



# Bridging the gap between cellular modalities with Inverse Optimal Transport

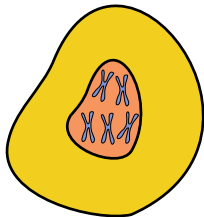
---

Jules Samaran

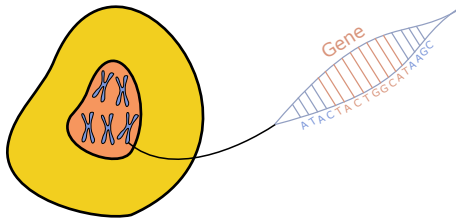
February 17th, 2025

GDR OT, Lyon

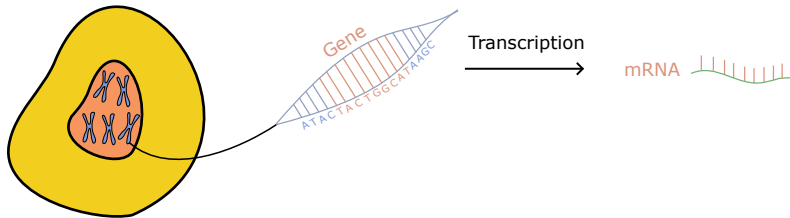
# The central dogma of molecular biology



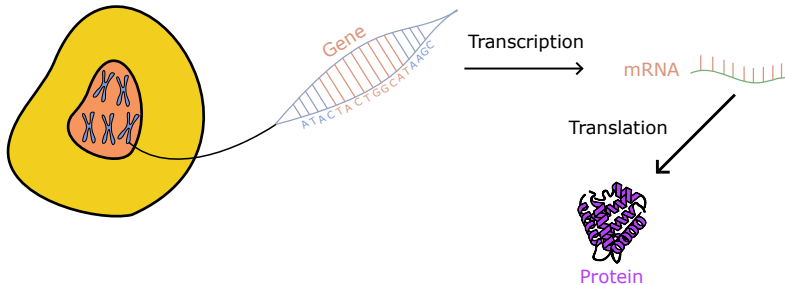
# The central dogma of molecular biology



# The central dogma of molecular biology



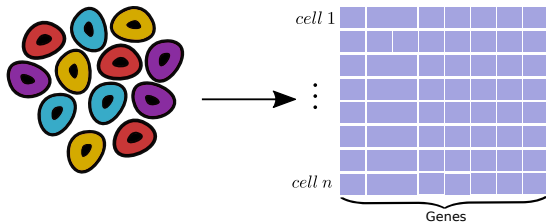
# The central dogma of molecular biology



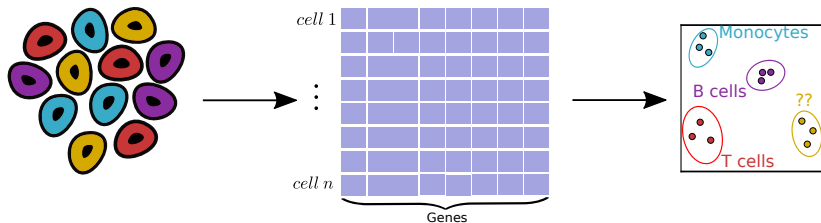
# Single-cell sequencing reveals cellular heterogeneity



