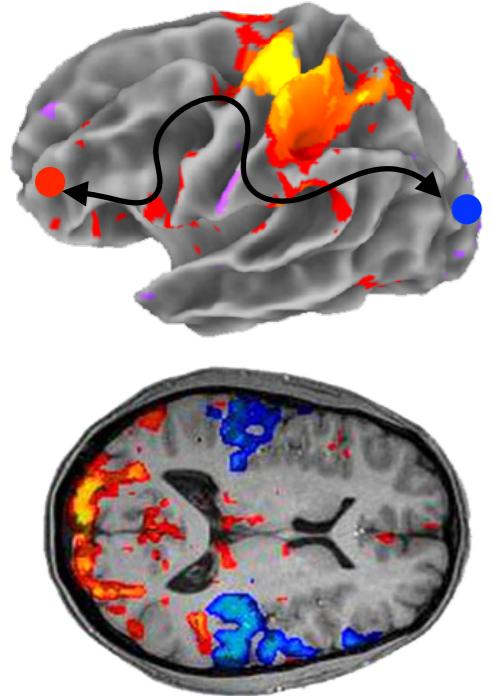


Optimal Transport Strikes Twice in Genomics

Gabriel Peyré



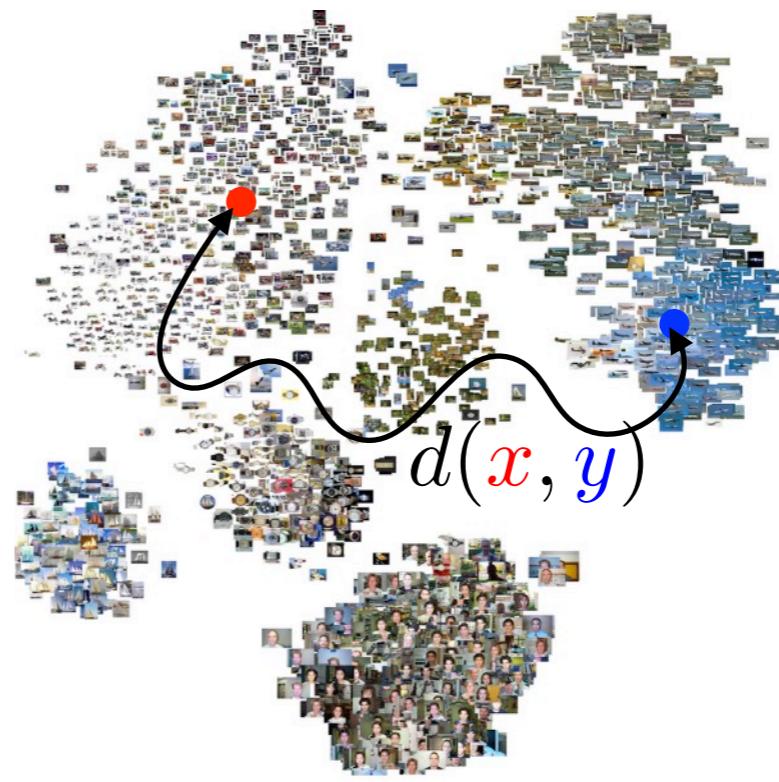
Comparing Distributions for Learning



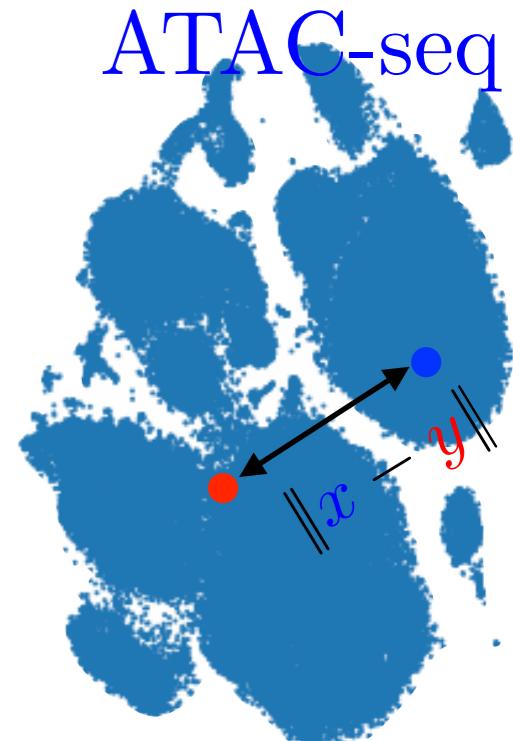
Imaging



NLP

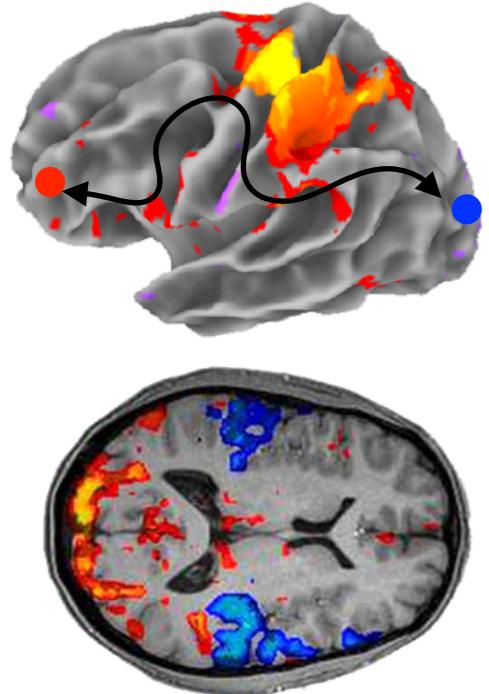


Vision



Genomics

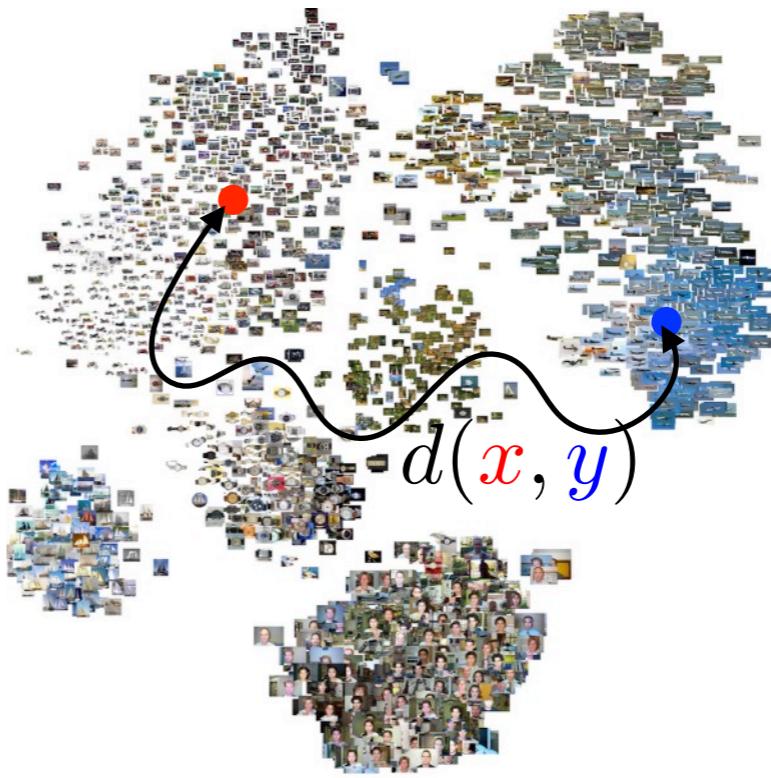
Comparing Distributions for Learning



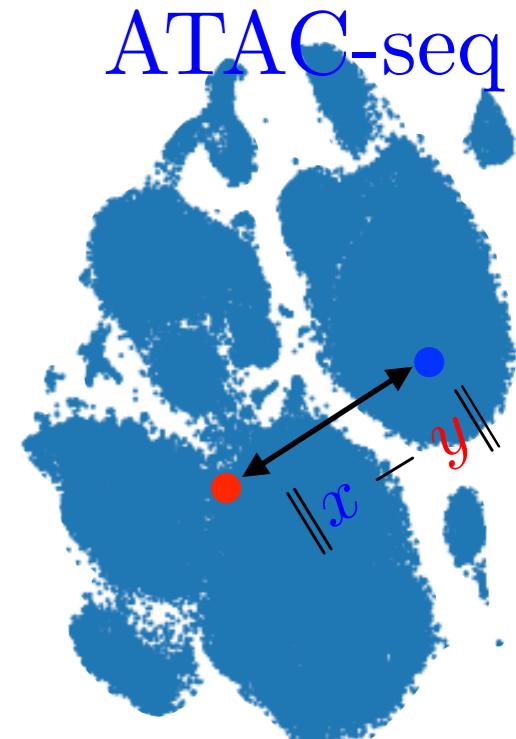
Imaging



NLP



Vision

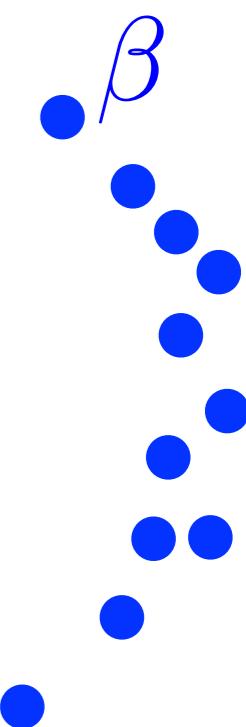
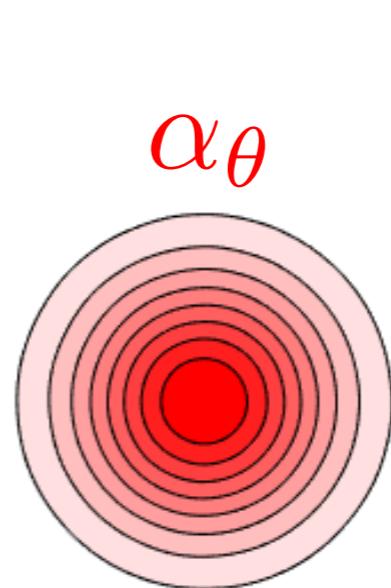


Genomics

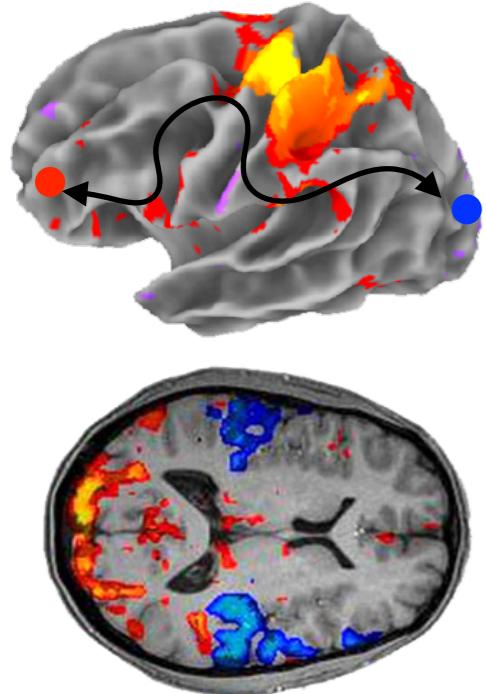
Unsupervised learning

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$



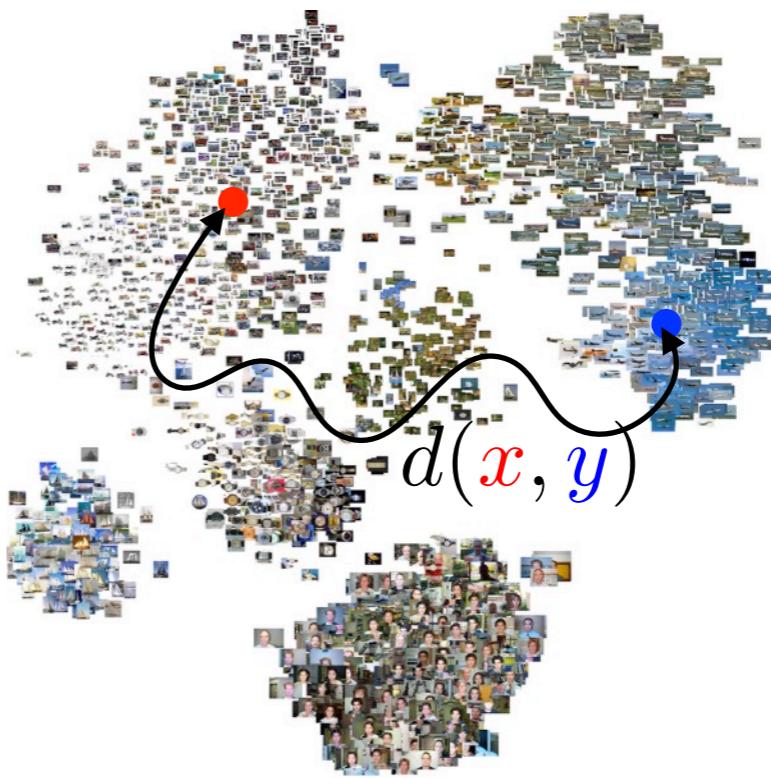
Comparing Distributions for Learning



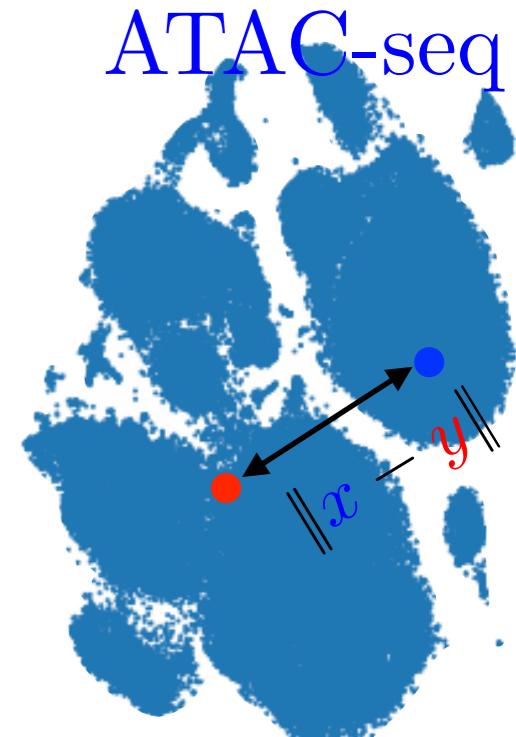
Imaging



NLP



Vision



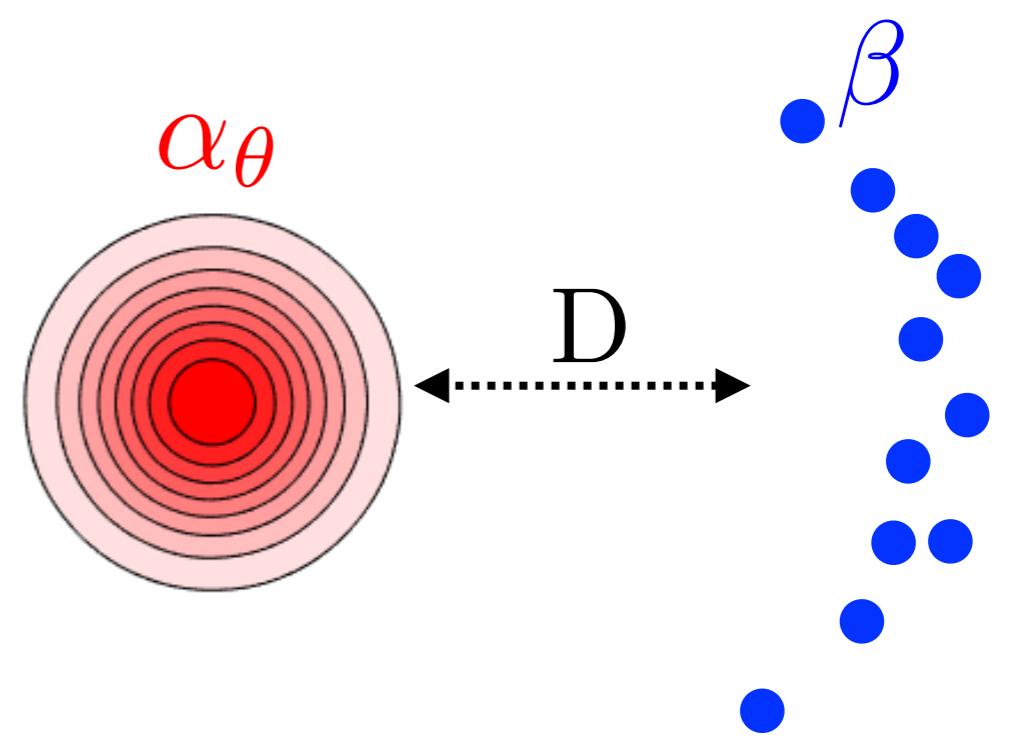
Genomics

Unsupervised learning

Observations: $\beta \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$

Parametric model: $\theta \mapsto \alpha_\theta$

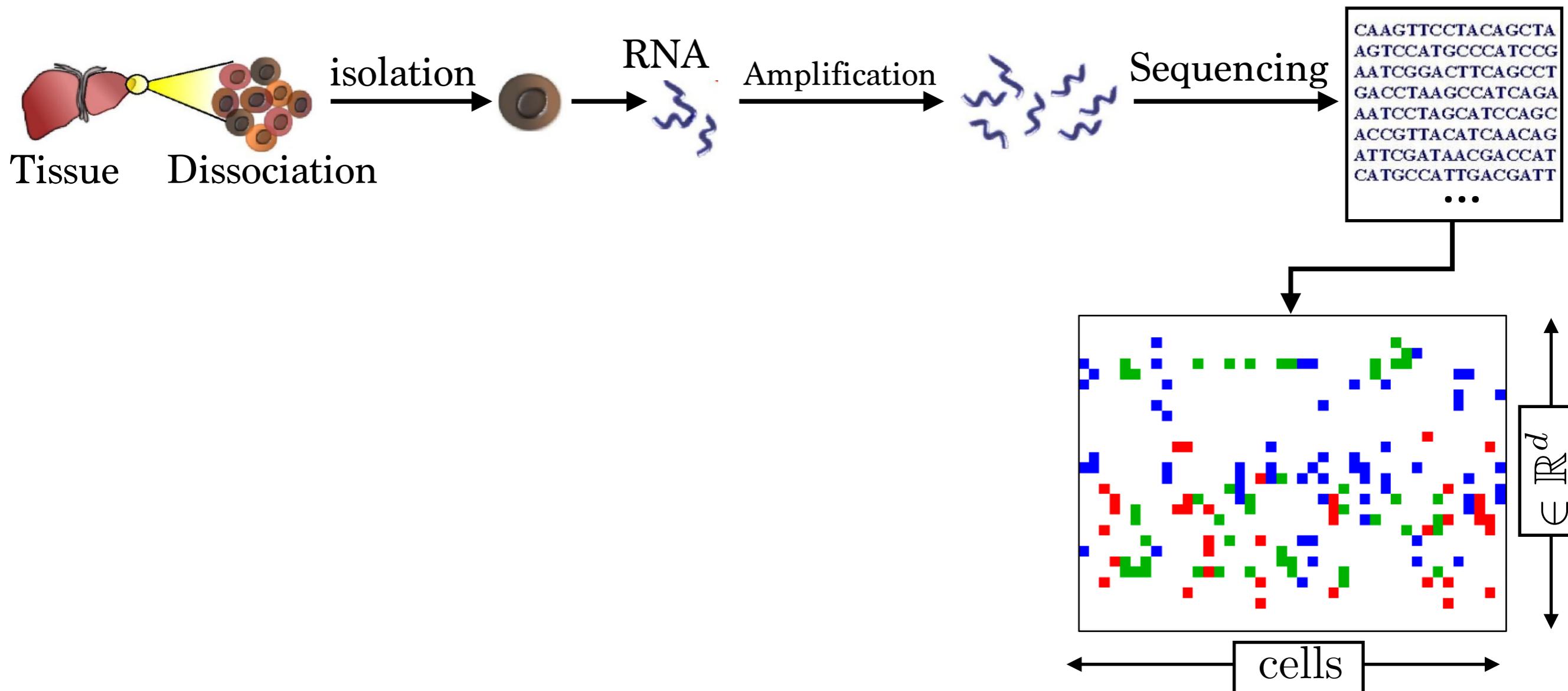
Density fitting: $\min_{\theta} D(\alpha_\theta, \beta)$
 → takes into account a metric d .



