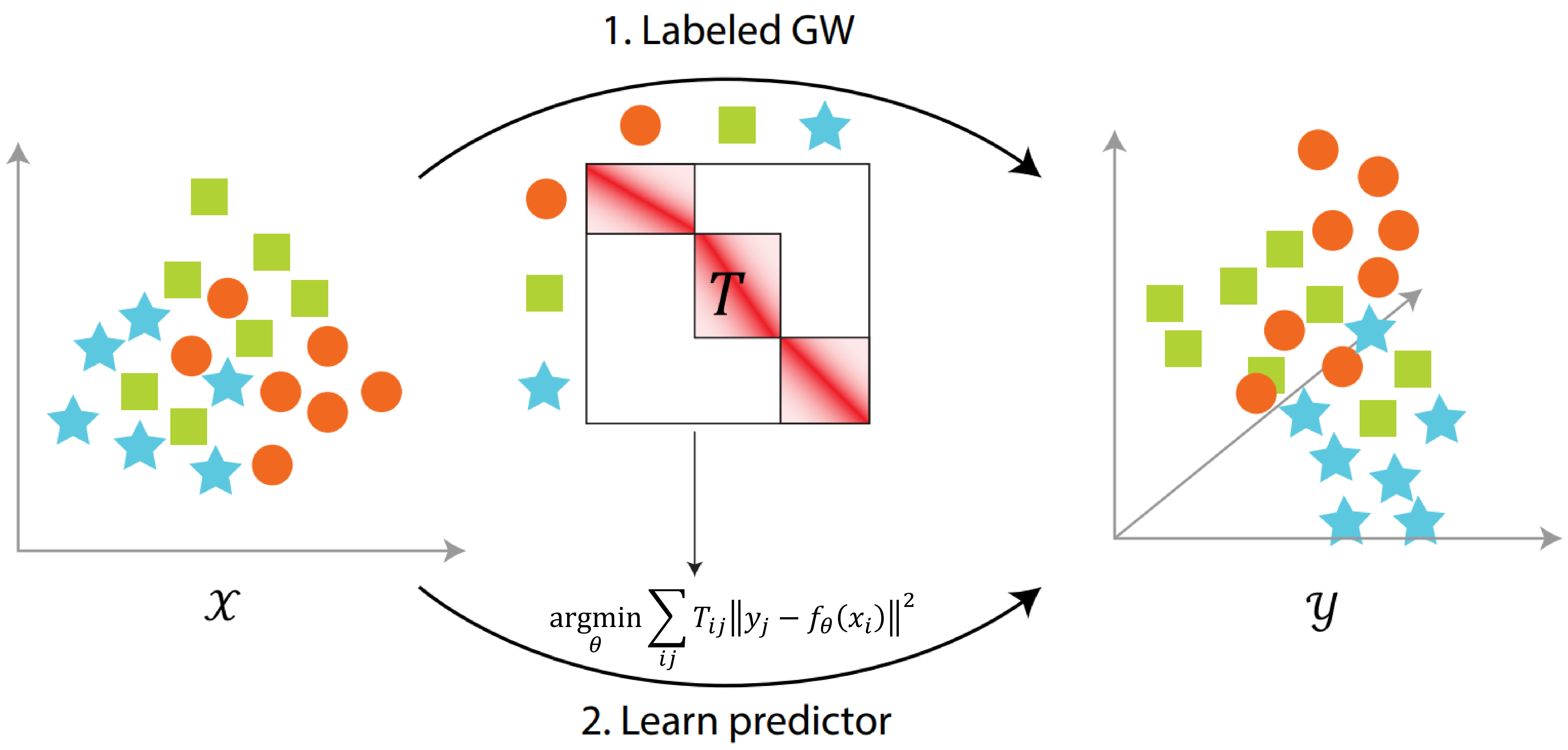
Cross-modality Matching and Prediction of Perturbation Responses with Labeled Gromov-Wasserstein Optimal Transport

