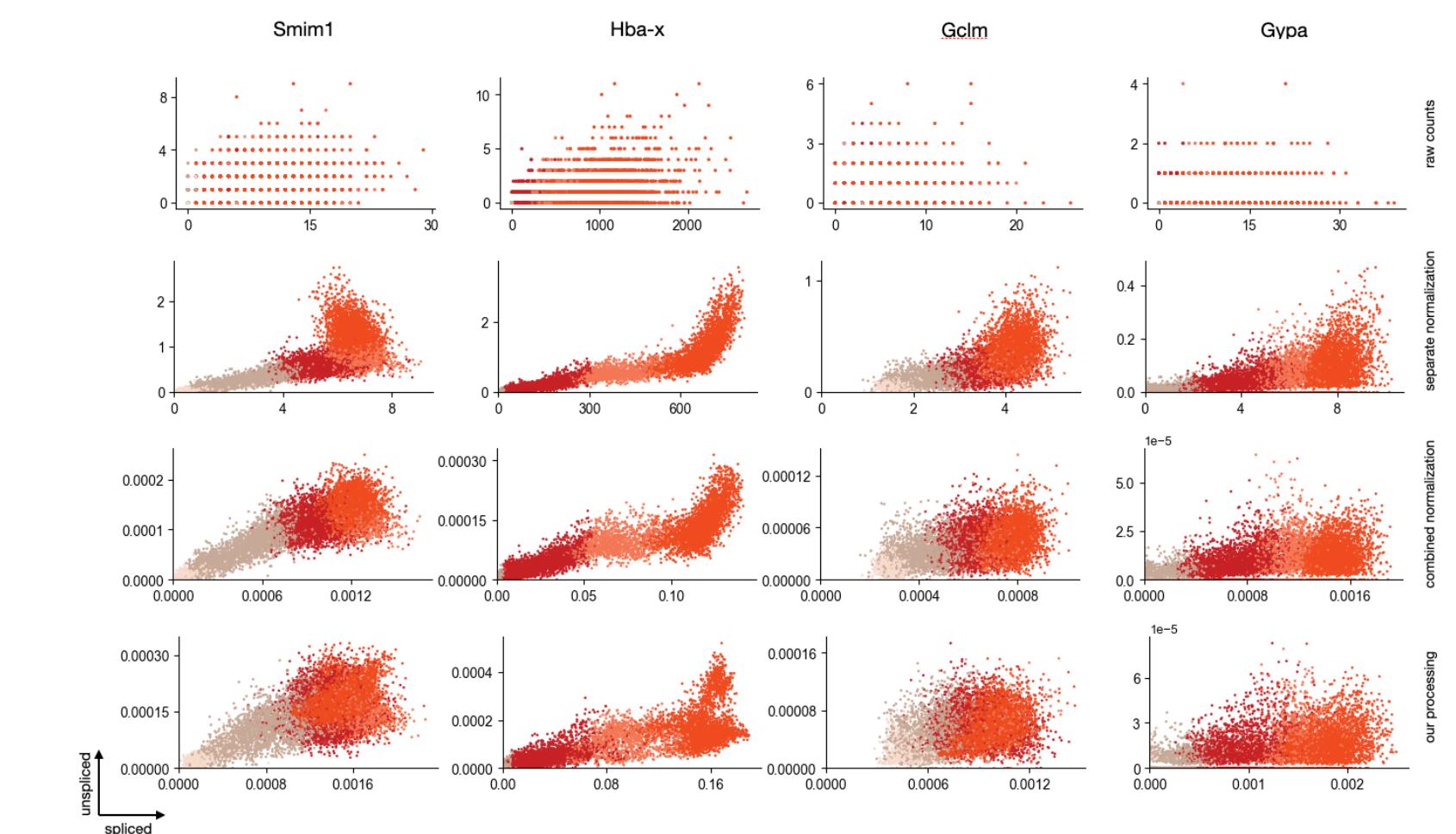
U-S counting and estimation from short-read scRNA-seq data

