

Optimal Transport for Single Cell Genomics

Gabriel Peyré

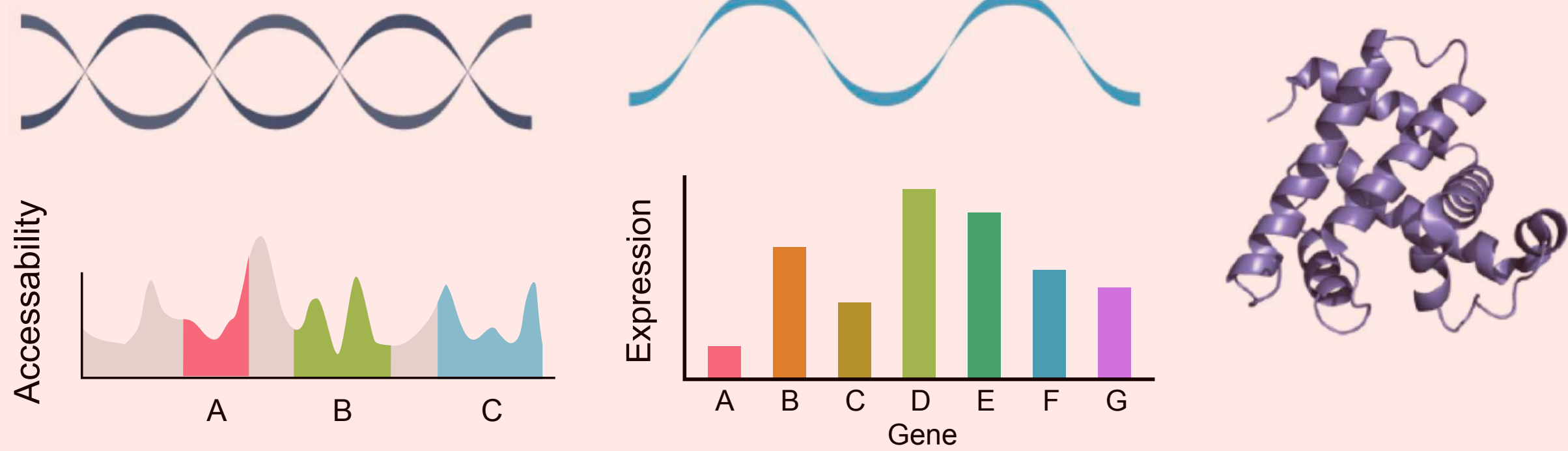


Geert-Jan
Huizing

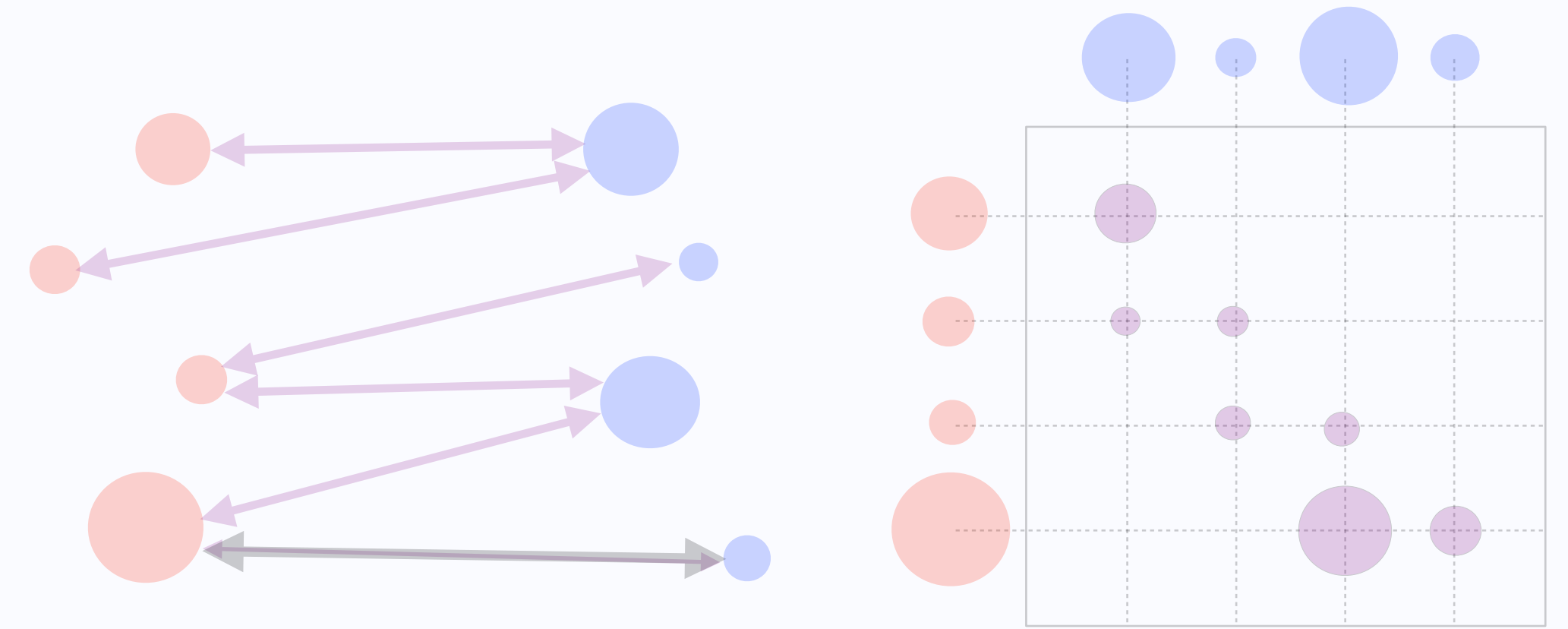
Laura
Cantini



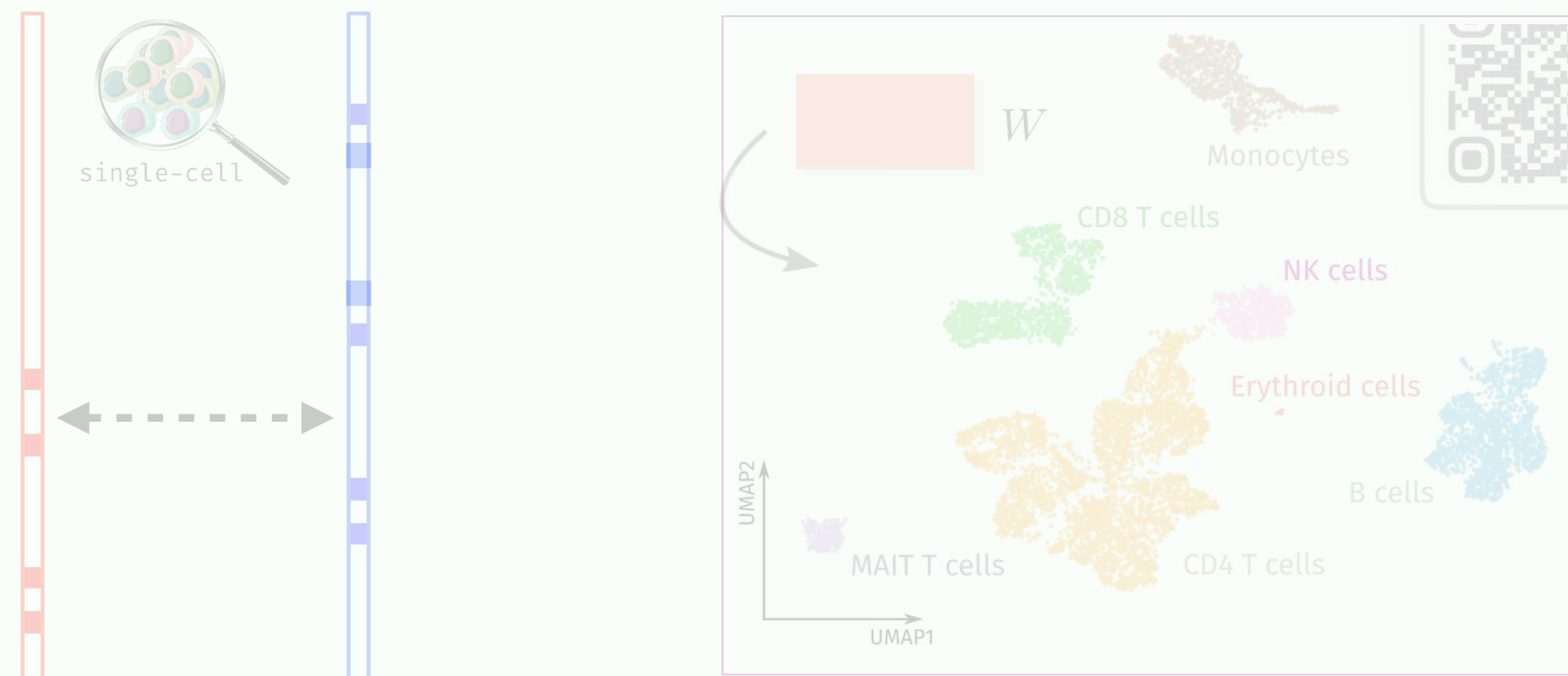
Single cell genomics



Optimal Transport



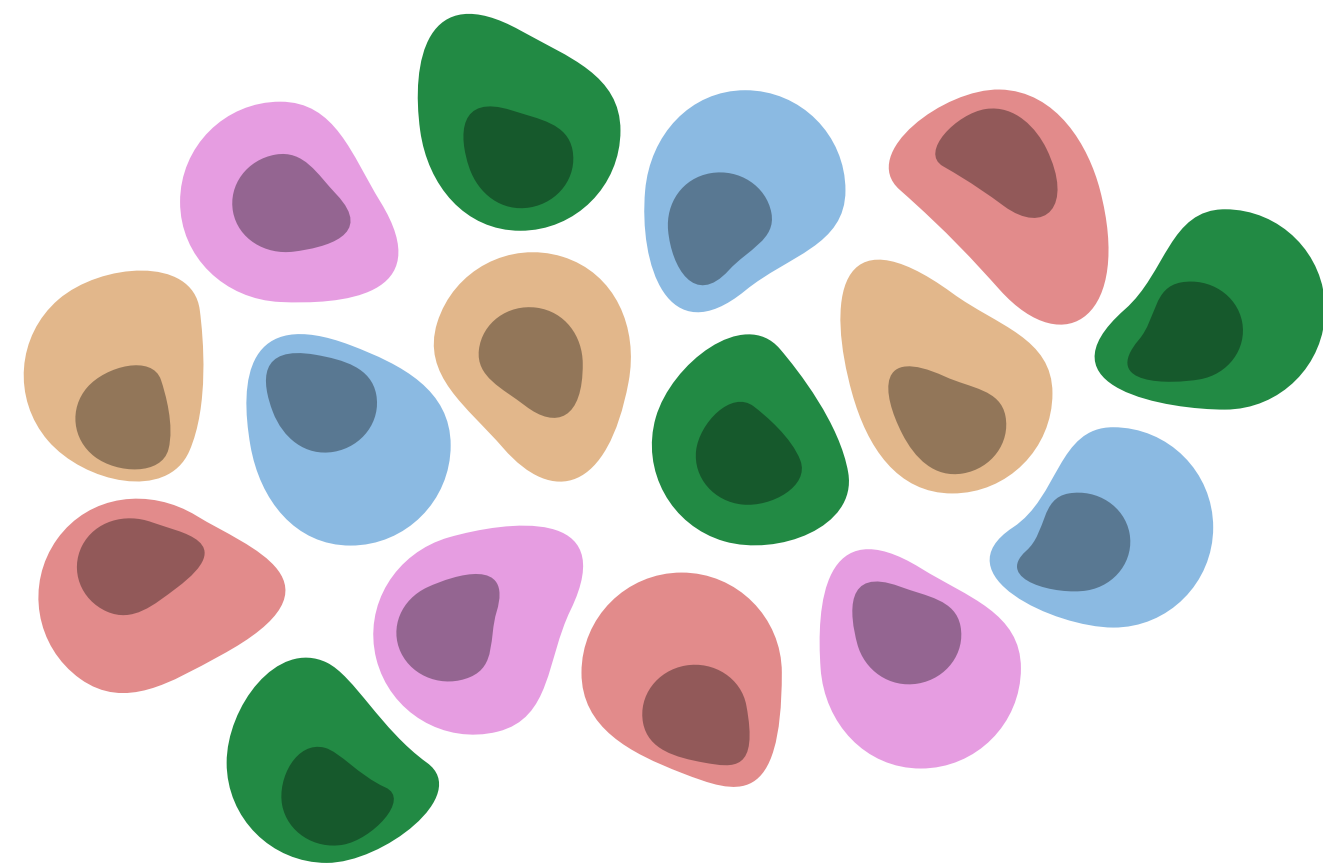
Comparing Cell-Cell with OT



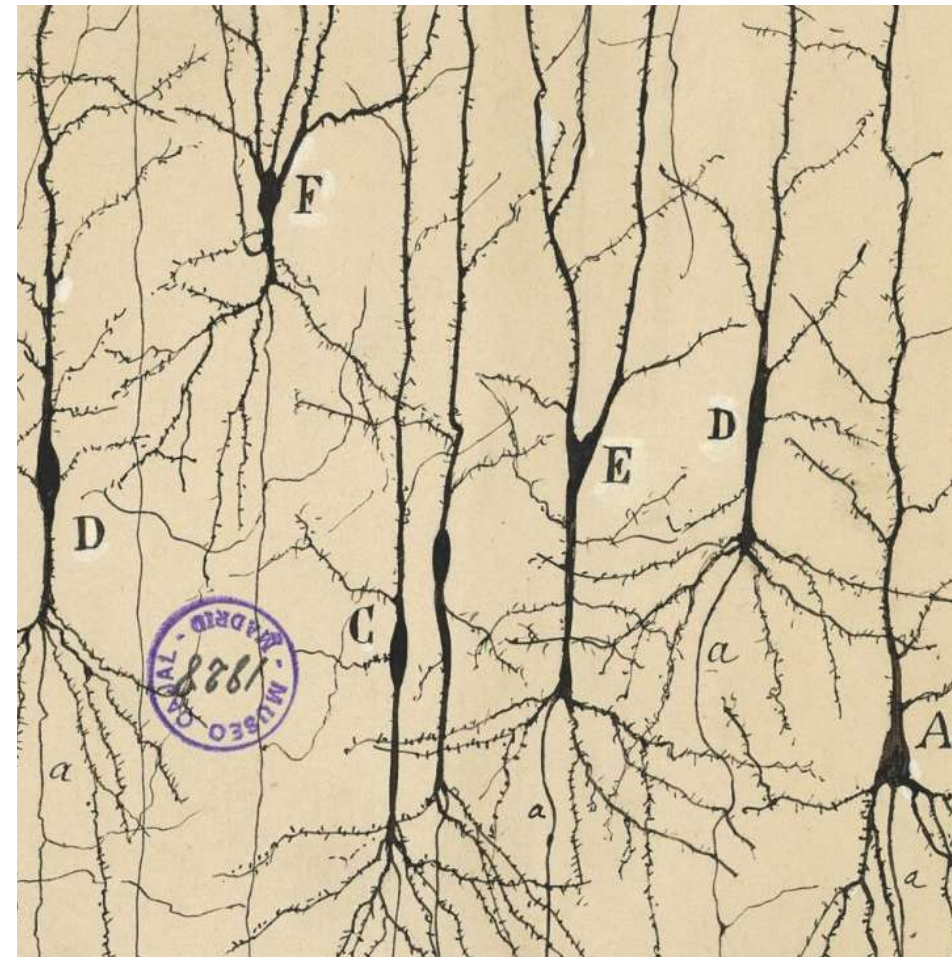
Tracking Cell Evolution with OT



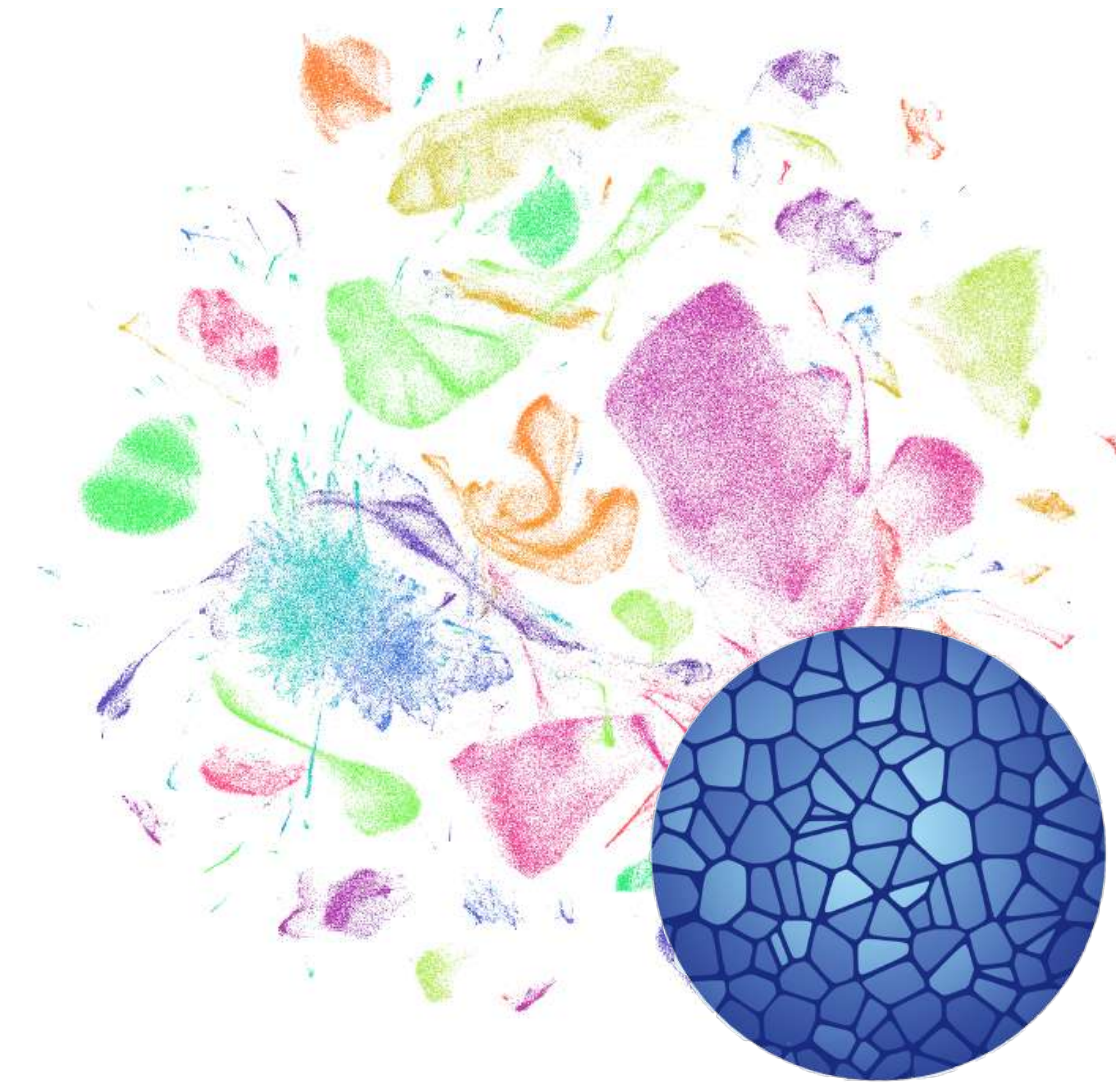
Unraveling cell diversity



$\sim 10^{13}$ cells in the human body, with vastly different functions.



Early efforts to cartography cell identity relied on microscopy.
[Ramón y Cajal, 1899]

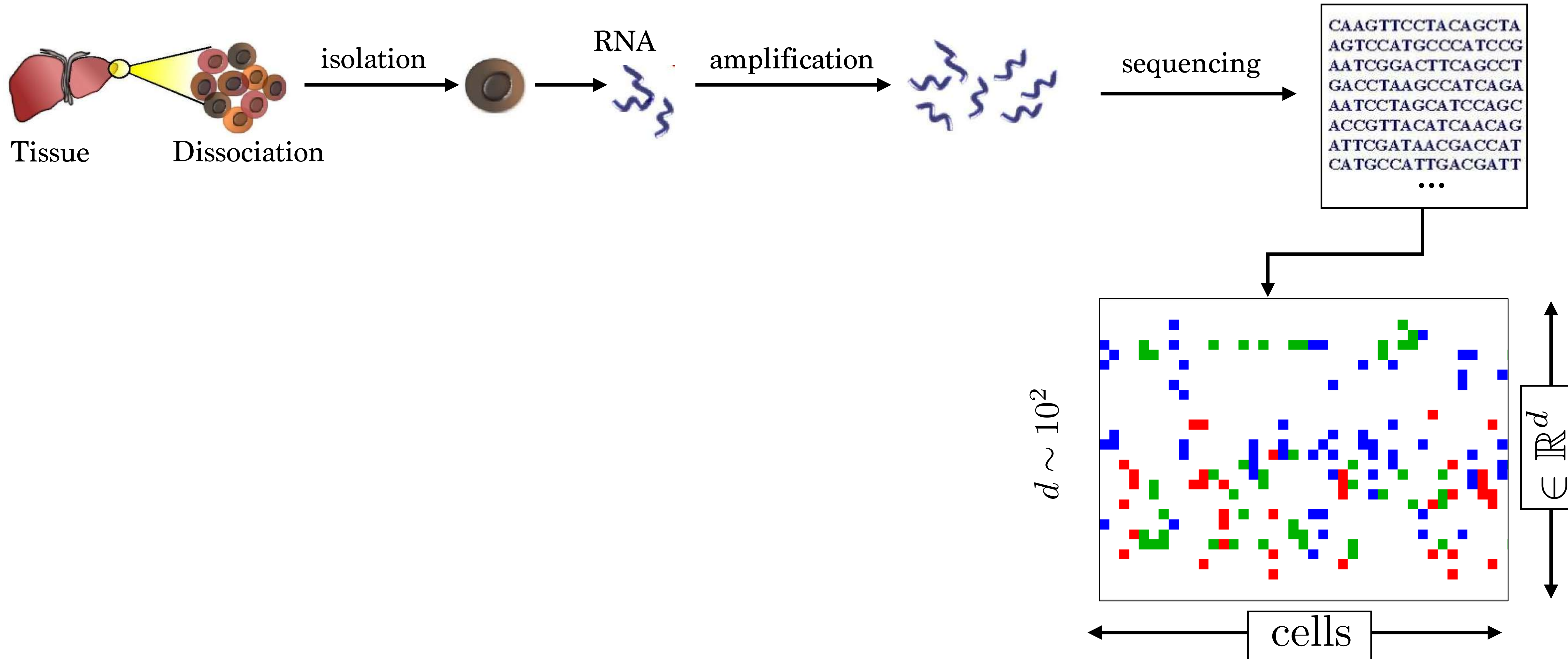


Recent initiatives measure the molecular profile of the cell.
[Regev et al., eLife, 2017]

Single Cell Multi-omics

Understanding cell diversity: many types, many states for each type.

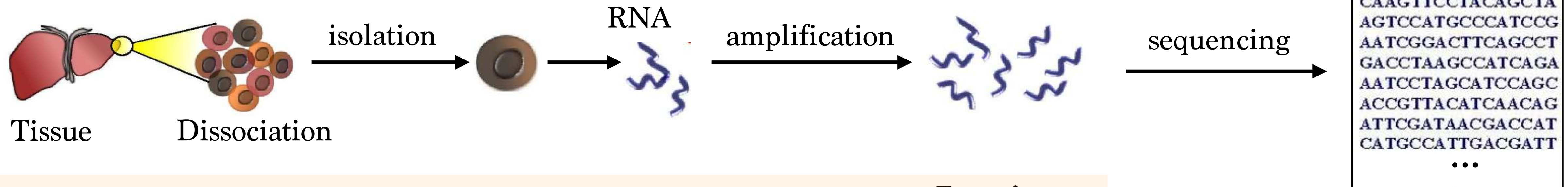
Applications: cancer mutations, dynamic of adaptation, development, ...