ICML AI4Science workshop

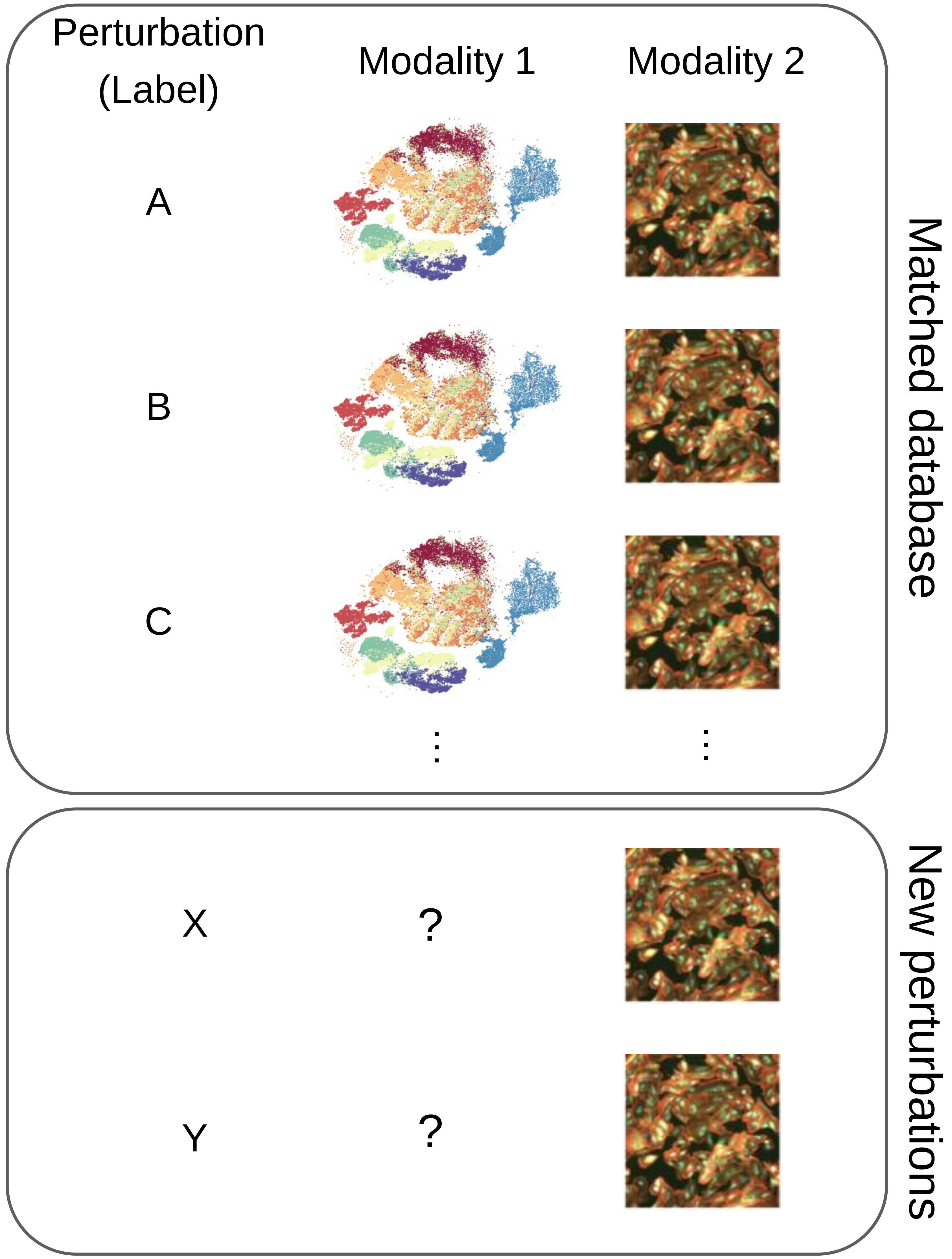


Jayoung Ryu, 7/26/2024

Genentech
A Member of the Roche Group
HARVARD MEDICAL SCHOOL
BLAVATNIK INSTITUTE BIOMEDICAL INFORMATICS

Cross-modality matching & prediction

- Perturbation profiling in different modalities has their own strengths & weaknesses
  - Modality-specific information
  - Scale vs resolution

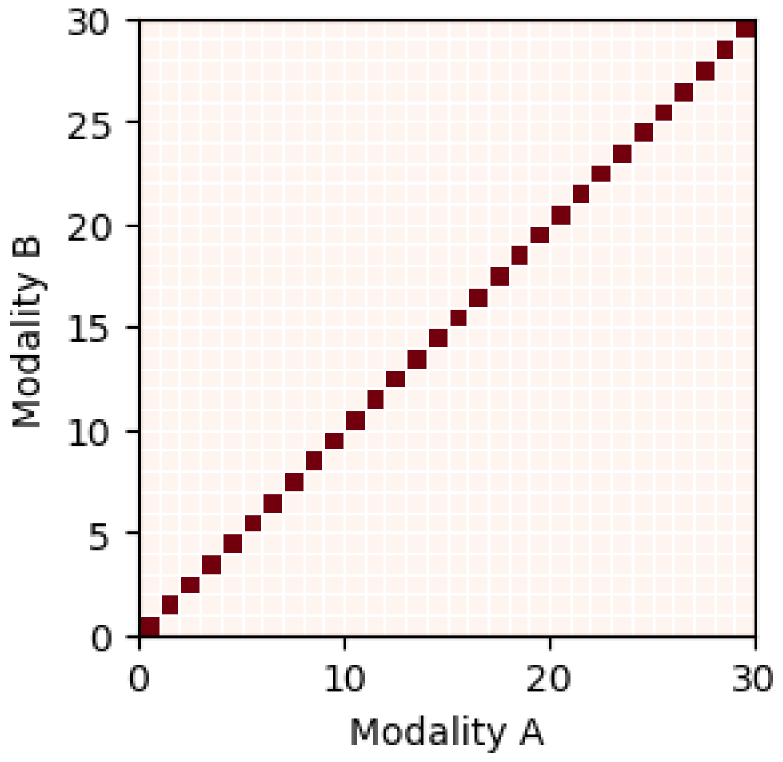


Cross-modality prediction as matching + prediction

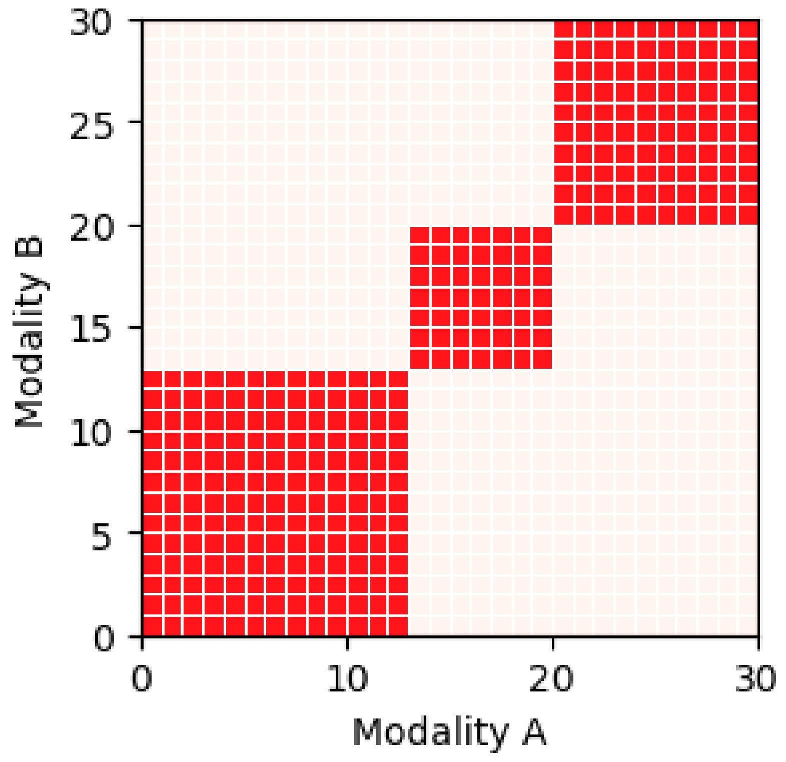
$$\underset{\theta}{\mathrm{arg}min} \sum_{ij} \pi_{ij} (y_j - f_\theta(x_i))^2$$