Single Cell Multi-omics

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

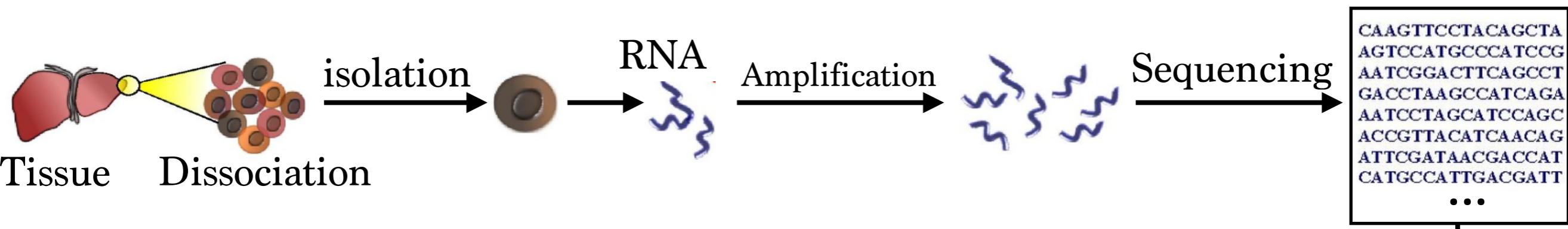
Applications: cancer mutations, dynamic of adaptation, development, ...



Single Cell Multi-omics

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

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



RNA: expression of the DNA. $d \sim 10^4$

ATAC: chromatin accessibility $d \sim 10^5$
→ *geometry of DNA*

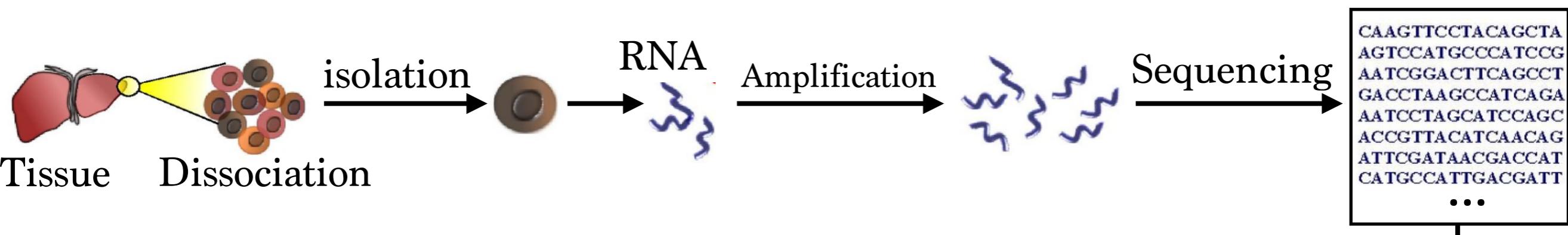
Methylation: presence of methyl groups
→ *cause of differential expression* $d \sim 10^5$

Proteome: presence of proteins $d \sim 10^2$
→ *result of expression*

Single Cell Multi-omics

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

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



RNA: expression of the DNA. $d \sim 10^4$

ATAC: chromatin accessibility $d \sim 10^5$
→ *geometry of DNA*

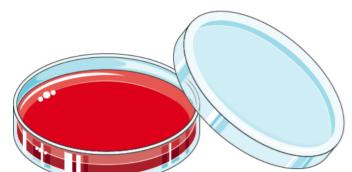
Methylation: presence of methyl groups
→ *cause of differential expression* $d \sim 10^5$

Proteome: presence of proteins $d \sim 10^2$
→ *result of expression*

Paired multi-omics:
→ 10X Multiome (RNA + ATAC)
→ CITE-seq (RNA + proteomes)

Un-paired multi-omics: next frontier ...

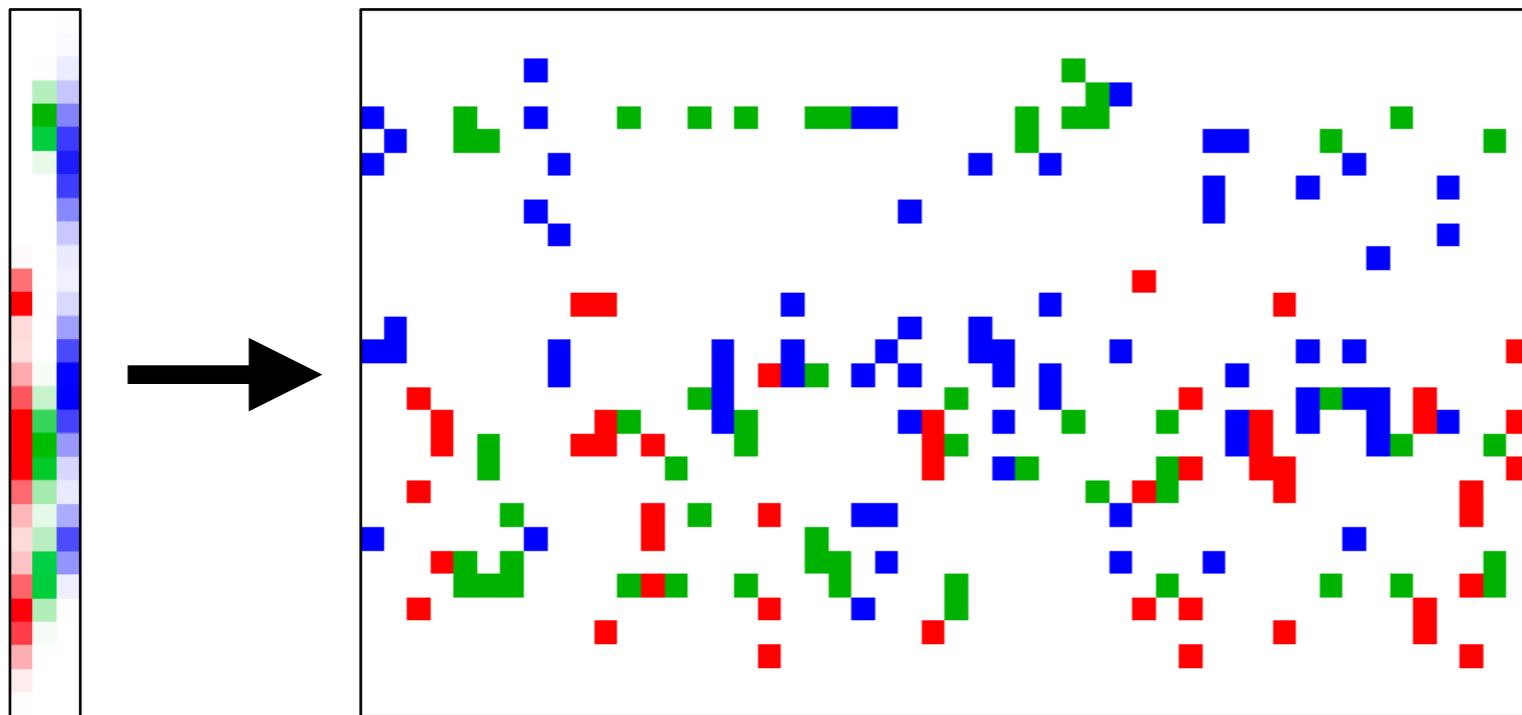
Comparing Distributions for Single Cells



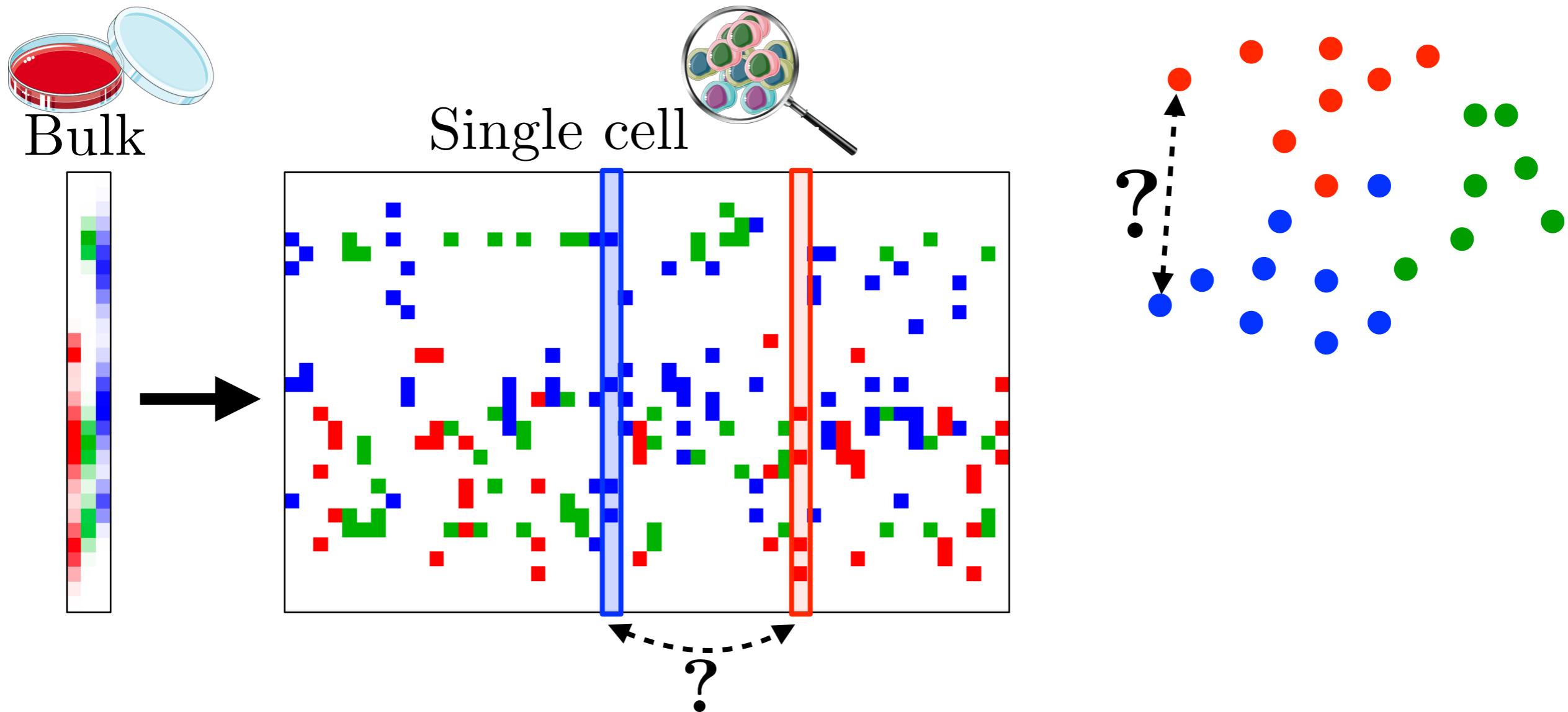
Bulk



Single cell



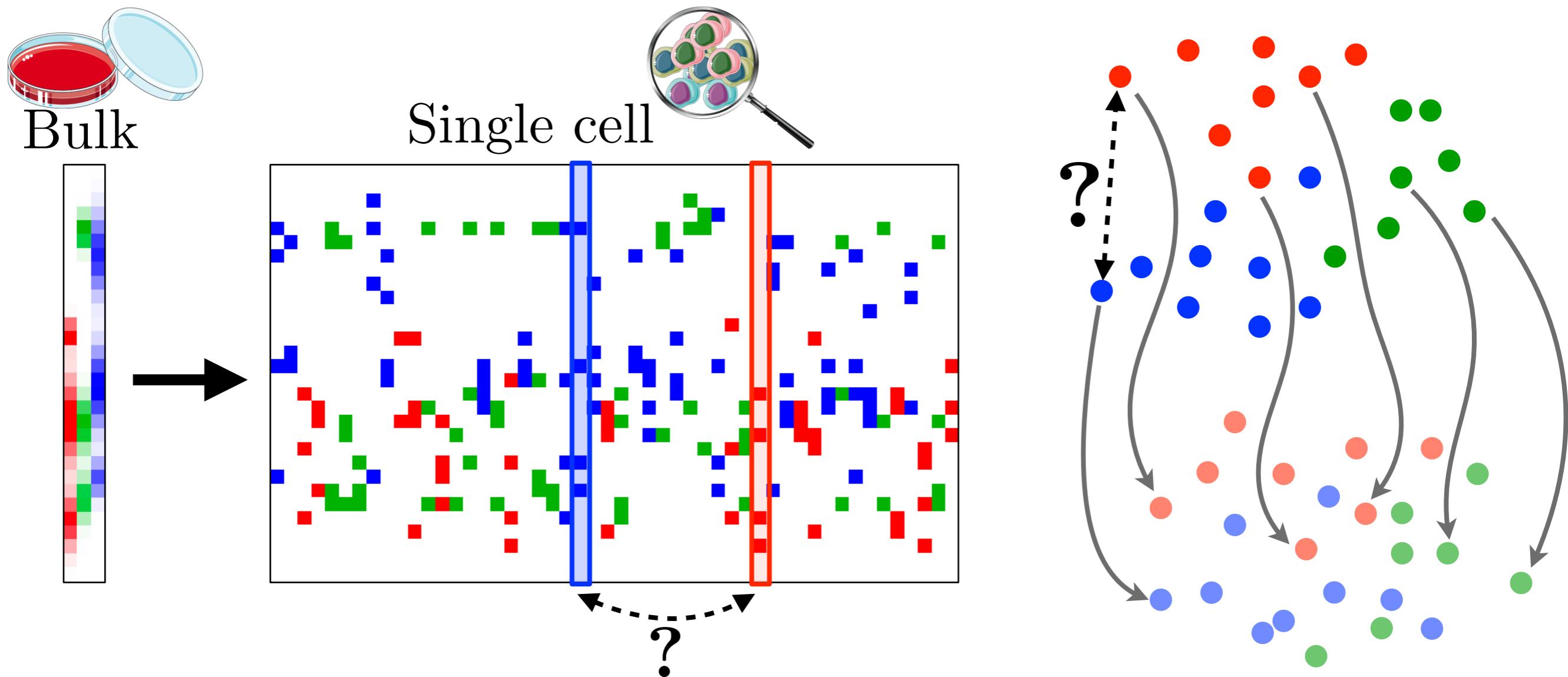
Comparing Distributions for Single Cells



Cluster cells from a single biopsy.

“Distance” between cells? → OT on genes’ space.

Comparing Distributions for Single Cells



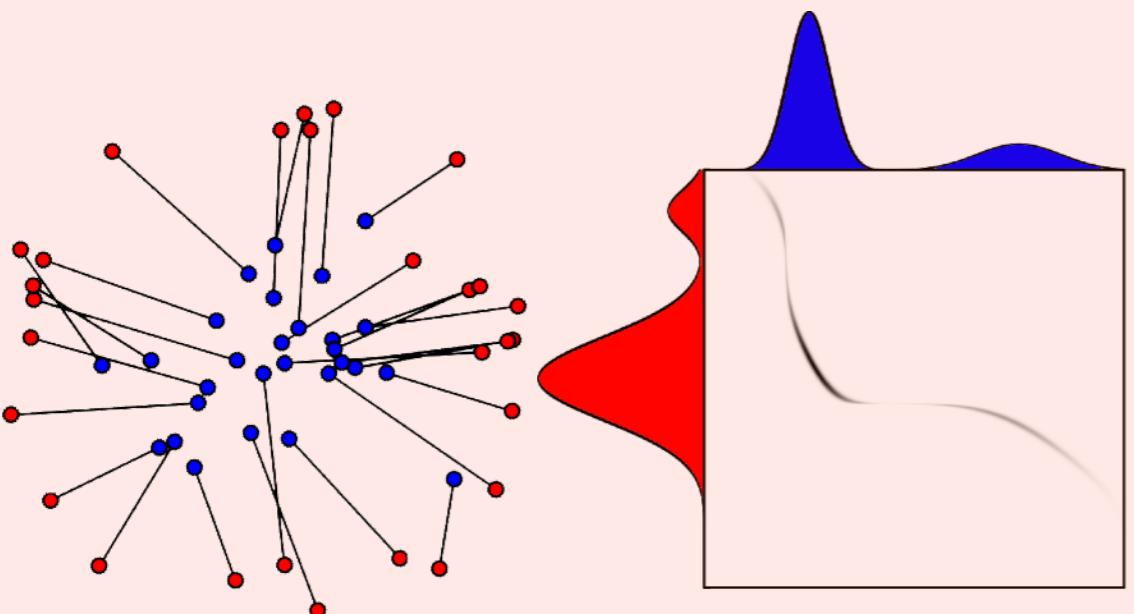
Cluster cells from a single biopsy.

“Distance” between cells? → OT on genes’ space.

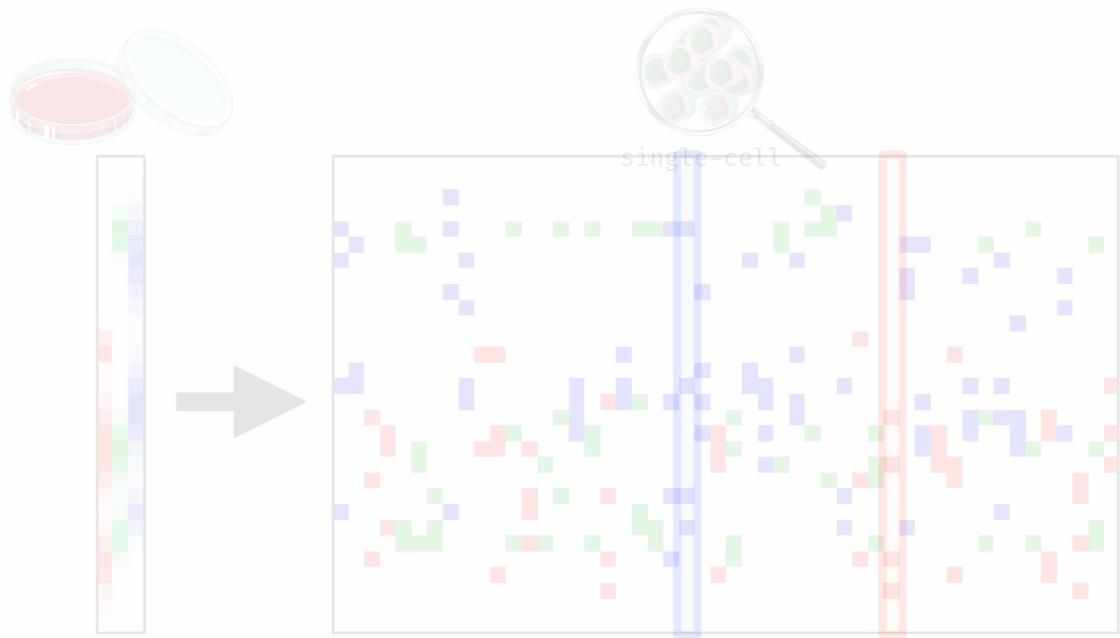
Match cells between two biopsies.

→ OT on cells’ space.