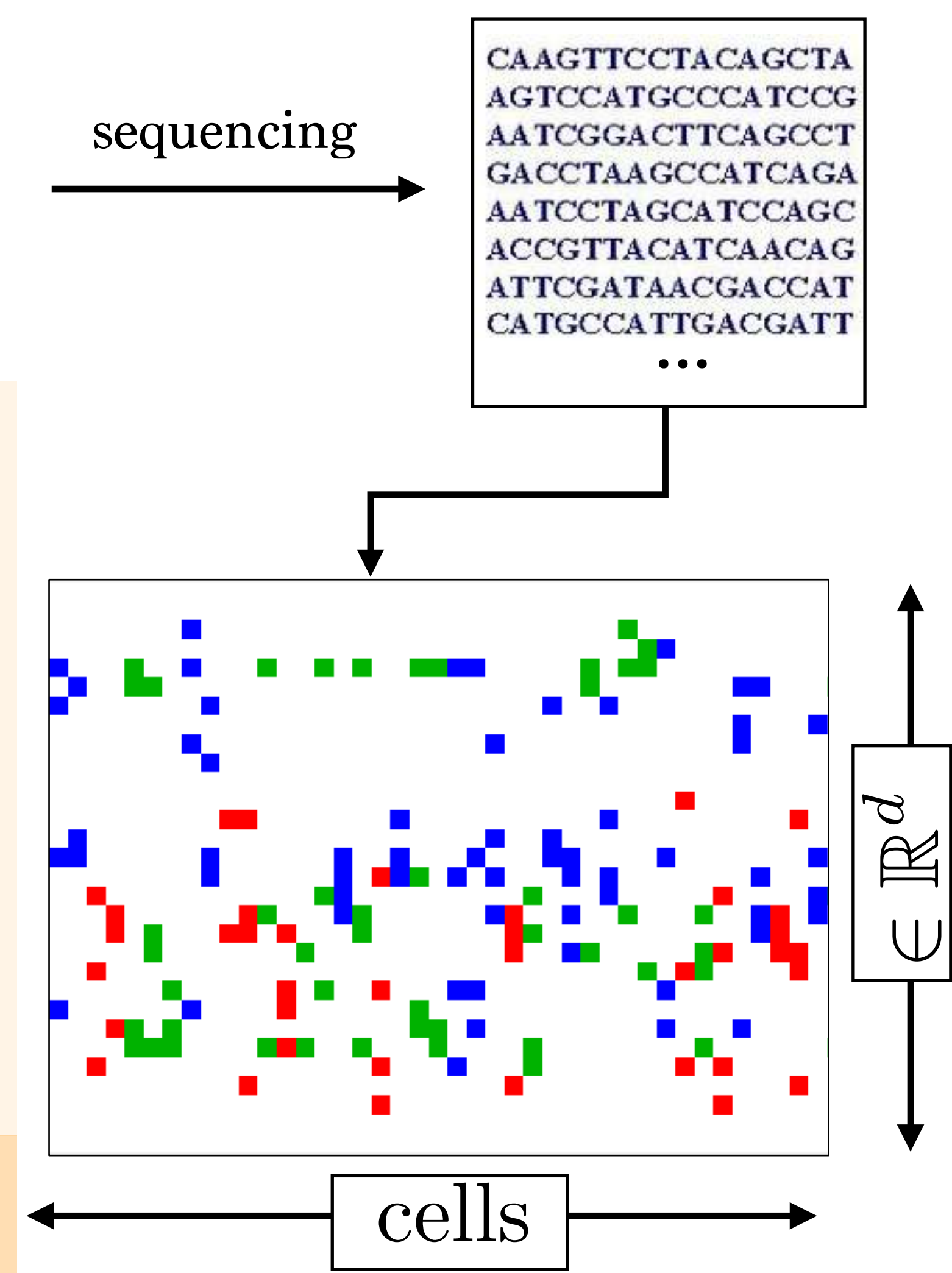
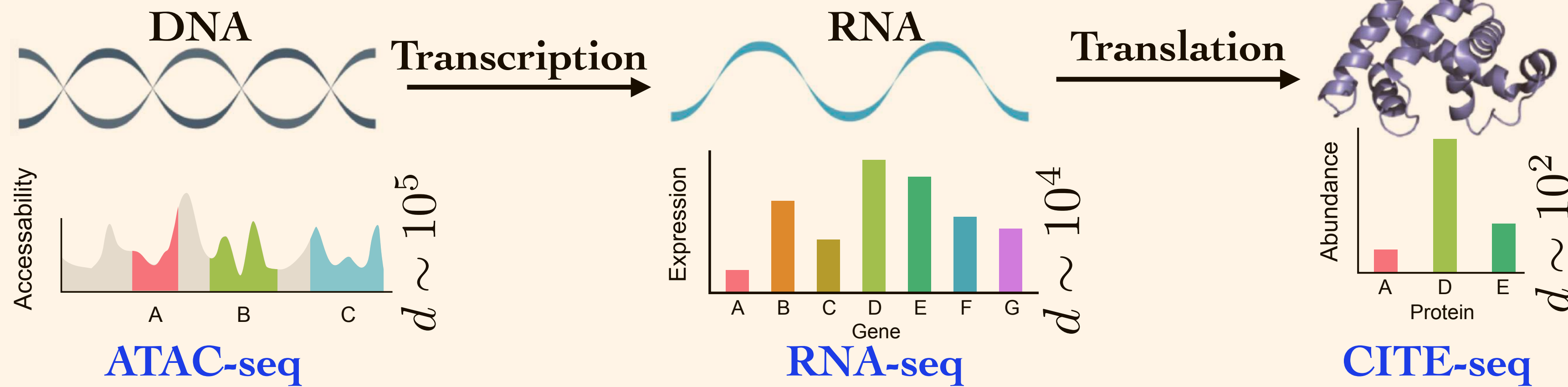
Single Cell Multi-omics

Understanding cell diversity: many types, many states for each type.

Applications: cancer mutations, dynamic of adaptation, development, ...

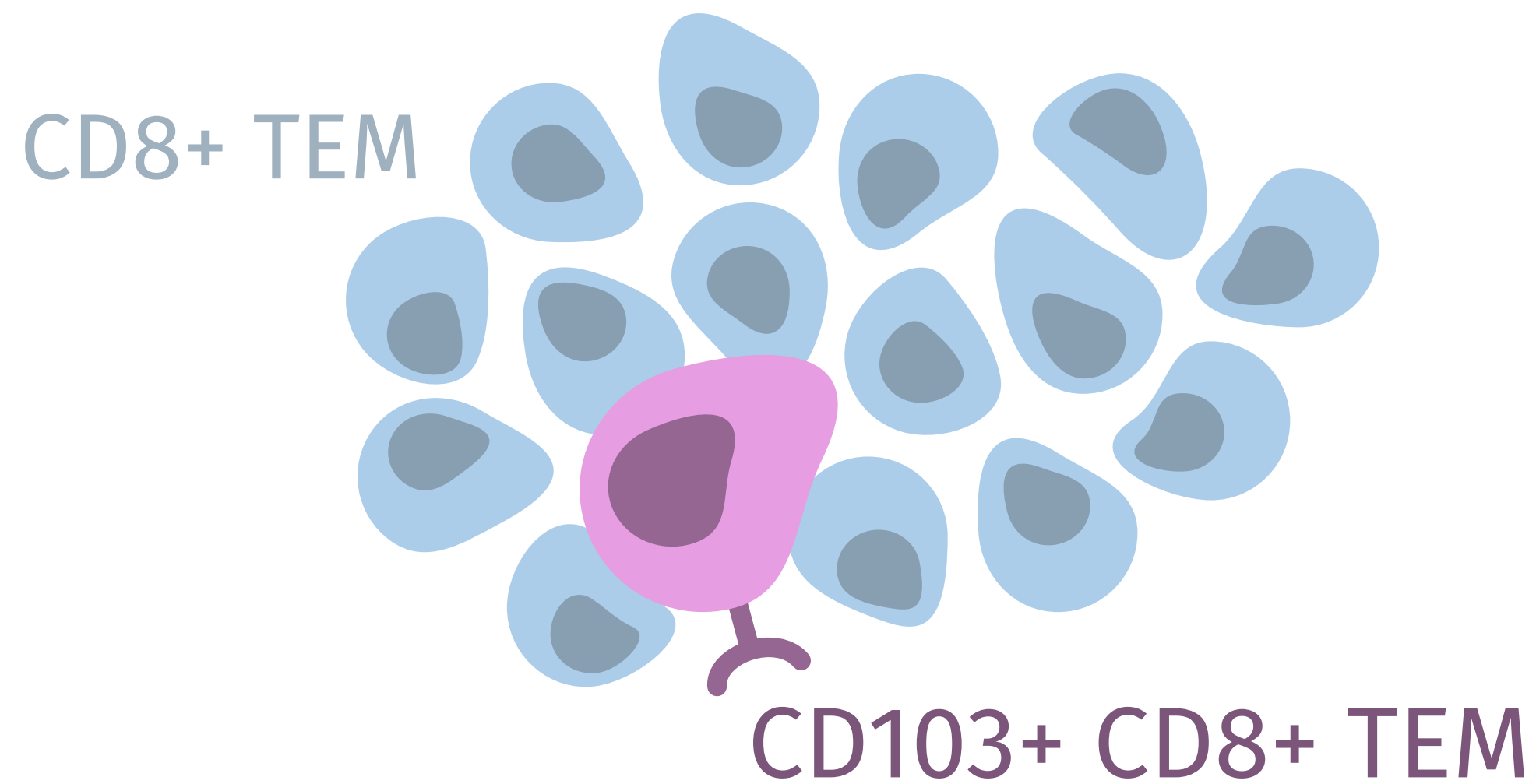


Different omics spaces:



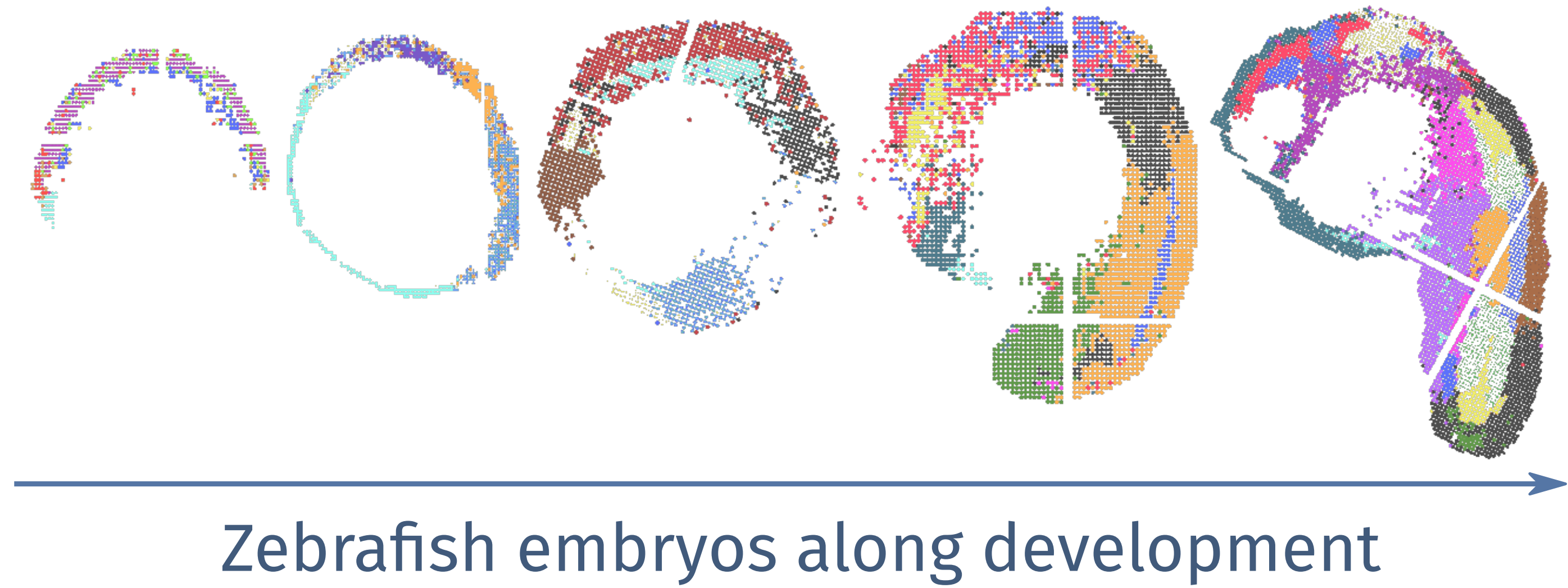
Multi-omics integration: next frontier ...

Examples of recent biological discoveries



The joint profiling of **gene expression and surface proteins** enabled to identify a new subpopulation of CD8 TEM cells.

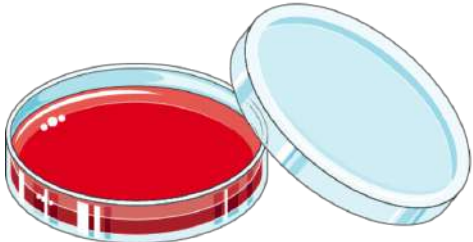
[Hao et al., Cell, 2021]



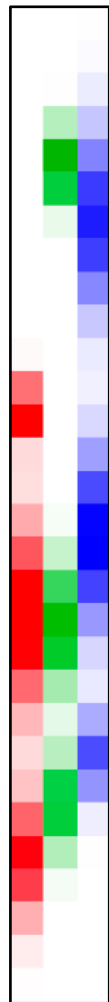
Spatial transcriptomics profiled across time have allowed to study development at unprecedented resolution.

[Liu et al., Developmental Cell, 2022]

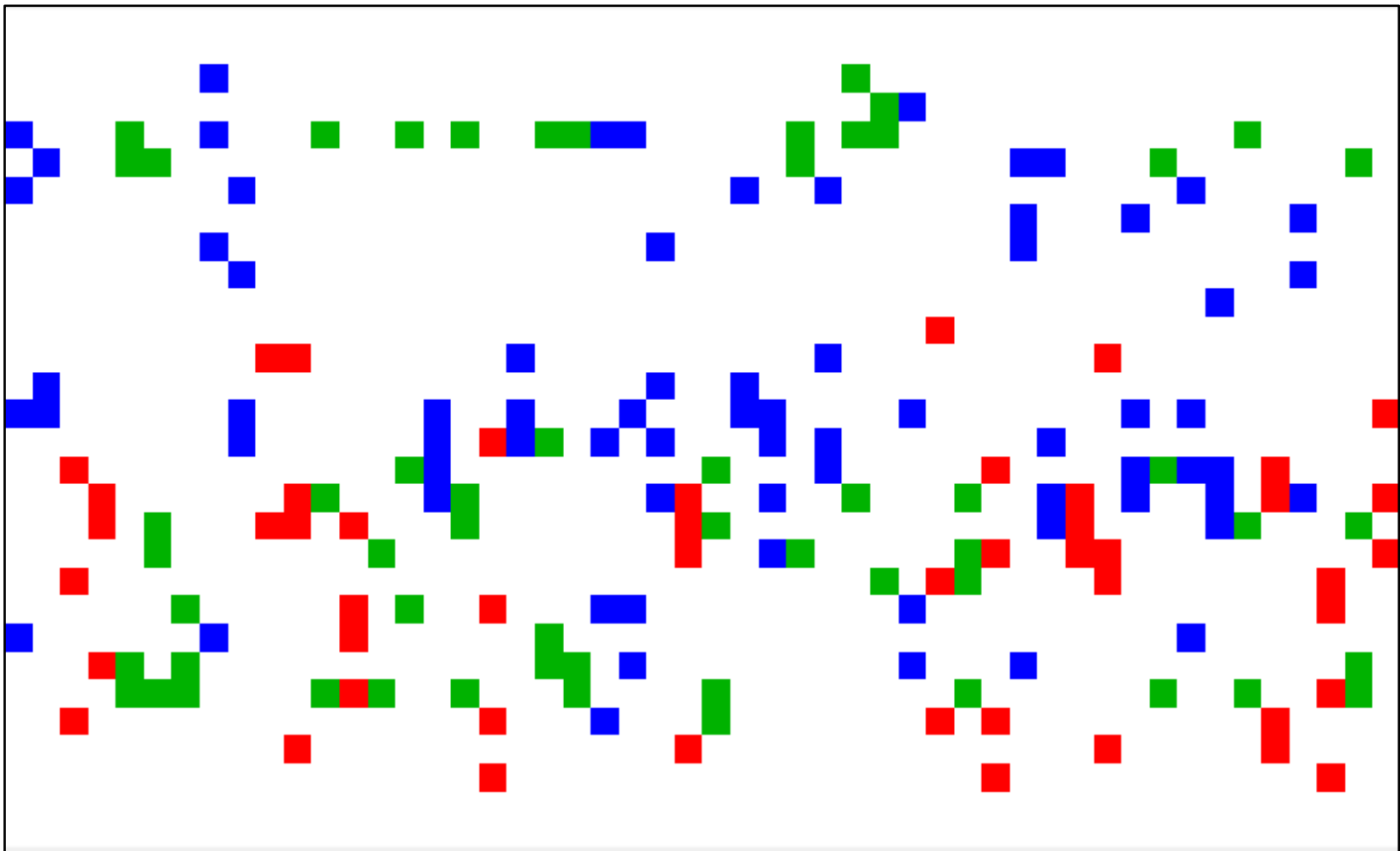
Comparing Distributions for Single Cells



Bulk



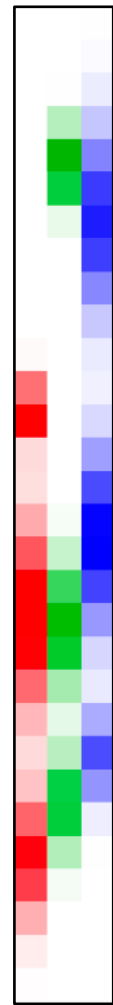
Single cell



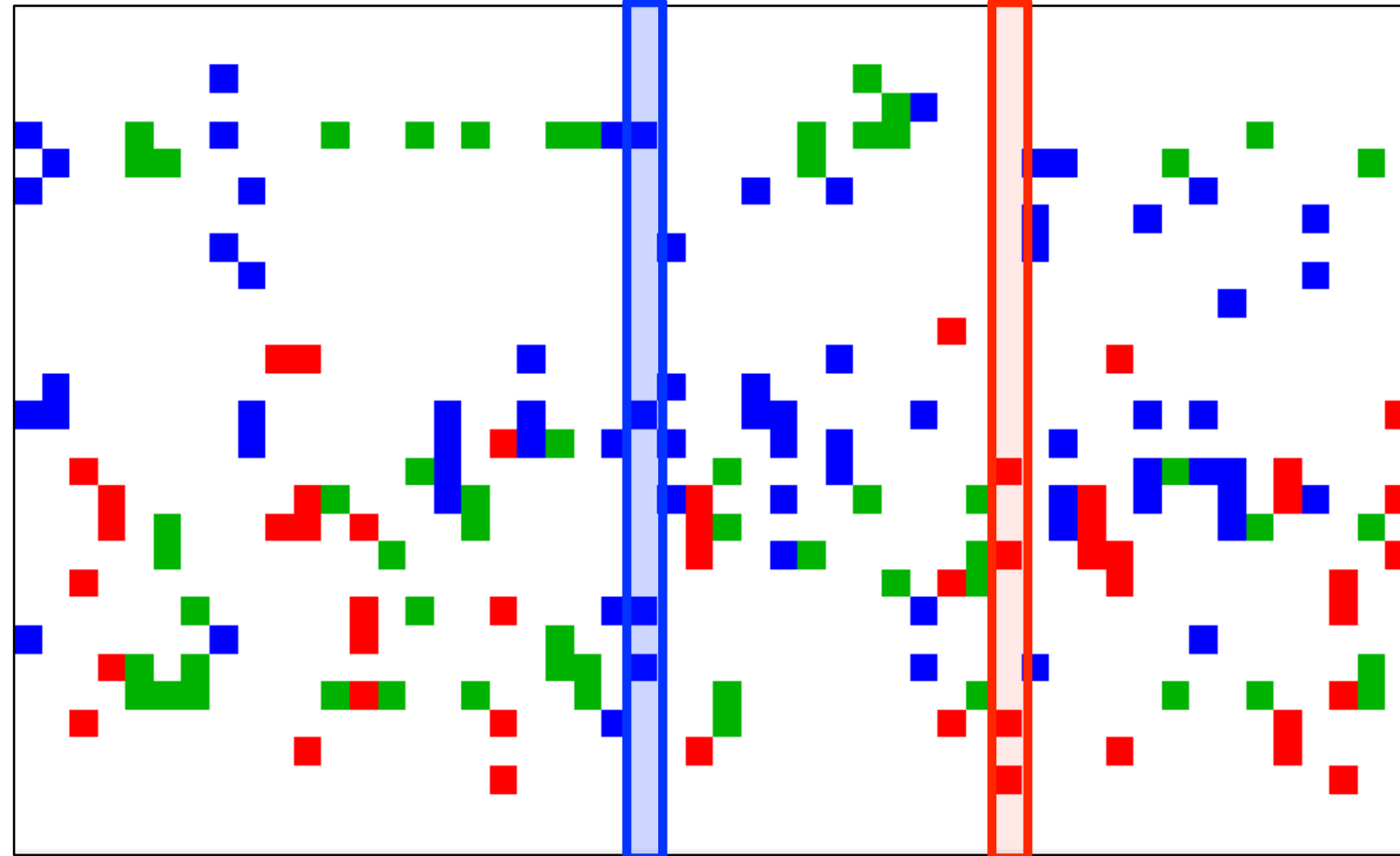
Comparing Distributions for Single Cells



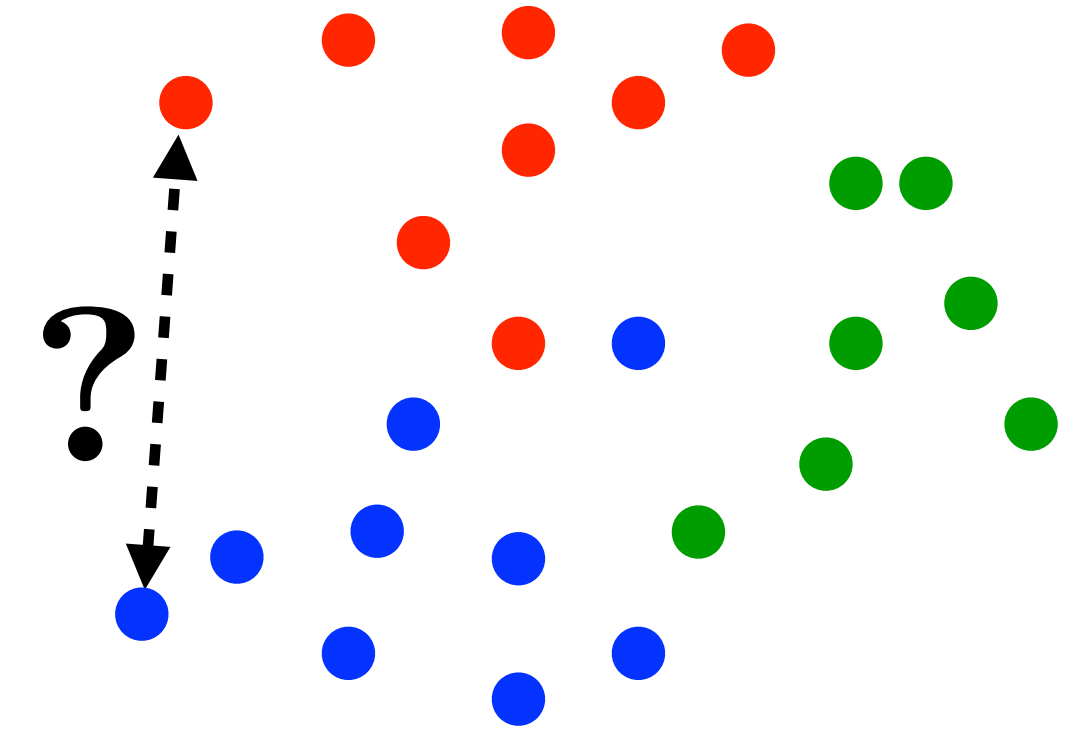
Bulk



Single cell



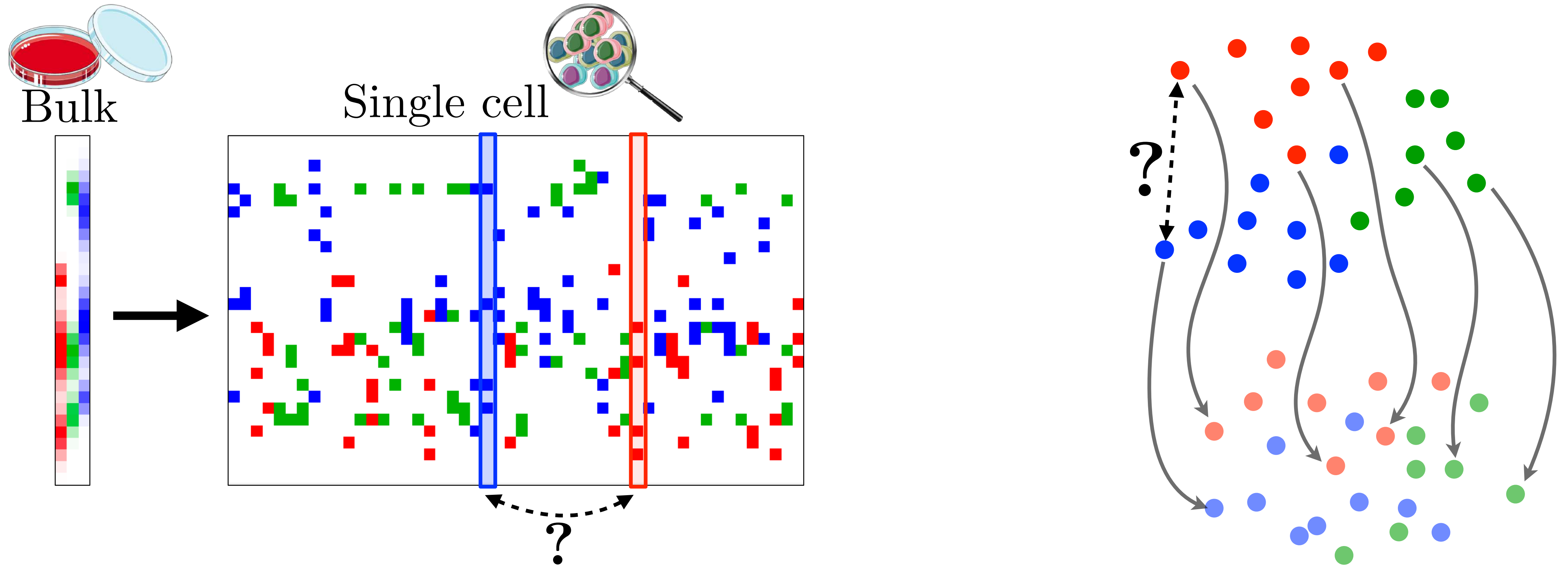
?