# Single-cell sequencing reveals cellular heterogeneity

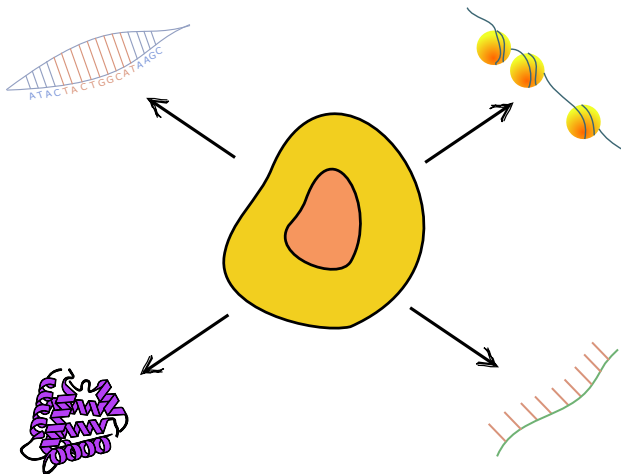


# Single-cell sequencing reveals cellular heterogeneity

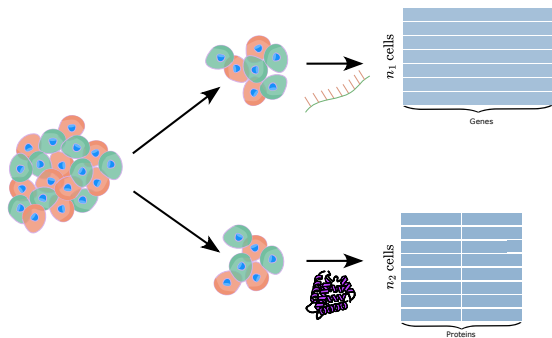




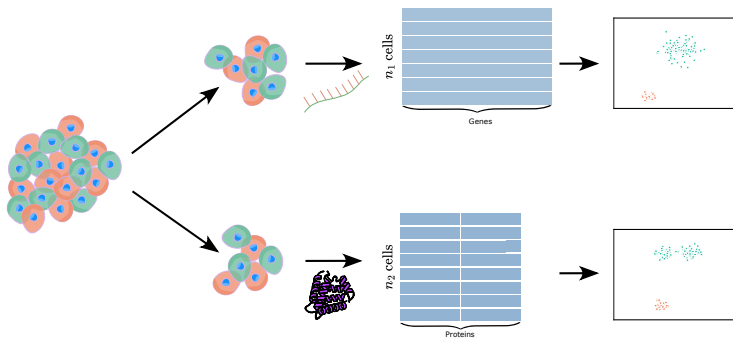
# Cellular biology is multimodal



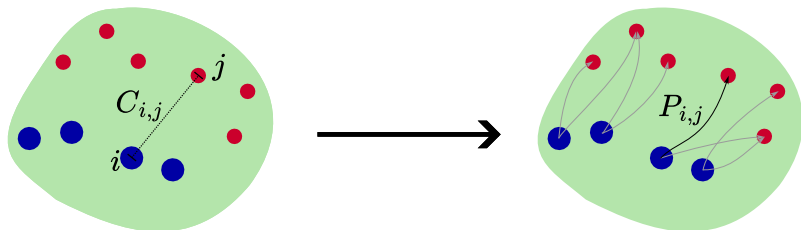
# Unpaired multimodal data are more frequent but challenging



# Common and complementary information across modalities



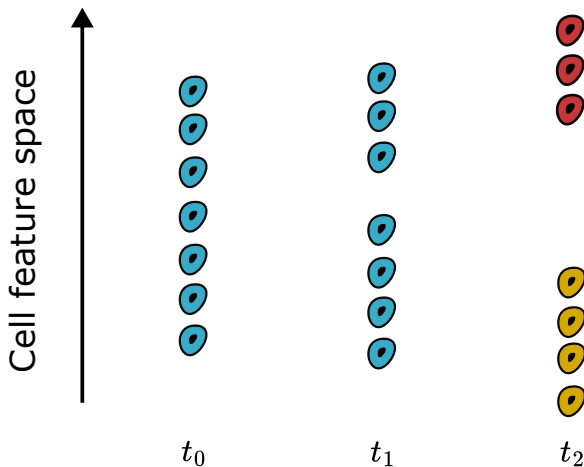
# Using Optimal Transport to compare cell populations



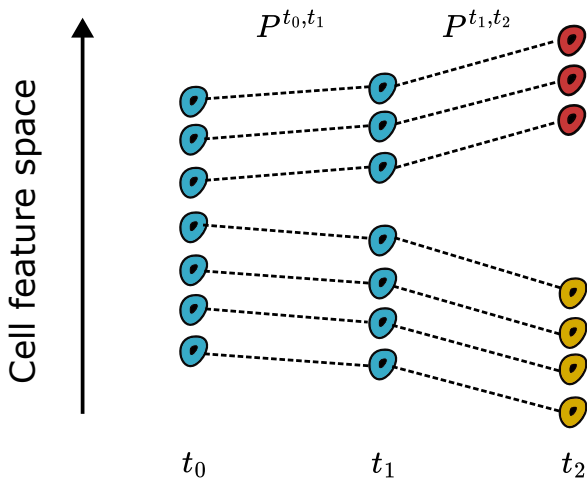
$$\mathcal{W}_\varepsilon(C) \stackrel{\text{def}}{=} \min_{P \in \Pi(n_1, n_2)} \sum_{i,j} P_{i,j} C_{i,j} - \varepsilon H(P)$$

with  $\Pi(n_1, n_2) \stackrel{\text{def}}{=} \{P \in \mathbb{R}_+^{n_1 \times n_2} \text{ s.t. } P\mathbf{1} = \frac{1}{n_1}\mathbf{1}, P^\top\mathbf{1} = \frac{1}{n_2}\mathbf{1}\}$

