

# Trajectory inference using entropic optimal transport

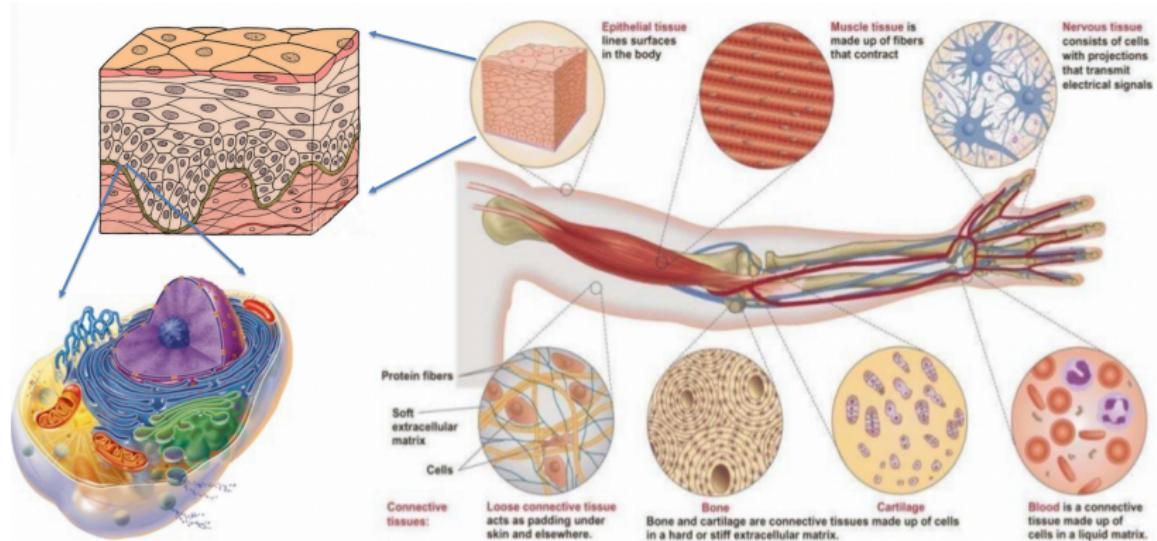
Cours Transport optimal & Machine Learning

Elias Ventre

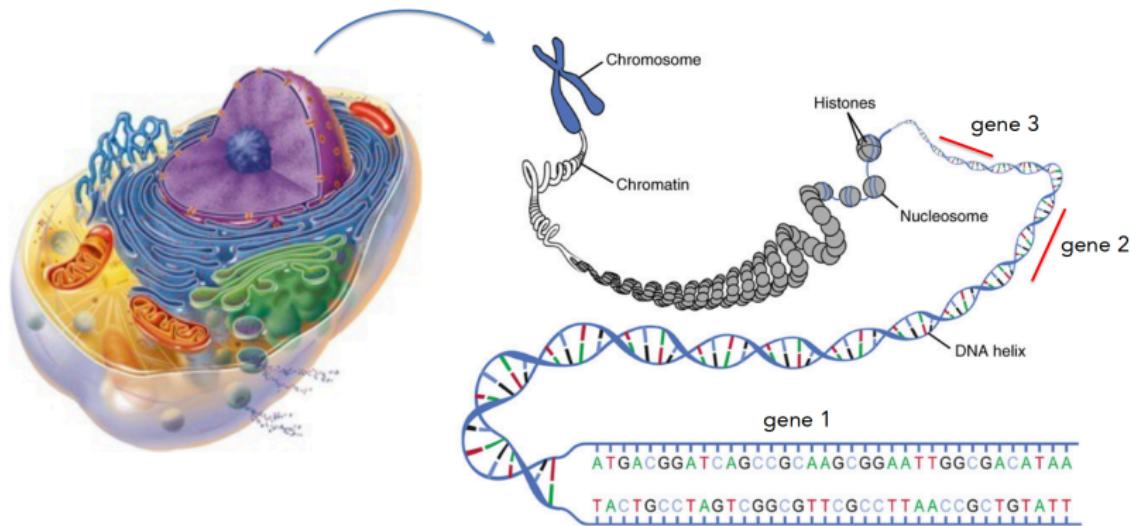
UCBL1

29 Janvier 2026

# Cell biology



# From cell to DNA



# Gene expression

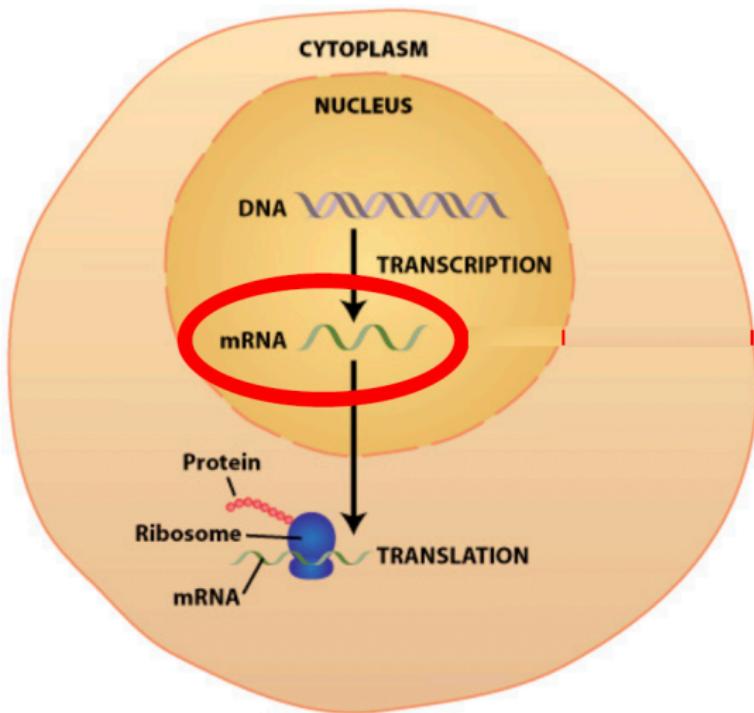


Image from G. Schiebinger.

# Gene expression measurement

- $X_t = (X_{1,t}, \dots, X_{d,t})$  = vector characterizing a cell  $X$  at time  $t$  in the **gene expression space**  $\mathcal{X} \subset \mathbb{R}^{d^+}$ .

