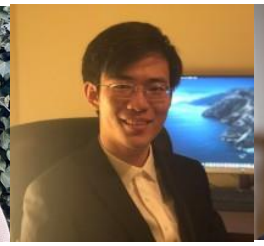
RECOMB
Seoul | 서울 | 2025

Learning Latent Trajectories in Developmental Time Series with Hidden-Markov OT

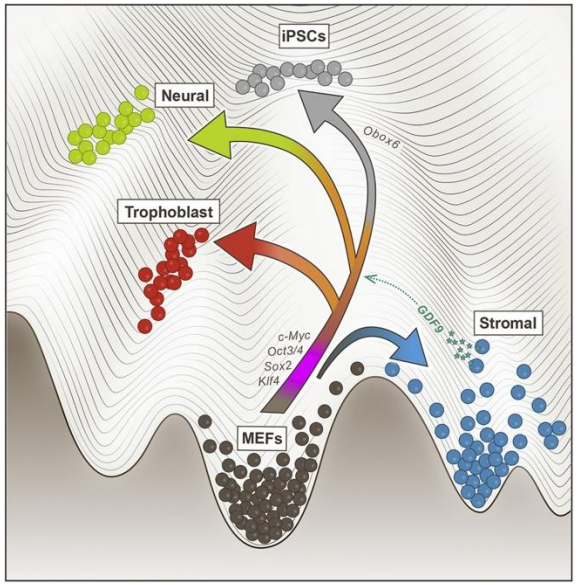
Peter Halmos*, **Julian Gold***, Xinhao Liu, and Ben Raphael

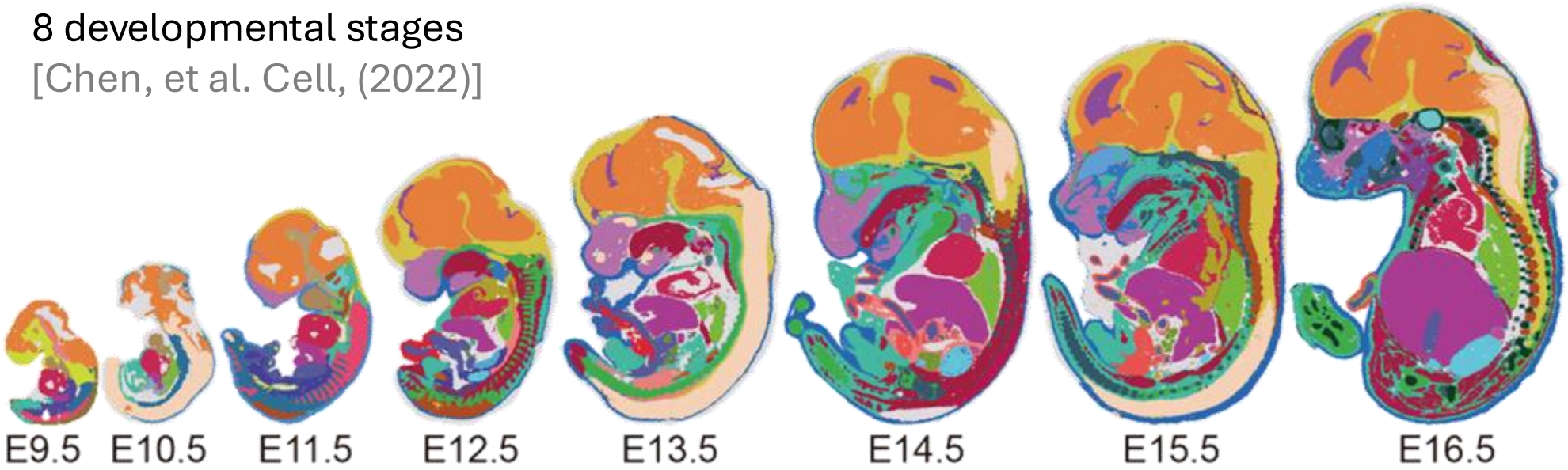

PRINCETON UNIVERSITY

Temporal and Spatiotemporal transcriptomics: Sequencing across multiple time points during developmental and reprogramming processes

Reprogramming of fibroblasts to induced pluripotent stem cells [Schiebinger, et al. Cell, (2019)]