Cluster cells from a single biopsy.

“Distance” between cells? \longrightarrow OT on genes' space.

Comparing Distributions for Single Cells



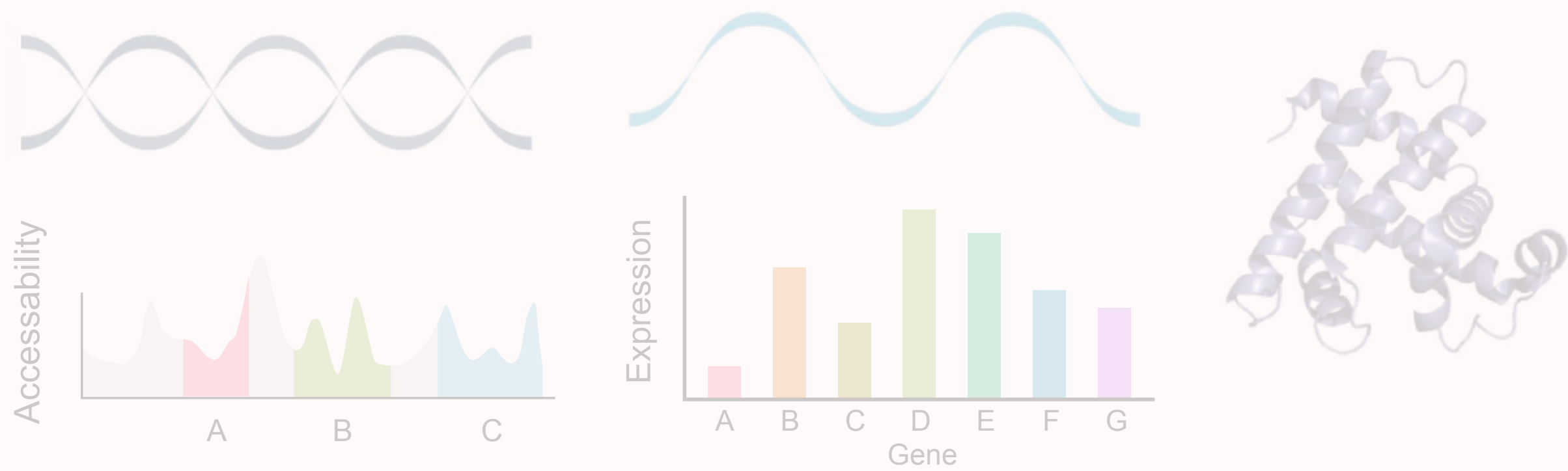
Cluster cells from a single biopsy.

"Distance" between cells? \longrightarrow OT on genes' space.

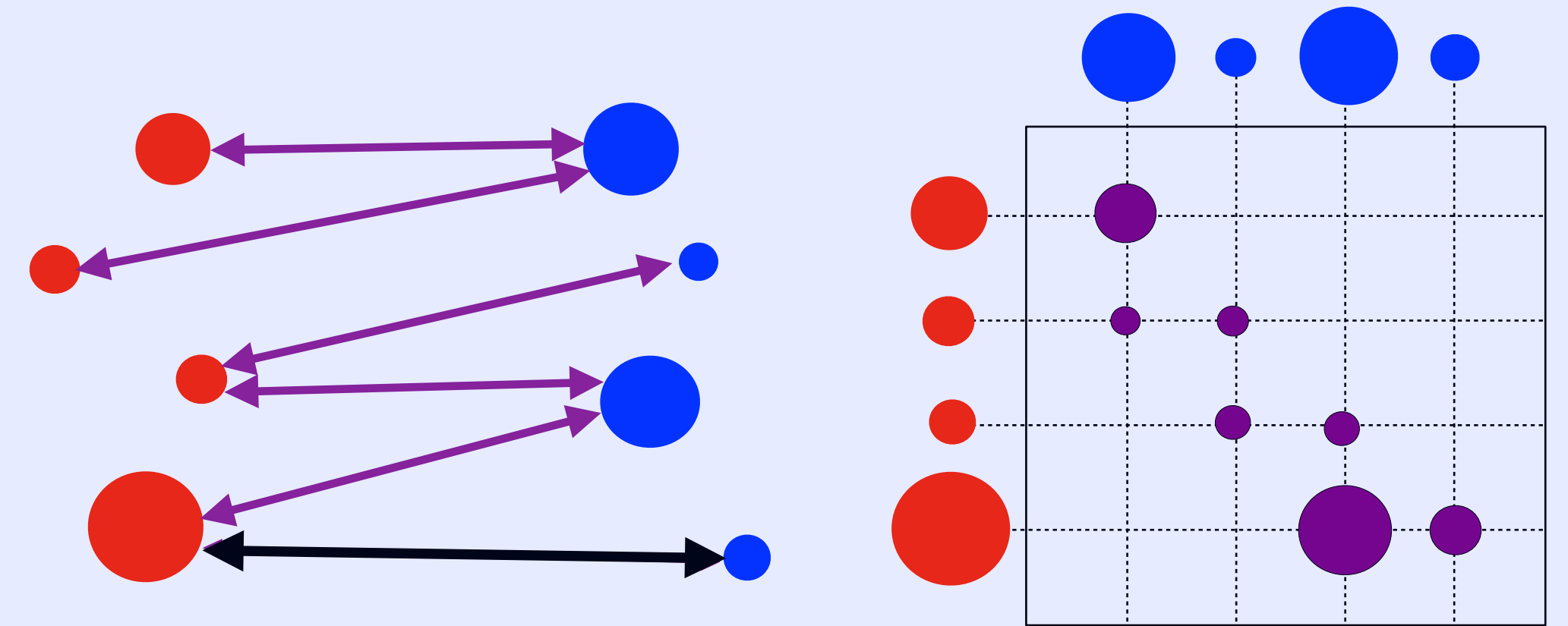
Match cells between two biopsies.

\longrightarrow OT on cells' space.

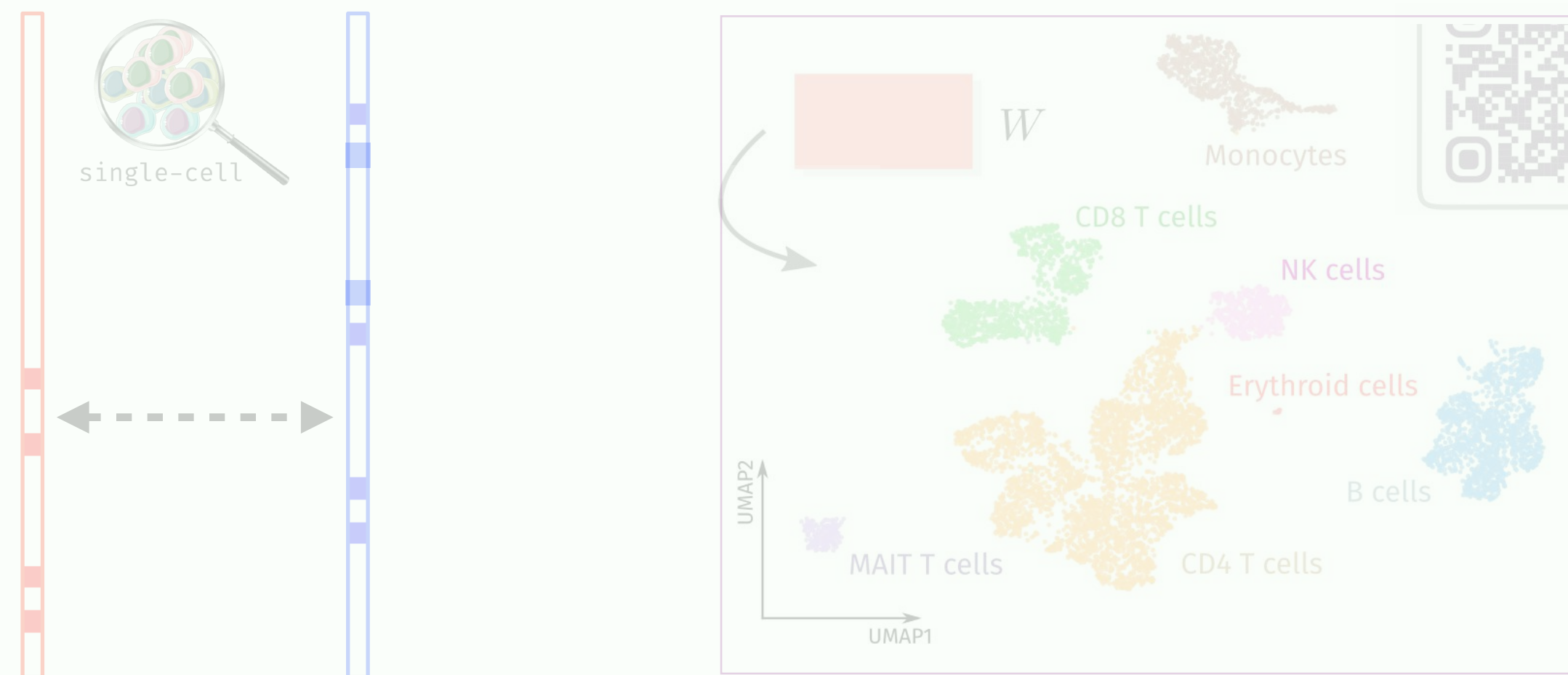
Single cell genomics



Optimal Transport



Comparing Cell-Cell with OT



Tracking Cell Evolution with OT



Monge's Problem

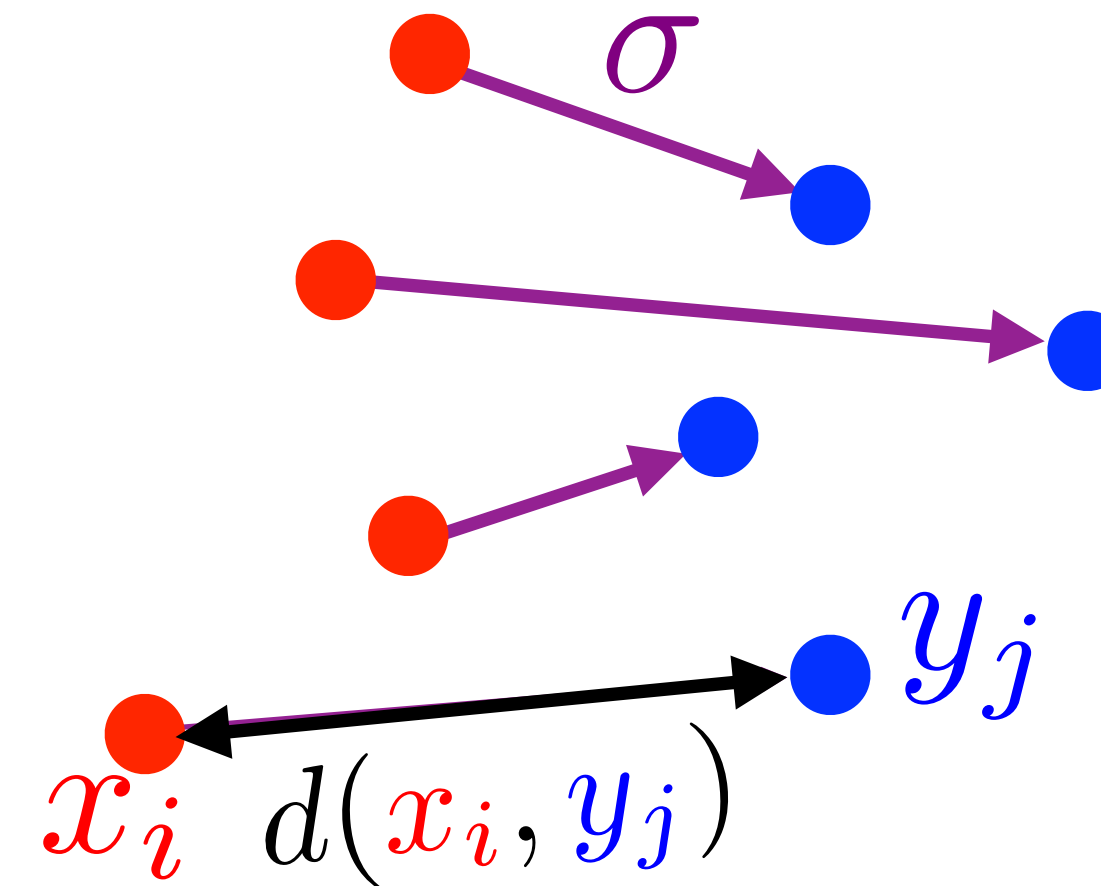
Points $(x_i)_i, (y_j)_j$

Permutation:

$$\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$$

Monge optimal matching:

$$D(X, Y) = \min_{\sigma} \sum_{i=1}^n d(x_i, y_{\sigma(i)})$$



[Monge 1784]

M É M O I R E
S U R L A
T H É O R I E D E S D É B L A I S
E T D E S R E M B L A I S.

Par M. M O N G E.

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'en suit 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 fera un *minimum*.

Monge's Problem

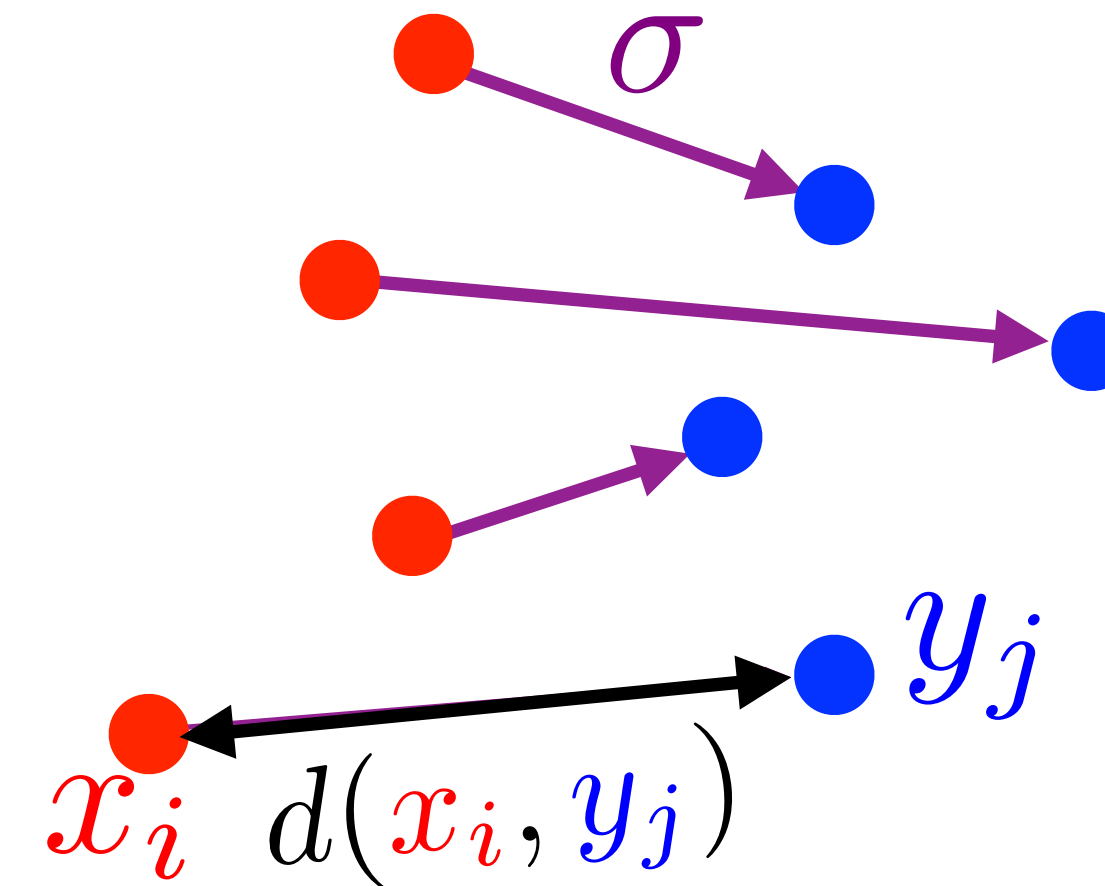
Points $(x_i)_i, (y_j)_j$

Permutation:

$$\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$$

Monge optimal matching:

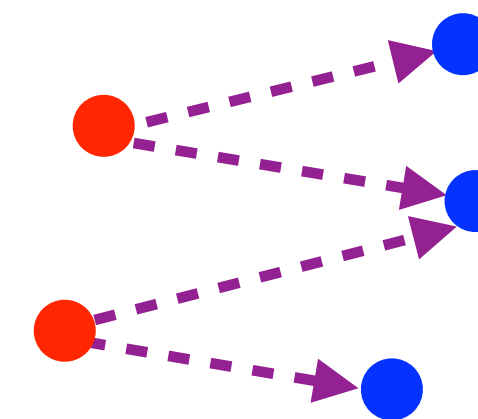
$$D(X, Y) = \min_{\sigma} \sum_{i=1}^n d(x_i, y_{\sigma(i)})$$



[Monge 1784]

→ Seems intractable: $n!$ possibilities.

→ Different number of points?



M É M O I R E
S U R L A
T H É O R I E D E S D É B L A I S
E T D E S R E M B L A I S.

Par M. M O N G E.

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'en suit 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 fera un *minimum*.

Kantorovitch's Formulation

Discrete distributions:

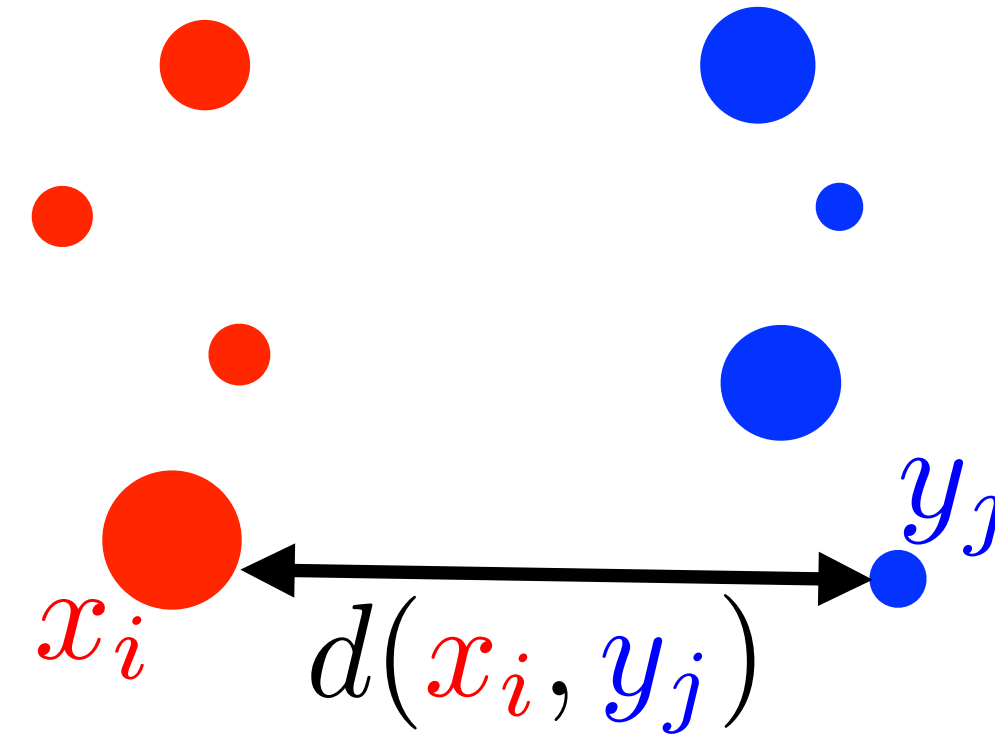
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

$$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0$.