Bam files visualisation by Genome Browser IGV

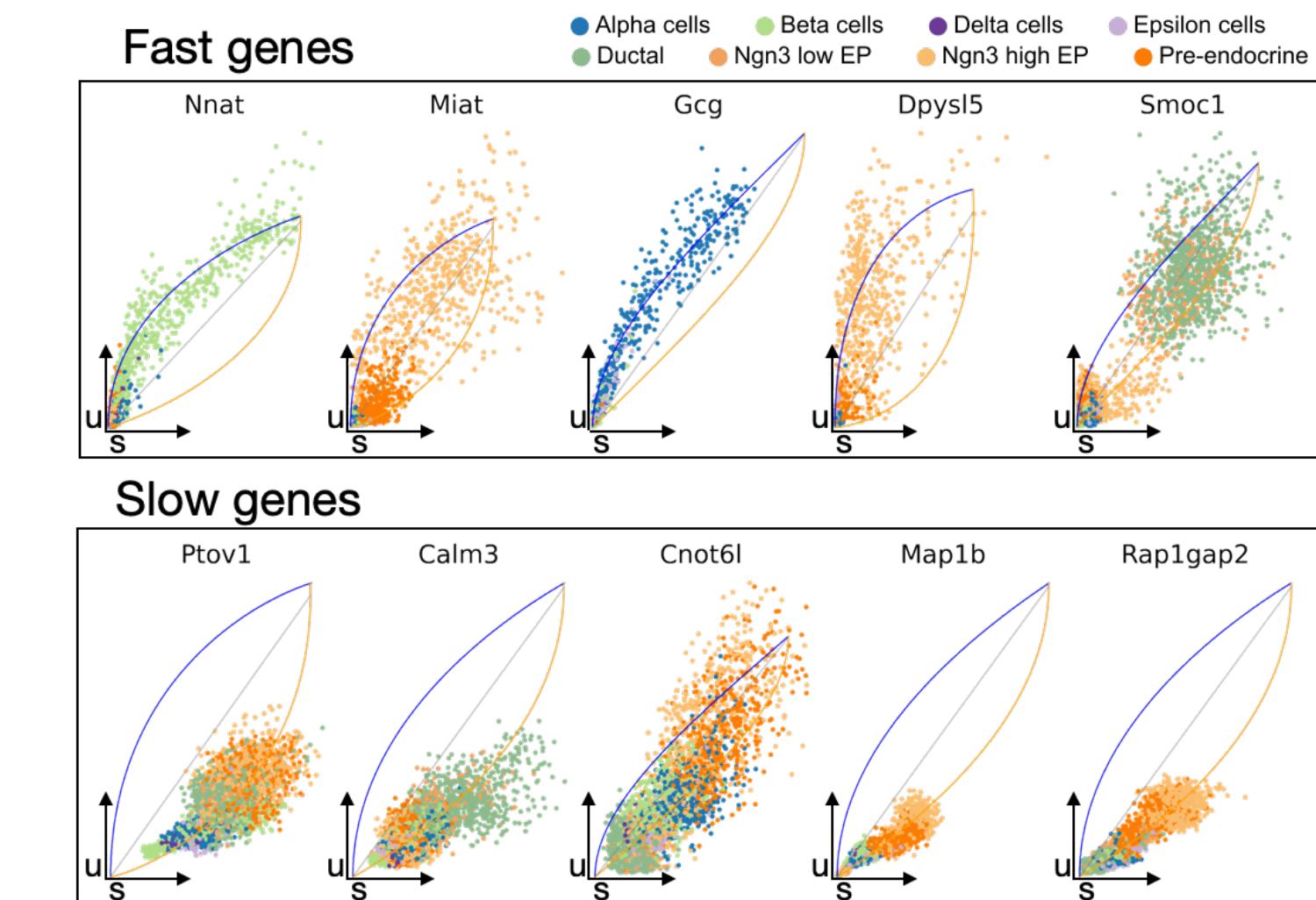
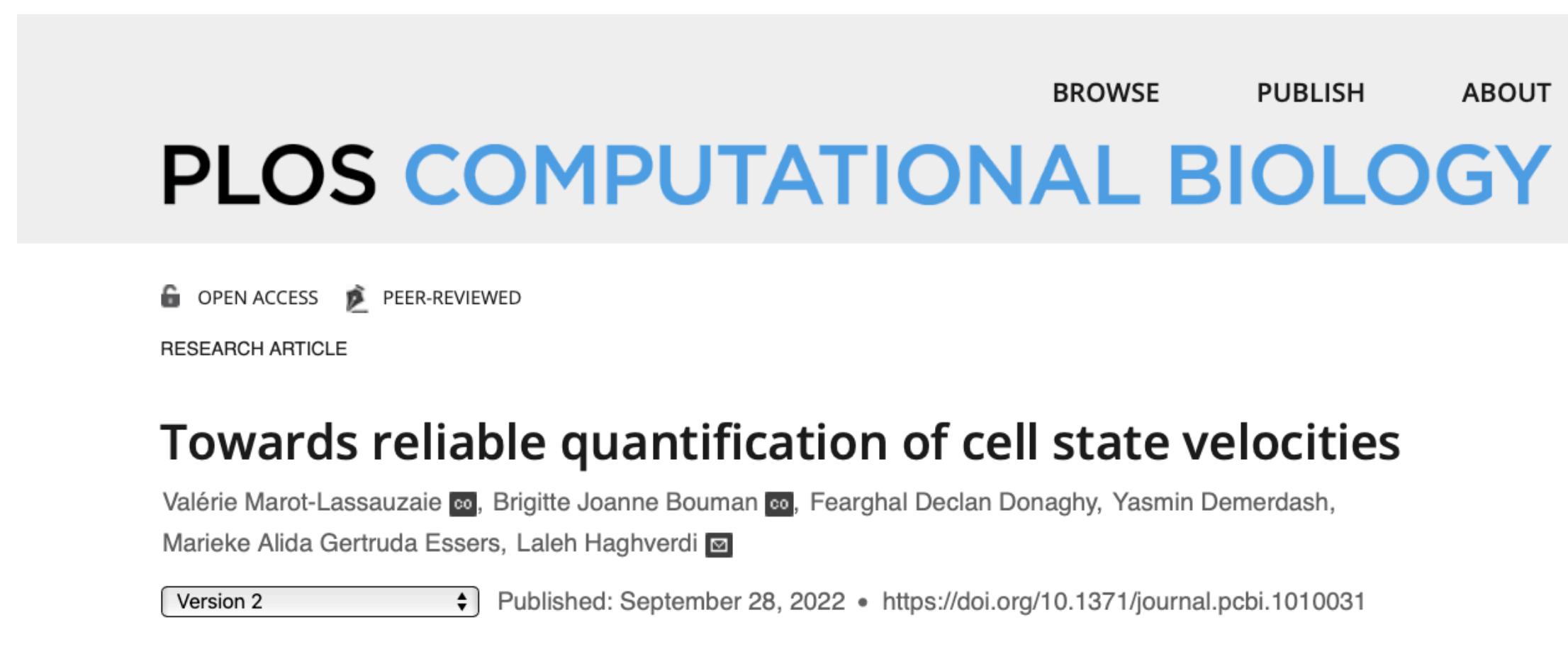


- Long read data (Oxford nanopore, PACbio etc.) rising

- Really bad data qualities
- Most reads from one side of the gene body → extrapolation of u-s counts
- Not covering many intron-exon region
- Most reads are spliced