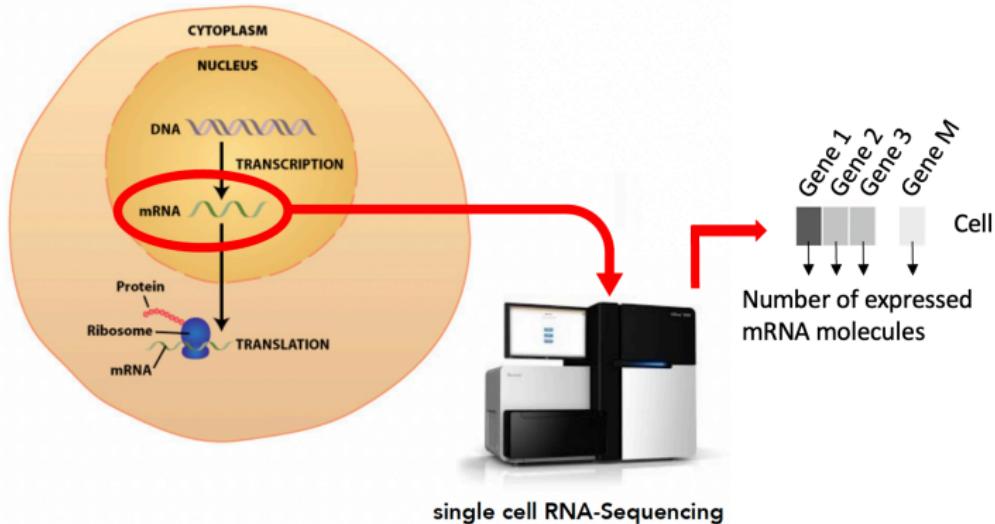


Image from G. Schiebinger.

# Vector of cell identity

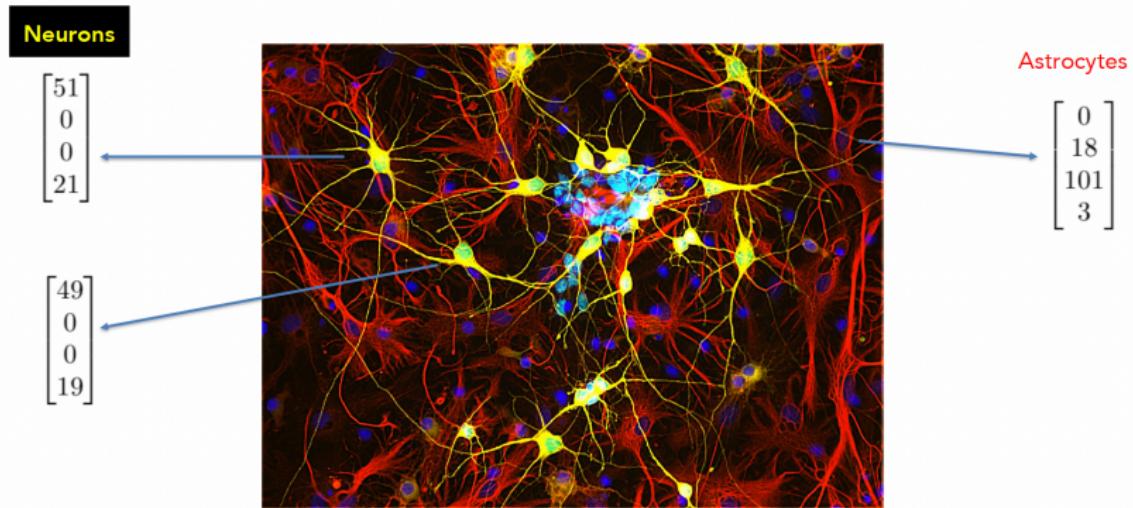
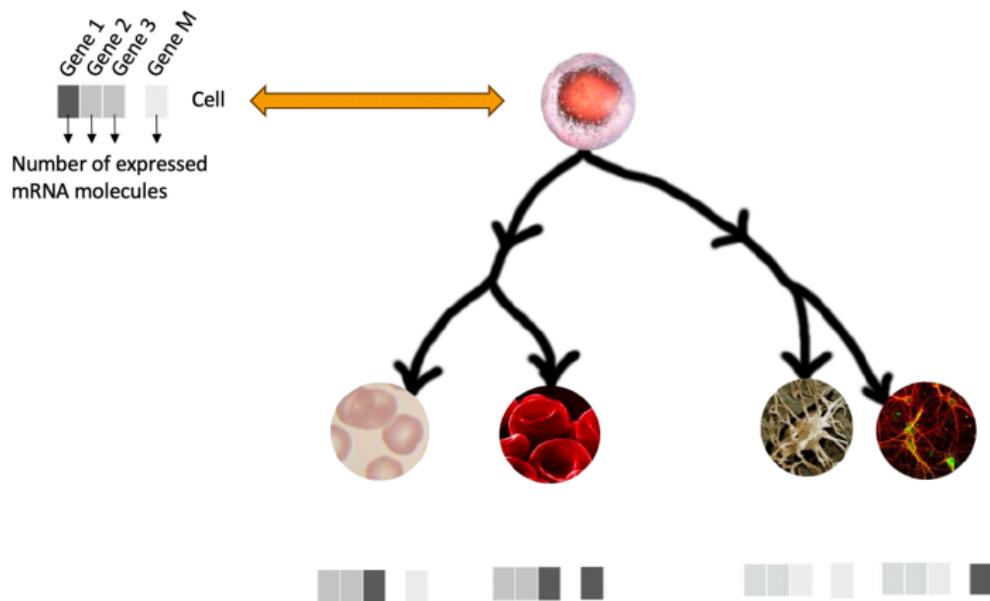
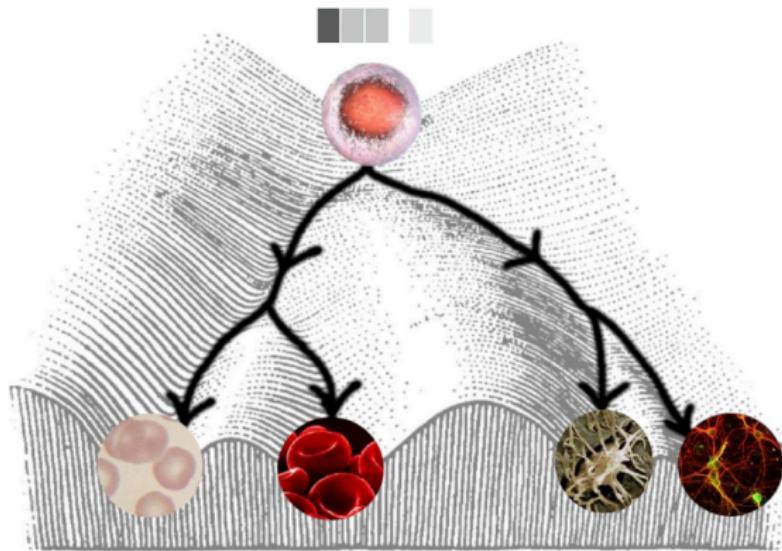


Image from G. Schiebinger.

