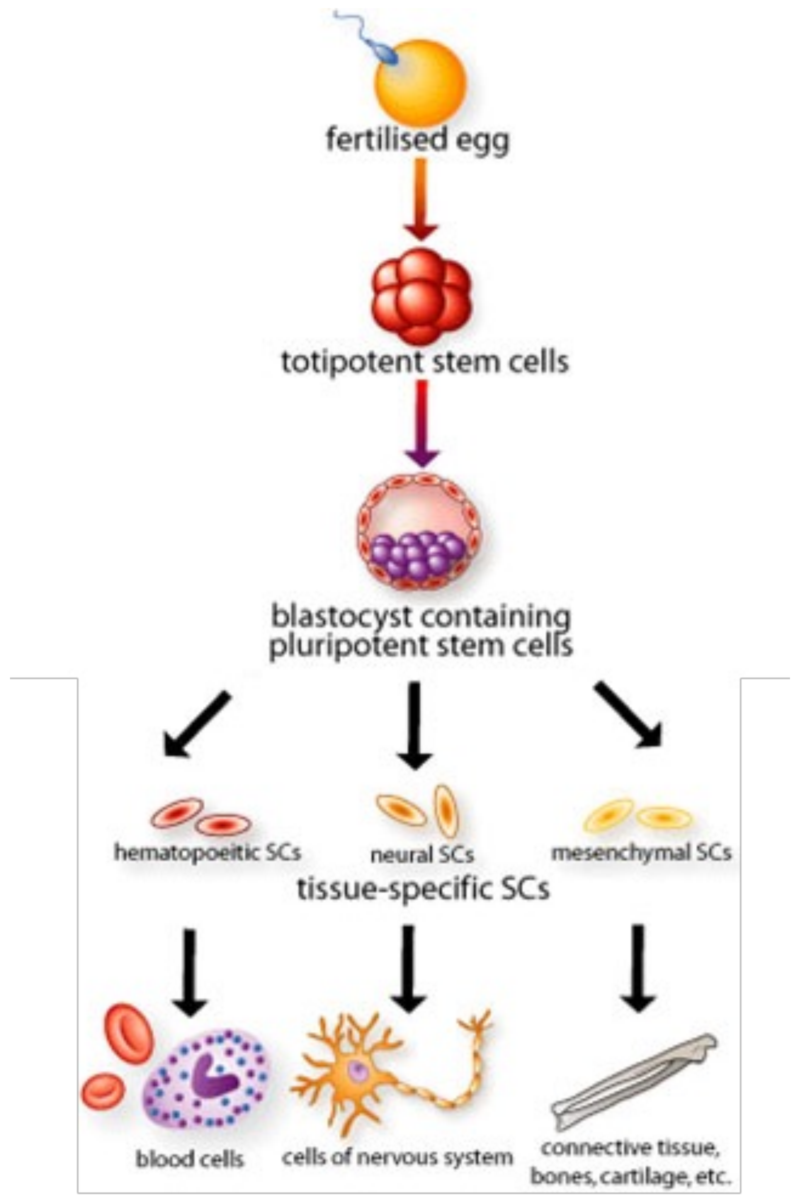
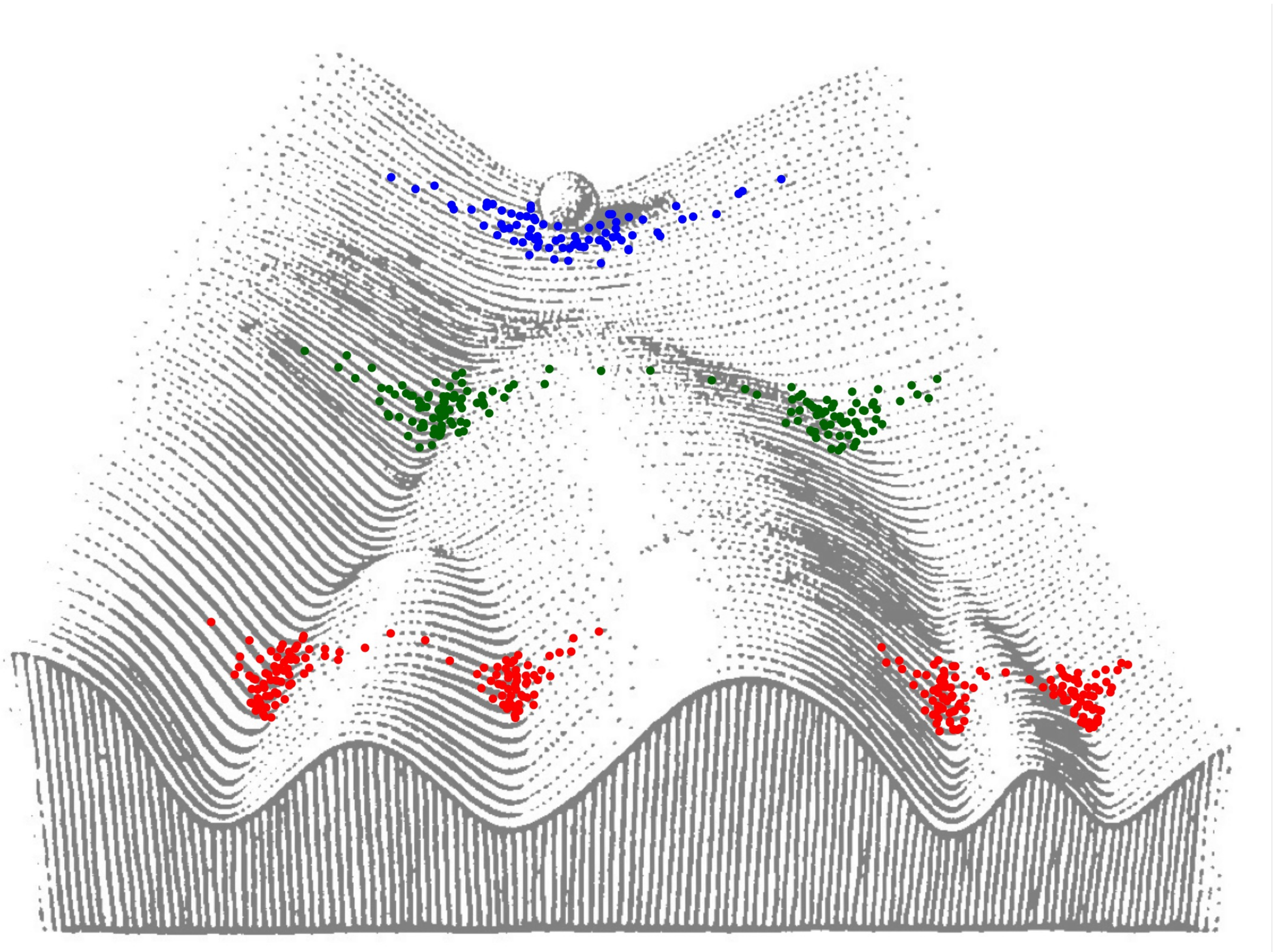
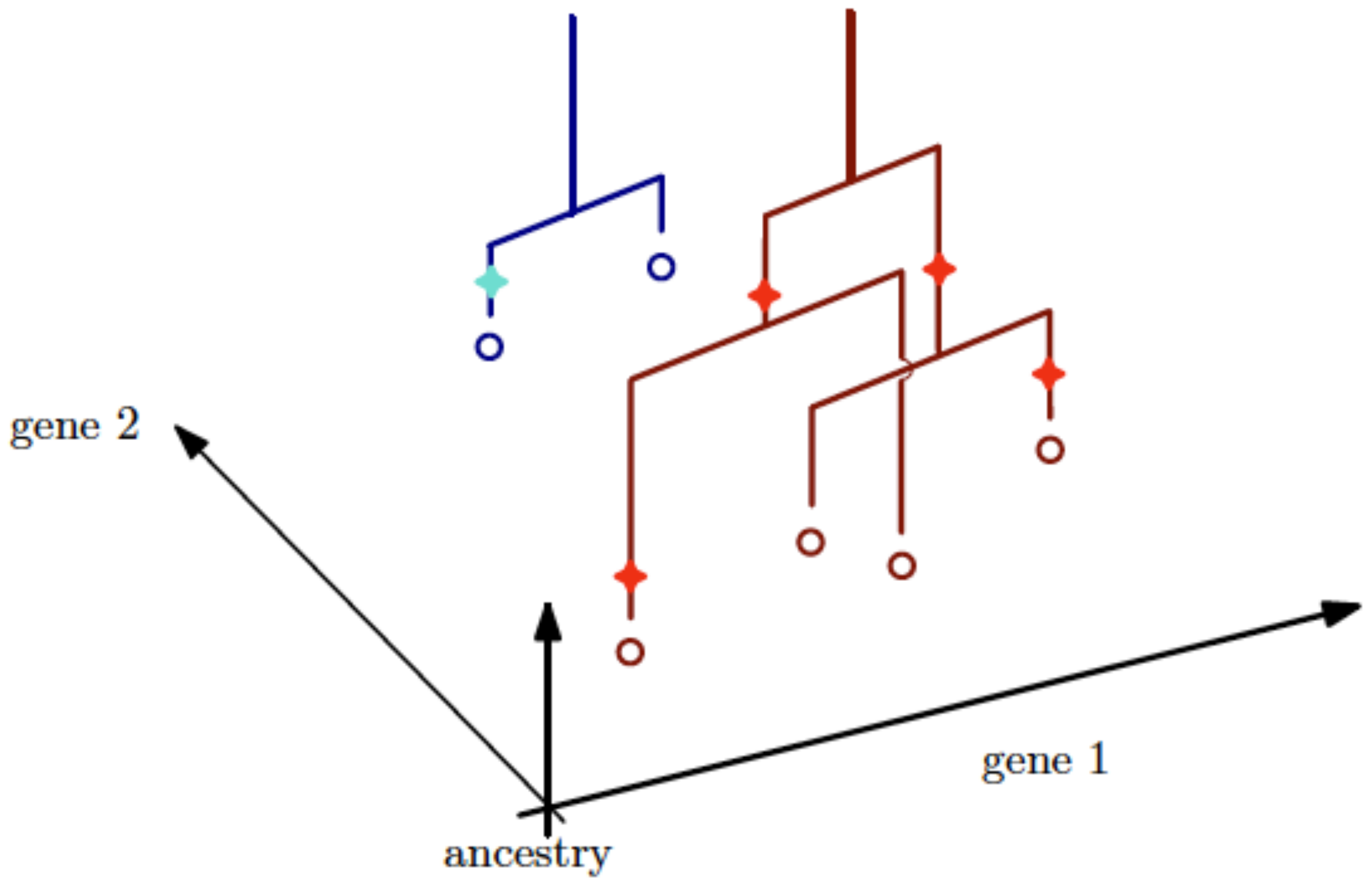
Towards a Mathematical Theory of Developmental Biology

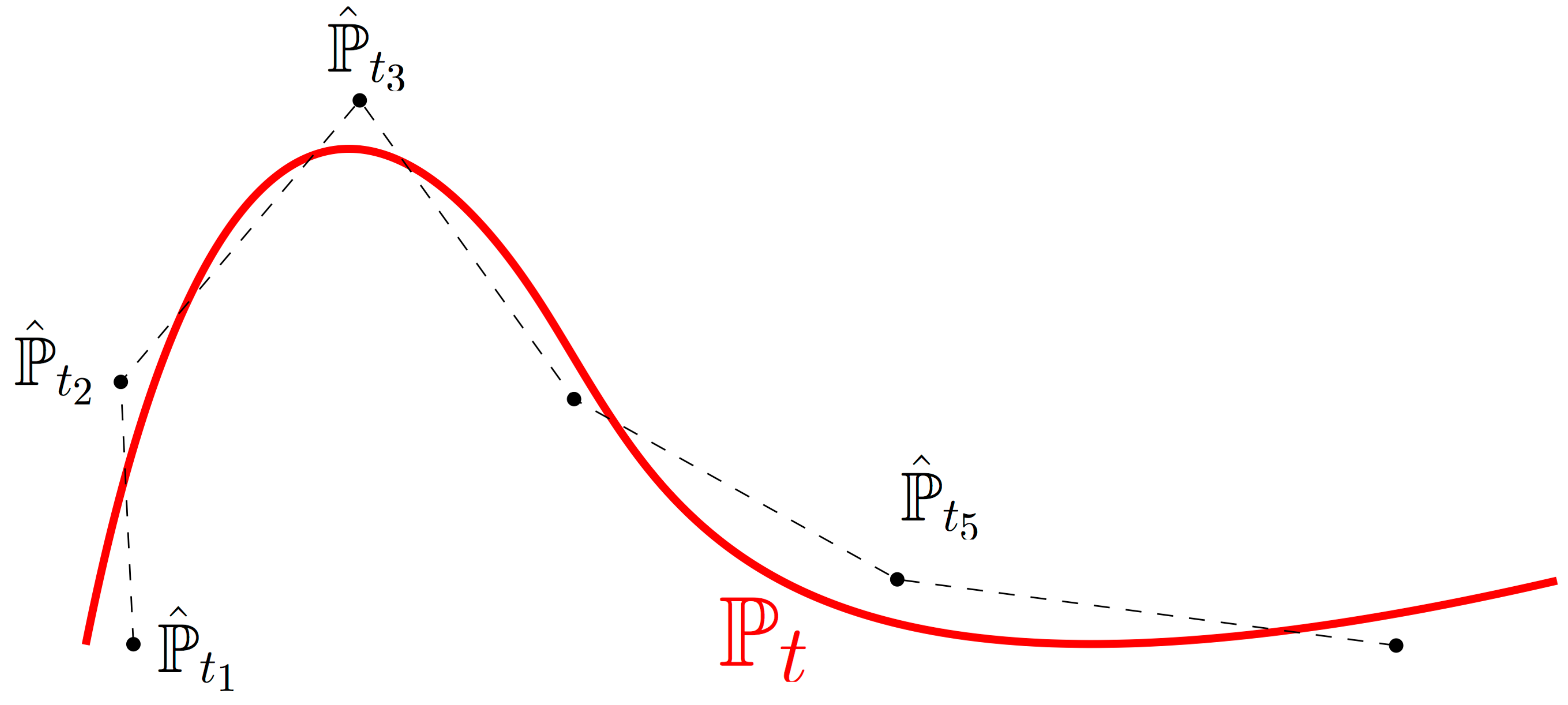


Geoffrey Schiebinger

University of British Columbia



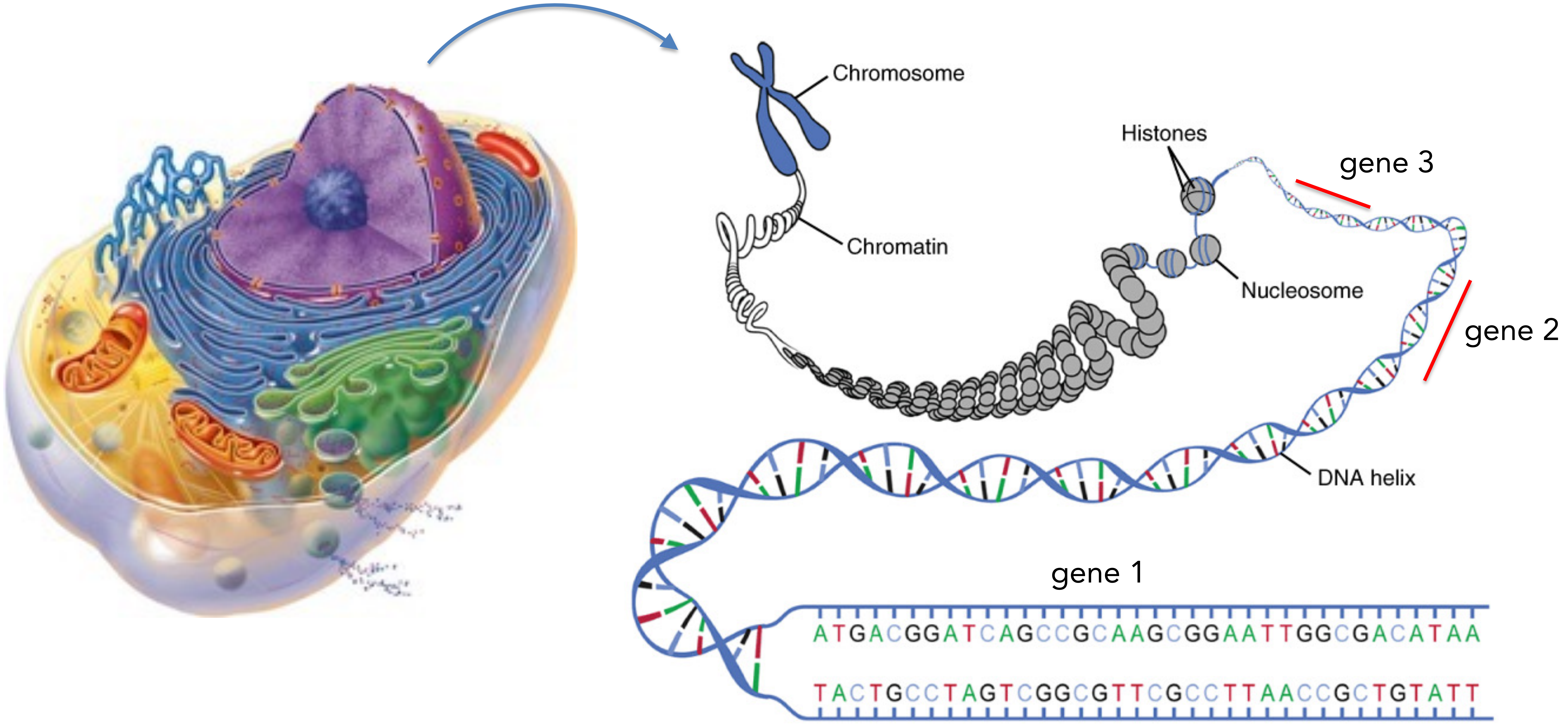




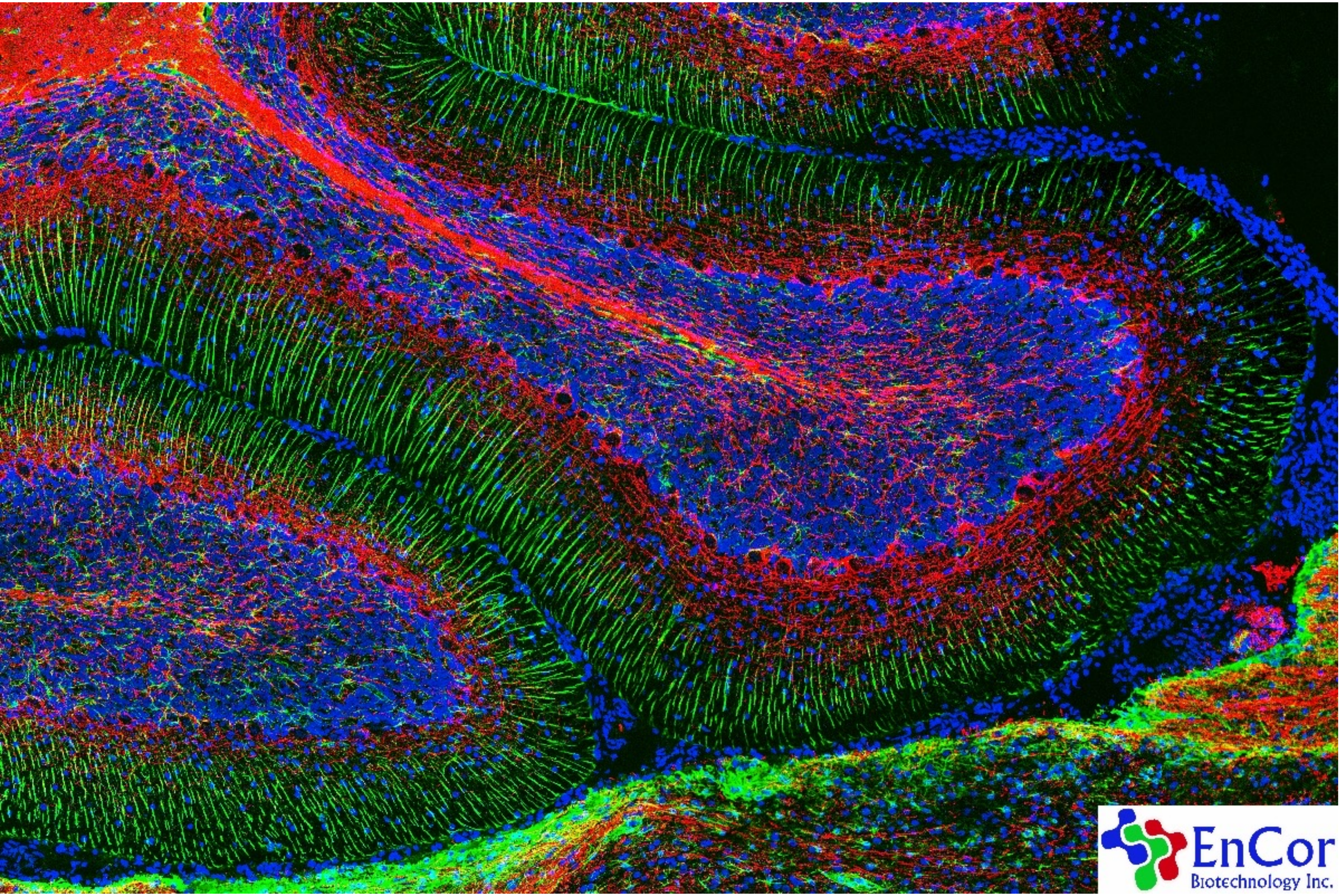
PIMS Kantorovich Summer School on Optimal Transport

Our bodies are made of 20 trillion cells

DNA stores the genetic code in 20,000 genes



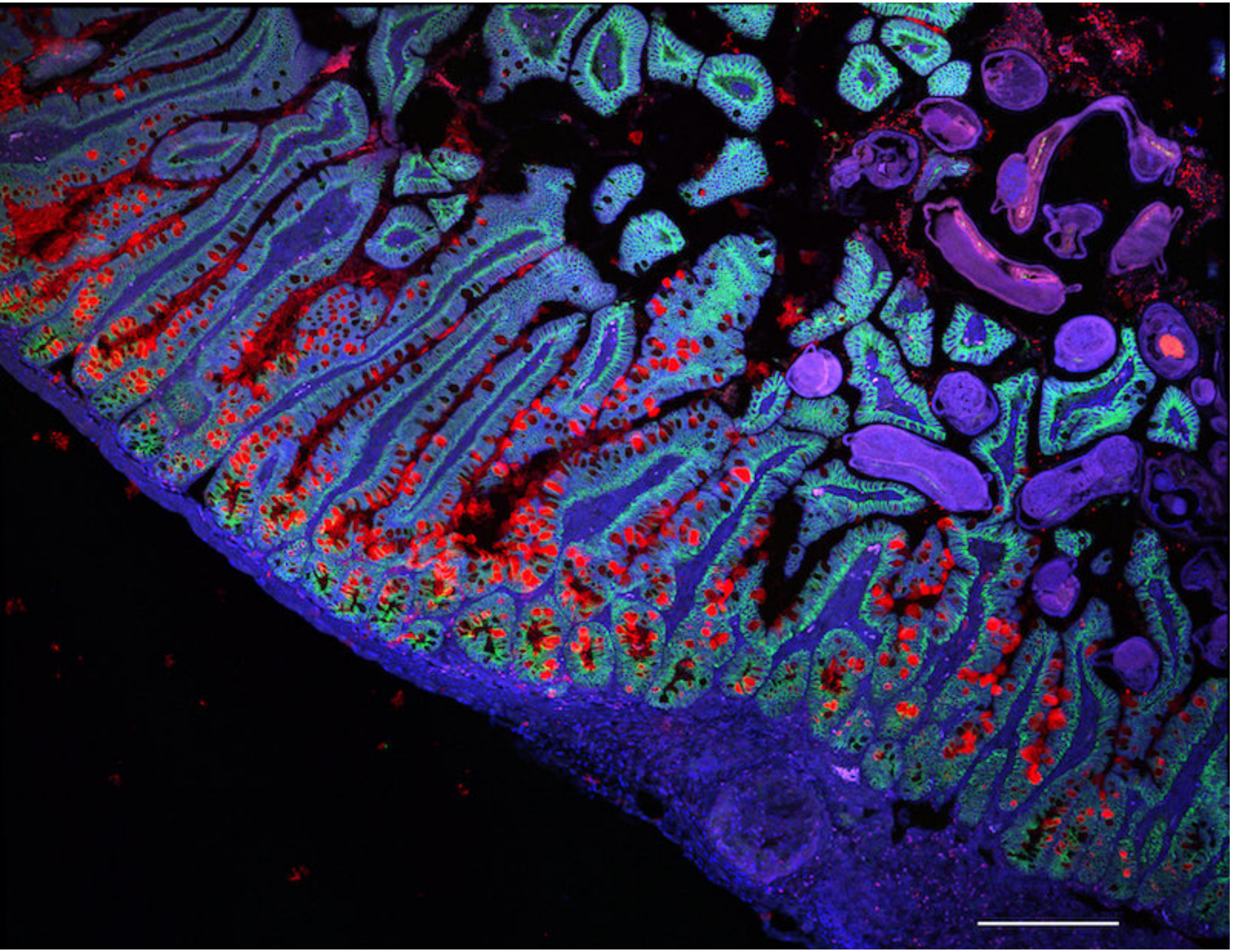
Each cell has the same DNA but expresses different genes



Cells in the brain perform different functions:

Neurons process information
Astrocytes support neurons

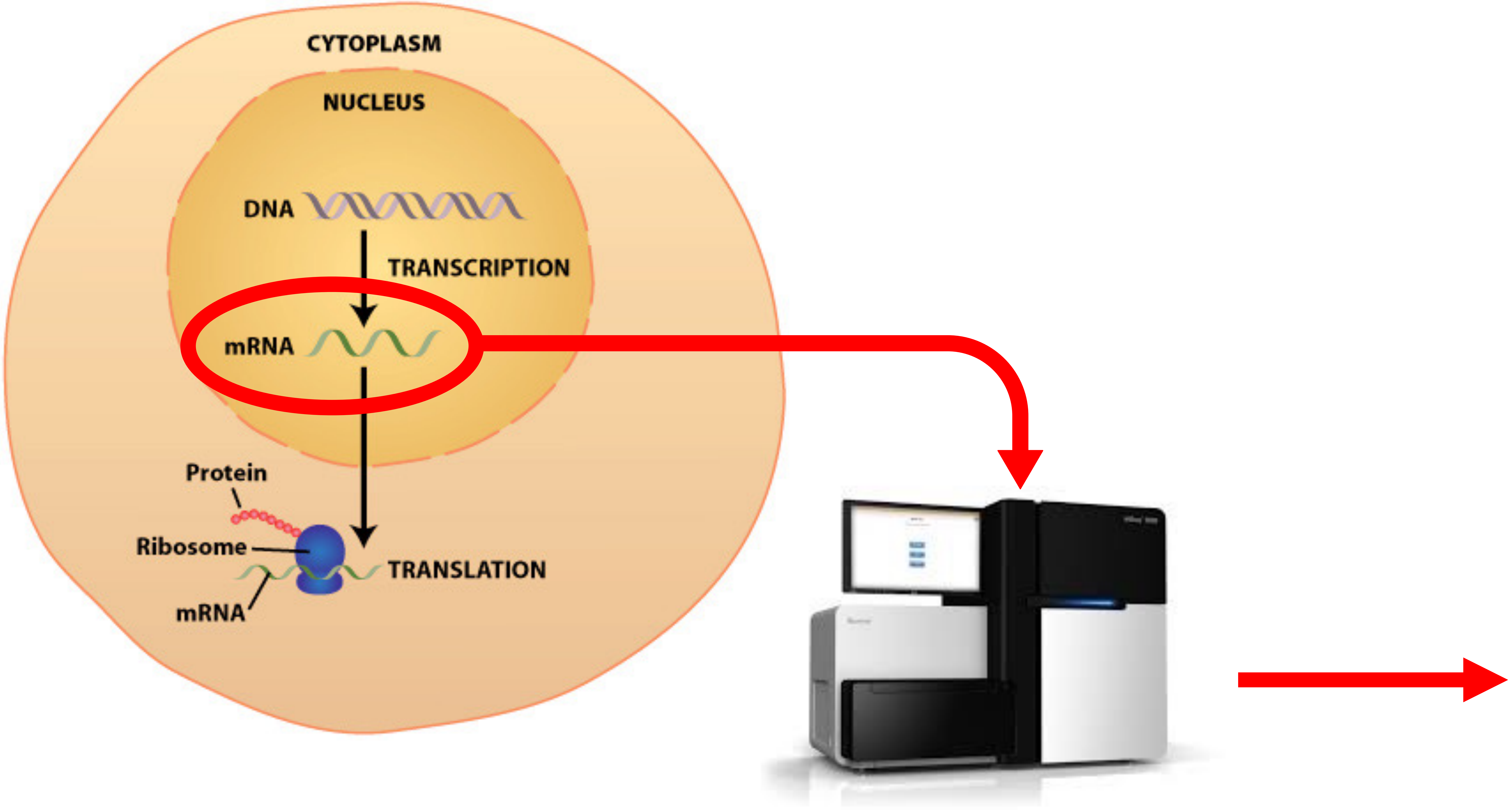
Each cell has the same DNA but expresses different genes



The intestine is the most regenerative organ in the human body.

It completely regenerates the lining every 5 days!

Single cell RNA sequencing

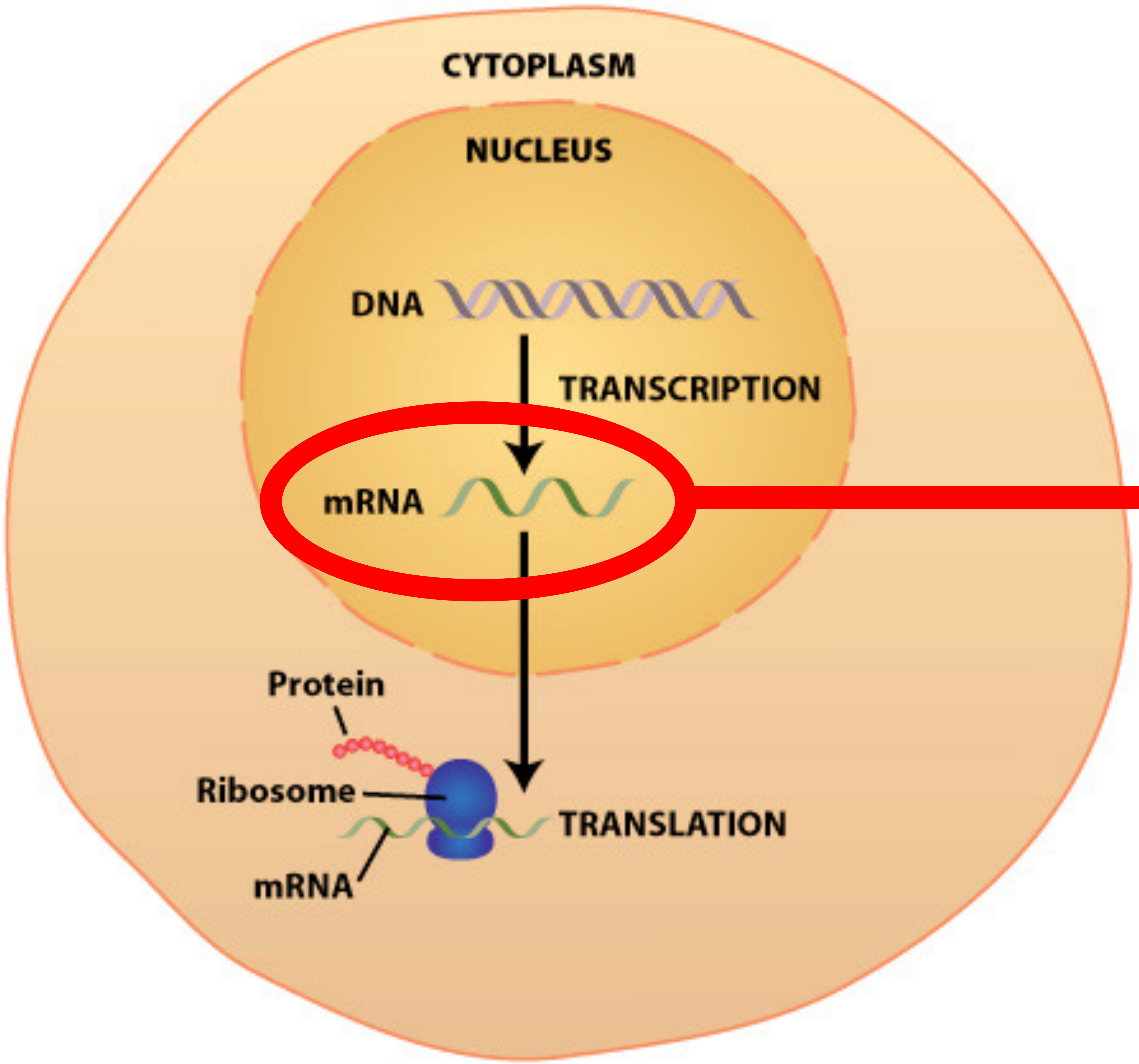


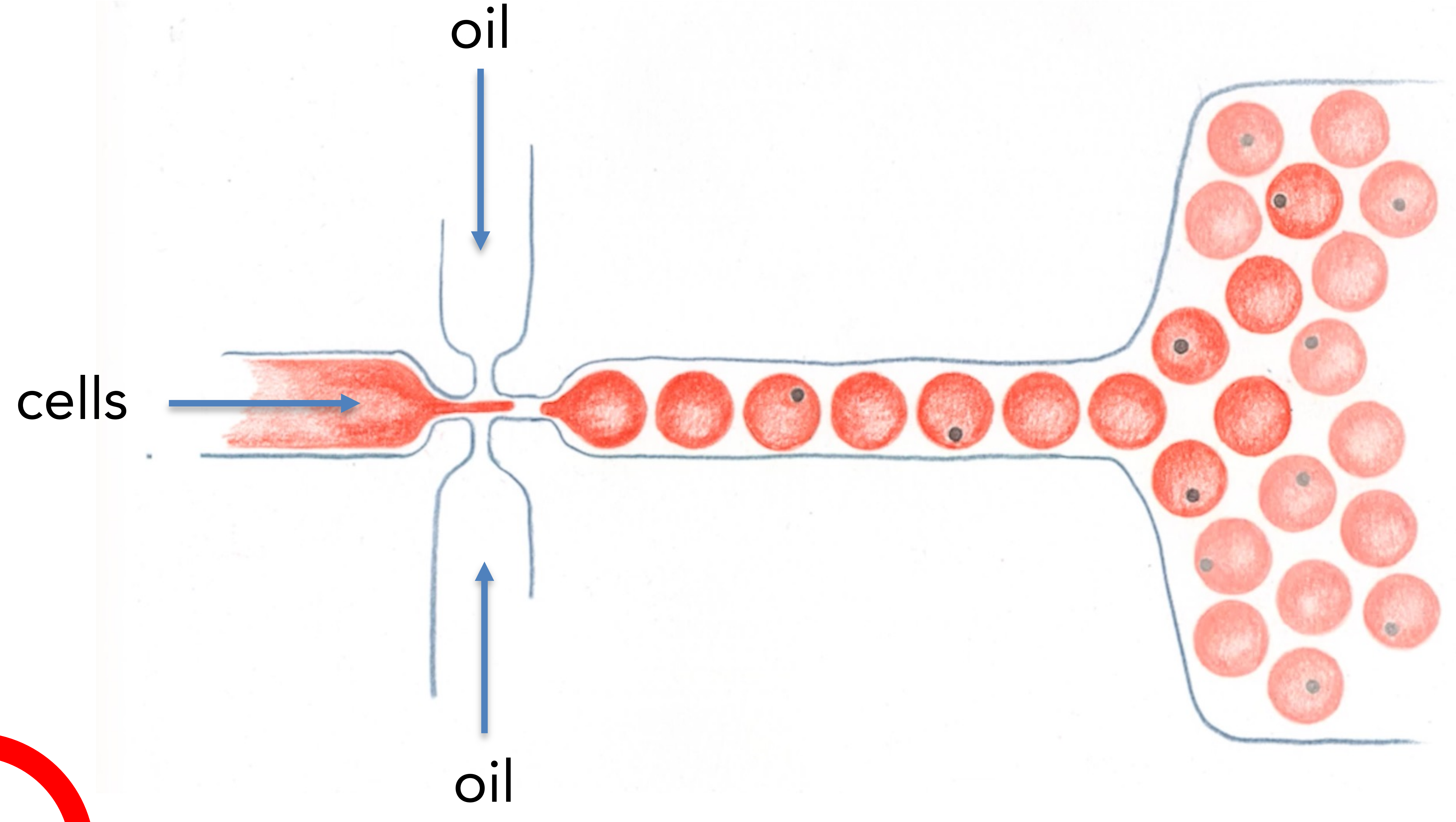
single cell RNA-Sequencing

Single cell expression profile

$$\begin{bmatrix} \text{num. mRNA gene}_1 \\ \text{num. mRNA gene}_2 \\ \vdots \\ \text{num. mRNA gene}_G \end{bmatrix}$$

Single cell RNA sequencing

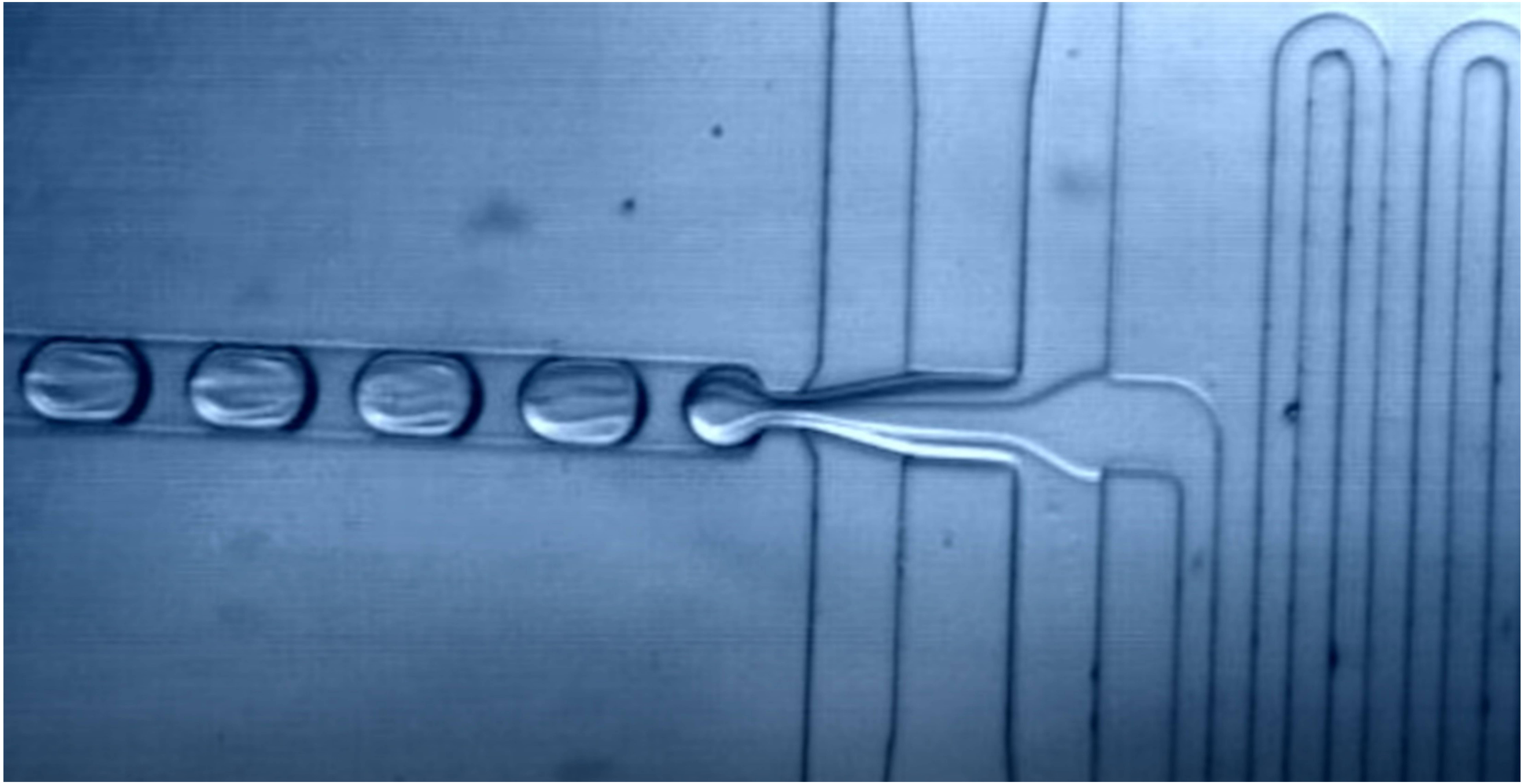




single cell RNA-Sequencing

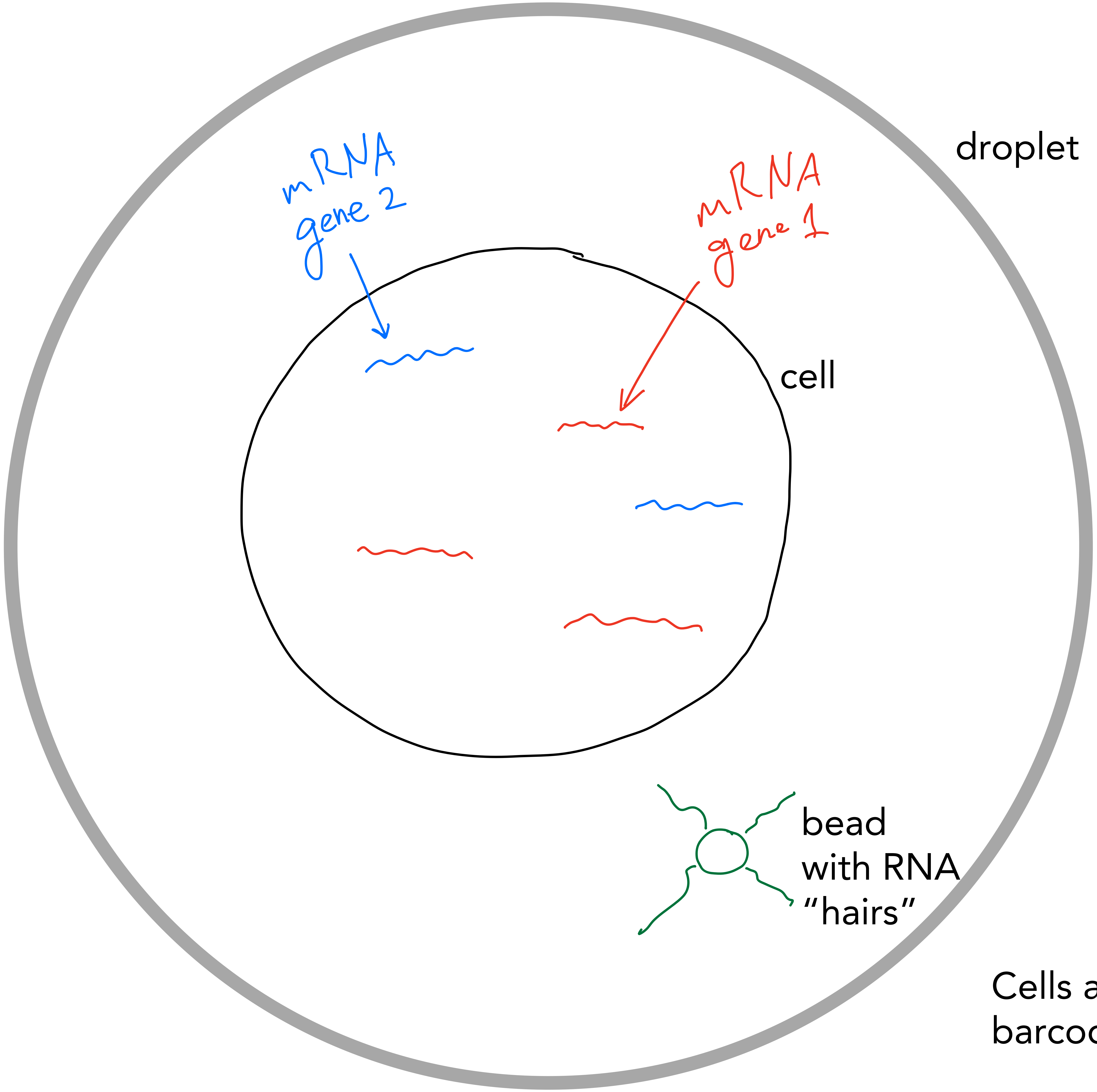
Single cell expression profile

$$\begin{bmatrix} \text{num. mRNA gene}_1 \\ \text{num. mRNA gene}_2 \\ \vdots \\ \text{num. mRNA gene}_G \end{bmatrix}$$



oil

droplet



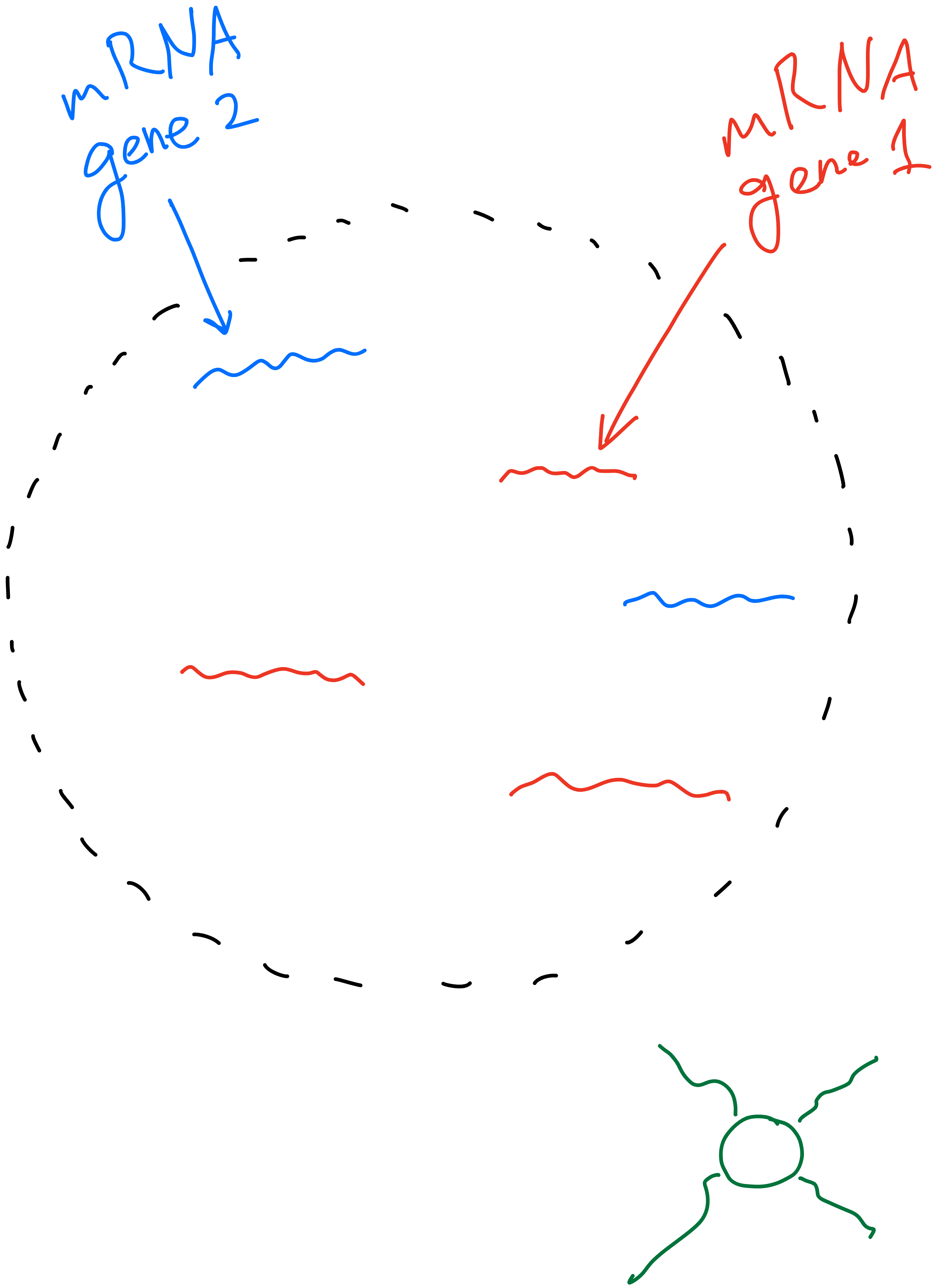
cell

oil

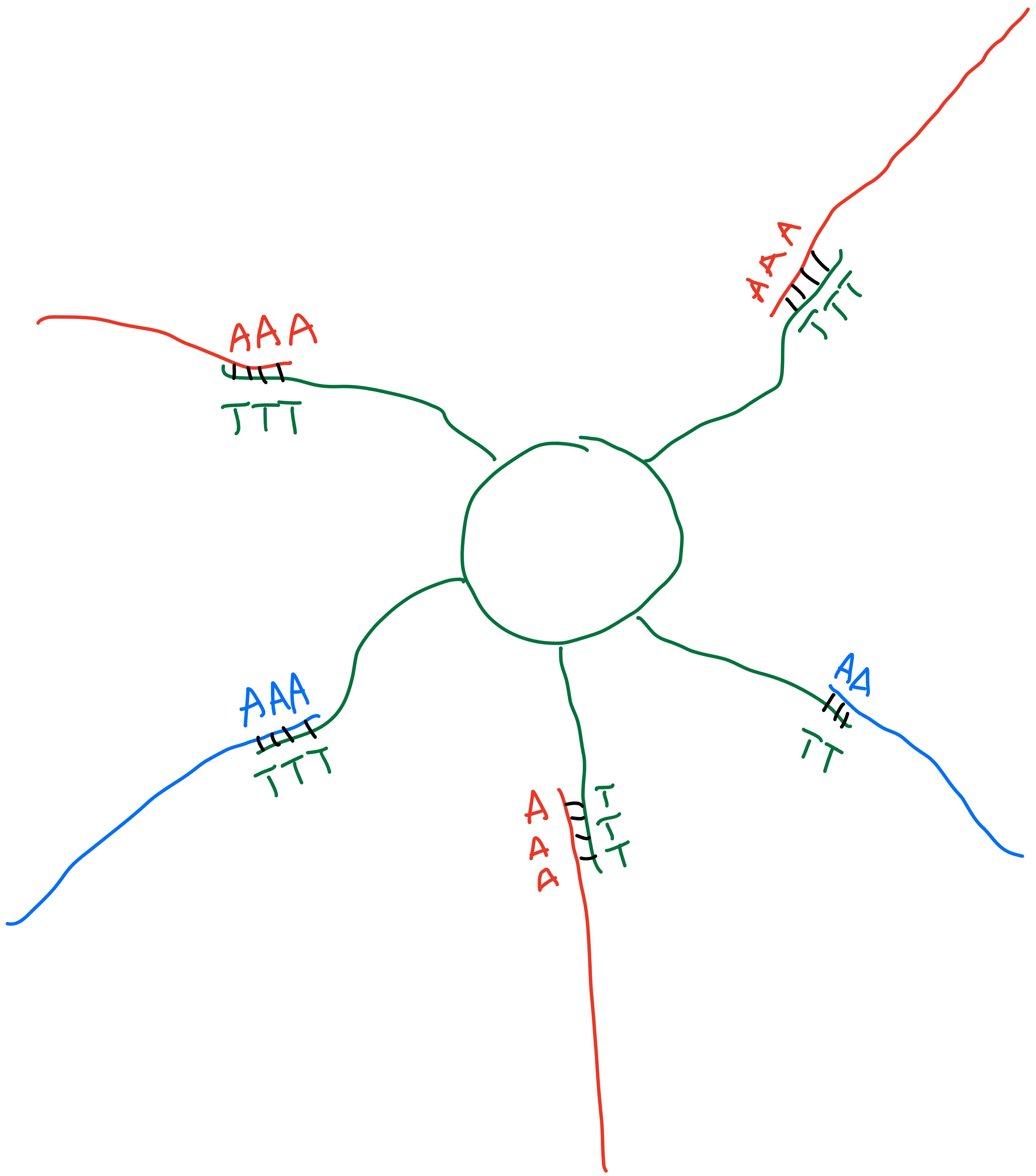
oil

bead
with RNA
“hairs”