Challenges of cell state velocities estimation

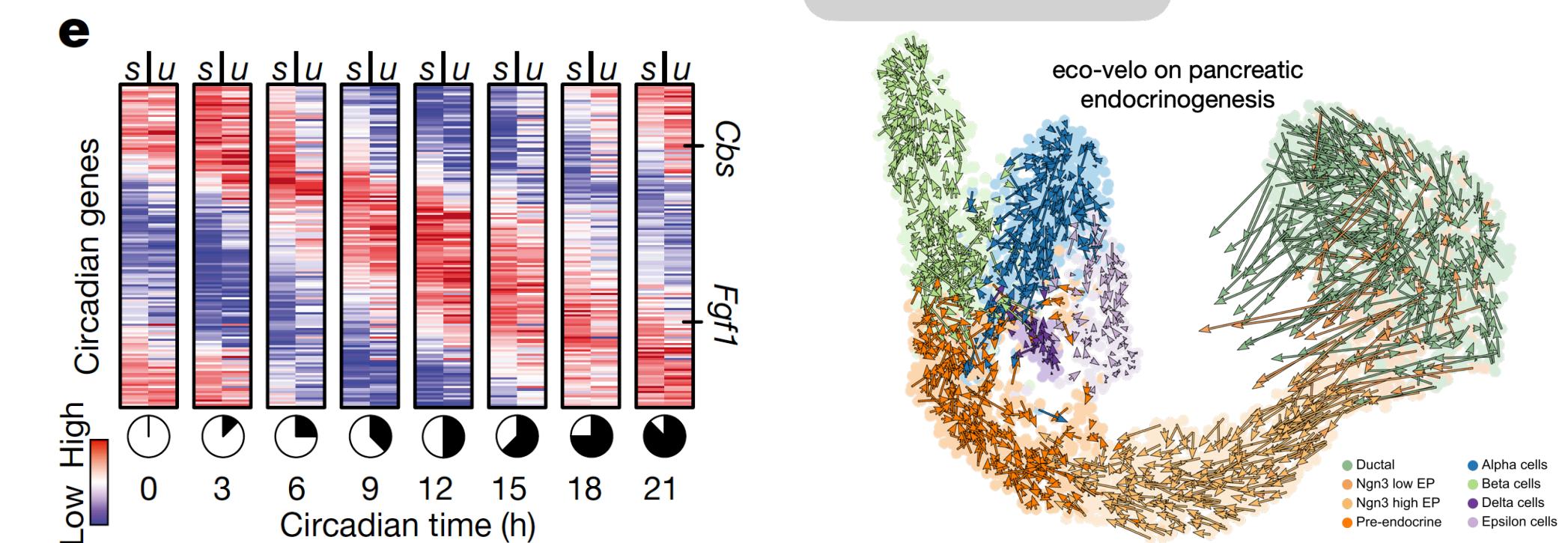


- Kappa-velo: address scale invariance by using cell densities

Circadian-associated genes in the mouse liver over a 24-h time course
[La Manno et al, Nature 2018]

- Eco-velo:

- Skip error-prune expensive gene-wise curve fitting
- u/s MNN matching in genes space
- Select gene set with similar splicing and degradation time scales

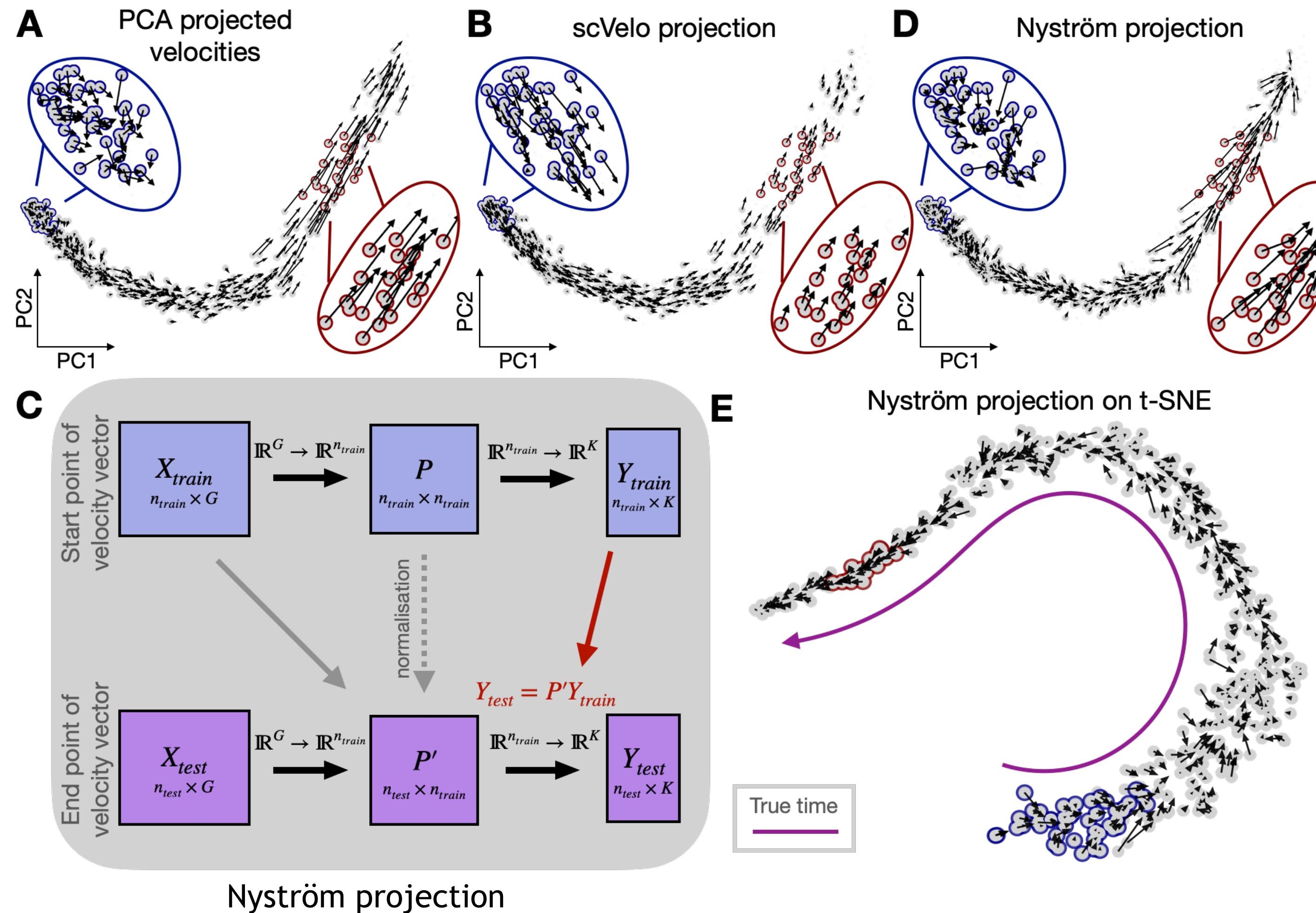


Exercises:

a) In what conditions (assumptions) from the ODE dynamics can we conclude that the amount of unspliced mRNA reads in time t approximates the spliced mRNA reads in time $t + \tau$?

Visualisation of cell state velocities

Marot-Lassauzaie, Valérie, et al. "Towards reliable quantification of cell state velocities." *PLoS Computational Biology* (2022)



- Previous (velocyto, scVelo) methods:

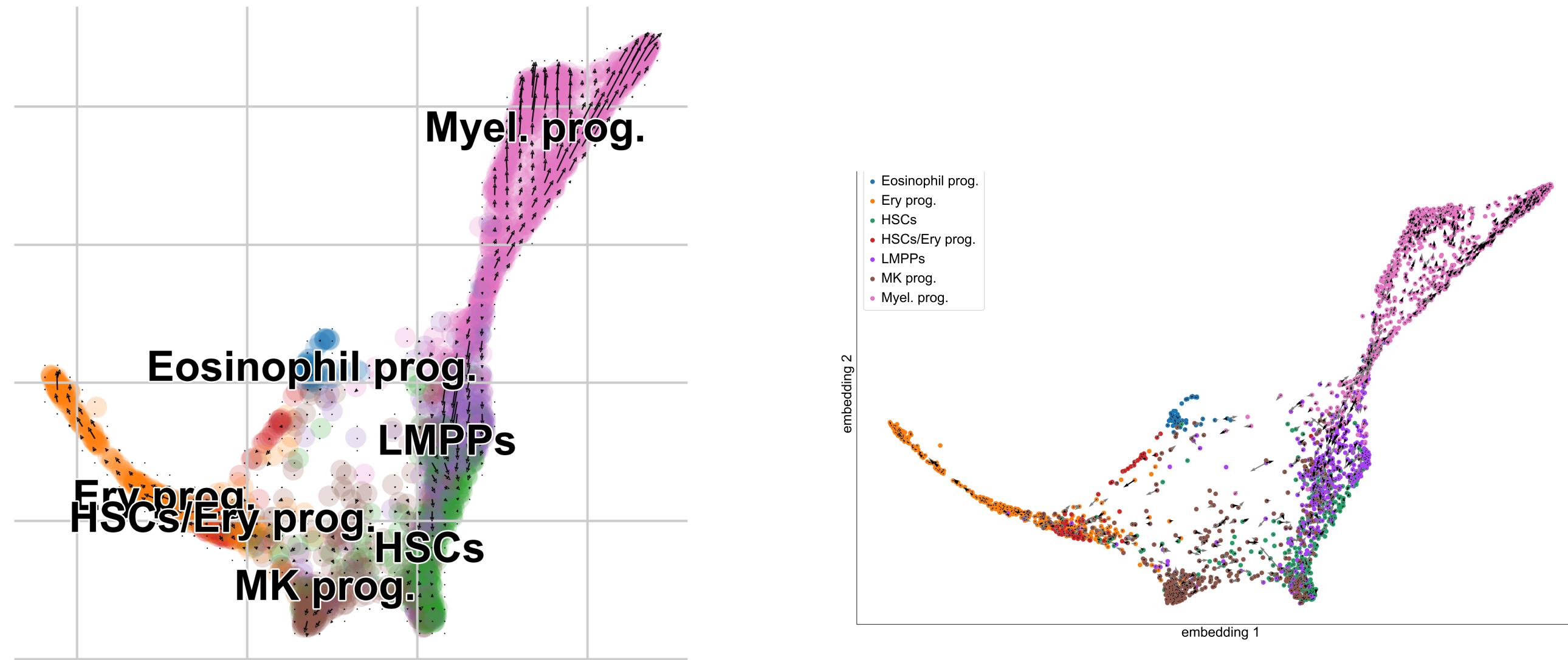
$$\Delta \vec{Y}_i = \sum_j (P_{ij} - \frac{1}{n}) \frac{\vec{Y}_j - \vec{Y}_i}{\|\vec{Y}_j - \vec{Y}_i\|}$$

$$P_{ij} = \exp\left(\frac{\cos \angle(\vec{s}_j - \vec{s}_i, \vec{v}_i)}{\sigma^2}\right)$$

- Unfaithful to:
 - Velocity vectors norm
 - Cell-to-cell variety of velocities (stochasticity and cell plasticity)
- General artefacts of out of distribution non-linear projection
- Use PCA when possible

eco-velo upgrade

- No error-prone gene-wise parameter fitting
- Applicable for u-s as well as unlabeled-labeled- mRNA
- Fixed time-scale (labelling time)
- Joint embedding of current old-new cell states —> also resolves out-of-distribution projection problems



- Preserve velocity heterogeneity and quantify the diffusion component
- Compare to velocities post-inflammation