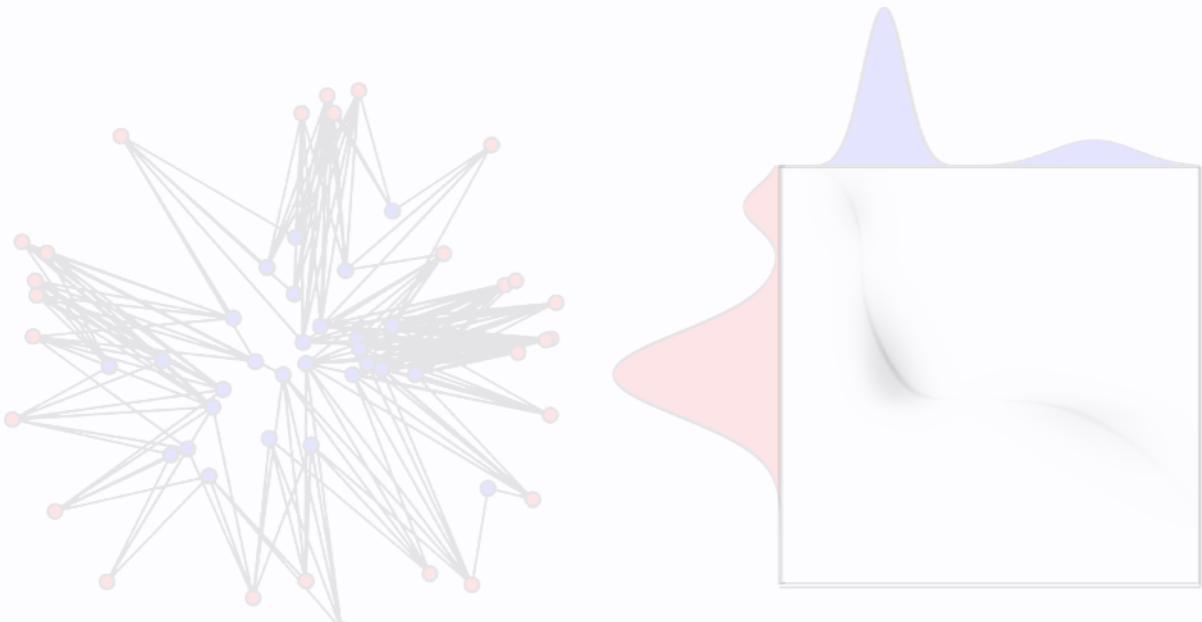
Optimal Transport



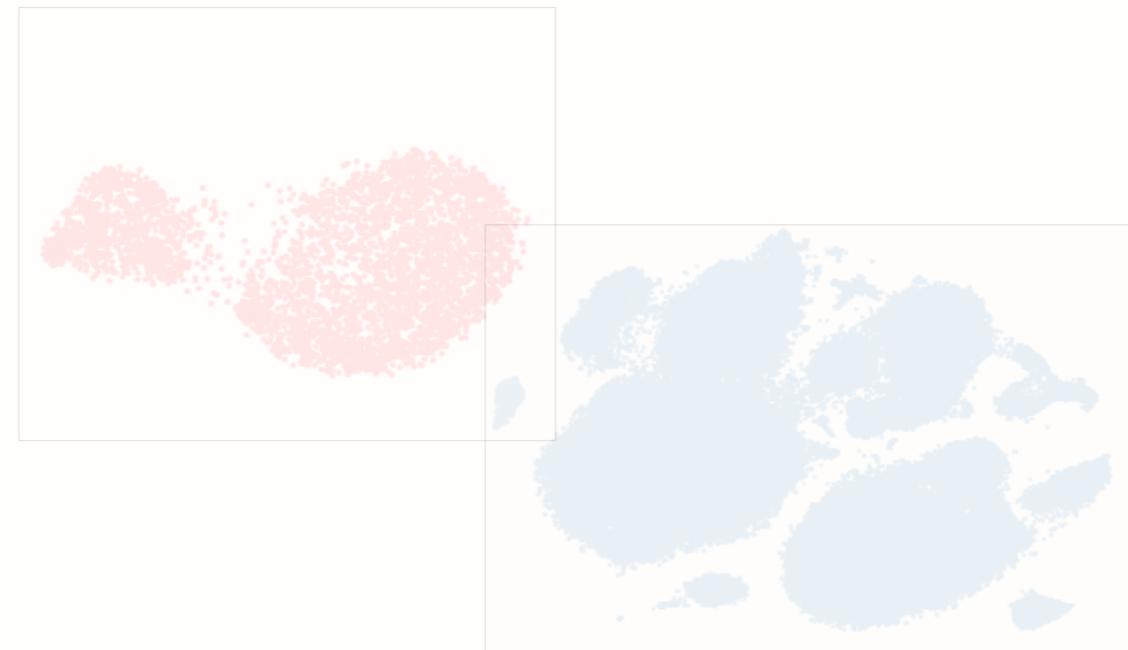
Single Cell Genomics



Entropic Regularization



Gromov Wasserstein

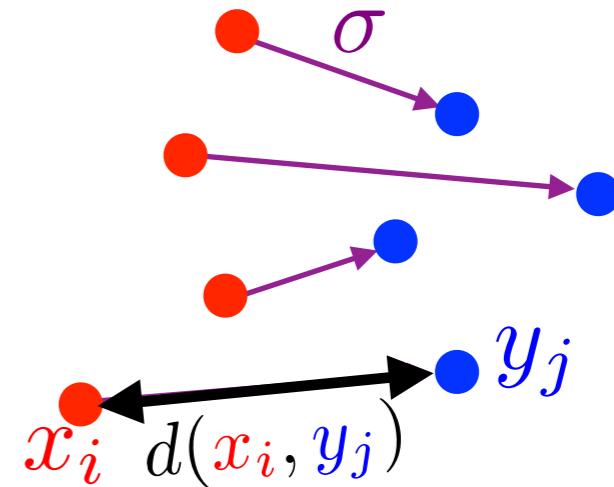


Monge's Problem

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]

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.

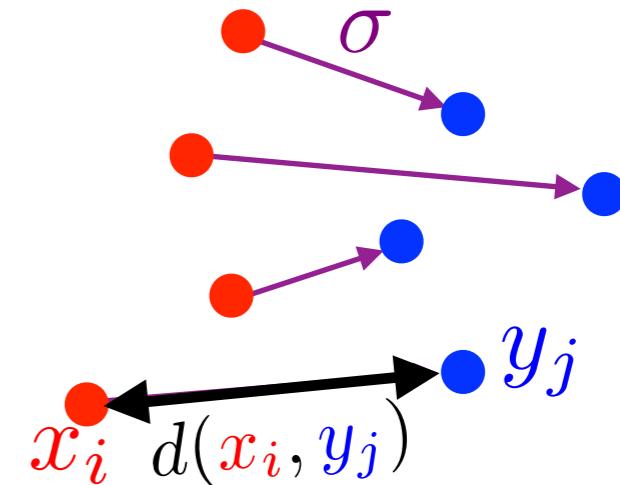
Le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, il s'enfuit 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 fera 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]

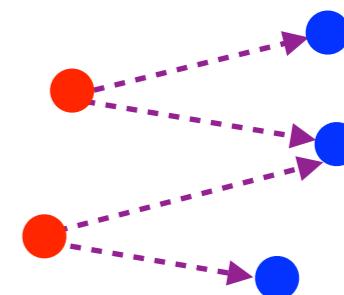
MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

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'enfuit 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 fera la moindre possible, & le prix du transport total fera un minimum.

→ Seems intractable: $n!$ possibilities.

→ Different number of points?



Kantorovitch's Formulation

Discrete distributions:

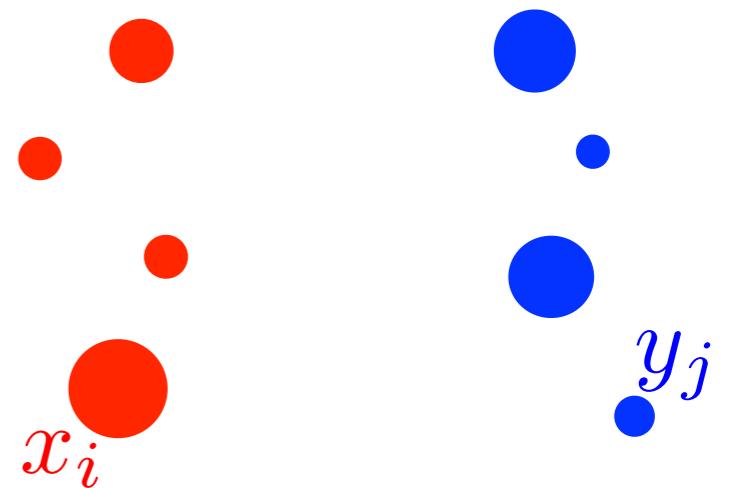
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

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

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:

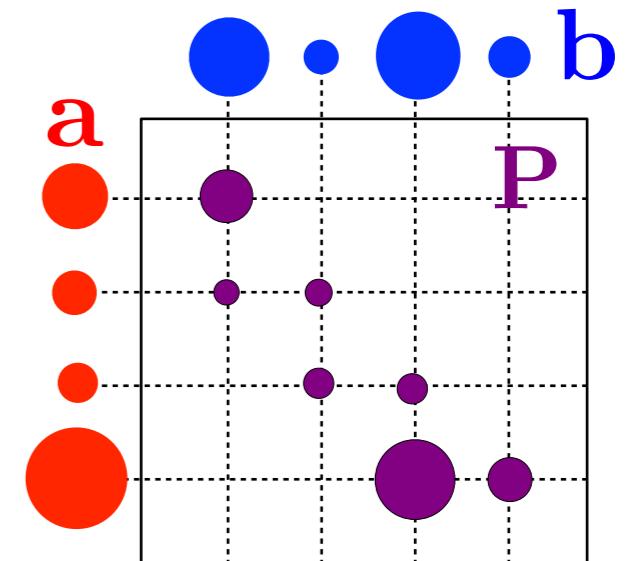
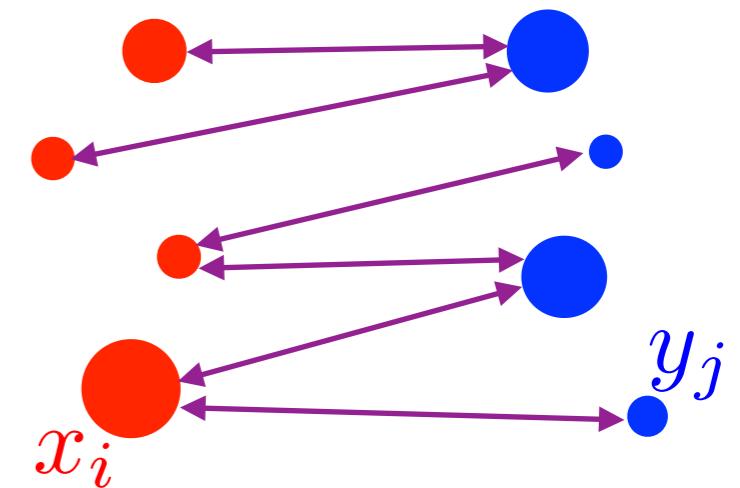
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

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

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:

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

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

$$\mathbf{P} \geq 0, \mathbf{P} \mathbf{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbf{1}_n = \mathbf{b}$$

Kantorovitch's Formulation

Discrete distributions:

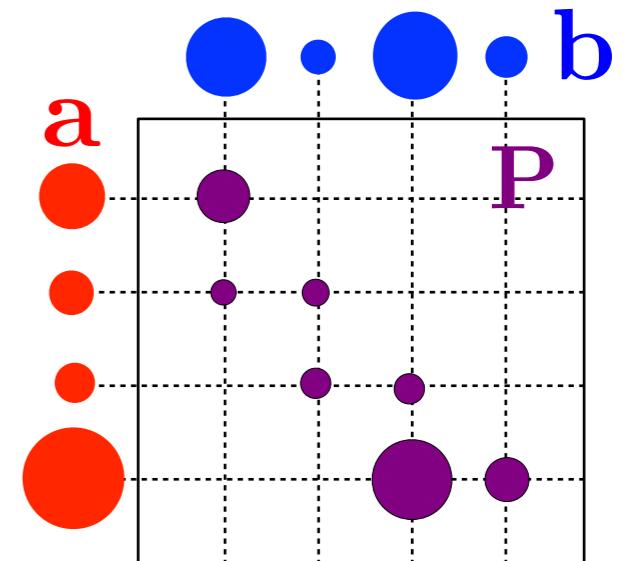
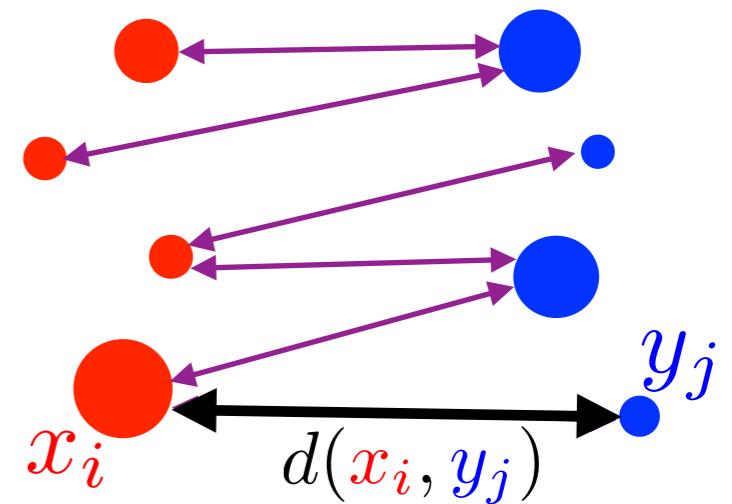
$$\alpha = \sum_{i=1}^n \mathbf{a}_i \delta_{x_i}$$

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

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

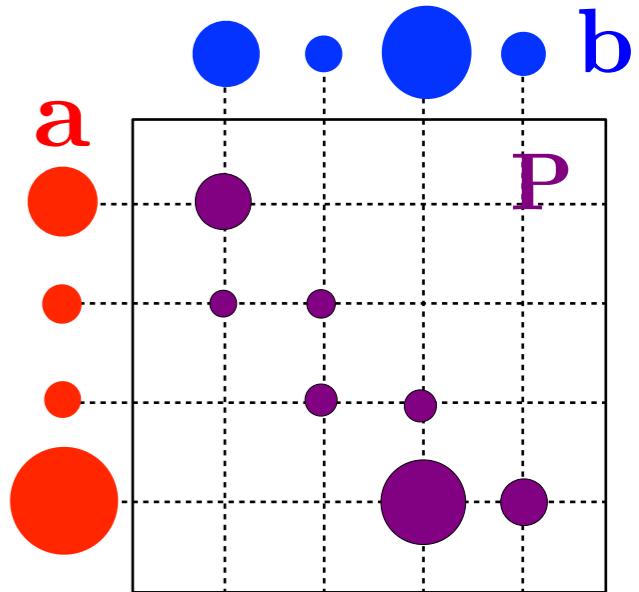
$\sum_i \mathbf{P}_{i,j} = \mathbf{b}_j$

[Kantorovich 1942]

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

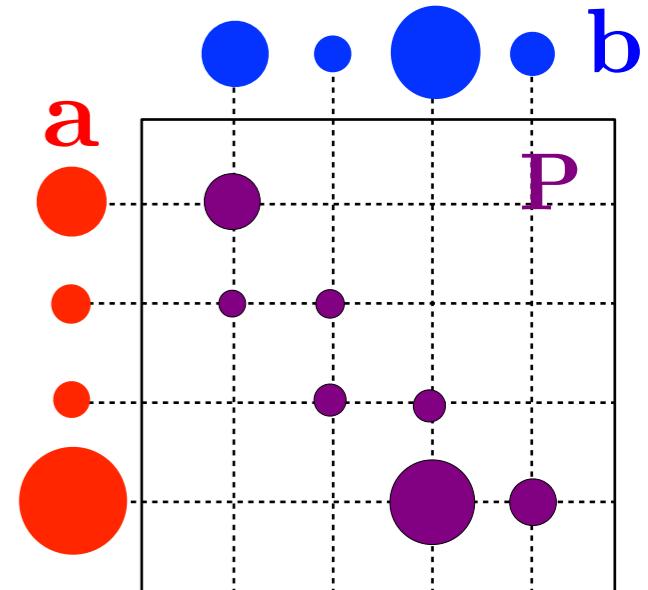
Optimal Transport Distances

$$W_p(\alpha, \beta) \stackrel{\text{def.}}{=} \left(\min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} \right)^{\frac{1}{p}}$$



Optimal Transport Distances

$$W_p(\alpha, \beta) \stackrel{\text{def.}}{=} \left(\min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} \right)^{\frac{1}{p}}$$



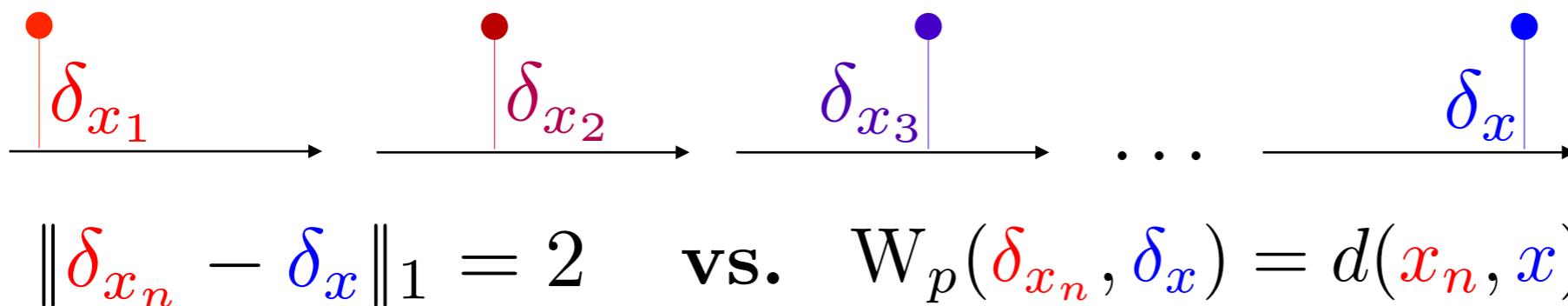
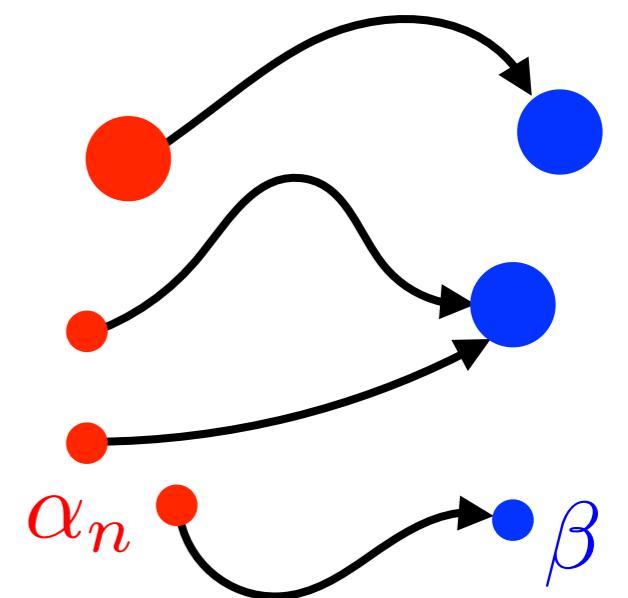
Theorem: W_p is a distance and

$$\alpha_n \rightarrow \beta \Leftrightarrow W_p(\alpha_n, \beta) \rightarrow 0$$