# OT for genomics: Waddington-OT



# OT for genomics: Waddington-OT



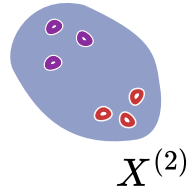
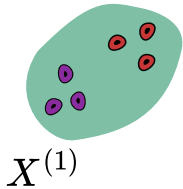
# Inverse Optimal Transport as a metric learning problem

Given a transport plan  $\hat{P}$ , can we find a cost  $C$  for which  $\hat{P}$  is the optimal transport plan?

$$iOT(\hat{P}) \stackrel{\text{def}}{=} \min_{C \in \mathbb{R}_+^{n_1 \times n_2}} \text{KL}(\hat{P} | P(C)) + \mathcal{R}(C)$$

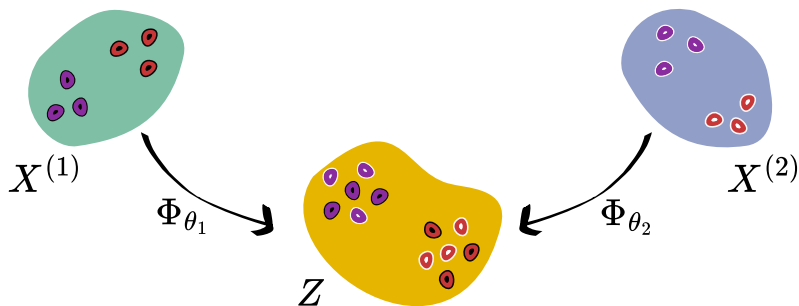
$$\text{with } P(C) = \arg \min_{Q \in \Pi(\mu_1, \mu_2)} \sum_{i,j} Q_{i,j} C_{i,j} - \varepsilon H(Q)$$

# Unpaired data are challenging

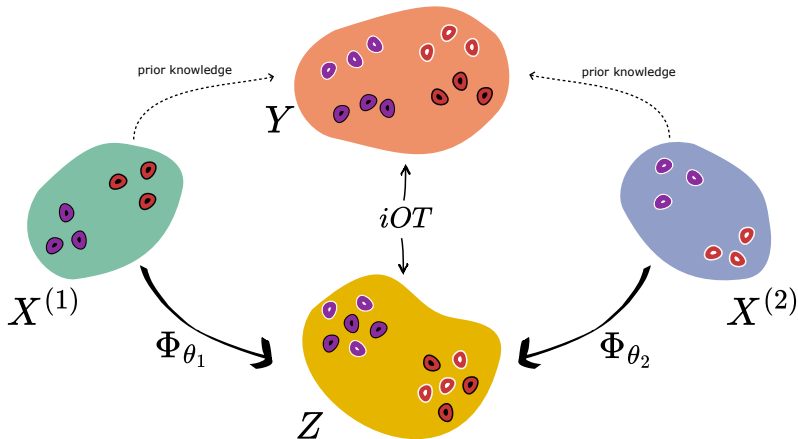




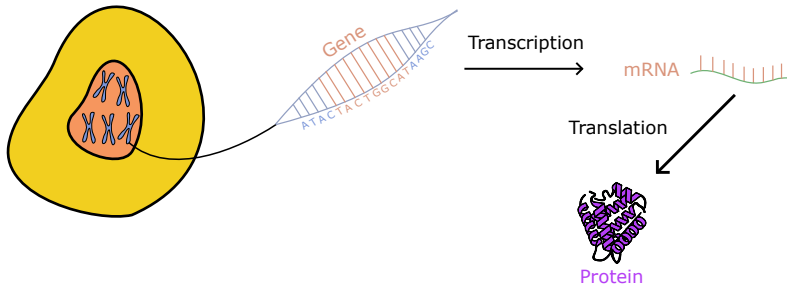
# Finding a shared low dimensional latent space