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$



Kantorovitch's Formulation

Discrete distributions:

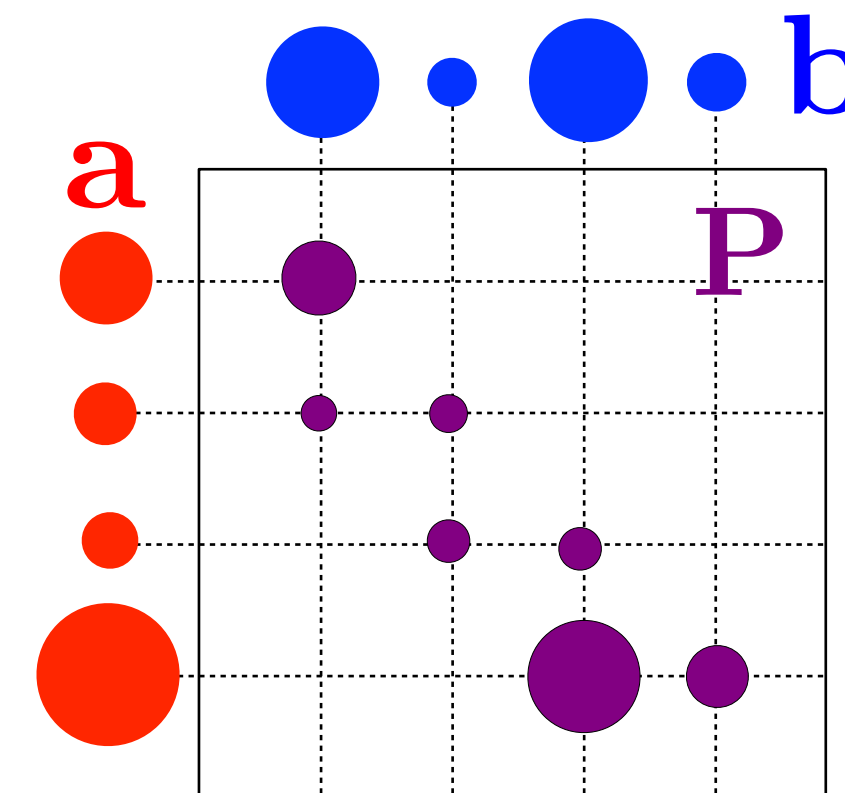
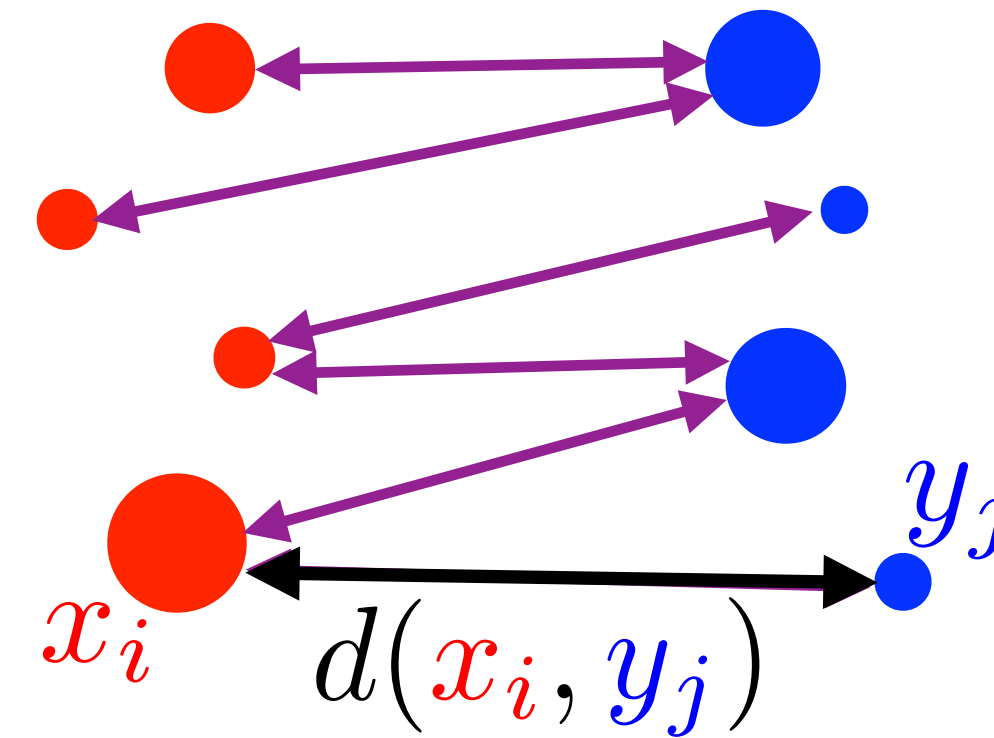
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

$$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0$.

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$



Couplings:

$$\sum_j P_{i,j} = \mathbf{a}_i$$

$$\sum_i P_{i,j} = \mathbf{b}_j$$

Leonid
Kantorovitch



Kantorovitch's Formulation

Discrete distributions:

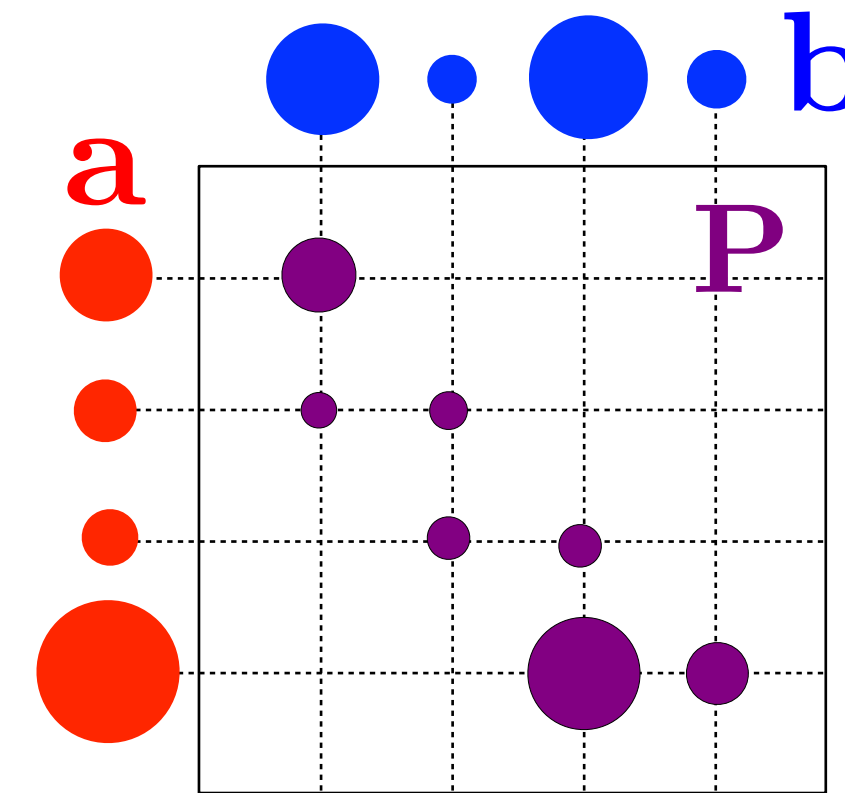
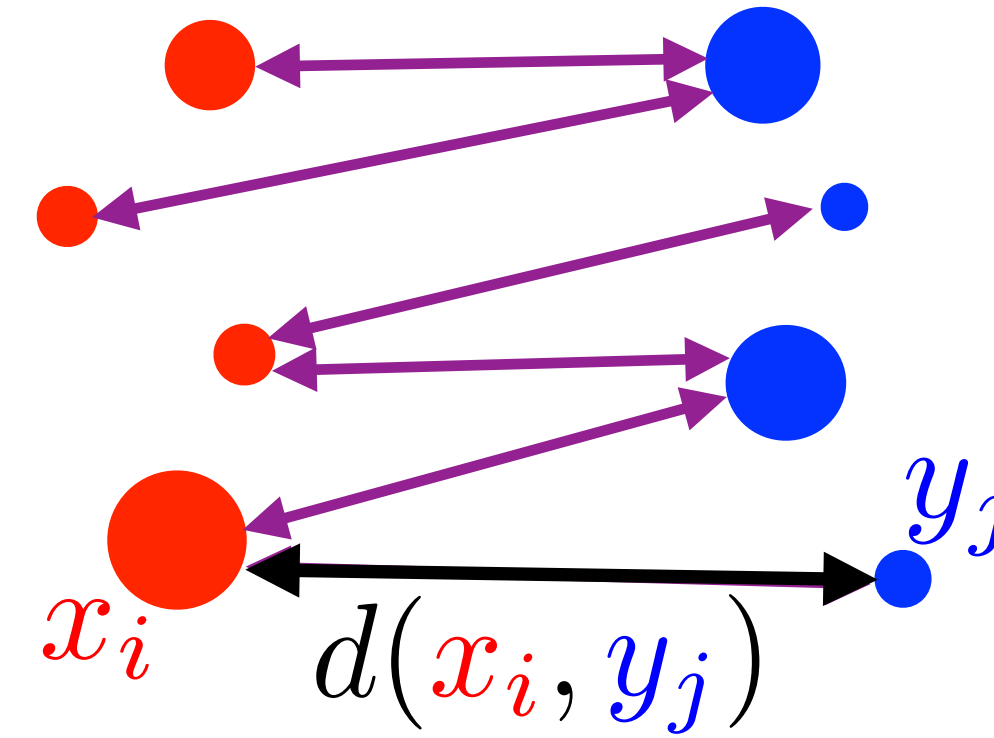
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

$$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$$

Points $(x_i)_i, (y_j)_j$

Weights $\mathbf{a}_i \geq 0, \mathbf{b}_j \geq 0$.

$$\sum_{i=1}^n \mathbf{a}_i = \sum_{j=1}^m \mathbf{b}_j = 1$$



Leonid Kantorovitch



Couplings:

$$\sum_j \mathbf{P}_{i,j} = \mathbf{a}_i$$

$$\sum_i \mathbf{P}_{i,j} = \mathbf{b}_j$$

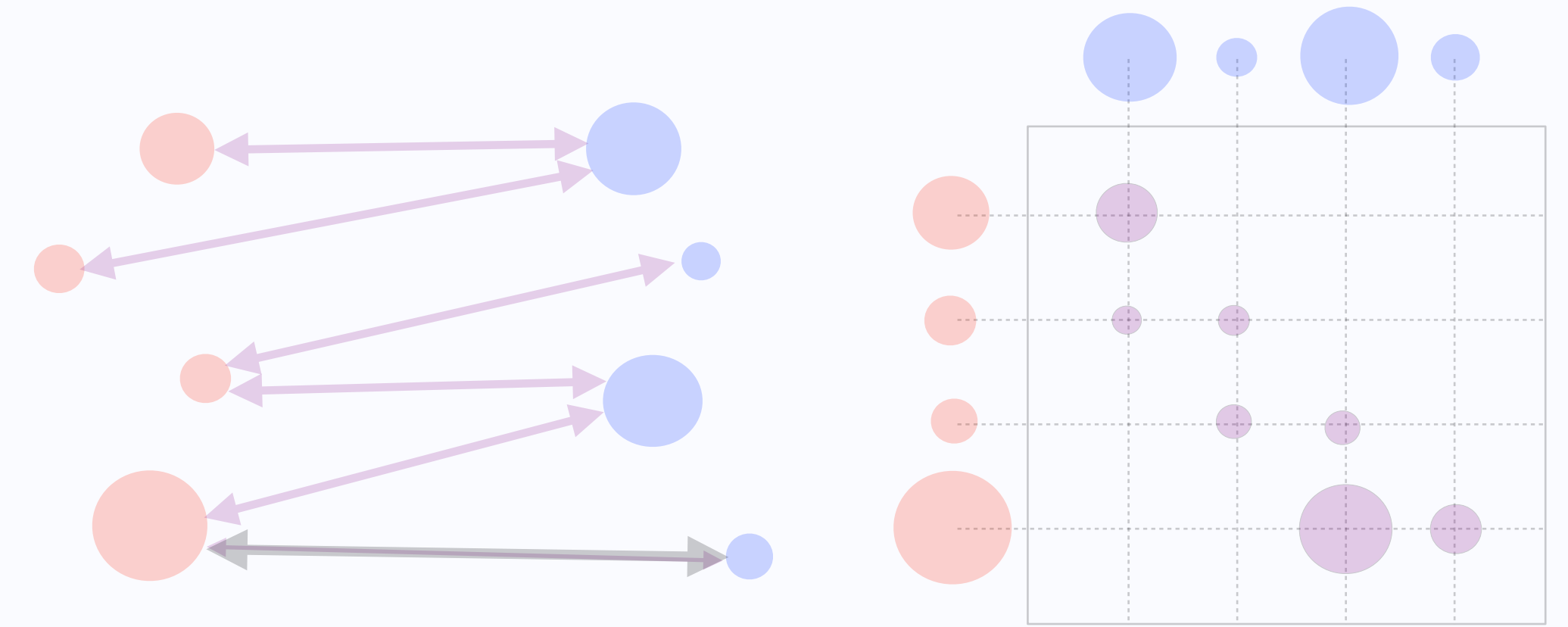
$$\min_{\mathbf{P}} \left\{ \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} ; \mathbf{P} \geq 0, \mathbf{P} \mathbf{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbf{1}_n = \mathbf{b} \right\}$$

[Kantorovitch 1942]

George Dantzig

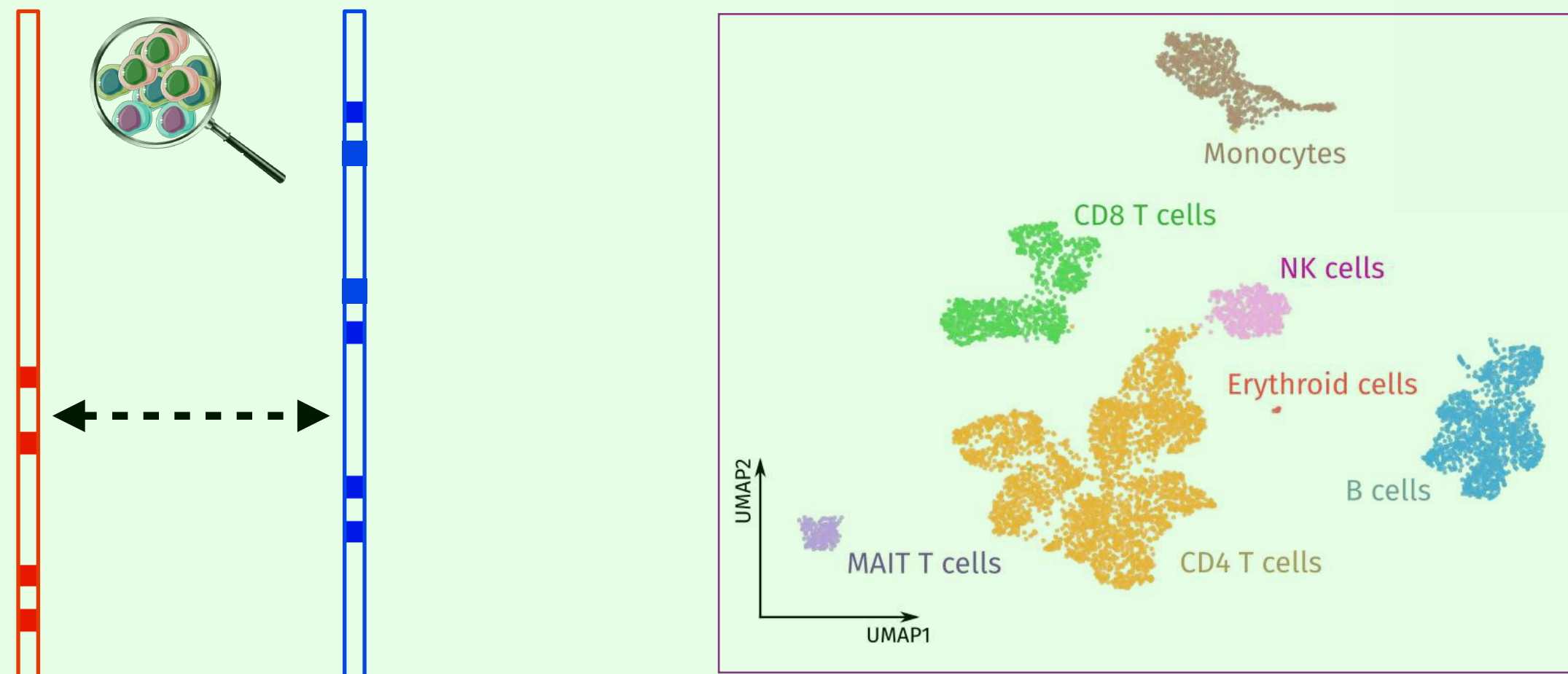
Single cell genomics

Optimal Transport

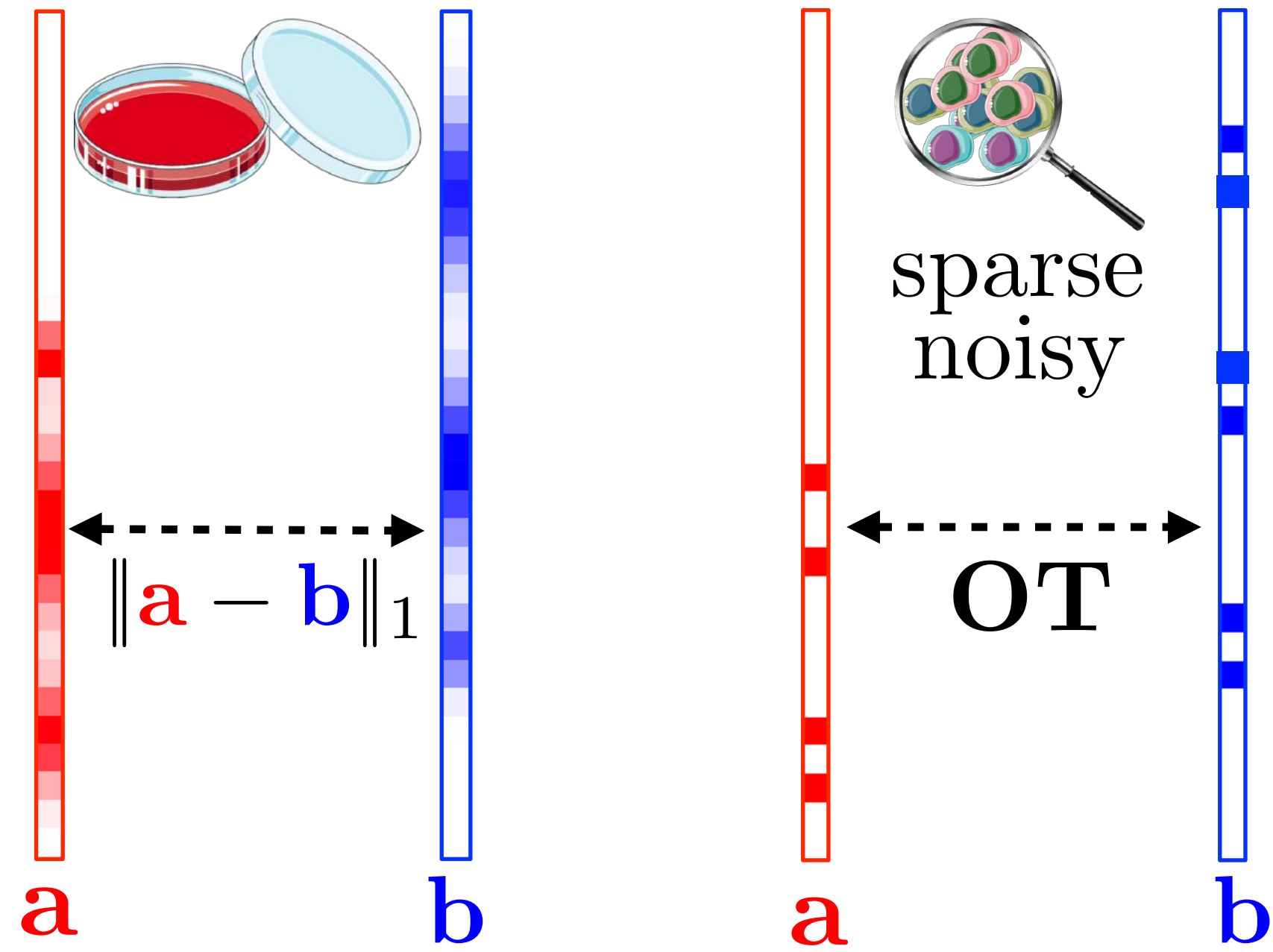


Comparing Cell-Cell with OT

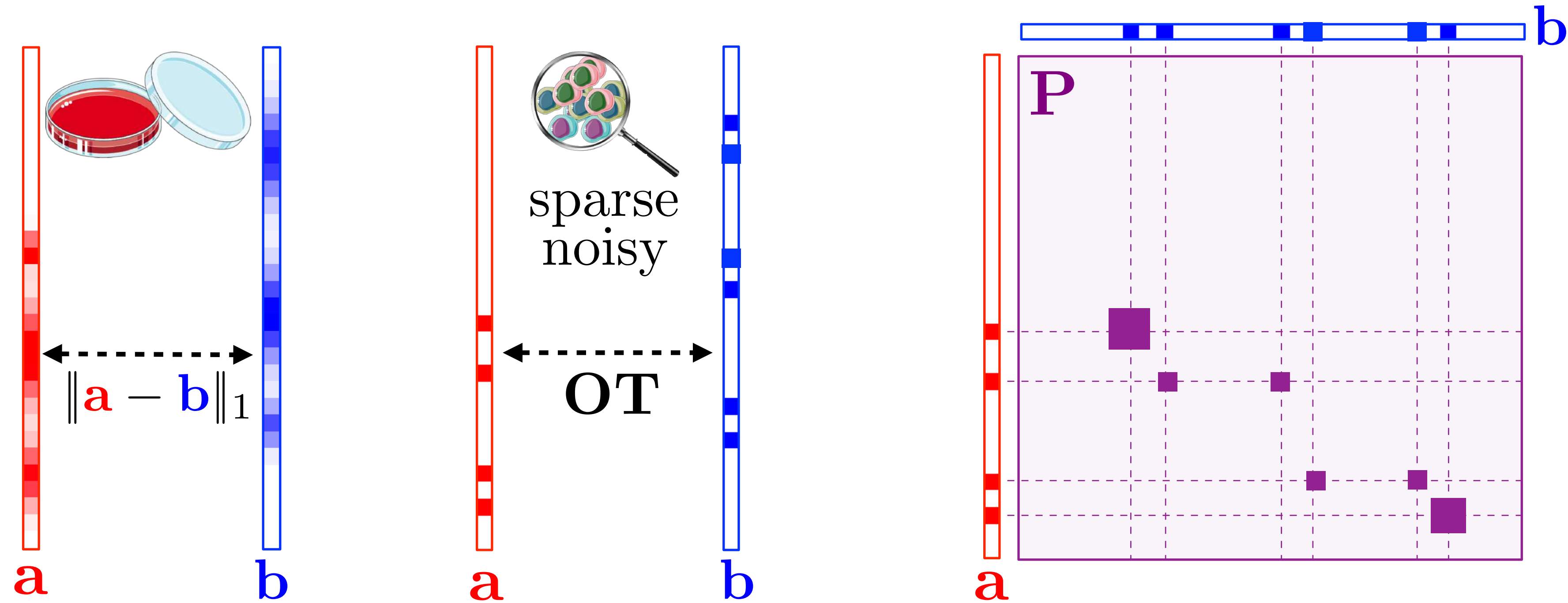
Tracking Cell Evolution with OT



OT for Cell to Cell Dissimilarity

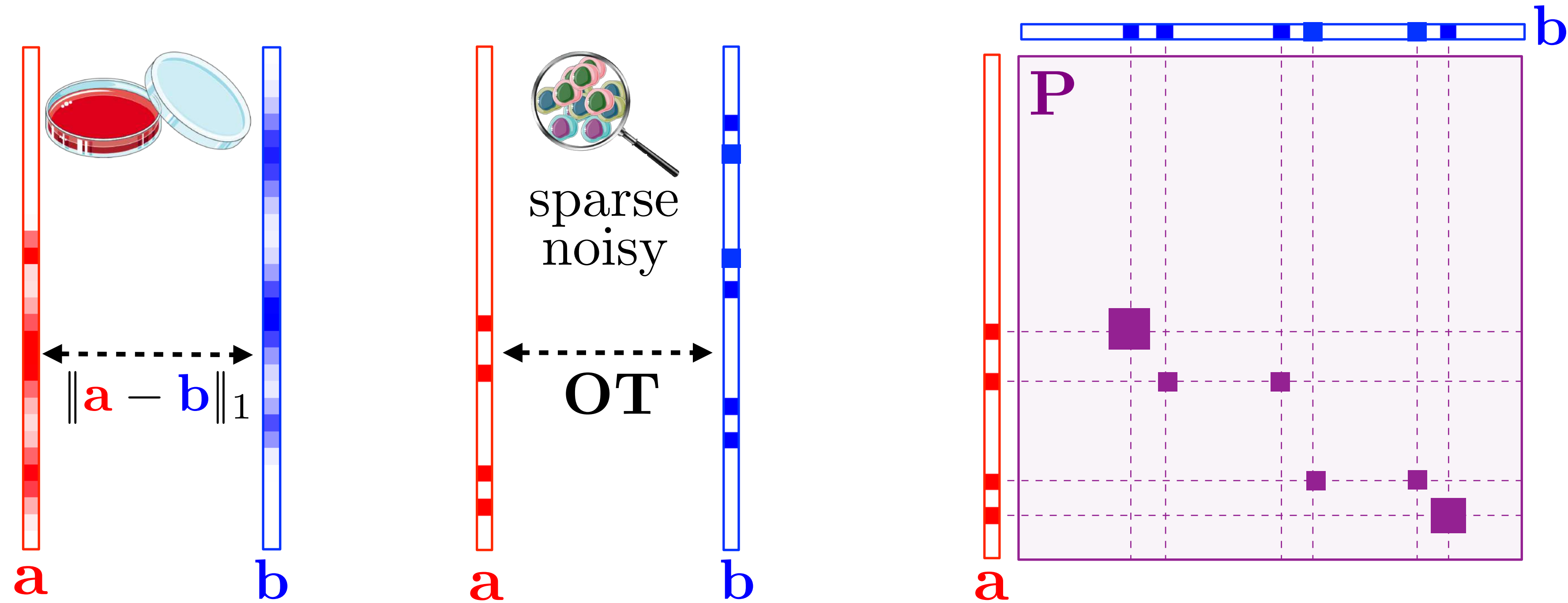


OT for Cell to Cell Dissimilarity



$$\mathbf{OT}(a, b) := \min_{\mathbf{P} \mathbf{1} = a, \mathbf{P}^\top \mathbf{1} = b} \sum_{i,j} d(x_i, x_j) \mathbf{P}_{i,j}$$