Matching

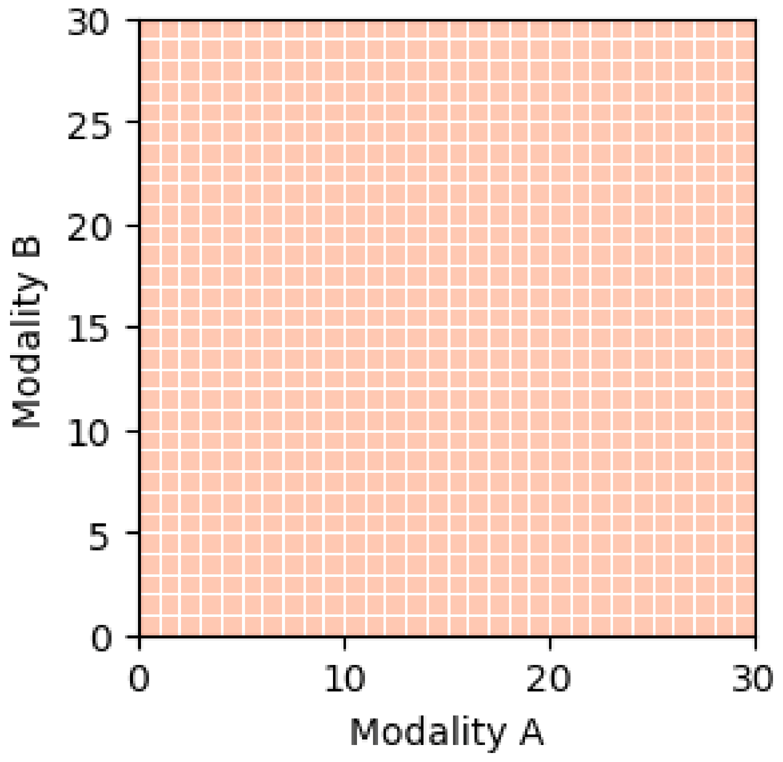
One-to-one



Group-to-group



All-to-all



f

Maps sample to sample

Maps group means to group means

Maps overall mean to overall mean

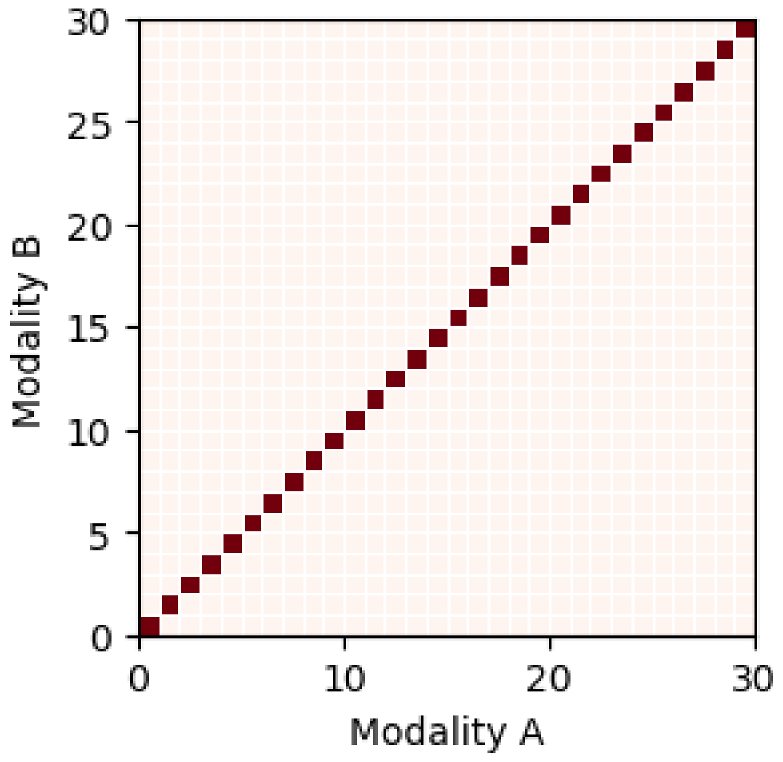
Sample label granularity
Prediction model accuracy

Cross-modality prediction as matching + prediction

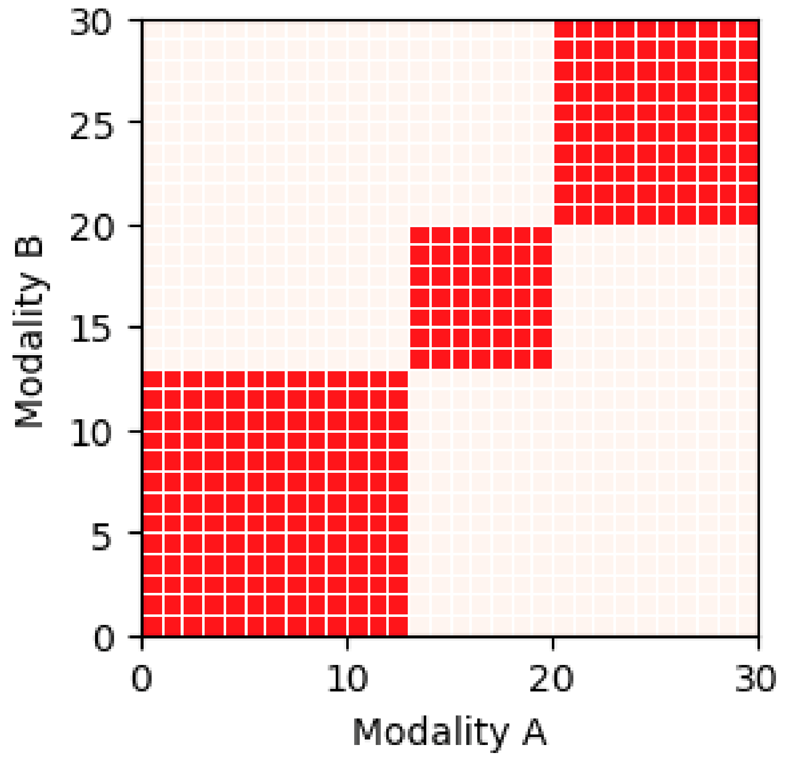
$$\underset{\theta}{\operatorname{argmin}} \sum_{ij} \pi_{ij} (y_j - f_\theta(x_i))^2$$