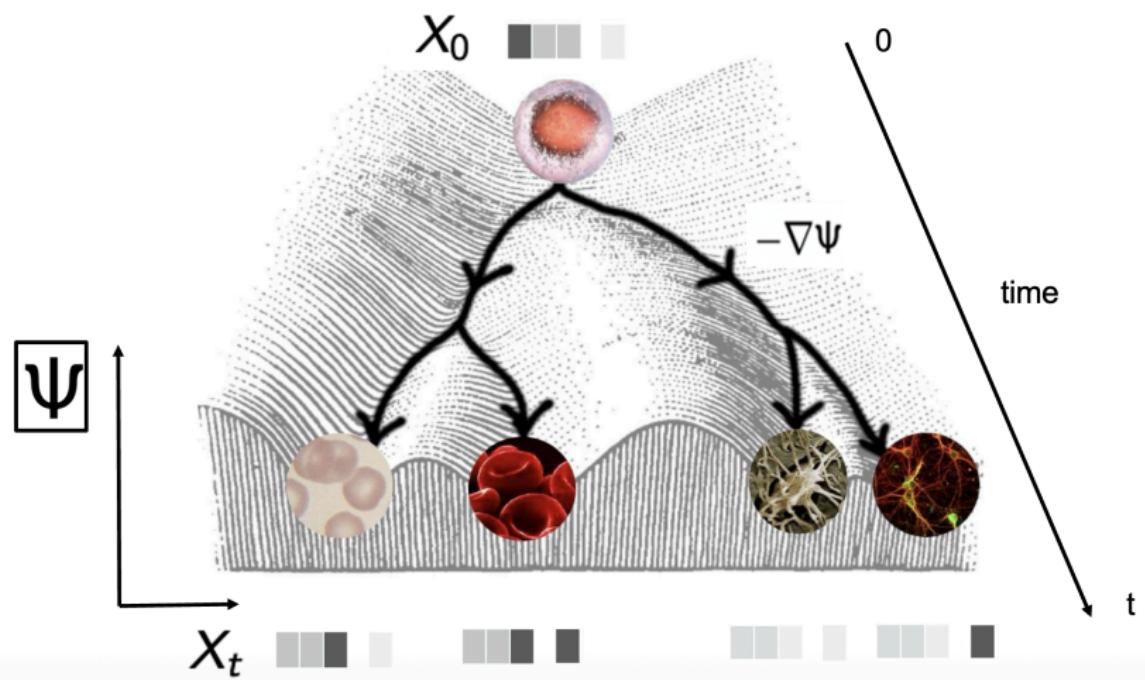
## Cell dynamics



# Cell dynamics



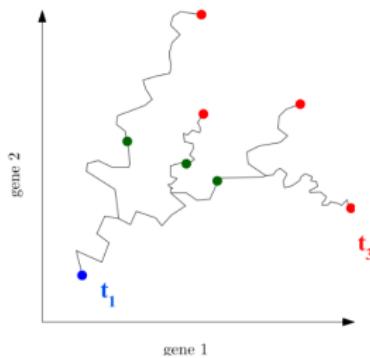
# Cell dynamics



# Stochastic process modeling cell dynamics

- **Characteristics:**

- Cells change gene expression over time;
- Cell division creates branching paths;
- Cell may communicate (via proteins);

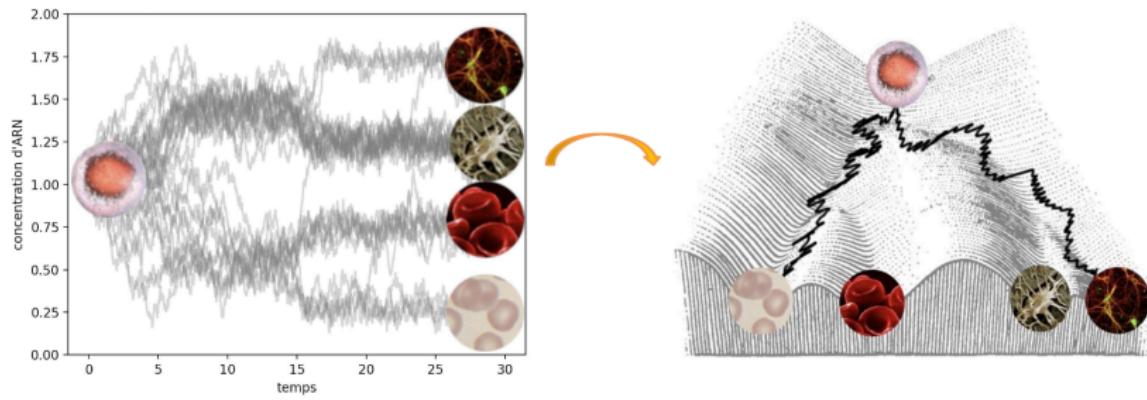


- **Fact:**

- Stochasticity plays a role in cell differentiation.

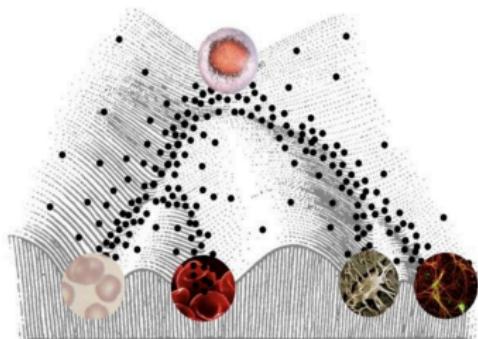
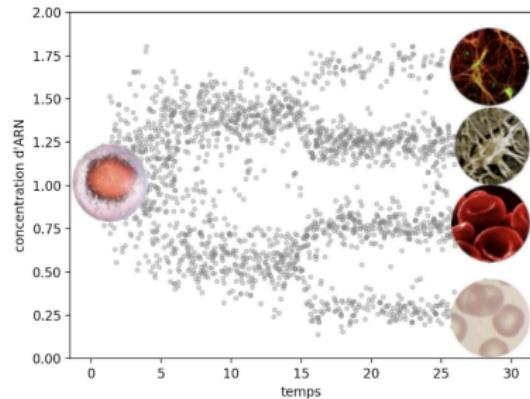
⇒ Cell dynamics can be described by a **stochastic process**, characterized by a **path measure**.