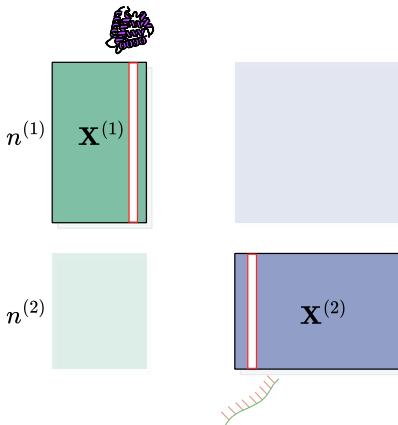
# A summary of our approach



# Prior biological knowledge: related features bridge modalities

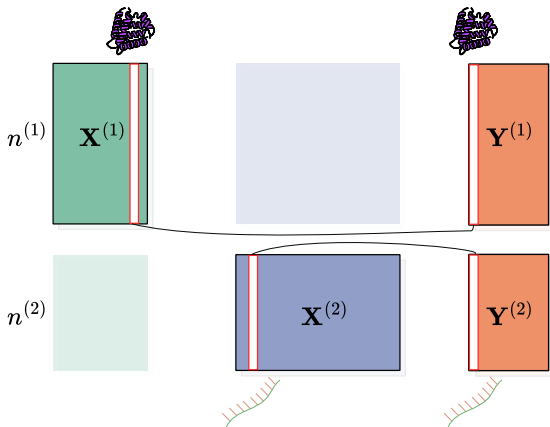


# Prior biological knowledge: related features bridge modalities



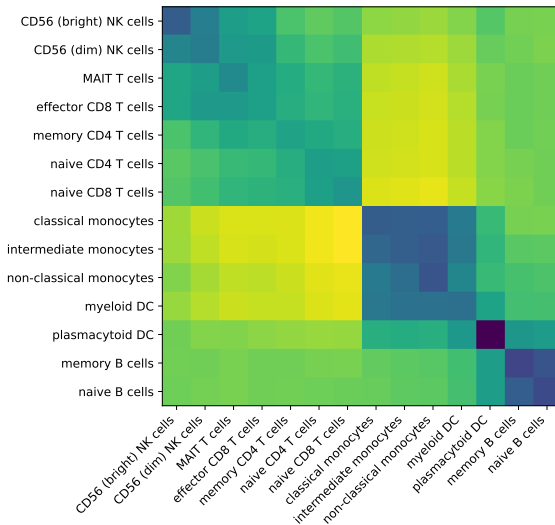
- proteins are related to their corresponding coding gene
- chromatin peaks are related to genes they're close to

# Prior biological knowledge: related features bridge modalities

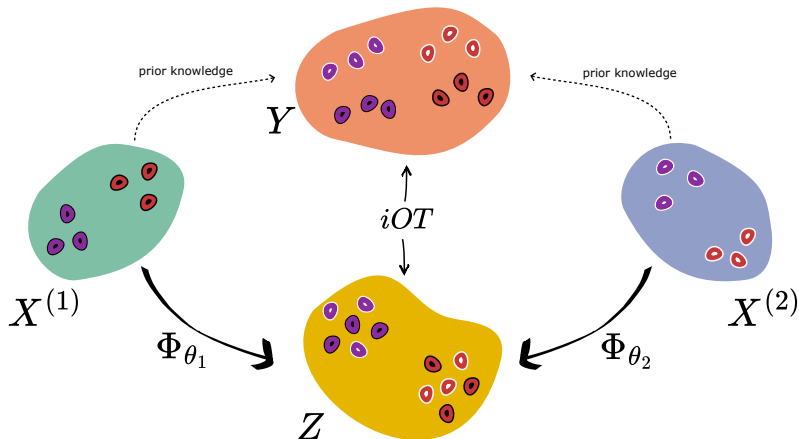


We can use these connections to translate modalities to "common" features

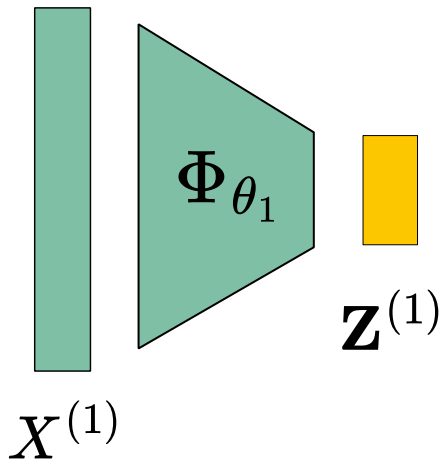
# Prior biological knowledge: related features bridge modalities