Matching

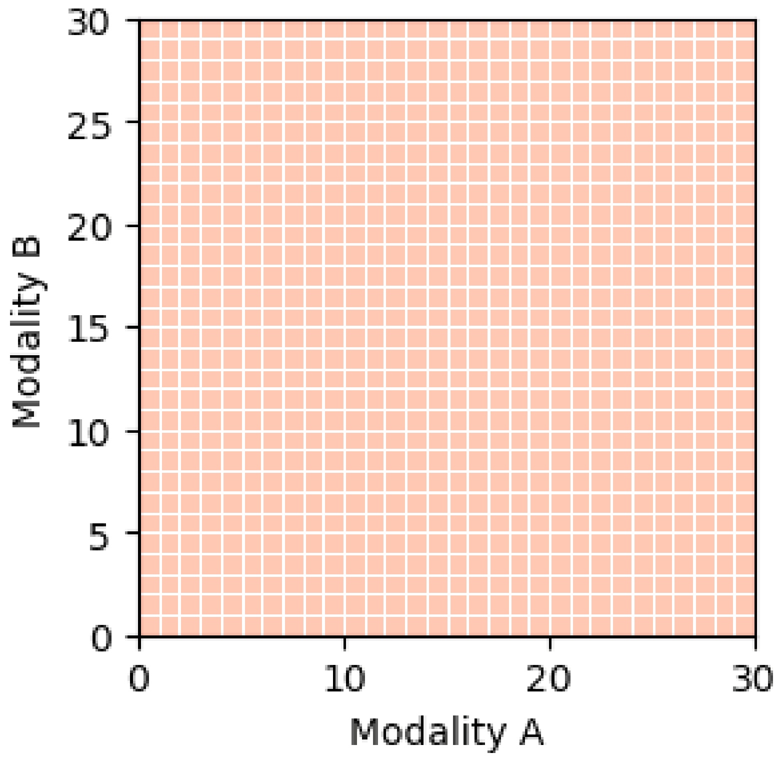
One-to-one



Group-to-group



All-to-all



How good of a matching can we get?
How much does this improve the prediction?

Sample label granularity
Prediction model accuracy

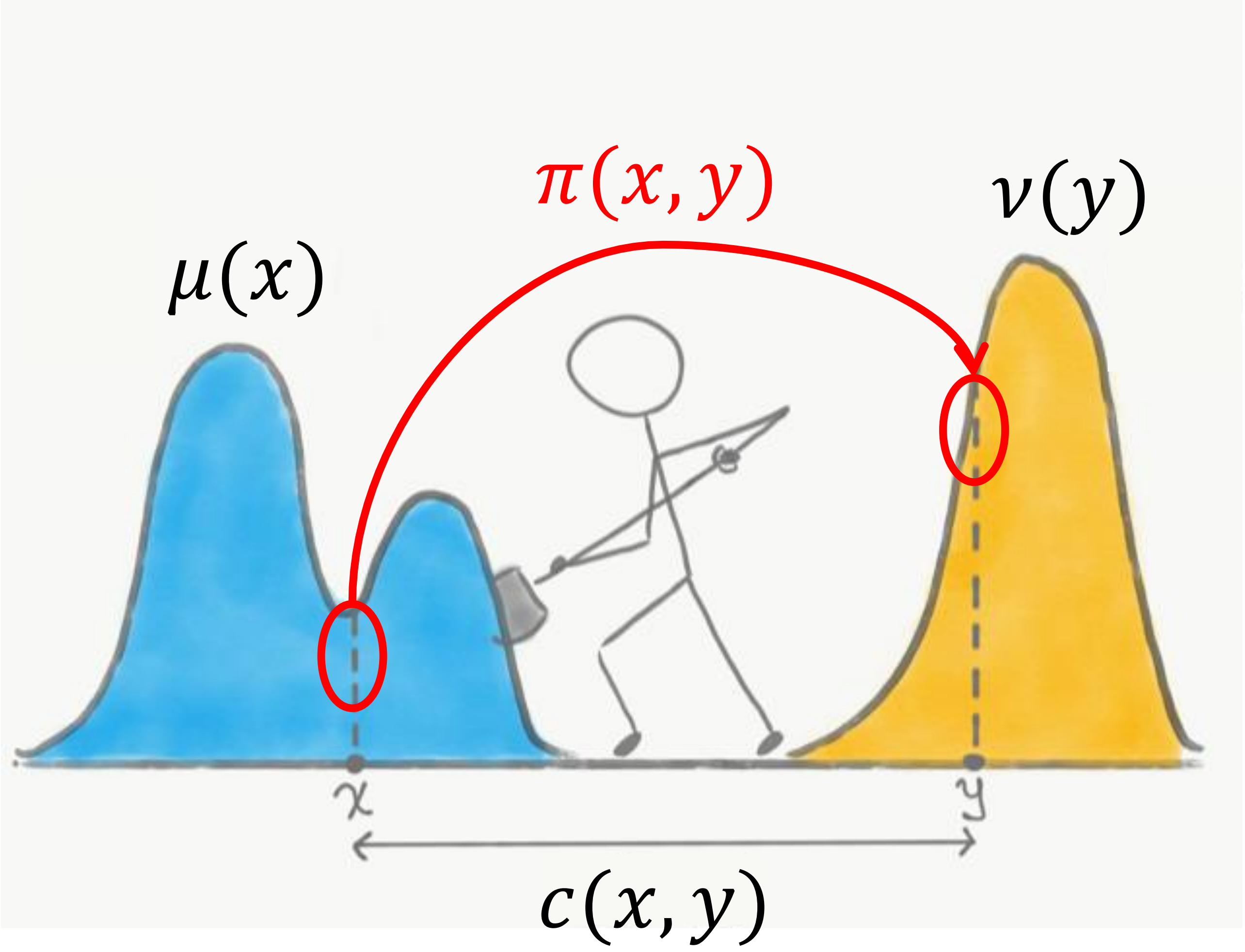
Background: Sample matching with Optimal Transport