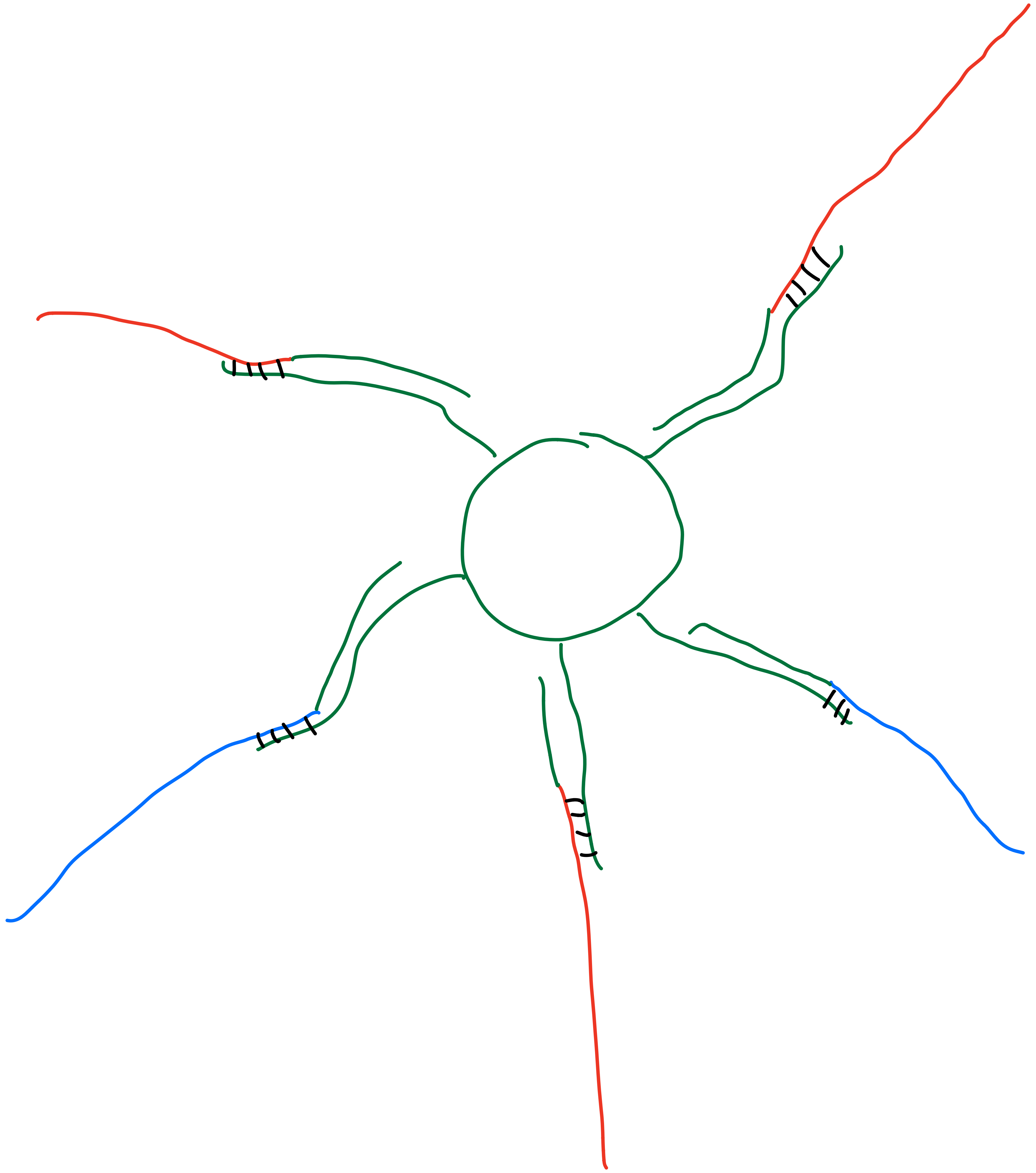
Cells are isolated together with barcoded beads in dropletts



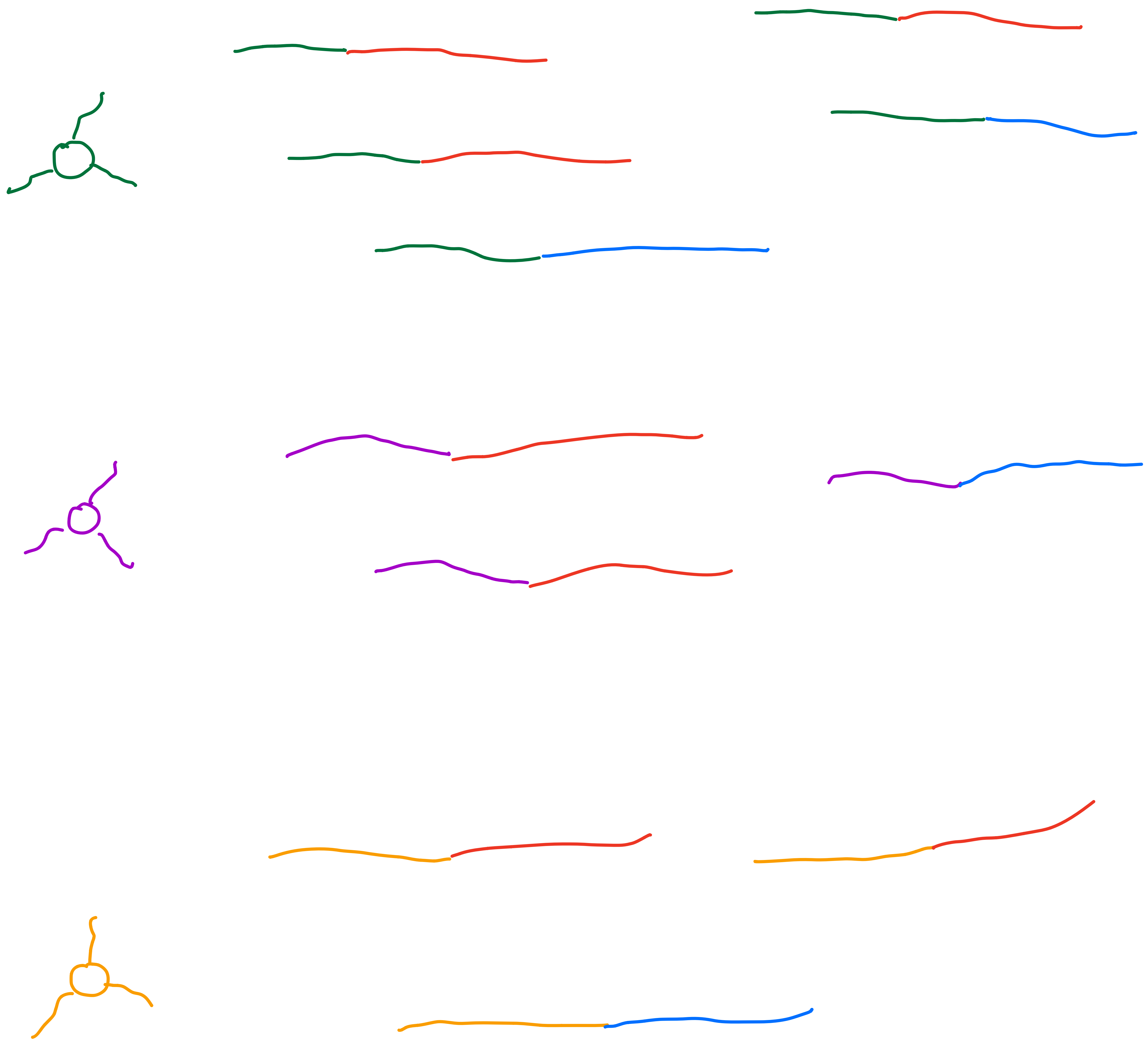
Cell membranes are lysed and
mRNA is captured by beads



mRNA is captured on beads



mRNA molecules are extended to include bead barcodes



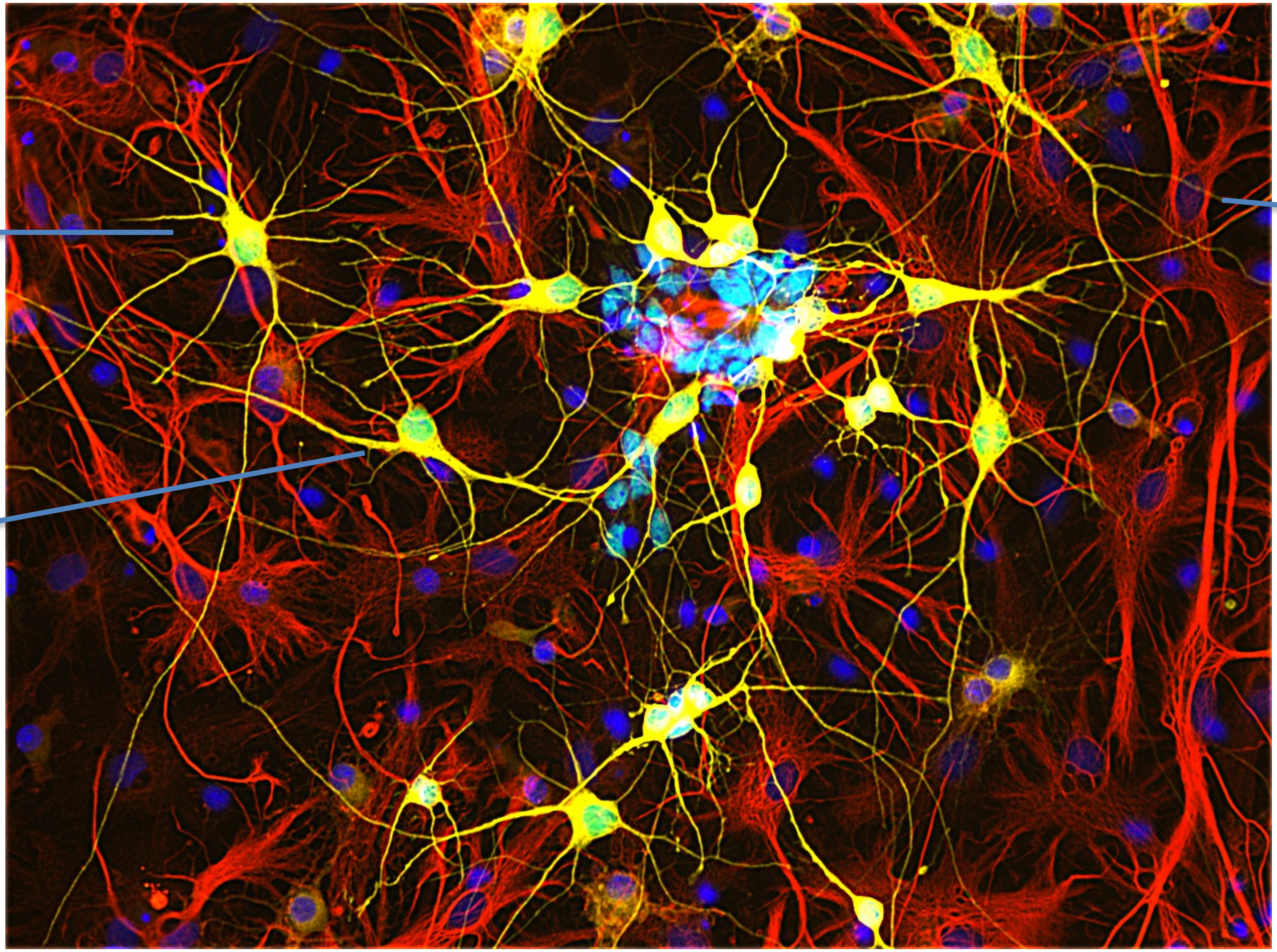
Barcoded mRNA molecules are then pooled together and sequenced

Single cell RNA seq reveals the vectors of cell identity

Neurons

$$\begin{bmatrix} 51 \\ 0 \\ 0 \\ 21 \end{bmatrix}$$

$$\begin{bmatrix} 49 \\ 0 \\ 0 \\ 19 \end{bmatrix}$$

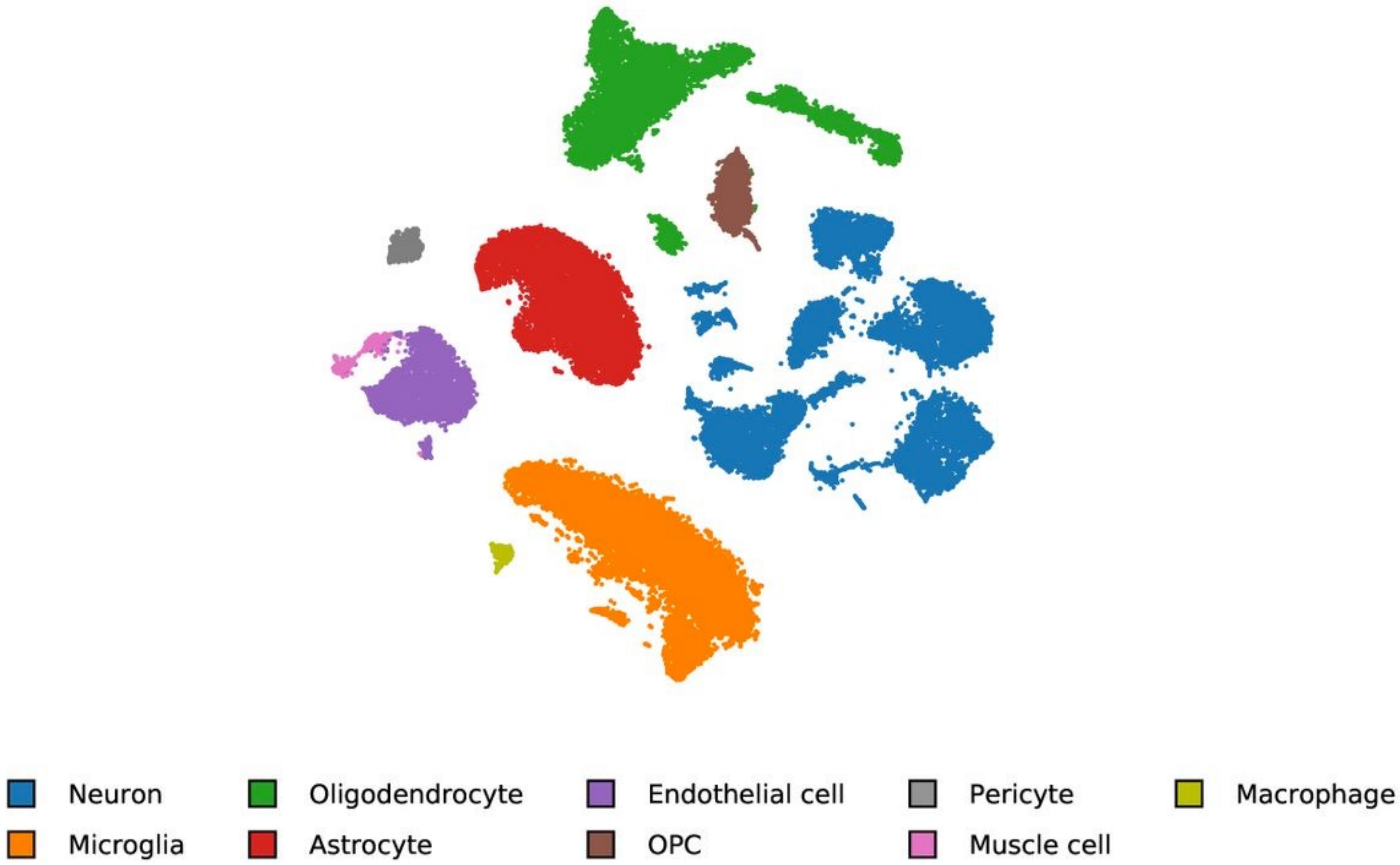


Astrocytes

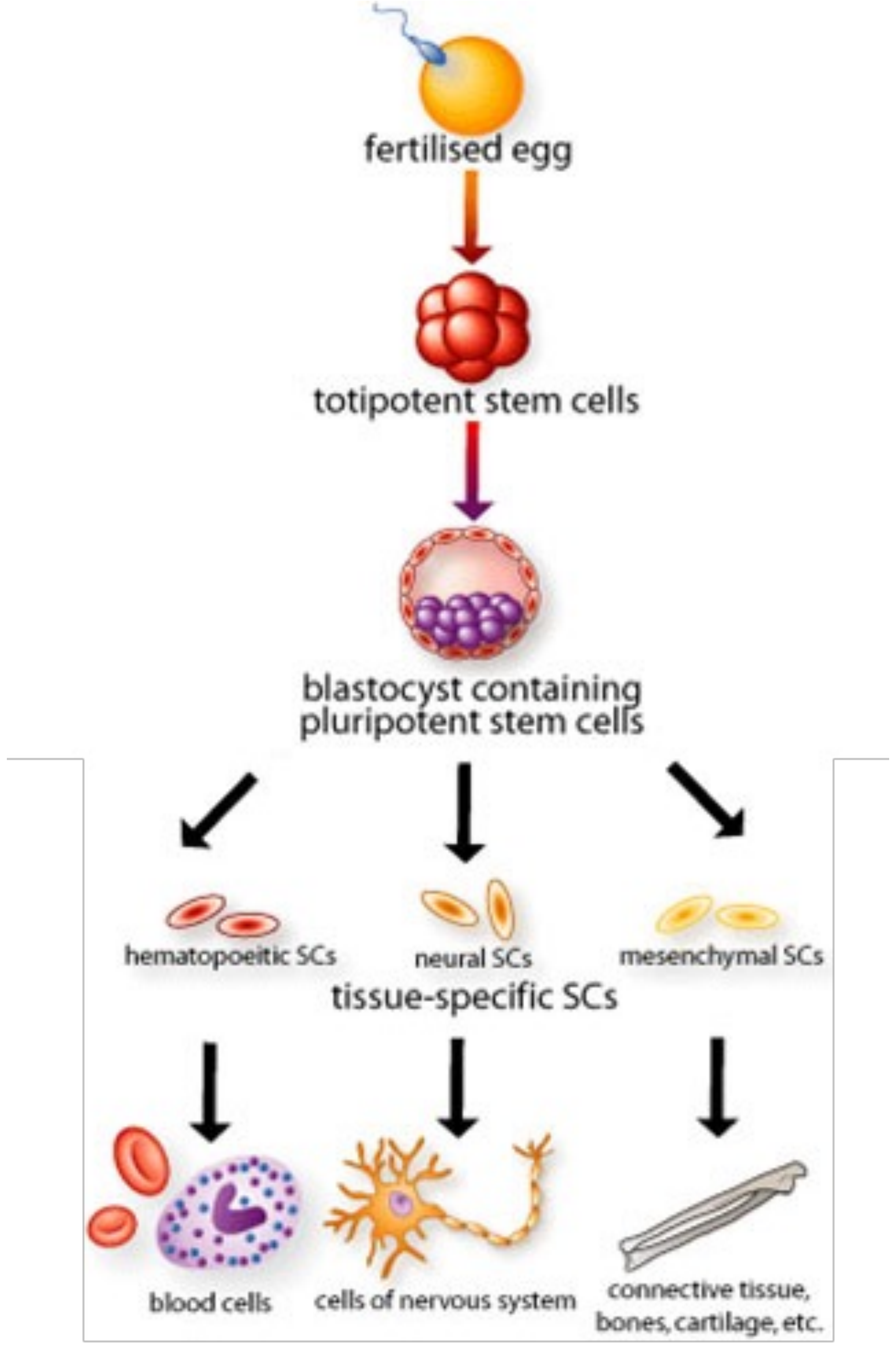
$$\begin{bmatrix} 0 \\ 18 \\ 101 \\ 3 \end{bmatrix}$$

Clustering groups cells by cell type

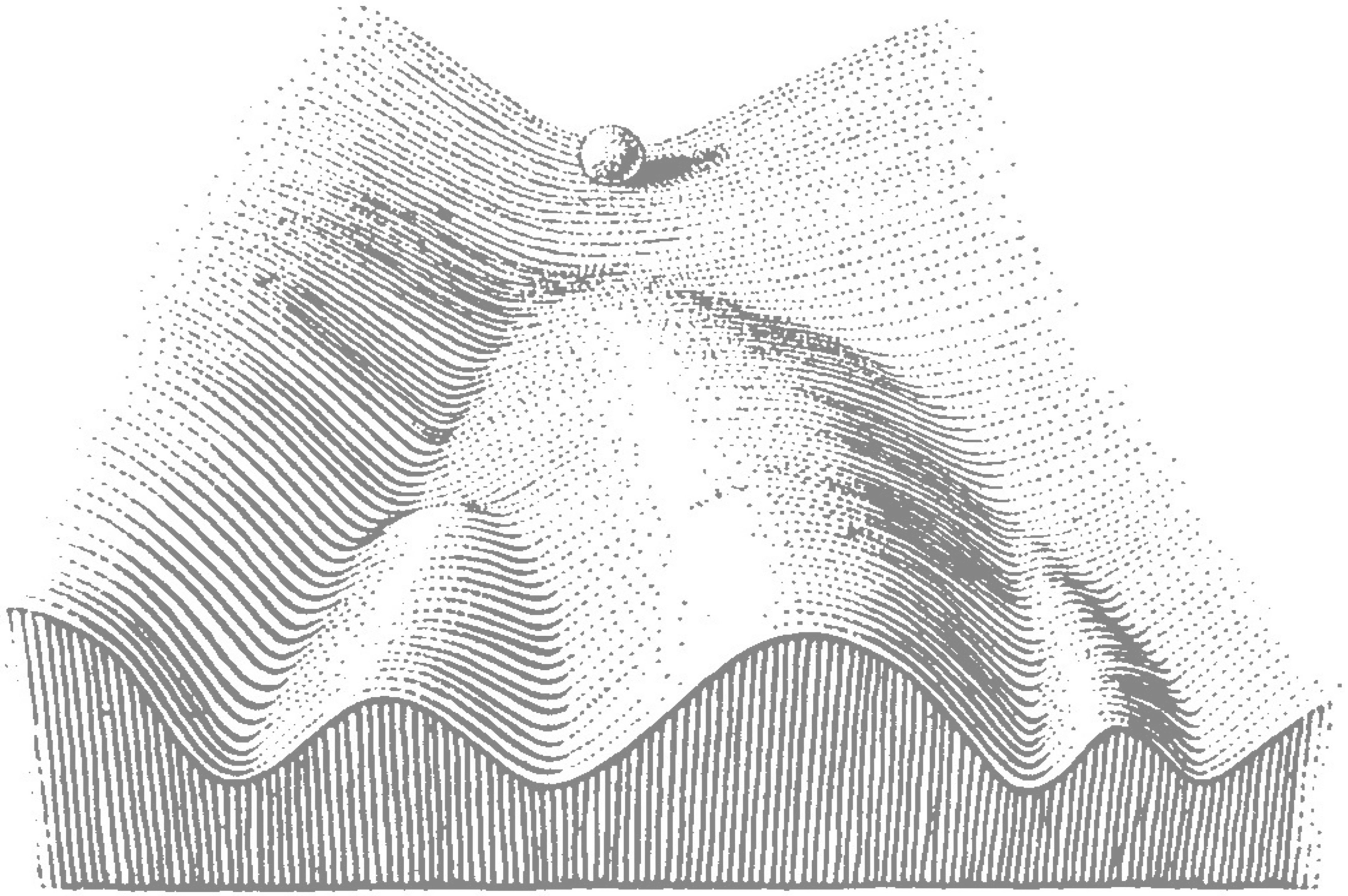
Differential expression reveals identity of each type



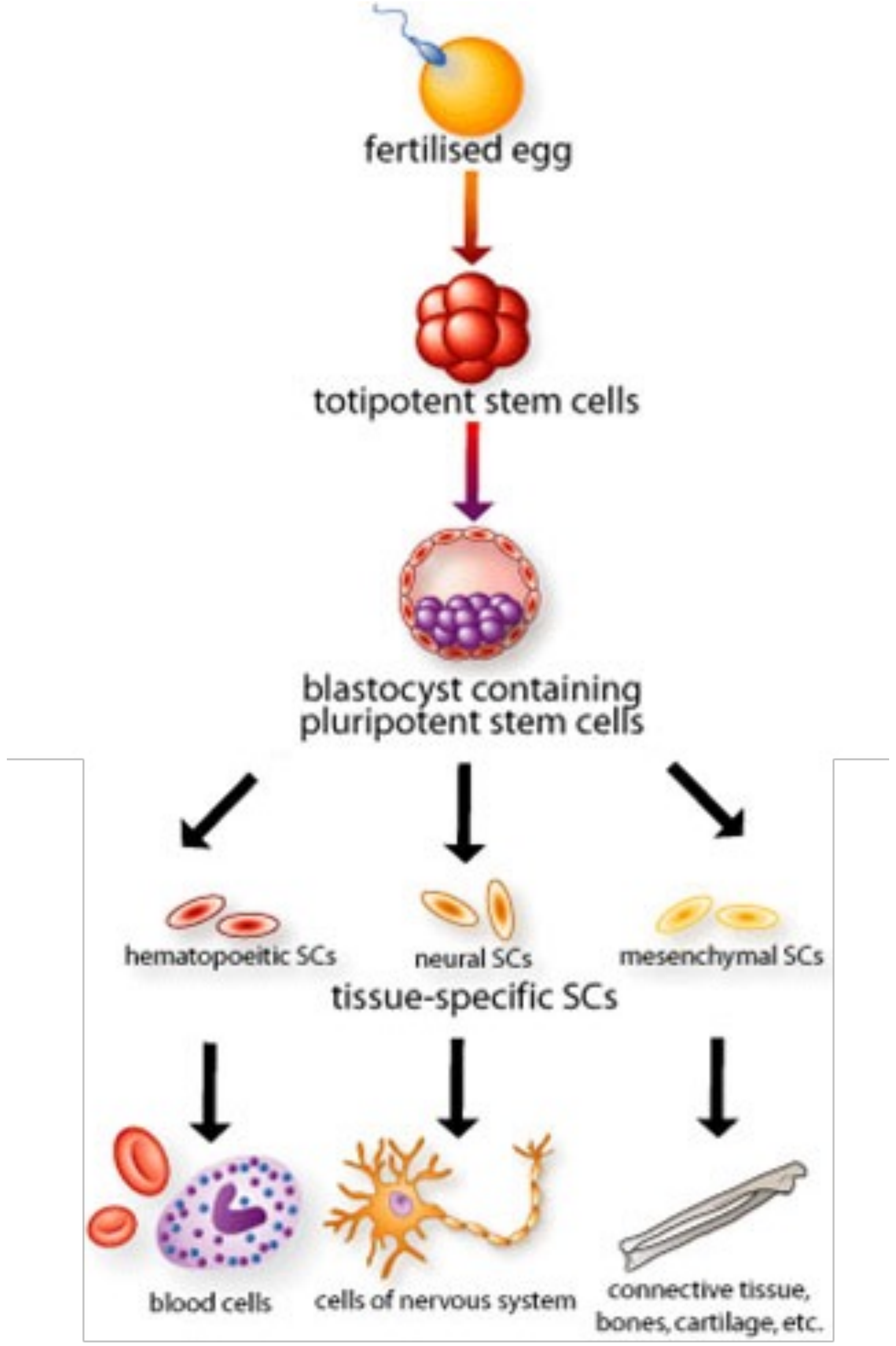
How do these different cell types arise?



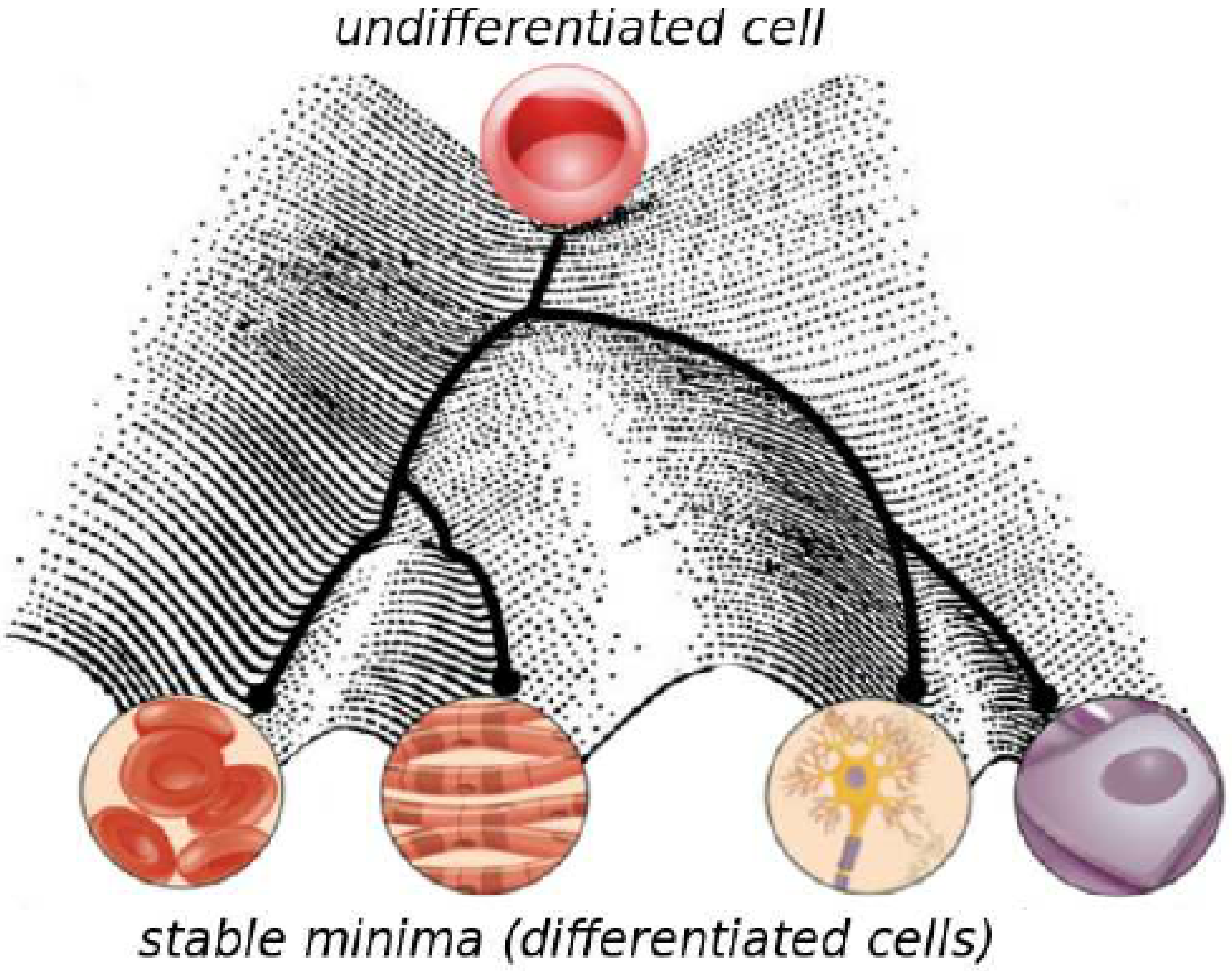
Waddington's **"Epigenetic Landscape"**



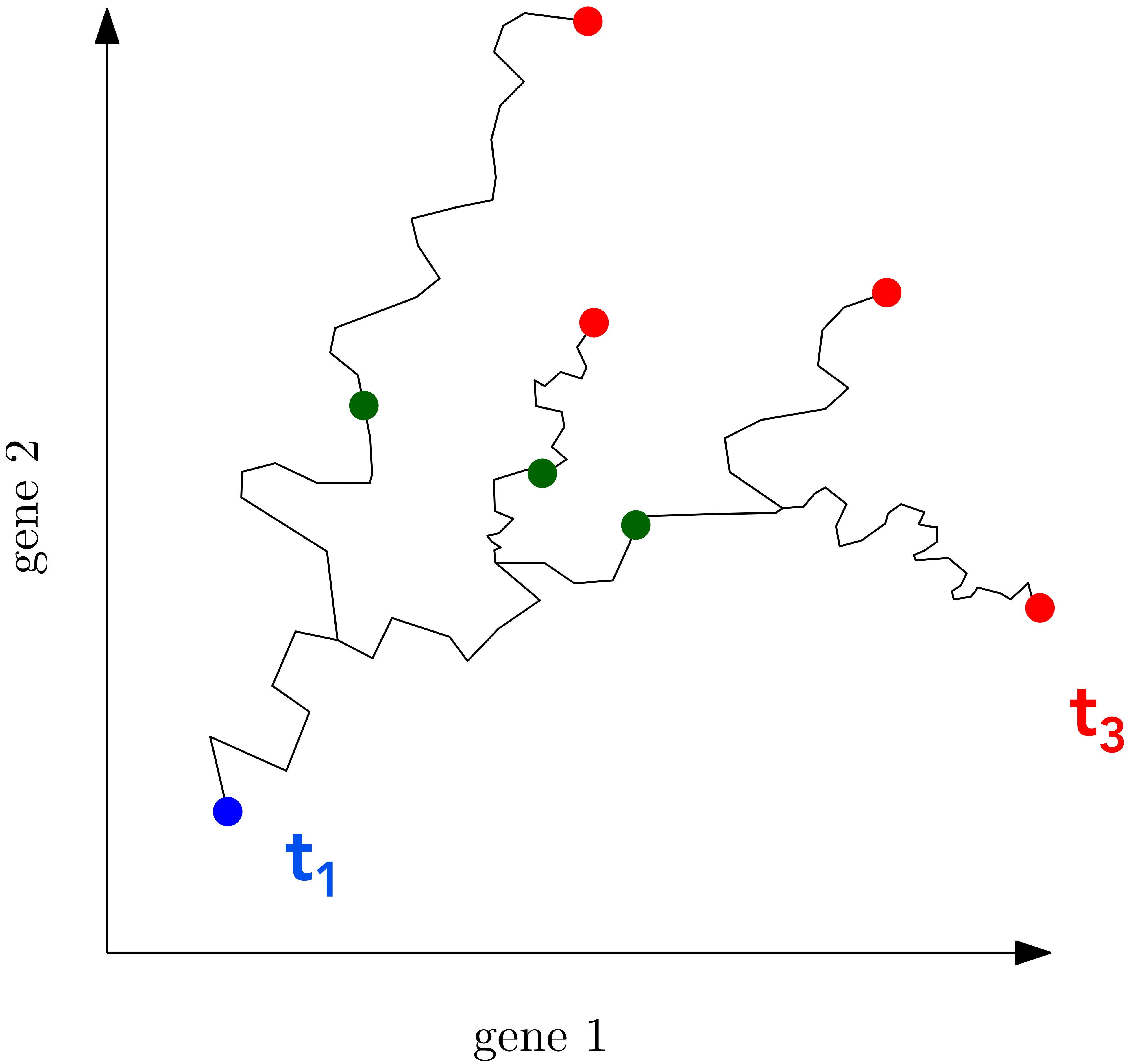
How do these different cell types arise?



Waddington's **"Epigenetic Landscape"**



Defining developmental trajectories

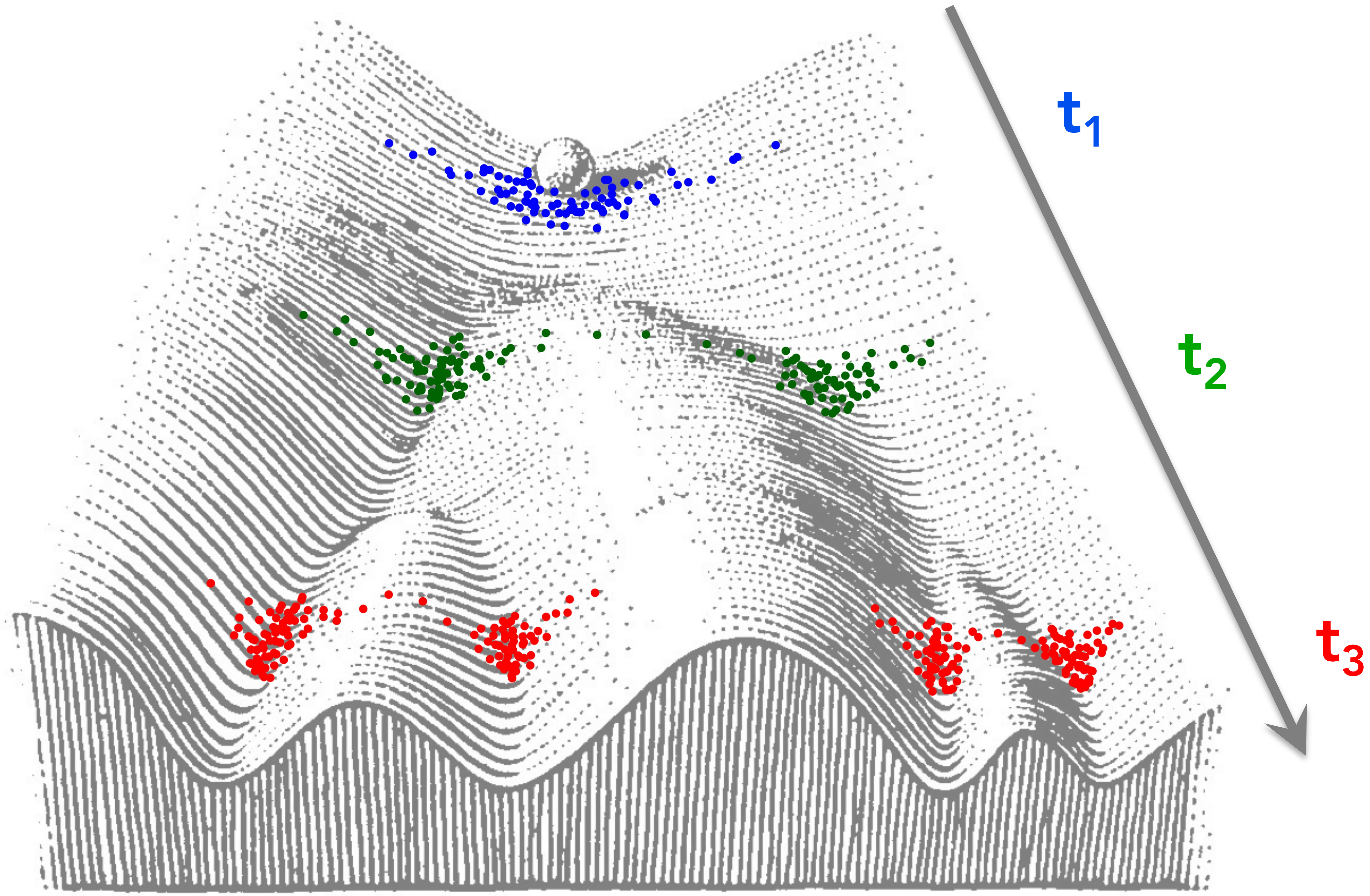


Cells change gene expression over time.

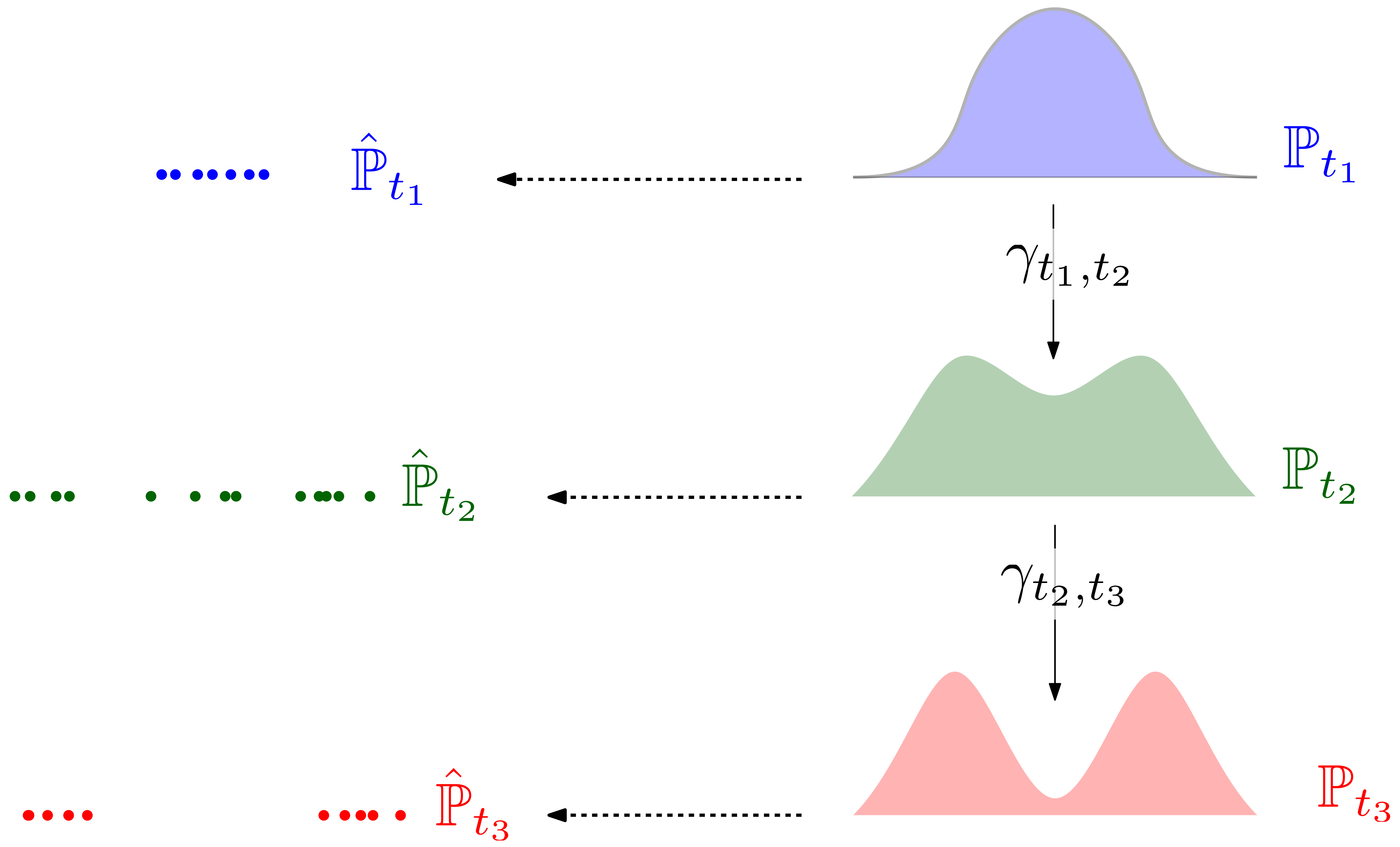
Cell division creates branching paths.

Measurement kills cells so we cannot observe paths!

Developmental time courses with scRNA-seq



Defining developmental stochastic processes

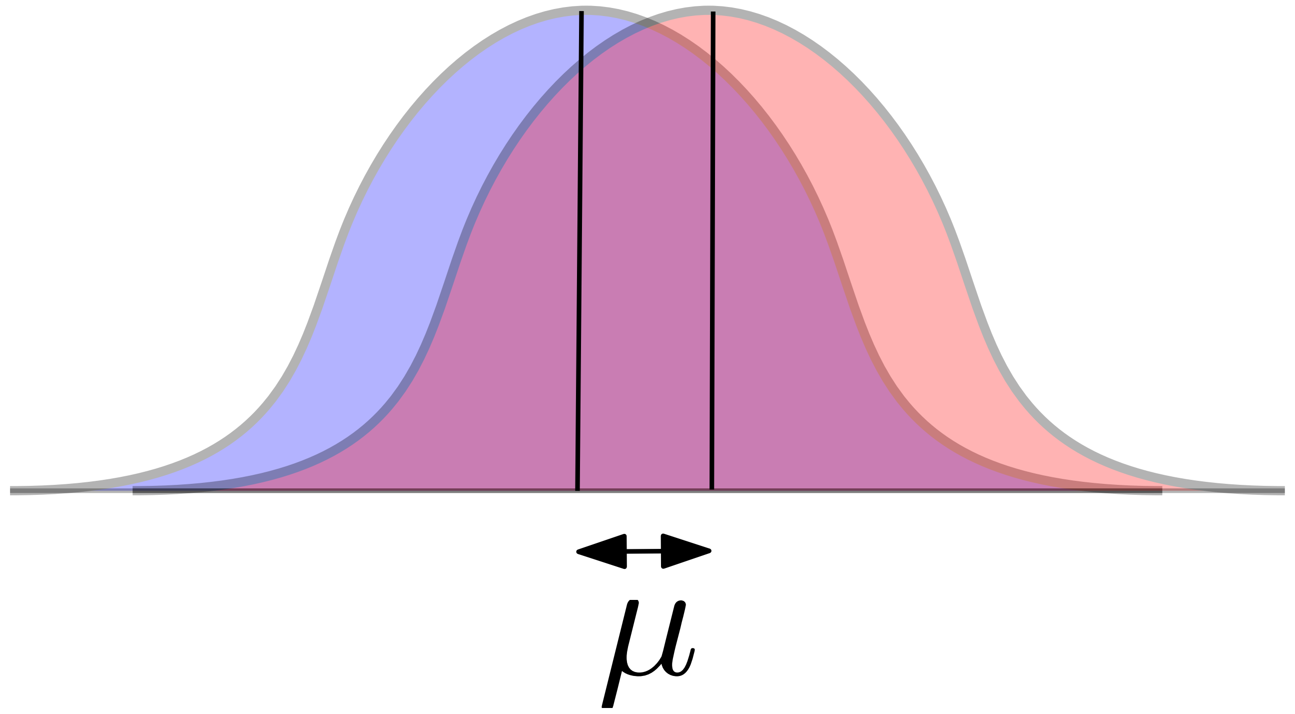


Goal: infer γ_{t_1,t_2} from independent samples.

How can we estimate joint distributions?

Example. Consider a pair of random variables

$$X_0 \sim \mathcal{N}(0, \sigma^2) \quad \text{and} \quad X_1 \sim \mathcal{N}(\mu, \sigma^2)$$



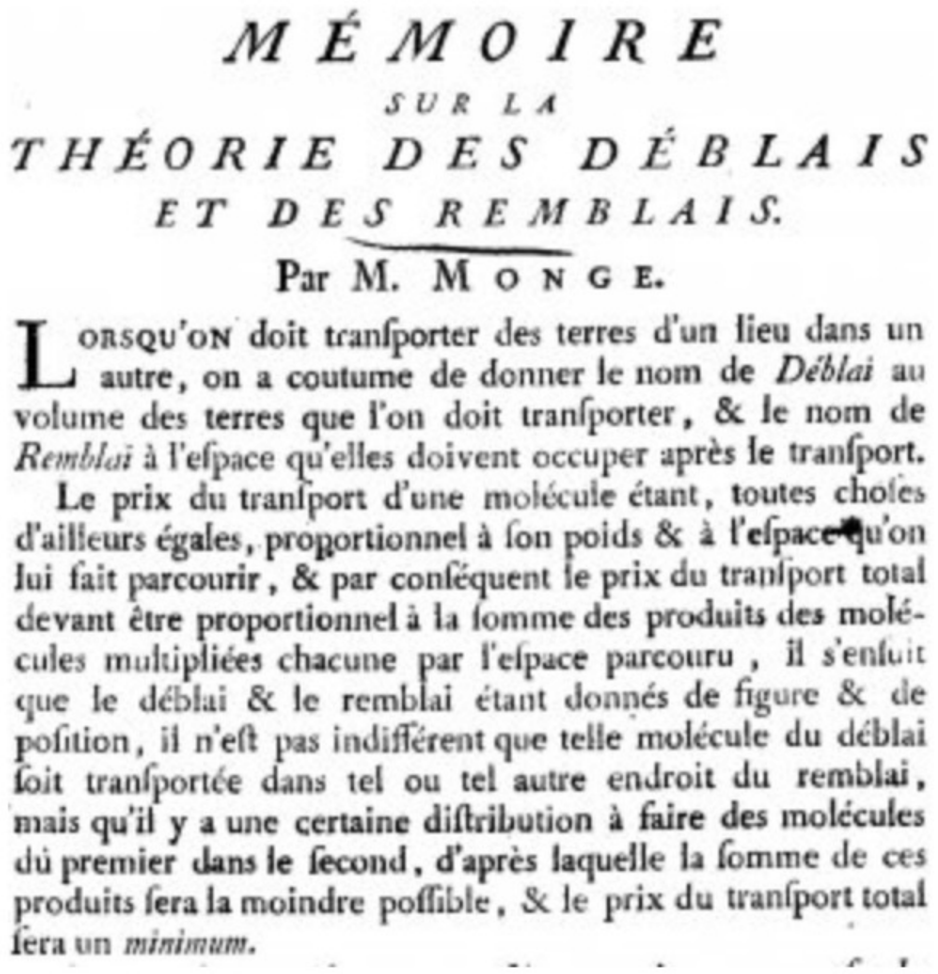
How can we estimate the joint distribution?