# Using autoencoders for dimension reduction

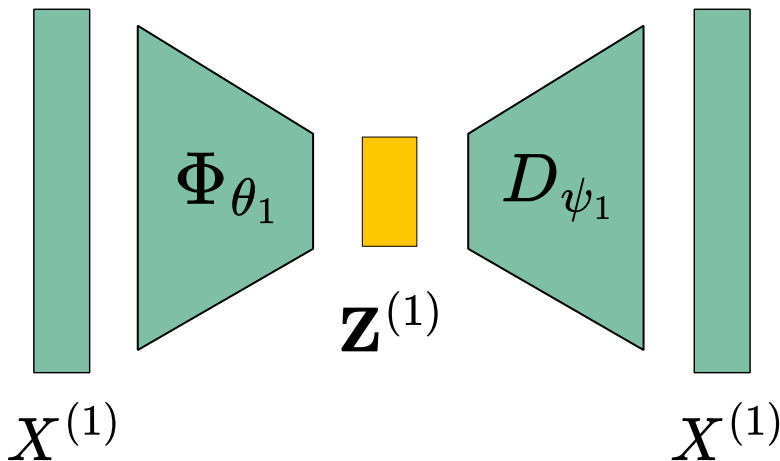


# Using autoencoders for dimension reduction

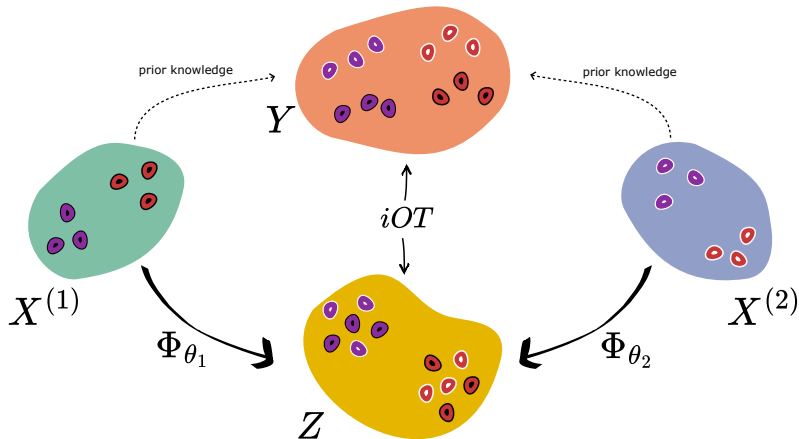




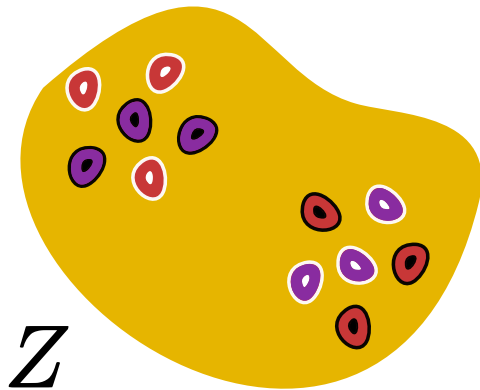
# Using autoencoders for dimension reduction



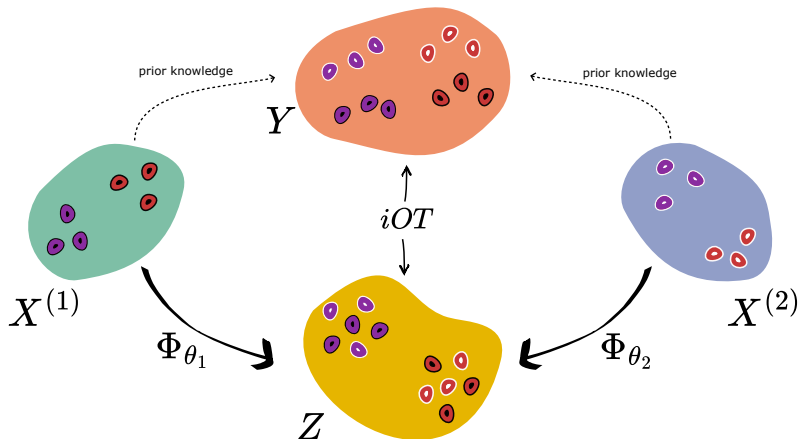
# Using autoencoders for dimension reduction