OT for Cell to Cell Dissimilarity

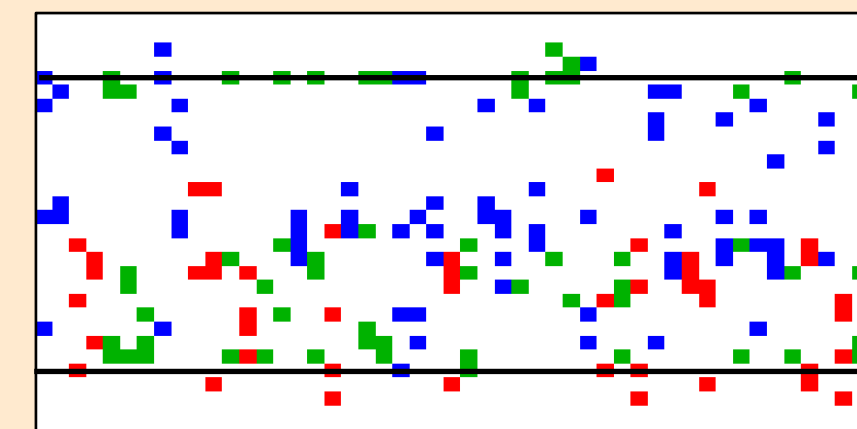


$$\text{OT}(a, b) := \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, x_j) \mathbf{P}_{i,j}$$

Choice of
gene's distance
 $d(x_i, x_j)$

Bio-inspired (e.g. via regulation networks)

Data-driven



$d(x_i, x_j)$

Paired Multi-omics Integration

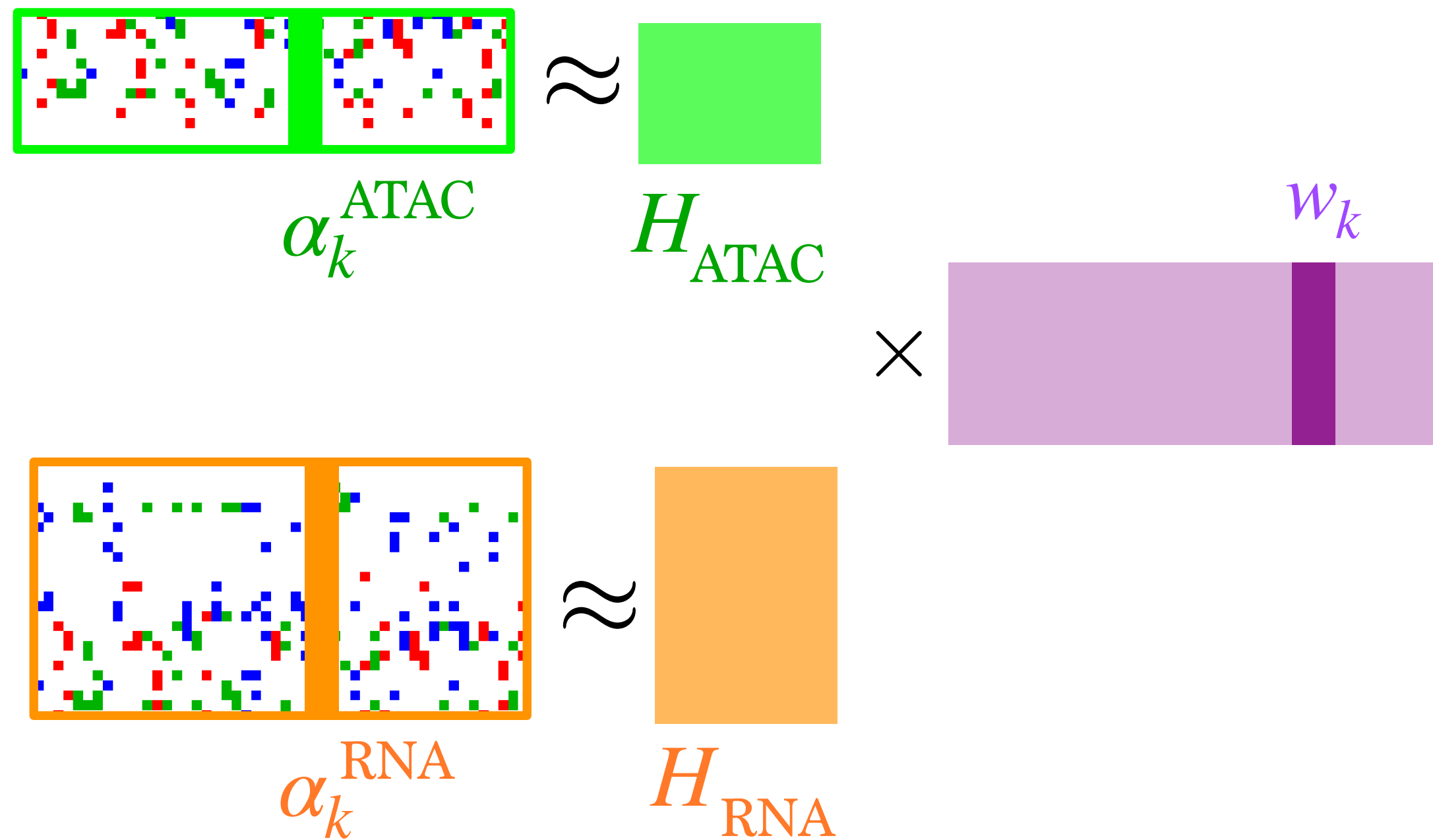
Mowgli

```
> pip install mowgli
```

Geert-Jan
Huizing

Laura
Cantini

$$\min_{(w_k), (H_m)} \left\{ \sum_m \sum_k \text{OT}(H_m w_k, \alpha_k^m) : w_k \geq 0, H_m \geq 0 \right\}$$



Paired Multi-omics Integration

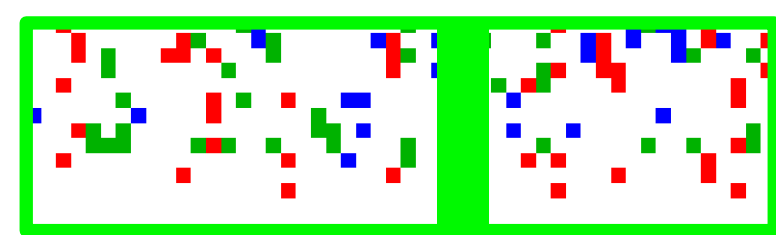
Mowgli

> pip install mowgli

Geert-Jan
Huizing

Laura
Cantini

$$\min_{(w_k), (H_m)} \left\{ \sum_m \sum_k \text{OT}(H_m w_k, \alpha_k^m) : w_k \geq 0, H_m \geq 0 \right\}$$



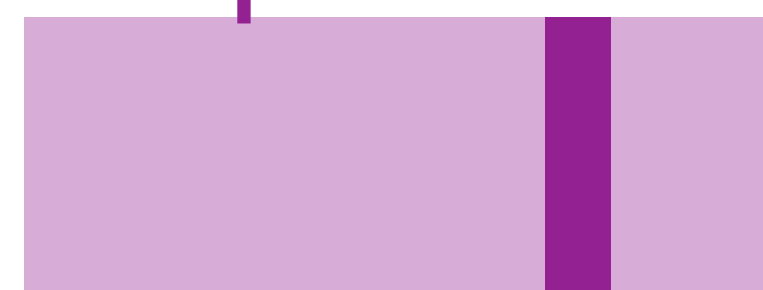
α_k^{ATAC}

\approx

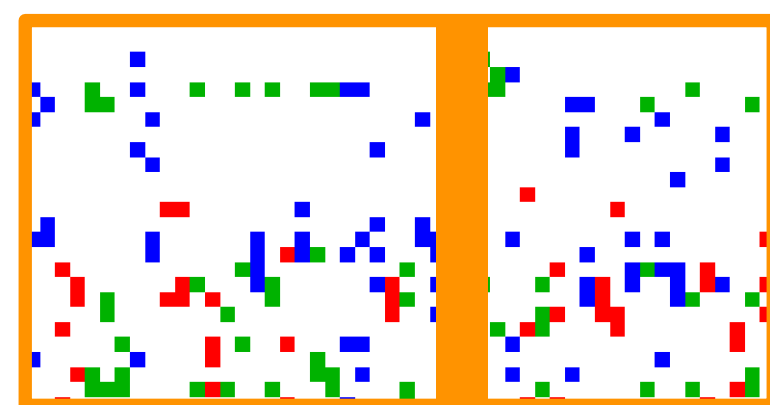


H_{ATAC}

\times



w_k



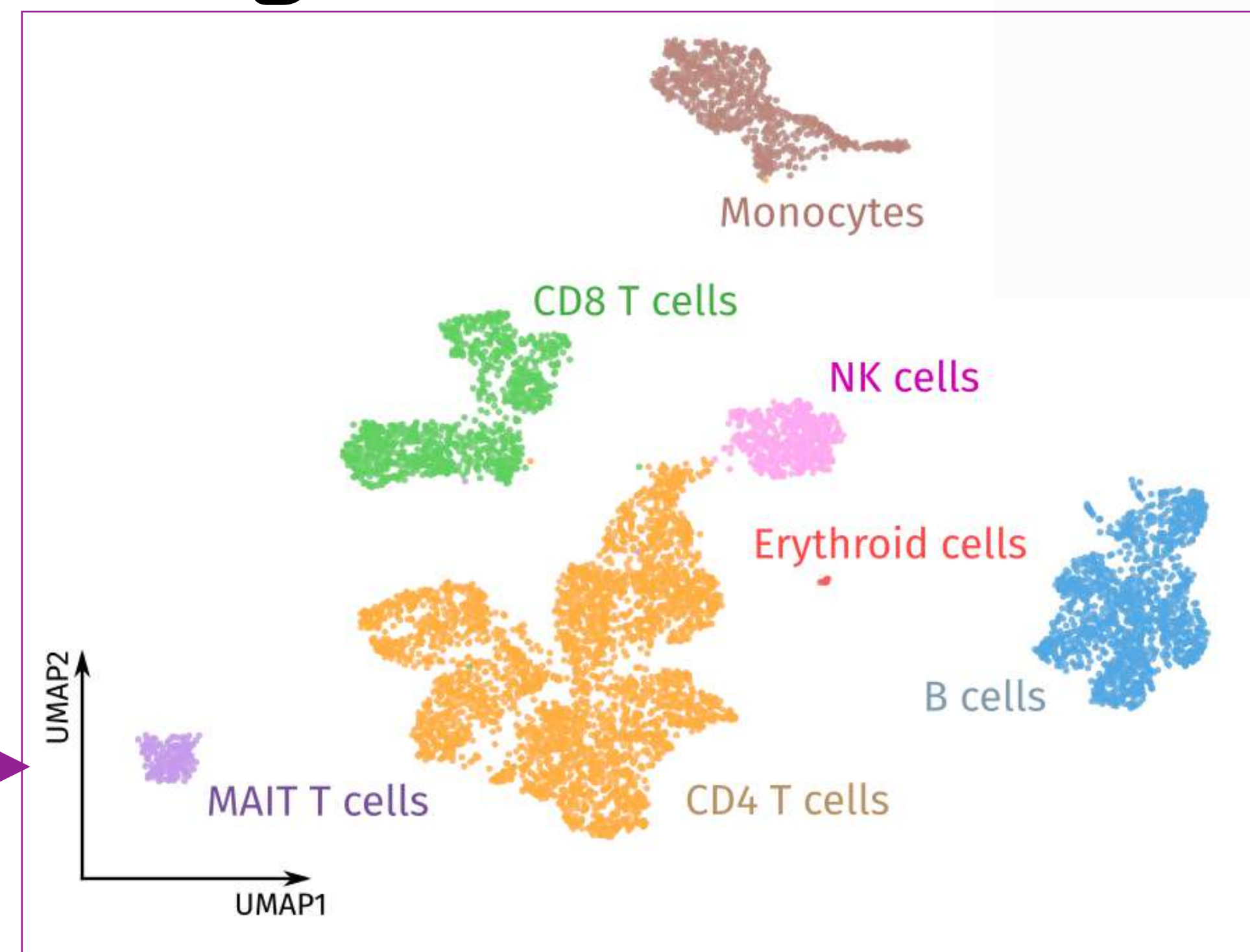
α_k^{RNA}

\approx



H_{RNA}

clustering



Paired Multi-omics Integration

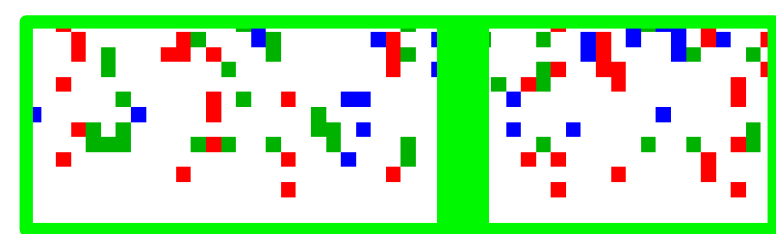
Mowgli

> pip install mowgli

Geert-Jan
Huizing

Laura
Cantini

$$\min_{(w_k), (H_m)} \left\{ \sum_m \sum_k \text{OT}(H_m w_k, \alpha_k^m) : w_k \geq 0, H_m \geq 0 \right\}$$



α_k^{ATAC}

\approx

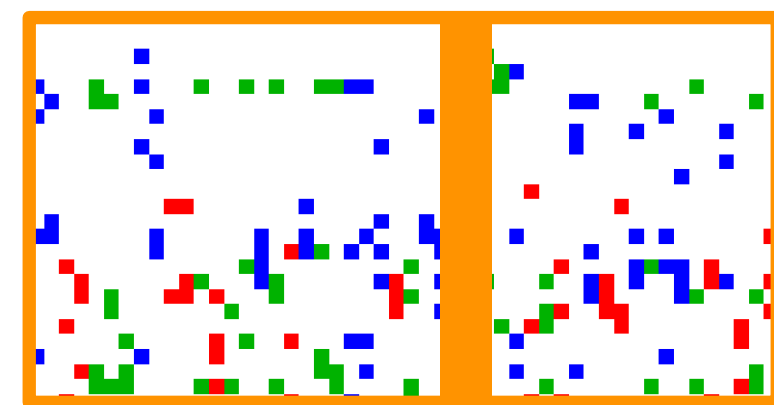
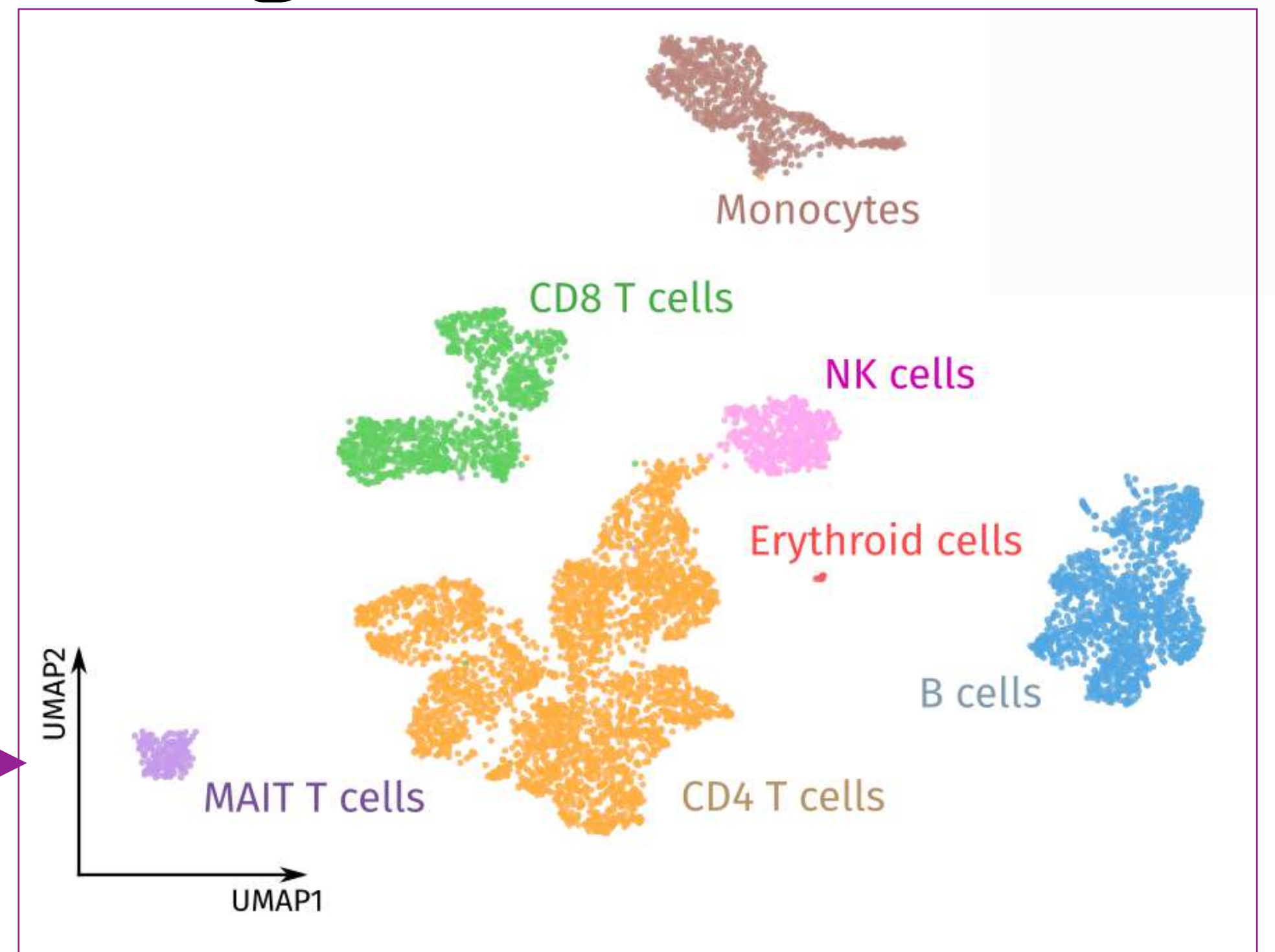