If X_0 is *close* to X_1, then estimate:

$$\hat{\gamma} \leftarrow \arg\min_{\pi} \mathbb{E}_{\pi} \|X_0 - X_1\|^2$$

Solution: $X_1 = X_0 + \mu$

Approach: optimal transport



MÉMOIRE

SUR LA

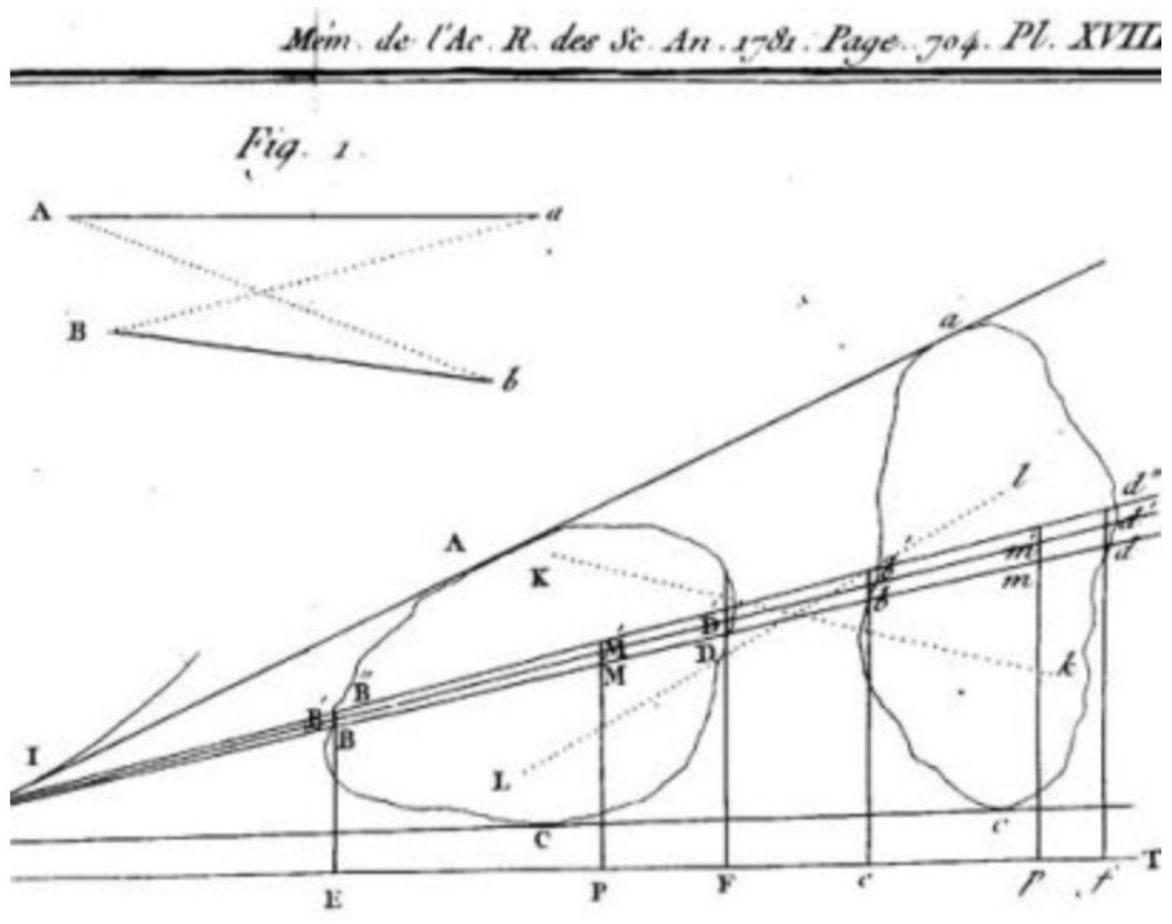
THÉORIE DES DÉBLAIS

ET DES REMBLAIS.

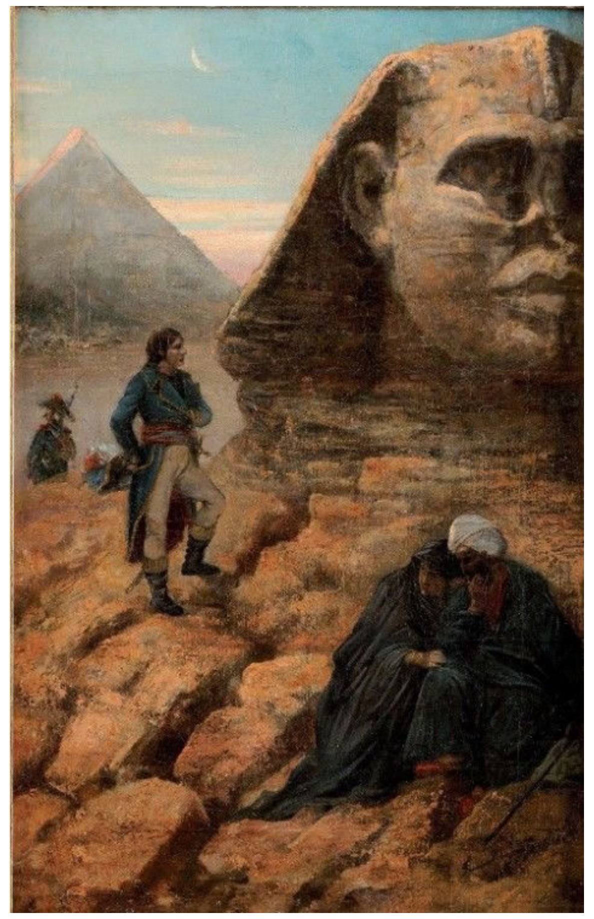
Par M. MONGE.

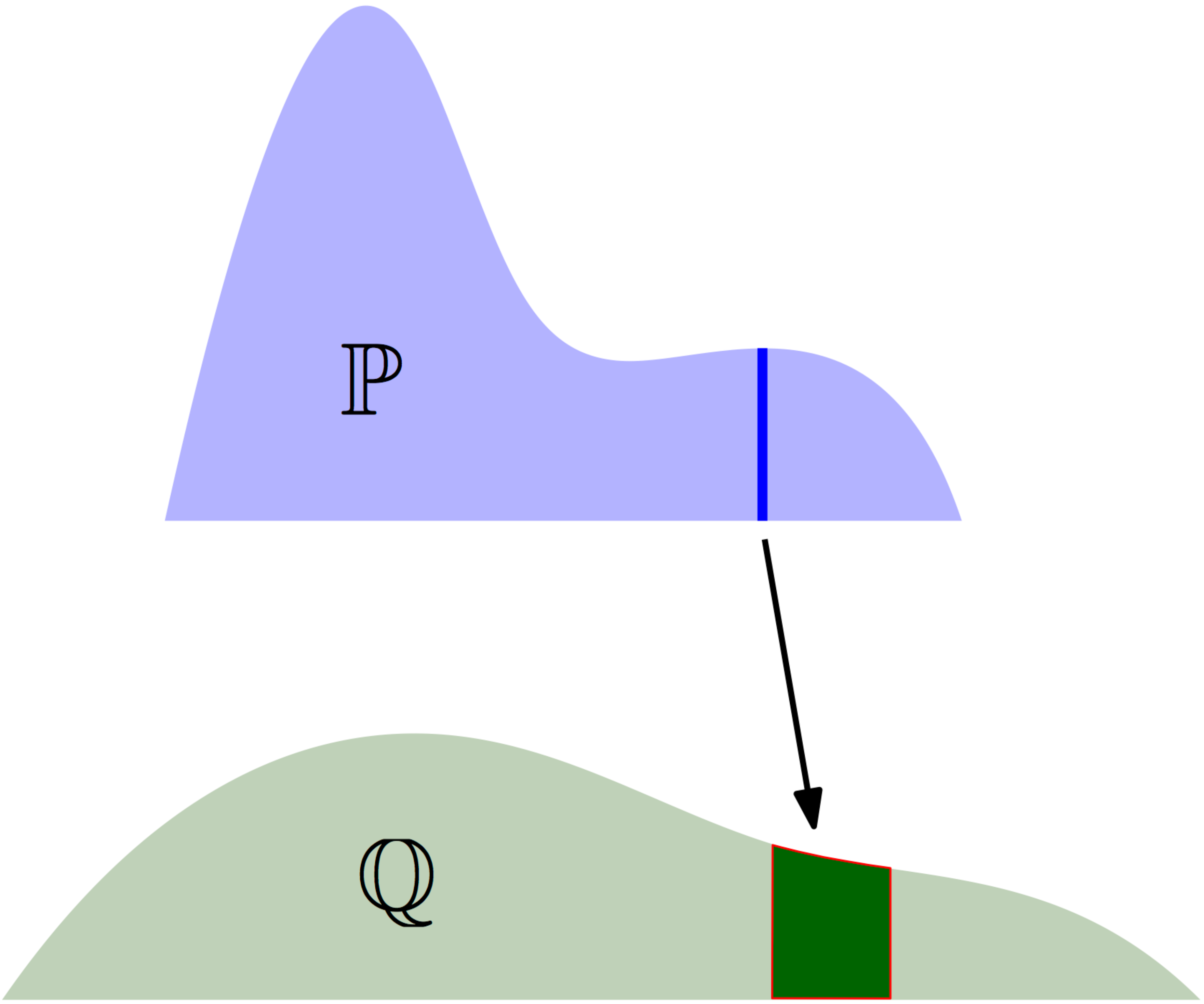
LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.



Gaspard Monge, 1781



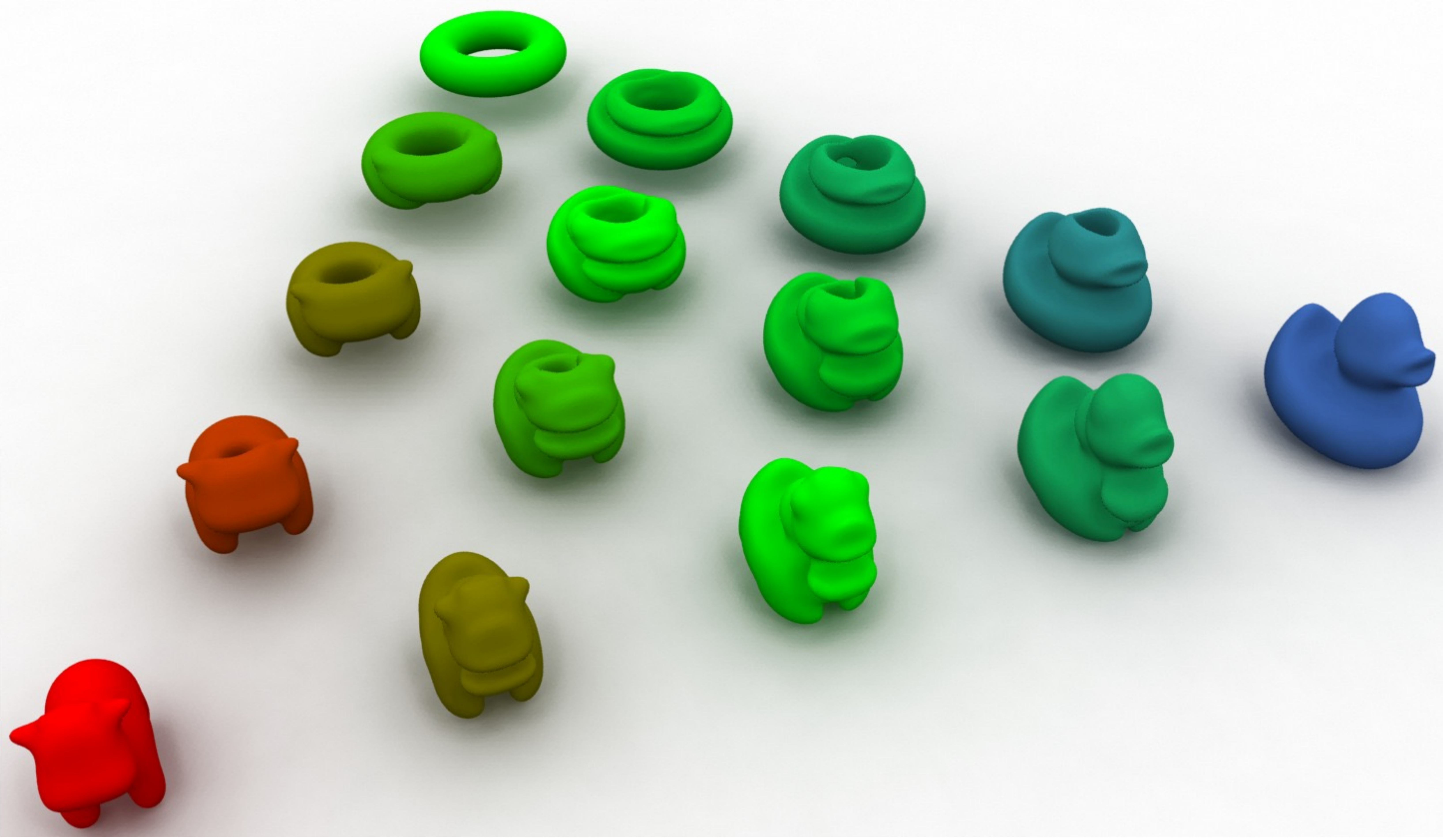


- Transport grain-by-grain to minimize total work
- Application to creating military fortifications
- Earliest anticipation of Linear Programming.

  Monge, 1781. Kantorovich, 1940.

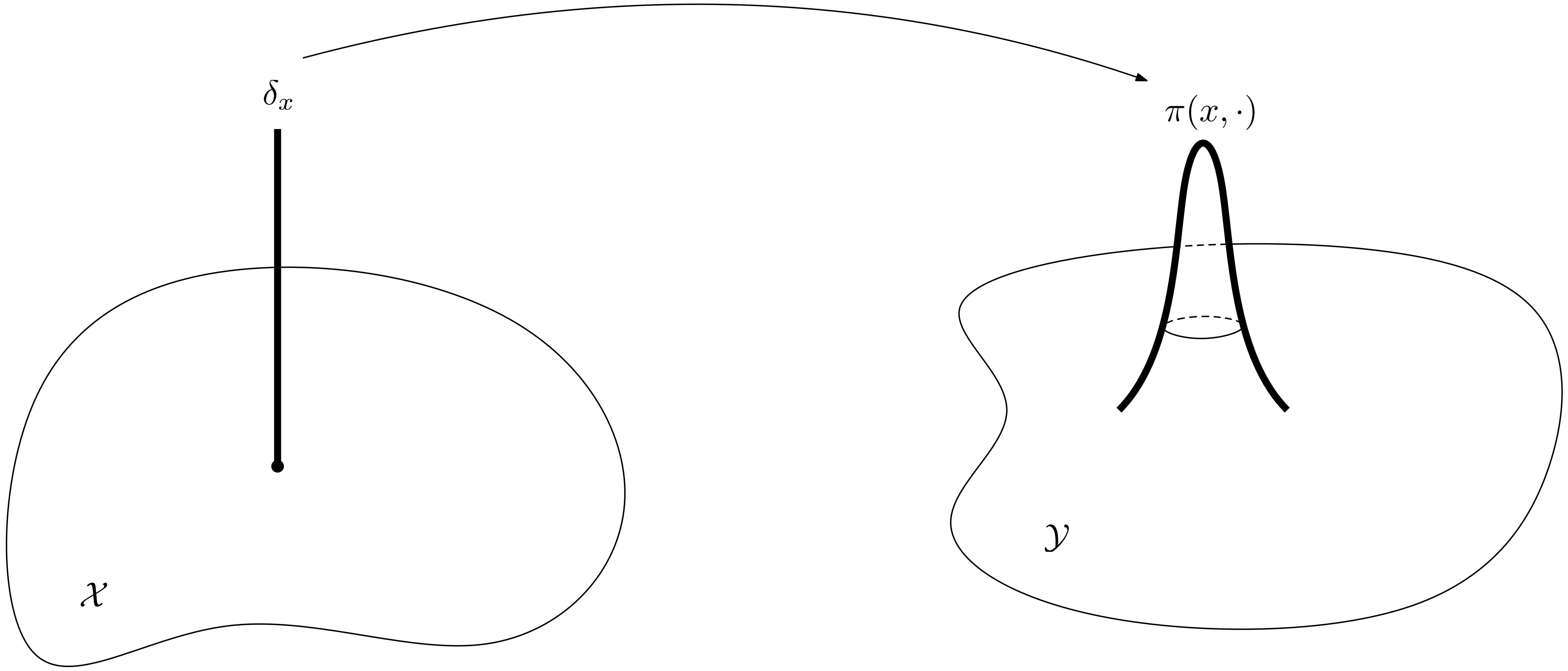
Optimal transport

computer graphics • pure math (PDE, differential geometry) • statistics • economics • fluid mechanics



Solomon et al. 2015

Transport plan redistributes mass

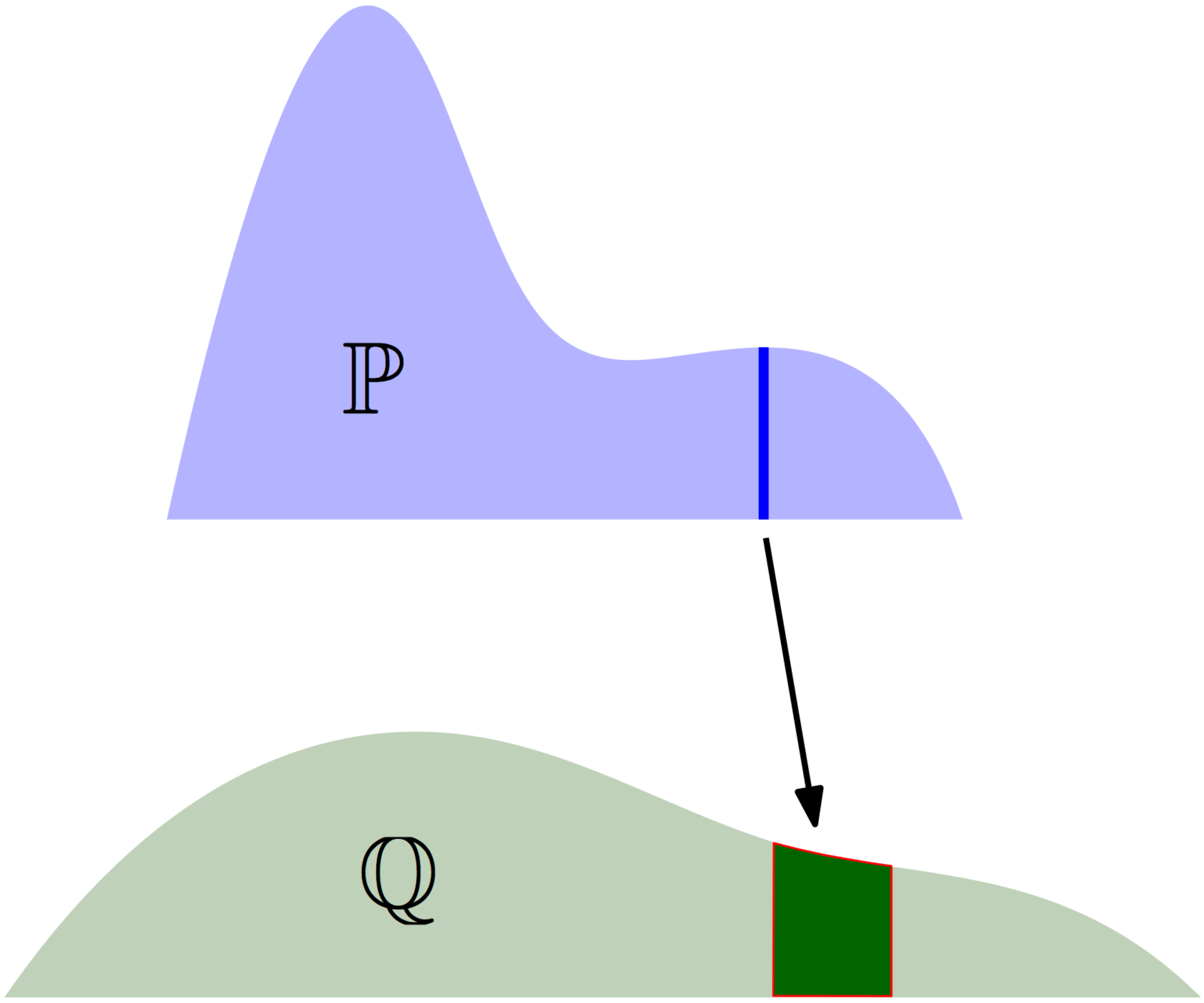


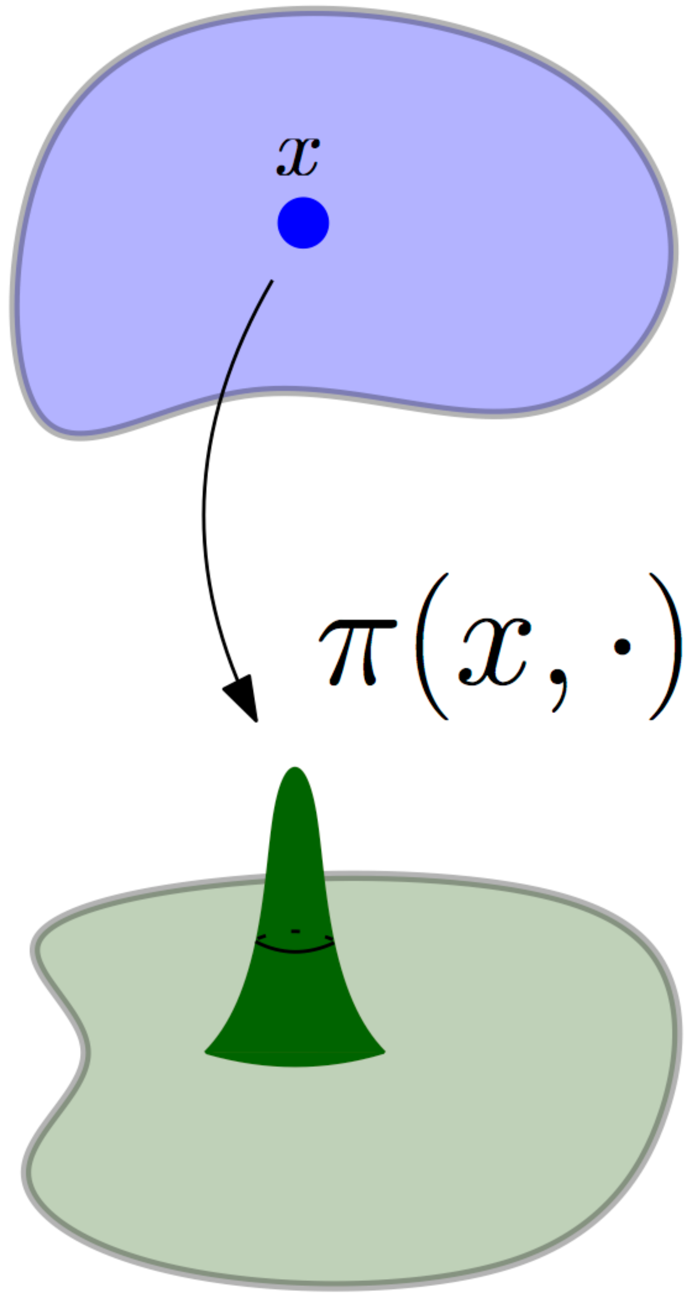
The ***optimal*** transport plan minimizes the expected cost:

$$\mathbb{E}_\pi c(x,y) = \iint c(x,y)\pi(x,y)dxdy.$$

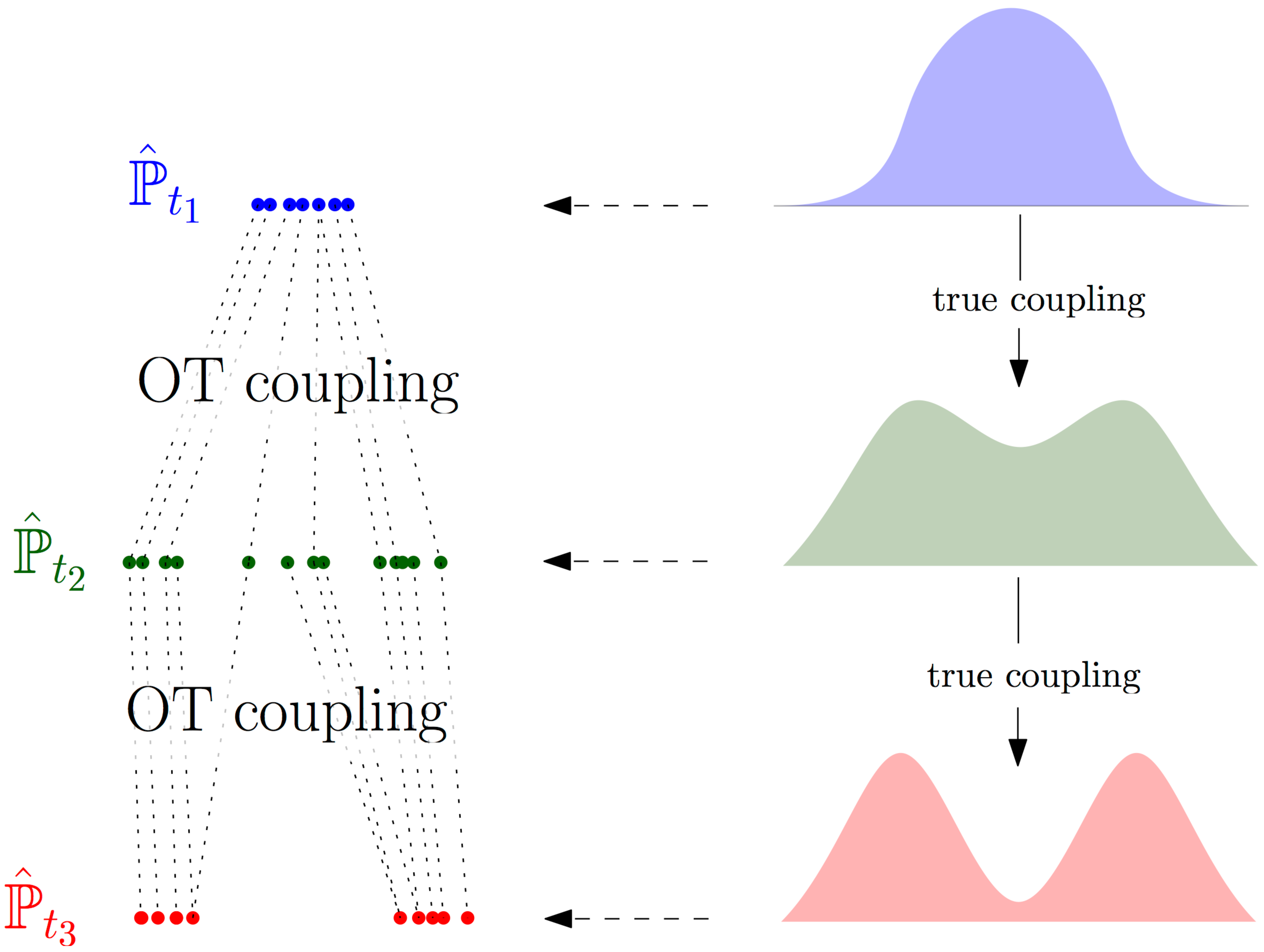
Optimal transport is a linear program

$$\begin{aligned} \underset{\pi}{\text{minimize}} \quad & \mathbb{E}_\pi \|x - y\|^2 \\ \text{subject to} \quad & \int \pi(x,y)\,dy = \mathbb{P}(x) \\ & \int \pi(x,y)\,dx = \mathbb{Q}(y) \end{aligned}$$



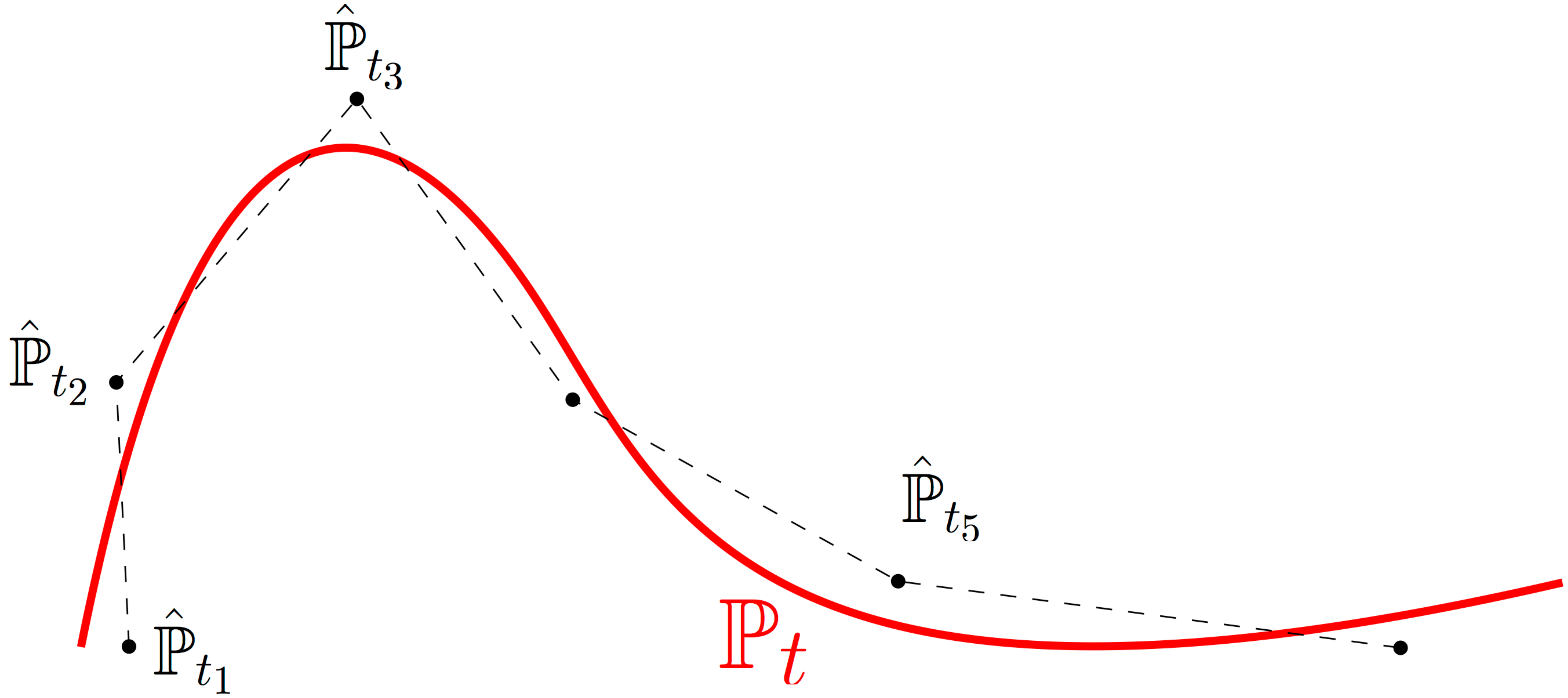


OT stochastic processes



OT stochastic processes

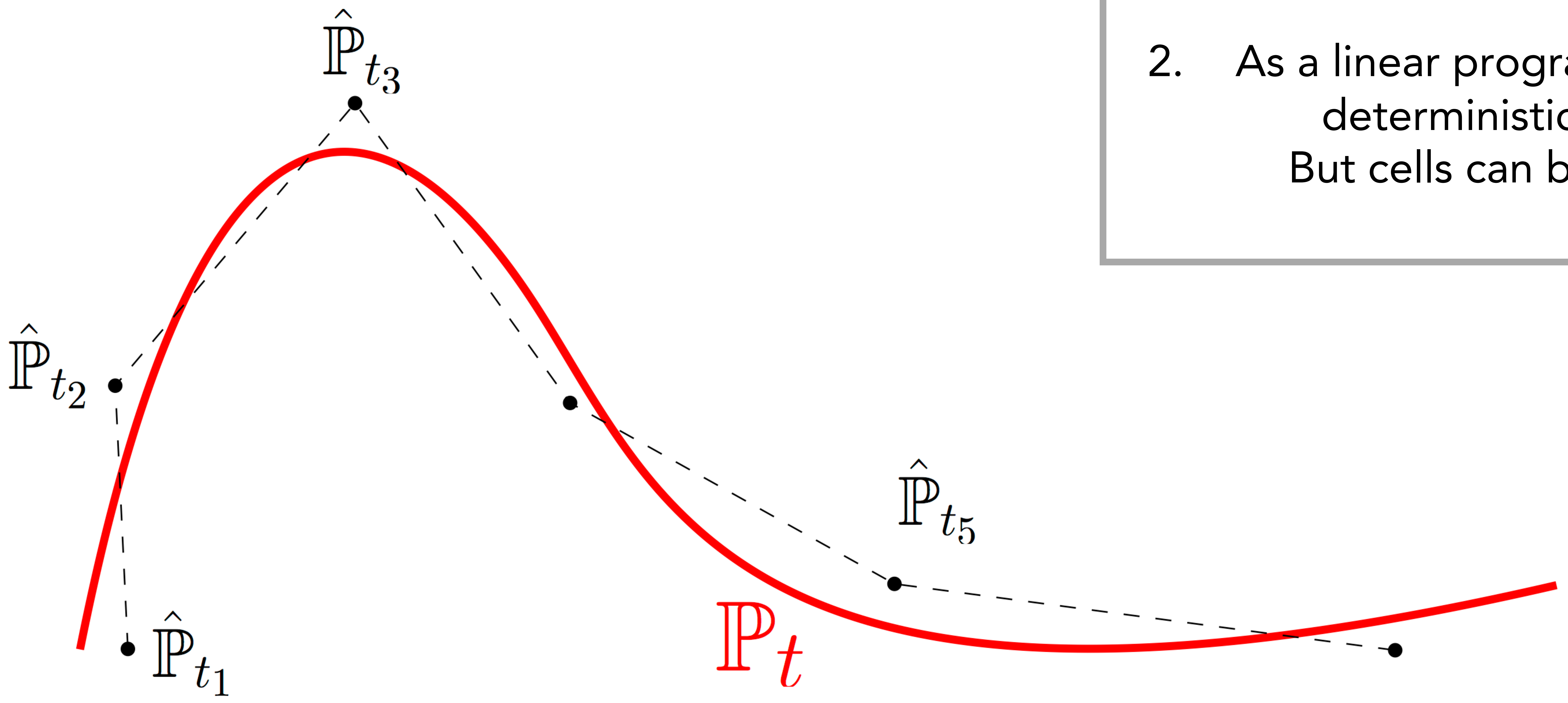
Intuition
In the space of distributions with the OT metric,
the process is **locally linear**:



OT stochastic processes

Intuition

In the space of distributions with the OT metric, the process is **locally linear**:



Biological considerations:

1. Classical OT conserves mass. But cells can proliferate!
2. As a linear program, OT gives a deterministic coupling. But cells can be partially fated!

Developmental temporal coupling with growth

Definition

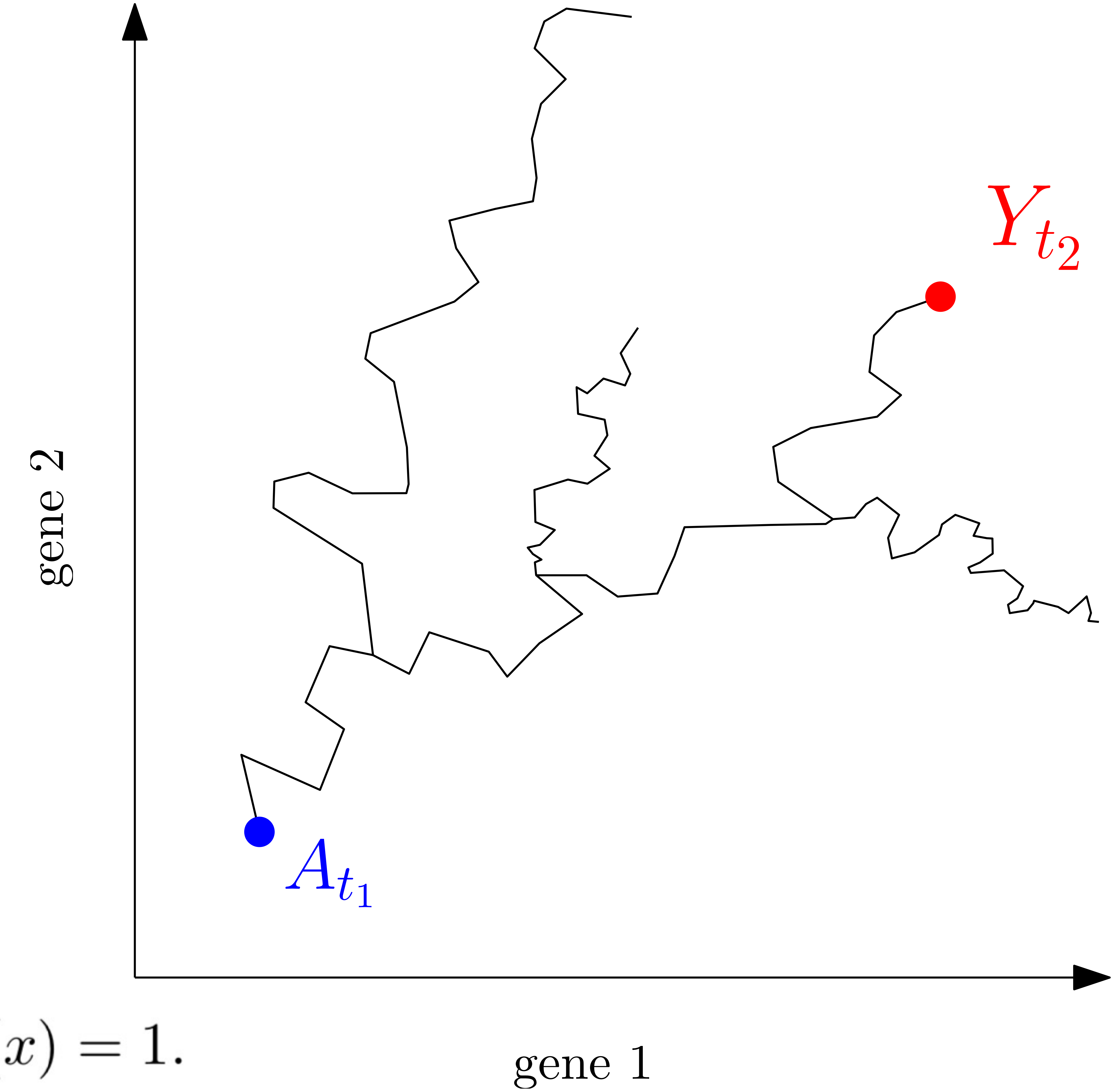
- Let Y_{t_2} denote a random cell at time t_2.
- Let A_{t_1} denote its ancestor at time t_1.
- γ_{t_1,t_2} is the distribution of (A_{t_1}, Y_{t_2}).

The marginals of γ_{t_1,t_2} are

$$\int \gamma_{t_1,t_2}(x,y)dx = \mathbb{P}_{t_2}(y)$$

$$\int \gamma_{t_1,t_2}(x,y)dy = \mathbb{P}_{t_1}(x)g(x)^{t_2-t_1}$$

for some growth function g with $\int g(x)^{\Delta_t} d\mathbb{P}_{t_1}(x) = 1$.

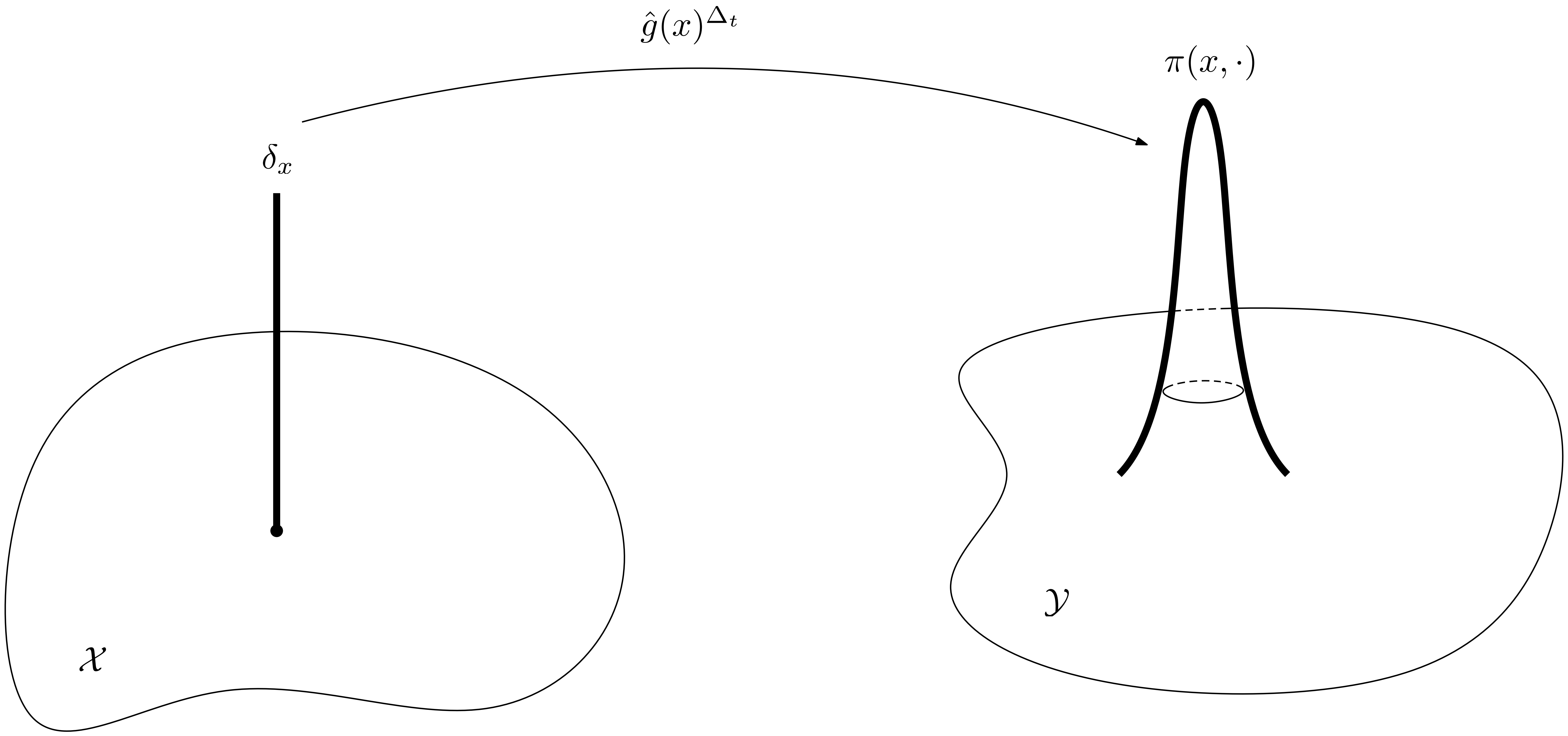


Optimal transport for proliferating cells

$$\begin{aligned}
\hat{\pi}_{t_1,t_2} = \quad & \underset{\pi}{\operatorname{argmin}} && \iint c(x,y)\pi(x,y)dxdy \\
& \text{subject to} && \int \pi(x,y)dx = d\hat{\mathbb{P}}_{t_2}(y) \\
& && \int \pi(x,y)dy = d\hat{\mathbb{P}}_{t_1}(x)\hat{g}(x)^{\Delta_t}
\end{aligned}$$

Optimal transport for proliferating cells

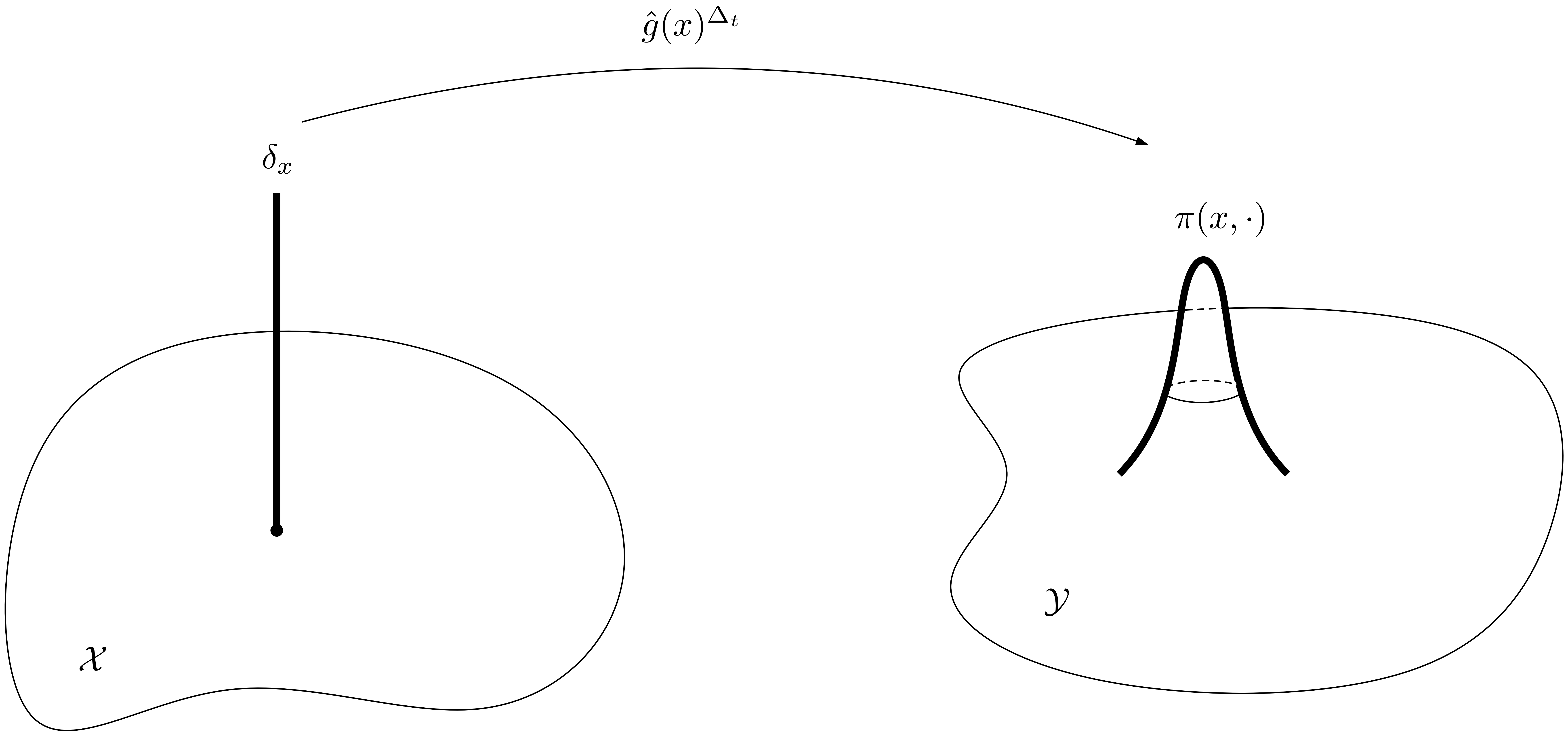
Estimate a growth function, $\hat{g}(x)^{\Delta_t}$, that scales transported mass



$\Delta_t = t_2 - t_1$

Optimal transport for proliferating cells

Estimate a growth function, $\hat{g}(x)^{\Delta_t}$, that scales the transported mass



$\Delta_t = t_2 - t_1$

Biological considerations:

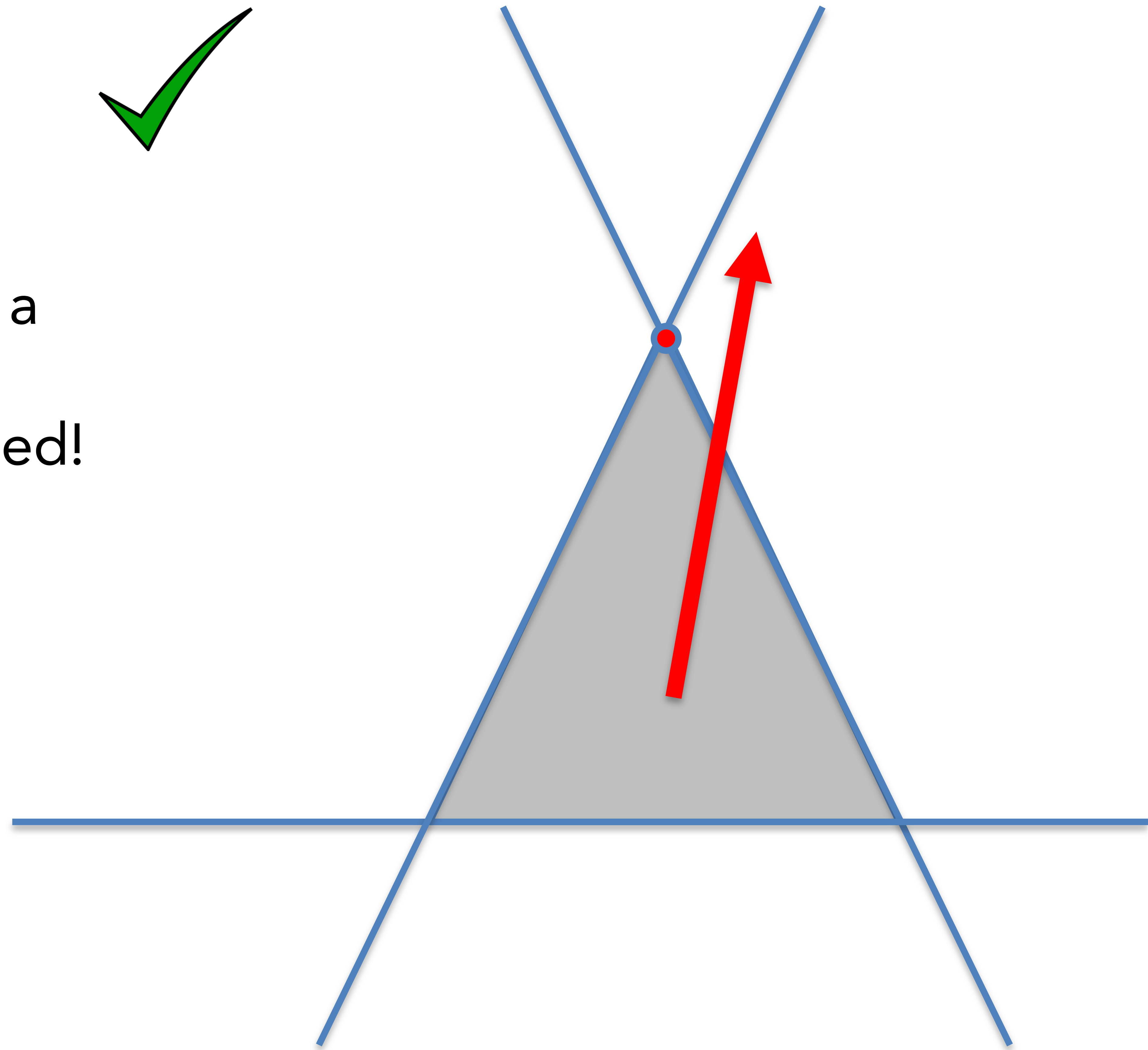
1. Classical OT conserves mass.
   But cells can proliferate! ✓

2. As a linear program, OT gives a deterministic coupling.
   But cells can be partially fated!

Biological considerations:

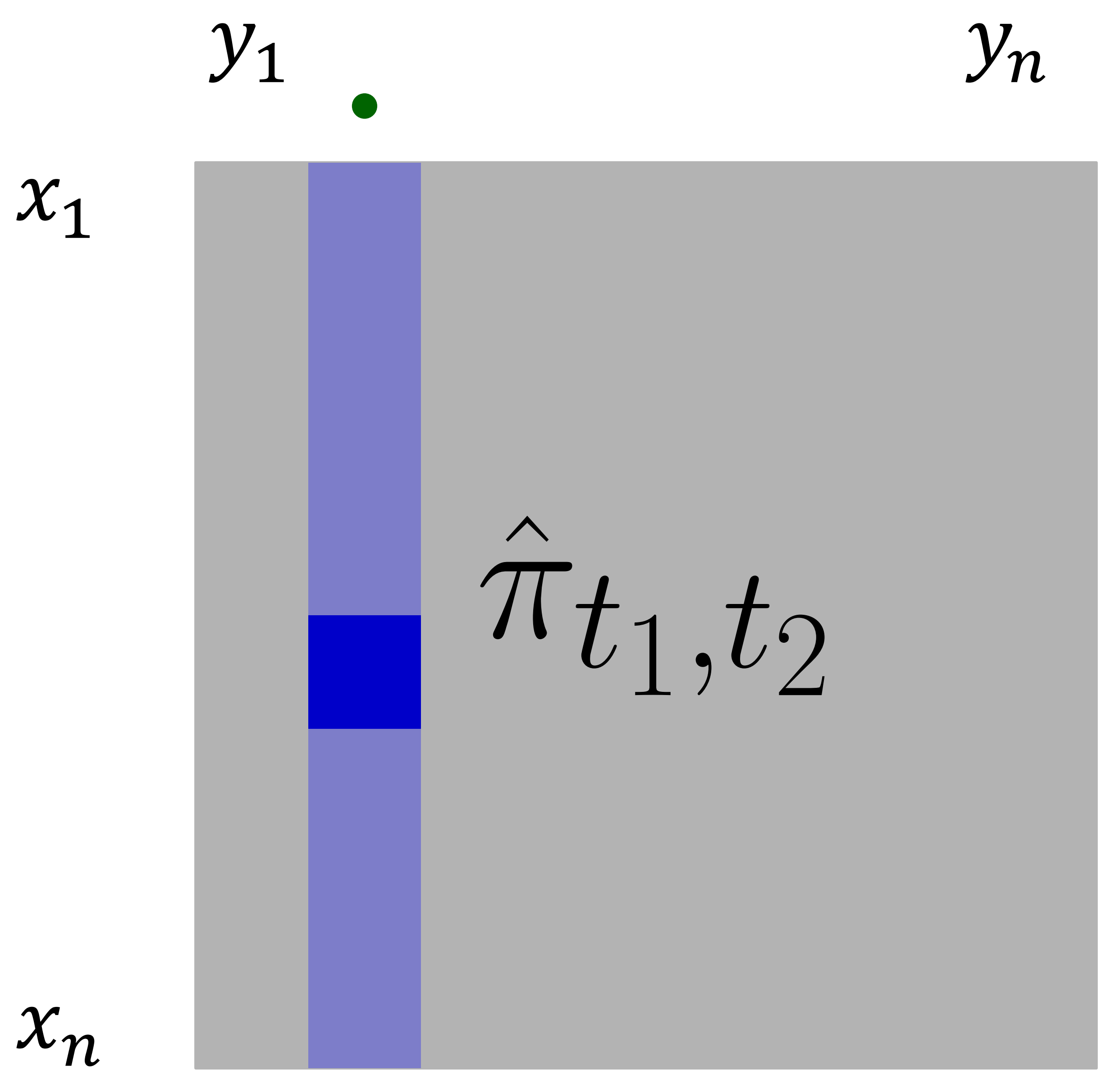
1. Classical OT conserves mass. ✓
   But cells can proliferate!