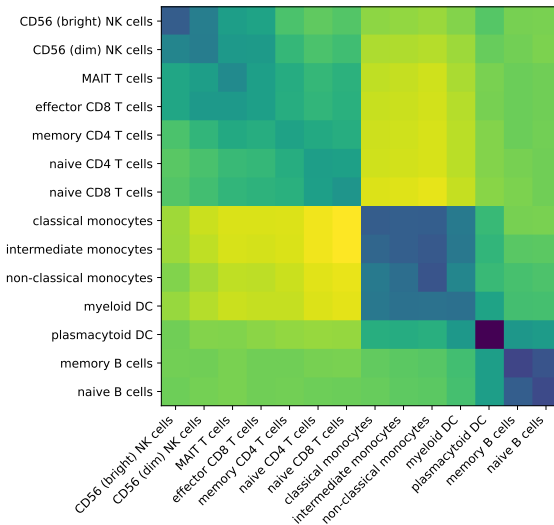
## Distinguishing spurious alignments



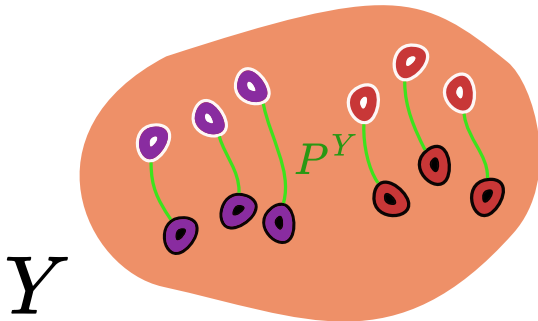
# Matching cells using OT on the prior cost



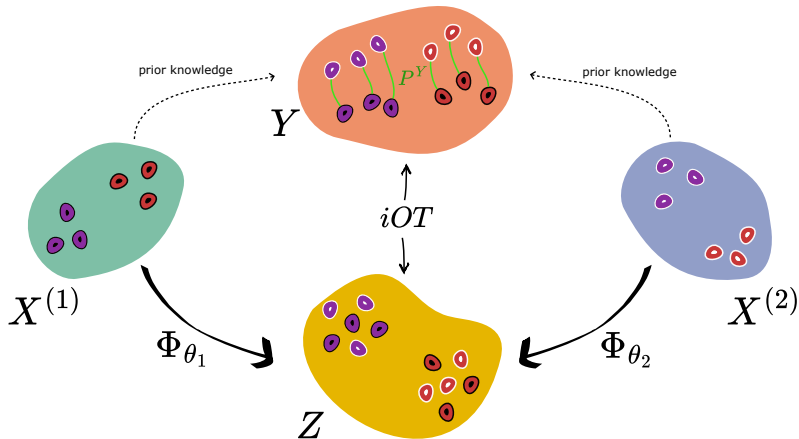
# Matching cells using OT on the prior cost



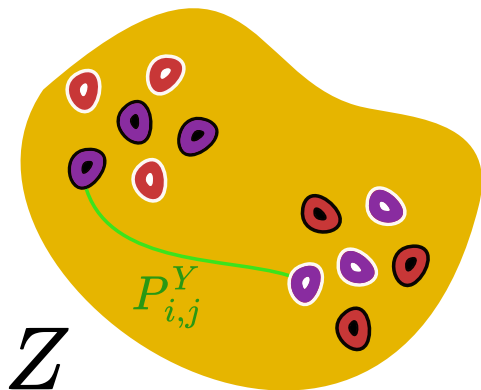
# Matching cells using OT on the prior cost