H_{ATAC}

\times

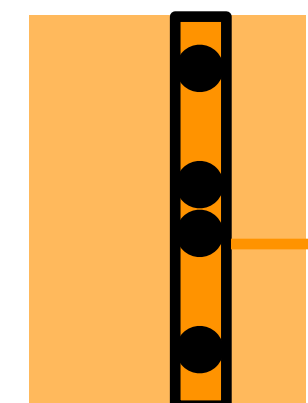


clustering



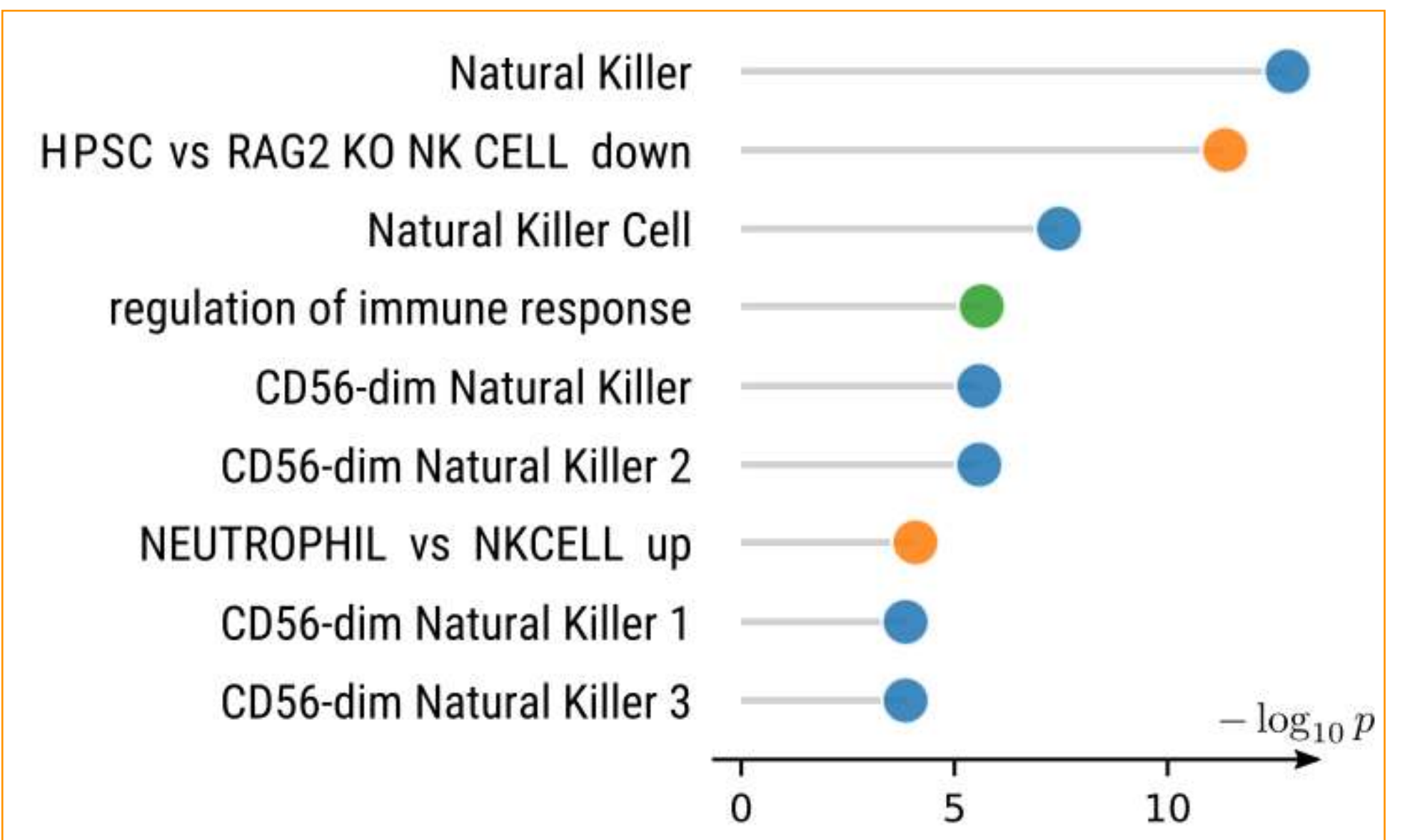
α_k^{RNA}

\approx



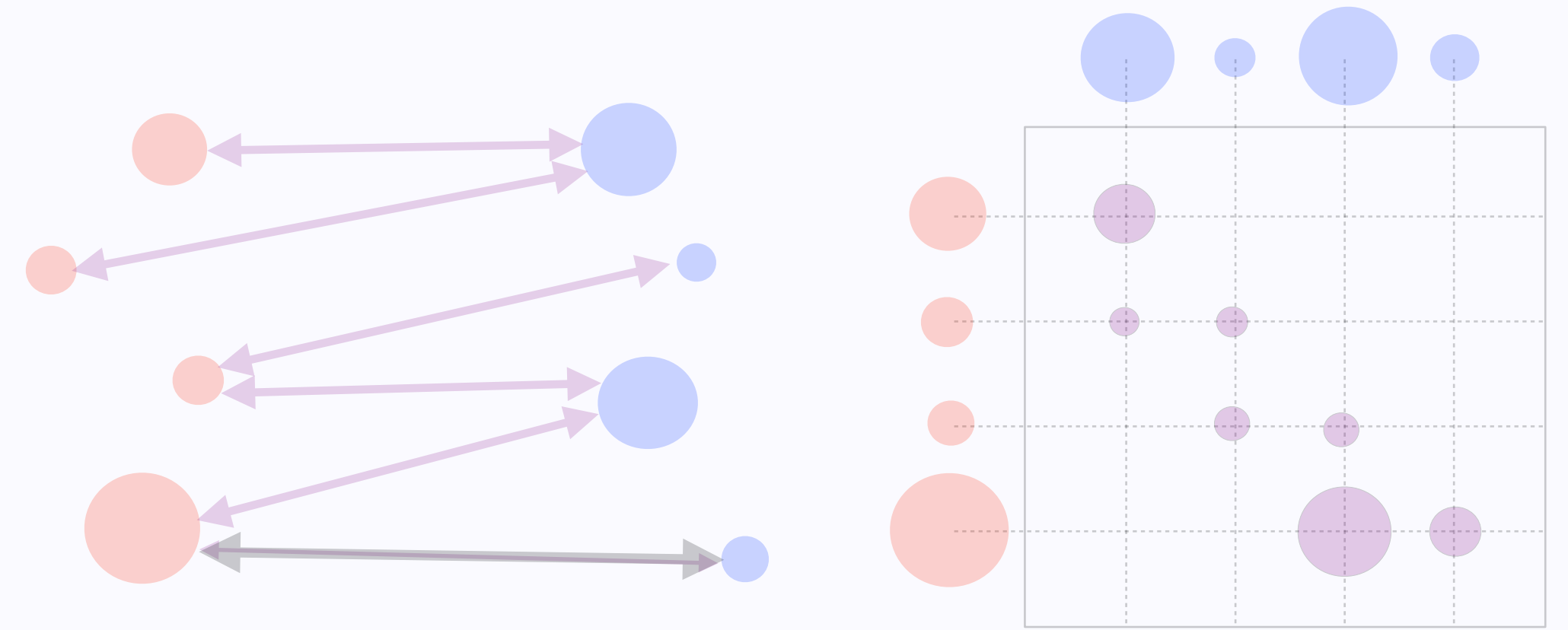
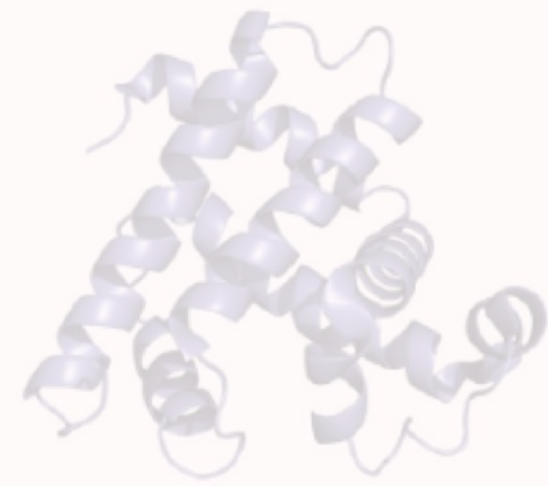
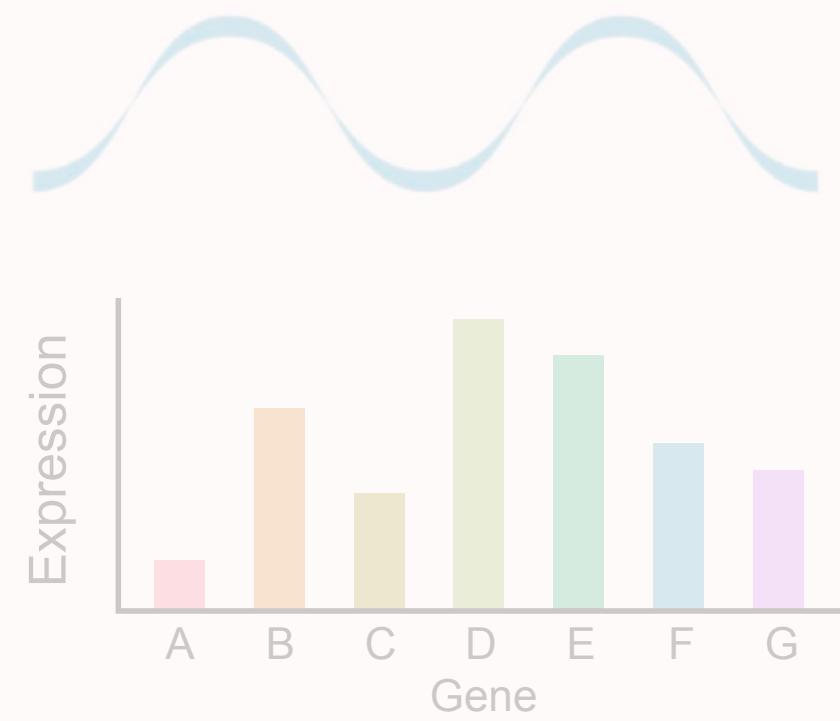
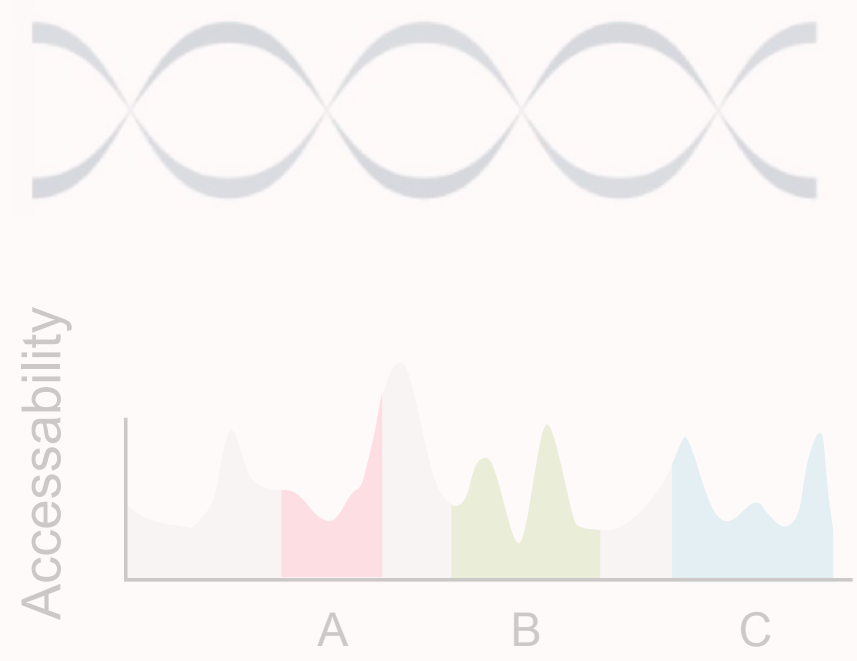
H_{RNA}

Gene set
enrichment



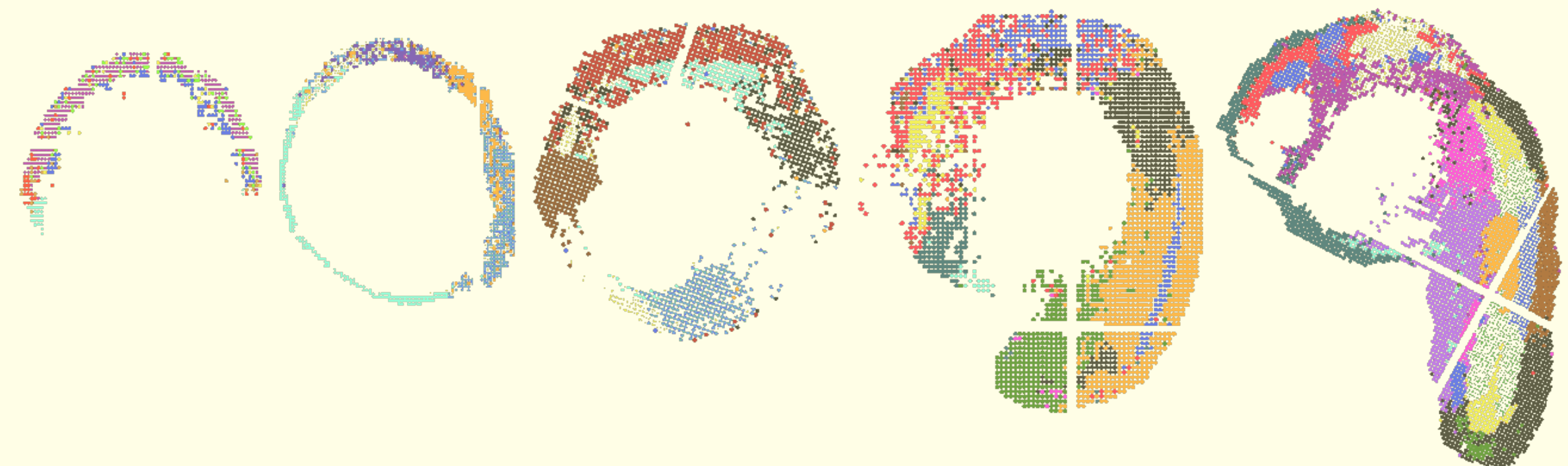
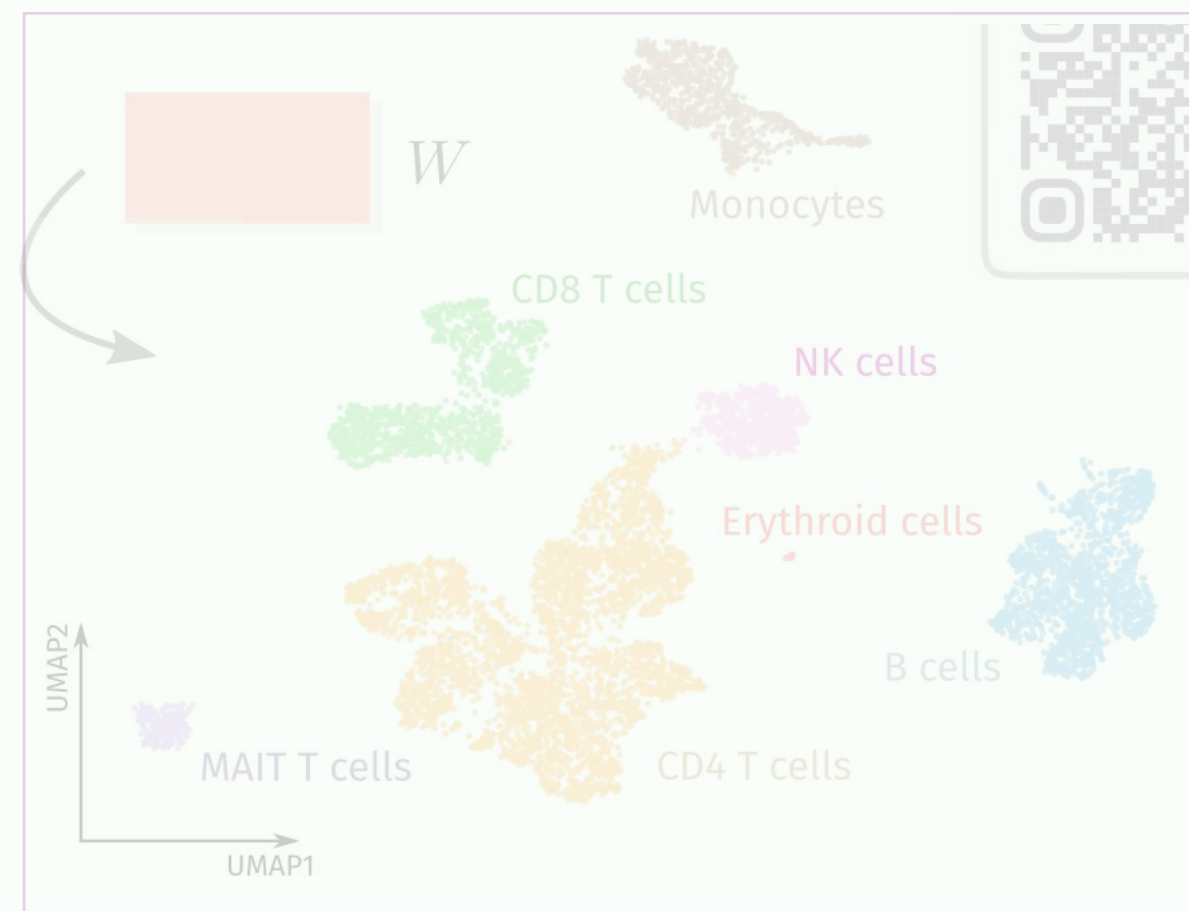
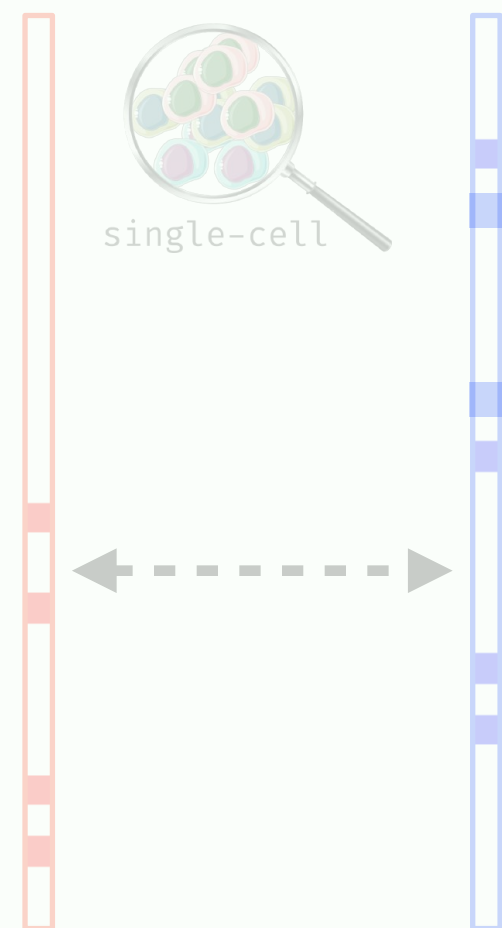
Single cell genomics

Optimal Transport



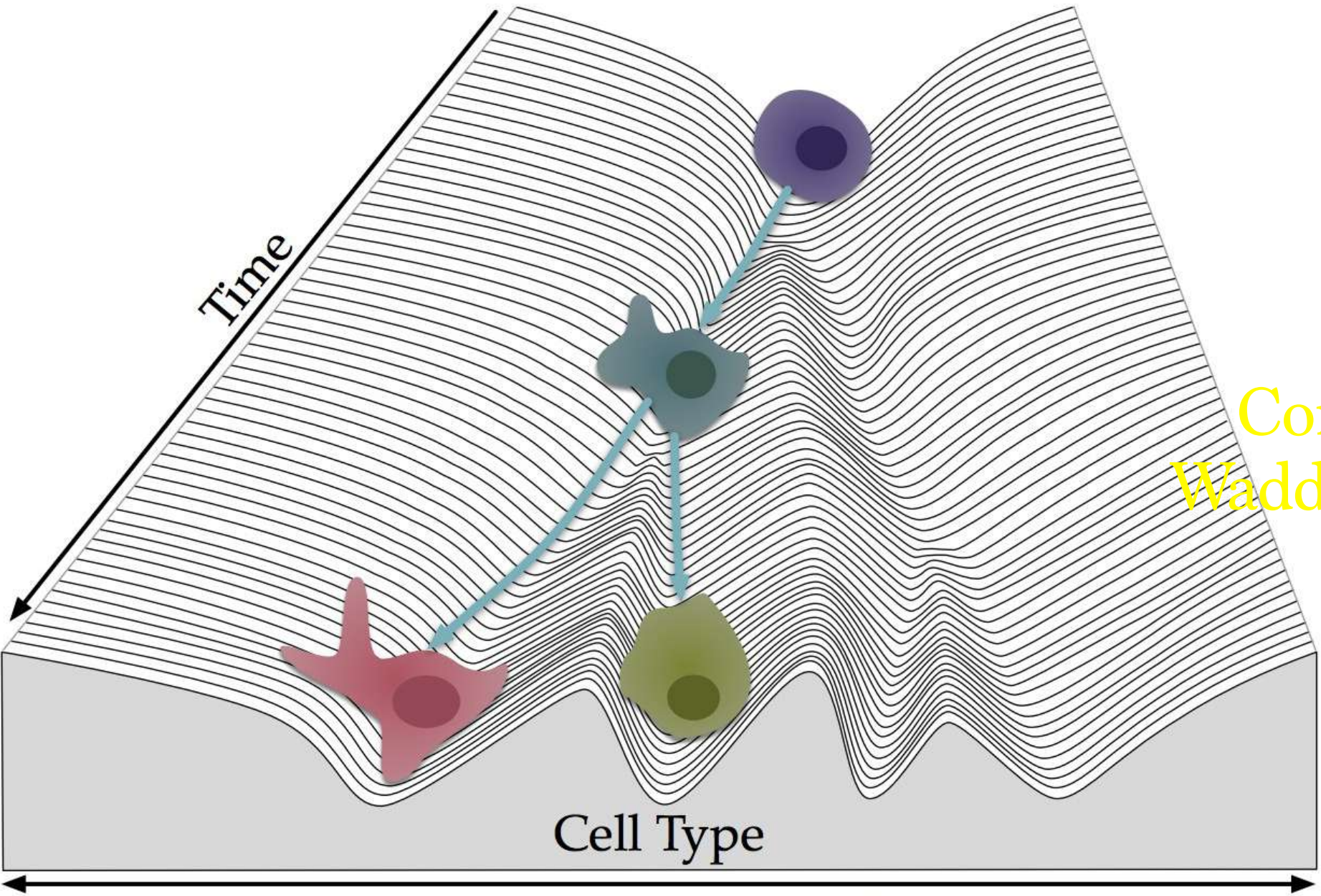
Comparing Cell-Cell with OT

Tracking Cell Evolution with OT

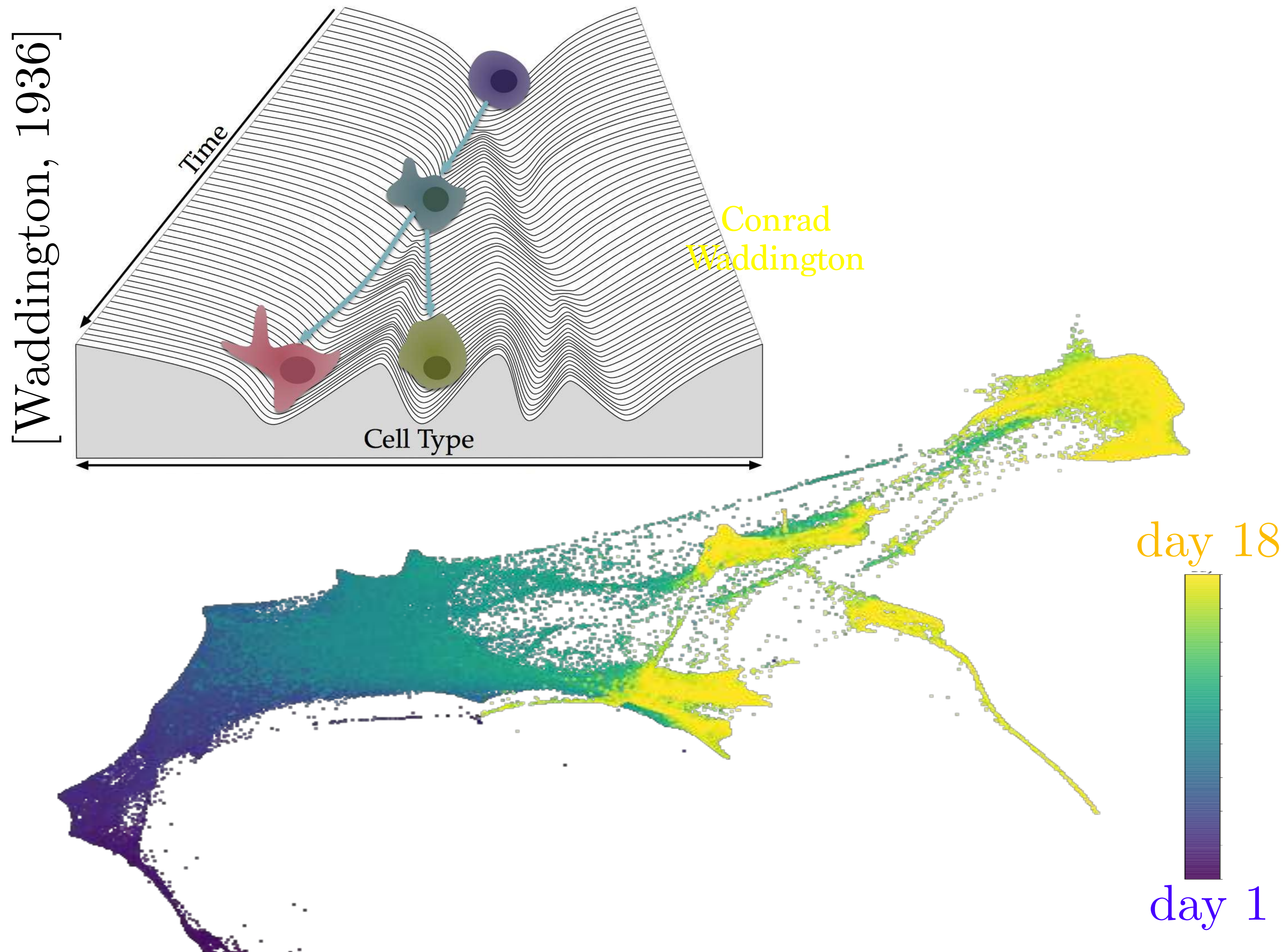


Optimal Transport Across Cells

[Waddington, 1936]