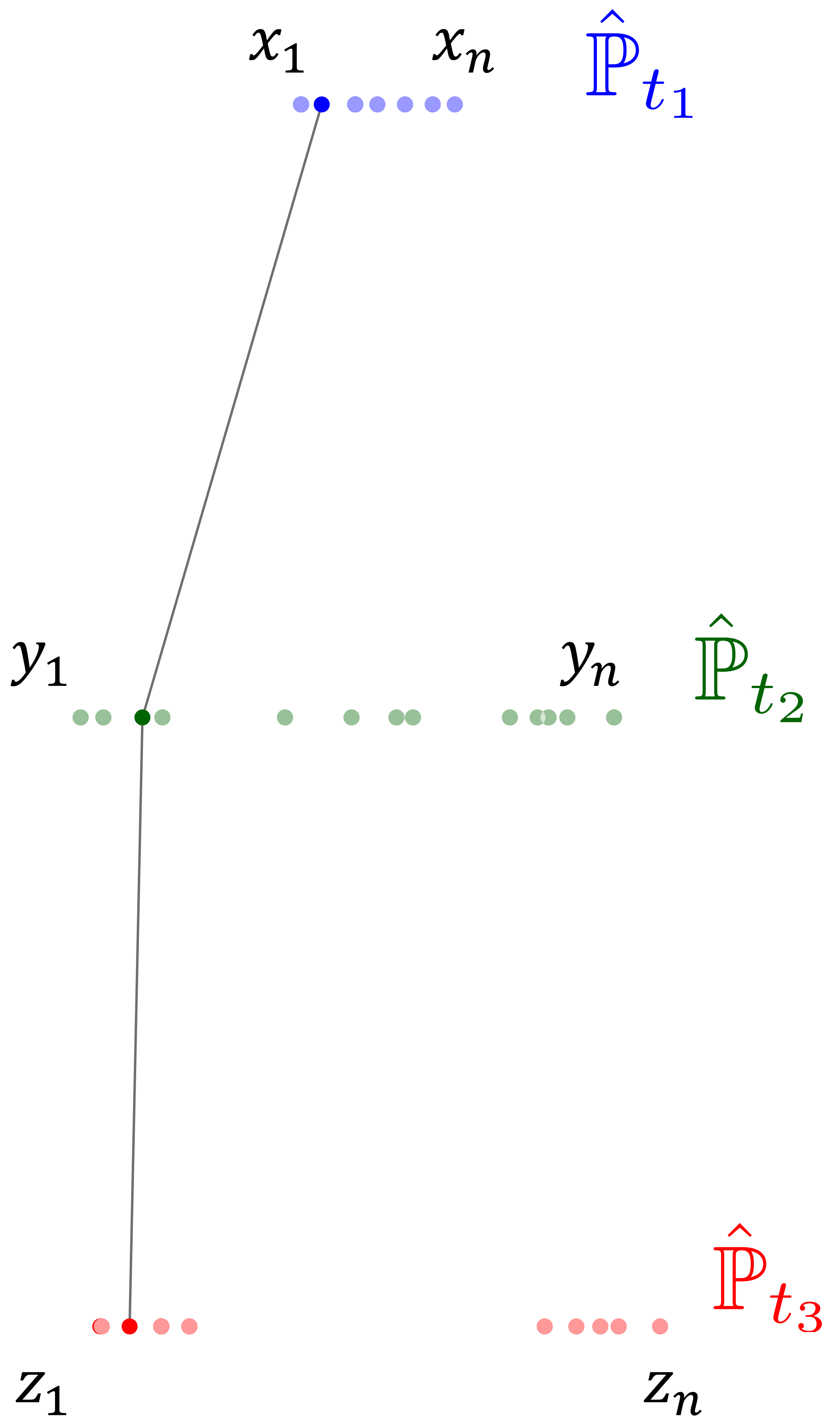
➡ 2. As a linear program, OT gives a deterministic coupling.
   But cells can be partially fated!



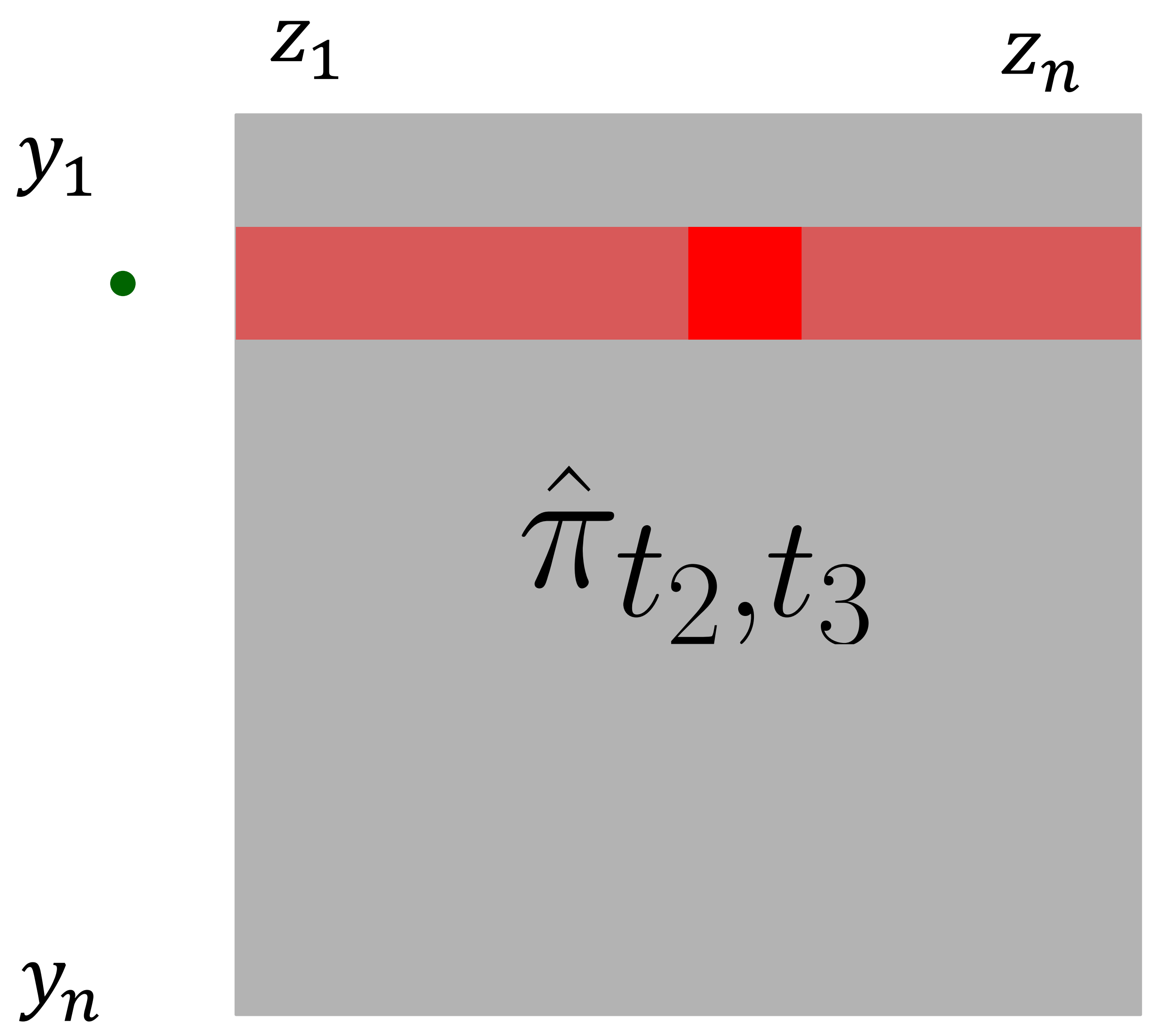
Ancestors, descendants, and entropy

columns specify **"ancestors"**





rows specify **"descendants"**



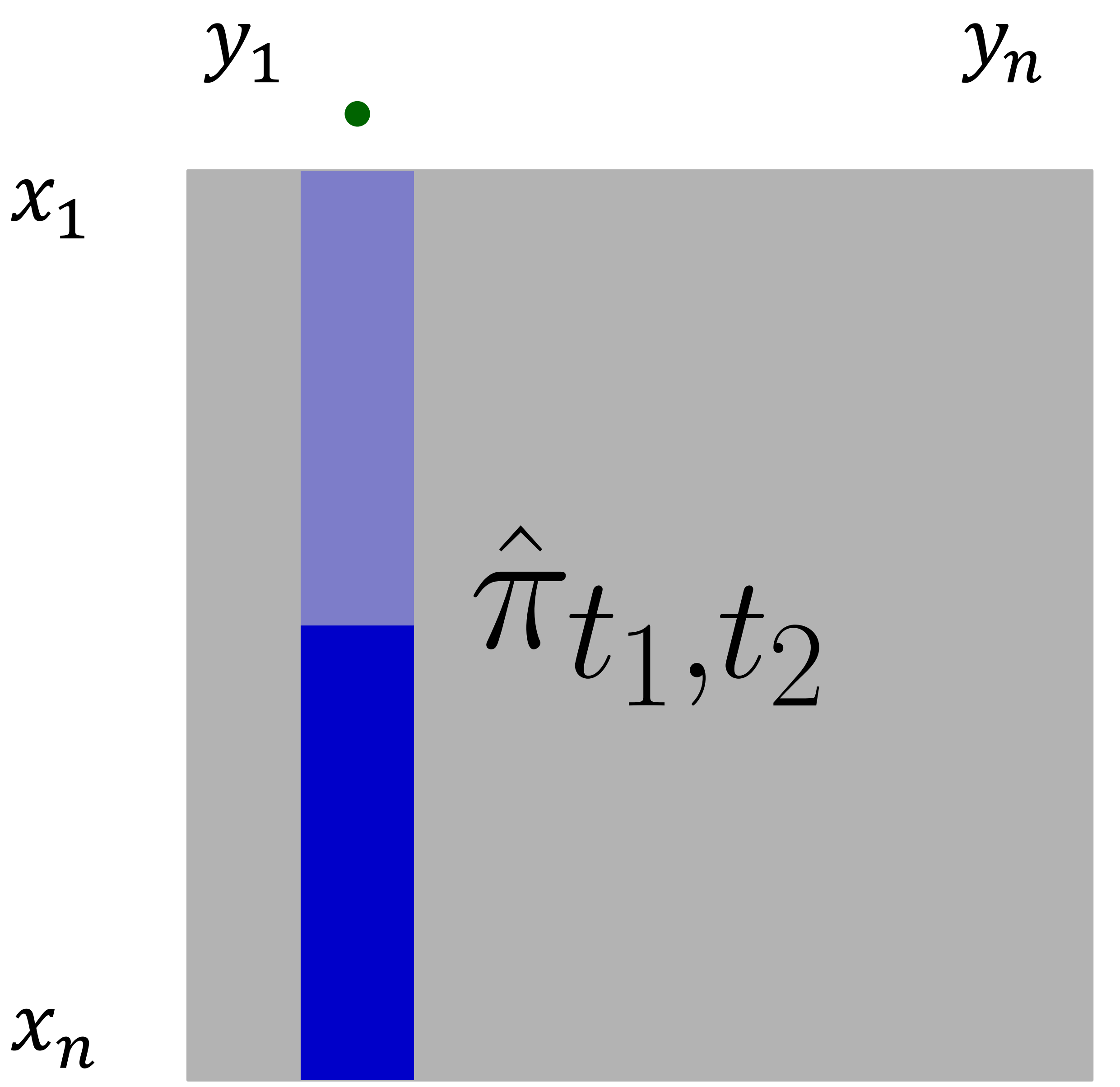
Entropy regularization

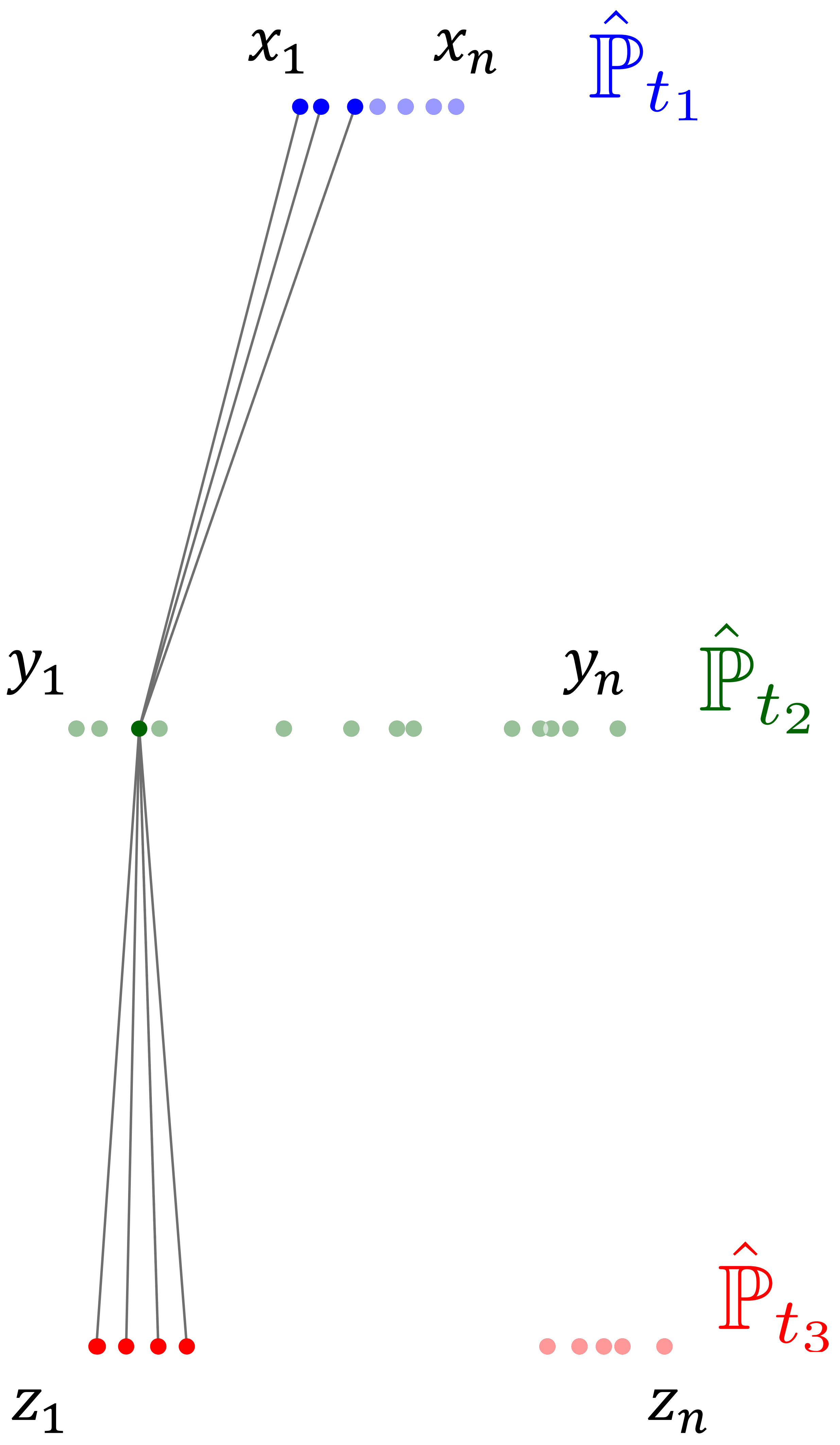
$$
\begin{aligned}
\hat{\pi}_{t_1,t_2} = \quad & \underset{\pi}{\operatorname{argmin}} && \iint c(x,y)\pi(x,y)dxdy - \epsilon\mathcal{H}(\pi) \\
& \text{subject to} && \int \pi(x,y)dx = d\hat{\mathbb{P}}_{t_2}(y) \\
& && \int \pi(x,y)dy = d\hat{\mathbb{P}}_{t_1}(x)\hat{g}(x)^{\Delta_t}
\end{aligned}
$$

Sinkhorn Distances: Lightspeed Computation of Optimal Transport. M. Cuturi (2013).

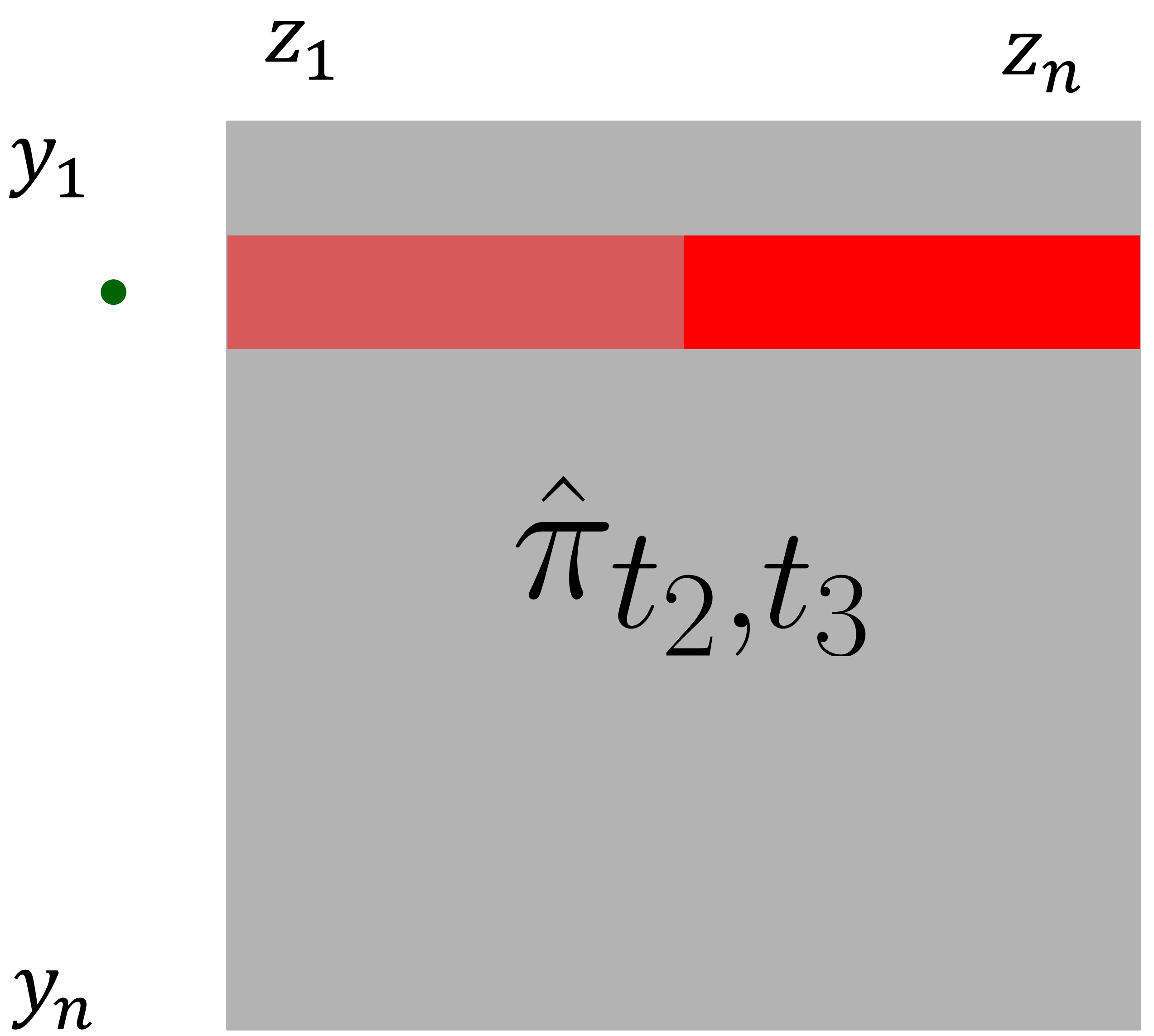
Ancestors, descendants, and entropy

columns specify **"ancestors"**





rows specify **"descendants"**



The benefits of entropic regularization

Biological

To what extent are cells fated?

Statistical

Given that we only have finitely many samples, we assign cells multiple descendants to avoid "overfitting".

Computational

Entropy regularization makes the optimization objective **strongly convex** and easy to solve quickly.

Theoretical justification: stay tuned for Part II

Interpreting entropy regularization: Schrodinger bridges

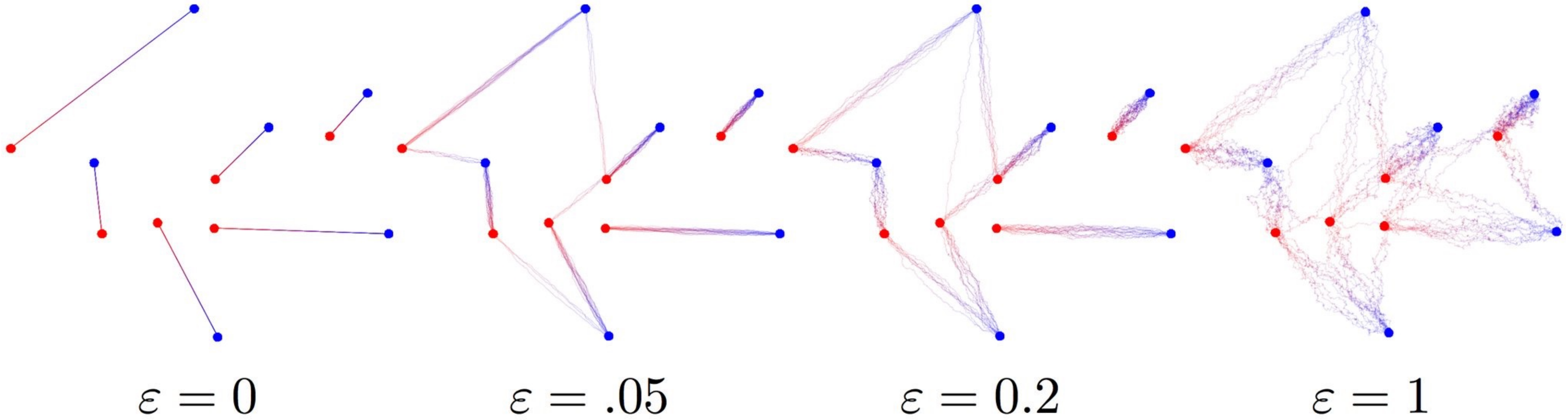


Photo courtesy of L. Chizat

Theorem: (Schrodinger, 1932)

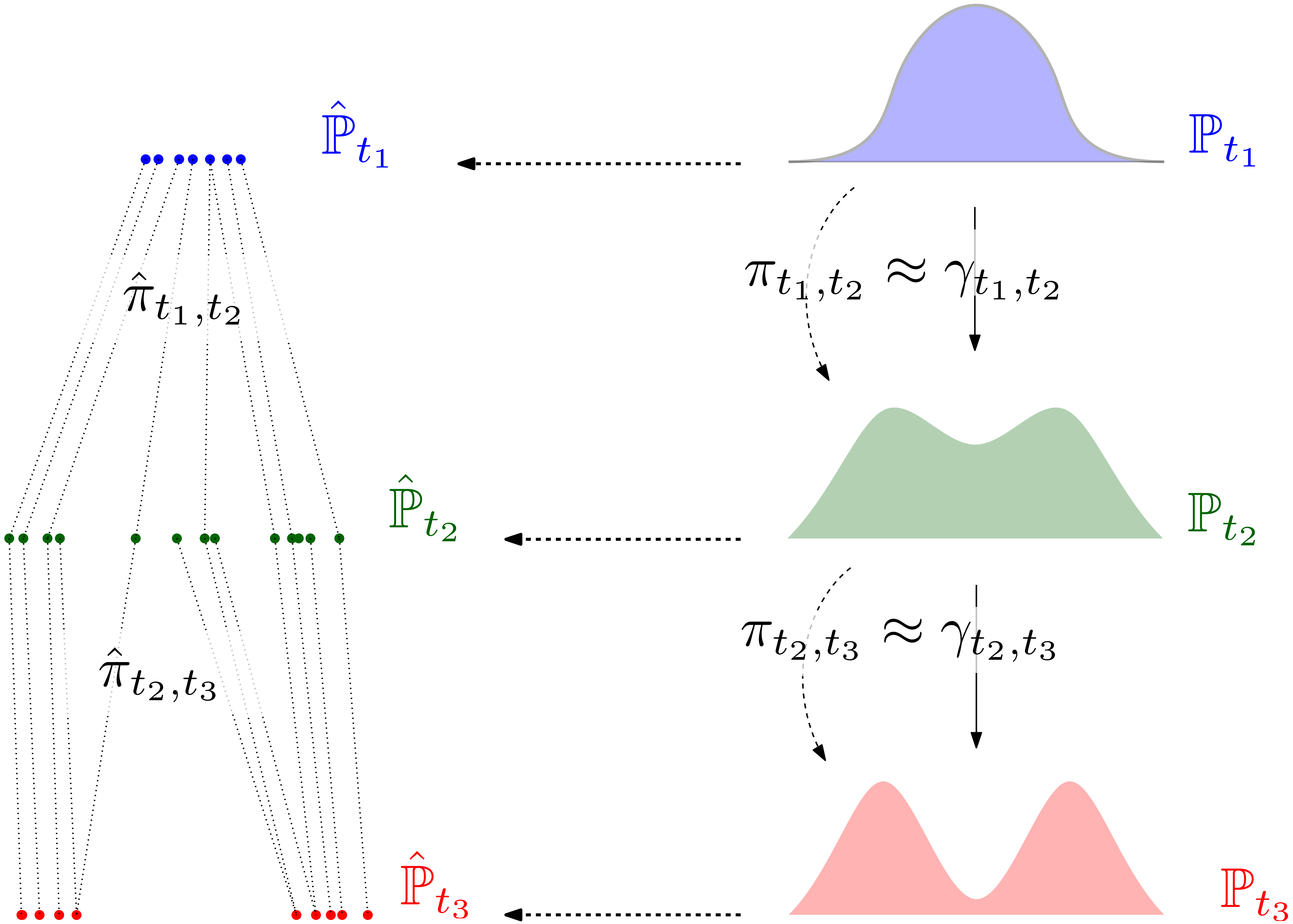
Entropically regularized transport map gives expected coupling of indistinguishable particles undergoing Brownian motion.

Unbalanced transport

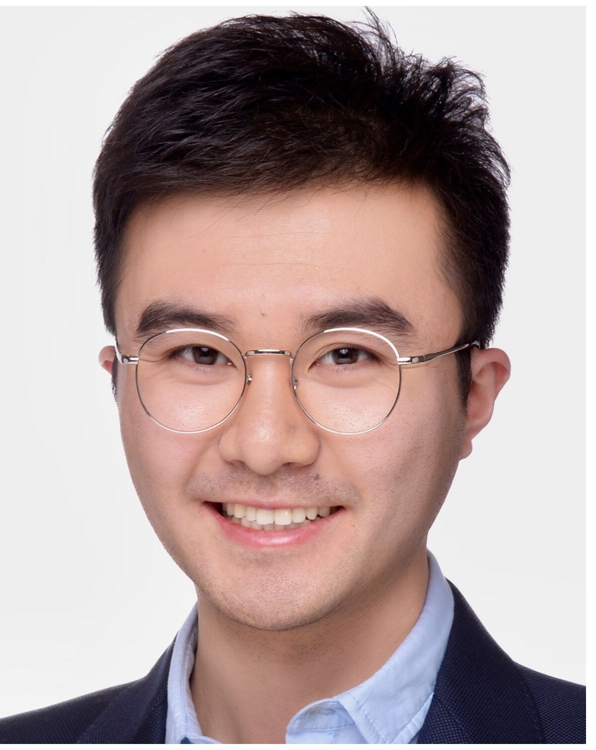
$$\begin{aligned} \hat{\pi}_{t_1,t_2} \quad = \quad \underset{\pi}{\text{argmin}} \quad & \iint c(x,y)\pi(x,y)dxdy - \epsilon\mathcal{H}(\pi) \\ \text{subject to} \quad & \int \pi(x,y)dx \approx d\hat{\mathbb{P}}_{t_2}(y) \\ & \int \pi(x,y)dy \approx d\hat{\mathbb{P}}_{t_1}(x)\hat{g}(x)^{\Delta_t} \end{aligned}$$

Scaling algorithms for unbalanced transport. L. Chizat et al., 2016.

Inferring developmental trajectories with optimal transport



Reprogramming fibroblasts to iPSCs



Jian Shu

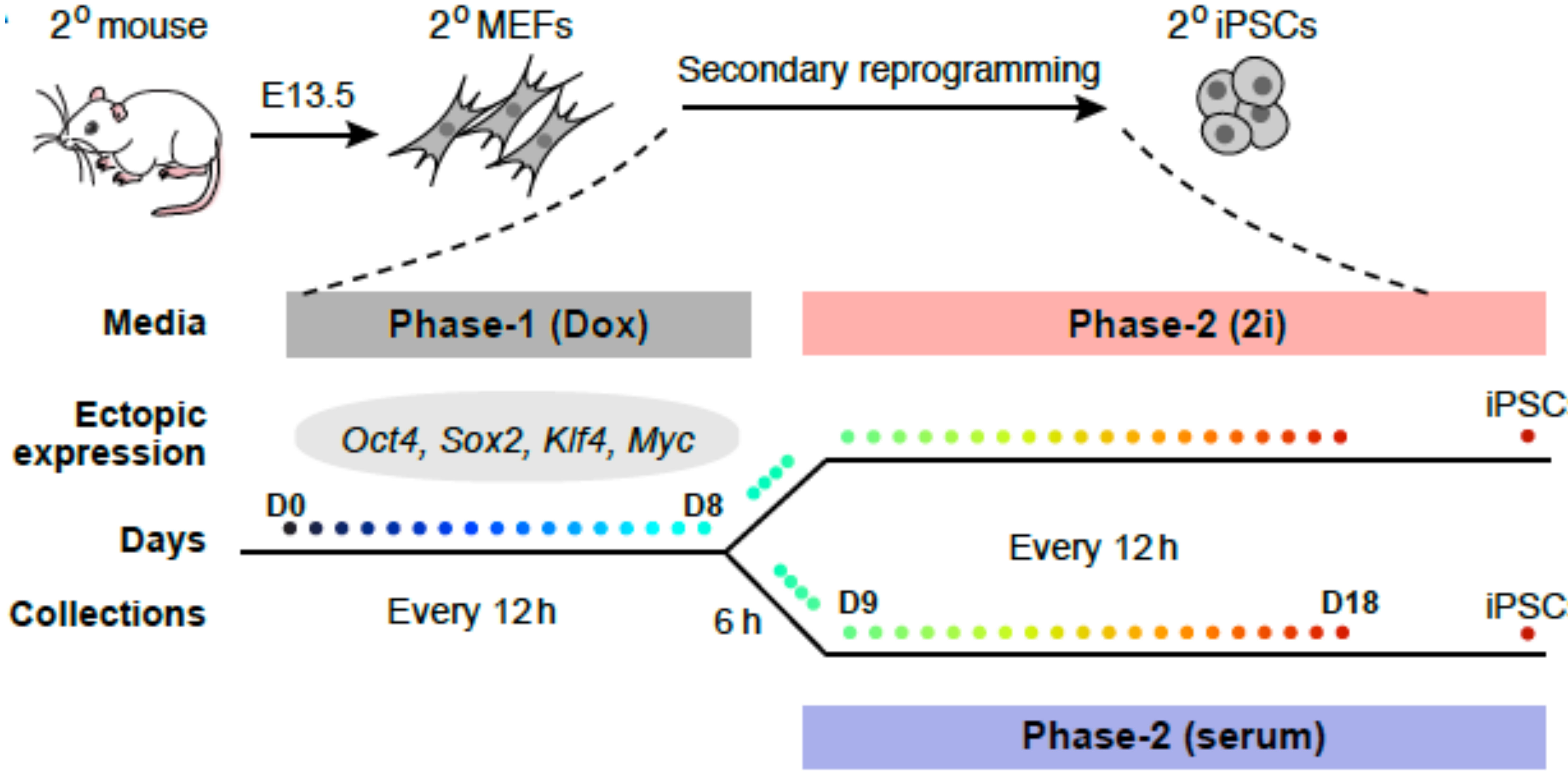
Marcin Tabaka

Brian Cleary

Philippe Rigollet

Aviv Regev

Eric Lander

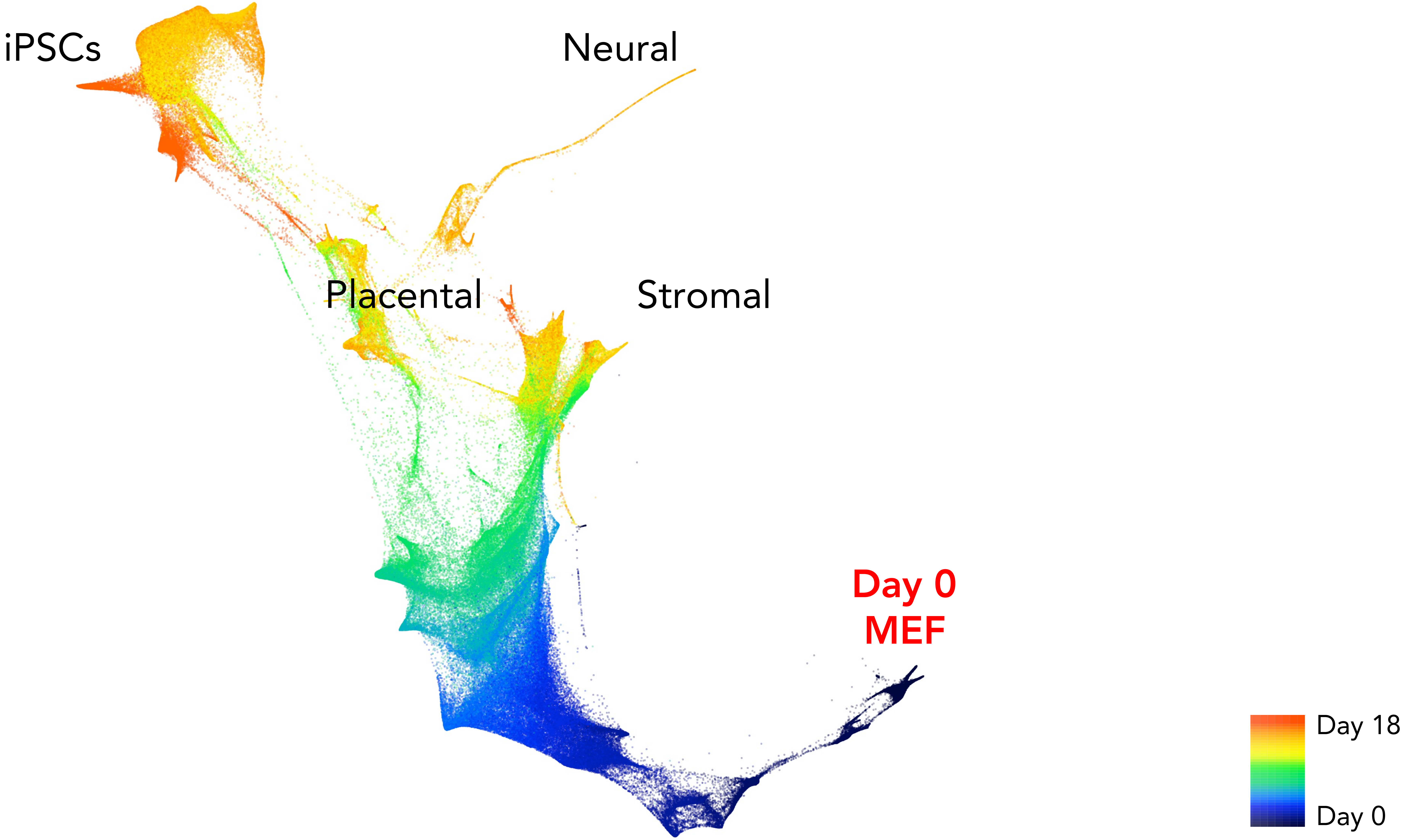


Schiebinger, Shu, Tabaka, Cleary, et al. Cell 2019

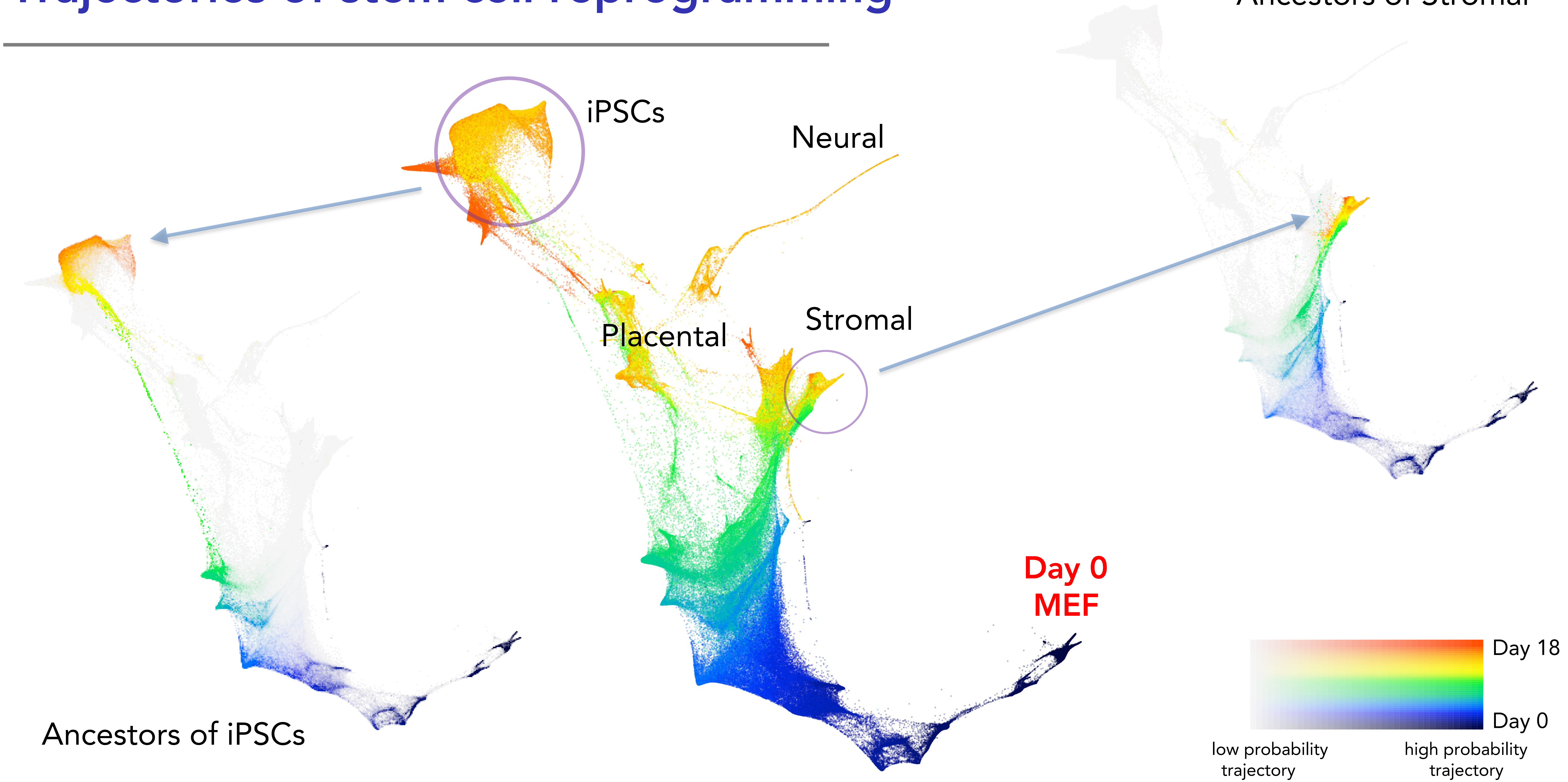
Day 0.0

315,000 cells
40 time points

Diverse cell types arise during reprogramming



Trajectories of stem cell reprogramming



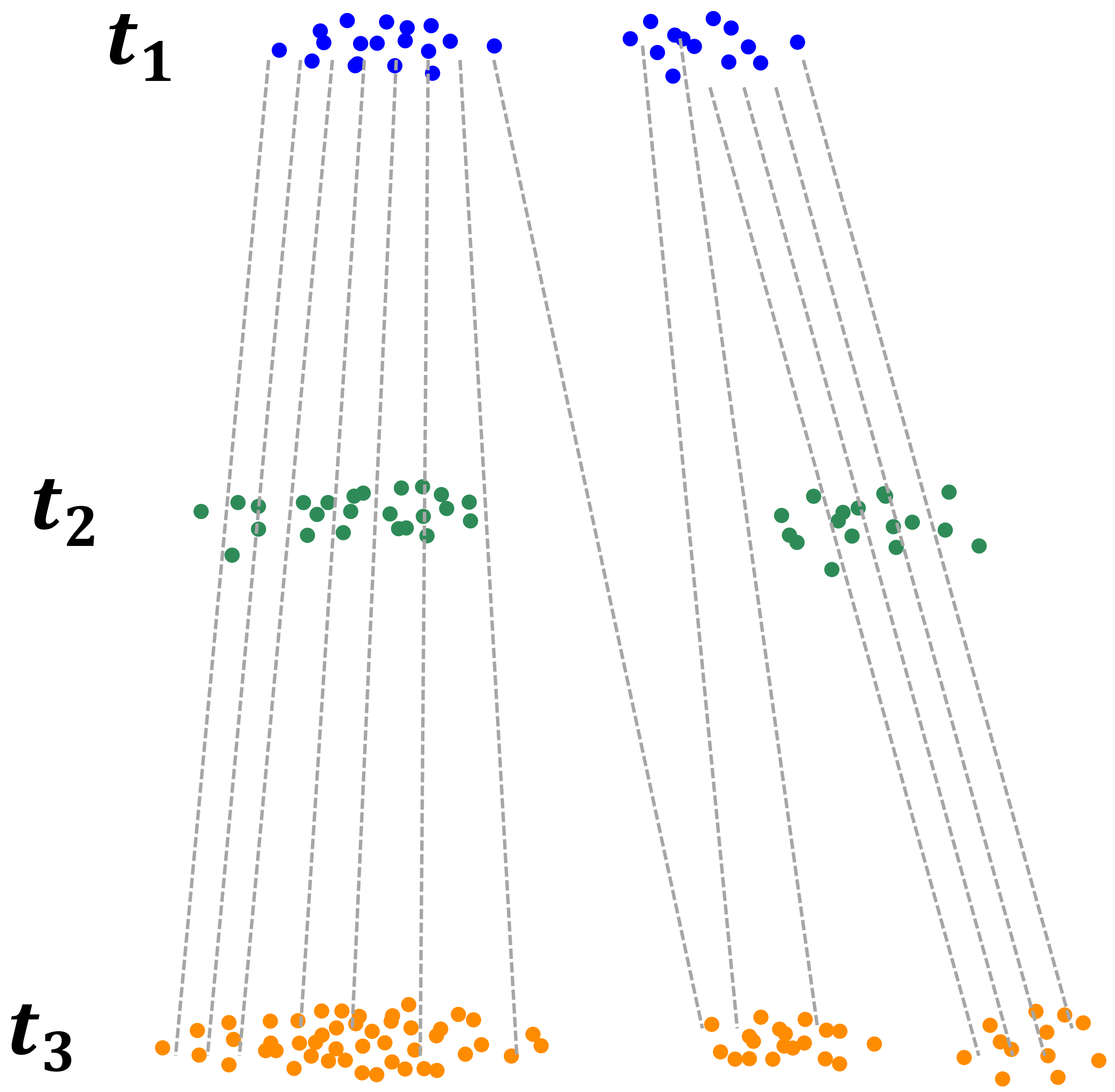

0.00

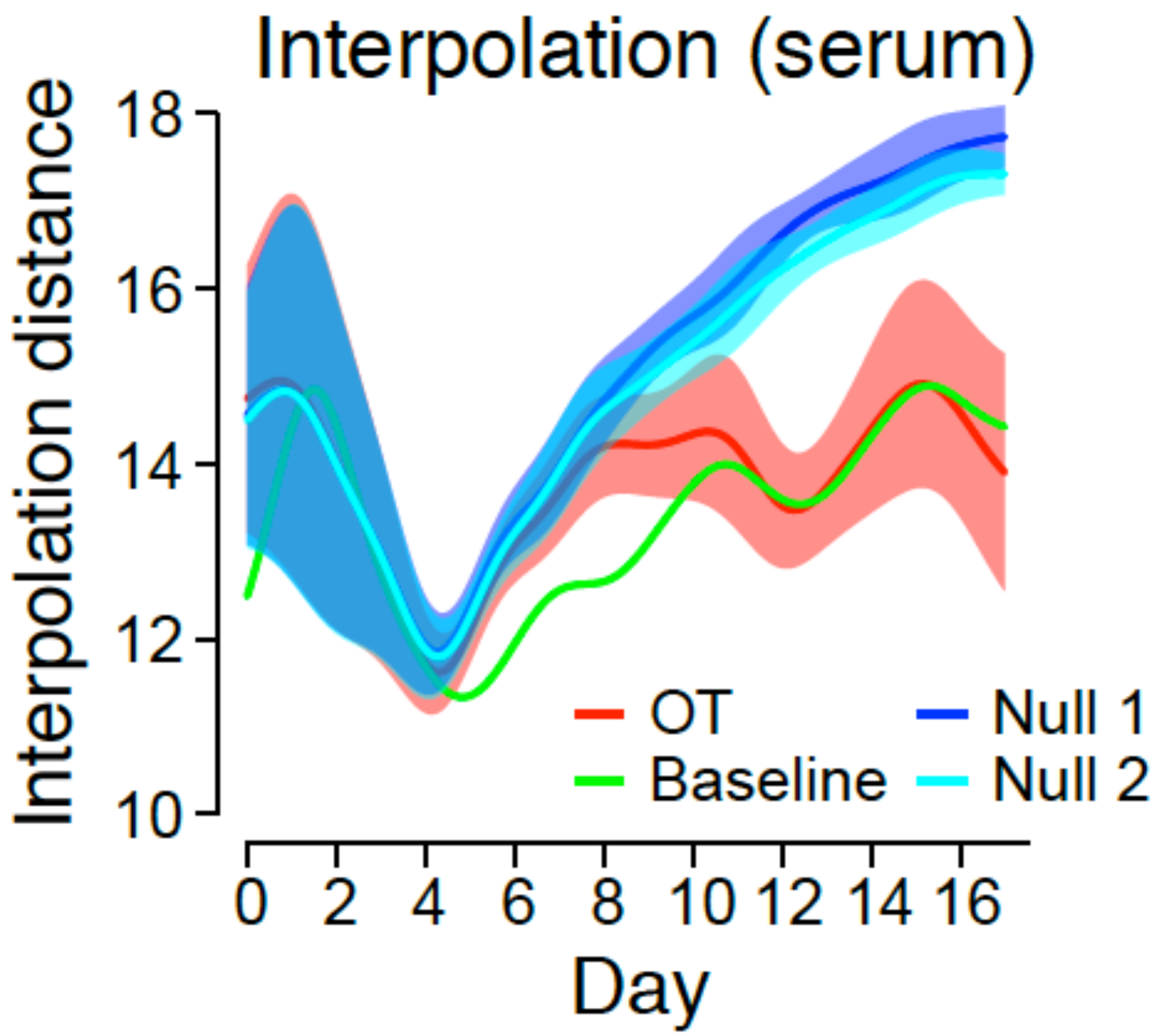
Validation by geodesic interpolation

Analysis:

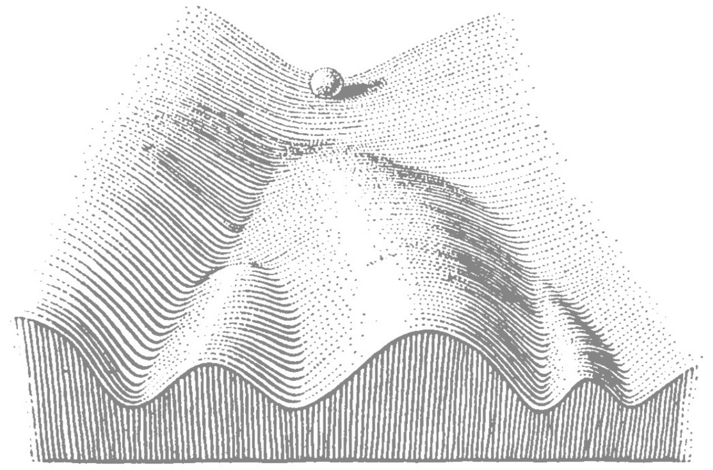
- Compute optimal transport from t_1 to t_3
- Infer distribution at time t_2
- Compare to observed distribution at t_2

Result: Inferred distribution at t_2 matches observed distribution at t_2 almost as well as one batch at t_2 matches other batch at t_2





Learning gene regulatory models



Model time dependence of $\mathbb{P}_t$ as arising from pushing it through a **differential equation**

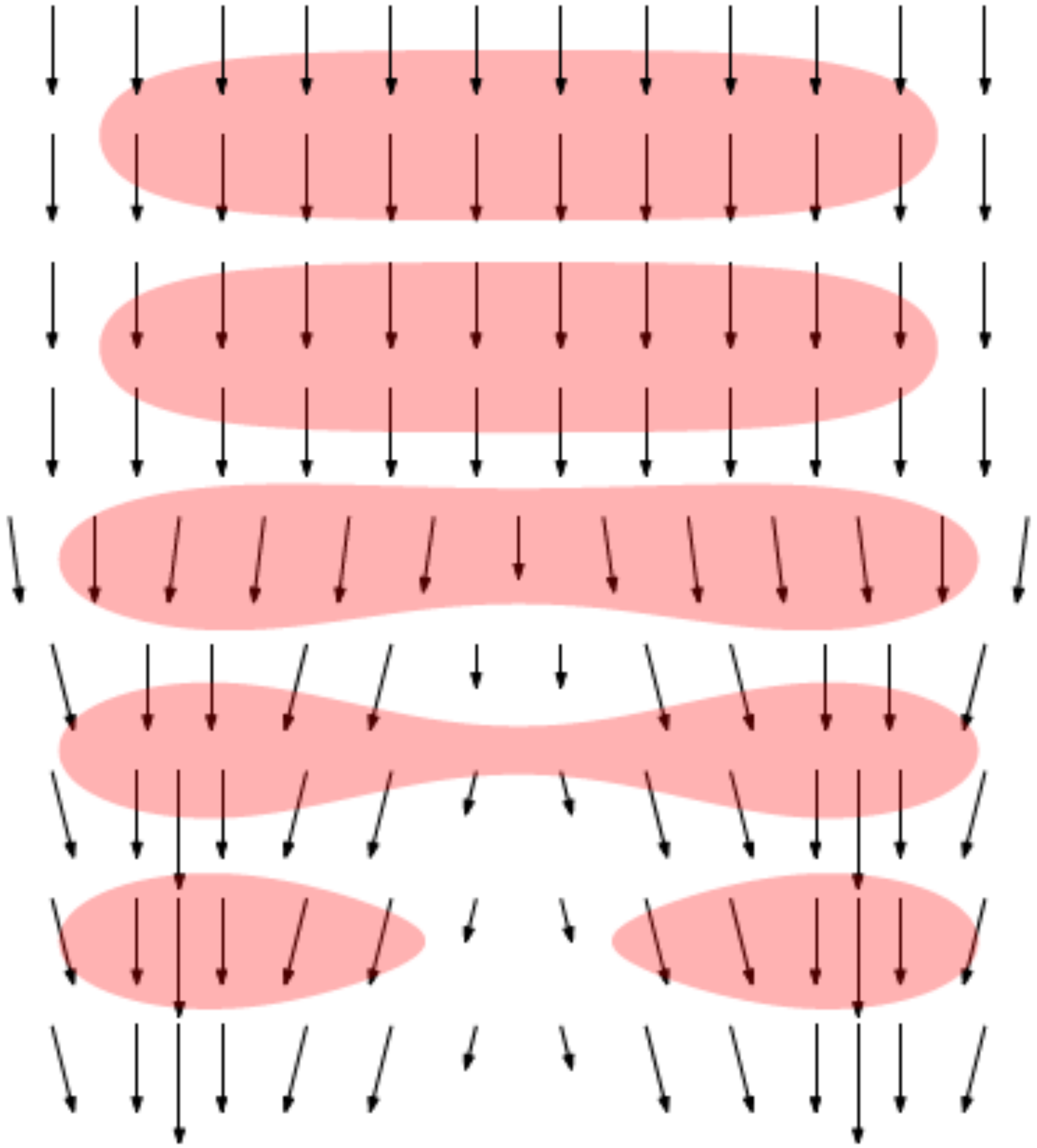
$$\dot{x} = f(x).$$

In discrete time

$$\frac{x_{t_2} - x_{t_1}}{\Delta_t} = f(x_{t_1}).$$

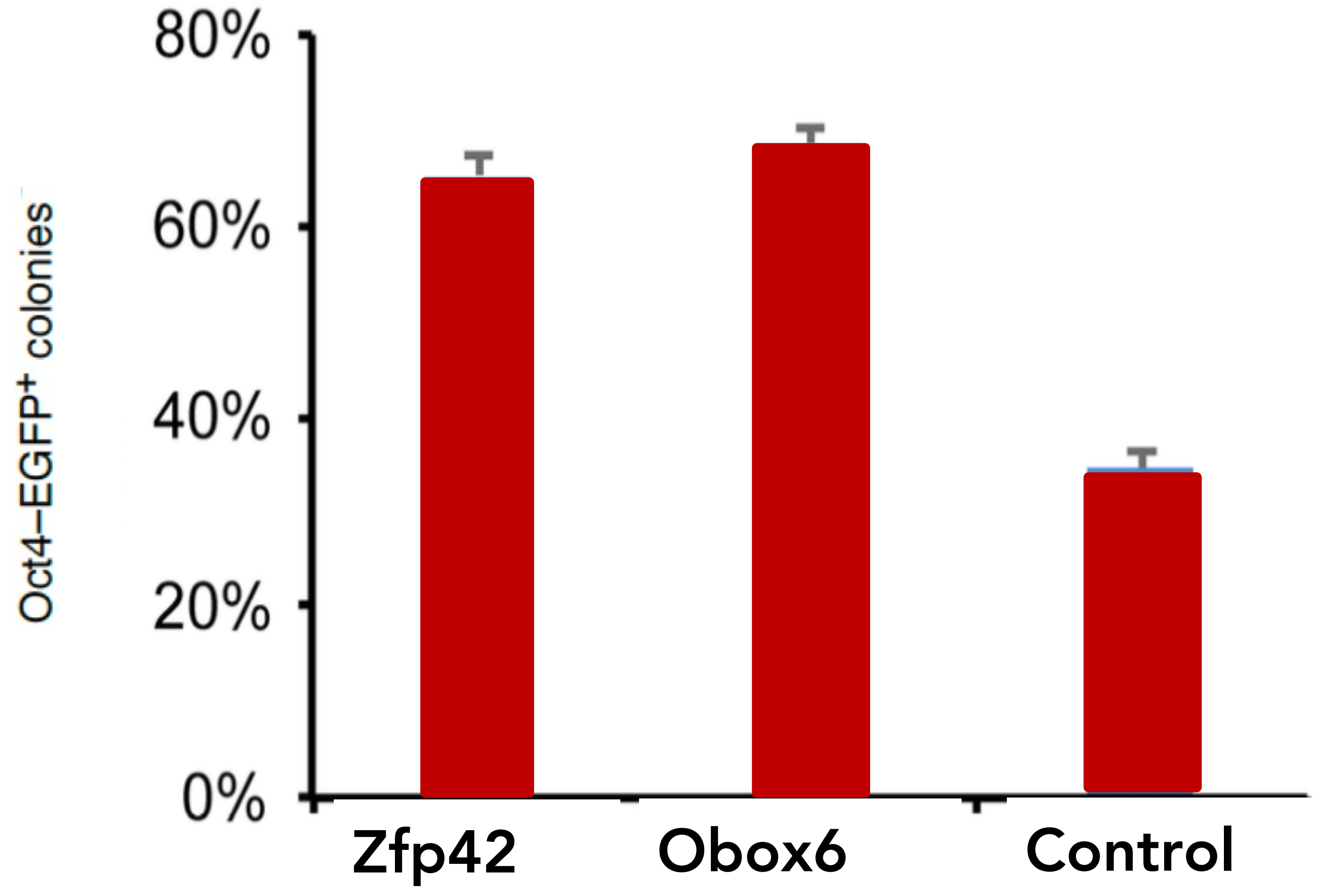
Use estimated couplings $\hat{\pi}$ to generate **training pairs** (x_{t_1}, x_{t_2}) and learn f via regression

$$\underset{f}{\text{minimize}} \qquad \mathbb{E}_{\pi} \|x_{t_2} - f(x_{t_1})\|^2 + \text{reg}(f)$$



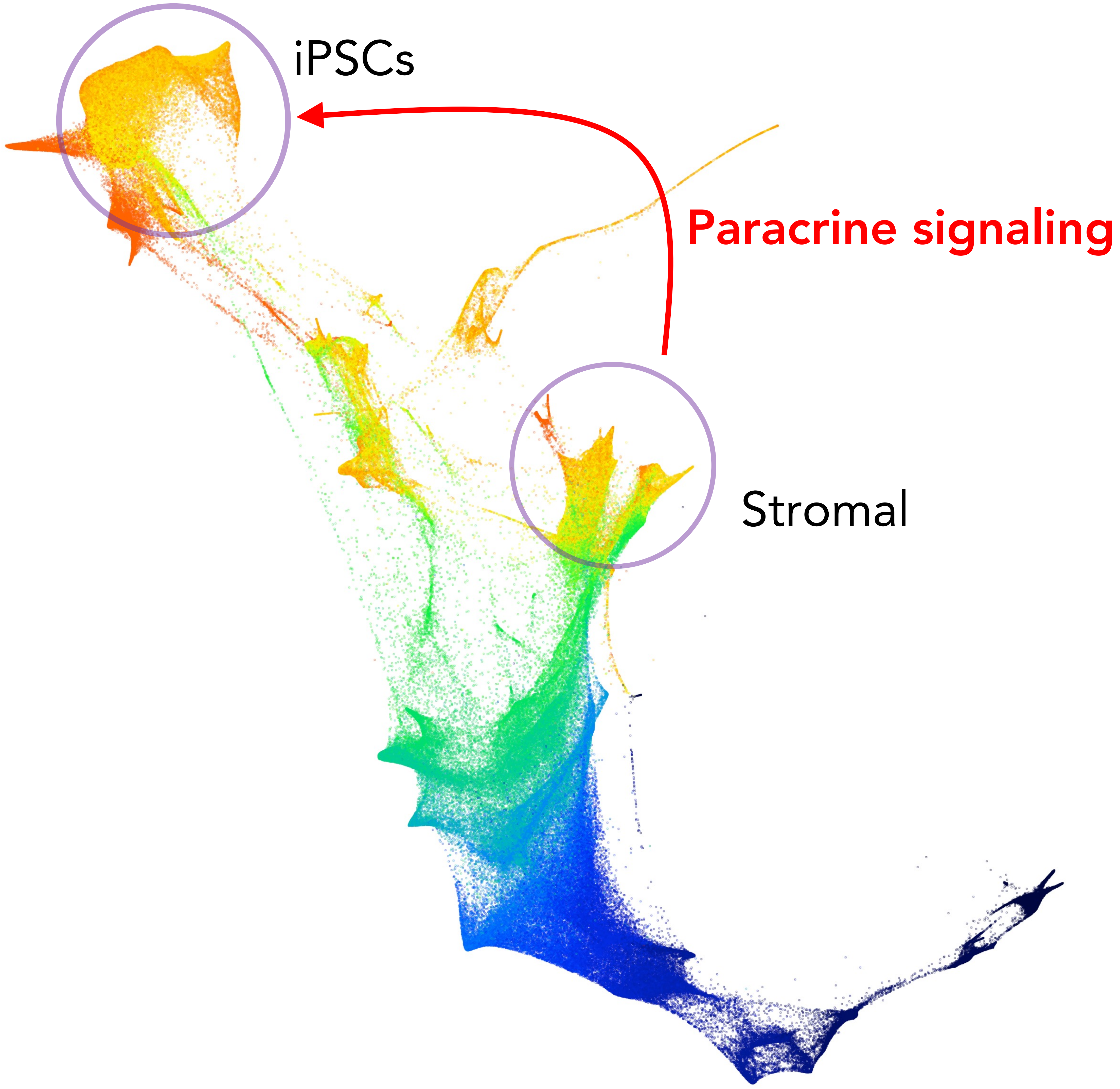
Learning gene regulatory models

Experimental Validation:
We get more stem cells!

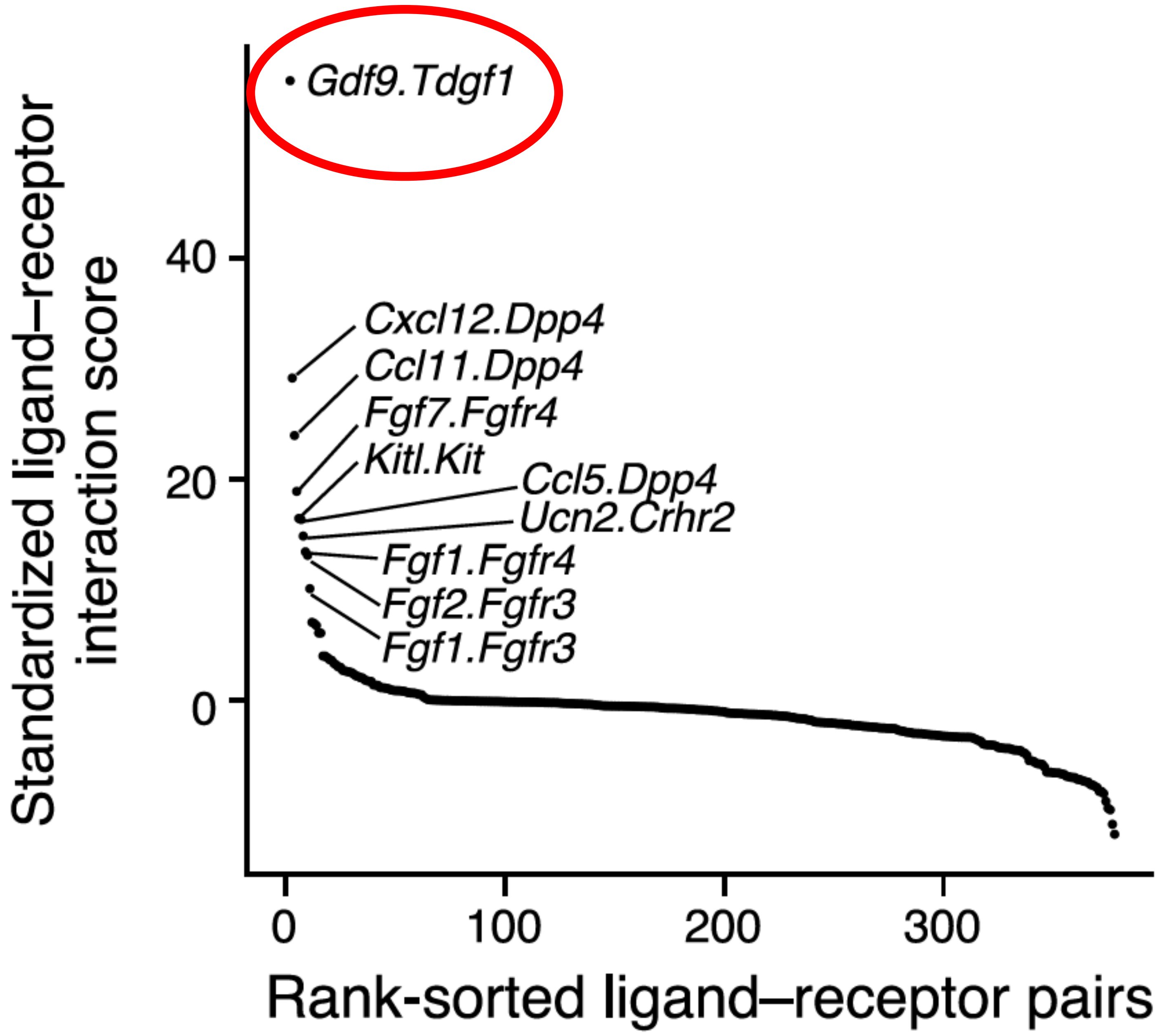


Prediction:
Obox6 drives cells towards iPSCs

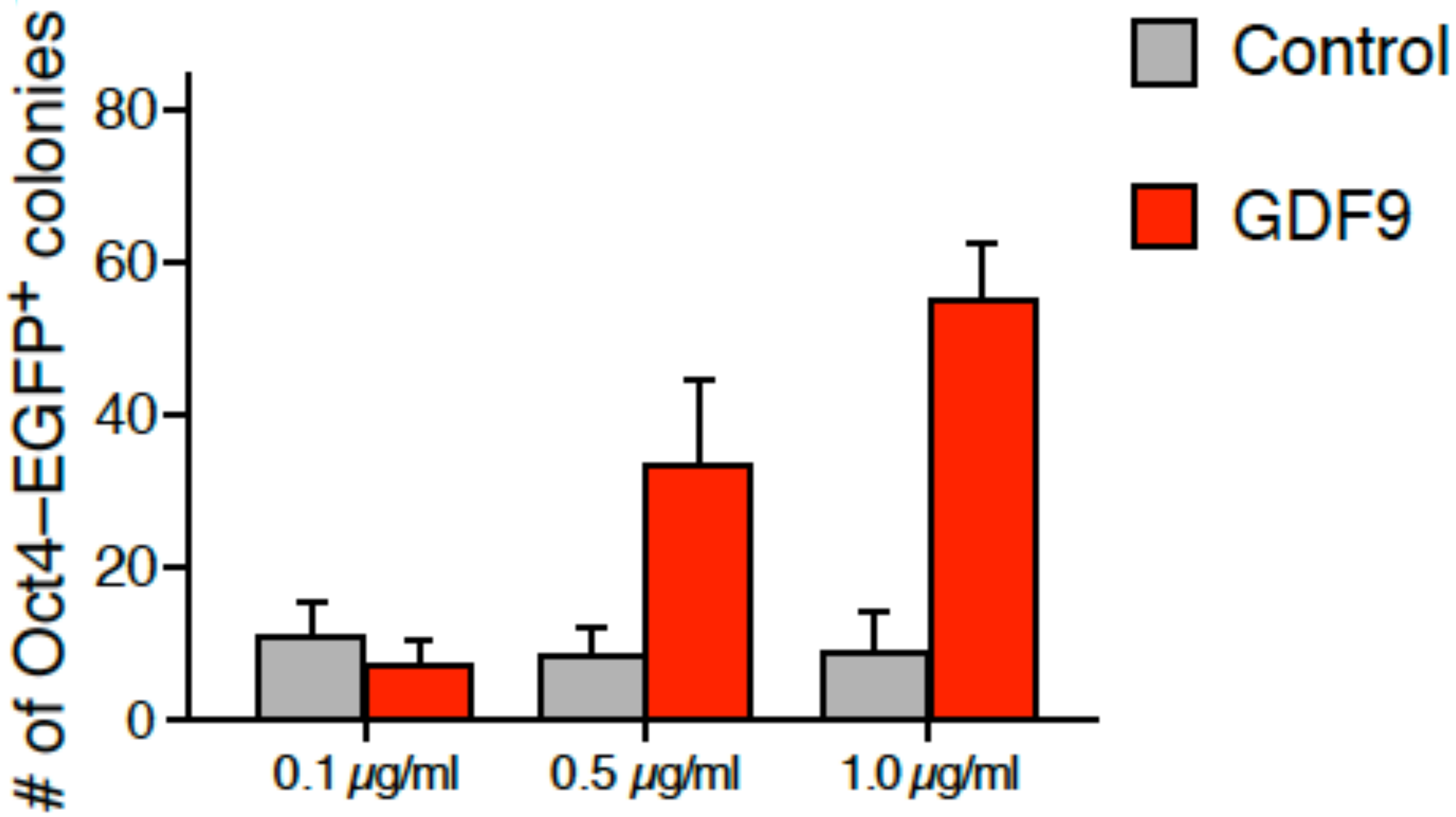
Cell-cell interactions



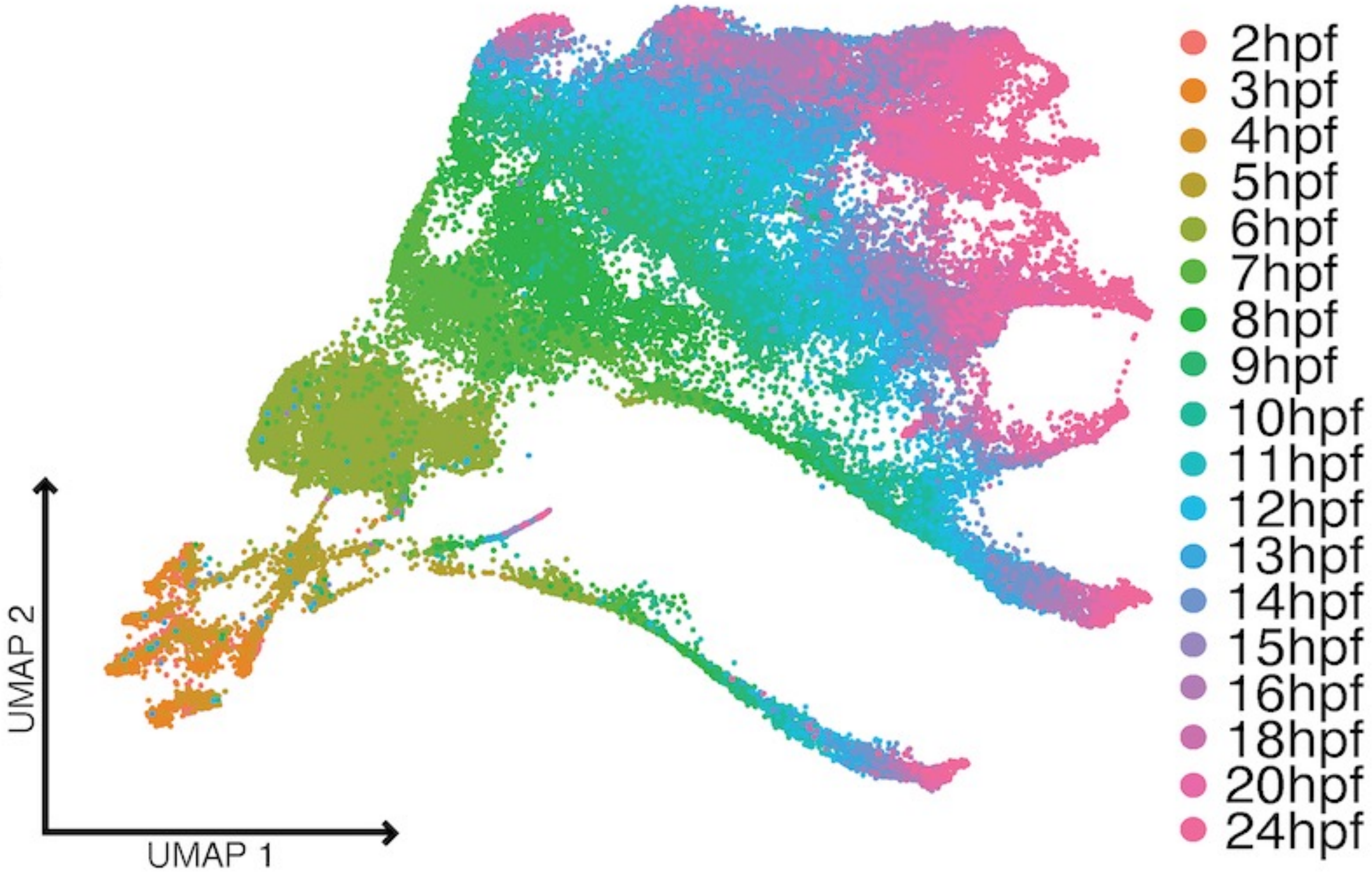
Predicted interactions between iPS and Stromal

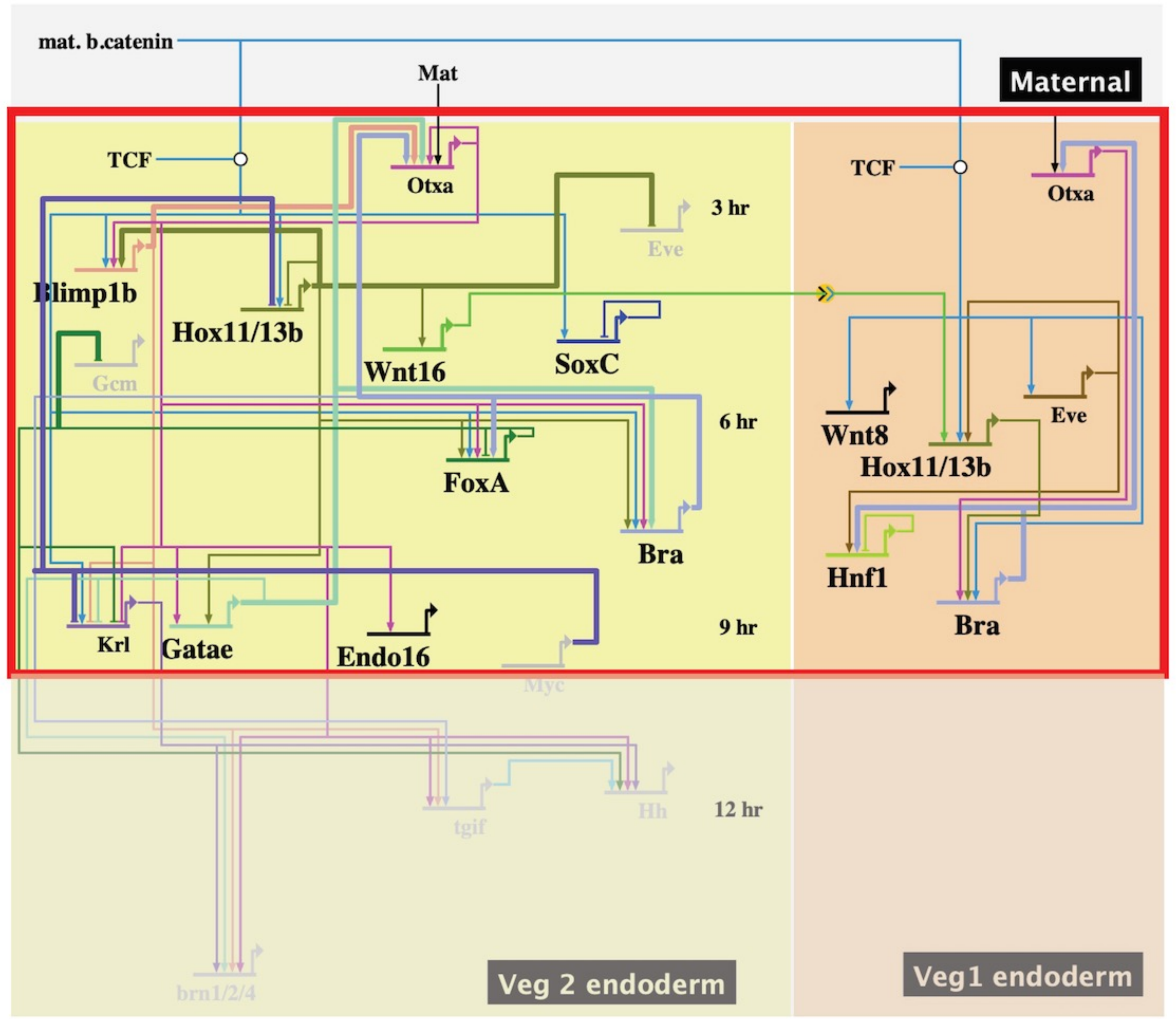


Experimental validation



Rediscovering Gene Regulation in Sea Urchin





AJ Massri

Dave
McClay

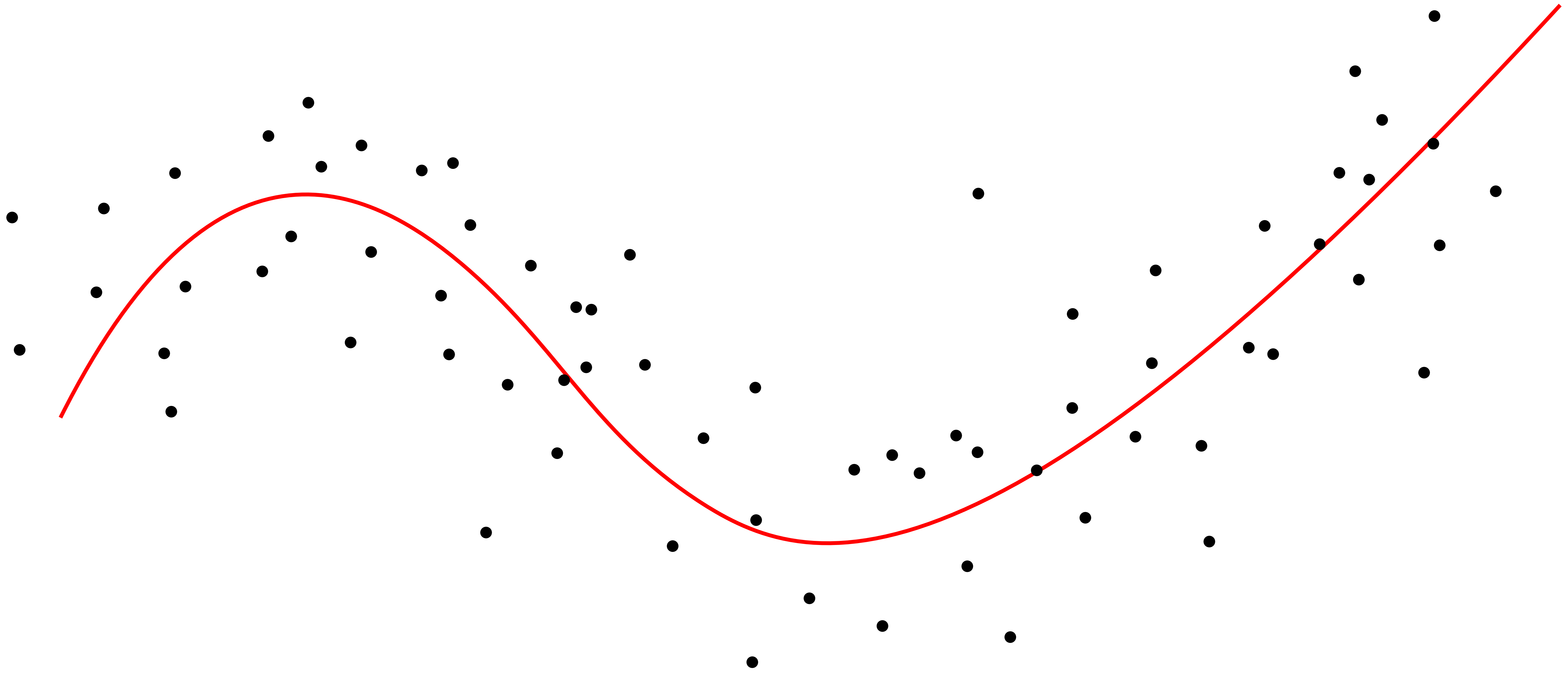
Greg
Wray

Duke University

Massri et al (2021, *Development*)

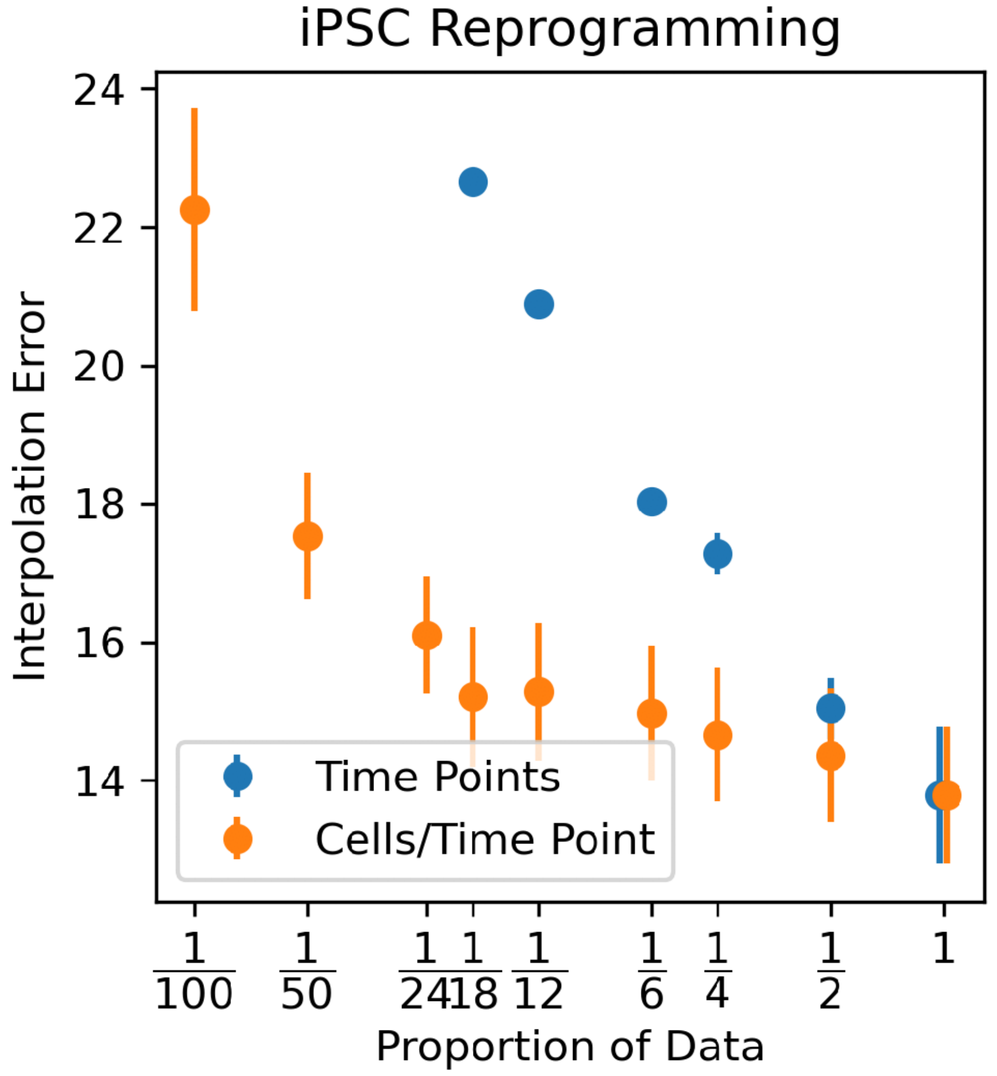
Perspective on experimental design

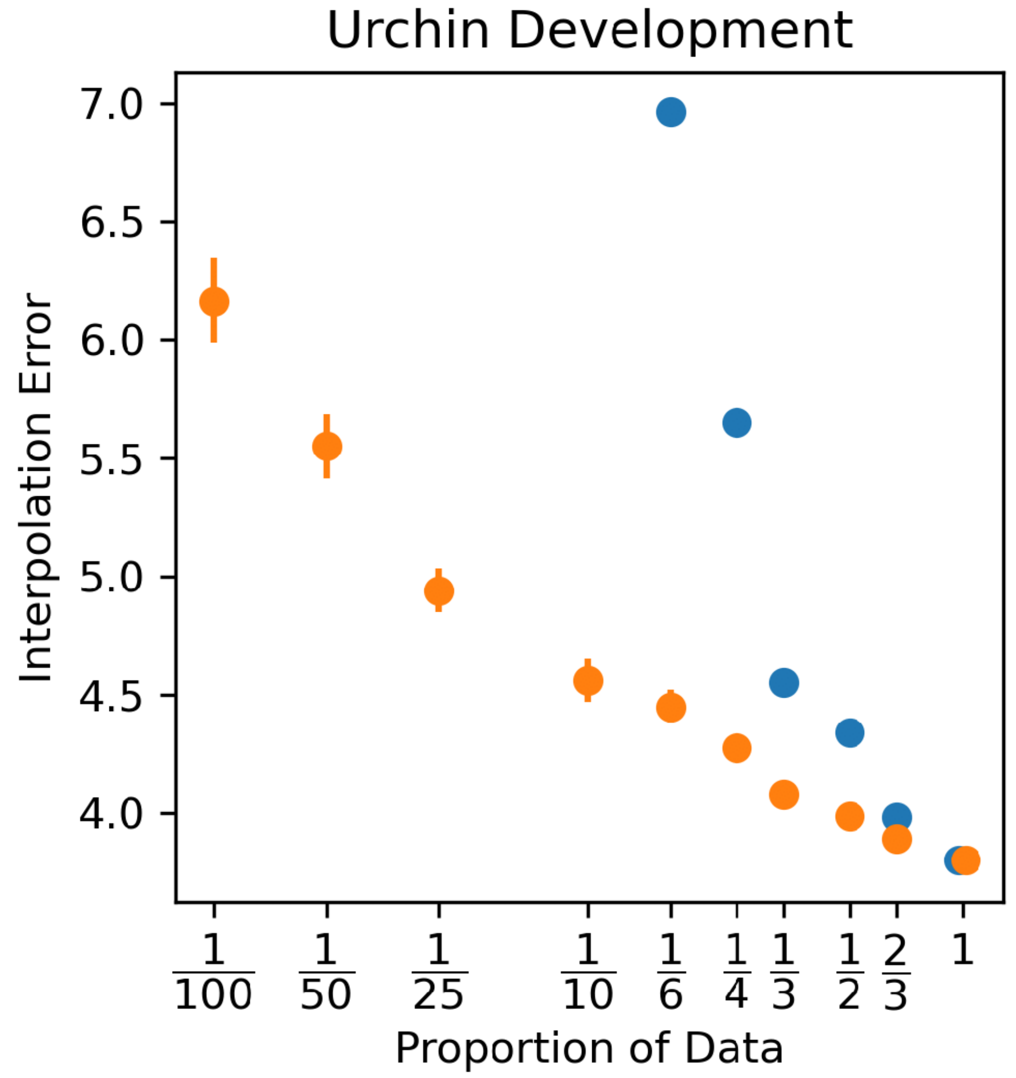
The number of time-points determines the “size of the dataset”.

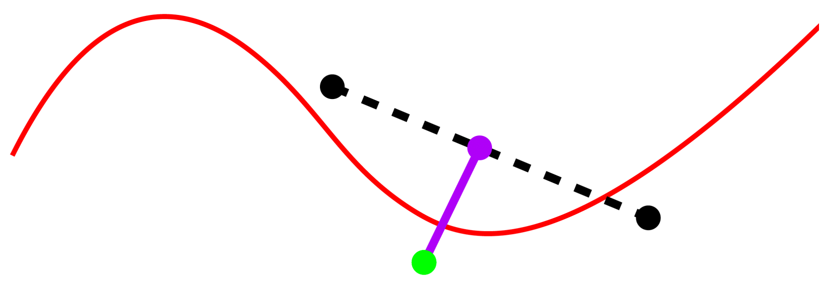


Each time-point is a “data point” along the curve.

The number of cells determines the noise level.

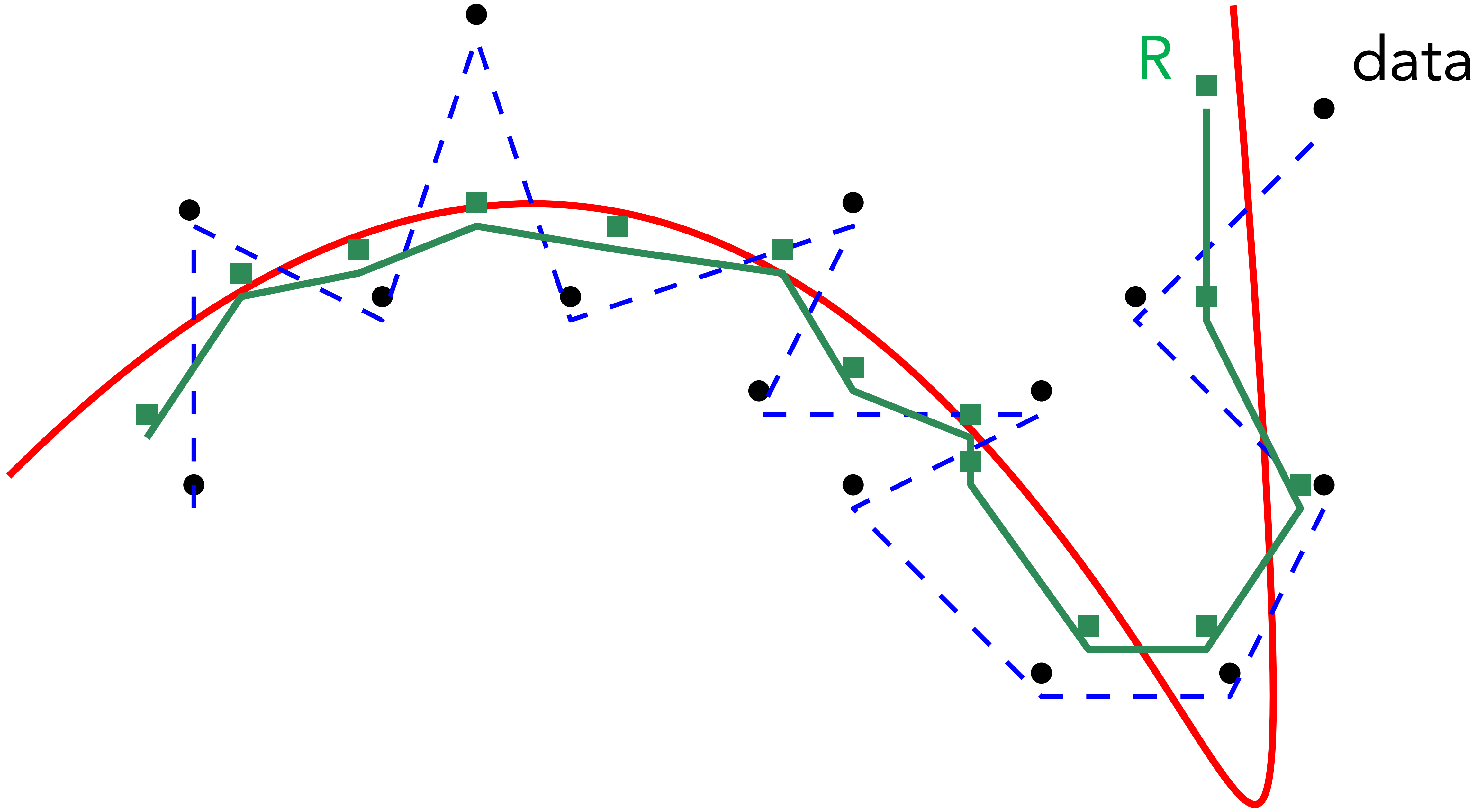


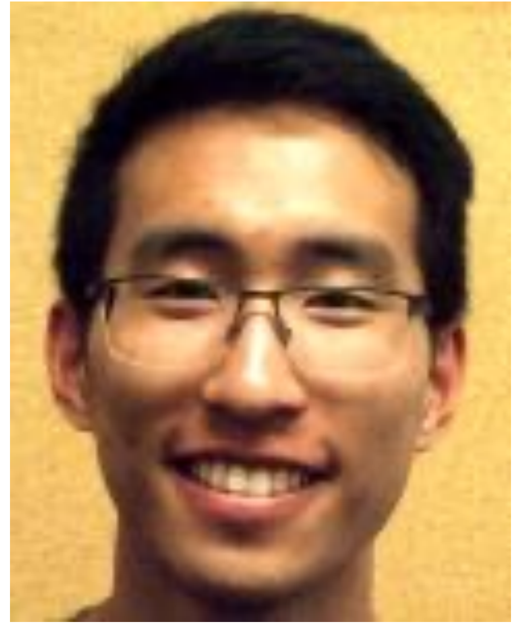




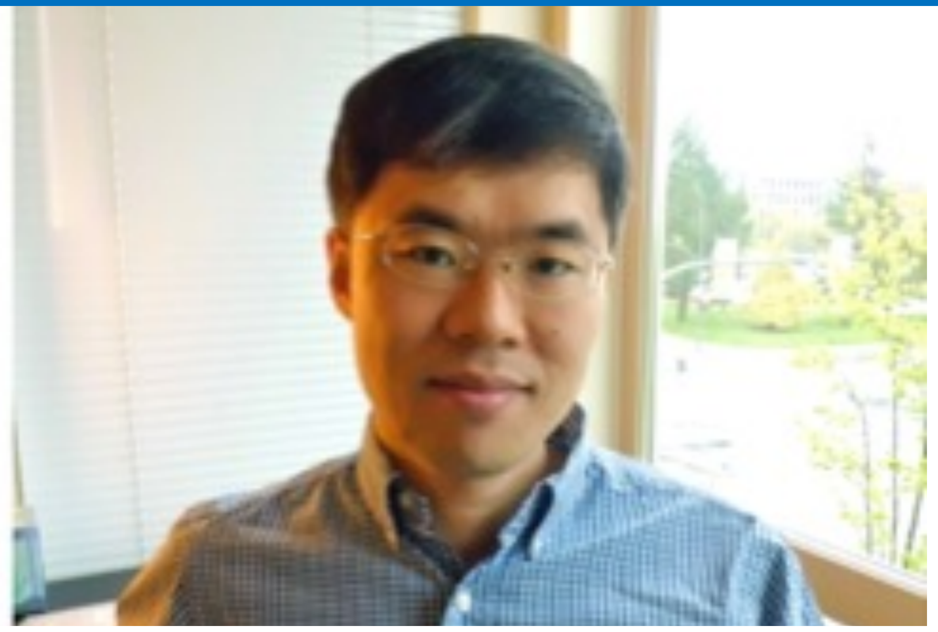
Schiebinger, Current Opinion in Systems Biology 2021

Wasserstein Regression



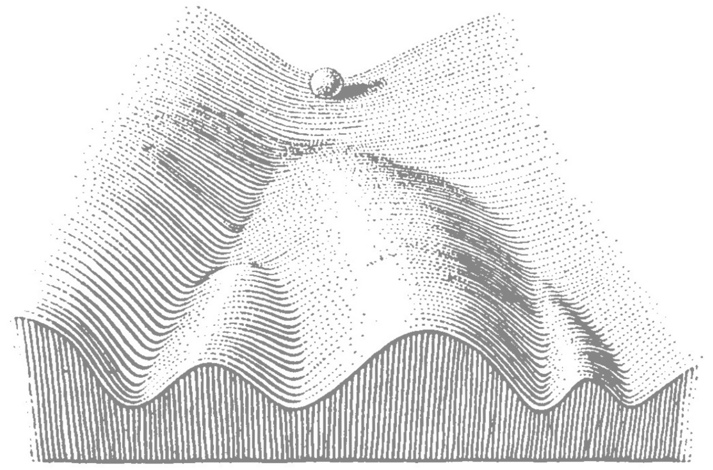

Stephen Zhang


Hugo Lavenant

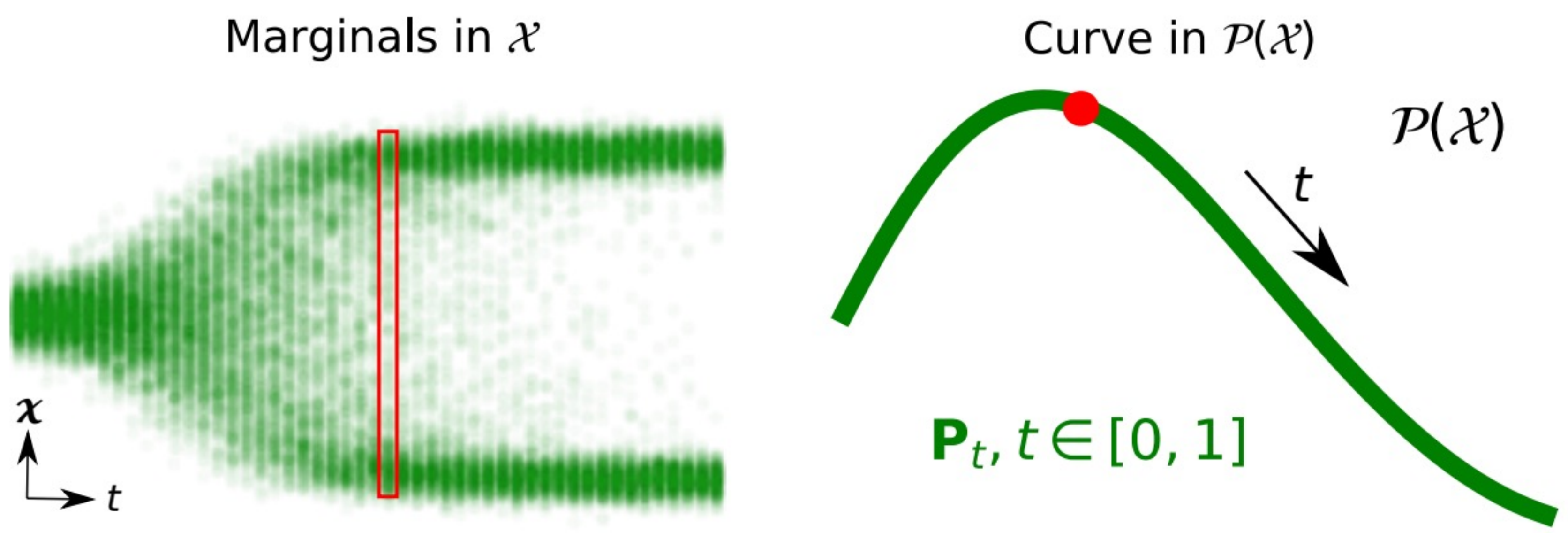

Young-Heon Kim

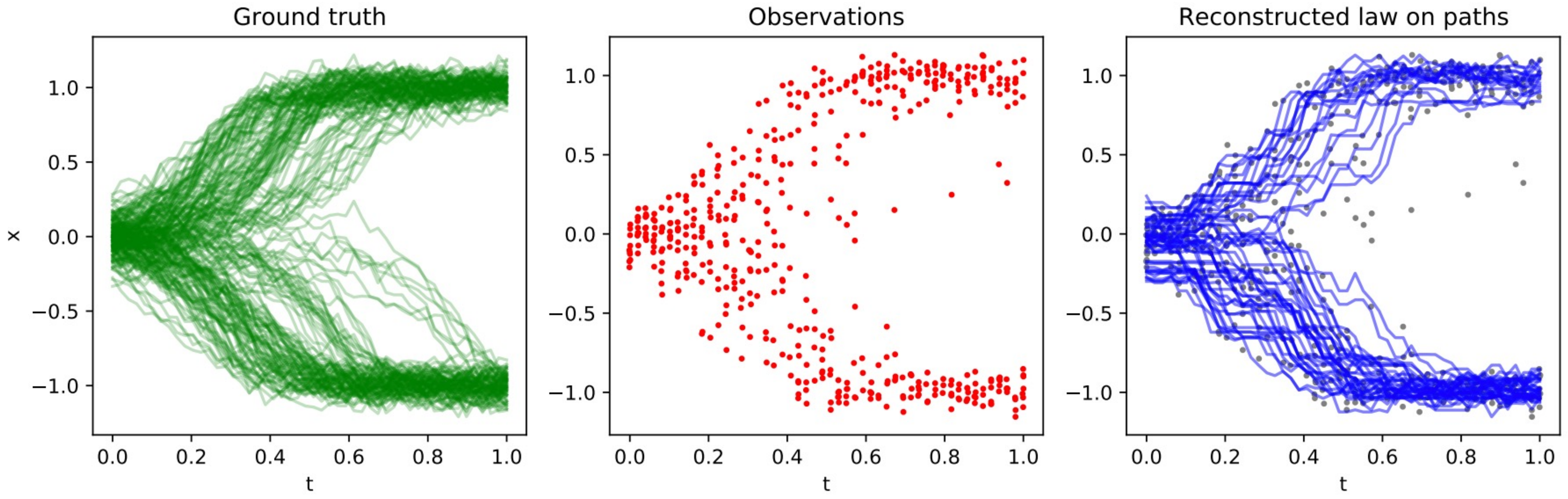
$$\min_{\mathbf{R}} \quad \underbrace{\lambda \mathrm{Reg}(\mathbf{R})}_{\text{penalise overfitting}} + \underbrace{\mathrm{DataFitting}(\mathbf{R})}_{\text{penalise deviating from data}}$$

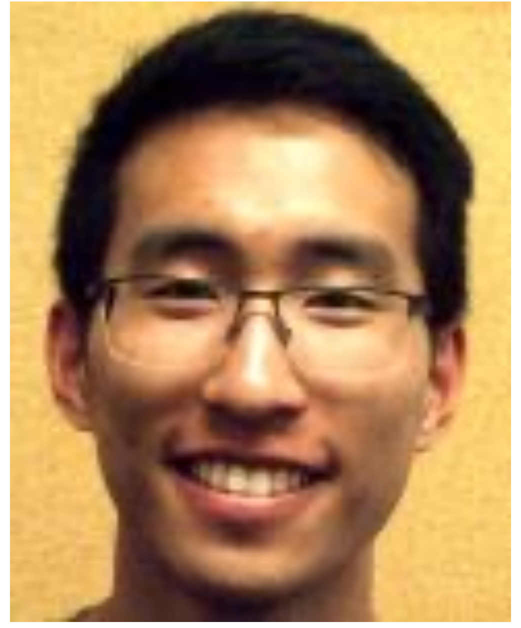
Theorem: Suppose the ground truth is generated by a diffusion + drift (with interactions). Then in the limit of infinitely many time-points (and small regularization), the optimal solution converges to the ground truth.