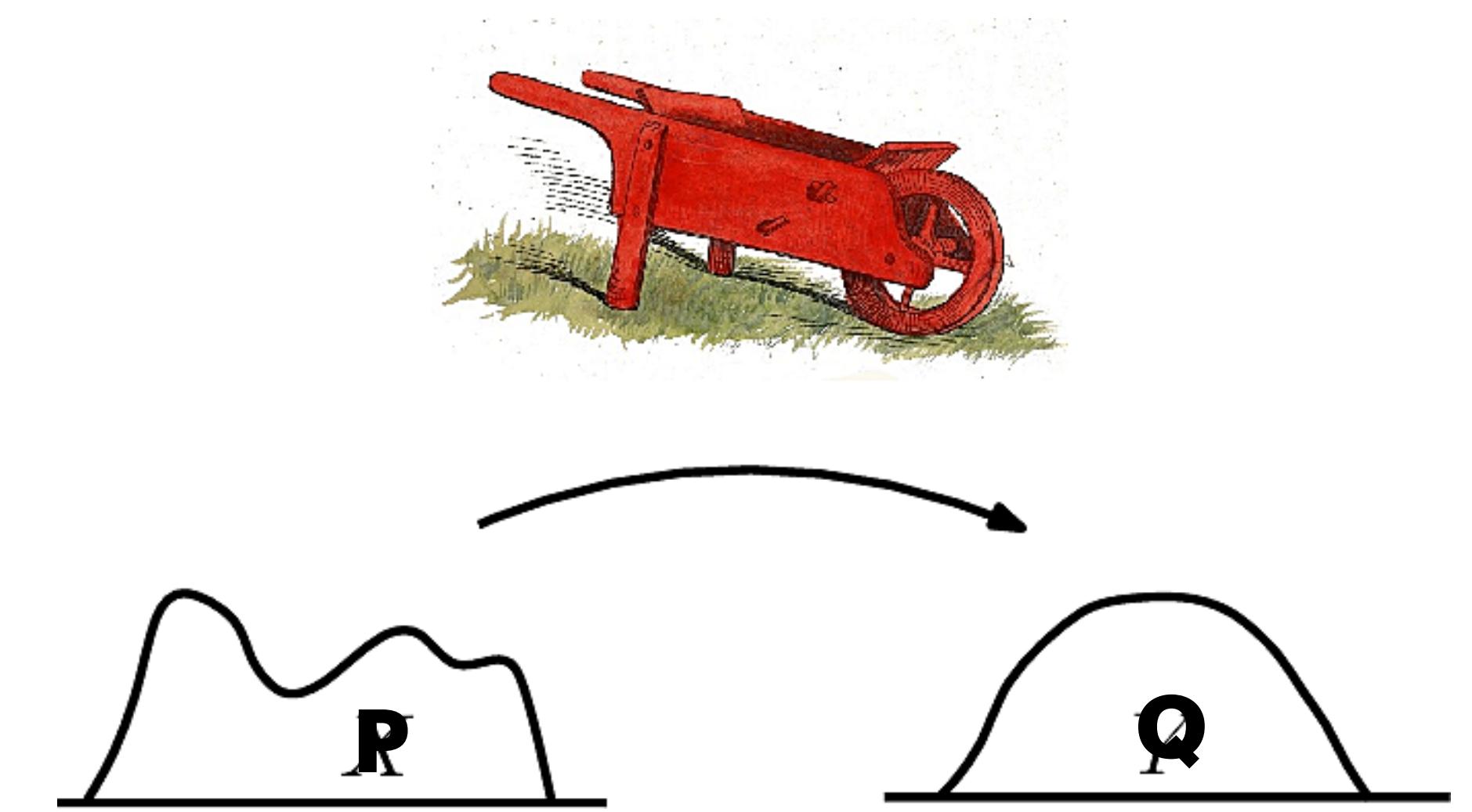
Exercises:

- a) If you have the velocity estimation for a group of cells in the neighbourhood of s , can you calculate the magnitude of the diffusion coefficient relative to the drift for position s in the Fokker-Planck equation?
- b) Use a time evolution simulations data (in 2D), calculate the velocity for each cell (e.g. in 5 time steps) to demonstrate with your proposition one can recapture the ground truth Fokker Planck coefficients.

Optimal Transport

Balanced case: Earth Mover's Distance

- Distance between (normal) distributions P and Q



$$EMD(P, Q) = \min_{f_{ij}} \left(\sum_{i \in p, j \in q} f_{ij} d_{ij} \right)$$

- With transport policy matrix F and distance matrix D between $i \in T_1$ and $j \in T_2$.
- 1D is easy using the Cumulative Distribution Functions of P and Q

$$EMD(P, Q) = \int_{-\infty}^{+\infty} \text{CDF}(P(x)) - \text{CDF}(Q(x)) dx$$

- More difficult to find the optimal flow in high-D
- More difficult to find the optimal flow in the unbalanced (i.e., unpreserved probability densities) case

Exercises:

a) Prove the following reaction holds in for EMD (P , Q) in 1D:

$$EMD(P, Q) = \int_{-\infty}^{+\infty} \text{CDF}(P(x)) - \text{CDF}(Q(x)) dx$$

b) Which step in the proof does not hold in higher dimensional space?

(Entropic regularised and unbalanced) Optimal Transport

- Entropy regularisation: add some randomness to the optimisation problem → soft flow (continuous values) probabilities π_{ij} instead of a permutation flow f_{ij} , efficient solution in high-D

$$f_{ij} \in \{0, 1\}$$

Cuturi M. Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems. 2013;26.