# Data to fit dynamics



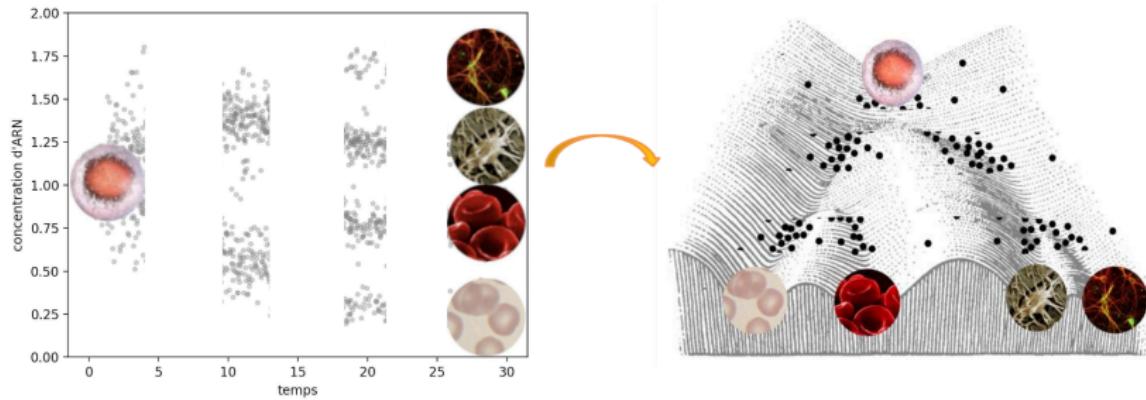
- Measurement kills cells so we cannot observe real paths to perform landscape reconstruction/model calibration.

# Data to fit dynamics



- Experimental constraints don't allow to get continuous temporal information.

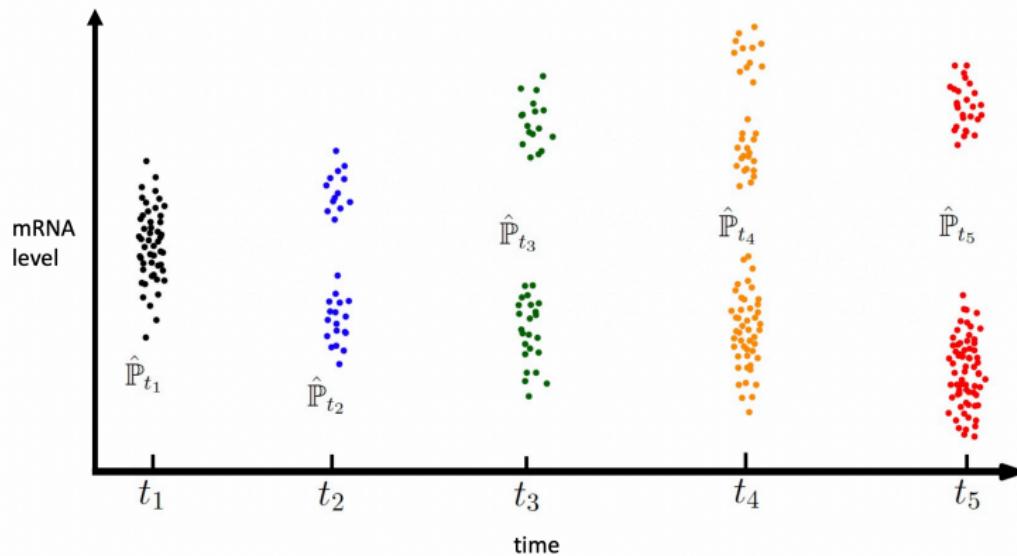
# Data to fit dynamics



- Thus, we only have access to snapshots of independent cells at some timepoints to reconstruct the process.

# Time-courses of single-cell datasets

- Basic data = **Samples of cells** representing some (empirical) distributions  $\hat{\mathbb{P}}_{t_i}$  between a time  $t_1$  and a time  $t_N$ .

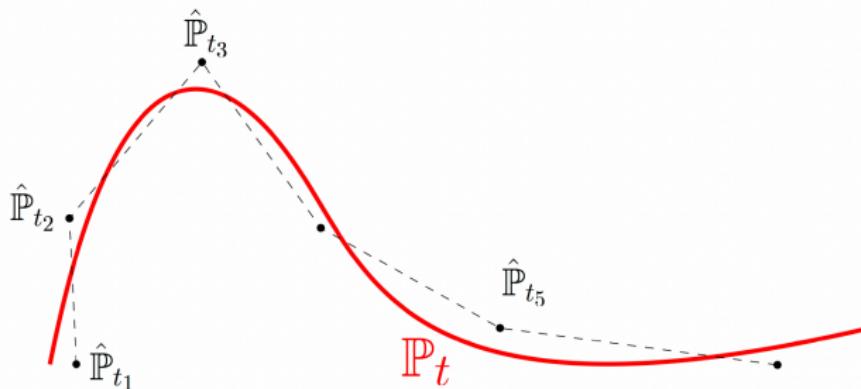


# Trajectory inference from single-cell data

- Assume for a moment that there's no branching or communication:

⇒ Aim of trajectory inference: reconstructing the

**path-measure  $\mathbb{P}$  from the empirical marginals**  $\hat{\mathbb{P}}_{t_i} := \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{X_j^i}$ .



# Trajectory inference from single-cell data

There are several challenges:

- We want to reconstruct a **measure on paths**  $\mathbb{P}$  from only (snapshots of) a **path of measures**  $\mathbb{P}_t$ :  
     $\Rightarrow$  We will need modeling assumption on the process (Session 1).
- The empirical marginals  $\hat{\mathbb{P}}_{t_i}$  are noisy estimations (in the space of measure) of the true temporal marginals  $\mathbb{P}_{t_i}$ :  
     $\Rightarrow$  We will need to formulate a **global problem** to optimize the marginals themselves (Session 2).