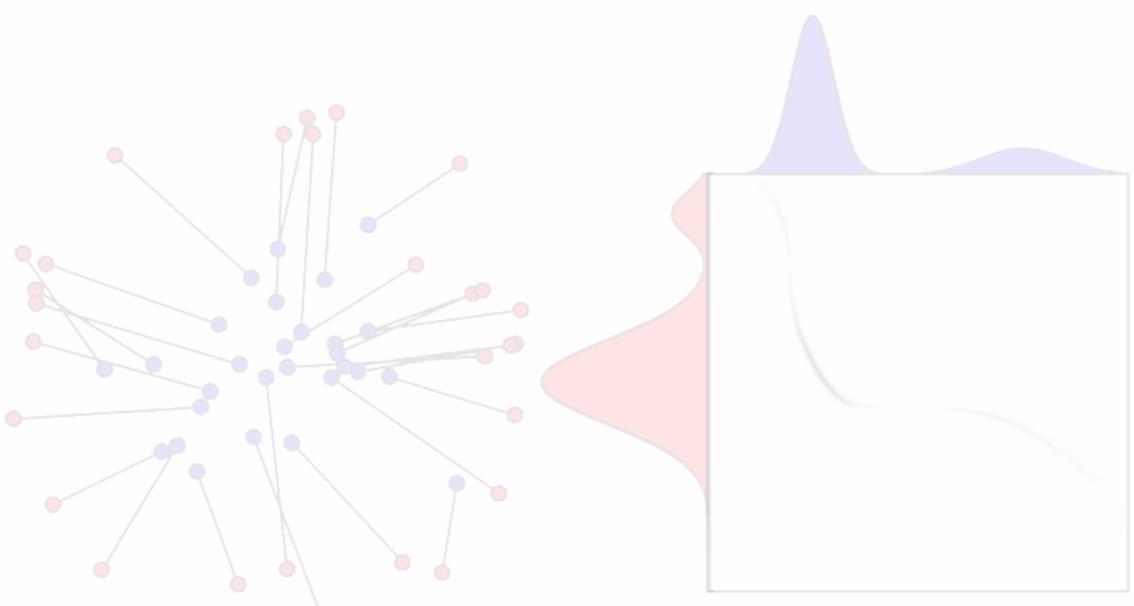
Convergence in law:

$$\alpha_n \rightarrow \beta$$

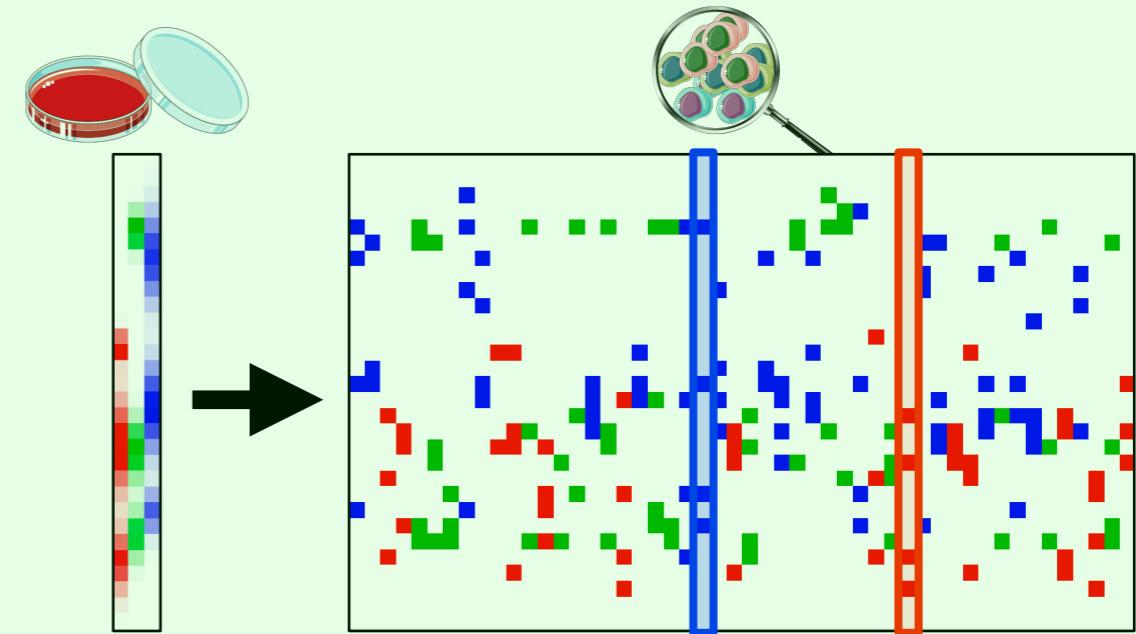
$$\Leftrightarrow \forall f \in \mathcal{C}(\mathcal{X}), \int_{\mathcal{X}} f d\alpha_n \rightarrow \int_{\mathcal{X}} f d\beta$$



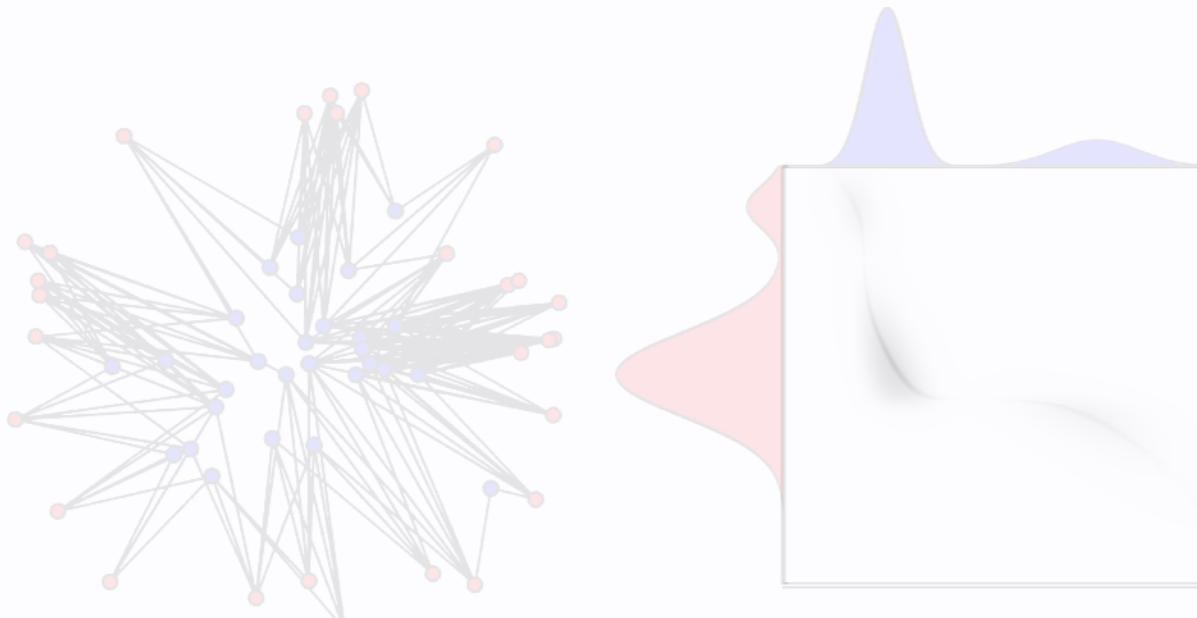
Optimal Transport



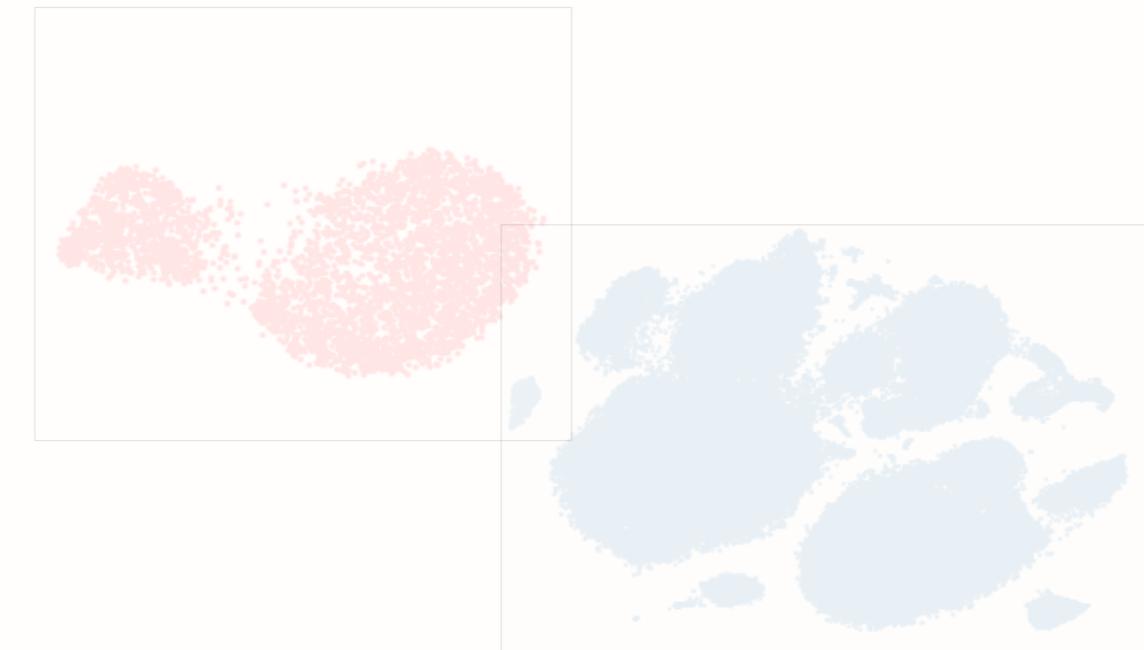
Single Cell Genomics



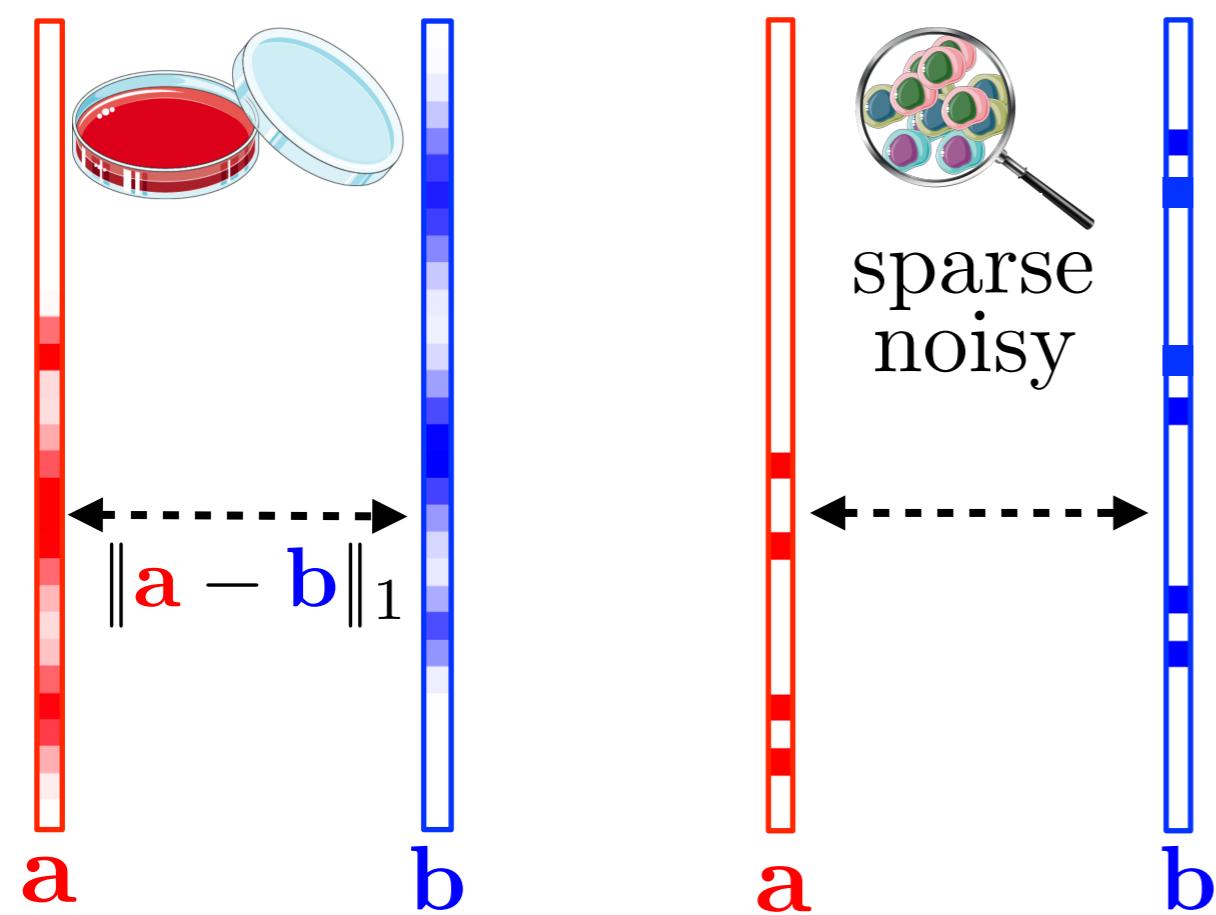
Entropic Regularization



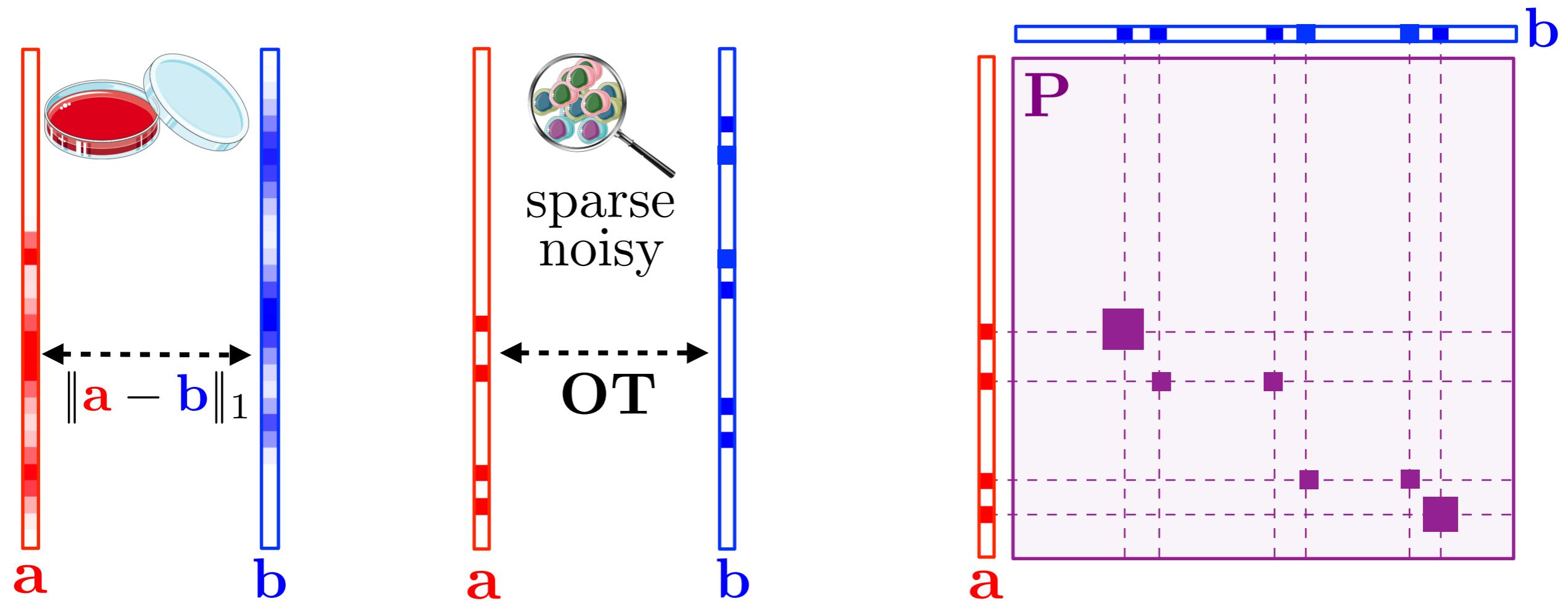
Gromov Wasserstein



OT for Cell to Cell Dissimilarity

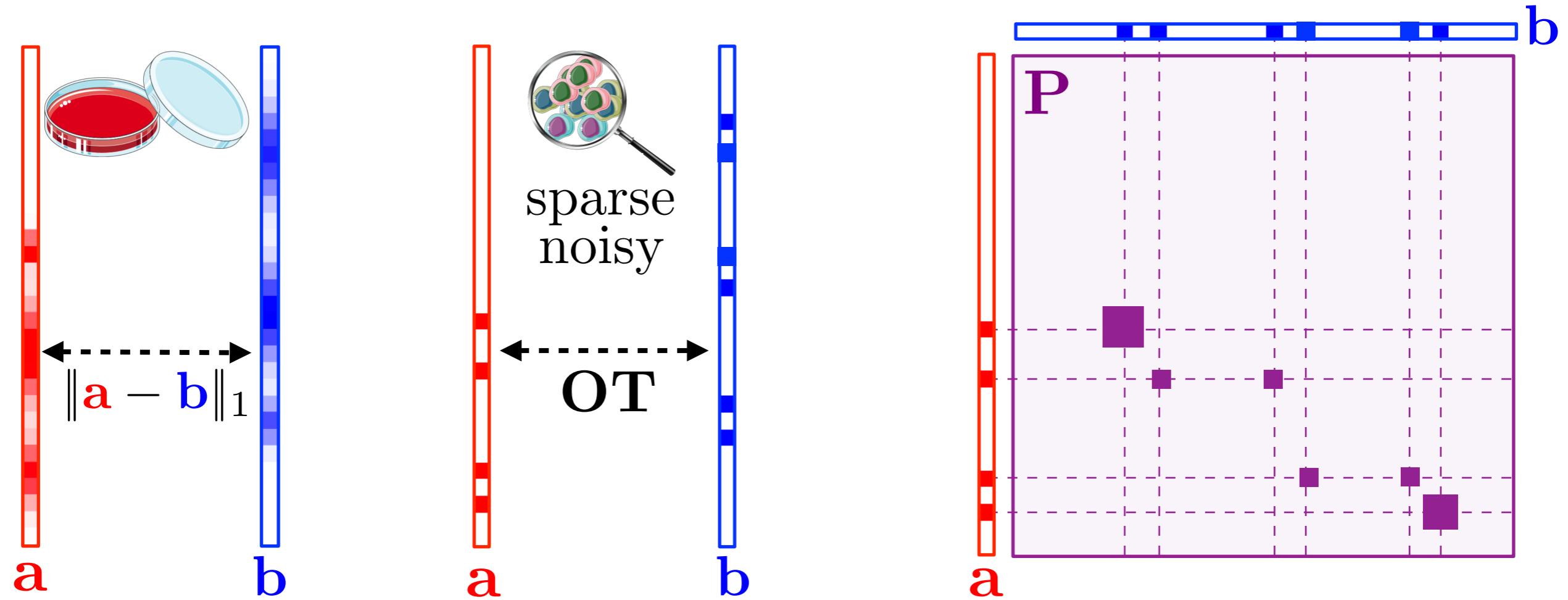


OT for Cell to Cell Dissimilarity

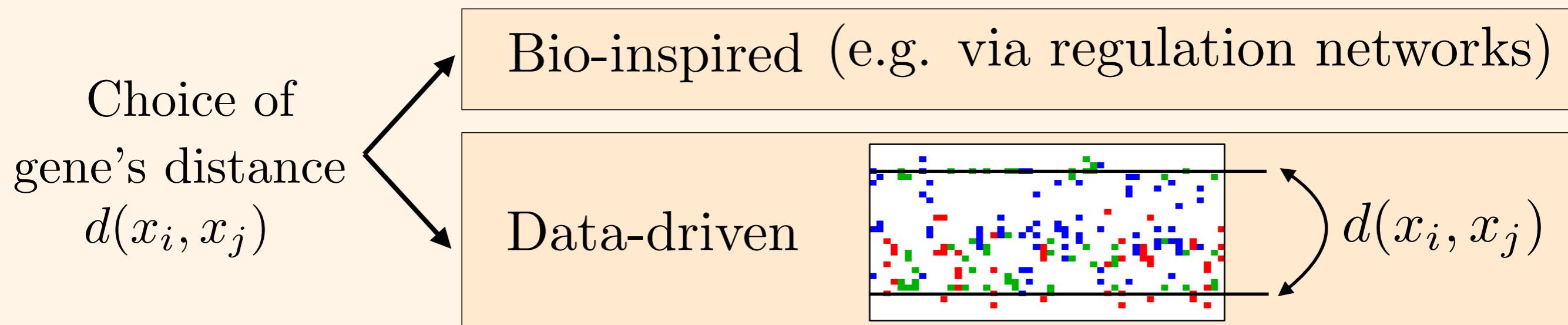


$$\min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, x_j) \mathbf{P}_{i,j}$$

OT for Cell to Cell Dissimilarity

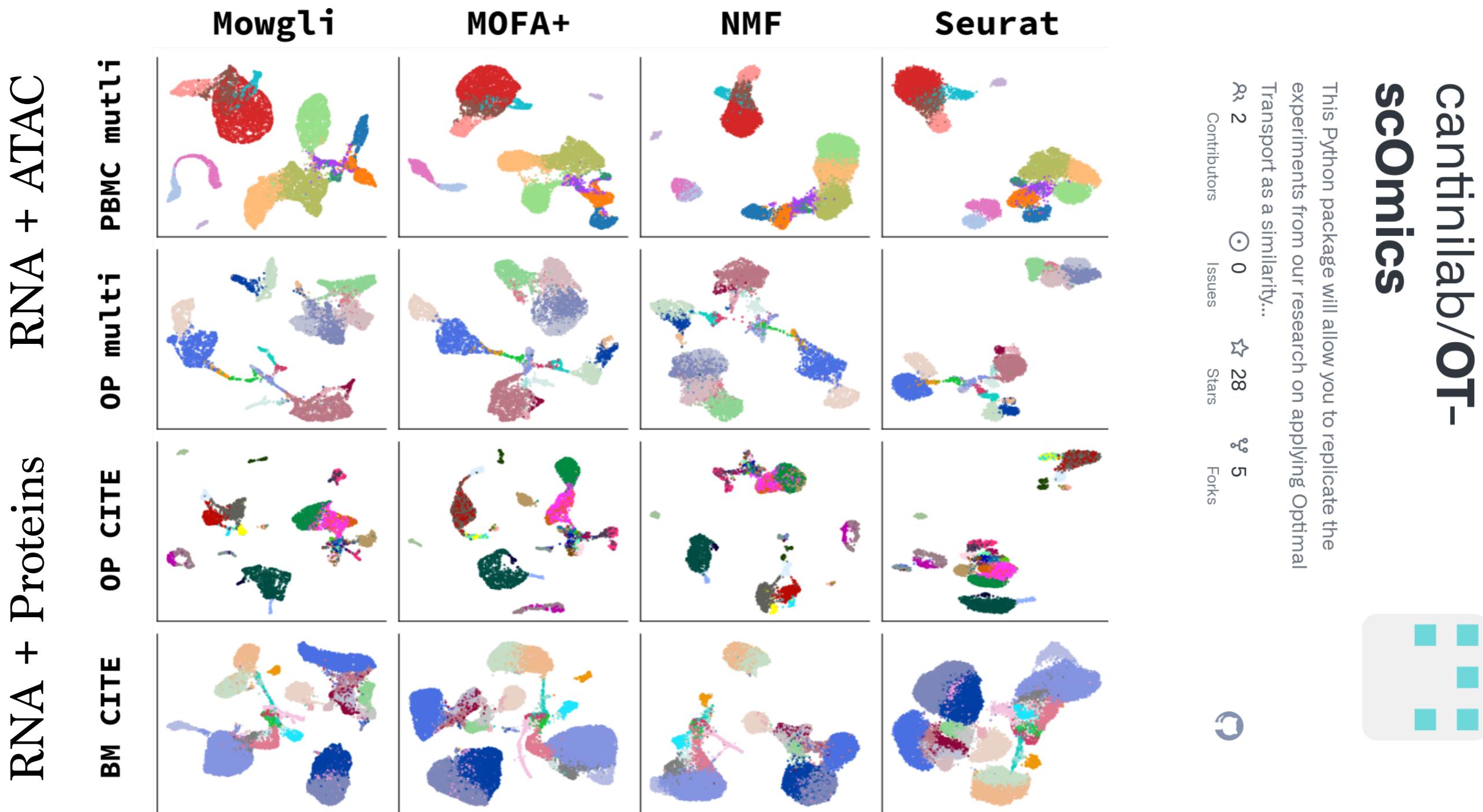


$$\min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, x_j) \mathbf{P}_{i,j}$$



Clustering Results

Mowgli: couple OT with non-negative matrix factorization.