- Unbalanced: no conservation of mass (sum of probabilities not equal to 1 for P and Q)

$$\pi_{ij} = \operatorname{argmin}_{\pi} \left(\sum_{i \in 1:N1, j \in 1:N2} c(s_i, s_j) \pi_{ij} - \epsilon \sum_{i \in 1:N1, j \in 1:N2} \pi_{ij} \log \pi_{ij} \right. \\ \left. + \beta_1 \tilde{KL} \left(\sum_{i \in 1:N1} \pi_{ij} || Q \right) + \beta_2 \tilde{KL} \left(\sum_{j \in 1:N2} \pi_{ij} || P \right) \right)$$

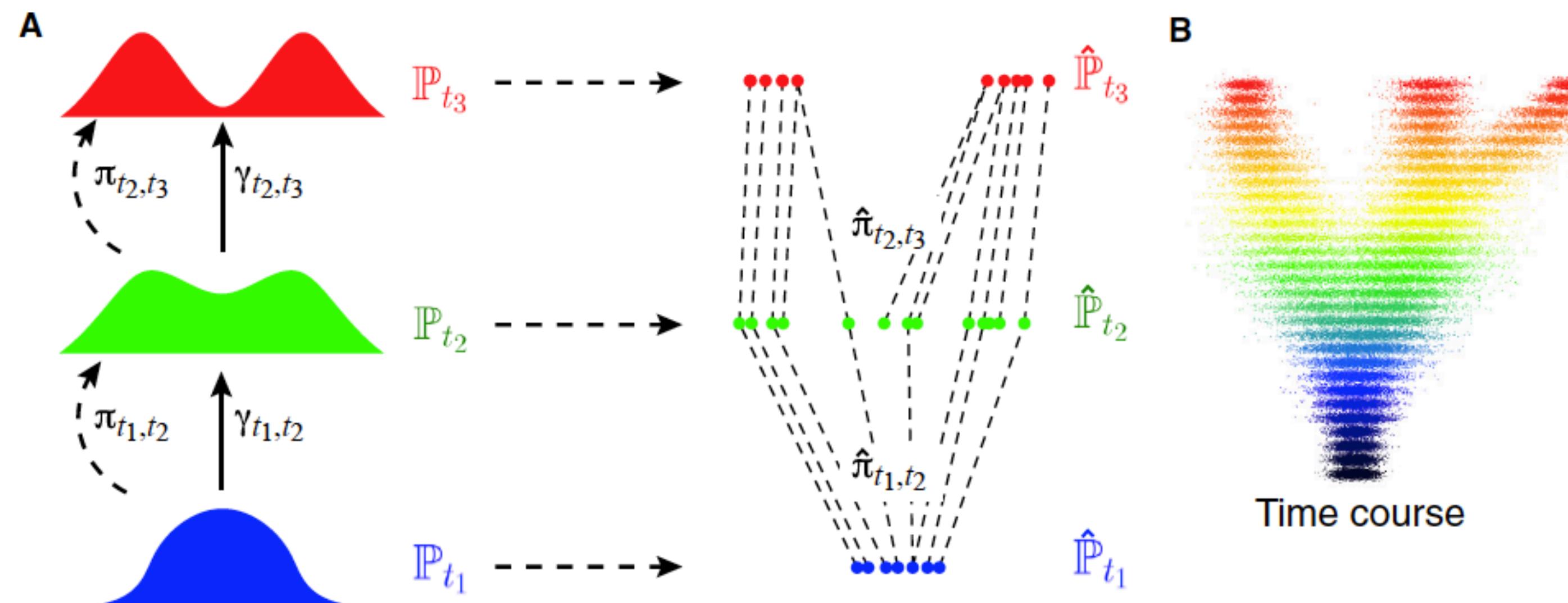
Exercises:

a) f is a strongly convex function if for a $\mu > 0$ we have:

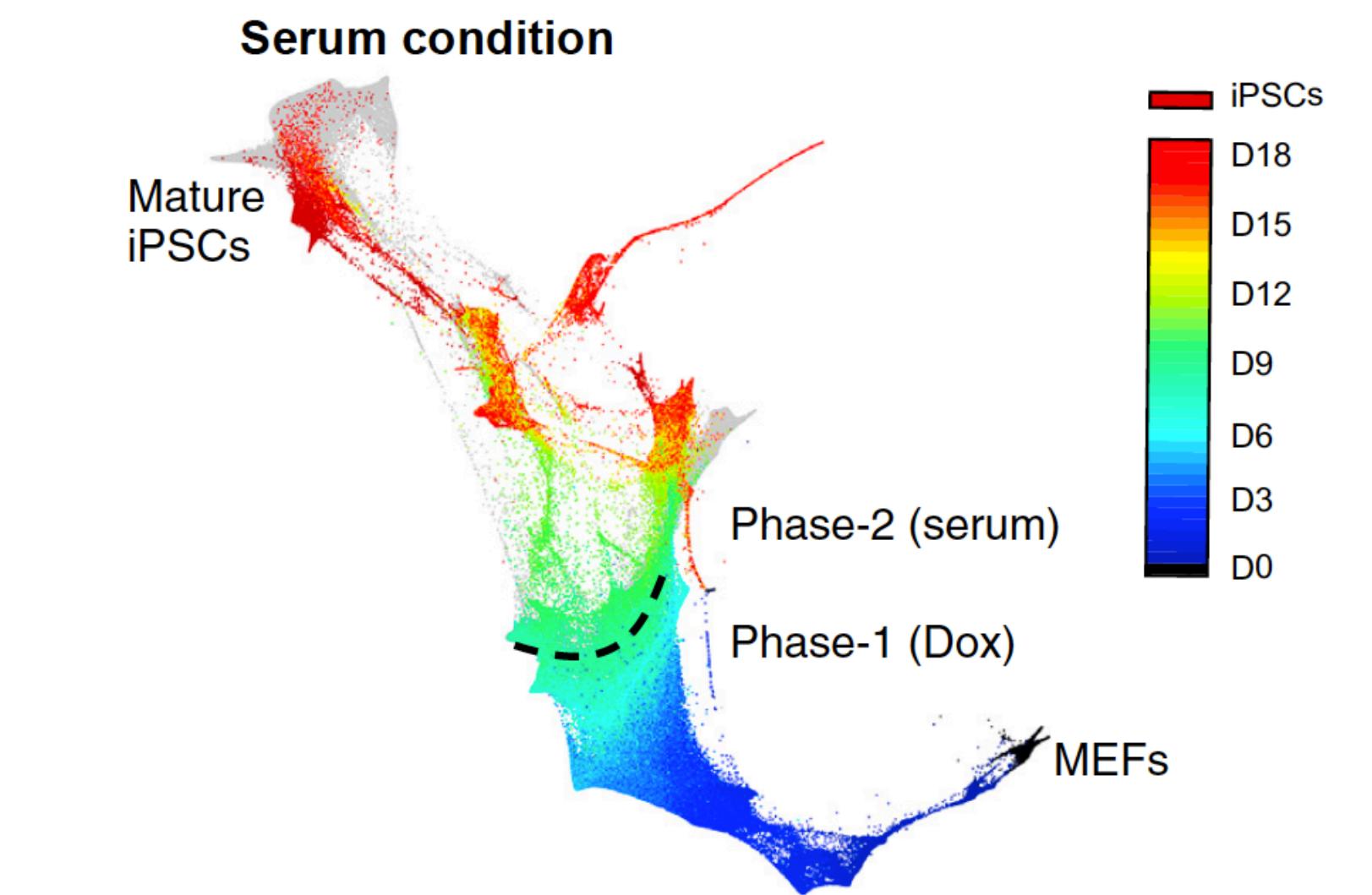
$$f(y) \geq f(x) + \nabla f(x)(y - x) + 2\mu(y - x)^2 \quad , \forall x, y$$