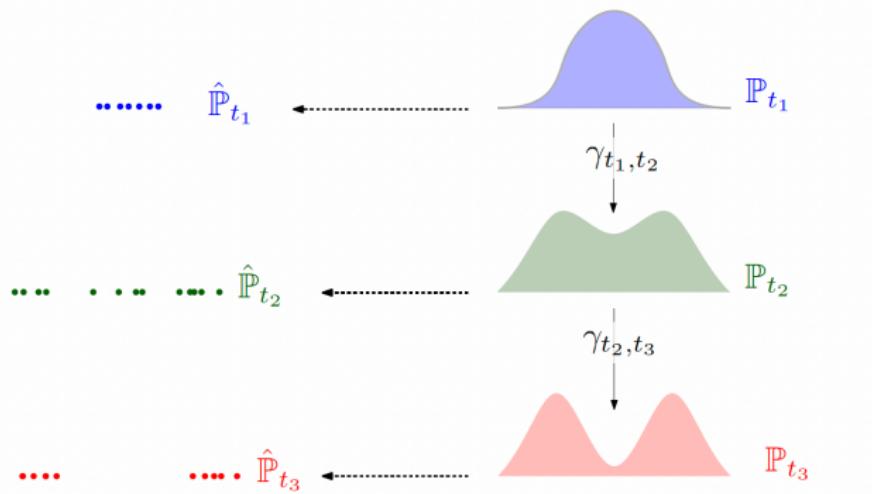
# Table of content

- **Session 1:** Using Entropic Optimal Transport for trajectory inference with perfect marginals.
- **Session 2:** Tackling noise in marginals by optimizing over measures.

## Session 1: Trajectory inference with perfect marginals

# Method for trajectory inference

- Aim  $\sim$  finding the **true coupling between the samples**:



- Assume that cell differentiation can be approximated by a **SDE**:

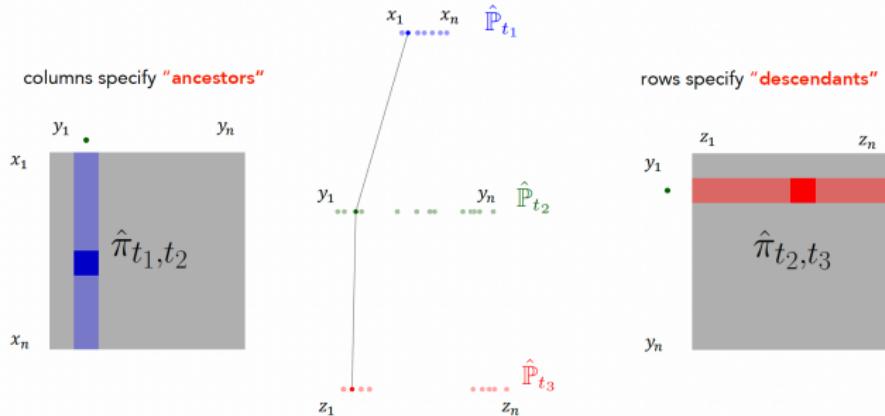
$$dX_t = -\nabla \Psi(t, X_t) dt + \sqrt{\sigma} dB_t;$$

# Coupling with Entropic optimal transport

- Schiebinger et al. (2019): use **entropic optimal transport** coupling between  $\hat{\mathbb{P}}_{t_i}$  and  $\hat{\mathbb{P}}_{t_{i+1}}$ :

$$\gamma_{t_i, t_{i+1}}^* := \arg \min_{\pi} \left\{ \underbrace{\mathbb{E}_{\pi} |X - Y|^2}_{\text{optimal transport}}, \right.$$

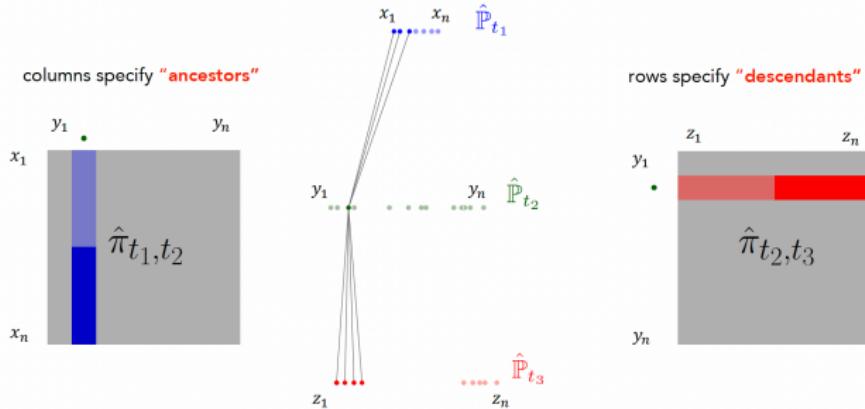
$$\text{s.t. } \int \pi(\cdot, dy) = \hat{\mathbb{P}}_{t_i}(\cdot), \int \pi(dx, \cdot) = \hat{\mathbb{P}}_{t_{i+1}}(\cdot) \left. \right\}.$$



# Coupling with Entropic optimal transport

- Schiebinger et al. (2019): use **entropic optimal transport** coupling between  $\hat{\mathbb{P}}_{t_i}$  and  $\hat{\mathbb{P}}_{t_{i+1}}$ :

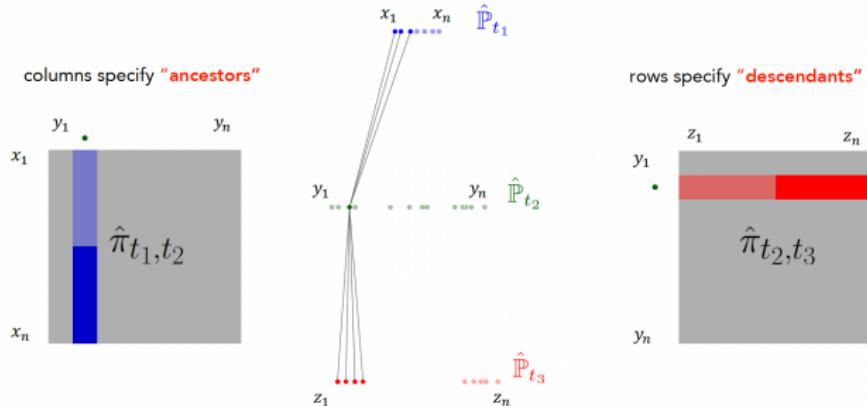
$$\begin{aligned} \gamma_{t_i, t_{i+1}}^* := \arg \min_{\pi} & \left\{ \underbrace{\mathbb{E}_{\pi} |X - Y|^2}_{\text{optimal transport}} + \underbrace{\sigma \Delta t_i \mathbf{H}(\pi)}_{\text{regularization}}, \right. \\ & \left. \text{s.t. } \int \pi(\cdot, dy) = \hat{\mathbb{P}}_{t_i}(\cdot), \int \pi(dx, \cdot) = \hat{\mathbb{P}}_{t_{i+1}}(\cdot) \right\}. \end{aligned}$$