- Matching within the **same space**: Optimal Transport

μ, ν: Probability distribution defined over $\mathcal{X}$, $\mathcal{Y}$

$c: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$

$$\Pi(\mu, \nu) = \left\{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}): \int_{\mathcal{Y}} \pi(x, y)\, dy = \mu(x), \int_{\mathcal{X}} \pi(x, y)\, dx = \nu(y)\right\}$$



https://www.microsoft.com/en-us/research/blog/measuring-dataset-similarity-using-optimal-transport/

$$\mathcal{OT}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$

Wasserstein distance
=Total minimum cost

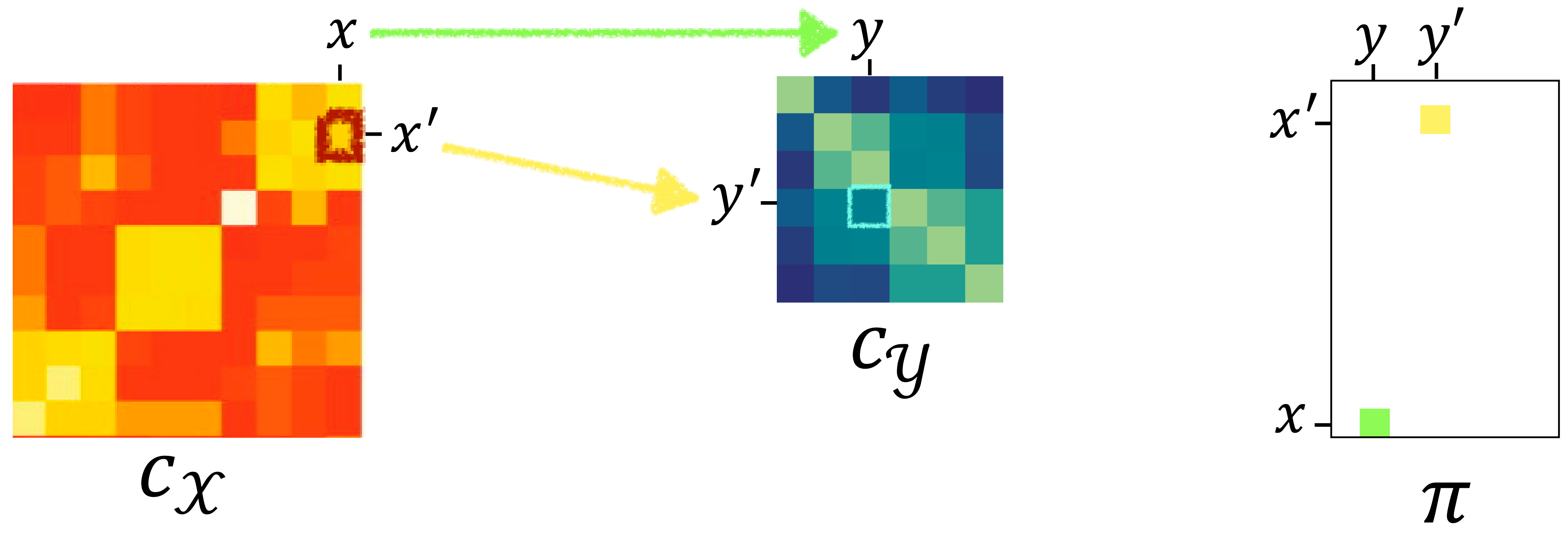
π^*: optimal [transport plan/coupling]

Background: Matching across different spaces

Gromov-Wasserstein Optimal Transport

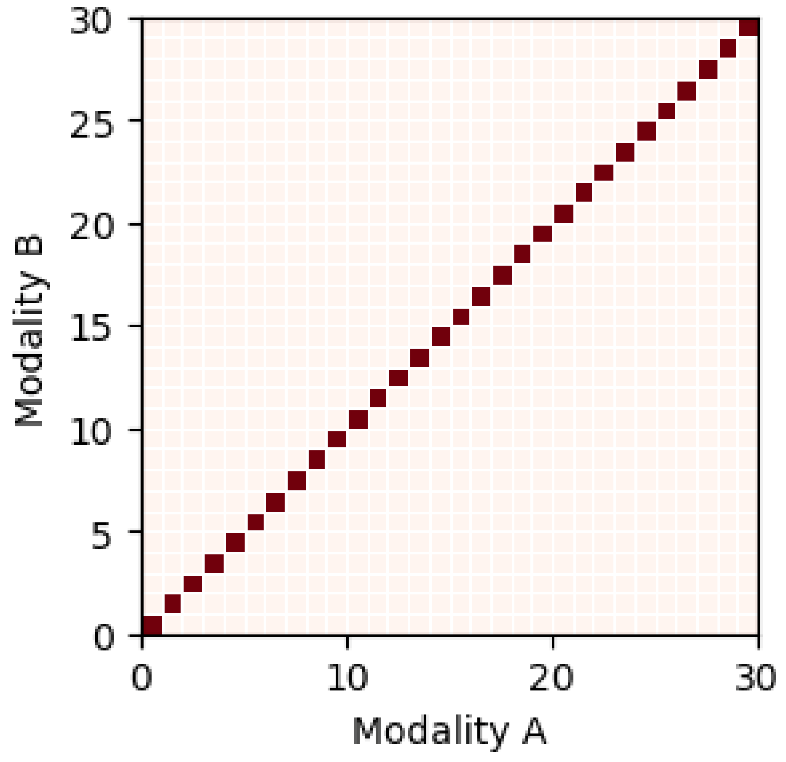
- Cost function $c: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$ does not make sense for $\mathcal{X}$ and $\mathcal{Y}$ are in different spaces.
- Instead of minimizing total distances of moving $\mathcal{X}$ to $\mathcal{Y}$, find π that minimizes differences of the within-space cost.

$$\mathcal{GW}(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}}\int_{\mathcal{X}\times\mathcal{Y}} c(c_{\mathcal{X}}(x,x'), c_{\mathcal{Y}}(y,y'))d\pi(x,y) \otimes \pi(x',y') \ -\epsilon H(\pi)$$