- b) Why in optimisation not only convexity but strong convexity is desirable?
- c) Prove that the EMD problem (without including the entropy (diffusion) term) is convex.
- d) Prove that the OT loss function with including the entropy (diffusion) term is strongly convex. Consider the balanced OT (without birth-death terms).

Unbalanced Optimal transport for time course experiments



Mouse Embryonic Fibroblasts



Schiebinger et al. Cell 2019

Optimal Transport relation with the diffusion-drift drift model

Π : transition matrix

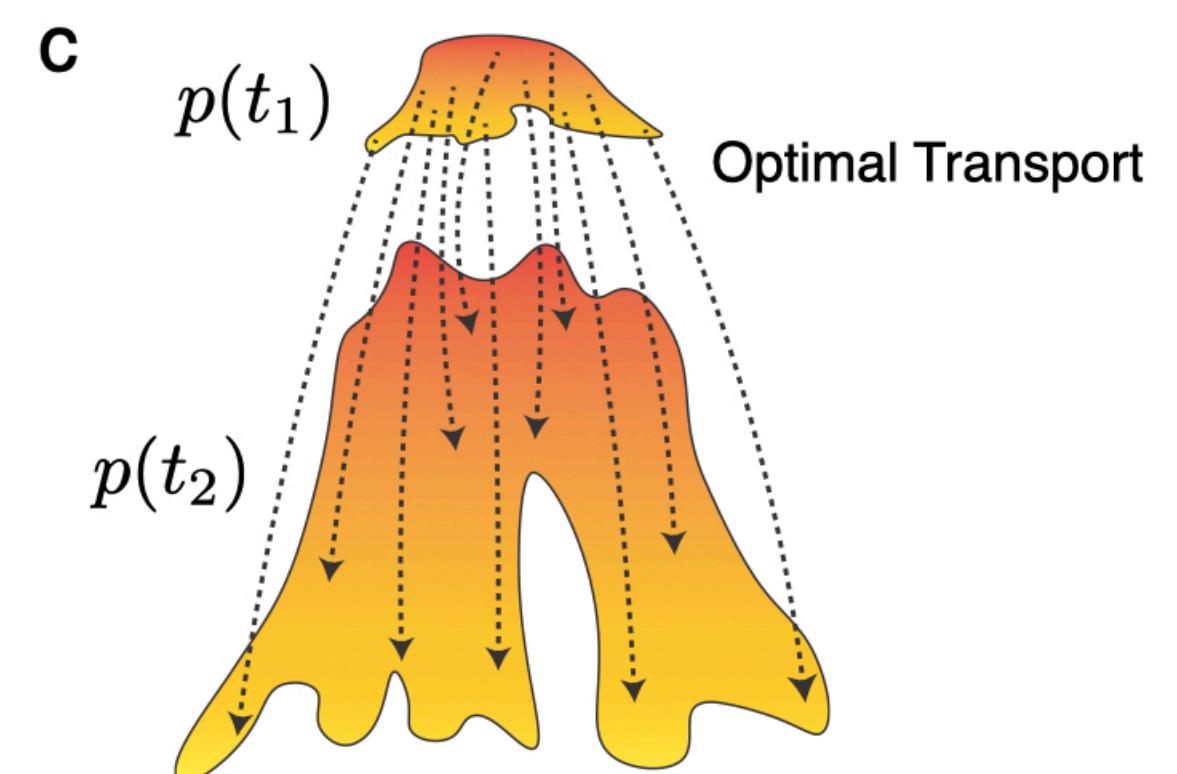
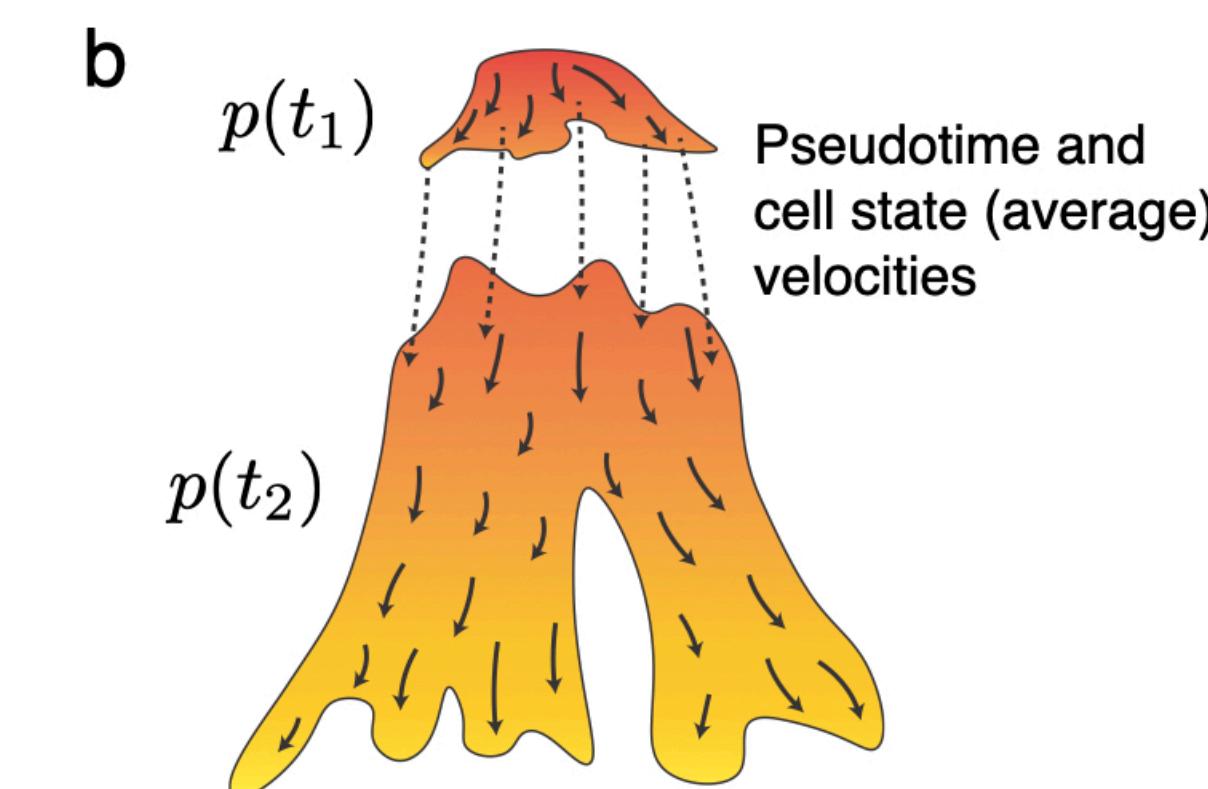
C : distance matrix