# Coupling with Entropic optimal transport

- Schiebinger et al. (2019): use **entropic optimal transport** coupling between  $\hat{\mathbb{P}}_{t_i}$  and  $\hat{\mathbb{P}}_{t_{i+1}}$ :

$$\gamma_{t_i, t_{i+1}}^* := \arg \min_{\pi} \left\{ \underbrace{\mathbb{E}_{\pi} |X - Y|^2}_{\text{optimal transport}} + \underbrace{\sigma \Delta t_i \mathbf{H}(\pi)}_{\text{regularization}} := \sigma \Delta t_i \mathbf{H} \left( \pi | e^{-\frac{|X-Y|^2}{\sigma \Delta t_i}} \right), \right. \\ \left. \text{s.t. } \int \pi(\cdot, dy) = \hat{\mathbb{P}}_{t_i}(\cdot), \int \pi(dx, \cdot) = \hat{\mathbb{P}}_{t_{i+1}}(\cdot) \right\}.$$



# EOT captures Epigenetic landscape

Proposition (Connection between EOT and SDEs)

For  $X \sim \mu$  and  $Y \sim \nu$ :

$$\begin{aligned} & \min_{\pi} \left\{ \mathbf{H} \left( \pi | e^{-\frac{|x-y|^2}{\sigma \Delta t_i}} \right) \text{ s.t } \int \pi(x, dy) = \mu(x), \int \pi(dx, y) = \nu(y) \right\} \\ &= \min_P \left\{ \mathbf{H}(P|W^\sigma) \text{ s.t } P_{t_i} = \mu, P_{t_{i+1}} = \nu \right\} \quad (W^\sigma = \text{Brownian motion}) \end{aligned}$$

$\implies \gamma_{t_i, t_{i+1}}^*(x, \cdot)$  corresponds to a **transition probability kernel**,  
associated to an **optimal stochastic process** w.r.t the Brownian motion:

# EOT captures Epigenetic landscape

Proposition (Connection between EOT and SDEs)

For  $X \sim \mu$  and  $Y \sim \nu$ :

$$\begin{aligned} & \min_{\pi} \left\{ \mathbf{H} \left( \pi | e^{-\frac{|x-y|^2}{\sigma \Delta t_i}} \right) \text{ s.t } \int \pi(x, dy) = \mu(x), \int \pi(dx, y) = \nu(y) \right\} \\ &= \min_P \left\{ \mathbf{H}(P|W^\sigma) \text{ s.t } P_{t_i} = \mu, P_{t_{i+1}} = \nu \right\} \quad (W^\sigma = \text{Brownian motion}) \\ &= \min_P \left\{ \mathbf{H}(P|W^\sigma) \text{ s.t } P = \text{path-measure of a SDE,} \right. \\ &\quad \left. \text{with gradient drift, diffusion } \sigma, \text{ joining } \mu \text{ at } t_i \text{ to } \nu \text{ at } t_{i+1} \right\} \end{aligned}$$

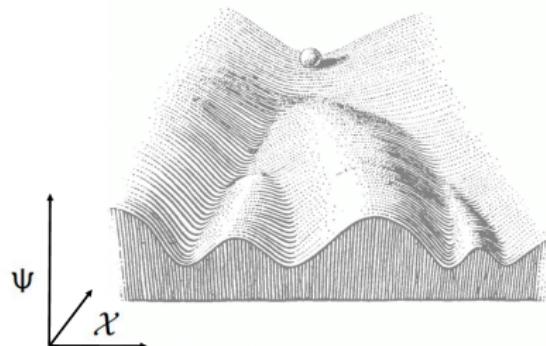
$\implies \gamma_{t_i, t_{i+1}}^*$  corresponds to **an optimal SDE with gradient drift!**

# EOT captures Epigenetic landscape

Proposition (Connection between EOT and SDEs)

$\gamma_{t_i, t_{i+1}}^* \leftrightarrow \min_P \{ \mathbf{H}(P|W^\sigma), \text{s.t } P = \text{path-measure of a SDE,}$   
*with gradient drift, diffusion  $\sigma$ , joining  $\mu$  at  $t_i$  to  $\nu$  at  $t_{i+1}$*  }.

Waddington's "Epigenetic Landscape"

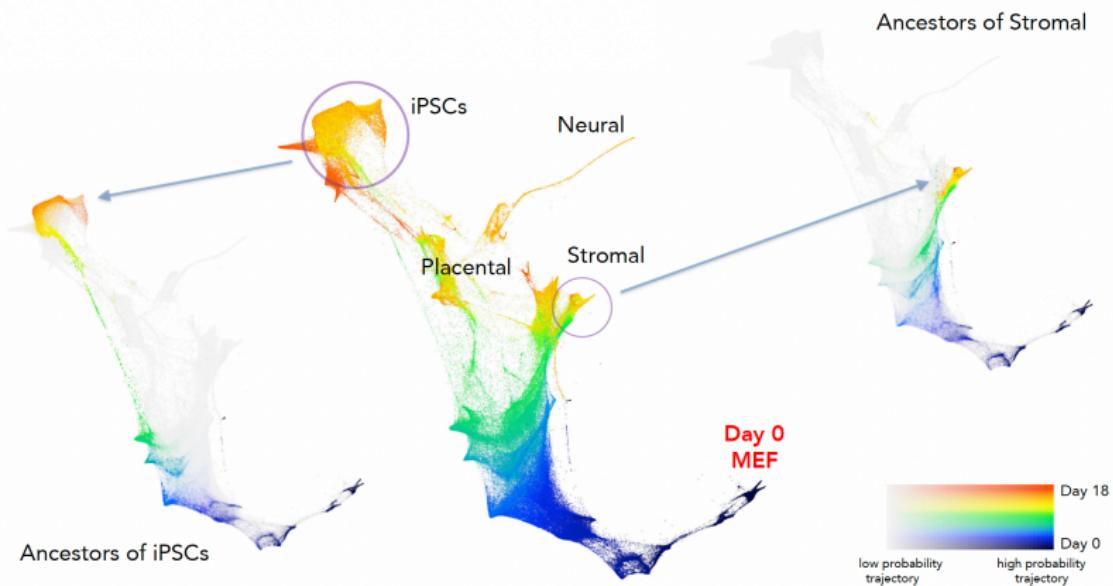


# Method for scRNA-seq data

- EOT is at the core of a popular method, **Waddington-OT**, that initiated a huge amount of works in single-cell analysis:
  - 1 fix  $\sigma$  by estimating the variance at each timepoint;
  - 2 solve EOT problem between each pair of timepoint;
  - 3 chain the optimal couplings to build trajectories.