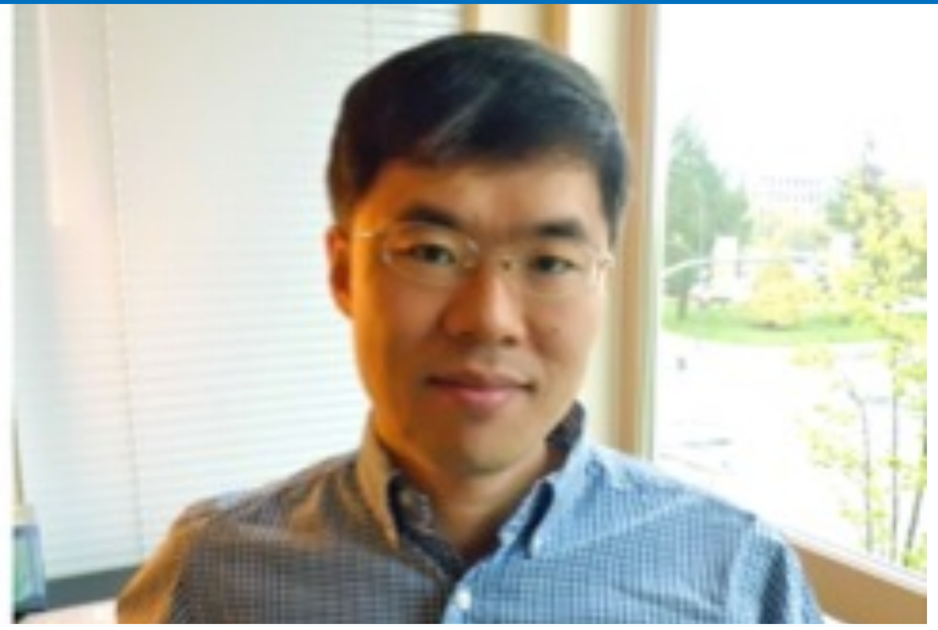
Wasserstein Regression





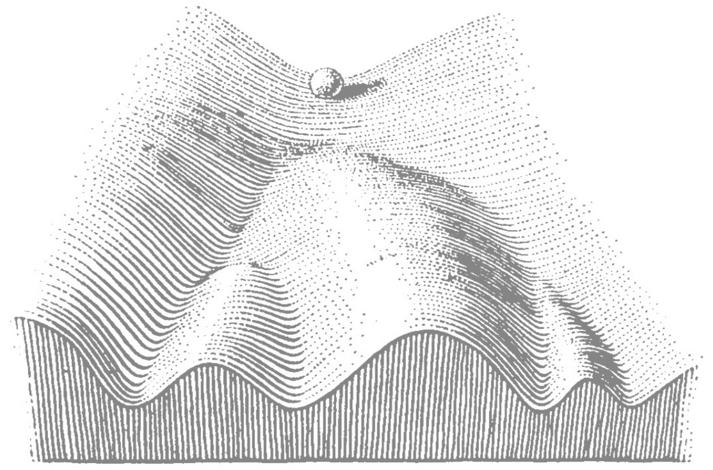

Stephen Zhang


Hugo Lavenant

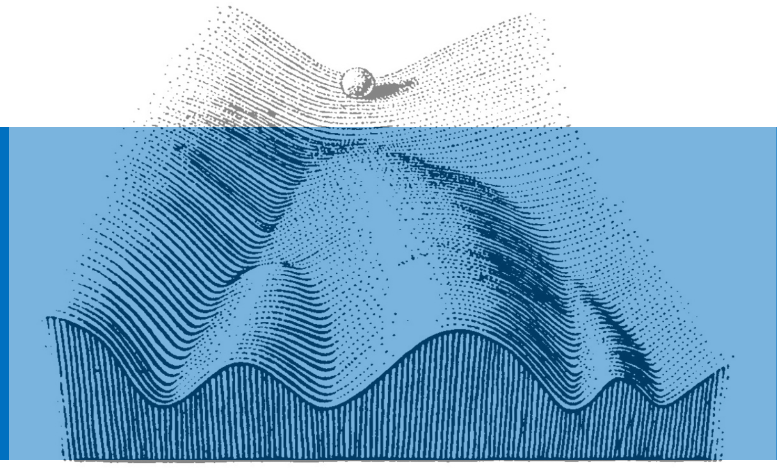

Young-Heon Kim

$$\min_{\mathbf{R}} \quad \underbrace{\lambda \mathrm{Reg}(\mathbf{R})}_{\text{penalise overfitting}} + \underbrace{\mathrm{DataFitting}(\mathbf{R})}_{\text{penalise deviating from data}}$$

Theorem: Suppose the ground truth is generated by a diffusion + drift (with interactions). Then in the limit of infinitely many time-points (and small regularization), the optimal solution converges to the ground truth.



OT captures Waddington's landscape



Theorem

Let **P** be law of trajectories from SDE:

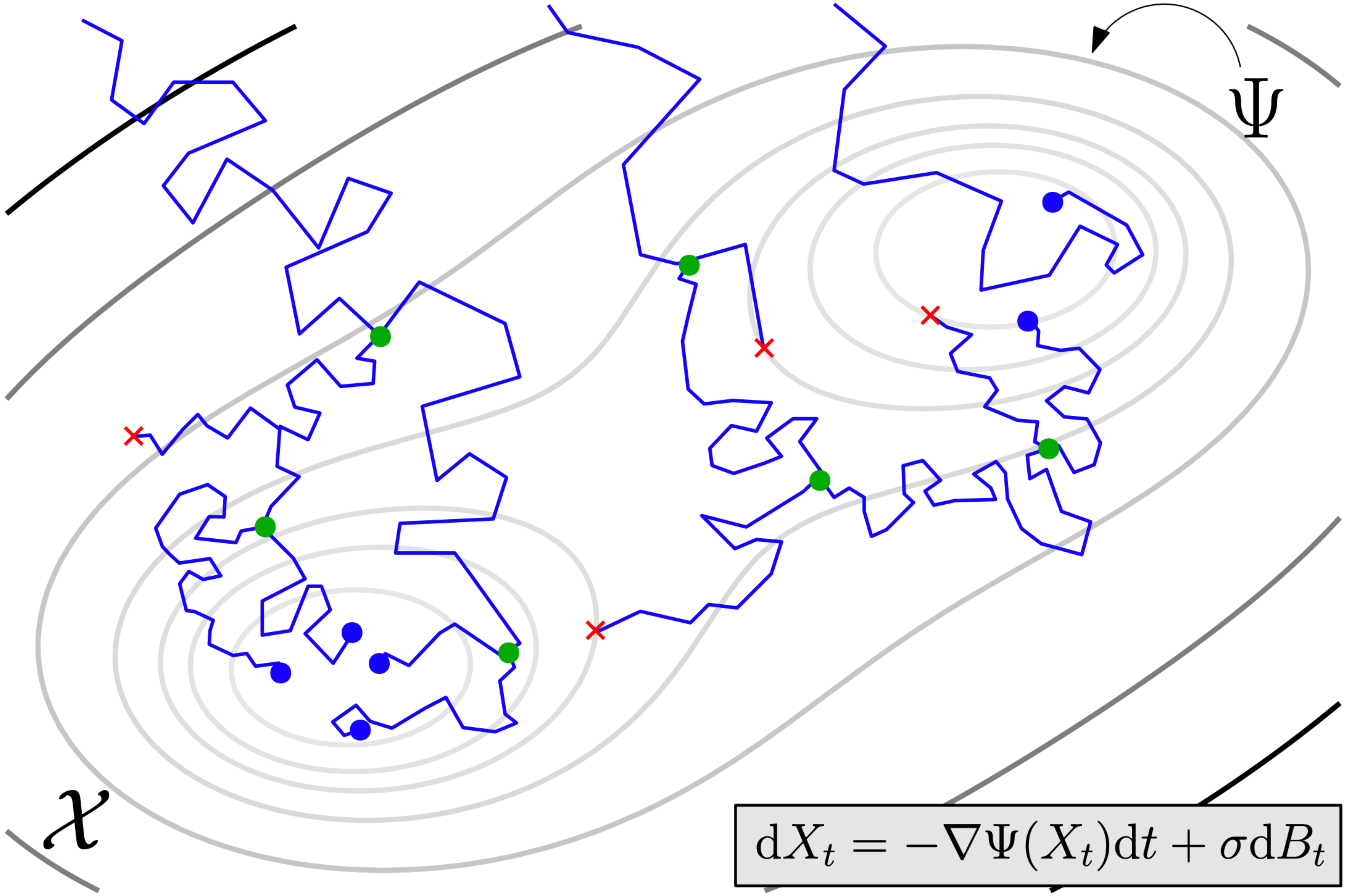
$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

Let $\mathbf{W}^\sigma$ be trajectories of Brownian motion.

For any other trajectories **R**, with the same marginals,

$$\mathrm{H}(\mathbf{P}|\mathbf{W}^\sigma) \leqslant \mathrm{H}(\mathbf{R}|\mathbf{W}^\sigma)$$

This means we can recover **P** from snapshots!



$$\mathrm{d}X_t = -\nabla\Psi(X_t)\mathrm{d}t + \sigma\mathrm{d}B_t$$

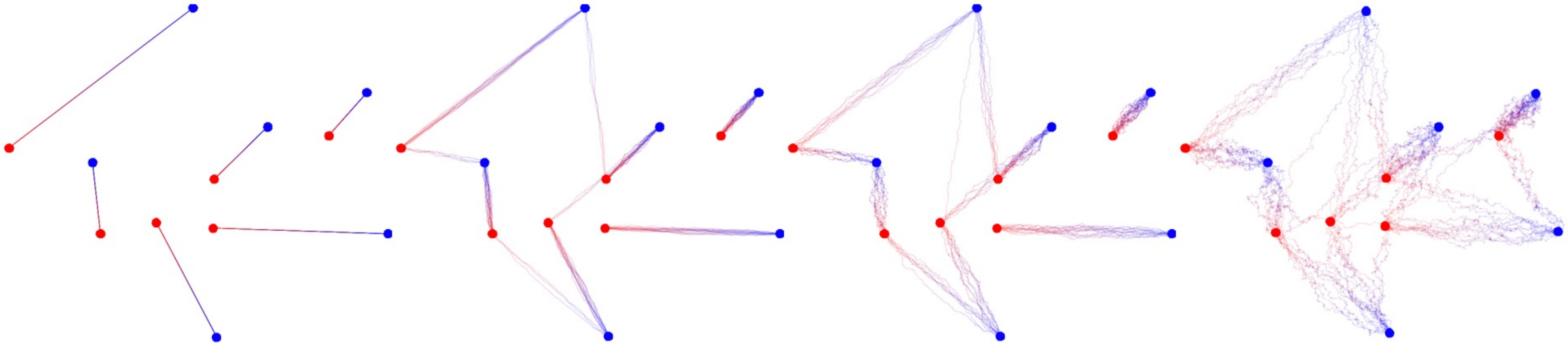
Lavenant et al, 2021 arxiv

Optimal transport and entropy minimization

Min Ent
over space of paths

Min Ent
over coupling matrices

Entropic OT
over coupling matrices



$\varepsilon = 0$ $\varepsilon = .05$ $\varepsilon = 0.2$ $\varepsilon = 1$

Optimal transport and entropy minimization

Min Ent
over space of paths

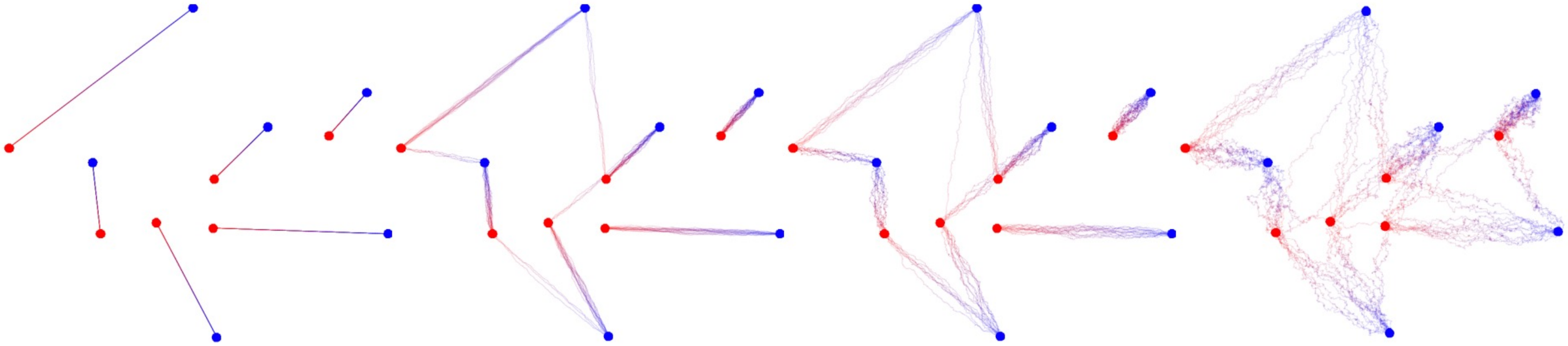
$$\underset{\pi\in\mathcal{P}(\Omega)}{\text{minimize}} \quad \mathrm{H}\left(\pi|W^{\epsilon}\right) \qquad \text{s.t.} \quad \pi_0 = \mu_0, \quad \pi_1 = \mu_1$$

Min Ent
over coupling matrices

$$\underset{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X})}{\text{minimize}} \quad \mathrm{H}\left(\pi|K\right) \qquad \text{s.t.} \quad \sum_y \pi(x,y) = \mu_0(x), \quad \sum_x \pi(x,y) = \mu_1(y)$$

Entropic OT
over coupling matrices

$$\underset{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X})}{\text{minimize}} \quad \sum_{x,y} c(x,y)\pi(x,y) + \epsilon H(\pi) \qquad \text{s.t.} \quad \sum_y \pi(x,y) = \mu_0(x), \quad \sum_x \pi(x,y) = \mu_1(y)$$



$\varepsilon = 0$ $\varepsilon = .05$ $\varepsilon = 0.2$ $\varepsilon = 1$

Optimal transport and entropy minimization

Min Ent
over space of paths

$$\underset{R\in\mathcal{P}(\Omega)}{\text{minimize}} \quad \mathrm{H}\left(R|W^{\epsilon}\right)$$

$$\text{s.t.} \quad R_0 = \mu_0$$

$$R_1 = \mu_1$$

$$\mathcal{H}(\pi|K) = \sum_{i,j} \pi_{i,j} \log \frac{\pi_{i,j}}{K_{i,j}}$$

Min Ent
over coupling matrices

$$K = e^{-\frac{C}{\epsilon}}$$

$$\underset{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X})}{\text{minimize}} \quad \mathrm{H}\left(\pi|K\right)$$

$$\text{s.t.} \quad \sum_y \pi(x,y) = \mu_0(x)$$

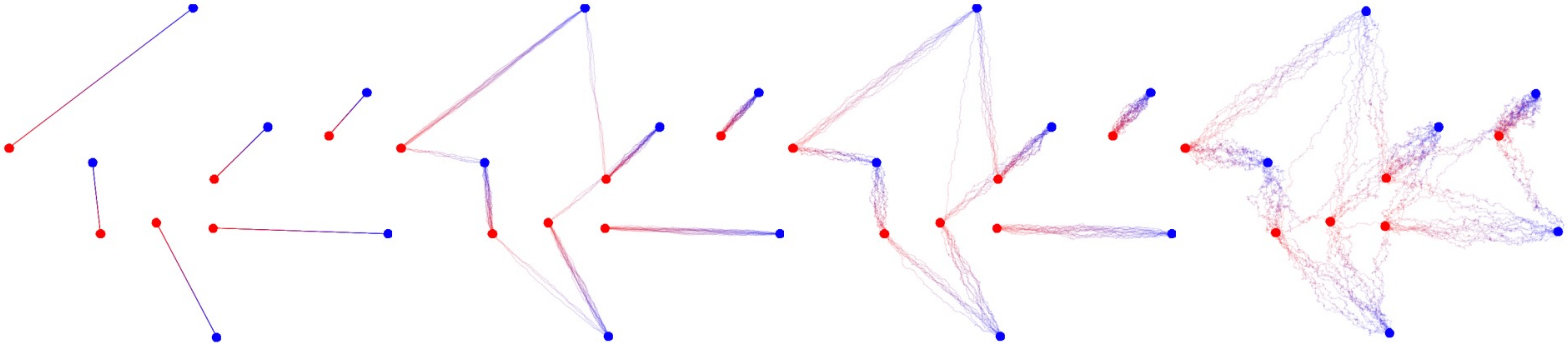
$$\sum_x \pi(x,y) = \mu_1(y)$$

Entropic OT
over coupling matrices

$$\underset{\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{X})}{\text{minimize}} \quad \sum_{x,y} c(x,y)\pi(x,y) + \epsilon H(\pi)$$

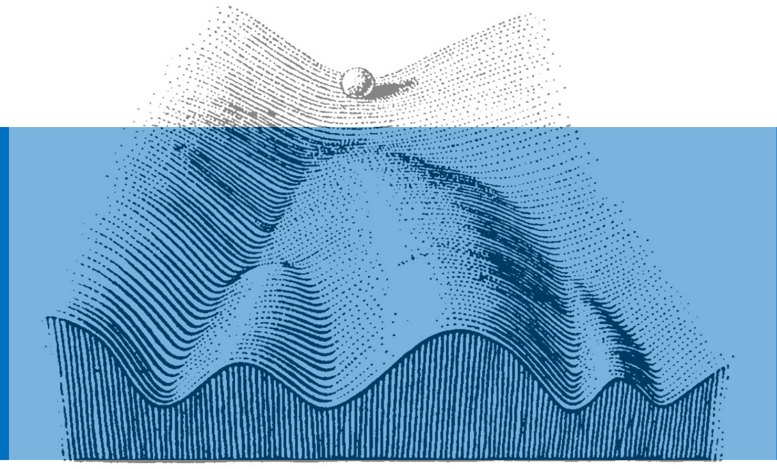
$$\text{s.t.} \quad \sum_y \pi(x,y) = \mu_0(x)$$

$$\sum_x \pi(x,y) = \mu_1(y)$$



$\varepsilon = 0$ $\quad$ $\varepsilon = .05$ $\quad$ $\varepsilon = 0.2$ $\quad$ $\varepsilon = 1$

OT captures Waddington's landscape



Theorem

Let **P** be law of trajectories from SDE:

$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

Let $\mathbf{W}^\sigma$ be trajectories of Brownian motion.

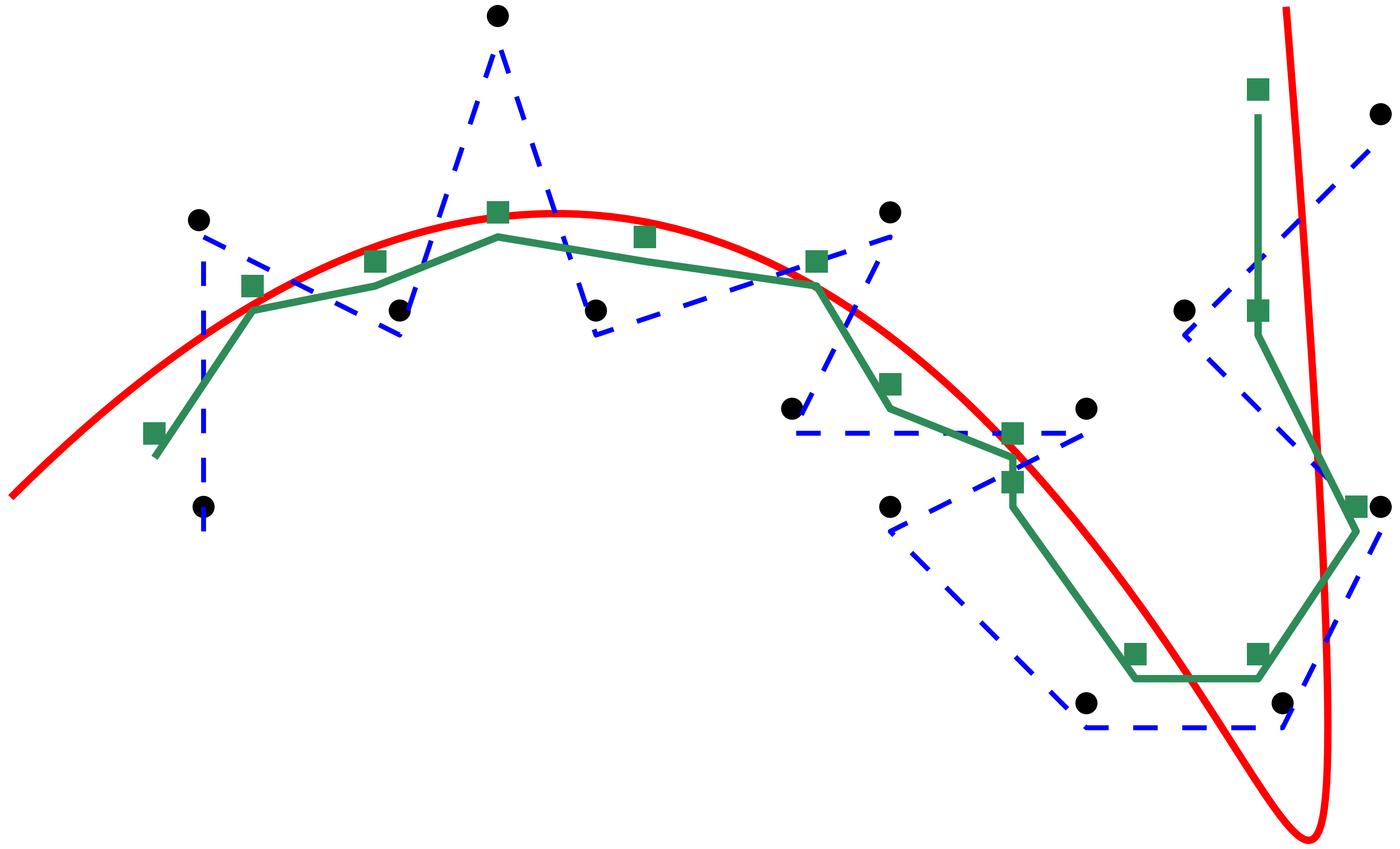
For any other trajectories **R**, with the same marginals,

$$\mathrm{H}(\mathbf{P}|\mathbf{W}^\sigma) \leqslant \mathrm{H}(\mathbf{R}|\mathbf{W}^\sigma)$$

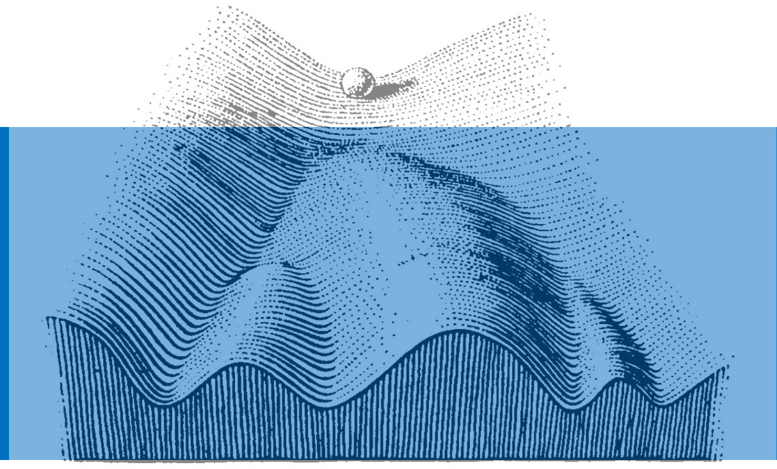
This means we can recover **P** from snapshots!

Lavenant et al, 2021 arxiv

Therefore "entropic OT" recovers the true trajectories from fully observed marginals



OT captures Waddington’s landscape



Theorem

Let **P** be law of trajectories from SDE:

$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

Let $\mathbf{W}^\sigma$ be trajectories of Brownian motion.

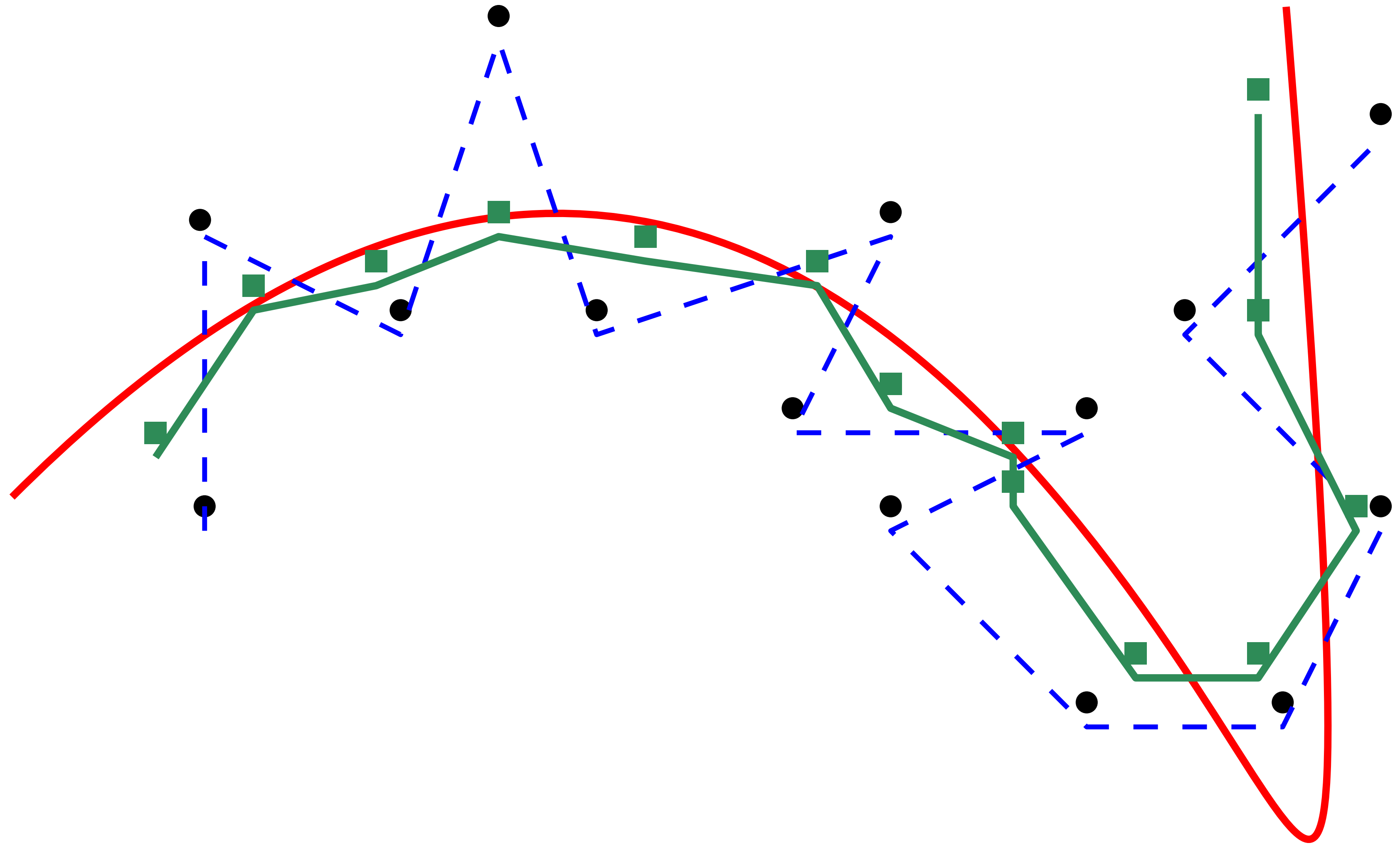
For any other trajectories **R**, with the same marginals,

$$\mathrm{H}(\mathbf{P}|\mathbf{W}^\sigma) \leqslant \mathrm{H}(\mathbf{R}|\mathbf{W}^\sigma)$$

This means we can recover **P** from snapshots!

Lavenant et al, 2021 arxiv

Therefore “entropic OT” recovers the true trajectories from fully observed marginals



What about finite data?

Trajectory inference via Min Ent (in path space)

Estimator Min-entropy relative to Wiener measure

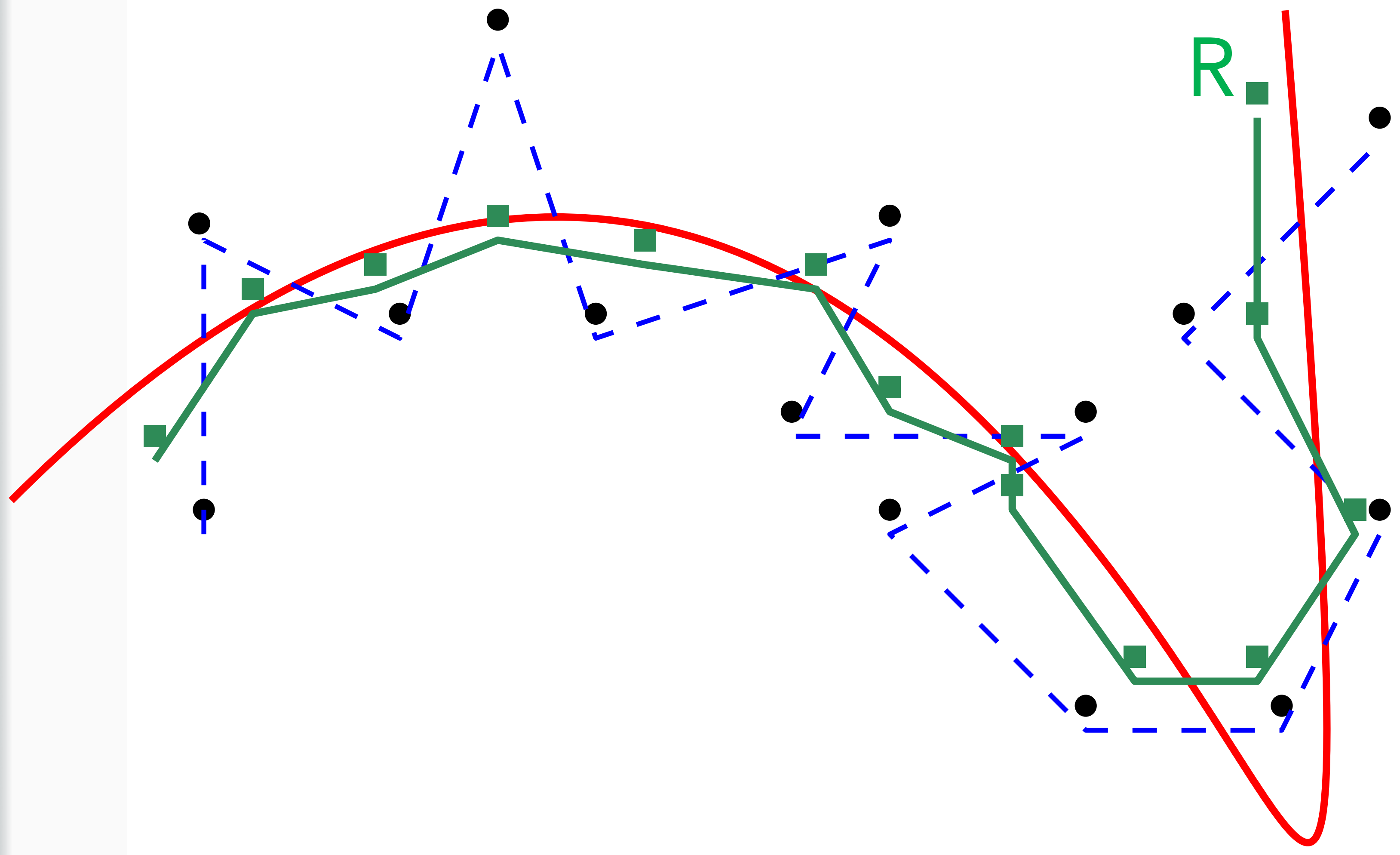
$$R^* := \operatorname*{argmin}_{R \in \mathcal{P}(\Omega)} \mathcal{F}(R), \quad \mathcal{F}(R) := \mathrm{Fit}_{\lambda,\sigma}(R_{t_1}, \ldots, R_{t_T}) + \tau H(R|W^\tau)$$

- $W^\tau \in \mathcal{P}(\Omega)$ is the law of the Brownian motion at temperature τ (reversible, reflected, on $\mathcal{X}$)
- $H(\mu|\nu) = \int \log(\,\mathrm{d}\mu/\,\mathrm{d}\nu)\,\mathrm{d}\mu$ is the relative entropy
- see next slide for $\mathrm{Fit}_{\lambda,\sigma}$

Theorem [Lavenant et al. 2021]

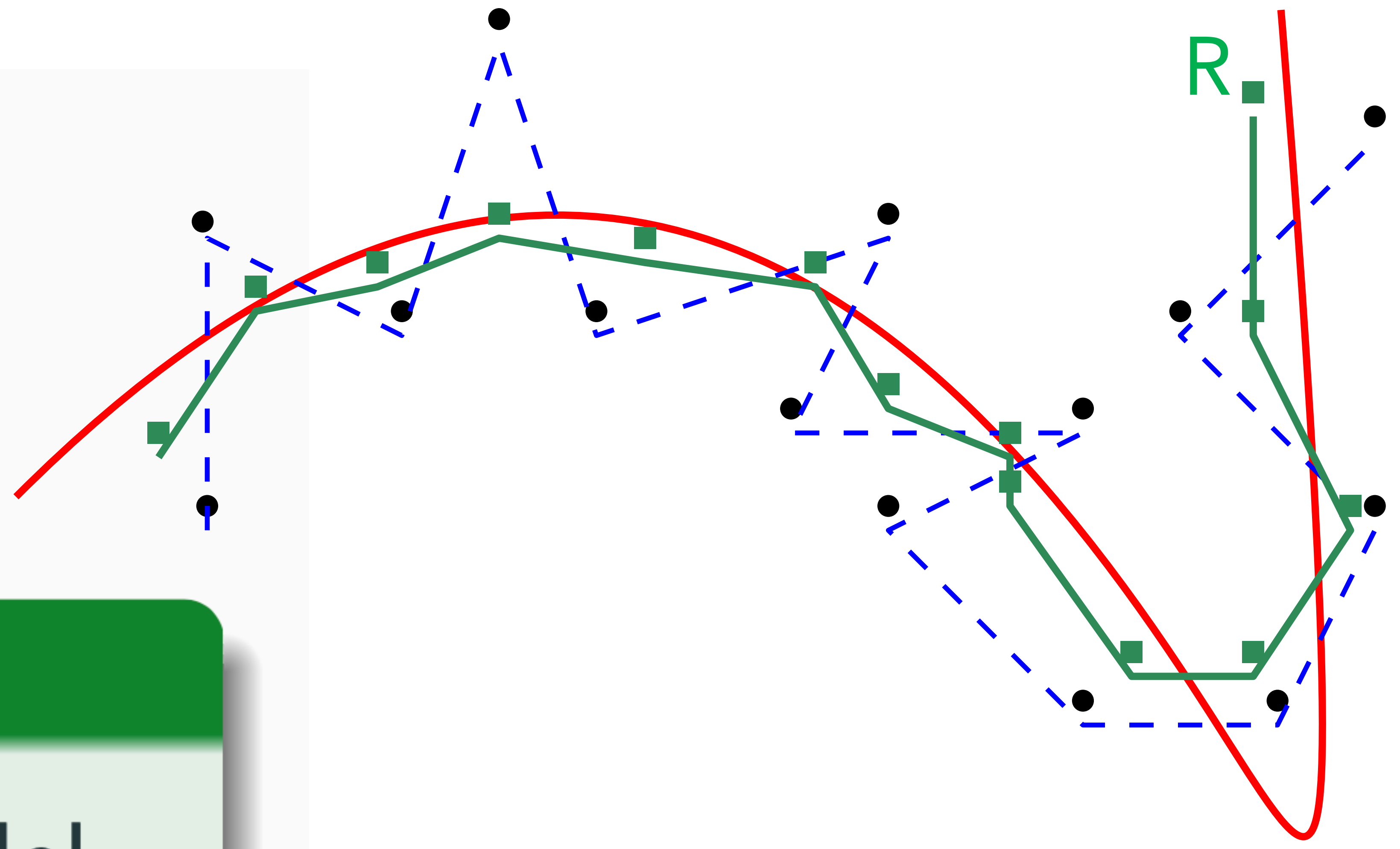
If $(t_i)_{i \in [T]}$ becomes dense in $[0,1]$ as T grows, then

$$\lim_{\lambda,\sigma \to 0} \lim_{T \to \infty} R^* = P \quad \text{weakly, a.s.}$$



Data fitting term

$$\mathrm{Fit}_{\lambda,\sigma}(R_{t_1},\dots,R_{t_T}) := \frac{1}{\lambda}\sum_{i=1}^{T}(\Delta t_i)\widetilde{\mathrm{Fit}}_\sigma(R_{t_i}|\hat{\mathbb{P}}_{t_i})$$



Log-likelihood fitting loss

Let $\widetilde{\mathrm{Fit}}_\sigma$ be the neg-log-likelihood under noisy observation model

$$\hat{X}_{t_i,j} = X_{t_i,j} + \sigma Z_{i,j}, \quad X_{t_i,j} \sim R_{t_i}, \quad Z_{i,j} \sim \mathcal{N}(0,1)$$

$$\widetilde{\mathrm{Fit}}_\sigma(R_{t_i}|\hat{\mathbb{P}}_{t_i}) := \int -\log\left(\int \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)\mathrm{d}R_{t_i}(x)\right)\mathrm{d}\hat{\mathbb{P}}_{t_i}(y)$$

- Convex, smooth in R: as nice as one could hope

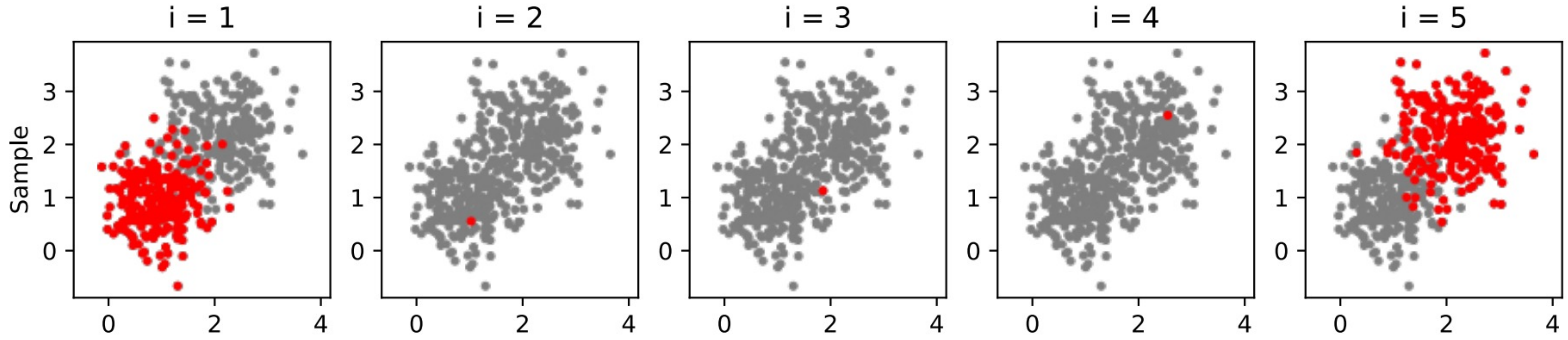
How can we solve this optimization problem?

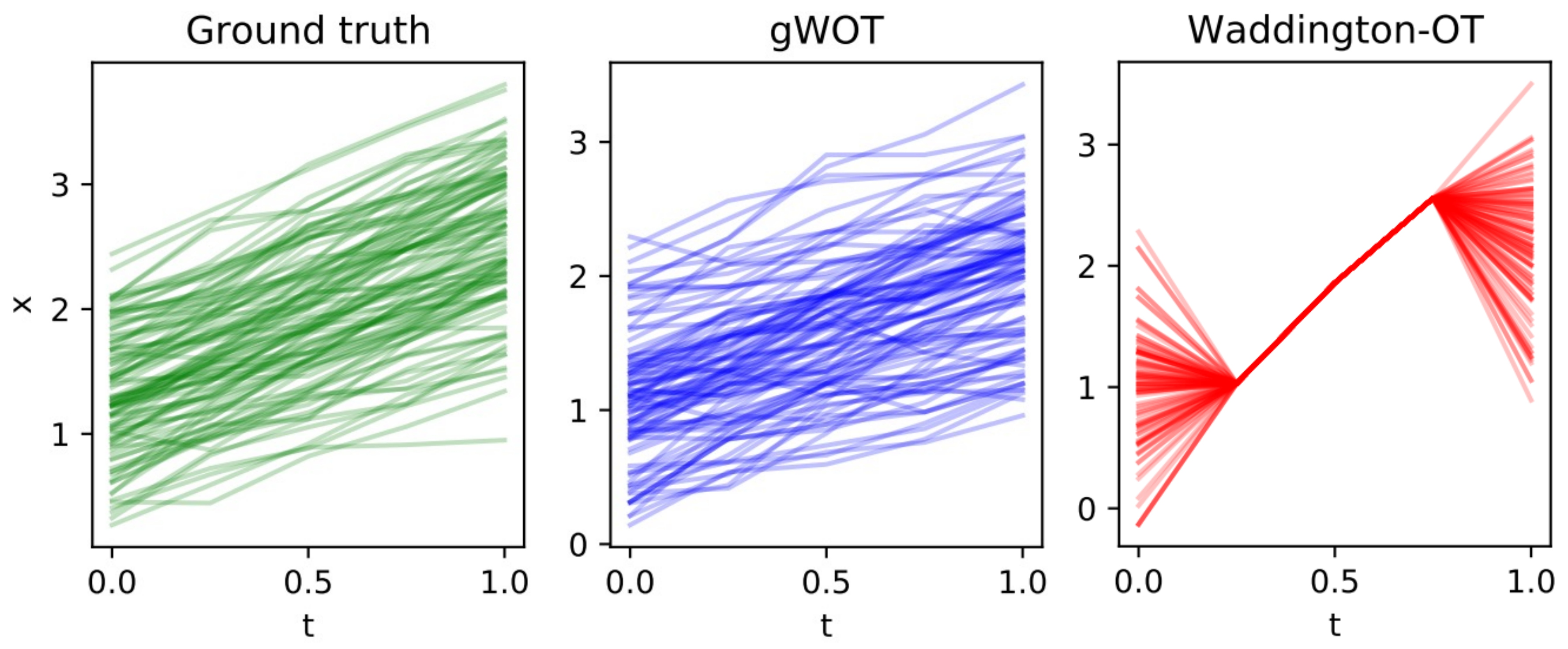
$$\mathcal{F}(R) := \mathrm{Fit}_{\lambda,\sigma}(R_{t_1},\ldots,R_{t_T}) + \tau H(R|W^\tau)$$

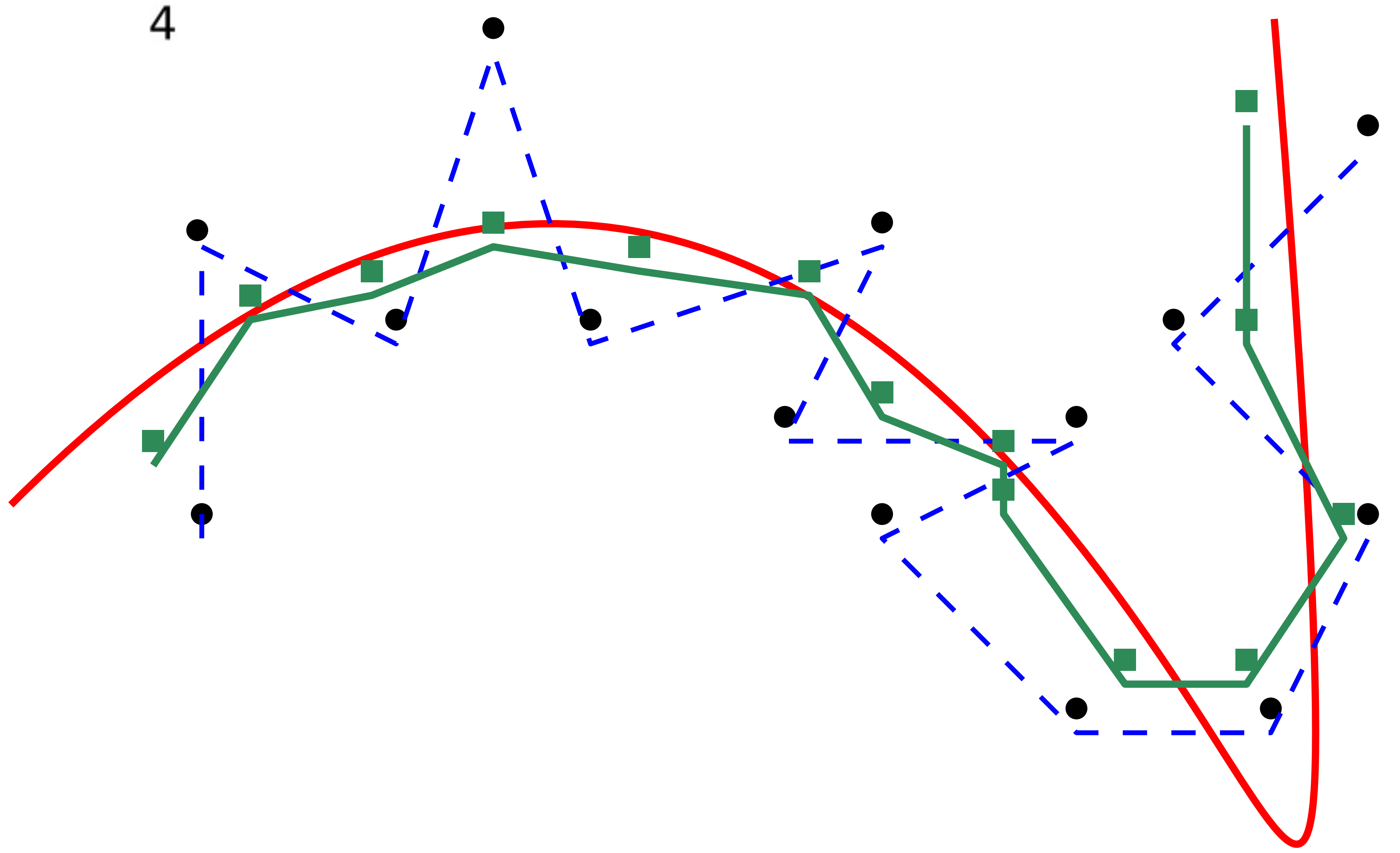
$$R^* := \operatorname*{argmin}_{R\in\mathcal{P}(\Omega)} \mathcal{F}(R).$$

- Well-posed convex optimization problem over $\mathcal{P}(\Omega)$
- Discretize then optimize approach is tractable...