\mathbb{P} : vector estimate of birth-death rate for the cells in T1

\mathbb{Q} : vector of 1 entries for all cells in T2

Under certain assumptions:

- Change in energy ~ Euclidean distance between cell states
- Same diffusion parameter for all cell states (i.e, positions)



Exercises:

a) in unbalanced OT, the following generalisation of KL is used instead of the standard KL to constrain the birth-death rates with initial estimates :

$$\tilde{KL}(R|S) = \sum_{i=1}^{N_1} \left(S_i \log \left(\frac{S_i}{R_i} \right) - S_i + R_i \right)$$

- b) What is the role of the $-S+R$ terms above?
- c) Can one impose the initial birth-death estimate by only $(R-S)^2$ and omit the first term (i.e., KL form)? Would the optimisation problem still be convex?

Diffusion-drift relation with Optimal Transport

Cells in T_1

- "Diffusion pseudotime robustly reconstructs lineage branching"
Haghverdi et al. Stem cell Reports 2023 (Supplemental Note 1)
- Discrete (matrix form) Diffusion-drift:

$$\Delta P_{(t)} = -P_{(t)} \Lambda (L^\alpha + W) \quad \Pi$$

$$P_{(t)} = P_{(t-1)} (I - \Lambda (L^\alpha + W))$$

$$P_{(t_1+t)} = P_{(t_1)} \Pi^t$$

- Given P at t1 and Q at (t1+t), Diffusion-drift Likelihood function L is:

P: State vector (N cell states) (1, 1, .., 1, 0, 0, .., 0)
 L: Laplacian matrix (undirected, normalised).
 α : Diffusion magnitude

W: Drift (directed, normalised)

Λ : Birth/death, diagonal matrix

$$\begin{aligned} L &= \sum_{i \in 1:N_1, k \in N-N_2:N} P_{1i} \quad (\Pi^t)_{ik} \quad Q_{k1} \\ &= \sum_{i \in 1:N_1, k \in N-N_2:N} P_{1i} \quad [(I - \Lambda(L^\alpha + W))^t]_{ik} \quad Q_{k1} \\ &= \sum_{i \in 1:N_1, j \in 1:N_2} \mathbb{P}_i \quad \hat{\pi}_{ij} \quad \mathbb{Q}_j \\ &\quad P_i \quad Q_j \end{aligned}$$

Diffusion-drift relation with Optimal Transport

- Log-likelihood optimisation of discrete diffusion-drift \rightarrow Optimal Transport

$$\pi_{ij} = \operatorname{argmin}_{\pi} \left(\sum_{i \in 1:N1, j \in 1:N2} c(s_i, s_j) \pi_{ij} - \epsilon \sum_{i \in 1:N1, j \in 1:N2} \pi_{ij} \log \pi_{ij} \right. \\ \left. + \beta_1 \tilde{KL} \left(\sum_{i \in 1:N1} \pi_{ij} || \mathbb{Q} \right) + \beta_2 \tilde{KL} \left(\sum_{j \in 1:N2} \pi_{ij} || \mathbb{P} \right) \right)$$

Drift **Diffusion**

Birth/date

Exercises:

- a) The relation between the Diffusion-Drift model and OT is only sketched in the supplemental note of Haghverdi & Ludwig 2022 (last two slides). Write the log-likelihood $\log(L)$ and show more precisely how maximisation take the form of OT cost minimisation. For simplicity (first) consider the balanced OT case with mass conservation.
- b) Under what assumption the change in energy of a particle (cell) $U_i - U_j$ in Fokker-Planck can be replaced by $c(i,j)$? Consider the drift matrix W_{ij} as $\exp(- (U_i - U_j)/k)$ with row normalisation.
- c) Can you imagine a scenario in cell differentiation that this assumption does not hold?
- d) Write a new version of OT loss function that relaxes this assumption and considers the true work (change in energy) for moving a particle from position s_i to s_j .

Seminar suggestion :

Wang, Shou-Wen, et al. "CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information." *Nature Biotechnology* 40.7 (2022): 1066-1074.

**Thank you for your
attention!**