## Step 1 of the TP

# Illustration with experimental scRNA-seq data



Schiebinger et al. 2019 - Cell.

# Predicting intermediate distributions

- A natural extension consists in fully reconstructing  $\mathbb{P}_t$  from these trajectories:

*How to predict biologically meaningful intermediate distributions between observed timepoints?*

⇒ Optimal transport provides a notion of **geodesics in the space of probability measures.**

# McCann interpolation (Wasserstein geodesics)

- Let  $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$  and  $\pi^*$  be an optimal transport plan for the quadratic cost.
- The **displacement interpolation** between  $\mu_0$  and  $\mu_1$  is defined by:

$$\mu_t := ((1-t)x + ty)_{\#} \pi^*, \quad t \in [0, 1].$$

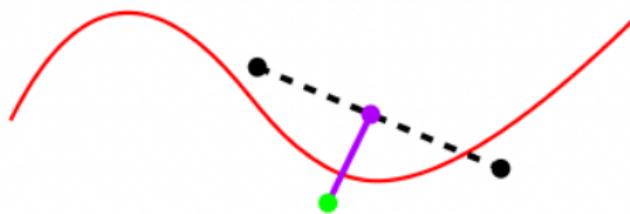
- $(\mu_t)_{t \in [0,1]}$  is a **constant-speed geodesic** for the Wasserstein-2 distance:

$$W_2(\mu_s, \mu_t) = |t - s| W_2(\mu_0, \mu_1).$$

- It corresponds exactly to the distribution arising from the **transport equation** associated to the OT problem between  $\mu_0$  and  $\mu_1$  (see Benamou-Brenier formula).

# Interpretation for trajectory inference

- The McCann interpolation provides a principled way to interpolate between empirical marginals.
- It provides a way of **testing predicted trajectories** by comparing interpolations learned from two distributions and the intermediate observation:



Step 2 of the TP

Now go on Github

<https://github.com/eliasventre/OT-cell-biology-TP1>

# Beyond gene expression: spatial structure

- In many datasets, cells are not only described by gene expression, but also by:
  - spatial positions,
  - tissue architecture,
  - cell-cell proximity graphs.
- Classical OT ignores this relational structure.  
⇒ Need for **structure-aware optimal transport**.

# Fused Gromov–Wasserstein transport

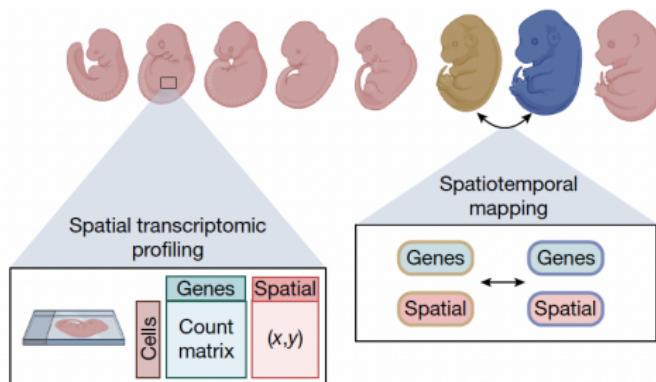
- Let:
  - $(X, C_X)$  and  $(Y, C_Y)$  be metric measure spaces (e.g. spatial graphs),
  - $M$  a feature-based cost (gene expression).
- The **Fused Gromov–Wasserstein** problem:

$$\min_{\pi} (1 - \alpha) \langle M, \pi \rangle + \alpha \sum_{i,j,k,l} |(C_X)_{ij} - (C_Y)_{kl}|^2 \pi_{ik} \pi_{jl}.$$

- Interpolates between:
  - feature matching (OT),
  - relational matching (GW).

# Application: moscot

- The **MOSCOT** framework (Klein et al., 2025) extend WOT, using FGW to:
  - align cells across time and space,
  - respect tissue geometry,
  - infer developmental trajectories.

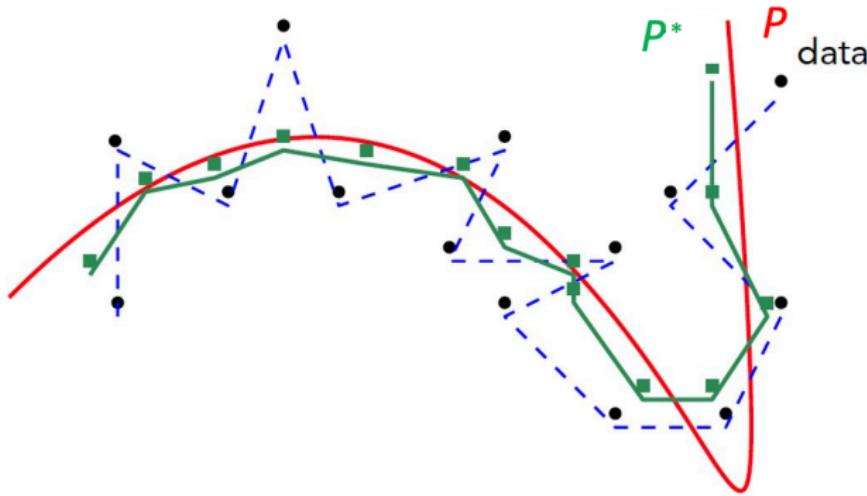


**Step 3 of the TP**

## Session 2: Tackling noise in marginals

# Need for regularization in case of noisy data

⇒ Need to **regularize the optimizer if the data are noisy**  
(= not enough cells per timepoint).



# Path-measure reconstruction (gWOT)

- **Method** = approximate  $\mathbb{P}$  by

$$P^{\lambda, N} := \arg \min_R \mathcal{F}(R) := \underbrace{\sigma \mathbf{H}(R | W^\sigma)}_{\text{regularization term}} + \underbrace{\frac{1}{\lambda} \sum_{i=1}^N \Delta t_i \mathbf{H}(\hat{\mathbb{P}}_{t_i} | R_{t_i})}_{\text{data fitting term}}$$

- $\lambda$  scales the **data fitting term**;
- $N$  is the **number of timepoints**.

⇒ The regularization term prevents from **overfitting**.