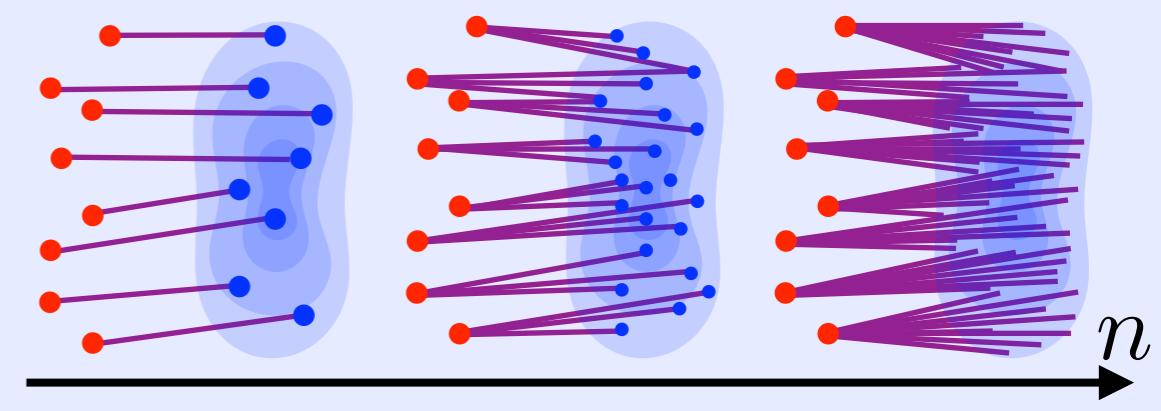
OT brings:

- Clustering quality improvement on ground truth.
- Better interpretability (pathway enrichment analysis)

Comparing Different Cells' Distributions

Curse of dimensionality:

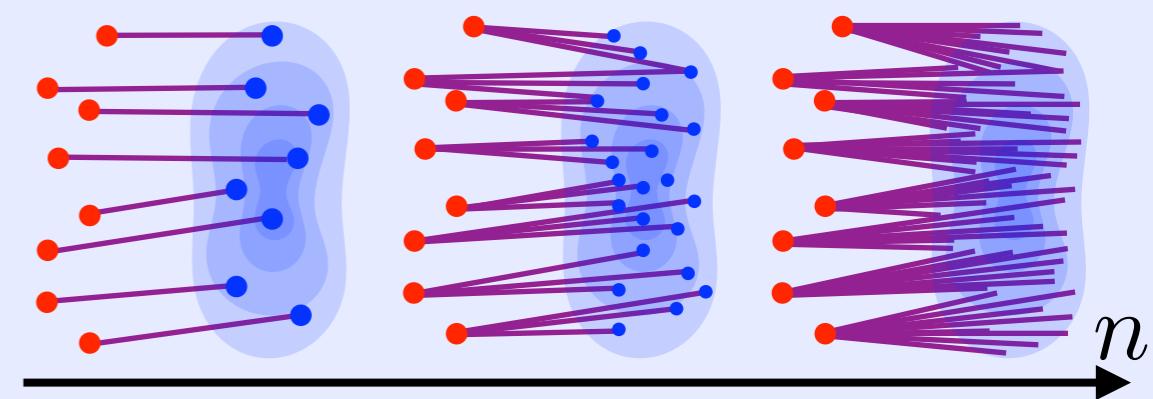
- Algorithms scale like $O(n^3)$
- n scales like $O(1/\text{precision}^d)$
- \rightsquigarrow Entropic regularization.



Comparing Different Cells' Distributions

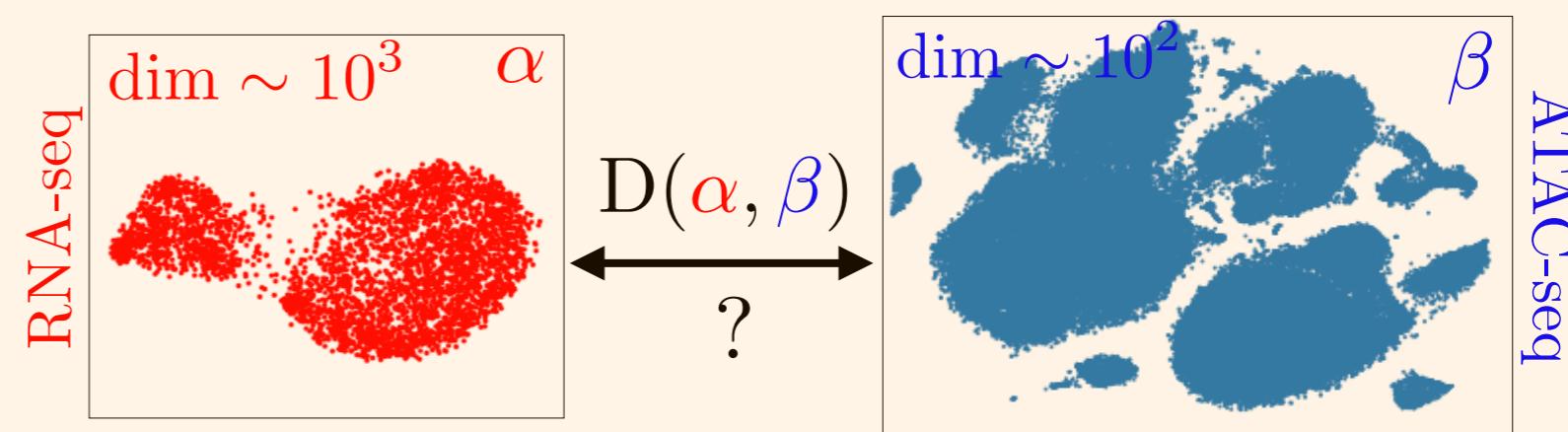
Curse of dimensionality:

- Algorithms scale like $O(n^3)$
- n scales like $O(1/\text{precision}^d)$
- \rightsquigarrow Entropic regularization.

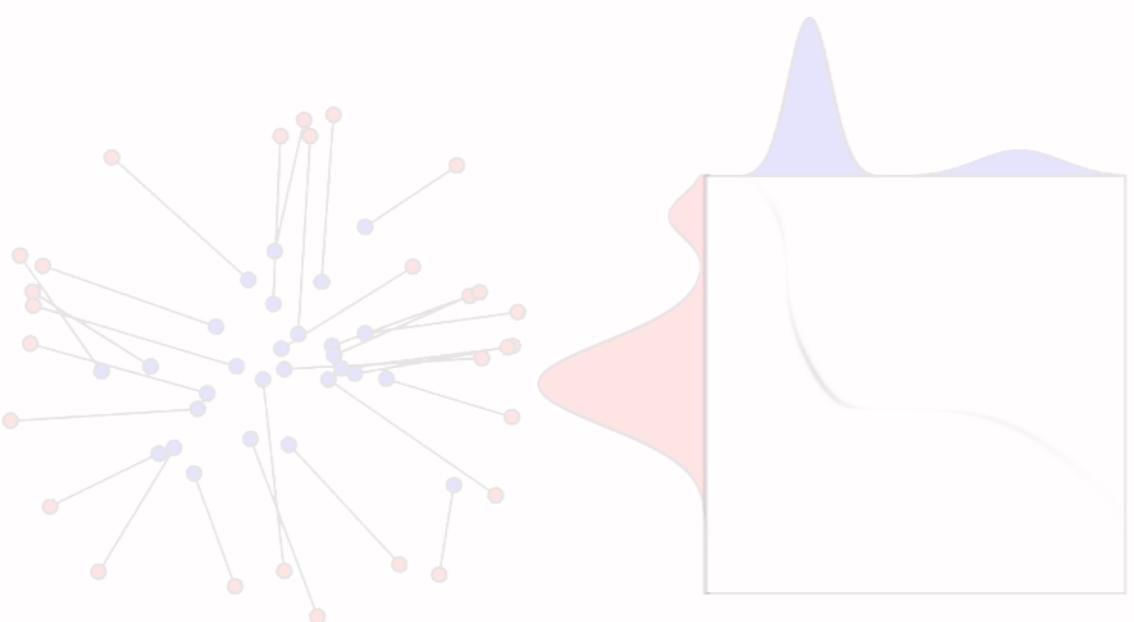


Transportation between different spaces:

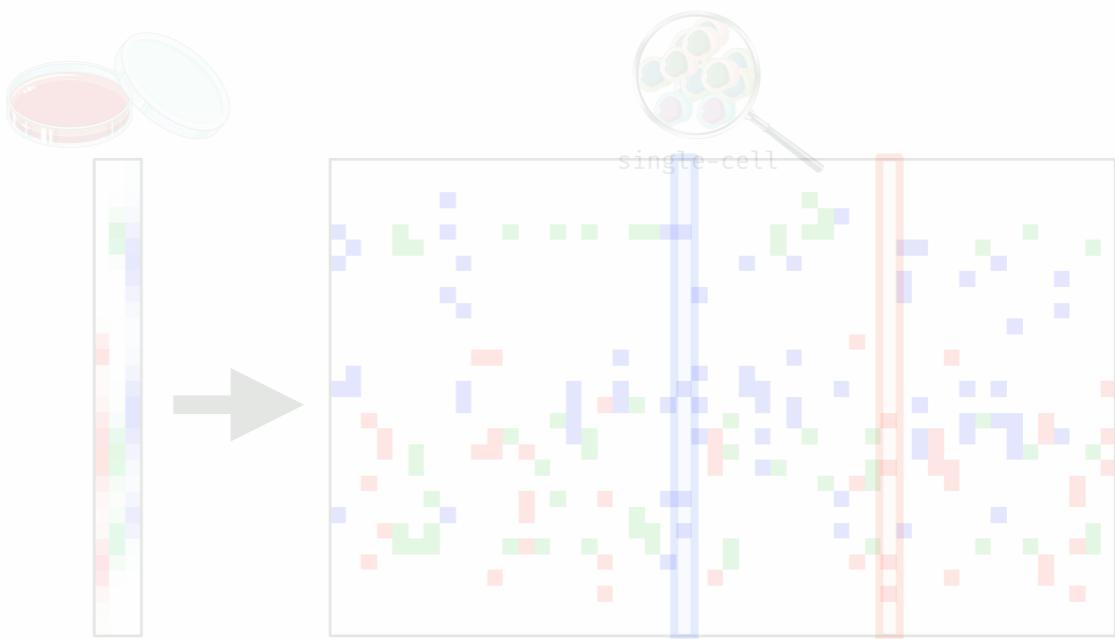
- No joint embedding.
- Different dimensions.
- \rightsquigarrow Gromov-Wasserstein.



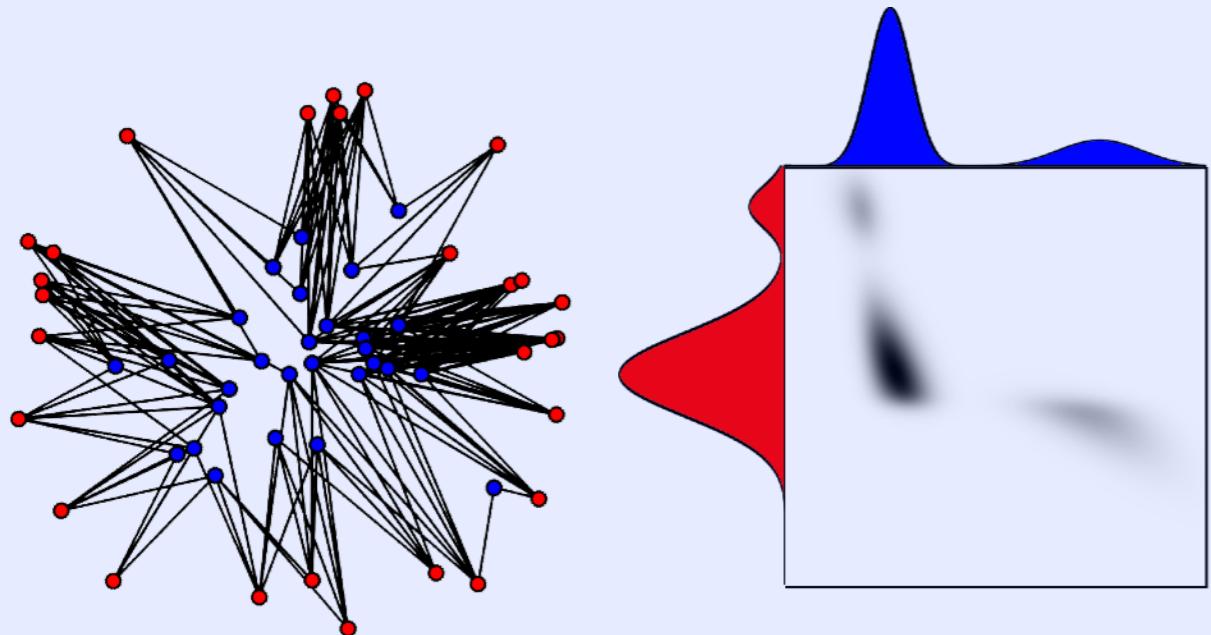
Optimal Transport



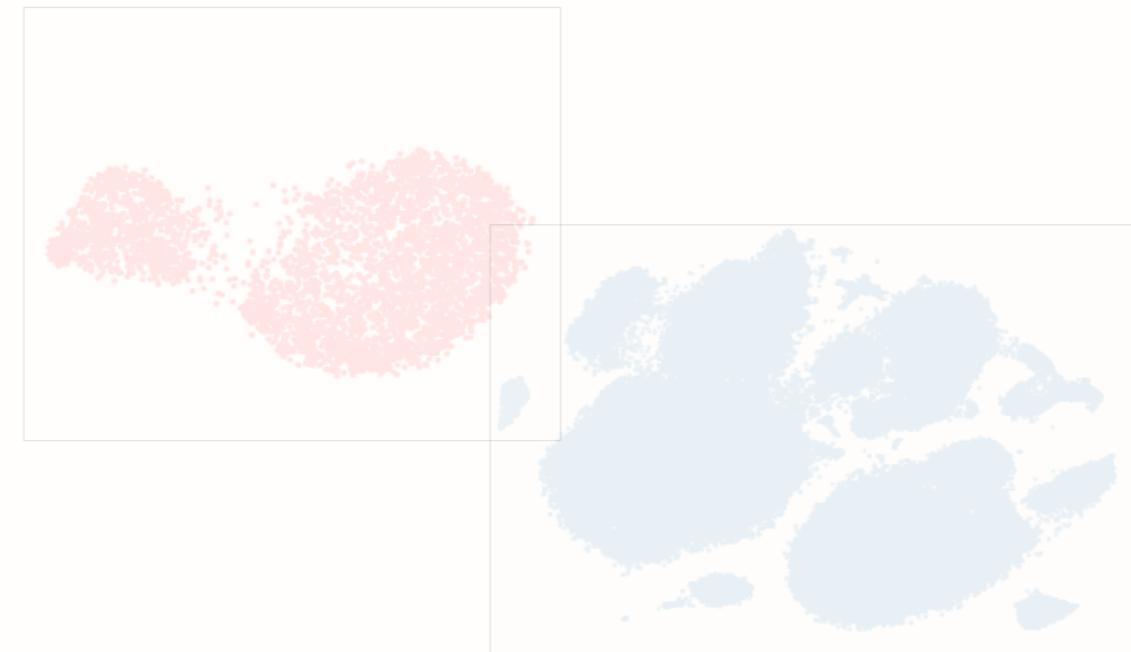
Single Cell Genomics



Entropic Regularization



Gromov Wasserstein



Entropic Regularization

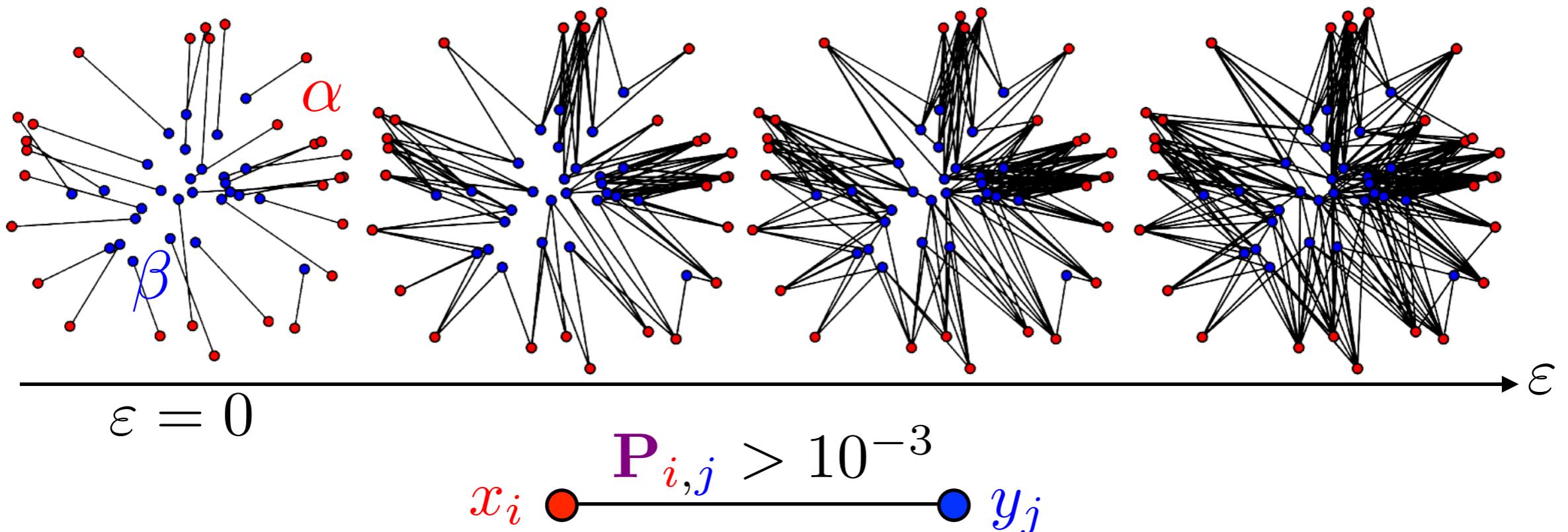
Schrödinger's problem:

[1931]