  $\rightsquigarrow$ reduction from $\mathcal{P}(\Omega)$ to $\mathcal{P}(\mathcal{X})^T$ thanks to the Markovian structure (Benamou et al. 2018), (Lavenant et al. 2021)





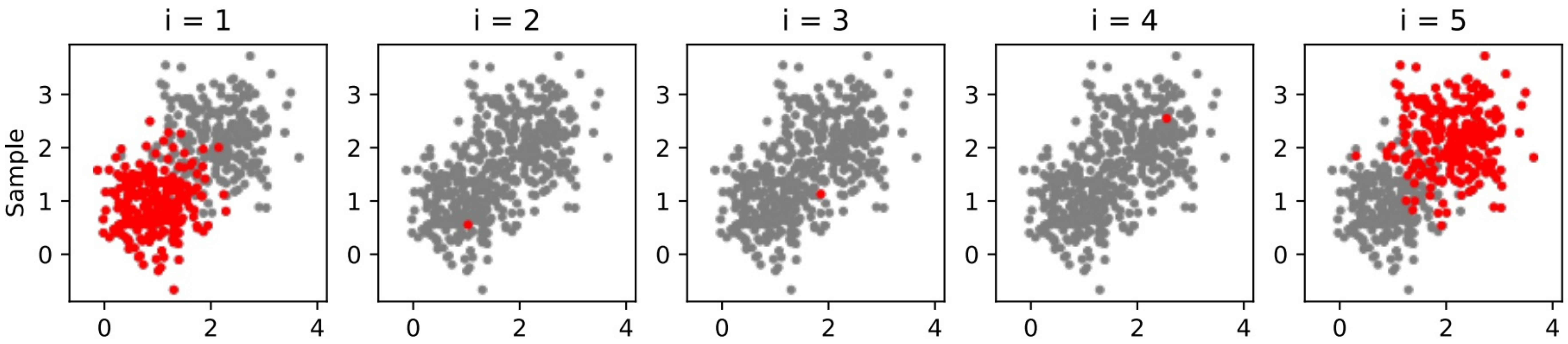


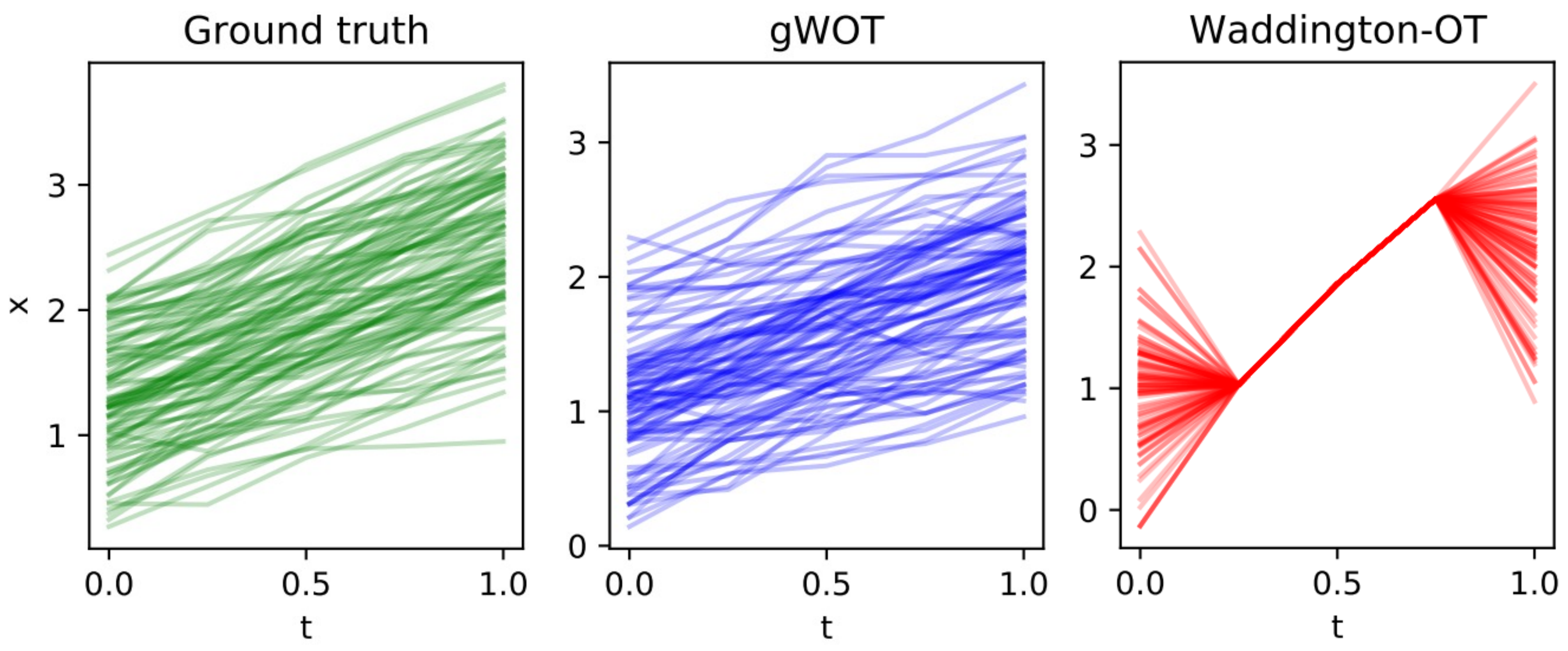
How can we solve this optimization problem?

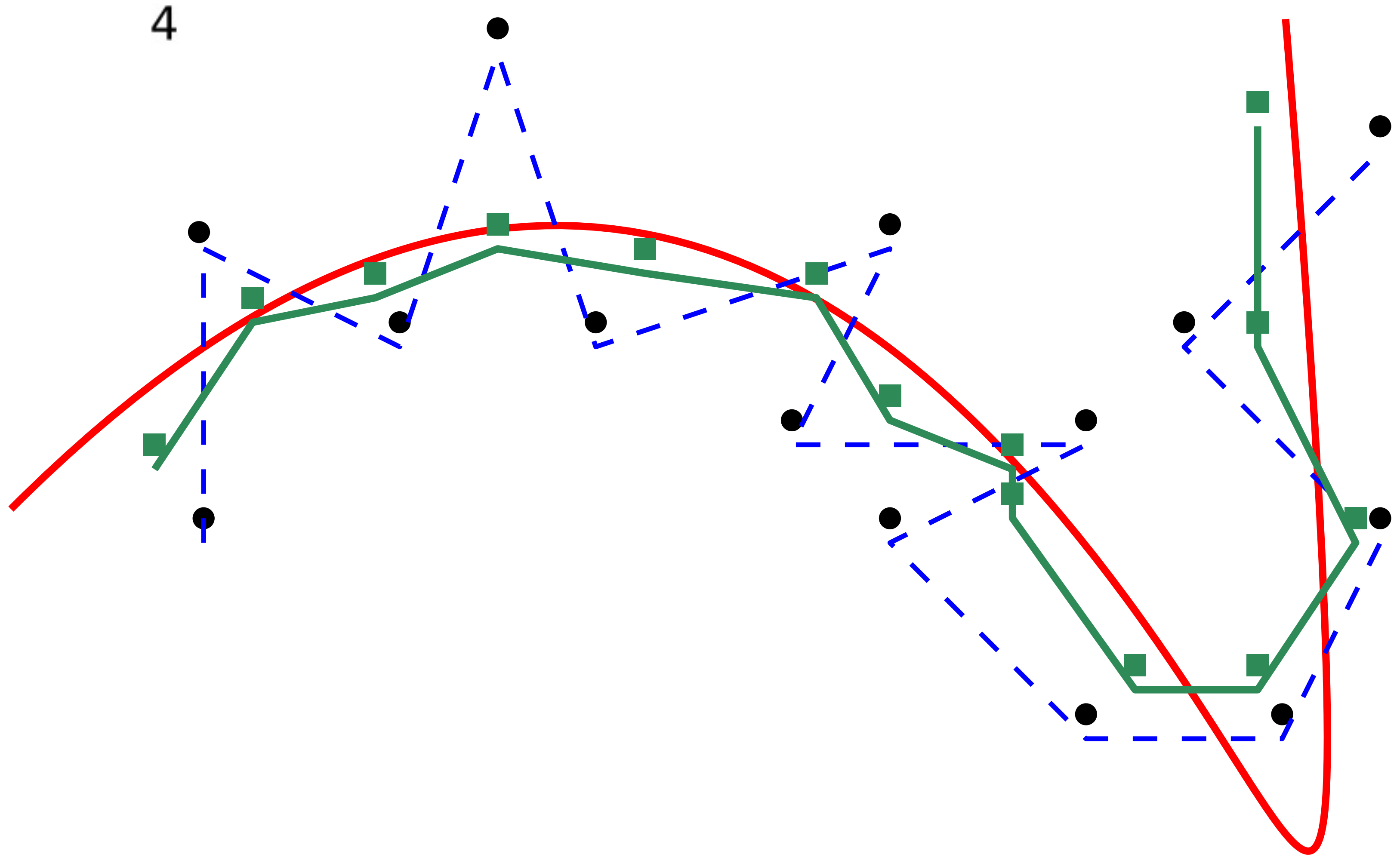
$$\mathcal{F}(R) := \mathrm{Fit}_{\lambda,\sigma}(R_{t_1},\dots,R_{t_T}) + \tau H(R|W^\tau)$$

$$R^* := \underset{R\in\mathcal{P}(\Omega)}{\operatorname{argmin}} \mathcal{F}(R)$$

- Well-posed convex optimization problem over $\mathcal{P}(\Omega)$
- Discretize then optimize approach is tractable...
  $\rightsquigarrow$ reduction from $\mathcal{P}(\Omega)$ to $\mathcal{P}(\mathcal{X})^T$ thanks to the Markovian structure (Benamou et al. 2018), (Lavenant et al. 2021)
- ... but not satisfying (curse of dimensionality)



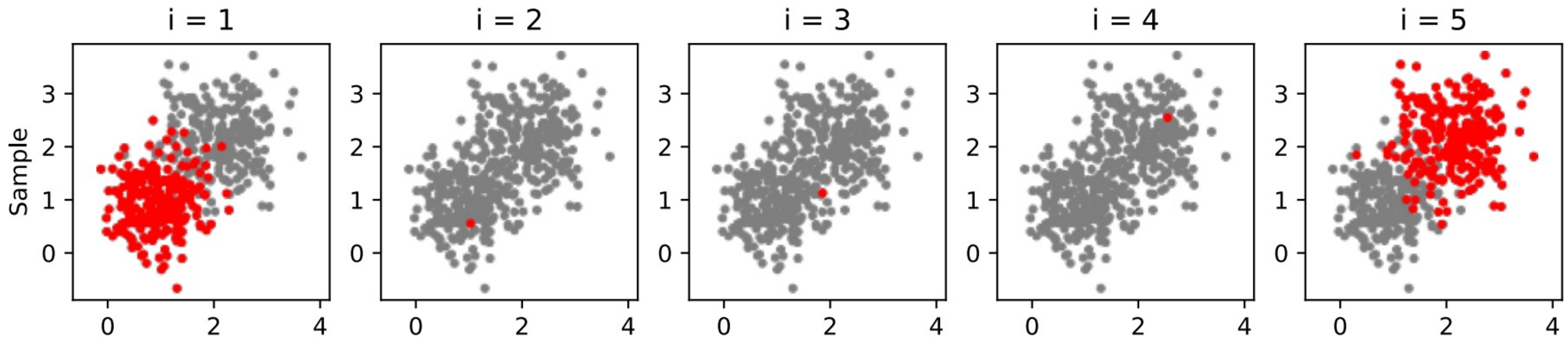


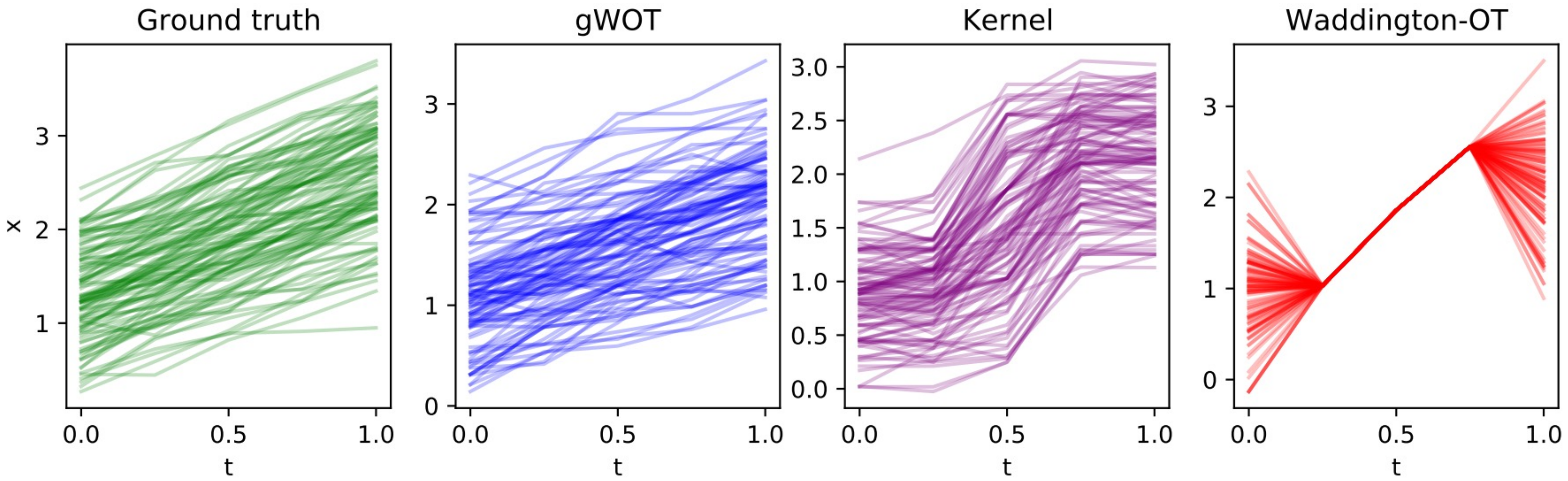


Aside: What about a simpler approach?

Locally-weighted averaging is a simple approach to function approximation:

$$R_{t_i} \propto \sum_{j=1}^{T} k_h(t_i - t_j)\hat{\mathbb{P}}_{t_j}$$





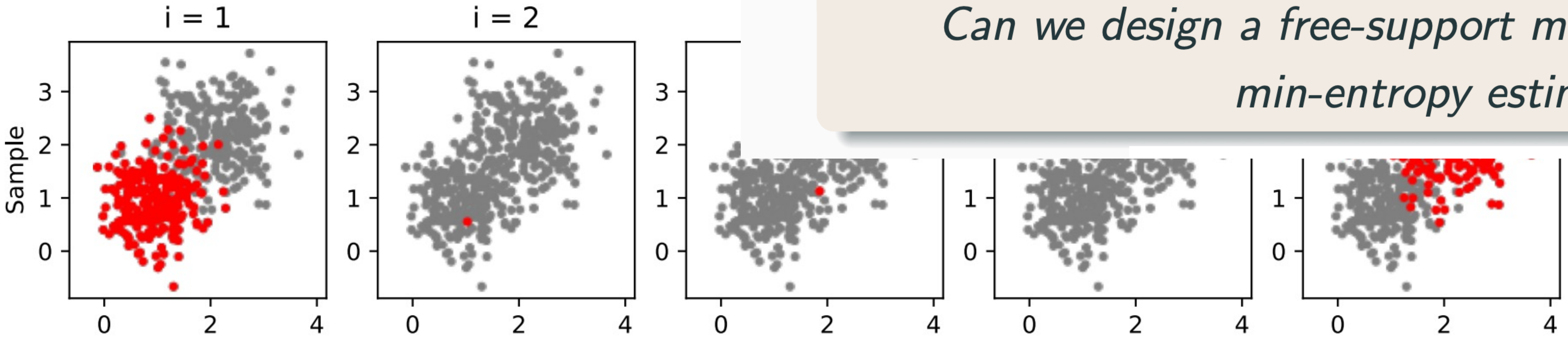
How can we solve this optimization problem?

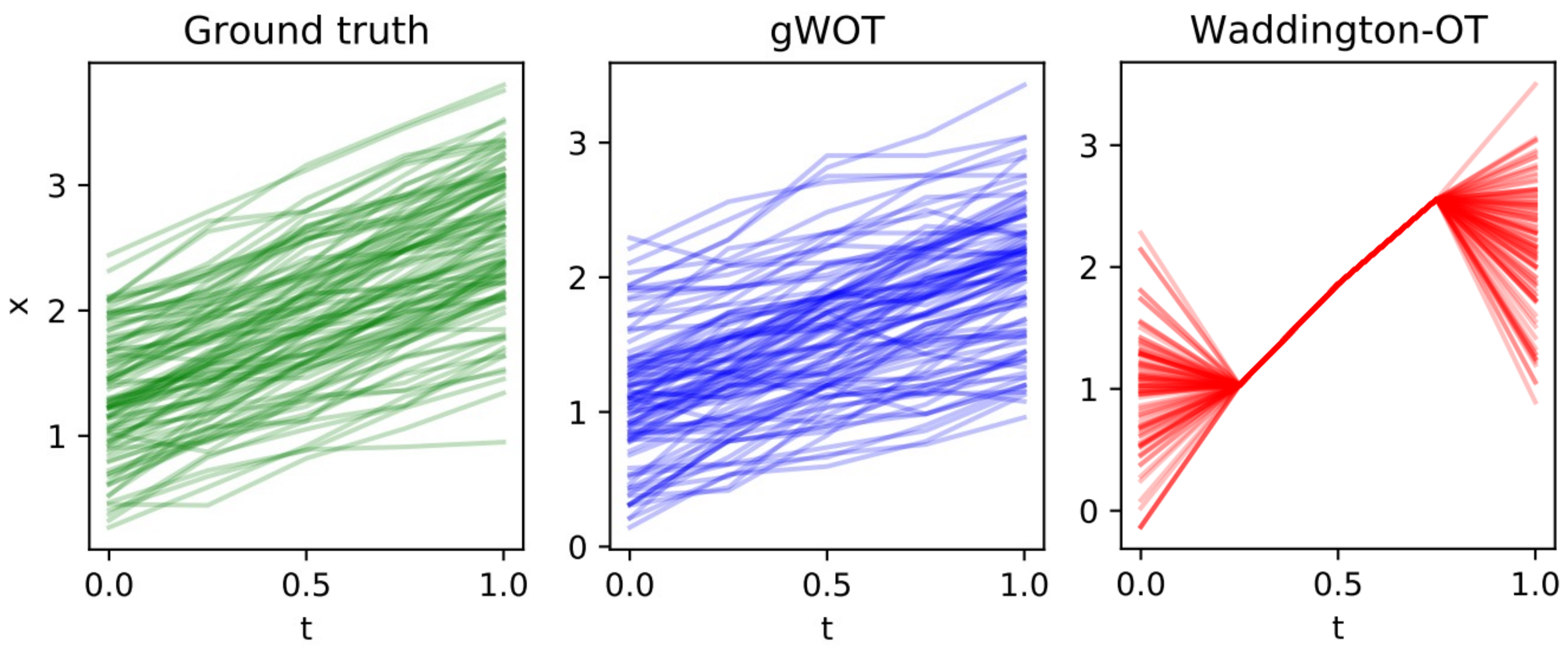
$$\mathcal{F}(R) := \mathrm{Fit}_{\lambda,\sigma}(R_{t_1}, \dots, R_{t_T}) + \tau H(R|W^\tau)$$

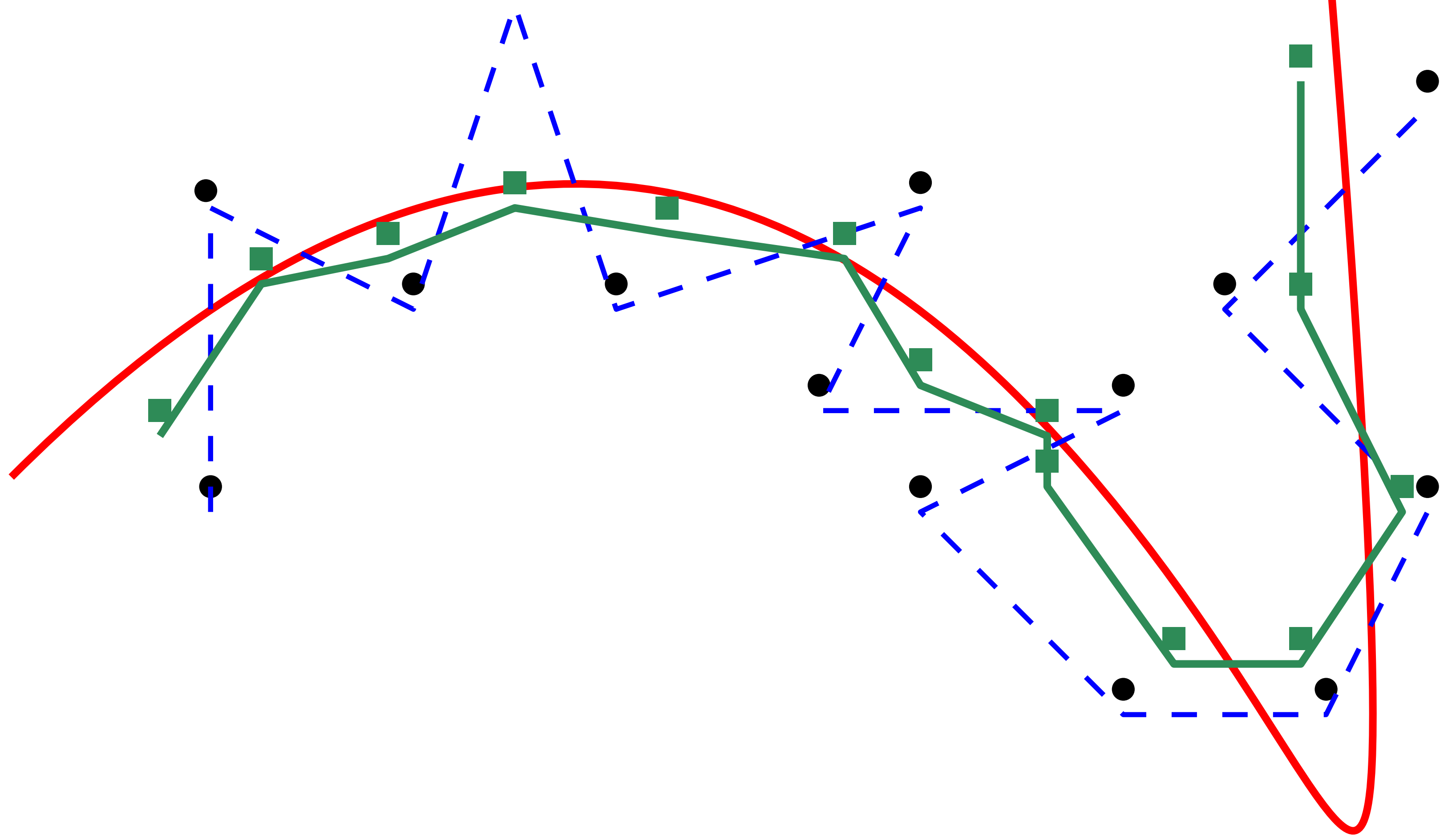
$$R^* := \operatorname*{argmin}_{R \in \mathcal{P}(\Omega)} \mathcal{F}(R).$$

- Well-posed convex optimization problem over $\mathcal{P}(\Omega)$
- Discretize then optimize approach is tractable...
  $\rightsquigarrow$ reduction from $\mathcal{P}(\Omega)$ to $\mathcal{P}(\mathcal{X})^T$ thanks to the Markovian structure (Benamou et al. 2018), (Lavenant et al. 2021)
- ... but not satisfying (curse of dimensionality)

> *Can we design a free-support method that computes the min-entropy estimator R^*?*







A definition

Let $\Pi(\mu, \nu)$ be the set of transport plans between $\mu, \nu \in \mathcal{P}(\mathcal{X})$.
$\mathcal{X} \subset \mathbb{R}^d$ compact.

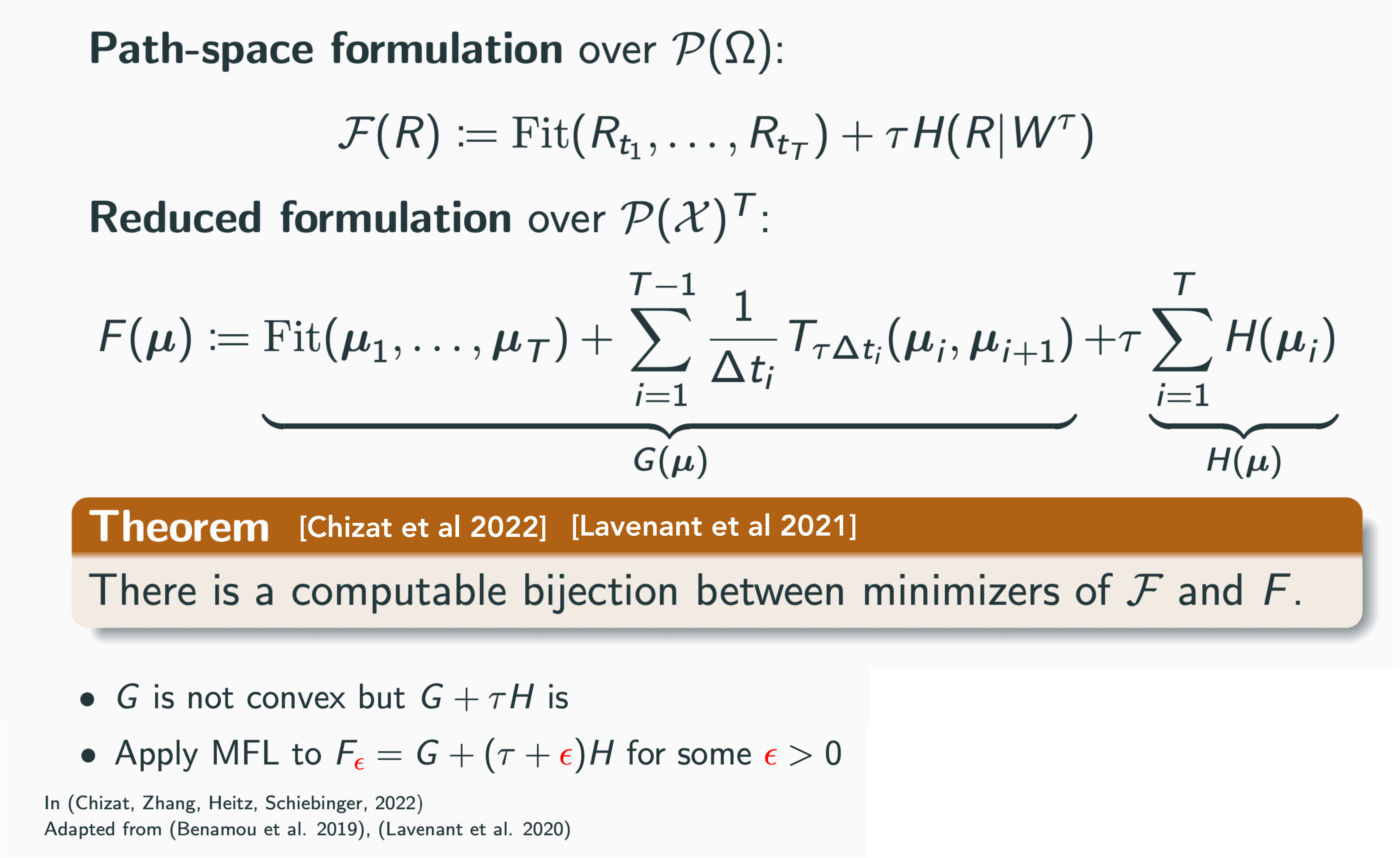
Entropic Optimal Transport

$$T_\tau(\mu, \nu) := \min_{\gamma \in \Pi(\mu,\nu)} \int c_\tau(x, y) \, \mathrm{d}\gamma(x, y) + \tau H(\gamma | \mu \otimes \nu)$$

where $c_\tau(x, y) \underset{\tau \to 0}{\longrightarrow} \frac{1}{2}\|y - x\|^2$ is the log-heat-kernel on $\mathcal{X}$.

- differentiable in (μ, ν),
- first variation given by the “stable” dual potentials (φ, ψ)
- T_τ is convex in μ and ν separately
- $T_\tau(\mu, \nu) + \tau H(\mu)$ is jointly convex

A "representer theorem"



Path-space formulation over $\mathcal{P}(\Omega)$:

$$\mathcal{F}(R) := \mathrm{Fit}(R_{t_1}, \dots, R_{t_T}) + \tau H(R|W^\tau)$$

Reduced formulation over $\mathcal{P}(\mathcal{X})^T$:

$$F(\boldsymbol{\mu}) := \underbrace{\mathrm{Fit}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T) + \sum_{i=1}^{T-1} \frac{1}{\Delta t_i} T_{\tau \Delta t_i}(\boldsymbol{\mu}_i, \boldsymbol{\mu}_{i+1})}_{G(\boldsymbol{\mu})} + \tau \underbrace{\sum_{i=1}^{T} H(\boldsymbol{\mu}_i)}_{H(\boldsymbol{\mu})}$$

Theorem [Chizat et al 2022] [Lavenant et al 2021]

There is a computable bijection between minimizers of $\mathcal{F}$ and F.

- G is not convex but $G + \tau H$ is
- Apply MFL to $F_\epsilon = G + (\tau + \epsilon) H$ for some $\epsilon > 0$

In (Chizat, Zhang, Heitz, Schiebinger, 2022)
Adapted from (Benamou et al. 2019), (Lavenant et al. 2020)

Langevin Dynamics

- **Goal**: given $V \in \mathcal{C}^2(\mathcal{X})$, sample from $\propto e^{-V/\tau}$, $\tau > 0$.
- **Noisy GD**:

$$X_{k+1} = -\eta \nabla V(X_k) + \sqrt{2\tau\eta} Z_k, \quad X_0 \sim \mu_0, \quad Z_k \overset{iid}{\sim} \mathcal{N}(0, I)$$

- As $\eta \to 0$, converges in law to a **Langevin Dynamics** ($t = k\eta$):

$$dX_t = -\nabla V(X_t)\, \mathrm{d}t + \sqrt{2\tau} dB_t, \quad X_0 \sim \mu_0, \quad B_t \text{ Brownian process}$$

- Moreover $\mu_t = \mathrm{Law}(X_t)$ follows the **Fokker-Planck equation**:

$$\partial_t \mu_t = \underbrace{\nabla \cdot (\mu_t \nabla V)}_{\text{drift}} + \underbrace{\tau \Delta \mu_t}_{\text{diffusion}}, \quad \mu_0 \text{ given}$$

NB: do not confuse optimization time vs biological time

Langevin Dynamics

Interpretation: Wasserstein gradient flow of

$$F_\tau(\mu) := \int V \,\mathrm{d}\mu + \tau H(\mu) = H(\mu|\mu^*_\tau)$$

where $\mu^*_\tau \propto e^{-V/\tau} \in \mathcal{P}(\mathcal{X})$, $H(\mu) = \int \log(\mathrm{d}\mu/\mathrm{d}x)\,\mathrm{d}\mu$ is the neg-entropy and $H(\mu|\nu) = \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\,\mathrm{d}\mu$ is the relative entropy

Theorem [Holley, Kusuoka, Stroock, 1989] and many more

- F_τ admits a unique minimizer $\mu^*_\tau \propto e^{-V/\tau}$.
- Assume that μ^*_τ satisfies a ρ_τ- *log-Sobolev inequality*, then

$$F_\tau(\mu_t) - F_\tau(\mu^*_\tau) \le e^{-2\tau\rho_\tau t}(F_\tau(\mu_0) - F_\tau(\mu^*_\tau)).$$

Langevin Dynamics

Interpretation: Wasserstein gradient flow of

$$F_\tau(\mu) := \int V \,\mathrm{d}\mu + \tau H(\mu) = H(\mu|\mu_\tau^*)$$

where $\mu_\tau^* \propto e^{-V/\tau} \in \mathcal{P}(\mathcal{X})$, $H(\mu) = \int \log(\mathrm{d}\mu/\mathrm{d}x)\,\mathrm{d}\mu$ is the neg-entropy and $H(\mu|\nu) = \int \log(\mathrm{d}\mu/\mathrm{d}\nu)\,\mathrm{d}\mu$ is the relative entropy

Theorem [Holley, Kusuoka, Stroock, 1989] and many more

- F_τ admits a unique minimizer $\mu_\tau^* \propto e^{-V/\tau}$.
- Assume that μ_τ^* satisfies a ρ_τ- *log-Sobolev inequality*, then

$$F_\tau(\mu_t) - F_\tau(\mu_\tau^*) \leq e^{-2\tau\rho_\tau t}(F_\tau(\mu_0) - F_\tau(\mu_\tau^*)).$$

Recent work on "Mean Field Langevin Dynamics" generalizes these ideas to general convex F **[Chizat 2022].**

Back to our trajectory inference problem

$$F(\mu) := \underbrace{\mathrm{Fit}(\mu_1, \dots, \mu_T) + \sum_{i=1}^{T-1} \frac{1}{\Delta t_i} T_{\tau \Delta t_i}(\mu_i, \mu_{i+1})}_{G(\mu)} + \tau \underbrace{\sum_{i=1}^{T} H(\mu_i)}_{H(\mu)}$$

- G is not convex but $G + \tau H$ is
- Apply MFL to $F_\epsilon = G + (\tau + \epsilon) H$ for some $\epsilon > 0$

In (Chizat, Zhang, Heitz, Schiebinger, 2022)
Adapted from (Benamou et al. 2019), (Lavenant et al. 2020)

Theorem [Chizat et al 2022]

If $\mathcal{X}$ is compact, the Mean-Field Langevin dynamics for F_ϵ is well-posed and converges exponentially to global minimizers.

Back to our trajectory inference problem

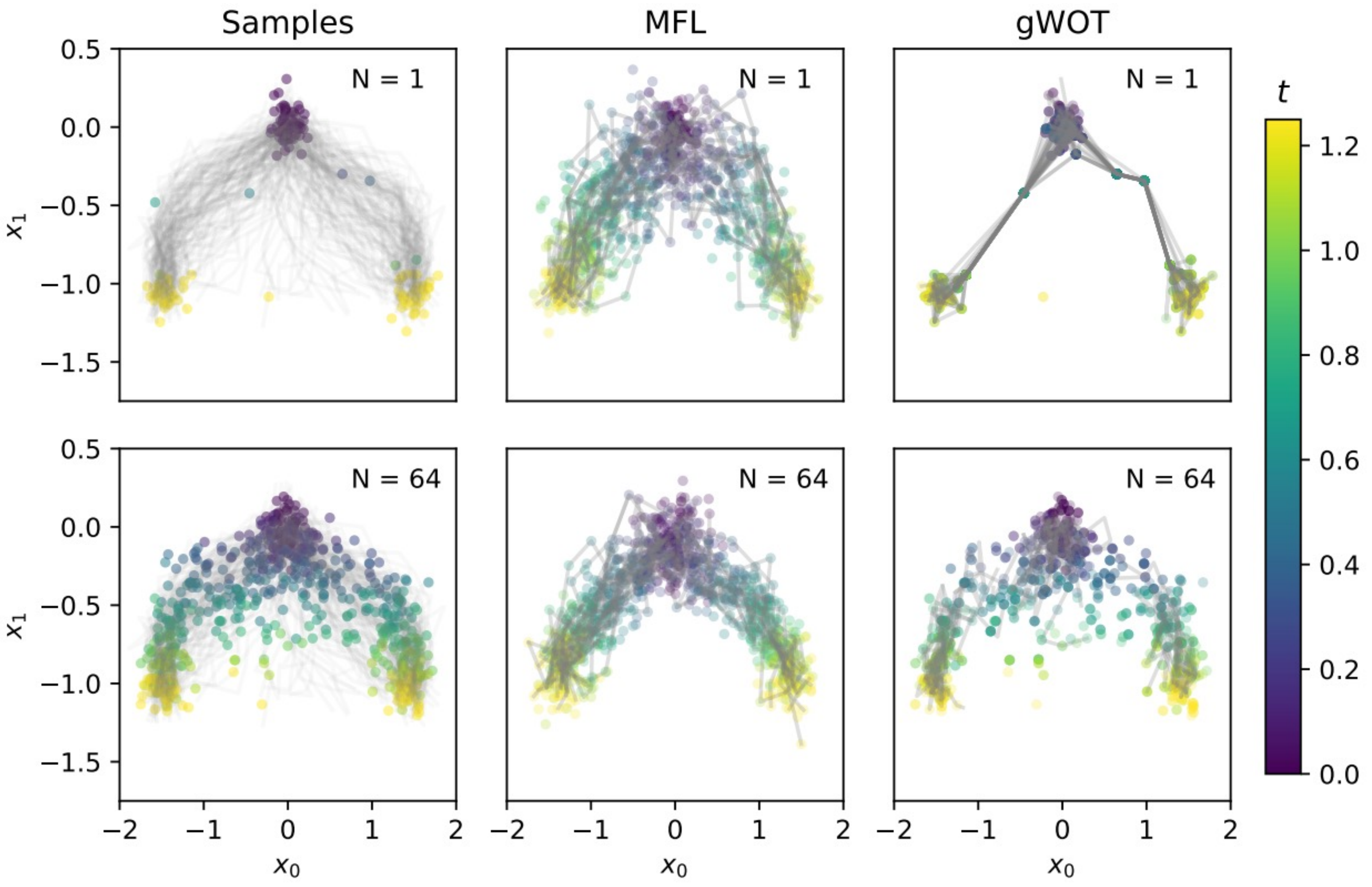
$$F(\boldsymbol{\mu}) := \underbrace{\mathrm{Fit}(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_T) + \sum_{i=1}^{T-1}\frac{1}{\Delta t_i} T_{\tau\Delta t_i}(\boldsymbol{\mu}_i,\boldsymbol{\mu}_{i+1})}_{G(\boldsymbol{\mu})} + \tau \underbrace{\sum_{i=1}^{T} H(\boldsymbol{\mu}_i)}_{H(\boldsymbol{\mu})}$$

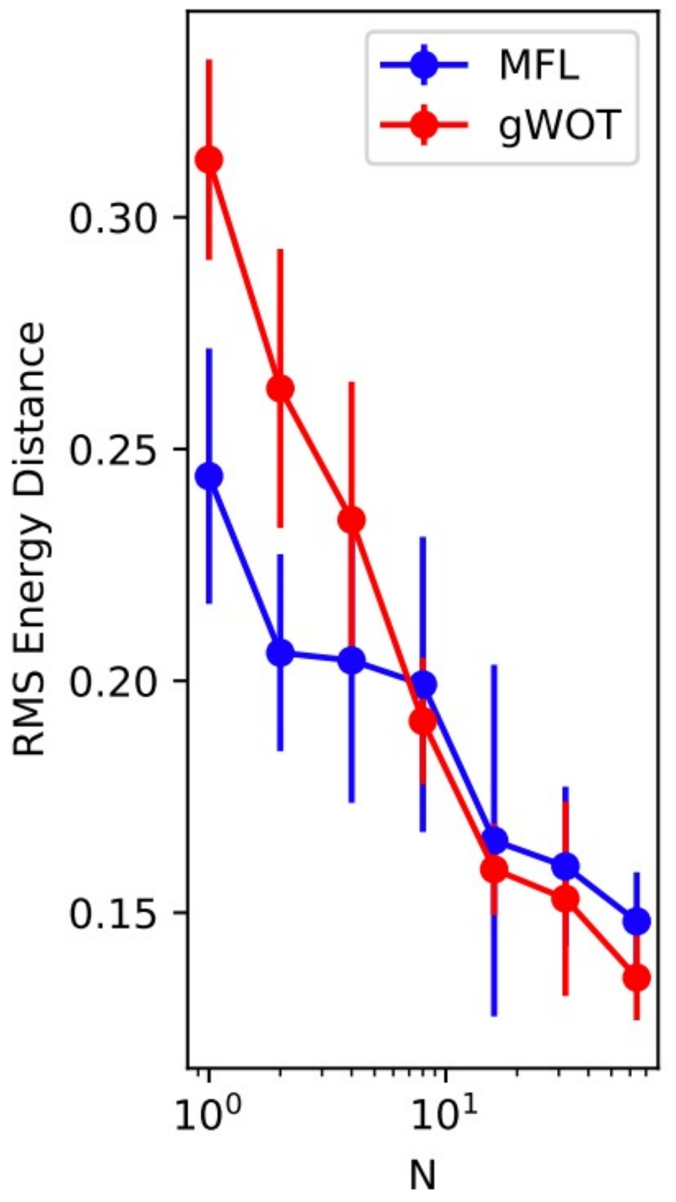
- G is not convex but $G + \tau H$ is
- Apply MFL to $F_\epsilon = G + (\tau + \epsilon)H$ for some $\epsilon > 0$

In (Chizat, Zhang, Heitz, Schiebinger, 2022)
Adapted from (Benamou et al. 2019), (Lavenant et al. 2020)

Theorem [Chizat et al 2022]

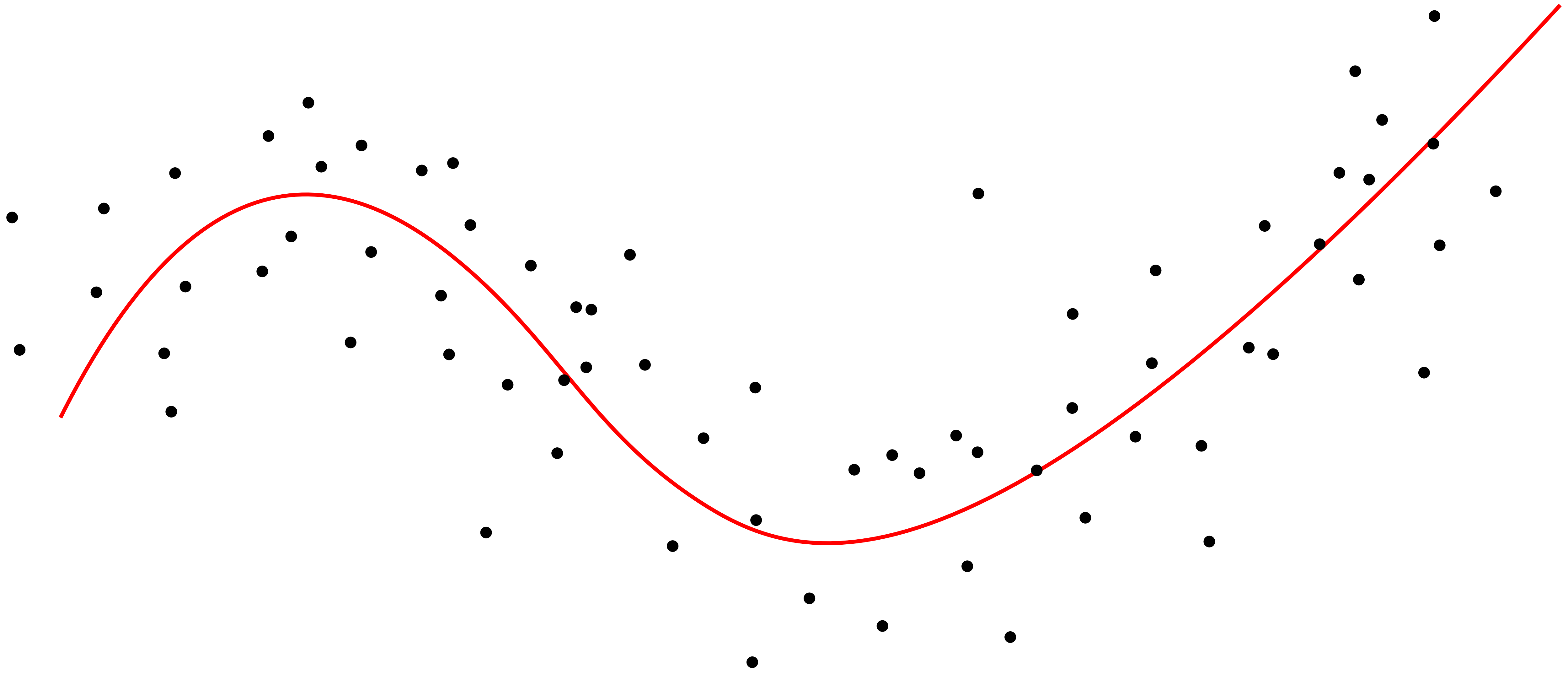
If $\mathcal{X}$ is compact, the Mean-Field
well-posed and converges expone





Perspective on experimental design

The number of time-points determines the "size of the dataset".



Nozomu Yachie

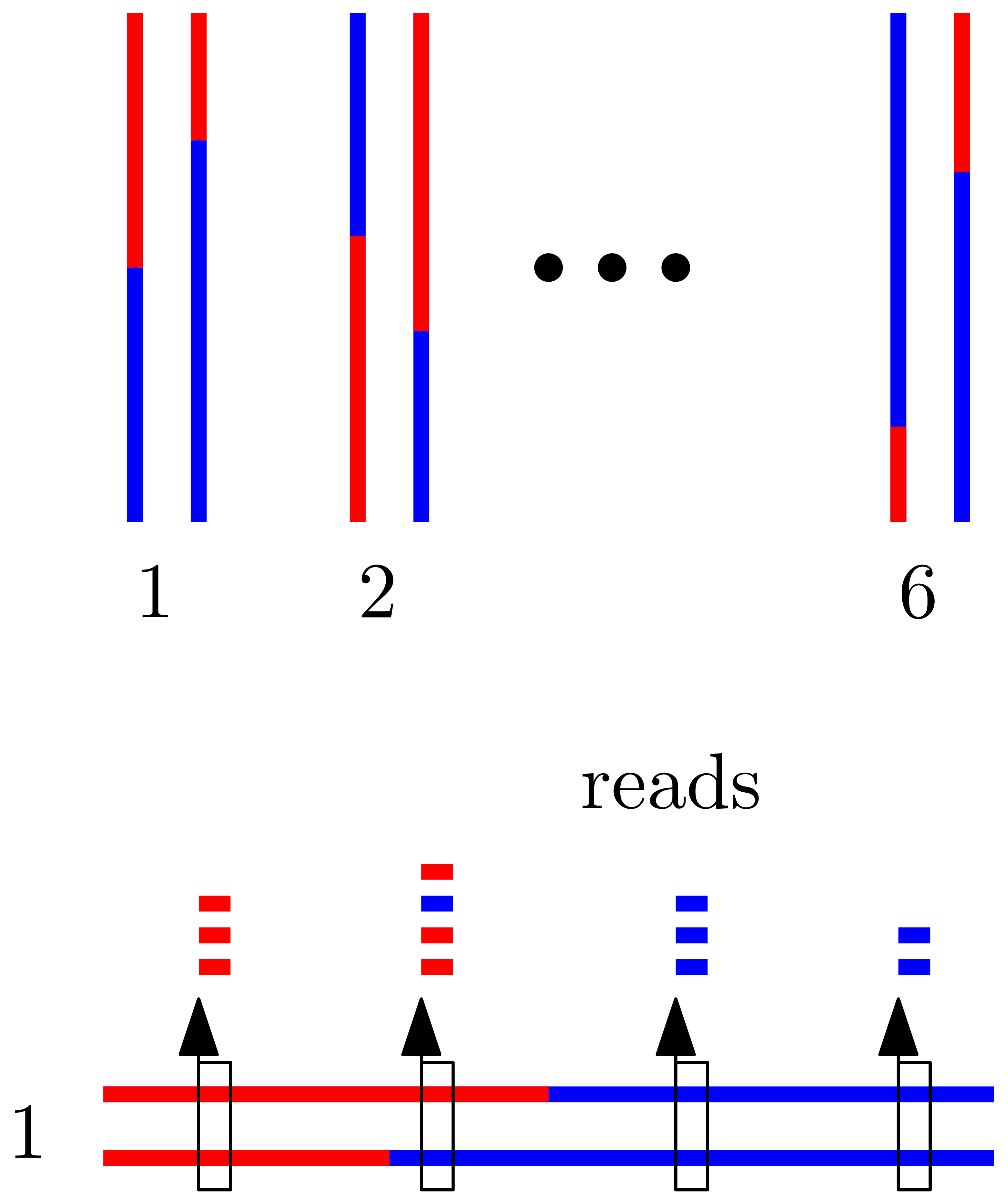
Kenji Sugioka

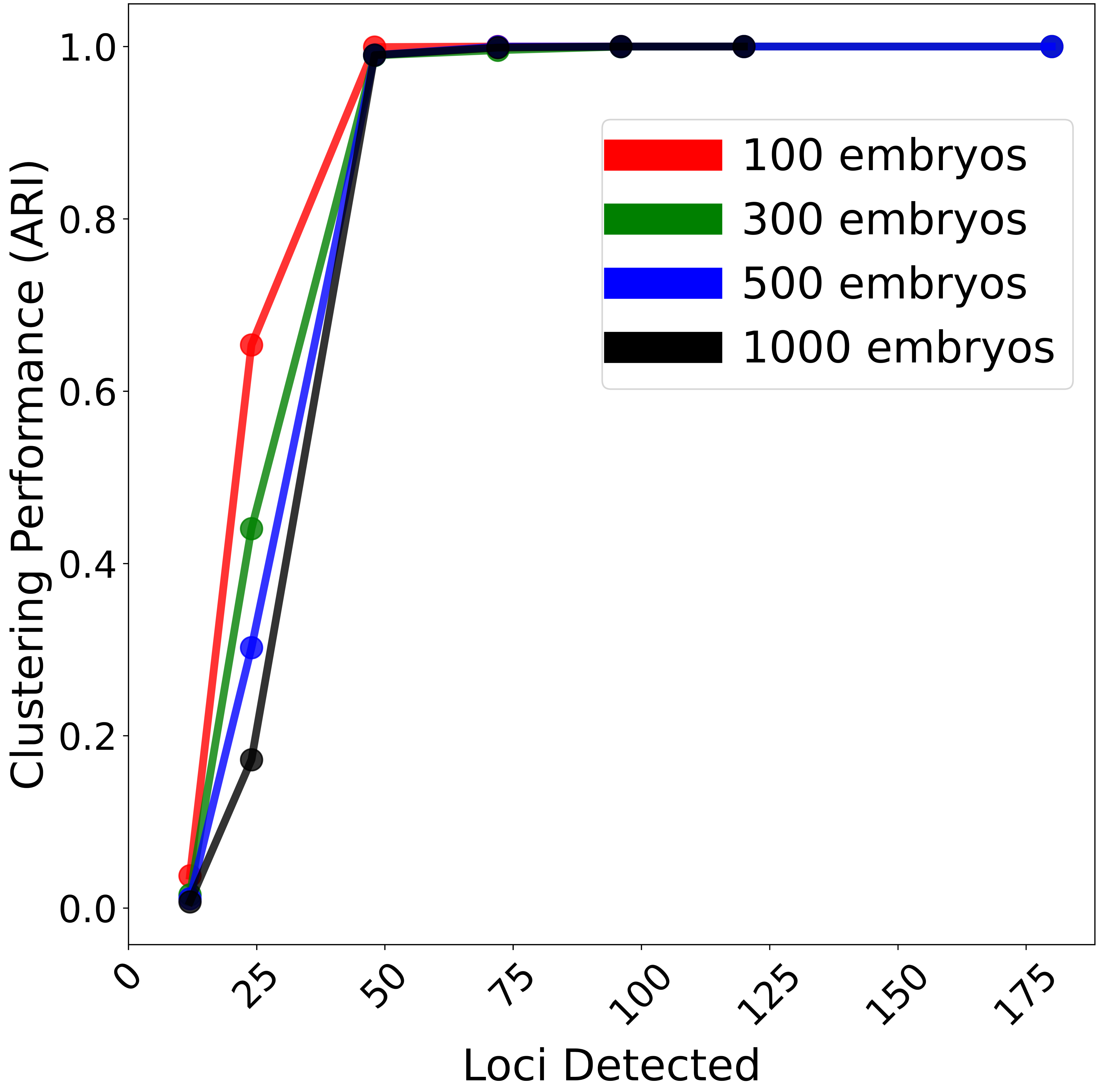
Each time-point is a "data point" along the curve.