Optimal Transport Across Cells

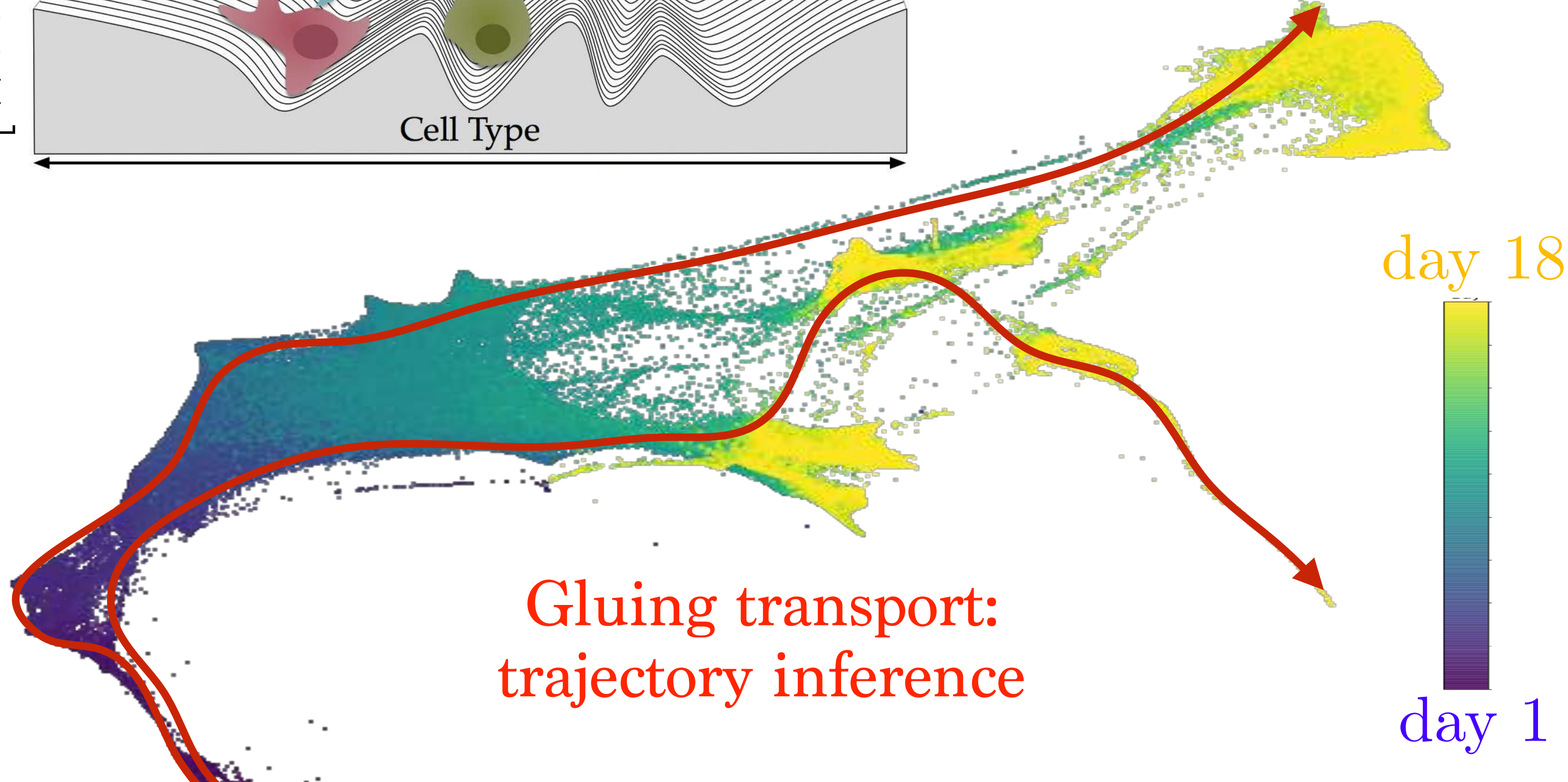
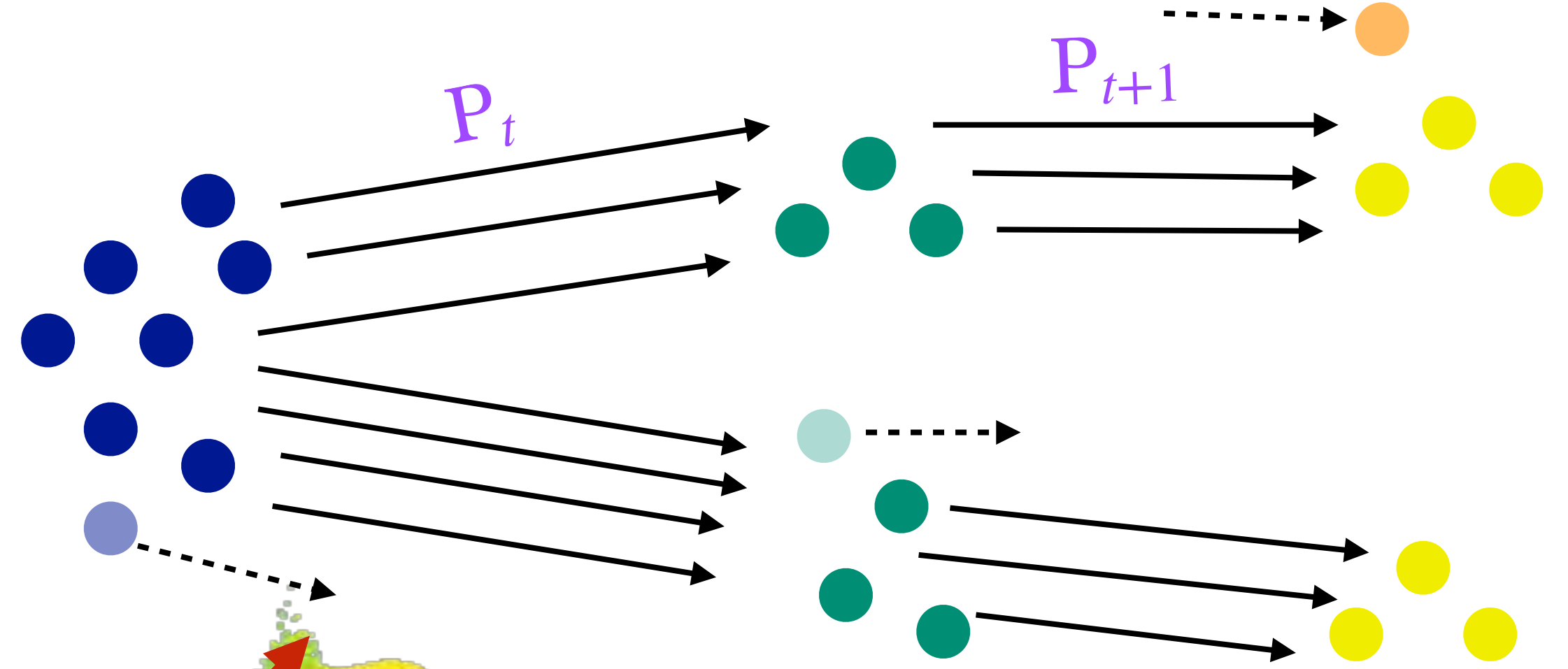
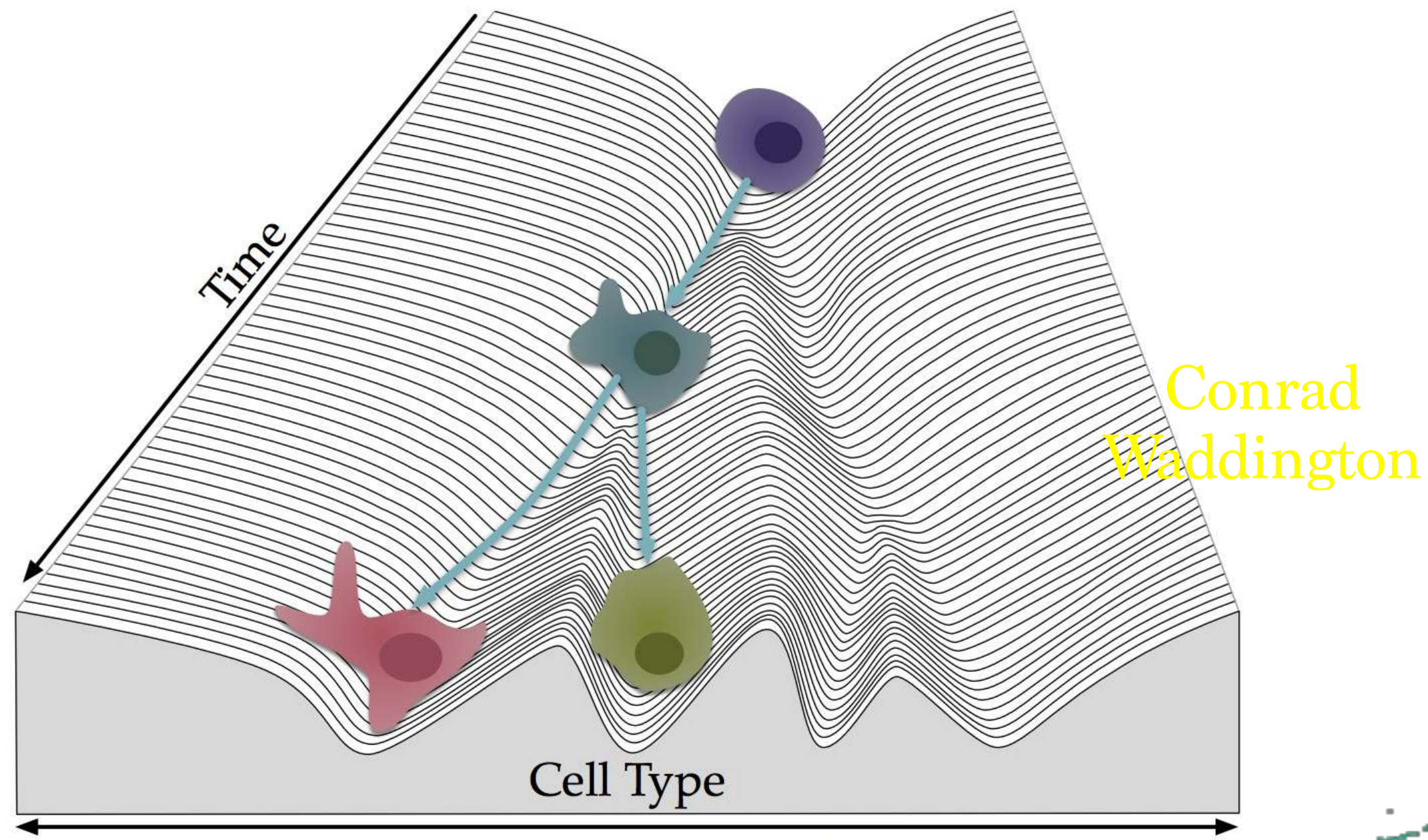


Geoffrey Schiebinger

[Schiebinger et al 2019]

Optimal Transport Across Cells

[Waddington, 1936]



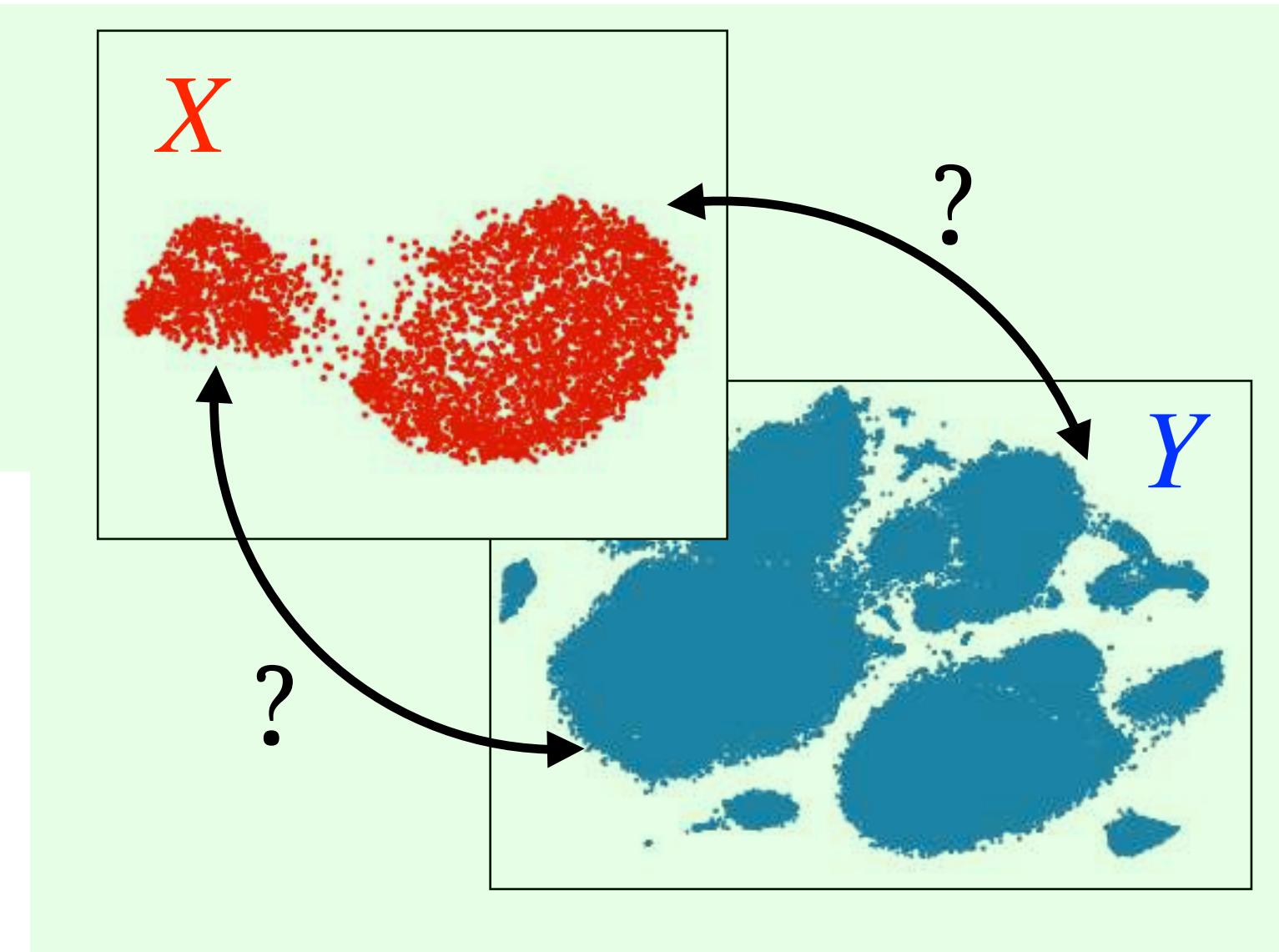
Geoffrey Schiebinger

[Schiebinger et al 2019]

Gromov-Wasserstein

Data in two different spaces X and Y .

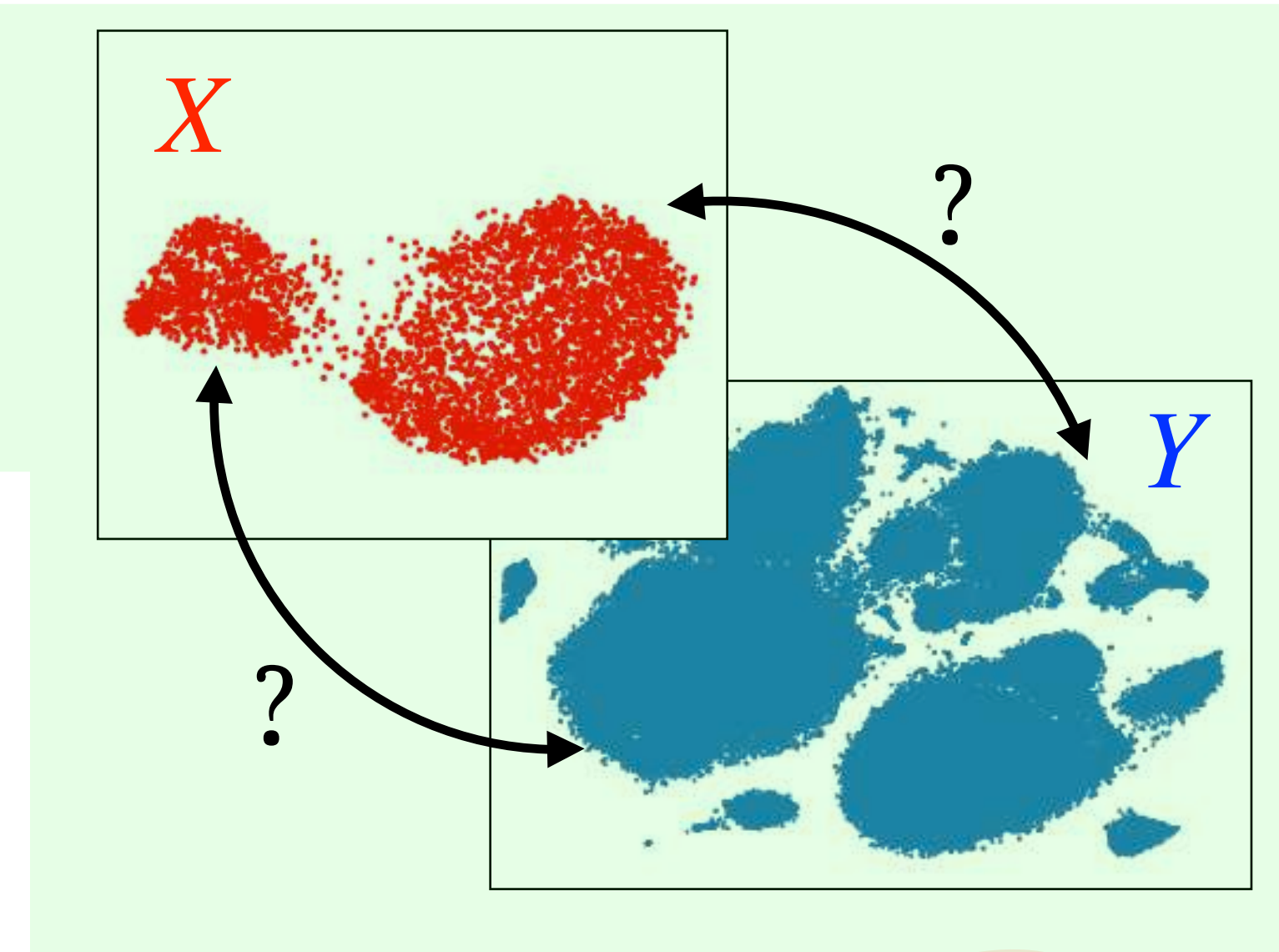
Inner distances d_X and d_Y



Gromov-Wasserstein

Data in two different spaces X and Y .

Inner distances d_X and d_Y

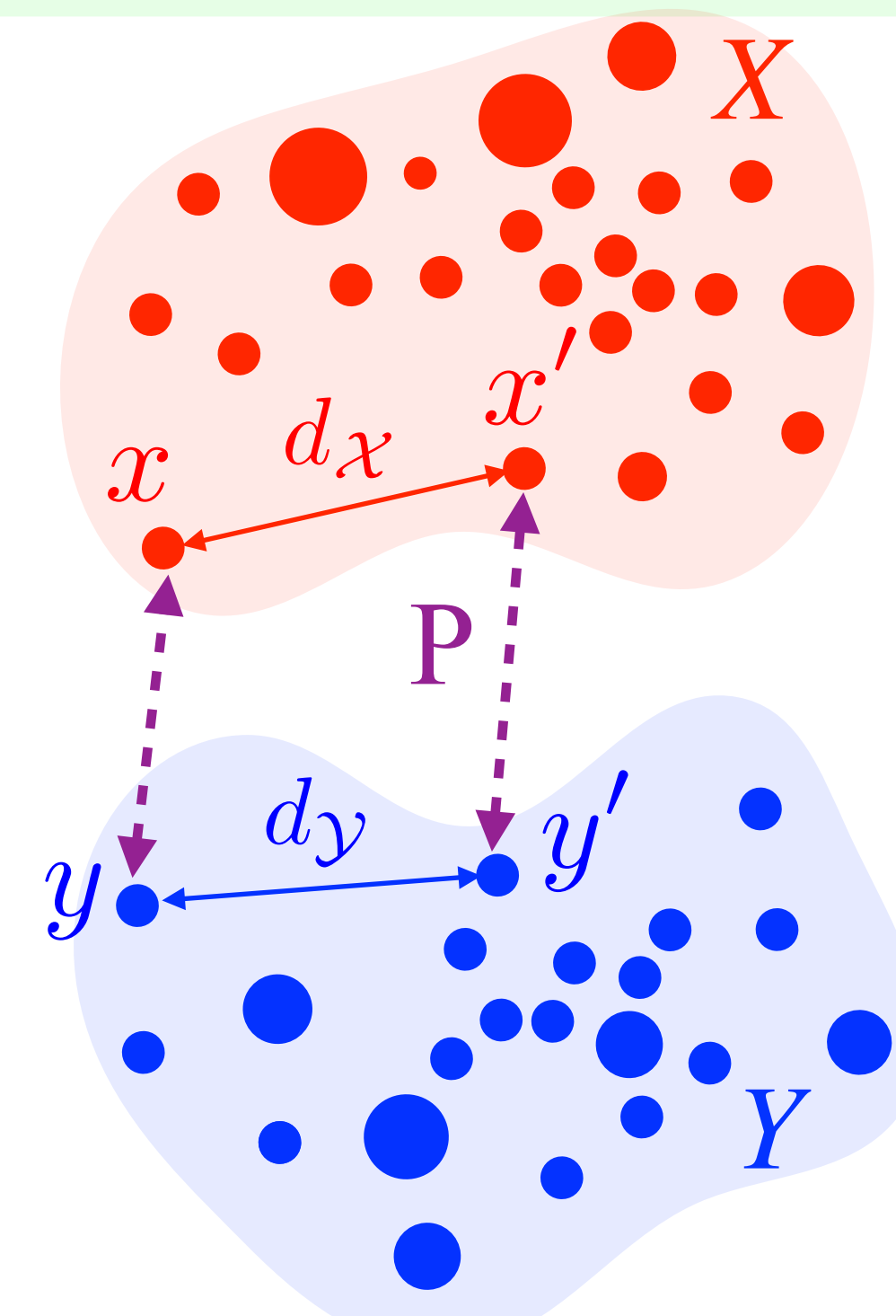


Gromov-Wasserstein problem:

$$\min_{\mathbf{P}} \sum_{i,j,i',j'} |d_X(x_i, x_{i'}) - d_Y(y_j, y_{j'})|^2 \mathbf{P}_{i,j} \mathbf{P}_{i',j'}$$

[Memoli 2011][Sturm 2011]

→ non-convex, NP-hard ...



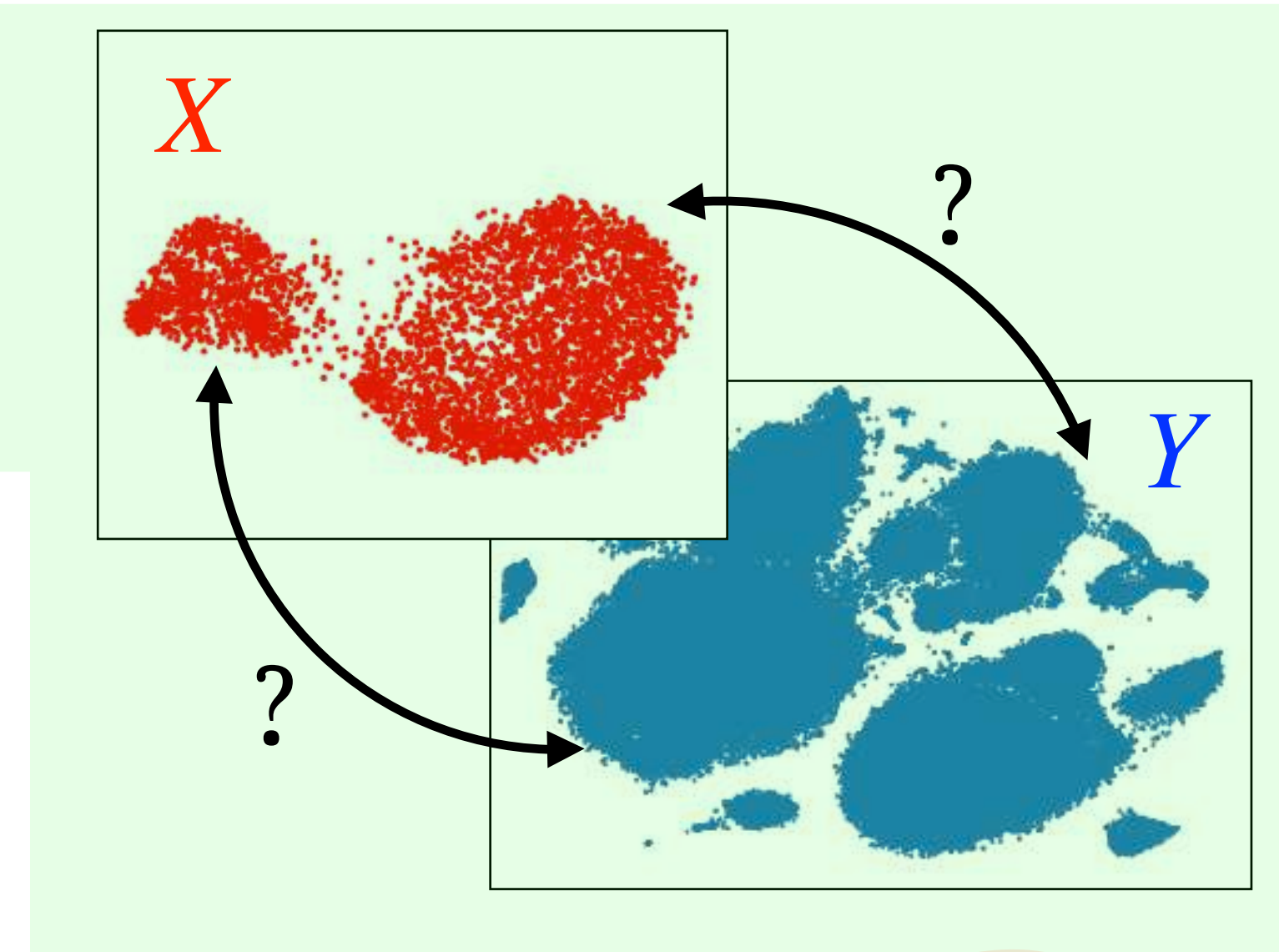
Facundo
Memoli

Karl-Theodor
Sturm

Gromov-Wasserstein

Data in two different spaces X and Y .