Erwin
Schrödinger

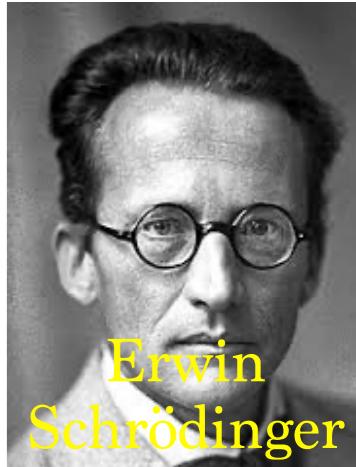
$$W_p^\varepsilon(\alpha, \beta)^p \stackrel{\text{def.}}{=} \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$



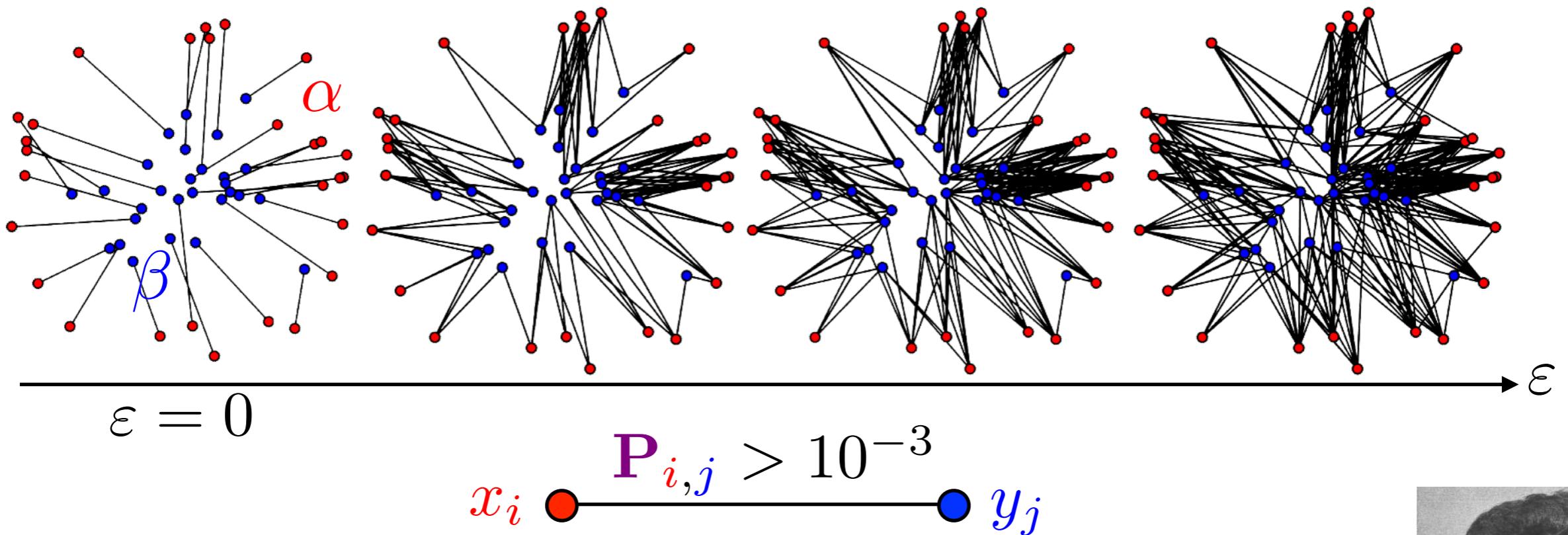
Entropic Regularization

Schrödinger's problem:

[1931]

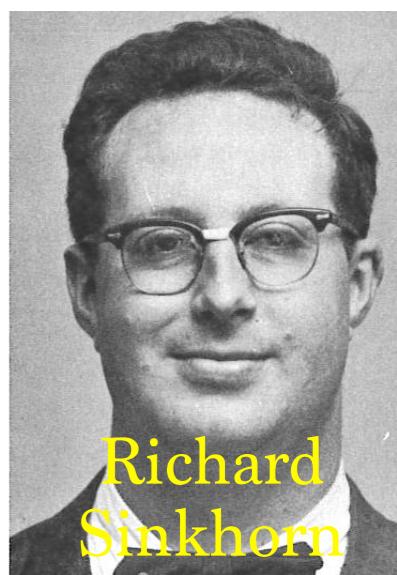


$$W_p^\varepsilon(\alpha, \beta)^p \stackrel{\text{def.}}{=} \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \sum_{i,j} d(x_i, y_j)^p \mathbf{P}_{i,j} + \varepsilon \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

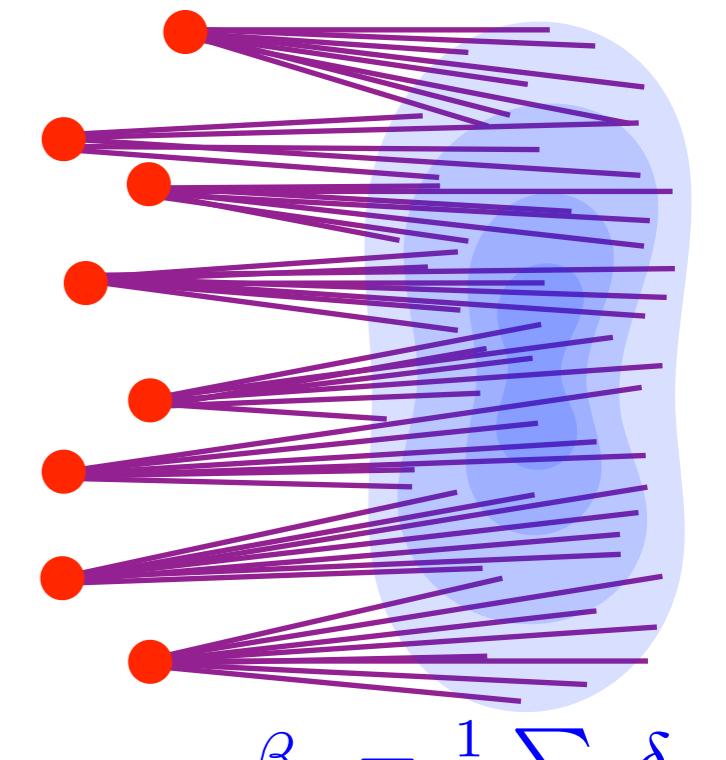
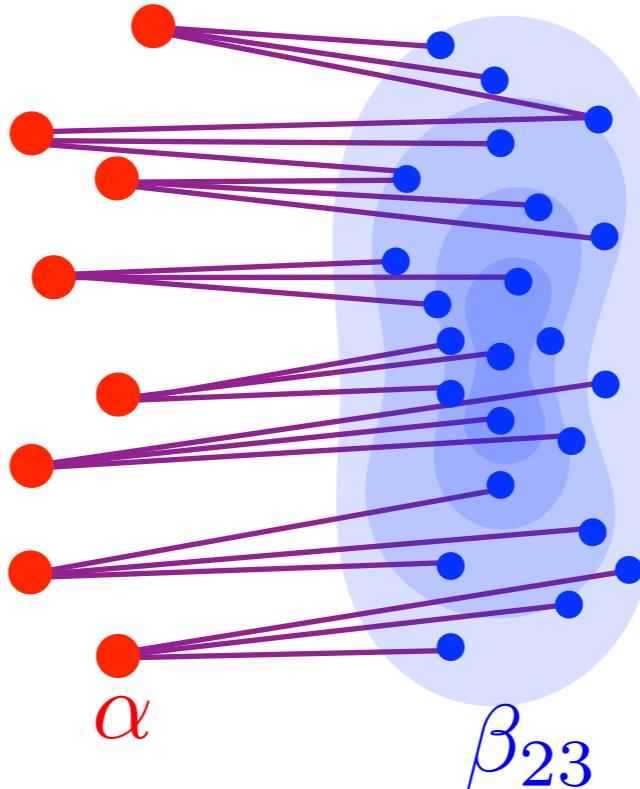
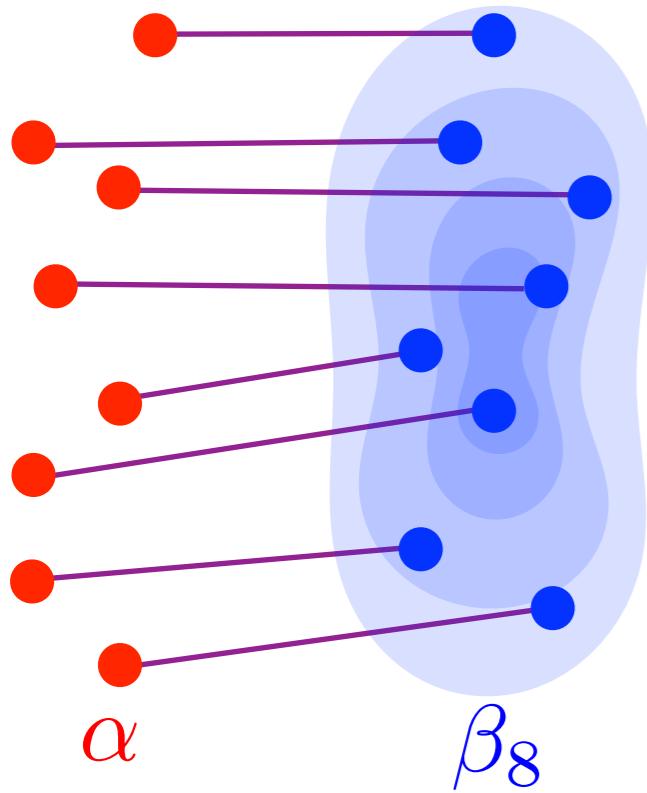


Sinkhorn's algorithm: $O(n^2/\varepsilon^2)$

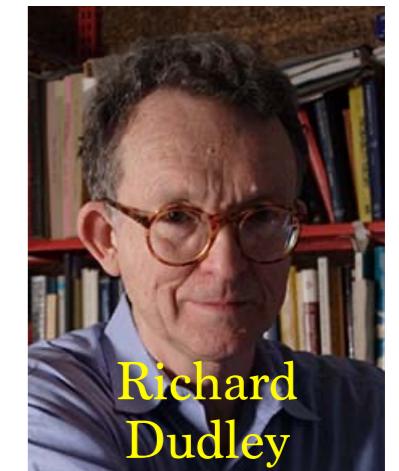
→ parallelizable, GPU-friendly [Cuturi 2013]



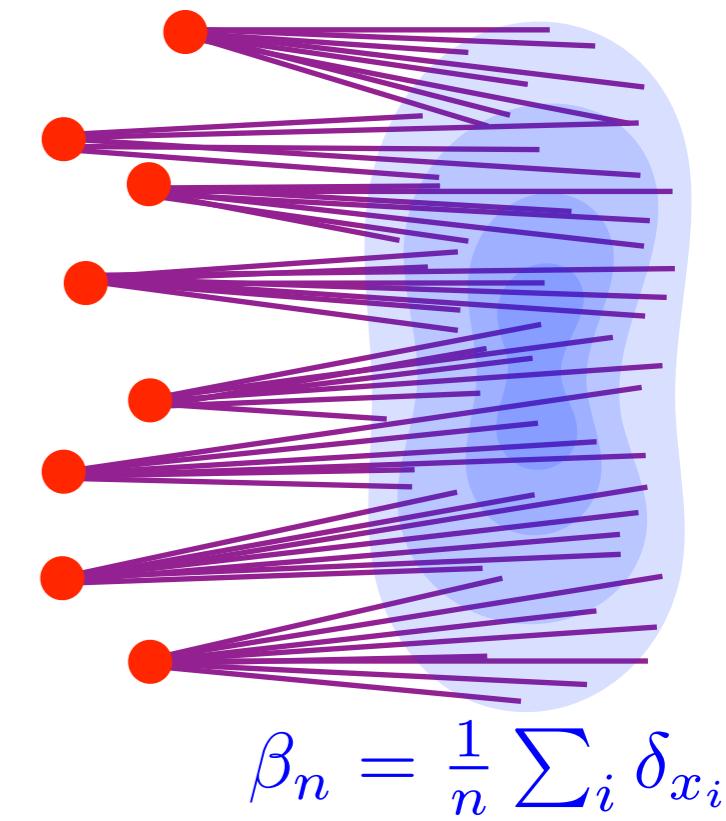
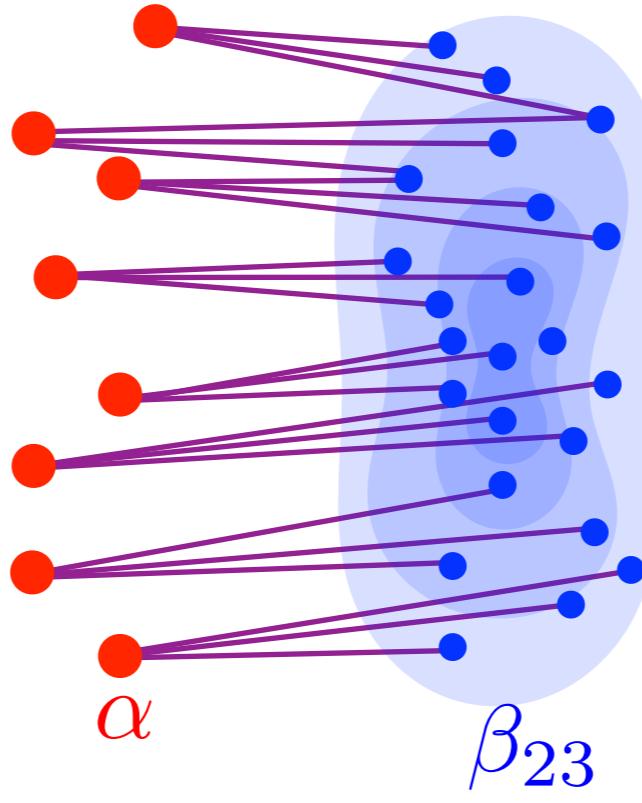
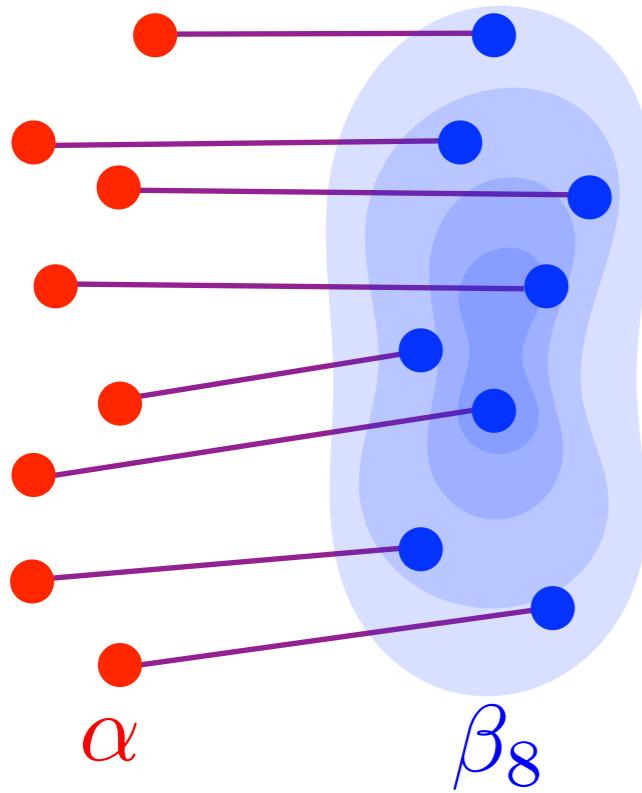
The Curse of Dimensionality



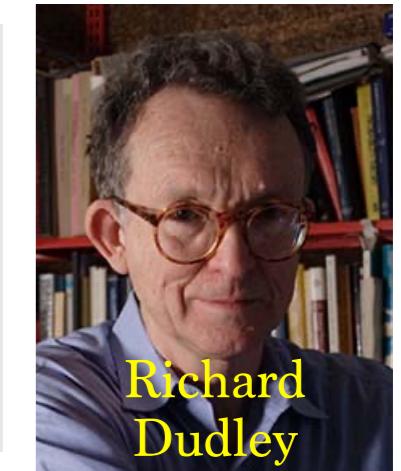
Theorem: $\mathbb{E}|W_p(\alpha, \beta_n) - W_p(\alpha, \beta_\infty)| \leq \delta$
[Dudley 1968] requires $n \sim (1/\delta)^{\text{dimension}}$



The Curse of Dimensionality



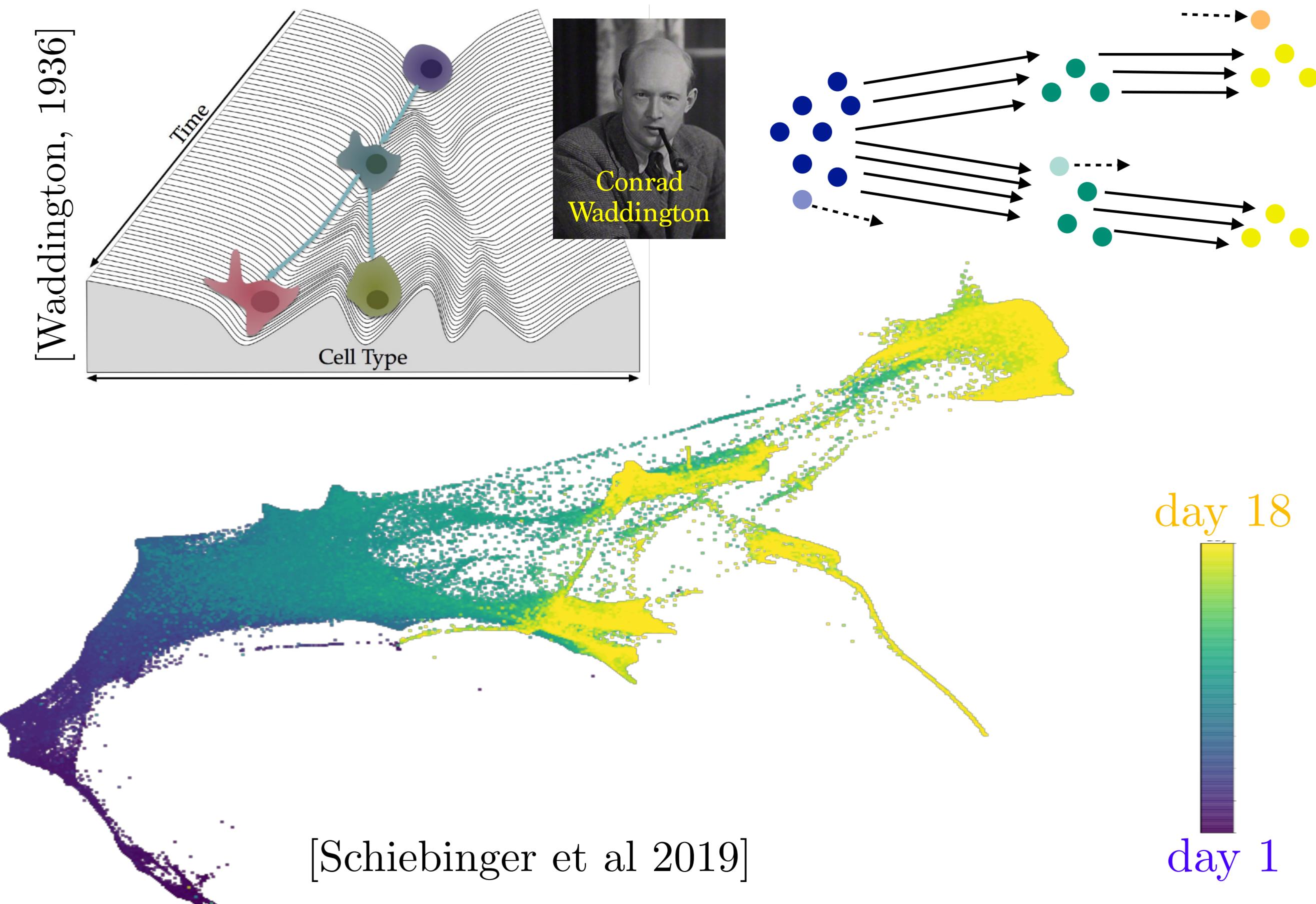
Theorem: $\mathbb{E}|W_p(\alpha, \beta_n) - W_p(\alpha, \beta_\infty)| \leq \delta$
[Dudley 1968] requires $n \sim (1/\delta)^{\text{dimension}}$