# Using iOT to align embeddings

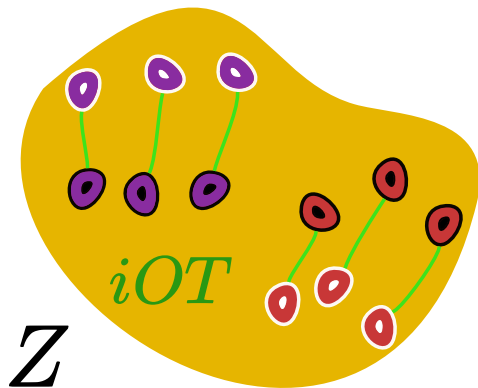


# Using iOT to align embeddings





## Using iOT to align embeddings



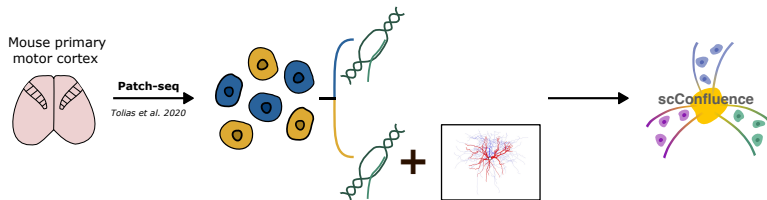
## Using iOT to align embeddings

**Goal:** Enforcing the similarity between  $C^Y$  and  $C^Z$ ,  
 $C_{i,j}^Z = \|Z_i^{(1)} - Z_j^{(2)}\|_2^2.$

**Approach:**

$$\begin{aligned}\mathcal{L}_{\text{align}} &= \text{KL}(P^Y | P^Z) + \mathcal{R}(C^Z) \\ &= \langle P^Y, C^Z \rangle + \lambda_{\text{reg}} \mathcal{W}_\varepsilon(C^Z)\end{aligned}$$

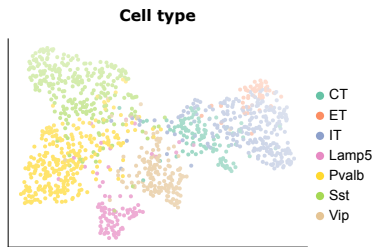
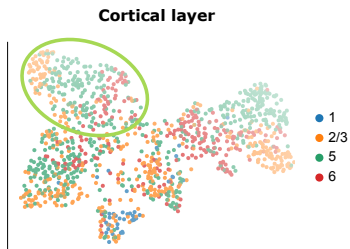
# Integrating scRNA-seq with neuronal morphologies



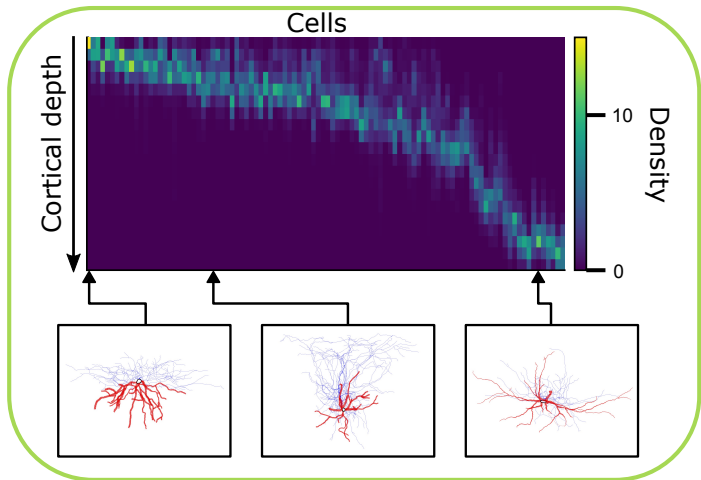
# Integrating scRNA-seq with neuronal morphologies



# Morphological heterogeneity in SST neurons across depths



# Morphological heterogeneity in SST neurons across depths



# scConfluence is now available

Article | [Open access](#) | Published: 05 September 2024

## scConfluence: single-cell diagonal integration with regularized Inverse Optimal Transport on weakly connected features

[Jules Samaran](#), [Gabriel Peyré](#) & [Laura Cantini](#) 

[Nature Communications](#) **15**, Article number: 7762 (2024) | [Cite this article](#)

cantinilab/**scconfluence** 

 Tests passing  codecov 87%  docs passing  pypi v0.1.0  code style black

```
pip install scconfluence
```

## Take home messages

- Massive unpaired datasets could be leveraged by integration
- Single-cell data present extremely diverse challenges
- Prior biological knowledge is essential to tackle these issues



# Thanks!