# Path-measure reconstruction

Proposition (Lavenant et al. 2021)

$$\lim_{\lambda \rightarrow 0, N \rightarrow \infty} P^{\lambda, N} = \mathbb{P}.$$

⇒ The convergence holds **regardless of the number of cells in  $\hat{\mathbb{P}}_{t_i}$ !**

- In practice, we compute directly the sequence of **optimal marginals**,

$$(\mathbb{P}_{t_1}^{\lambda, N}, \dots, \mathbb{P}_{t_N}^{\lambda, N}),$$

using i) the **analogy between relative entropy and EOT**, and ii) **particle systems for optimization over measures**.

# Optimization over probability measures

- In Session 2, we aim to minimize the functional:

$$\mathcal{F}(\mu_1, \dots, \mu_N) \quad \text{with } \mu_i \in \mathcal{P}(\mathbb{R}^d).$$

- This is a **nonlinear optimization problem in the space of measures**.
- Wasserstein space provides:
  - a geometry,
  - a notion of gradient,
  - a gradient flow interpretation.

# Mean-Field Langevin viewpoint

- Represent a measure  $\mu$  by an empirical particle system:

$$\mu^m = \frac{1}{m} \sum_{j=1}^m \delta_{X^j}.$$

- Perform **stochastic gradient descent** on particles:

$$X_{k+1}^j = X_k^j - \eta \nabla \frac{\delta \mathcal{F}}{\delta \mu}(\mu_k)(X_k^j) + \sqrt{2\eta} \xi_k^j.$$

- As  $m \rightarrow \infty$ :

⇒ convergence to a **gradient flow in measure space**.

# Gradient descent

- **Numerical method** for minimizing the previous functional: the family  $(\mu_k^{(1)}, \dots, \mu_k^{(N)})$  defined by **the stochastic gradient descent** system:

$$\begin{cases} \hat{X}_j^{(i)}[k+1] = \hat{X}_j^{(i)}[k] - \eta \nabla V^{(i)}[\hat{\mu}[k]](\hat{X}_j^{(i)}[k]) + \sqrt{2\eta(\tau + \epsilon)} Z_{j,k}^{(i)}, & \hat{X}_j^{(i)}[0] \stackrel{iid}{\sim} \mu_0^{(i)} \\ \hat{\mu}^{(i)}[k] = \frac{1}{m} \sum_{j=1}^m \delta_{\hat{X}_j^{(i)}[k]}, & i \in [T] \end{cases}$$

converges to the sequence of optimal marginals  $(\mathbb{P}_{t_1}^{\lambda,N}, \dots, \mathbb{P}_{t_N}^{\lambda,N})$  as  $k, m \rightarrow \infty$  (Chizat et al. 2022).

⇒ This provides a computable way of estimating **the optimal empirical distributions!**

# Differentiable Sinkhorn in practice

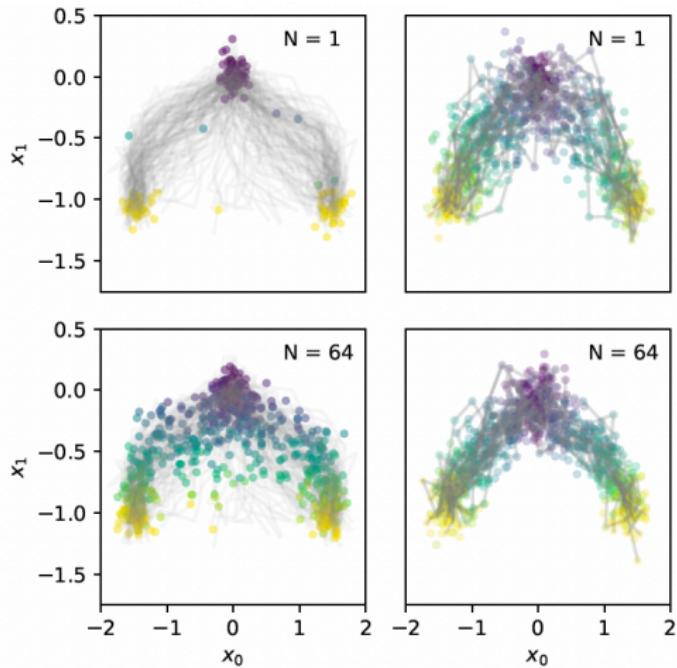
- Entropic OT admits a smooth dual formulation.
- The Sinkhorn algorithm defines a **differentiable map**:

$$(\mu, \nu) \mapsto \gamma_\varepsilon^*(\mu, \nu).$$

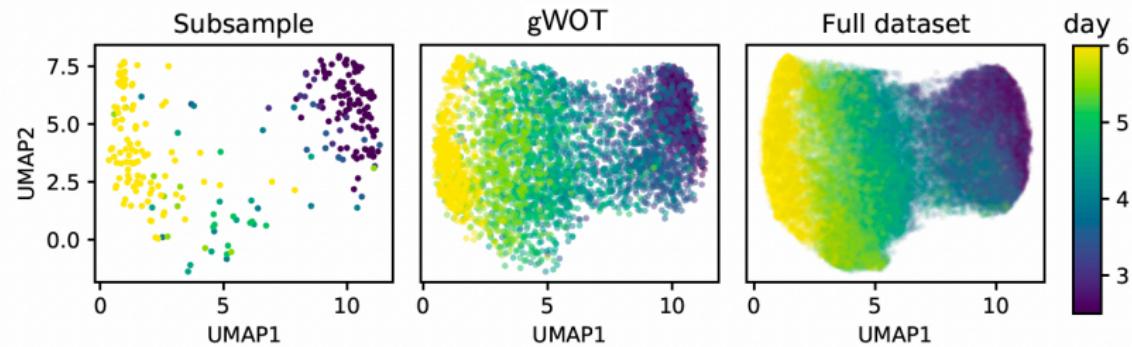
- In practice:
  - gradients computed via automatic differentiation,
  - enables end-to-end optimization,
  - scalable to large datasets.

⇒ This will be the computational backbone of Session 2.

# Application to simulated data



# Application to experimental data (iPSC reprogramming)



Chizat et al. 2022 - NeurIPS

# References

## Session 1:

- Schiebinger, G., Shu, J., Tabaka, M., ... & Lander, E. S. (2019). **Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming.** Cell.
- Klein, D., Palla, G., Lange, M., ... & Theis, F. J. (2025). **Mapping cells through time and space with moscot.** Nature.

## Session 2:

- Lavenant, H., Zhang, S., Kim, Y. H., Schiebinger, G. (2021). **Towards a mathematical theory of trajectory inference.** arXiv.
- Chizat, L., Zhang, S., Heitz, M., Schiebinger, G. (2022). **Trajectory inference via mean-field langevin in path space.** NeurIPS.