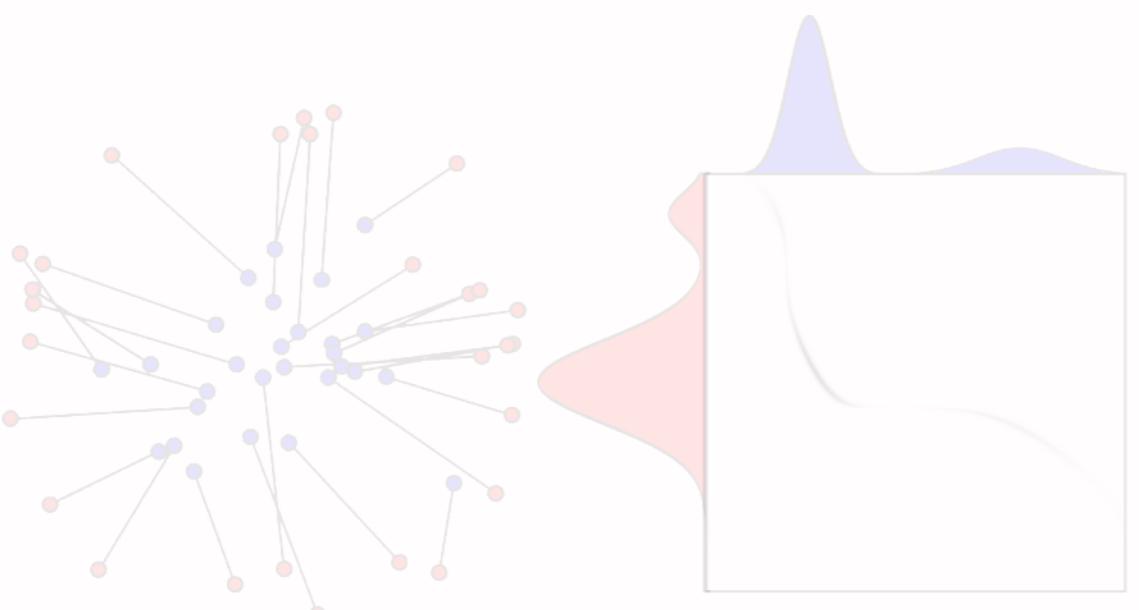
Theorem: $\mathbb{E}|W_p^\varepsilon(\alpha, \beta_n) - W_p^\varepsilon(\alpha, \beta_\infty)| \leq \delta$
[Genevay 2019] requires $n \sim (1/\varepsilon)^{\text{dimension}} \times (1/\delta)^2$



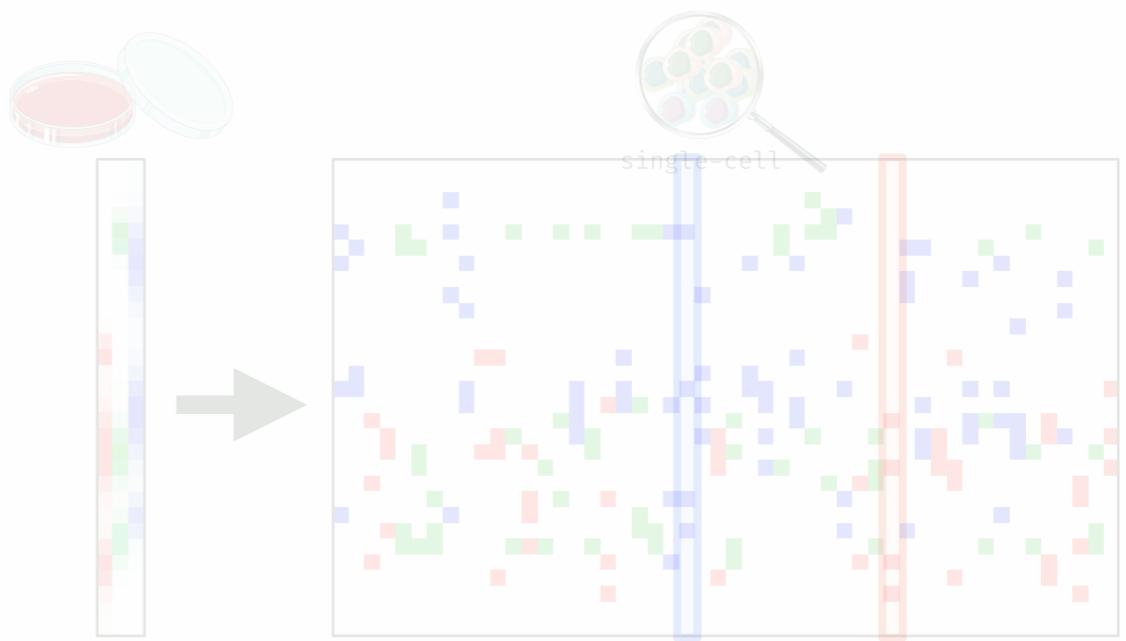
Trajectory Inference with OT



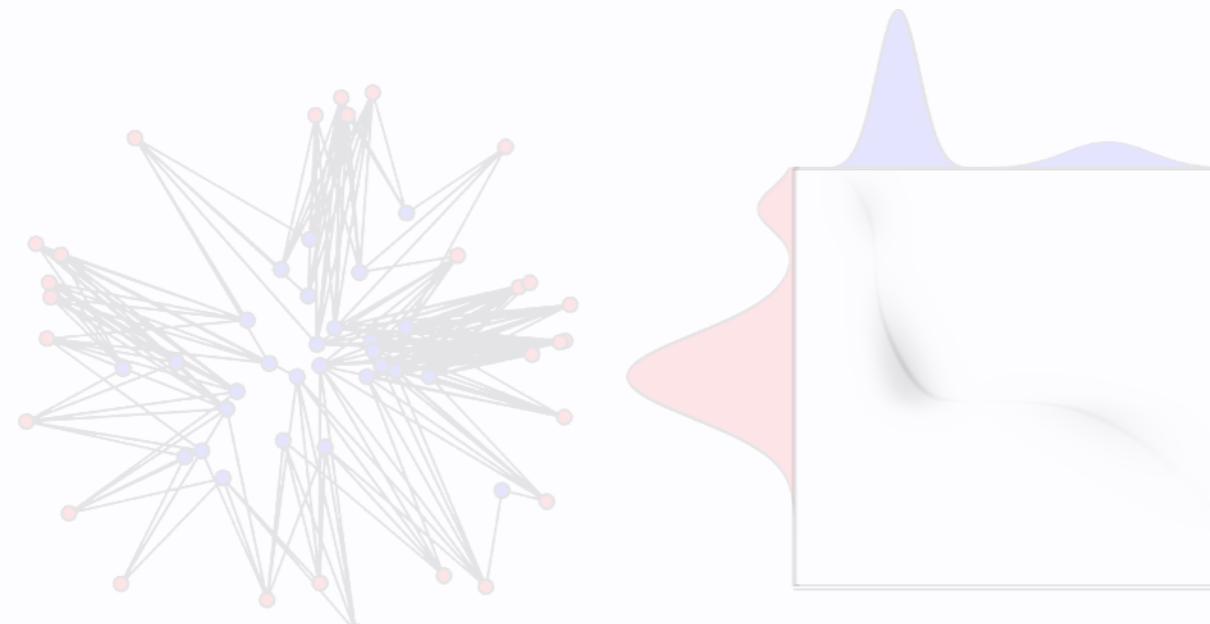
Optimal Transport



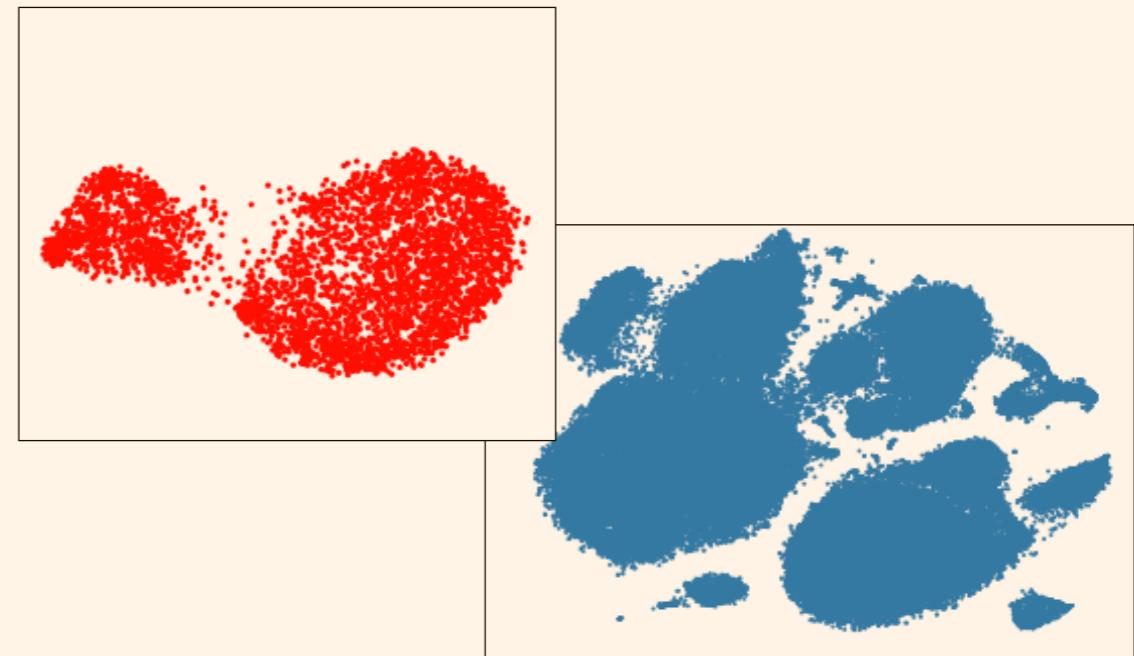
Single Cell Genomics



Entropic Regularization



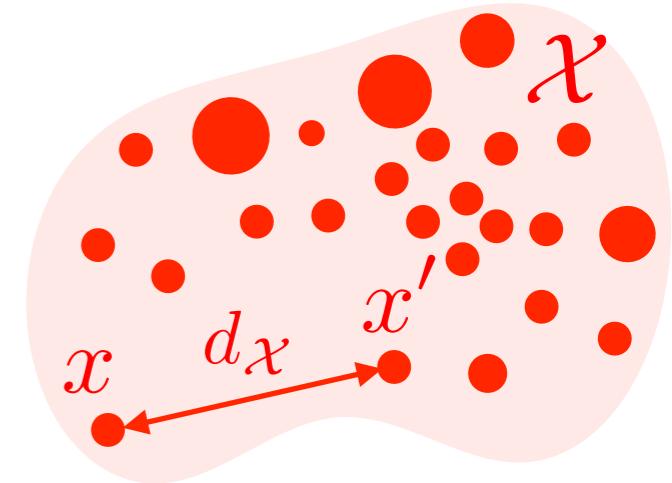
Gromov Wasserstein



Gromov-Wasserstein

Metric measure space $\mathbb{X} \triangleq (\mathcal{X}, \alpha, d_{\mathcal{X}})$:

$\alpha \in \mathcal{M}_+^1(\mathcal{X})$ and $d_{\mathcal{X}}$ distance on \mathcal{X} .



Gromov-Wasserstein

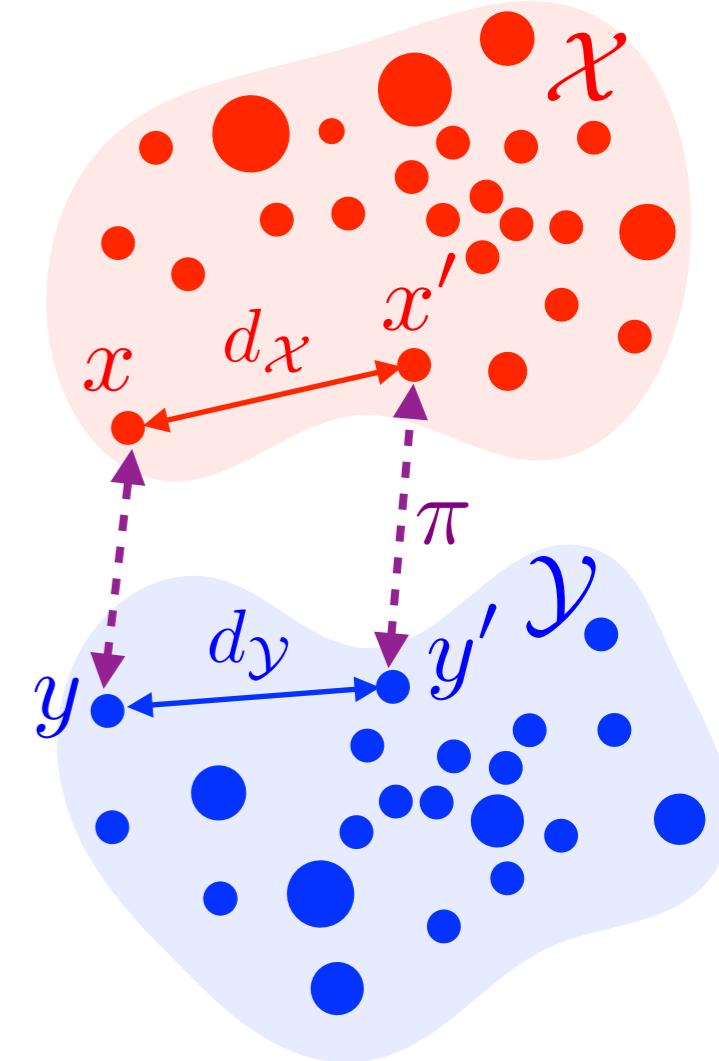
Metric measure space $\mathbb{X} \triangleq (\mathcal{X}, \alpha, d_{\mathcal{X}})$:

$\alpha \in \mathcal{M}_+^1(\mathcal{X})$ and $d_{\mathcal{X}}$ distance on \mathcal{X} .

$$\text{GW}_2^2(\mathbb{X}, \mathbb{Y}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \langle \mathcal{Q}\mathbf{P}, \mathbf{P} \rangle$$

$$\stackrel{\text{def.}}{=} \sum_{i, i', j, j'} |d_{\mathcal{X}}(x_i, x_{i'}) - d_{\mathcal{Y}}(y_j, y_{j'})|^2 \mathbf{P}_{i,j} \mathbf{P}_{i',j'}$$

[Memoli 2011][Sturm 2011]



Gromov-Wasserstein

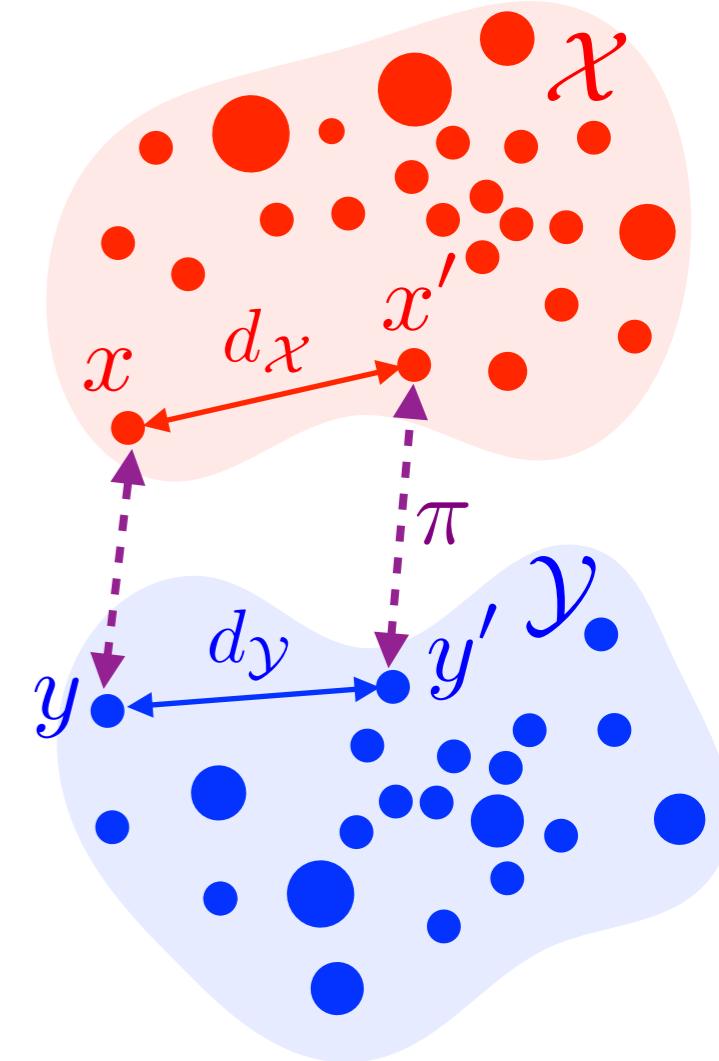
Metric measure space $\mathbb{X} \triangleq (\mathcal{X}, \alpha, d_{\mathcal{X}})$:

$\alpha \in \mathcal{M}_+^1(\mathcal{X})$ and $d_{\mathcal{X}}$ distance on \mathcal{X} .

$$\text{GW}_2^2(\mathbb{X}, \mathbb{Y}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \langle \mathcal{Q}\mathbf{P}, \mathbf{P} \rangle$$

$$\stackrel{\text{def.}}{=} \sum_{i, i', j, j'} |d_{\mathcal{X}}(x_i, x_{i'}) - d_{\mathcal{Y}}(y_j, y_{j'})|^2 \mathbf{P}_{i,j} \mathbf{P}_{i',j'}$$

[Memoli 2011][Sturm 2011]



→ non-convex, NP-hard ...

if $d_{\mathcal{X}} = \|\cdot\|$, $d_{\mathcal{Y}} = \|\cdot\|$: concave!

Gromov-Wasserstein

Metric measure space $\mathbb{X} \triangleq (\mathcal{X}, \alpha, d_{\mathcal{X}})$:

$\alpha \in \mathcal{M}_+^1(\mathcal{X})$ and $d_{\mathcal{X}}$ distance on \mathcal{X} .

$$\text{GW}_2^2(\mathbb{X}, \mathbb{Y}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \langle \mathcal{Q}\mathbf{P}, \mathbf{P} \rangle$$

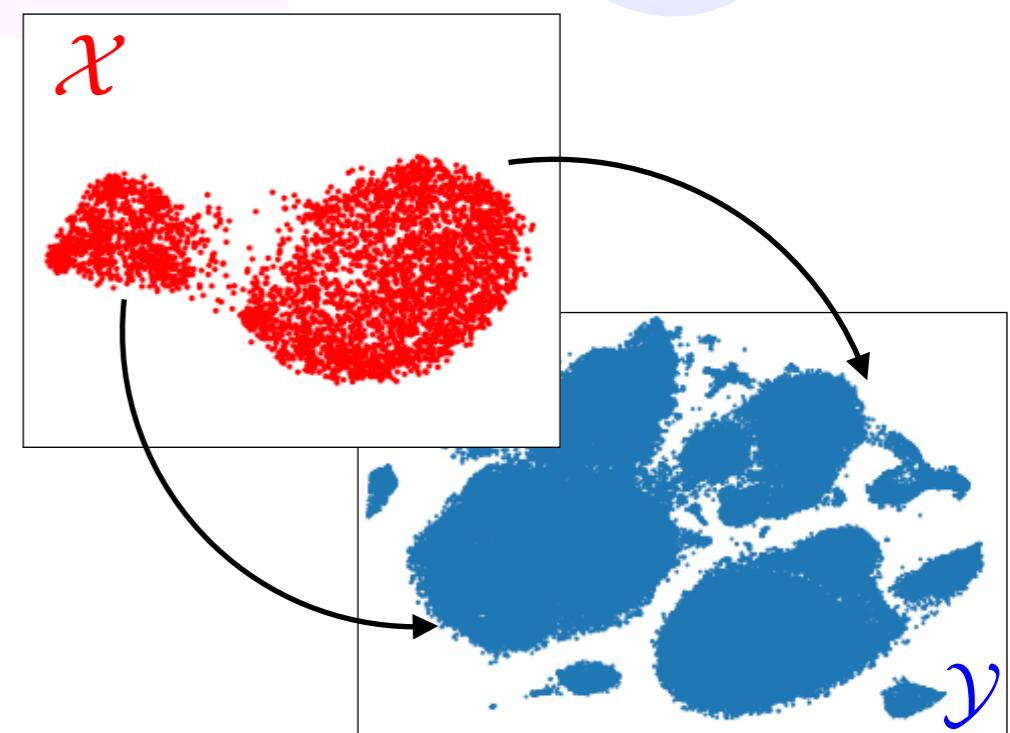
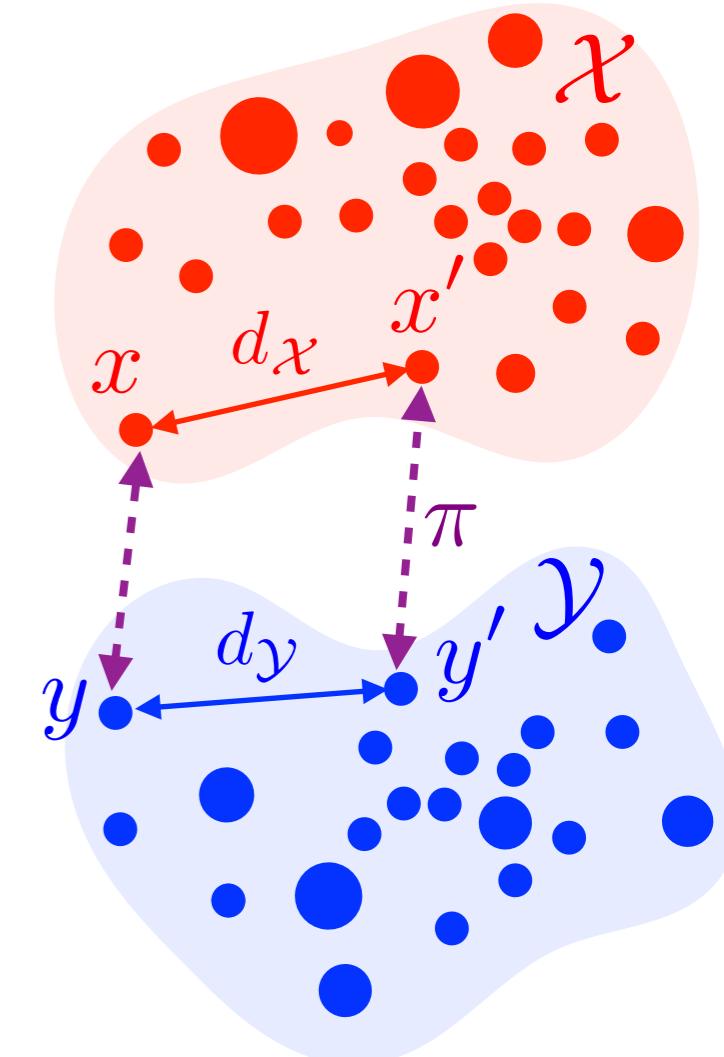
$$\stackrel{\text{def.}}{=} \sum_{i, i', j, j'} |d_{\mathcal{X}}(x_i, x_{i'}) - d_{\mathcal{Y}}(y_j, y_{j'})|^2 \mathbf{P}_{i,j} \mathbf{P}_{i',j'}$$

[Memoli 2011][Sturm 2011]

→ non-convex, NP-hard ...

if $d_{\mathcal{X}} = \|\cdot\|$, $d_{\mathcal{Y}} = \|\cdot\|$: concave!