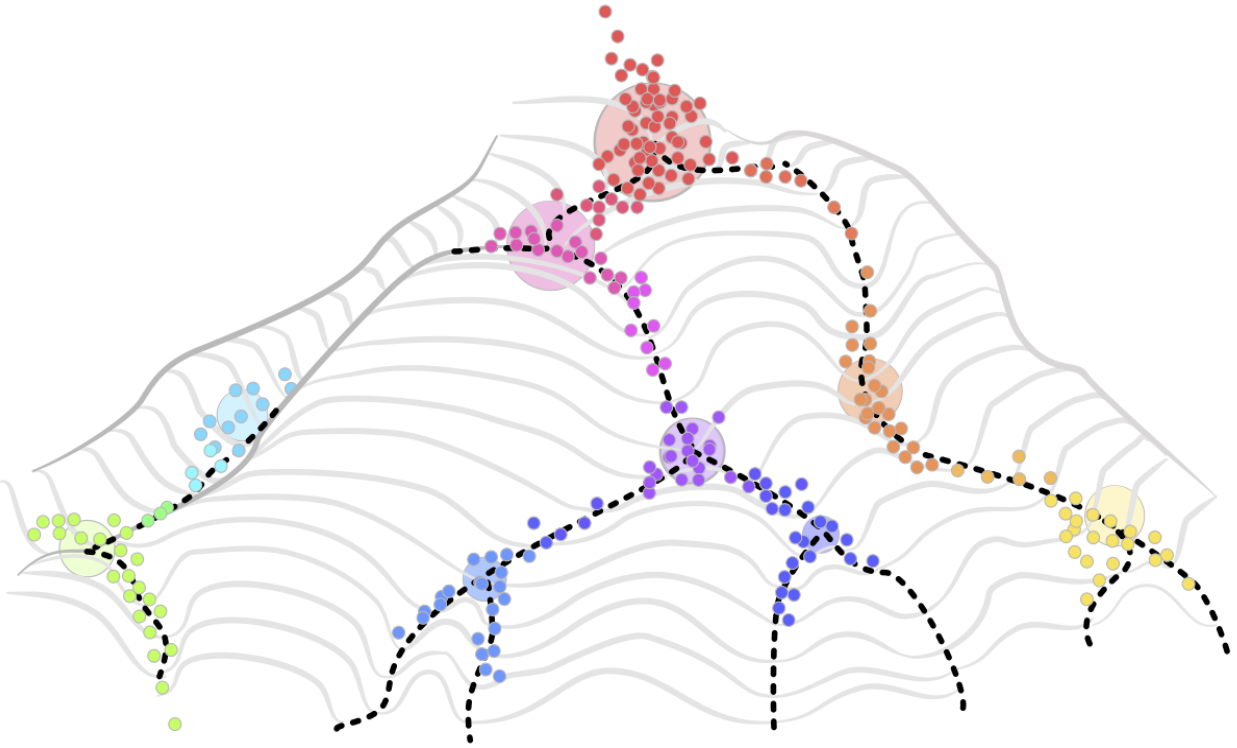
Spatial Transcriptomics of mouse embryos across 8 developmental stages
[Chen, et al. Cell, (2022)]



**And others!*
(Pijuan-Sala et al., Nature, 2019)
(Liu et al. *Developmental Cell*, 2022)

...

Temporal and Spatiotemporal transcriptomics: Opens up the Analysis of Fundamental Biological Questions!



The "Waddington Landscape"
Photo cred: (Waddington, 1957)

Questions:

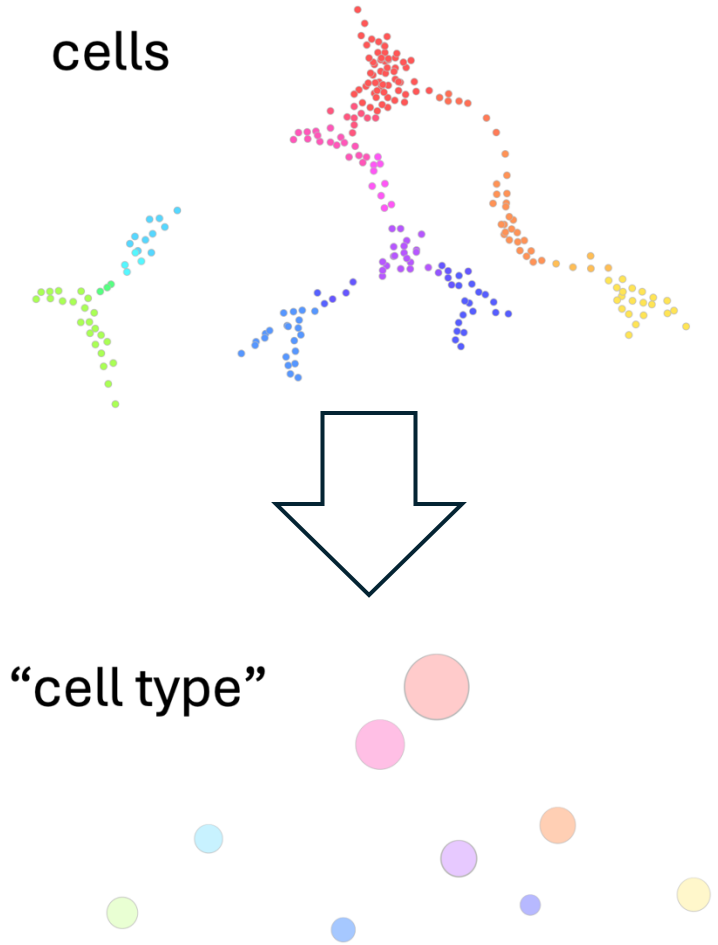
1. Ancestor-descendant relationships between cells across two timepoints?
2. Cell-states or types which index the temporal process of development?
3. Trajectories between these cell types?