Inner distances d_X and d_Y



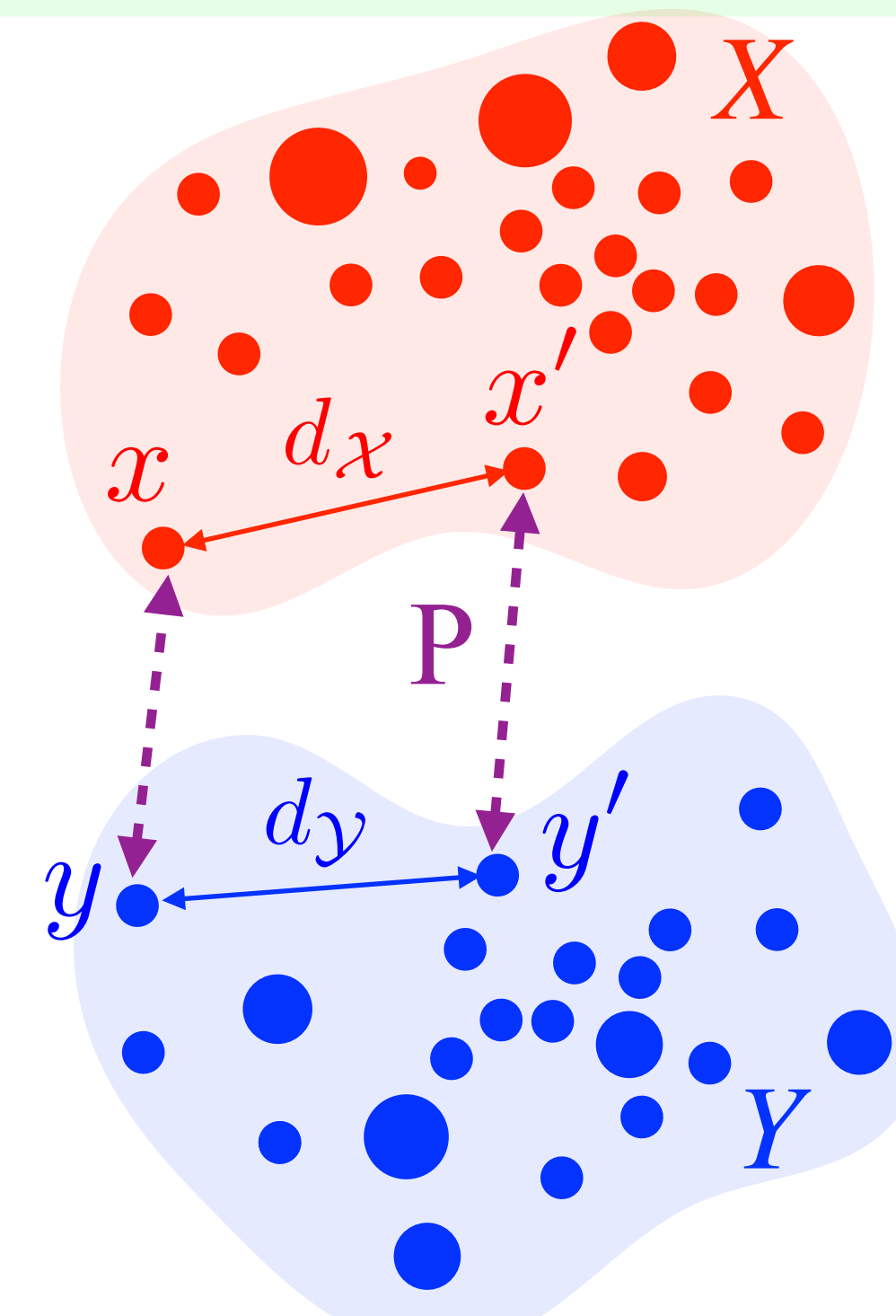
Gromov-Wasserstein problem:

$$\min_{\mathbf{P}} \sum_{i,j,i',j'} |d_X(x_i, x_{i'}) - d_Y(y_j, y_{j'})|^2 \mathbf{P}_{i,j} \mathbf{P}_{i',j'}$$

[Memoli 2011][Sturm 2011]

→ non-convex, NP-hard ...

Facundo
Memoli



Karl-Theodor
Sturm

Theorem: GW is a distance up to isometries.

Multi-Omics Integration

SCOT: single-cell alignment with optimal transport
[Demetci et al 2022]

Pinar
Demetci

RNA-seq

ATAC-seq

Gromov-Wasserstein

α

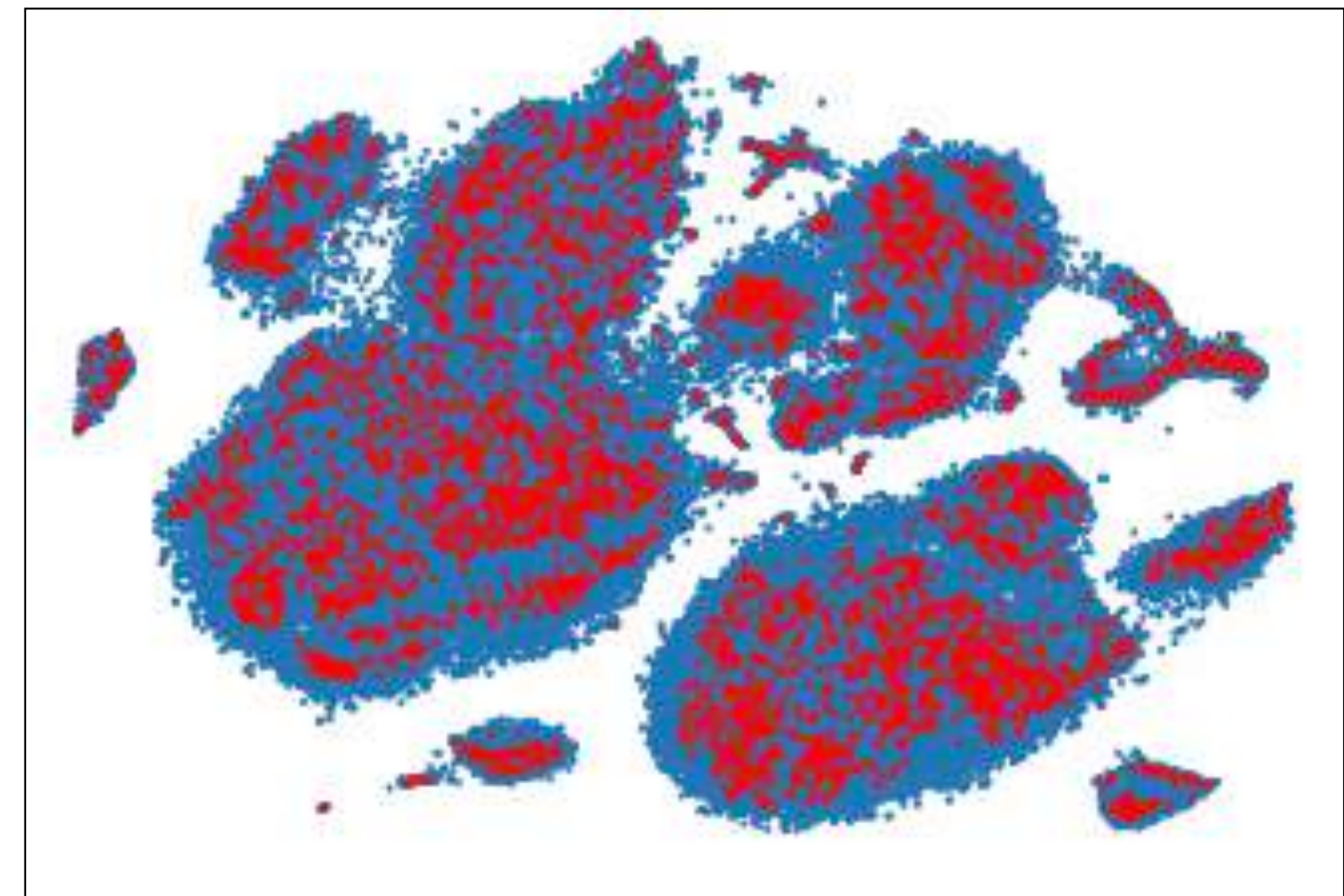
β

X

Y

$\dim \sim 10^3$ (genes)

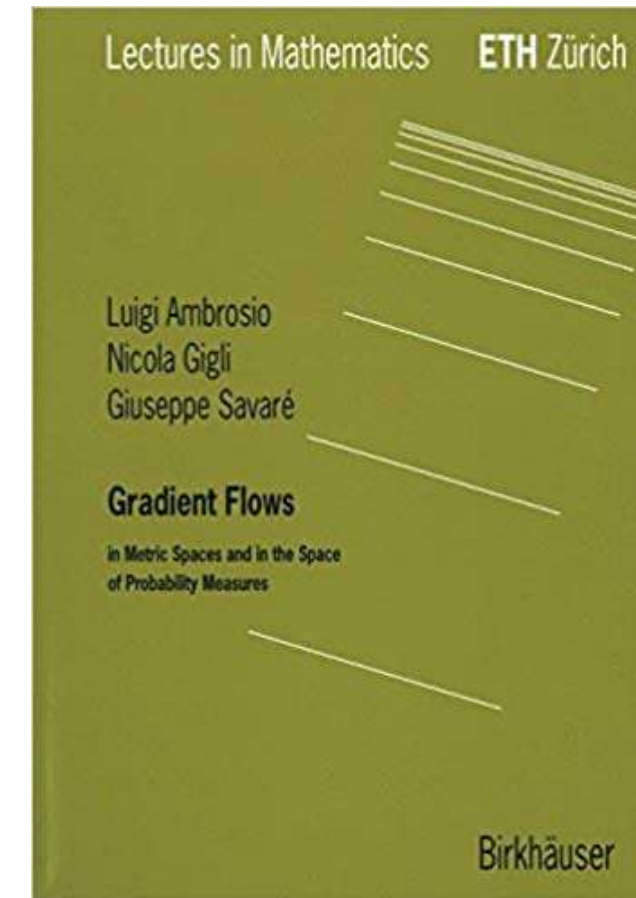
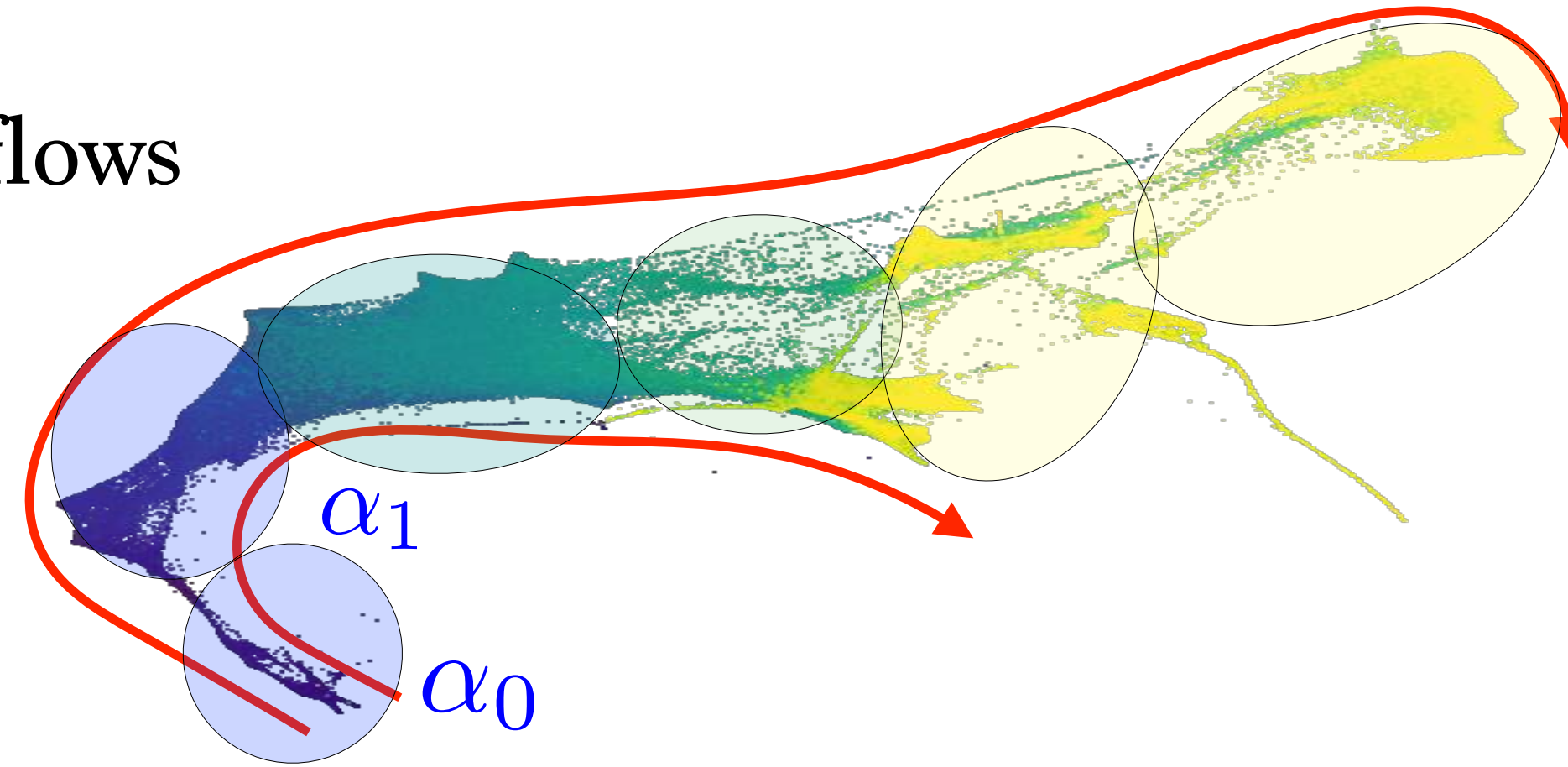
$\dim \sim 10^2$ (peaks)



Conclusion: gradient flows for genomics

Better modeling the dynamics: use gradient flows

$$\alpha_{t+1} = \operatorname{argmin}_{\beta} \operatorname{OT}(\alpha_t, \beta) + \tau f(\beta)$$



Luigi
Ambrosio

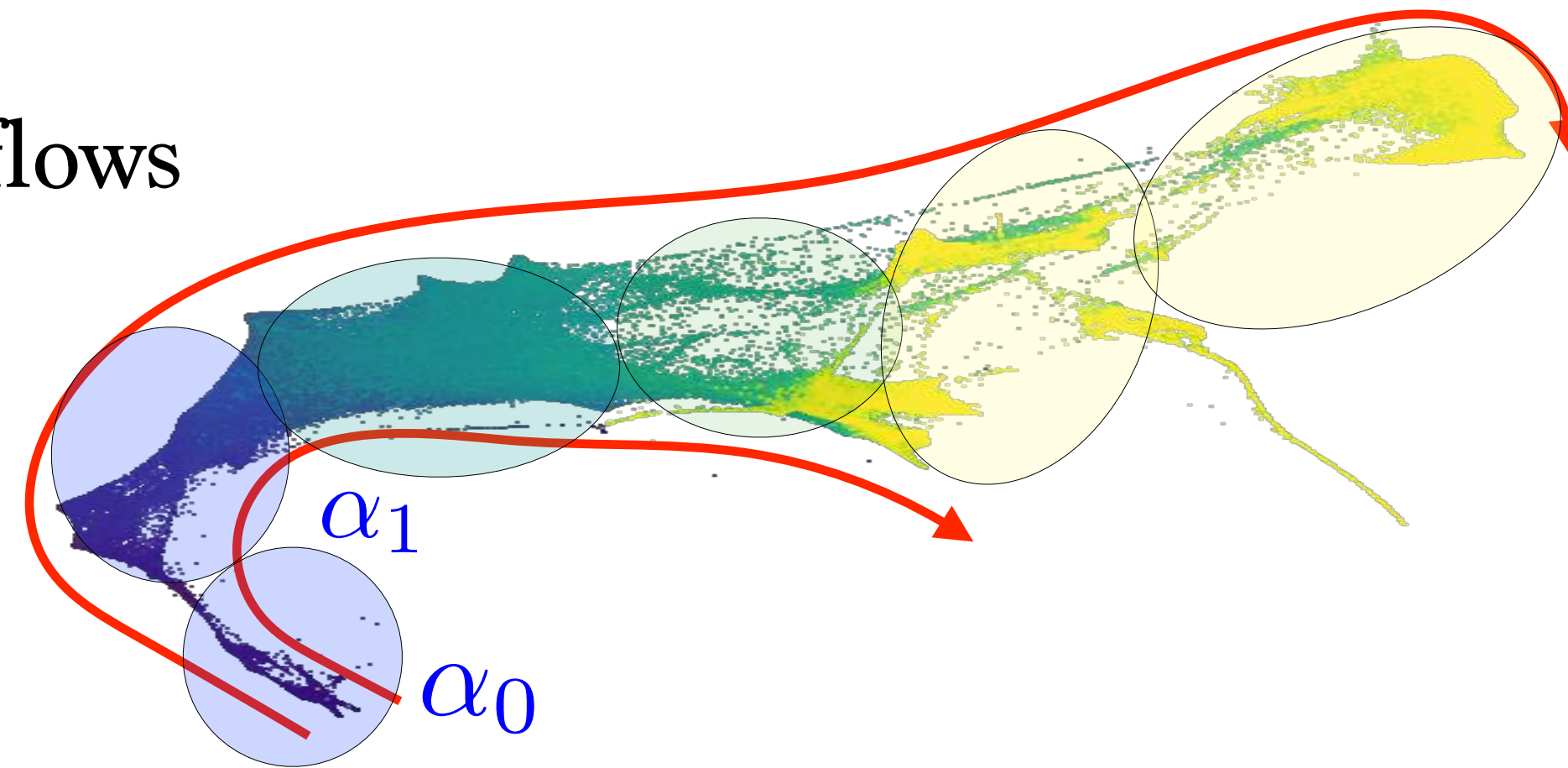
Giuseppe
Savare

Nicola
Gigli

Conclusion: gradient flows for genomics

Better modeling the dynamics: use gradient flows

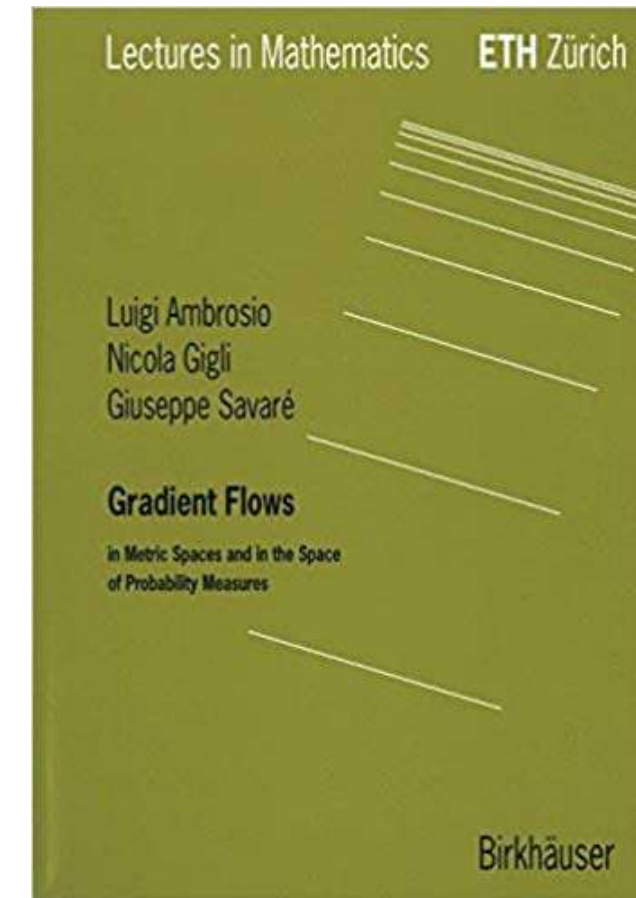
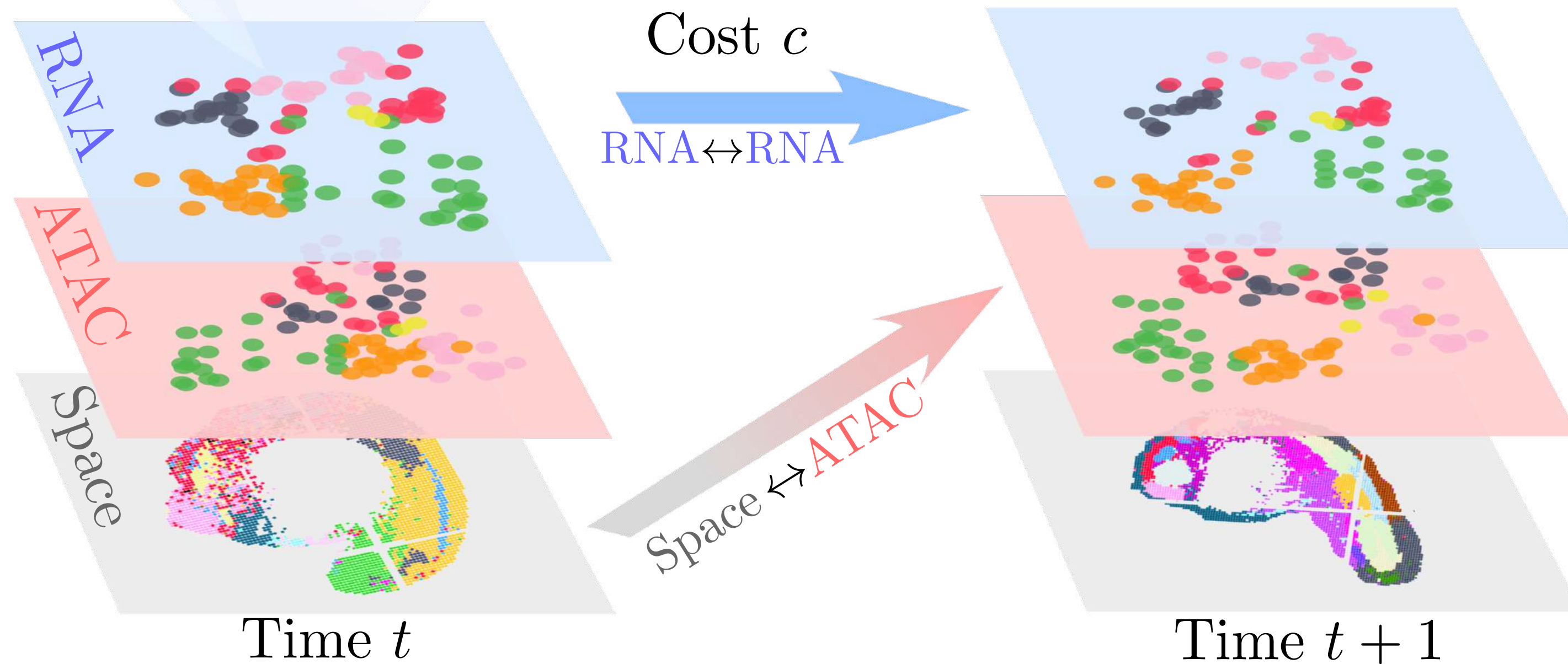
$$\alpha_{t+1} = \operatorname{argmin}_{\beta} \operatorname{OT}(\alpha_t, \beta) + \tau f(\beta)$$



Potential f



Key issues: learning the potential $f(\beta)$
Integrate several omics



Luigi
Ambrosio

Giuseppe
Savare

Nicola
Gigli