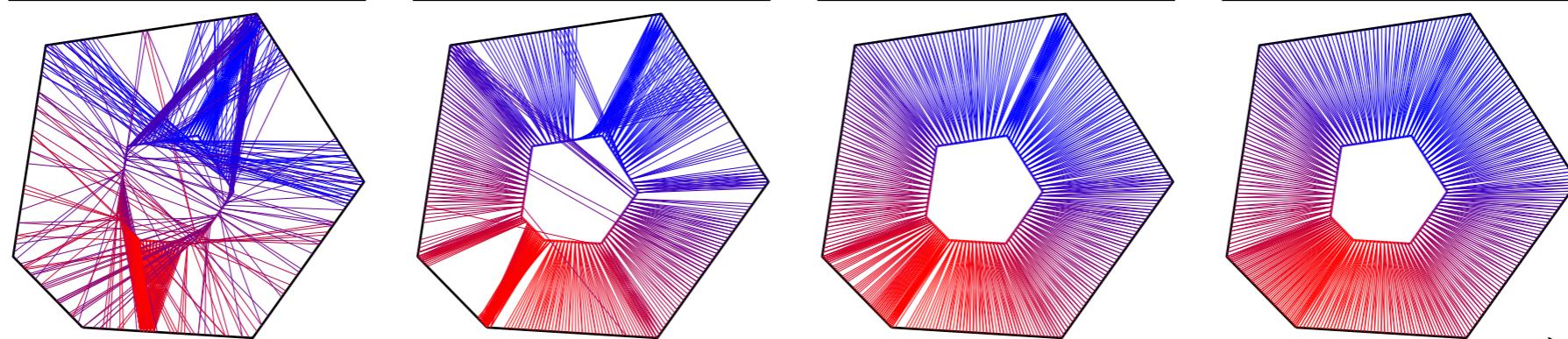
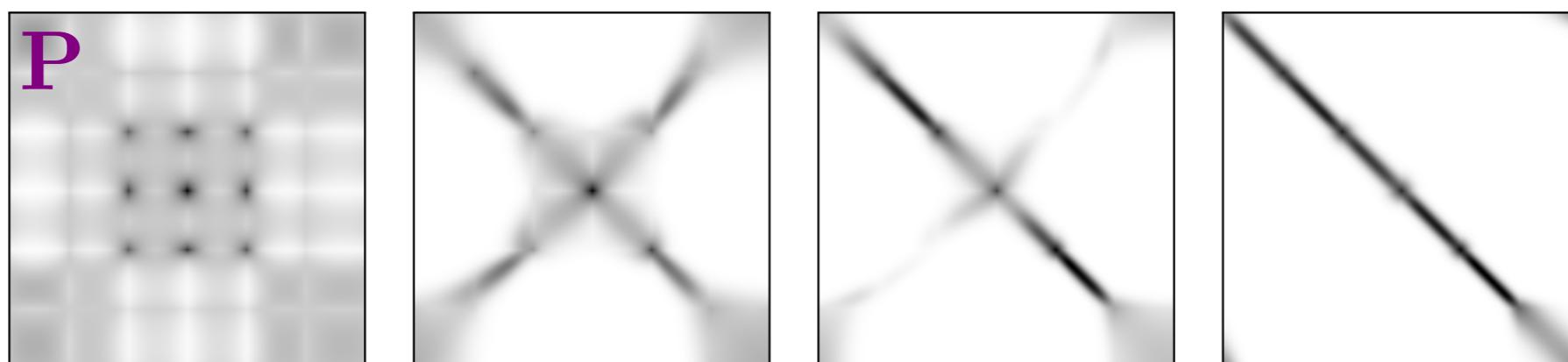
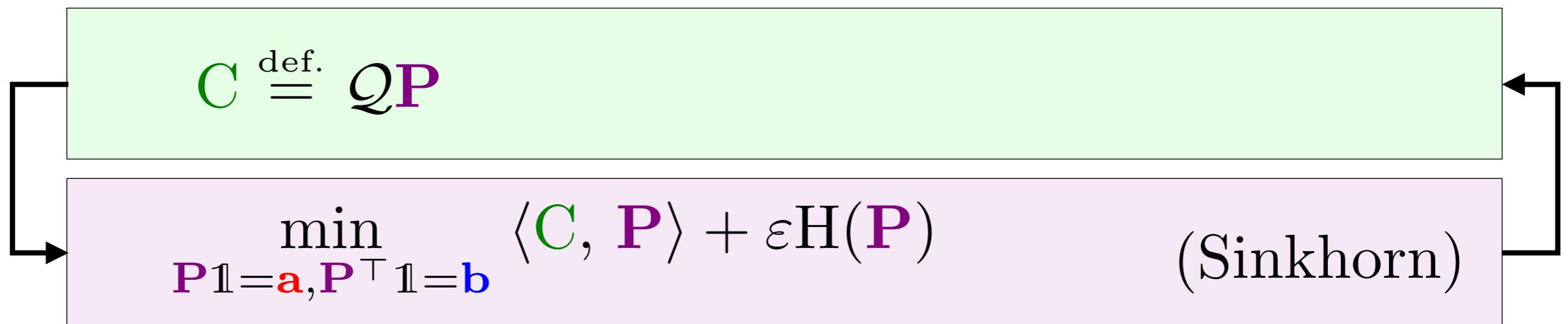
Theorem: GW is a distance up to isometries.



Schrodinger GW

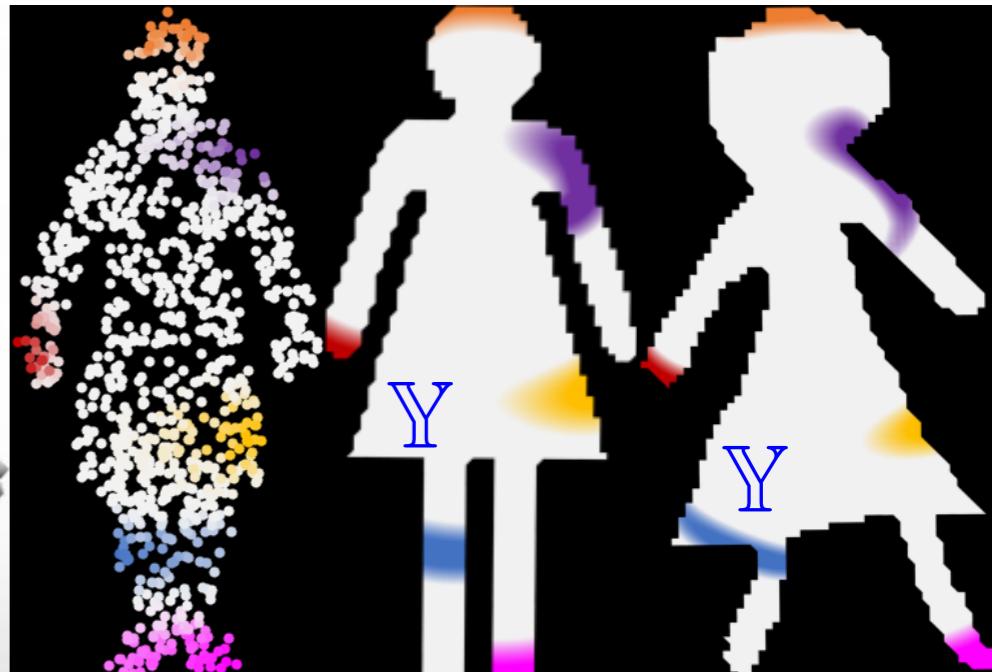
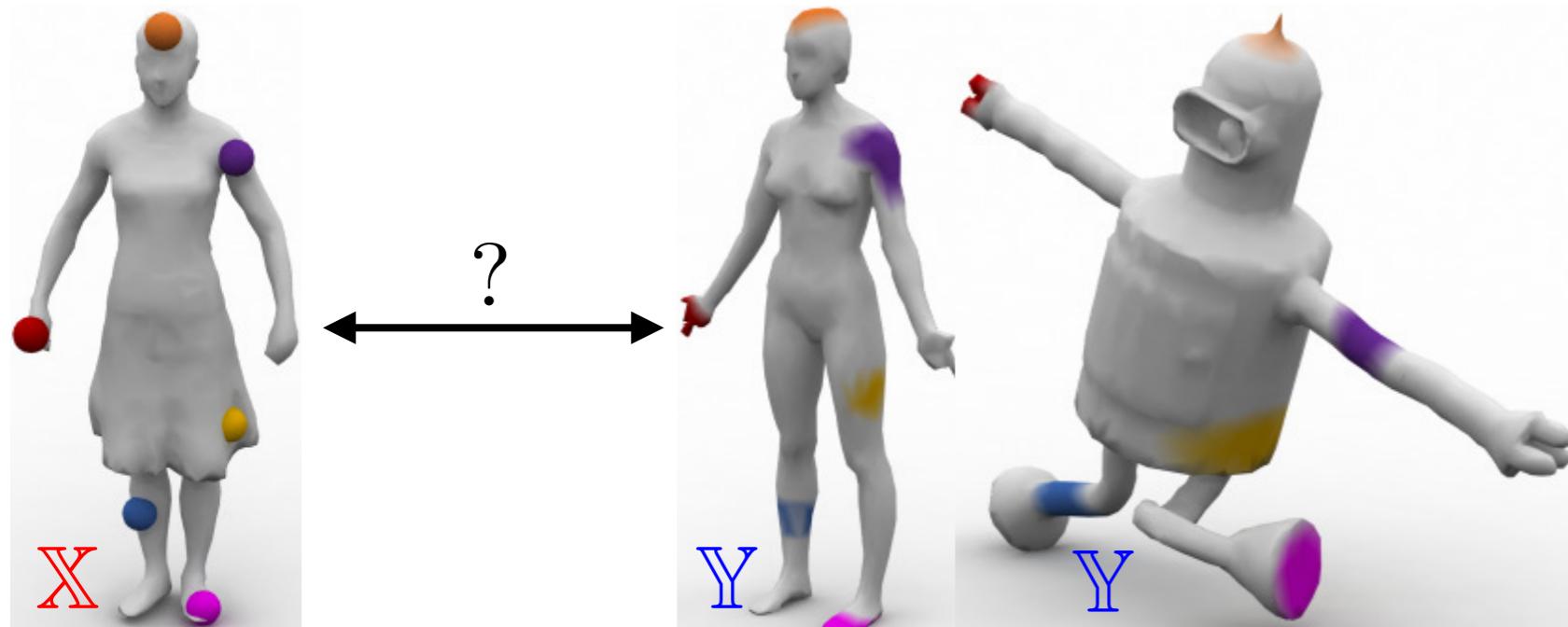
$$\min_{\mathbf{P} \mathbf{1} = \mathbf{a}, \mathbf{P}^\top \mathbf{1} = \mathbf{b}} \langle Q\mathbf{P}, \mathbf{P} \rangle + \varepsilon H(\mathbf{P})$$

DC-programming / Konno's relaxation / Frank-Wolfe / ...



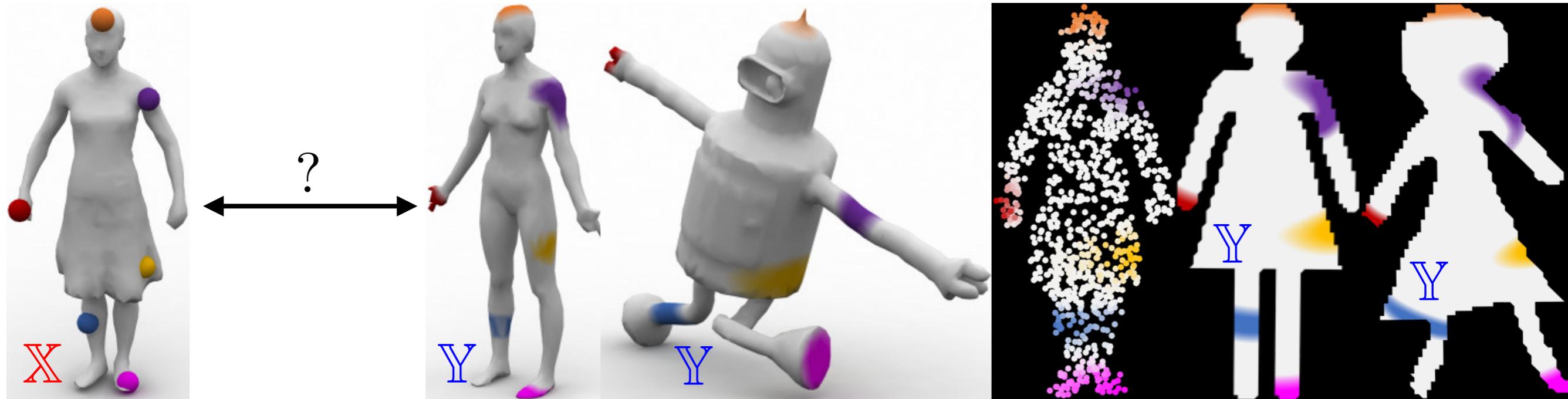
Examples of Applications

Shape registration:

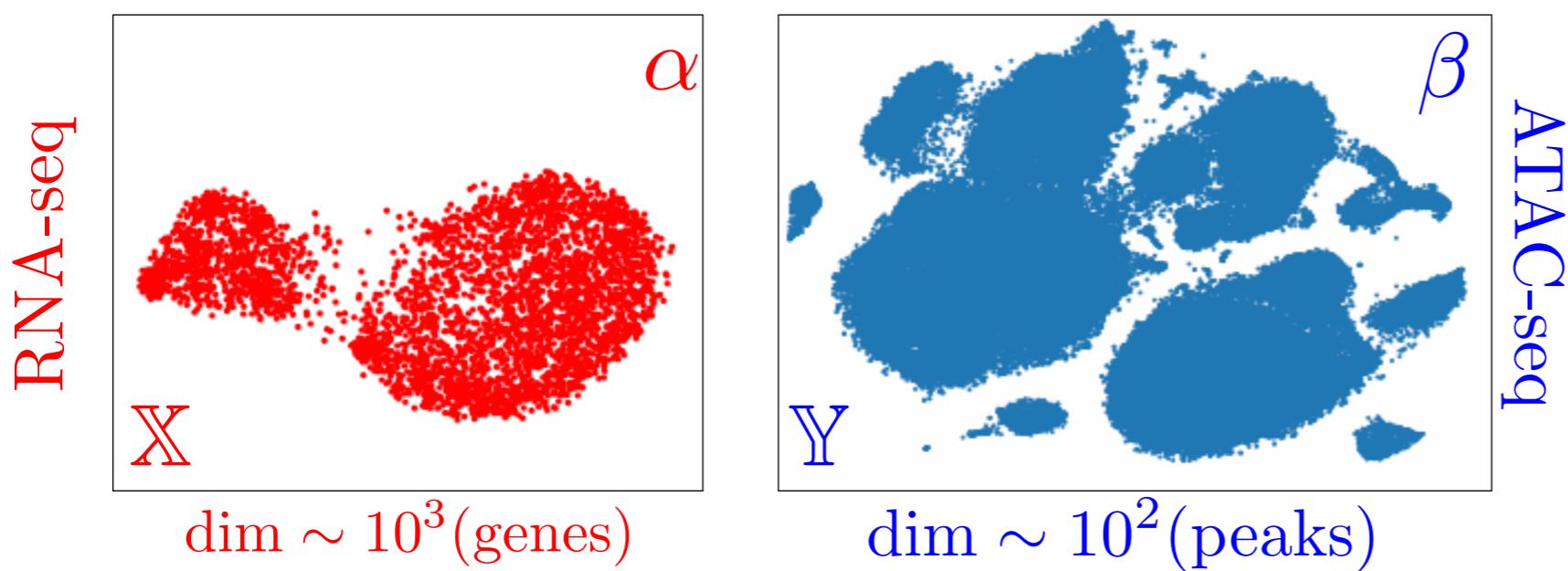


Examples of Applications

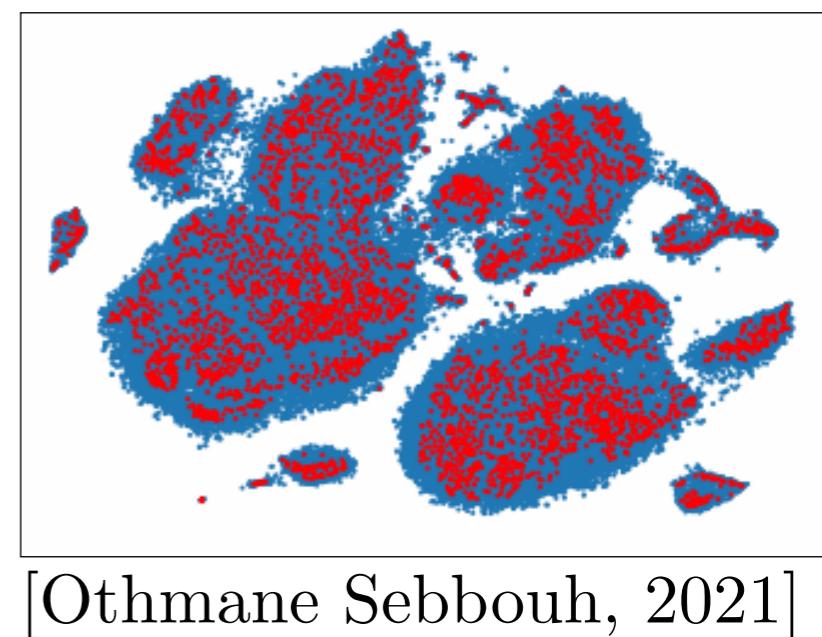
Shape registration:



Single-cell multi-omics



Gromov-Wasserstein
registration



Open Problems!

Toward high-dimensional OT:

- Geometrical properties of the Sinkhorn divergence ?
- Gradient flows (single cell evolution, . . .)

Gromov Wasserstein:

- Existence of Monge maps for $\varepsilon = 0$?
- Taylor expansions when $\varepsilon \rightarrow 0$