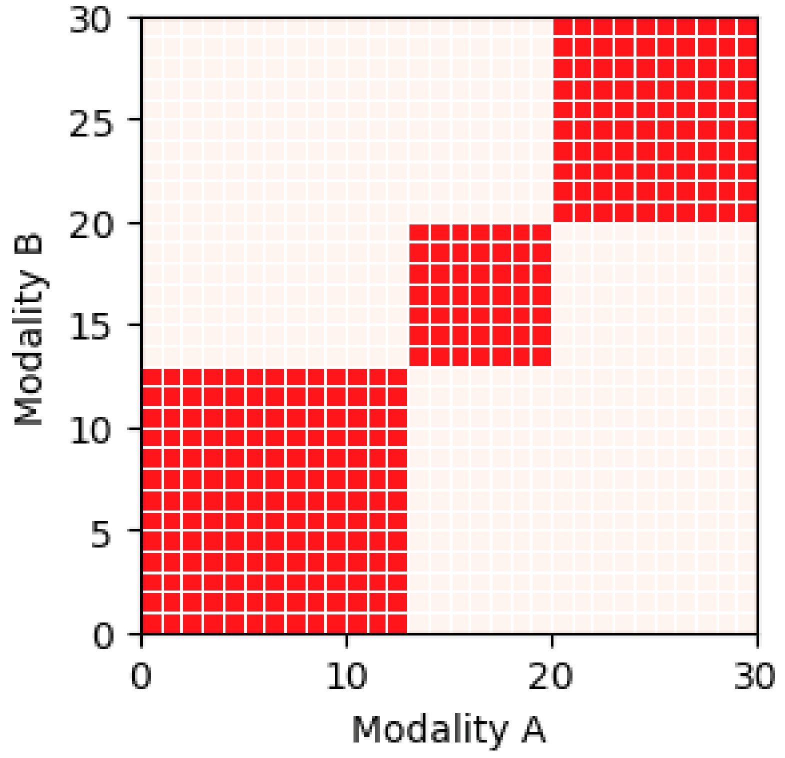


Cross-modality prediction as matching + prediction

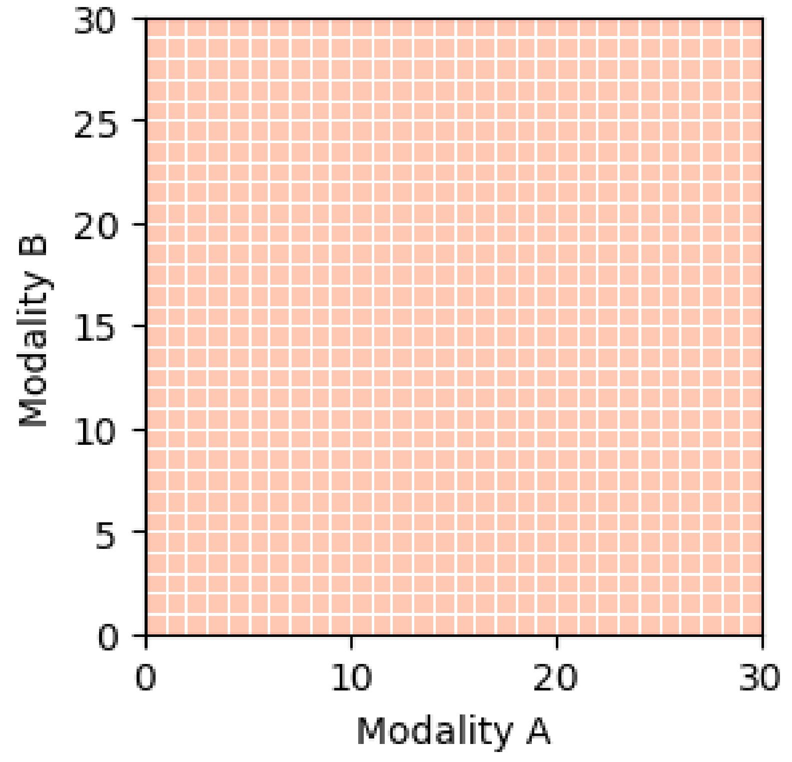
One-to-one

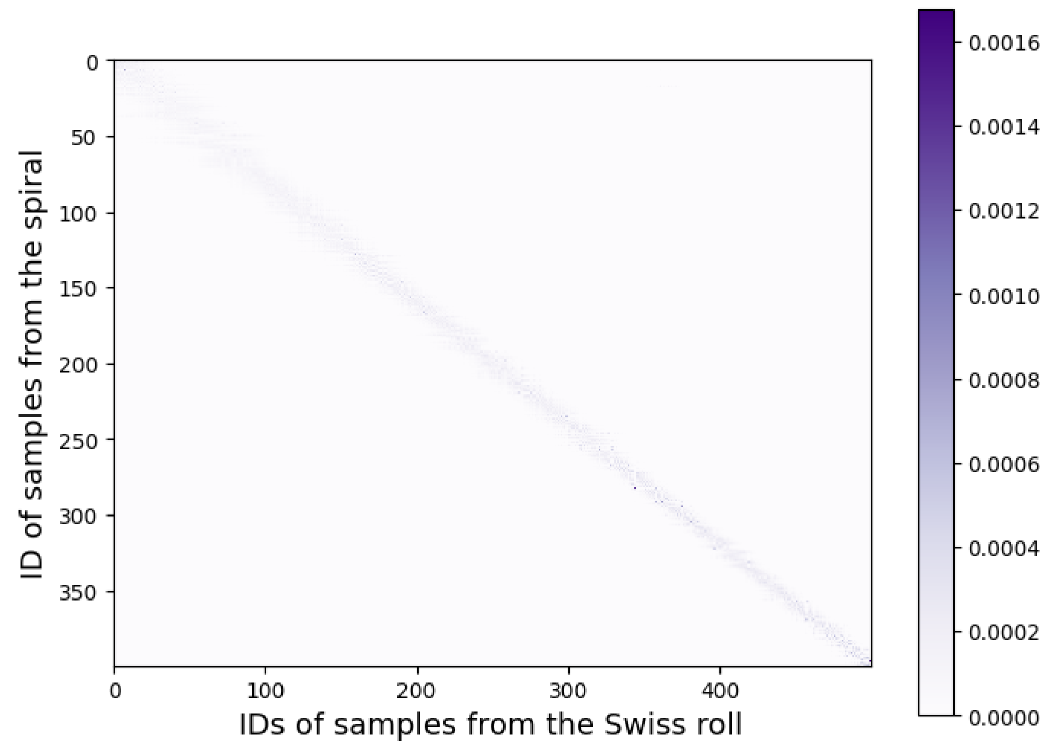


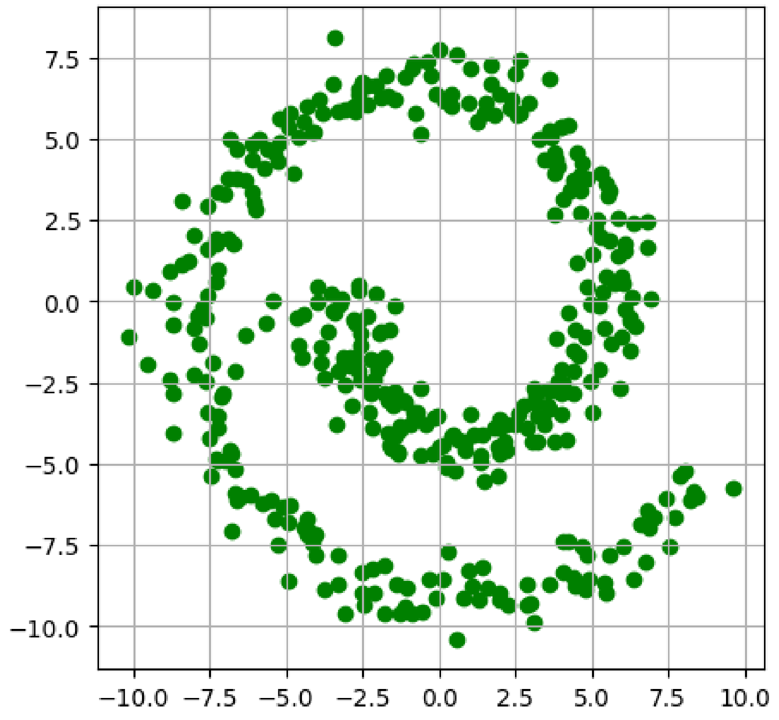
Group-to-group

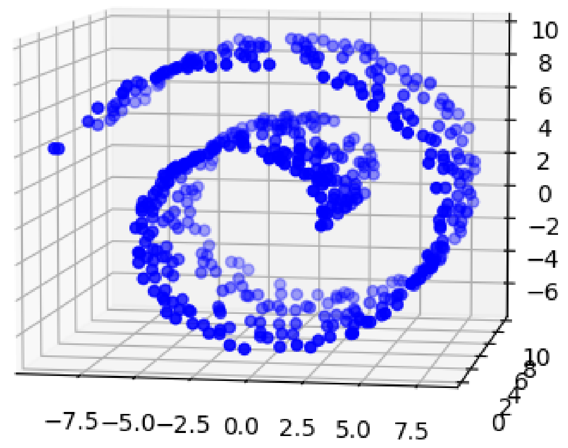


All-to-all









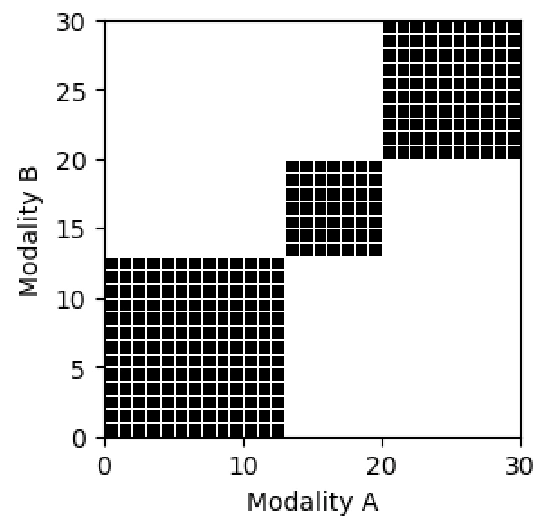
https://ott-jax.readthedocs.io/en/latest/tutorials/gromov_wasserstein.html

Labeled Entropic Gromov-Wasserstein Optimal Transport

Contribution:

- Entropic GW with labels will solve OT problem with constrained Π^l.
- We show **Π^l-constrained OT still can be solved with Sinkhorn-Knopp algorithm**.
- Implemented in OTT-jax

$$\pi^{k+1} \leftarrow \underset{\pi \in \Pi^l}{\operatorname{argmin}} \int_{\mathcal{X}\times\mathcal{Y}} f\left(c_X, c_Y, \pi^k\right)\pi(x,y) - \epsilon H(\pi)$$



$$B = \left[l_{x_i} = l_{y_j}\right]_{ij}$$

$$\Pi^l = \{\pi \in \Pi | \pi_{ij} > 0 \ \Rightarrow B_{ij} = 1\}$$

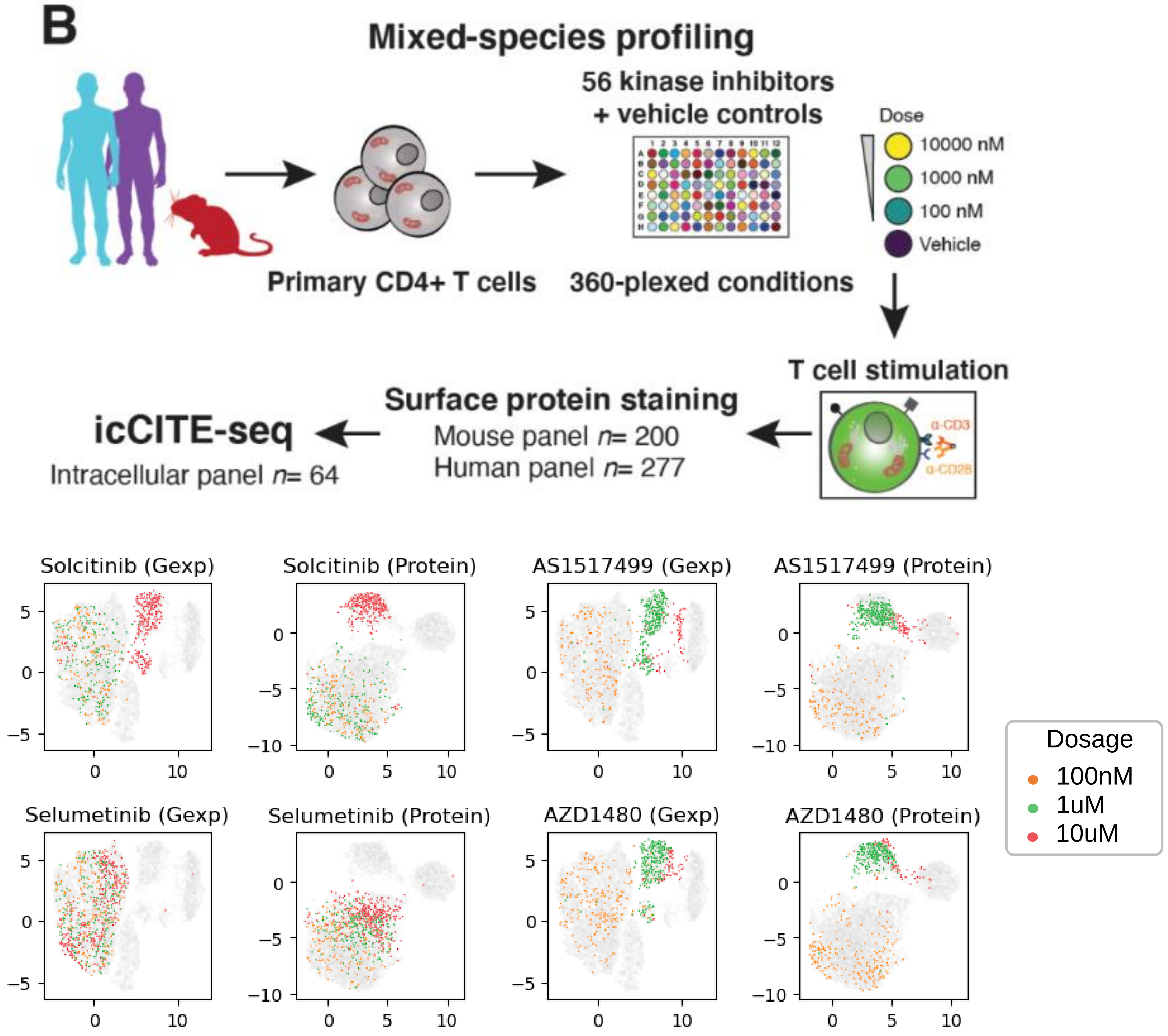
Algorithm 1 Computation of l-constrained coupling for EGW