The number of cells determines the noise level.

Collecting thousands of time-points with scRNA-seq + lineage tracing in C. elegans.

Embryo barcoding for high-density temporal profiling



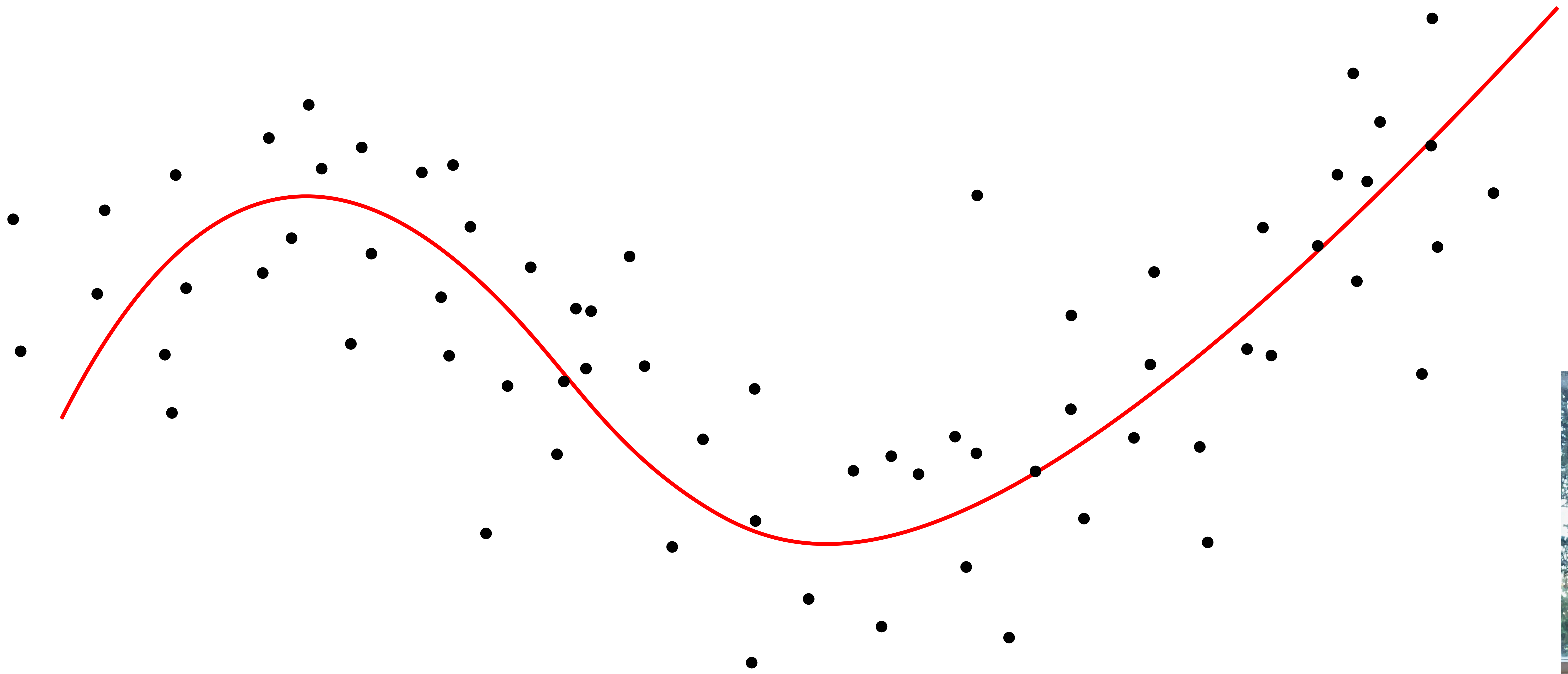


Kenji Sugioka
UBC Zoology

2021 CIHR project grant proposal by Schiebinger, Sugioka and Yachie

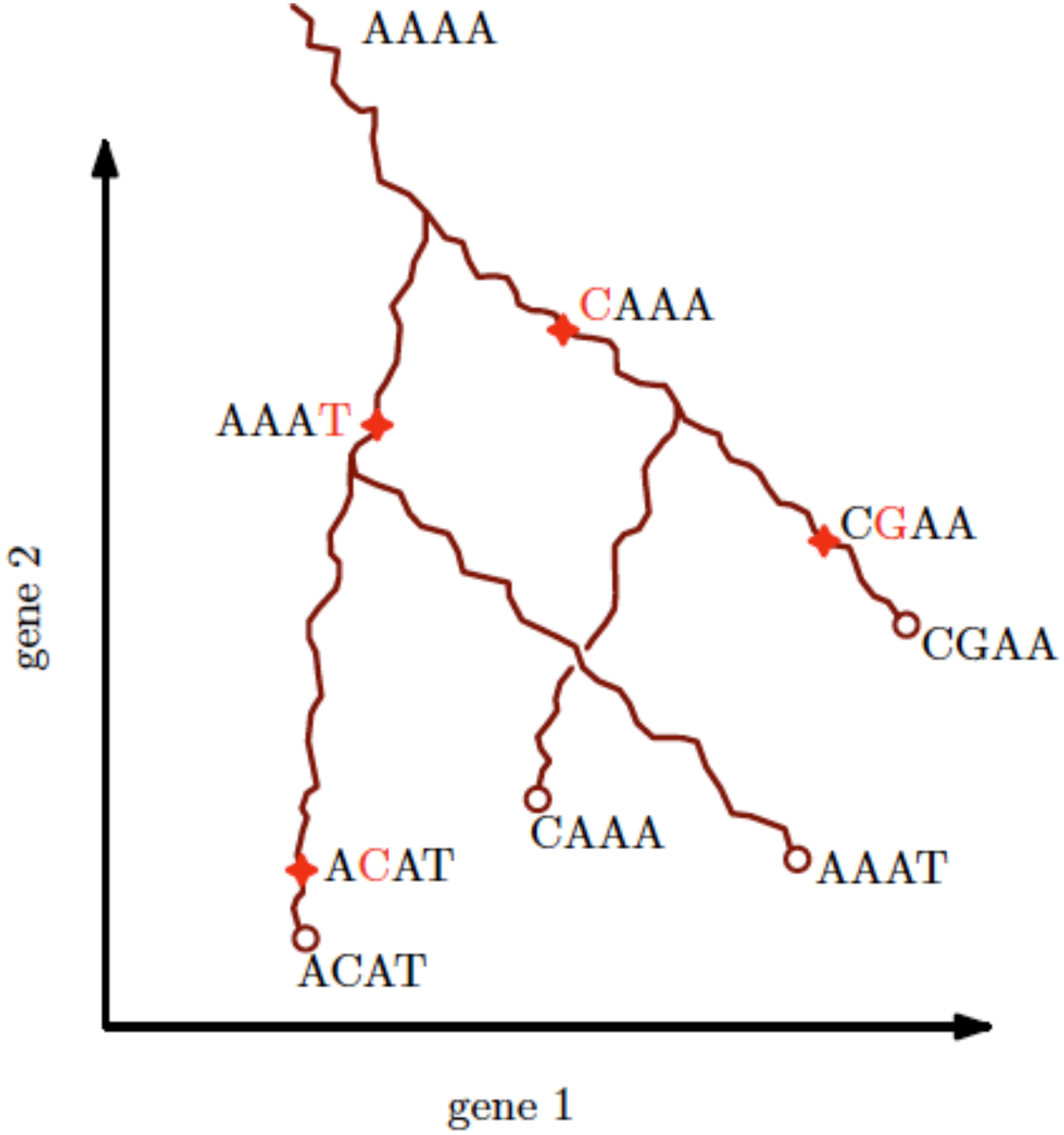
Nozomu Yachie, UBC SBME

Order embryos in developmental time



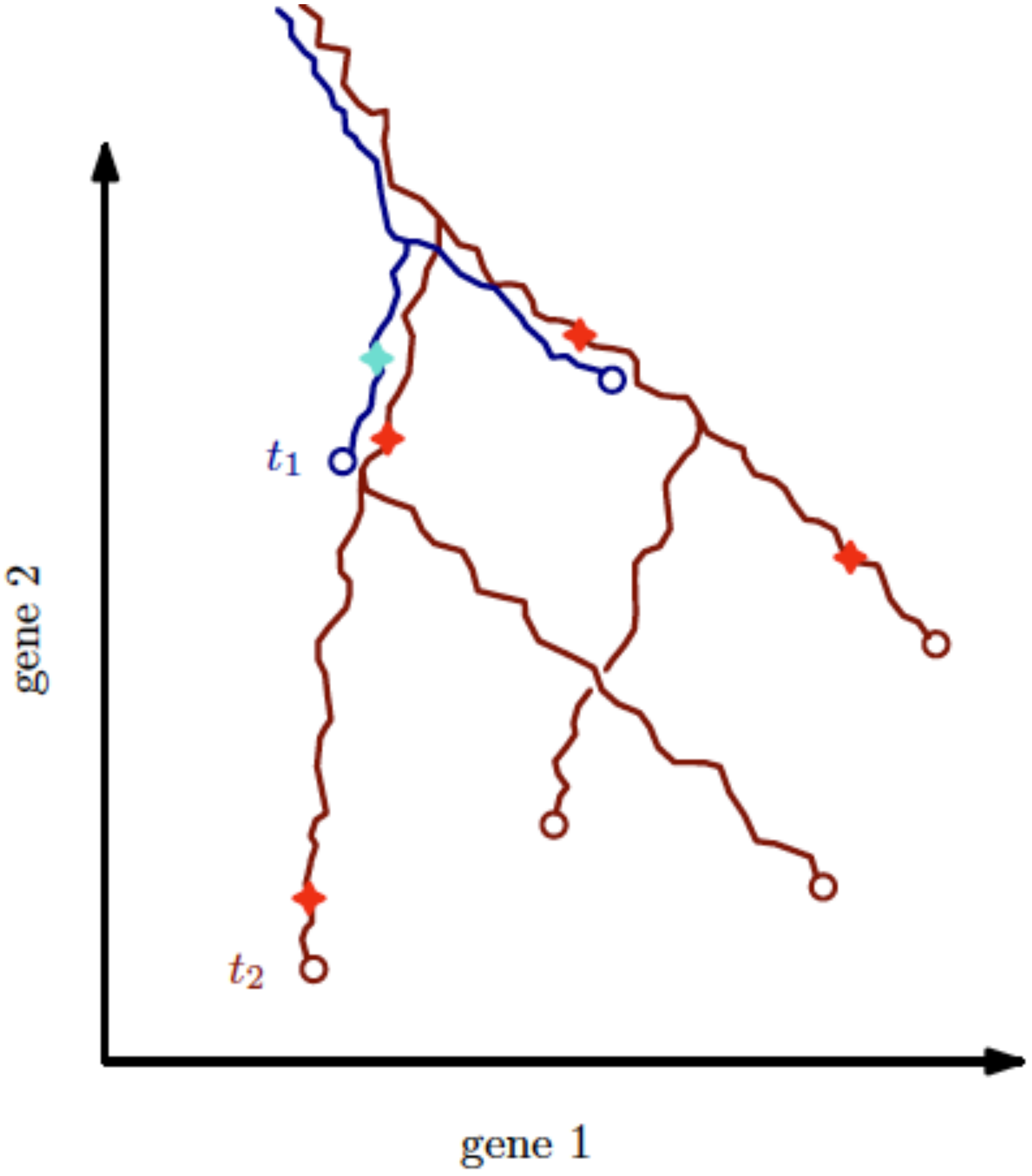
Roomina Zendehboodi

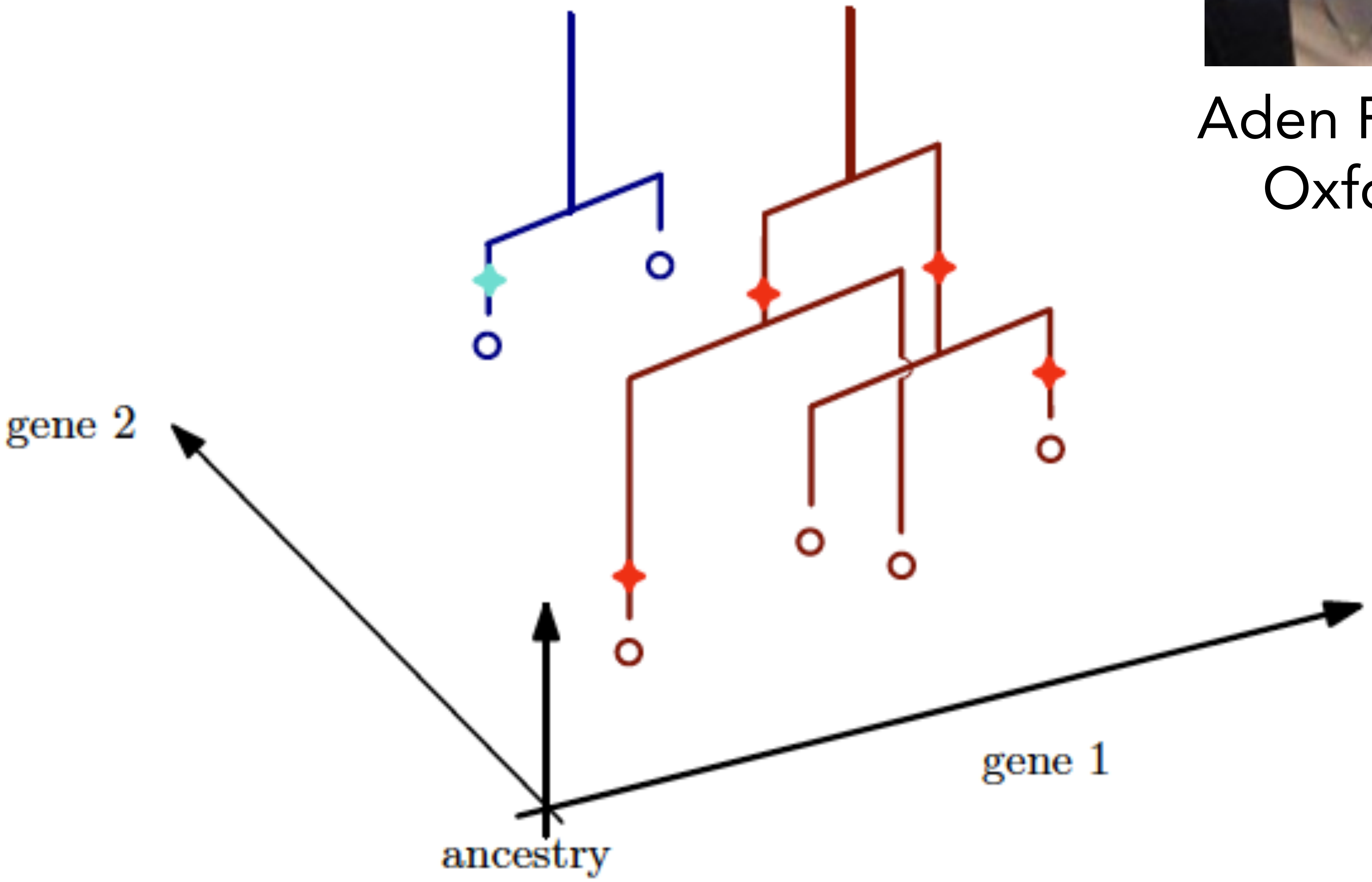
Lineage Tracing + Trajectory Inference



Aden Forrow
Oxford

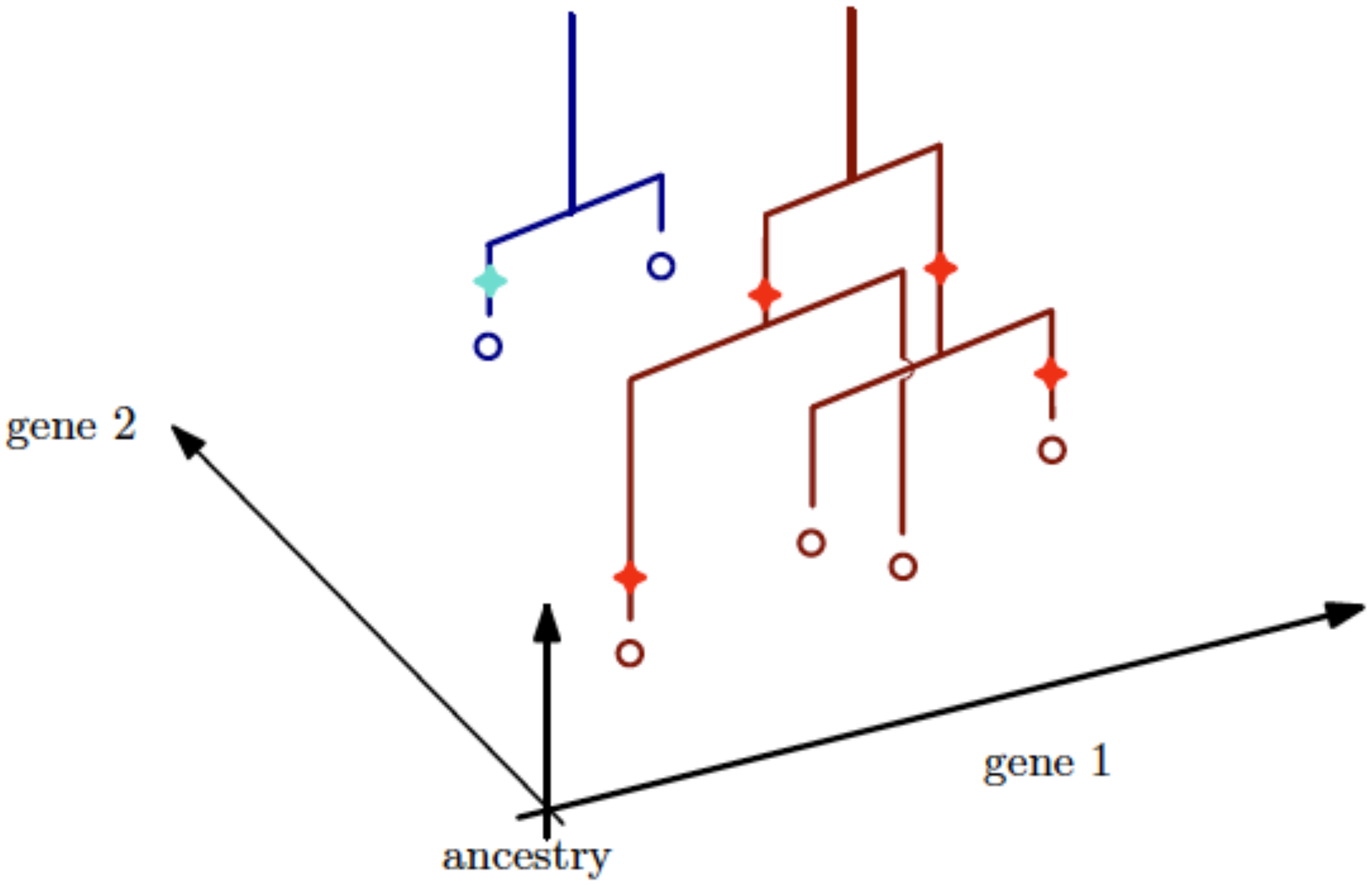
Lineage Tracing + Trajectory Inference

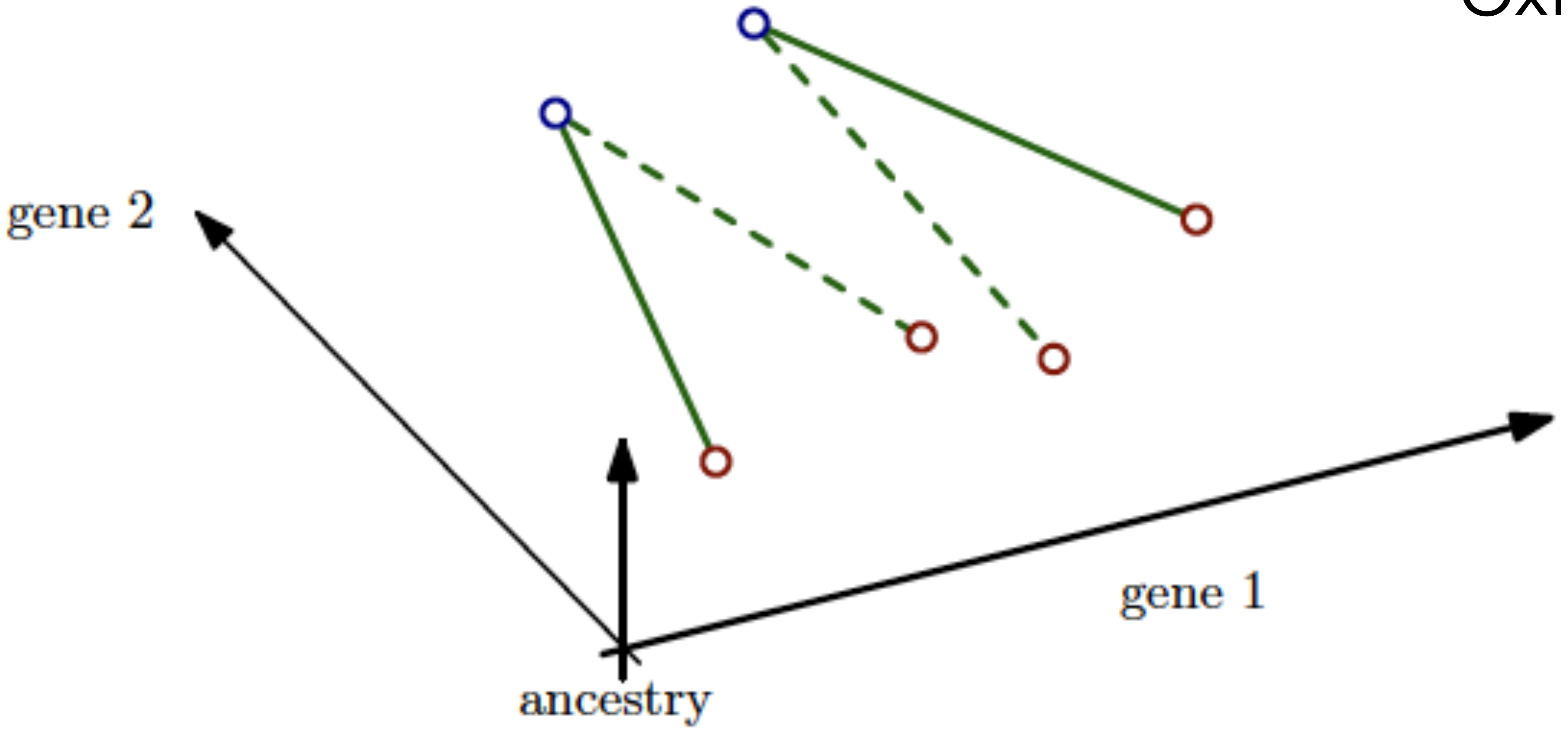




Aden Forrow
Oxford

Lineage Tracing + Trajectory Inference

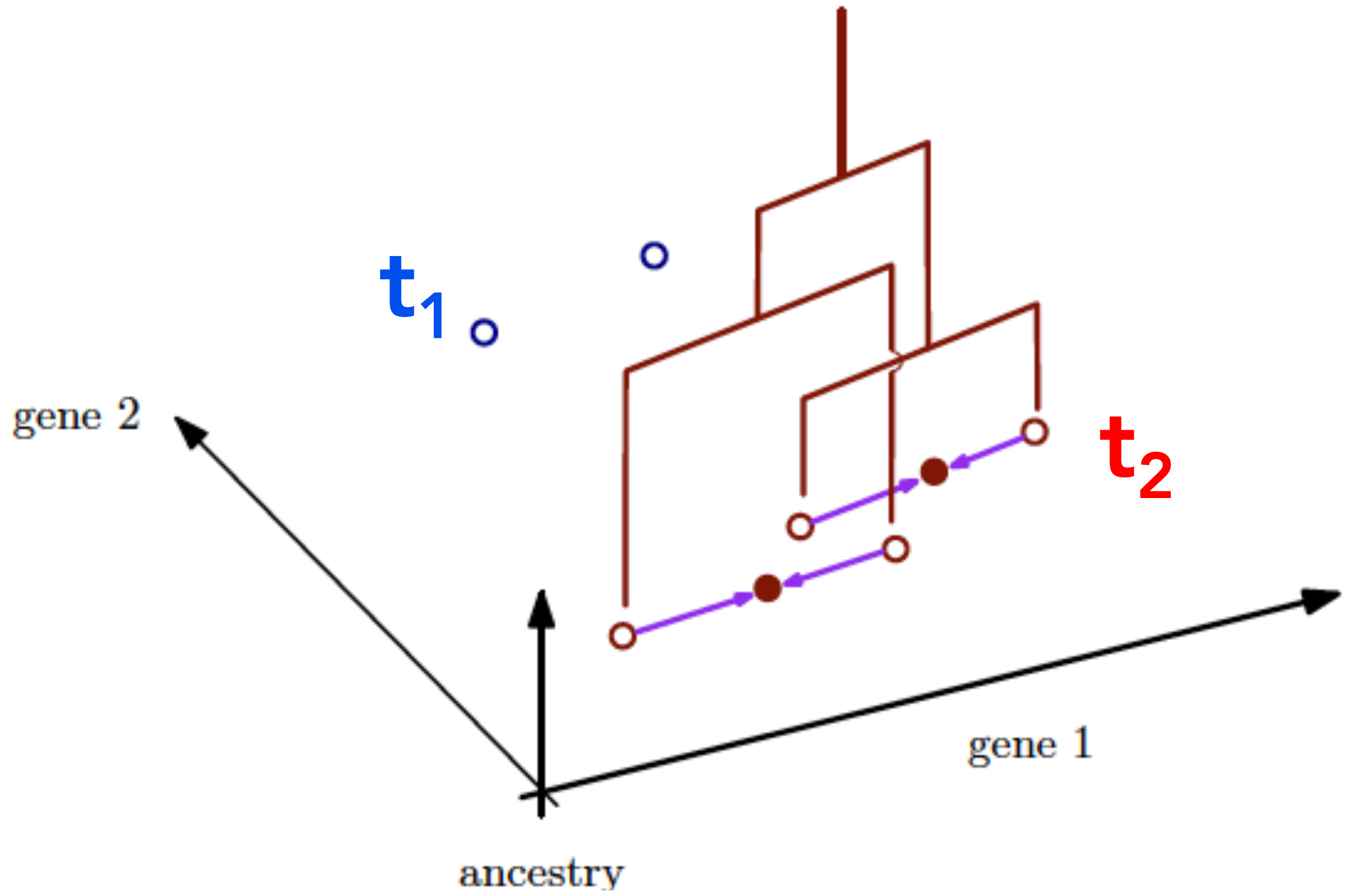




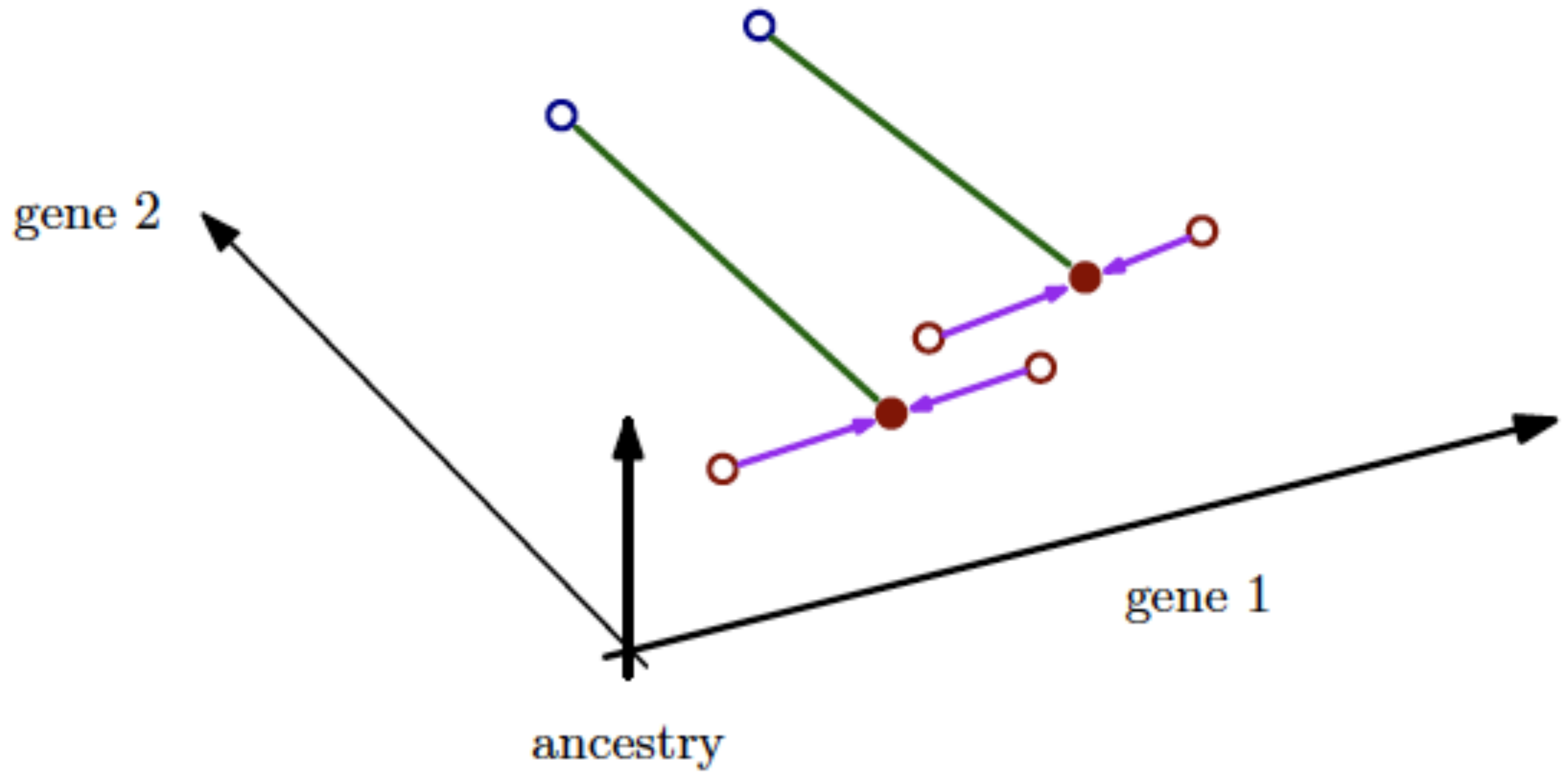
Aden Forrow
Oxford

Lineage-OT

Step 1: adjust t_2 cells based on lineage

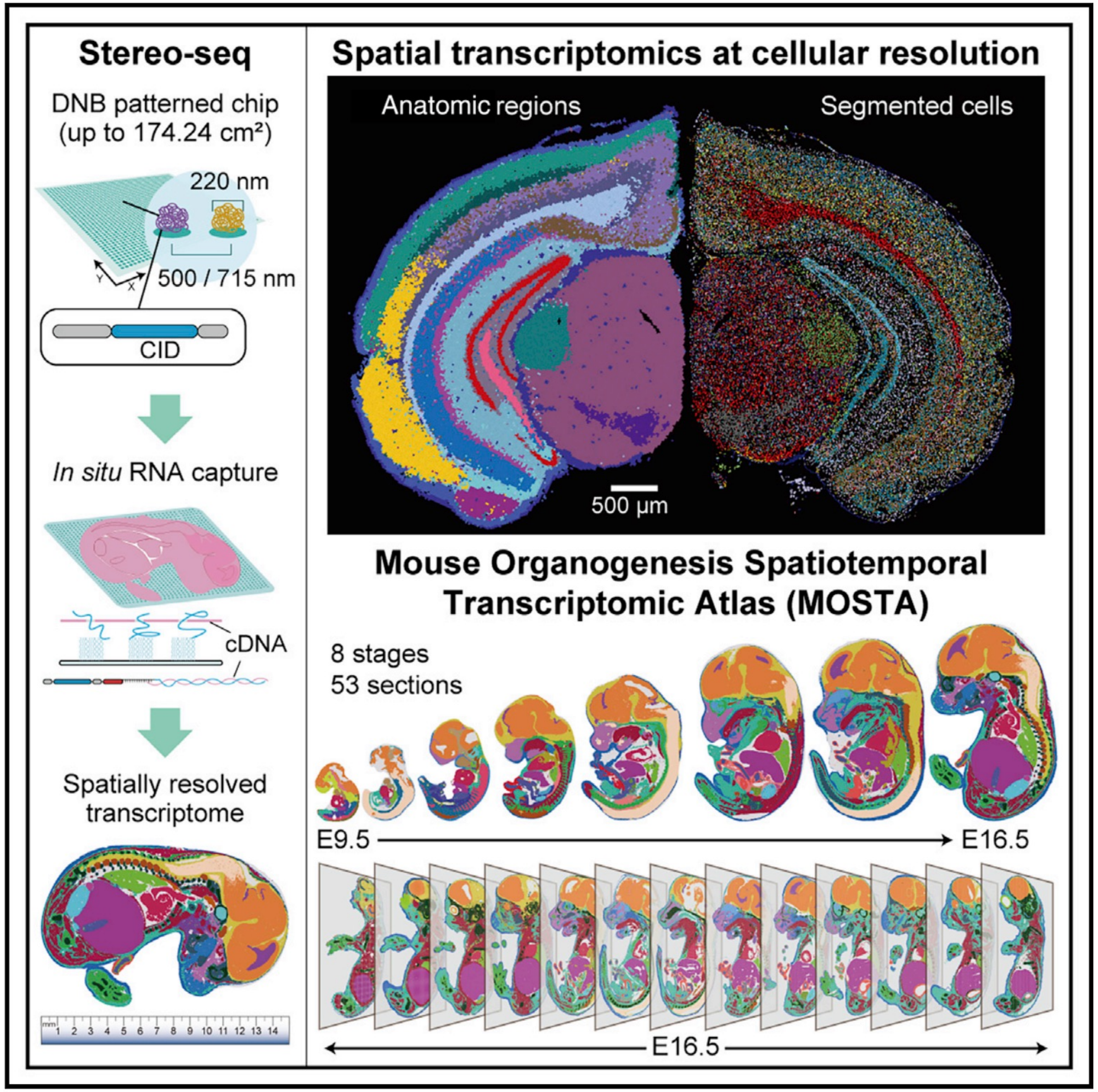


Step 2: connect ancestors to descendants with OT.

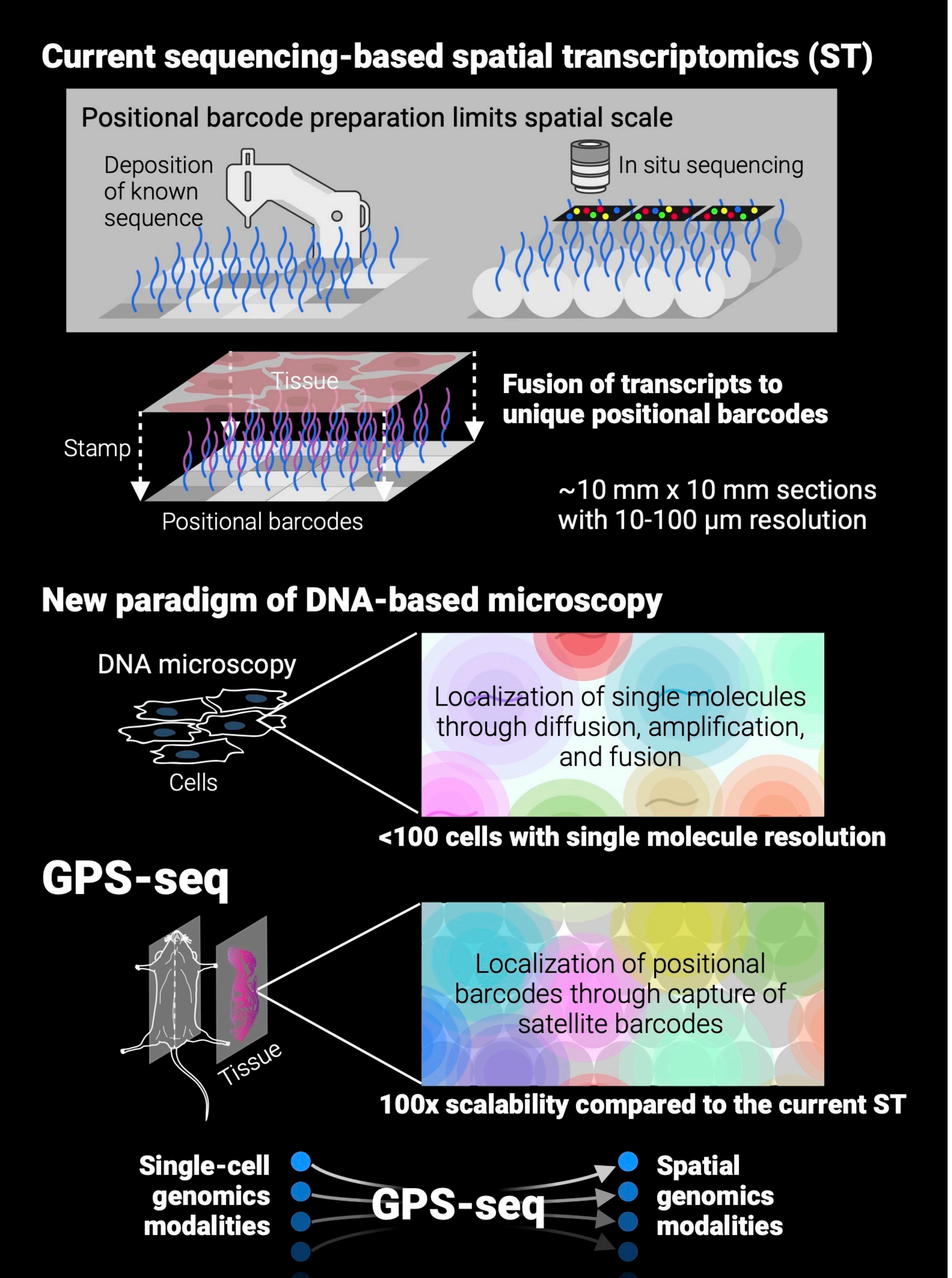


Forrow and Schiebinger (2021, *Nature Communications*)

Spatial transcriptomics



A DNA-based Global Positioning System



Current sequencing-based spatial transcriptomics (ST)

Positional barcode preparation limits spatial scale

Fusion of transcripts to unique positional barcodes

~10 mm x 10 mm sections with 10-100 µm resolution

New paradigm of DNA-based microscopy

Localization of single molecules through diffusion, amplification, and fusion

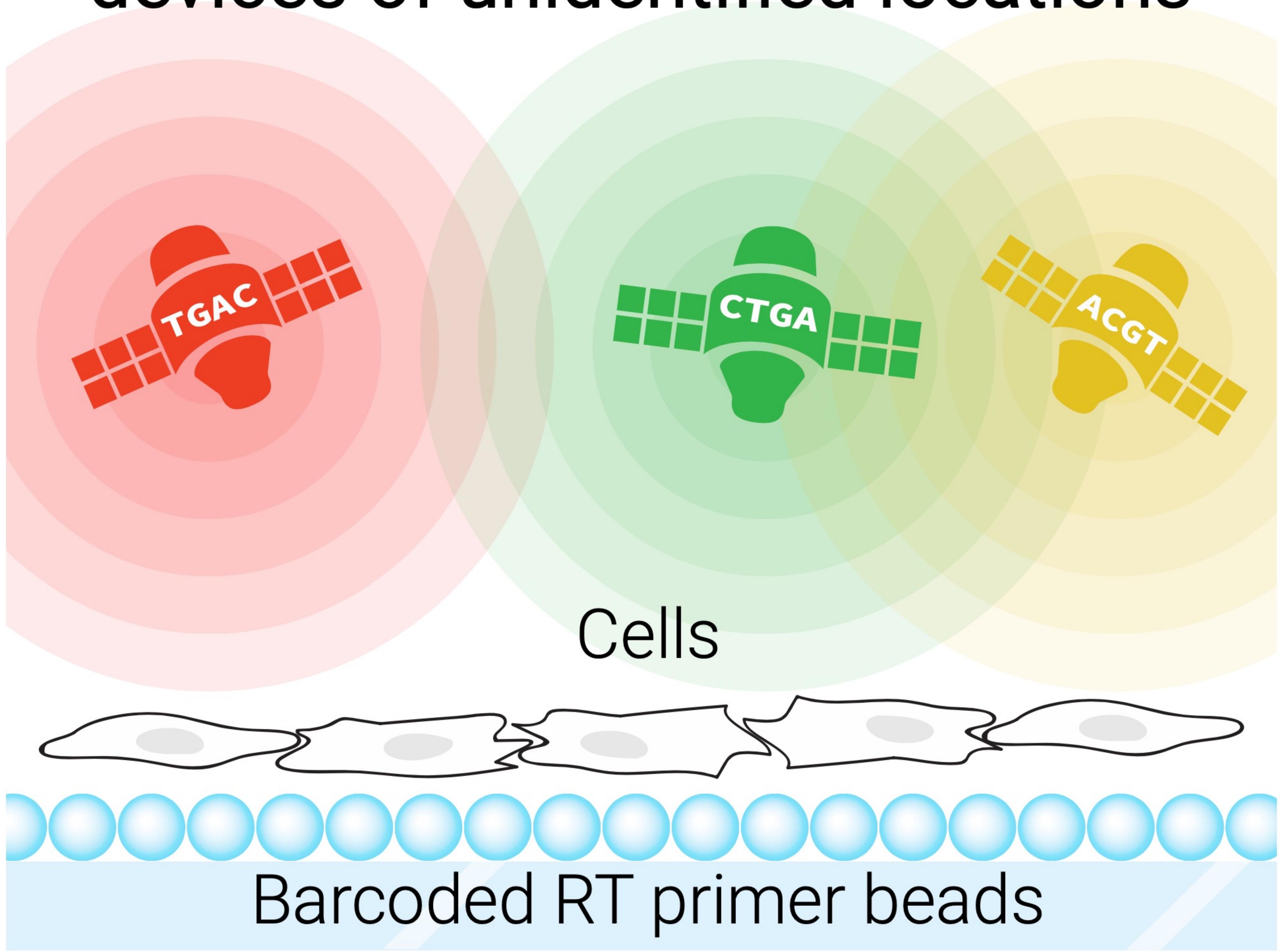
<100 cells with single molecule resolution

GPS-seq

Localization of positional barcodes through capture of satellite barcodes

100x scalability compared to the current ST

Diffusion of barcodes from satellite devices of unidentified locations



Greenstreet et al, bioRxiv 2022

GPS-seq

Transcriptome

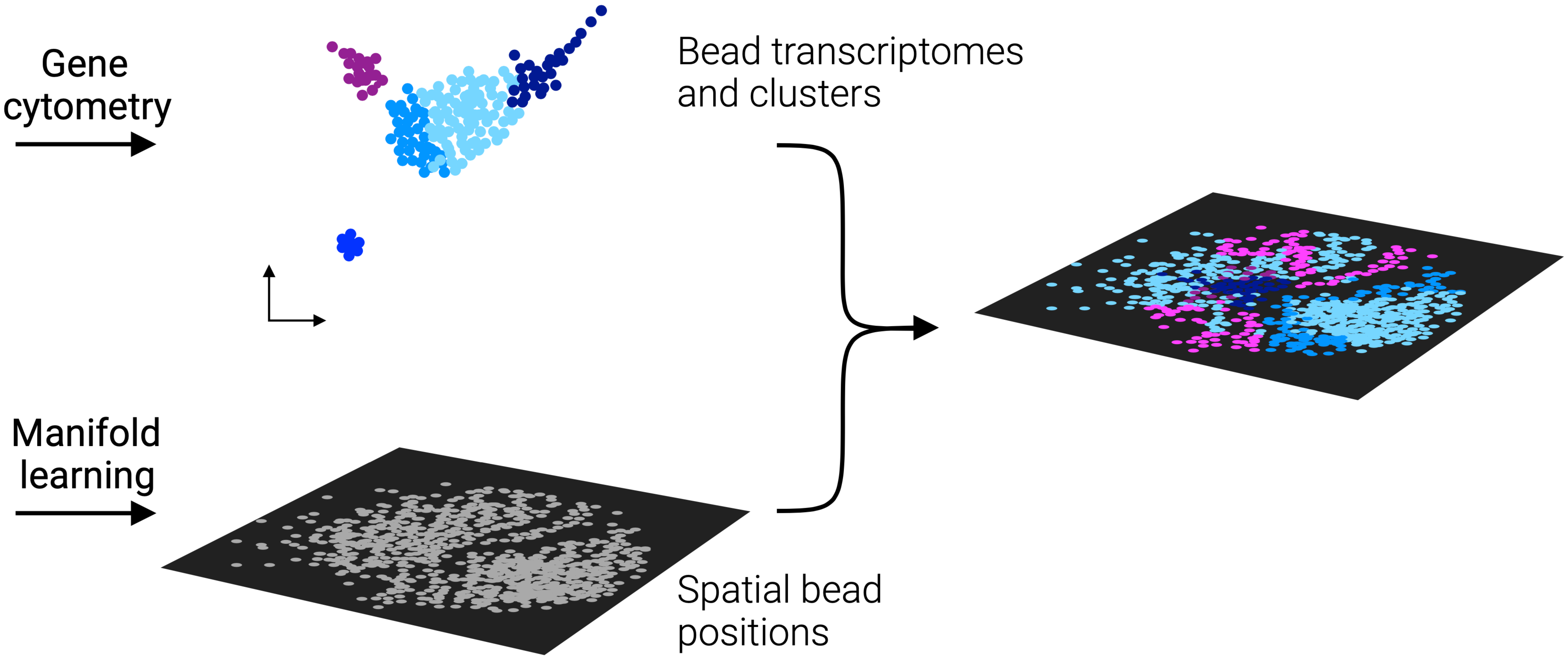
Bead BCs

4	34	9	32
3	0	67	2
0	43	1	78
17	5	0	3

Satellite BCs

Bead BCs

0	5	2	4
6	0	0	1
0	0	1	0
0	0	0	23



Gene cytometry

Manifold learning

Bead transcriptomes and clusters

Spatial bead positions

Greenstreet et al, bioRxiv 2022

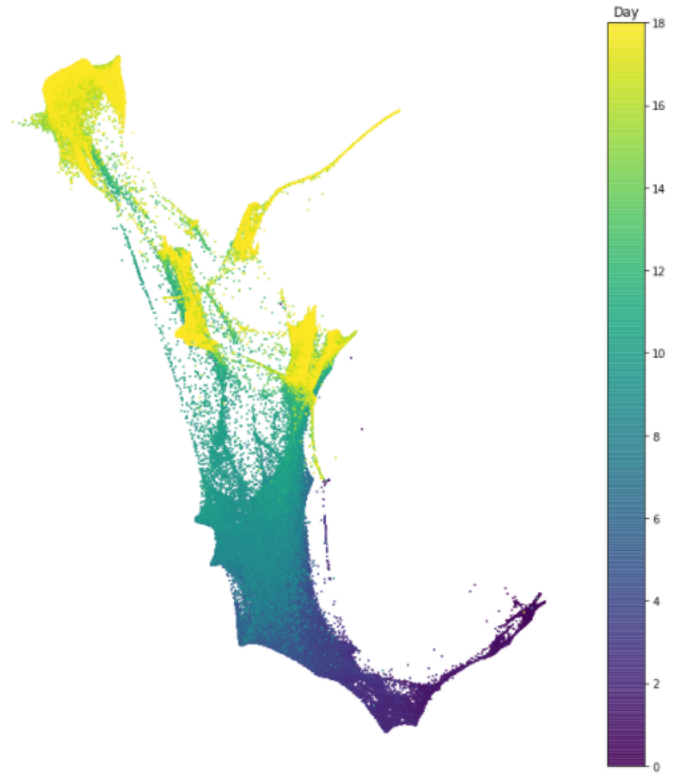
Waddington-OT Tutorial

- Introduction
- **Visualizing and Exploring the Data**
  - Notebook 1: Visualizing and exploring the data
- Inferring temporal couplings with optimal transport
- Interpreting transport maps
- Validation and Experimental Design
- Inferring gene regulatory networks

Visualizing and Exploring the Data

In this section we explore the time-series of reprogramming from Schiebinger et al. 2019. The dataset consists of 39 time points collected over 18 days of reprogramming. In the following notebook we visualize the dataset in two dimensions, and we examine patterns of gene expression programs.

Notebook 1: Visualizing and exploring the data



In this notebook we visualize the data with the force layout embedding. This is a graph visualization tool which we apply to layout a nearest neighbor graph constructed from our single cell gene expression data. There is a node for each cell, and each cell is connected to its k nearest neighbors. Then the cells are arranged in 2D so that cells connected by an edge attract, and cells not connected by an edge repel each other. This visualization is used many times throughout the tutorial.

To get a basic idea of the lay of the land, we examine patterns of gene expression programs. To do this, we score each cell according to expression of a dictionary of **gene signatures**; in other words we test whether the set of genes in a signature is significantly expressed in each cell. Based on these gene signatures, we define sets of cells. In the following notebooks, we will use optimal transport to examine the developmental trajectories leading to these cell sets.

- View notebook 1

References

Towards a mathematical theory of trajectory inference

https://arxiv.org/abs/2102.09204

Trajectory Inference via Mean-field Langevin in Path Space

https://arxiv.org/abs/2205.07146

LineageOT is a unified framework for lineage tracing and trajectory inference

https://www.nature.com/articles/s41467-021-25133-1

Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming

https://www.sciencedirect.com/science/article/pii/S009286741930039X

Developmental single-cell transcriptomics in the Lytechinus variegatus sea urchin embryo

https://journals.biologists.com/dev/article/148/19/dev198614/272307/Developmental-single-cell-transcriptomics-in-the

NSERC
CRSNG

CIHR IRSC

Thank you!

BURROUGHS
WELLCOME
FUND

New Frontiers in Research Fund
Fonds Nouvelles frontières en recherche

Want to learn more?

contact: geoff@math.ubc.ca

Positions available!! (postdocs and students)