Input: $C_{\mathcal{X}}, C_{\mathcal{Y}}, \epsilon, B, p, q$
Initialize T.
repeat
// compute $c_s = c(C_{\mathcal{X}}, C_{\mathcal{Y}}) \otimes T$ as in (2).
$c_s \leftarrow \sum_{k=1}^{L} c^k_{C_{\mathcal{X}},C_{\mathcal{Y}}} - h_1(C^k_{\mathcal{X}})T^k h_2(C^k_{\mathcal{Y}})^T$.
// Sinkhorn iterations to compute $\mathcal{T}^l(c_s, p, q)$
Initialize $a \leftarrow \mathbb{1}$, set $K \leftarrow e^{-c_s/\epsilon} \otimes B$.
repeat
$b \leftarrow \frac{q}{K^T a}, a \leftarrow \frac{q}{Kb}$
until convergence
Update $T \leftarrow diag(a)(e^{-c_s/\epsilon} \otimes B)diag(b)$
until convergence

Experiment on Perturb-CITE-seq dataset

In collaboration with Dr. Kelvin Chen, Osaka University

- 13 selected kinase inhibitors with large effects, total 8486 cells
  - 3 dosages: 100nM, 1uM, 10uM
- Predicting 2000 genes (with highest variability) from 123 proteins



Results (Perturb-CITE-seq)

Label-aware approach works better than label-agnostic or per-label matching

Table 1. Evaluation metrics of OT and GW approaches for sample matching, prediction, and feature matching tasks.

Method		Matching			Prediction						Feature
		Bary FOSCTTM (↓)	Dosage match (↑)	Mean rank	R_v (↑)	ρ_v (↑)	R_s (↑)	ρ_s (↑)	MSE (↓)	Mean rank	Enrich-ment(↑)
Perfect		0	1	-	0.107	0.118	0.163	0.149	0.258	-	6.95
By dosage		0.239	1	-	0.0812	0.0448	0.0903	0.0863	0.264	-	5.16
Uniform per label		0.298	0.357	-	0.0794	0.0403	0.0761	0.0781	0.264	-	1.85
EOT	no label	0.428	0.040	9	0.0482	0.007	0.0068	0.0063	0.287	7	1.10
	per label	0.336	0.346	5	0.0544	0.0239	0.0345	0.0307	0.283	5.2	1.26
ECOOT	no label	0.414	0.049	8	0.053	0.0207	0.0395	0.0408	0.282	5	1.07
	labeled	0.270	**0.456**	2	**0.0852**	**0.0523**	**0.0854**	<u>0.0778</u>	<u>0.265</u>	1.6	<u>5.31</u>
EGWOT	no label	0.373	0.068	7	0.0631	0.0227	0.0302	0.034	0.282	4.8	3.74
	per label	0.332	0.381	4	0.0785	<u>0.0449</u>	0.0737	0.0737	0.265	2.6	1.26
	labeled	0.283	<u>0.452</u>	3	<u>0.0836</u>	0.044	**0.0854**	**0.0825**	**0.264**	1.8	**19.8**
DAVAE	no label	**0.231**	0.206	3	0.0342	-0.0069	0.0006	-0.0001	0.33	8	-
	labeled	<u>0.242</u>	0.205	4	0.0182	-0.0079	-0.0016	-0.0014	0.332	9	-

Takeaways

- Labeled GW improves matching & prediction results for coarse-labeled data
- **Input for matching**
  - Cost matrix should be valid, major modality-specific should be removed prior to GW
    - Imaging modality & sequencing-specific variations
    - Effect of different latent representations to calculate matching (representation learning of images)
- **Interpretability**
  - Learn feature-feature transport in the raw space as in Co-OT, based on learned sample-sample matching

Takeaways

- Labeled GW improves matching & prediction results for perturbation data
- **Input for matching**
  - Cost matrix should be valid, major modality-specific should be removed prior to GW
    - Imaging modality & sequencing-specific variations
    - Effect of different latent representations to calculate matching (representation learning of images)
- **Interpretability**
  - Learn feature-feature transport in the raw space as in Co-OT, based on learned sample-sample matching

Takeaways

- Labeled GW improves matching & prediction results for perturbation data
- **Input for matching**
  - Cost matrix should be valid, major modality-specific should be removed prior to GW
    - Imaging modality & sequencing-specific variations
    - Effect of different latent representations to calculate matching (representation learning of images)
- **Interpretability**
  - Learn feature-feature transport in the raw space as in Co-OT, based on learned sample-sample matching

Acknowledgements

Romain Lopez

Taka Kudo

Charlotte Bunne

Aviv Regev

Luca Pinello

Thanks for listening!

Paper

Perturb-OT