Limitations:

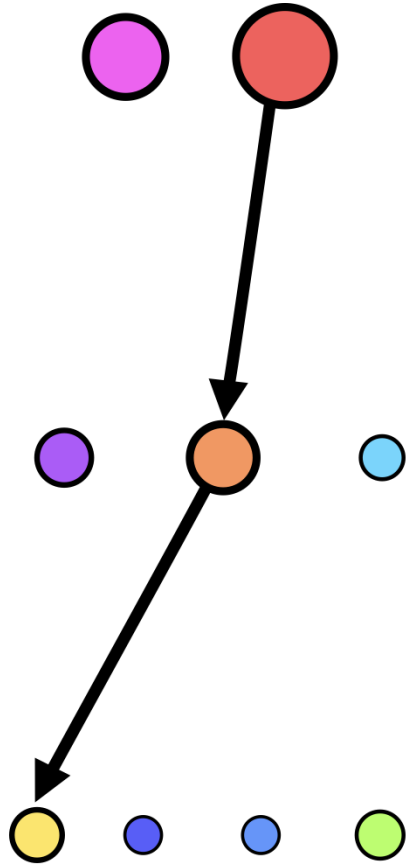
- Technology is destructive – each sample from a different individual
- Do not have ground-truth trajectories!

Differentiation maps and Cell Types (States)

A *cell-type* is a coarse-graining of cells into clusters.



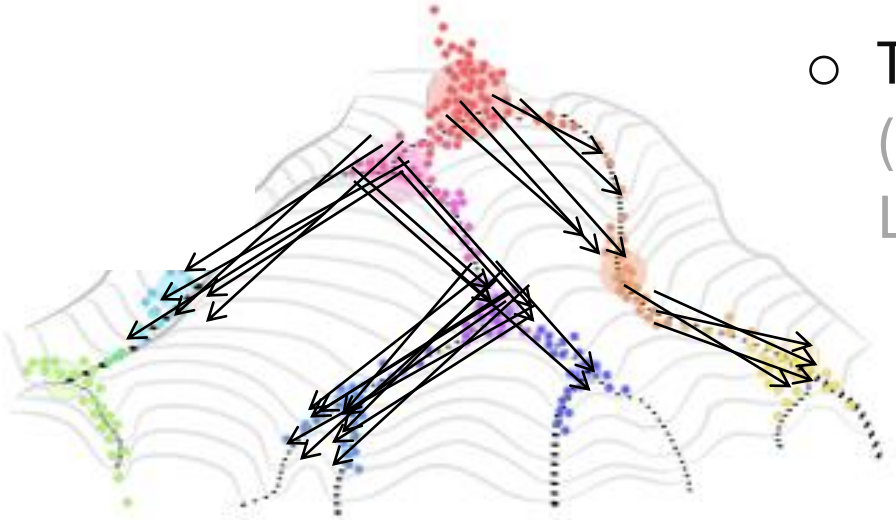
A differentiation map is a directed acyclic graph giving the ancestral relationship between cell-types



Existing Methods Infer Cell-Cell Coupling Independent of Cell Type

1. What are the ancestor-descendant relationship between cells across two timepoints?

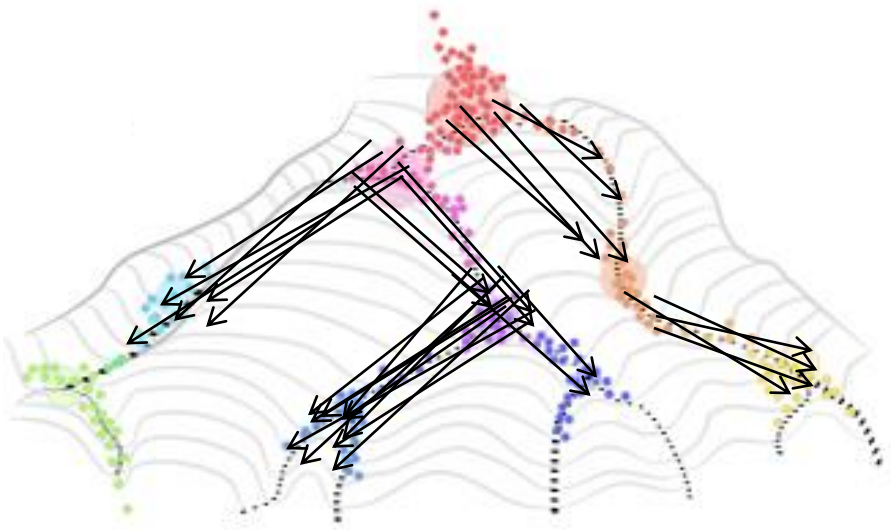
Cell-to-cell Coupling (Waddington OT, DeST-OT moscot)



- Many existing methods model the dynamics of cells using the technique of *optimal transport* (OT).
- These infer the least-cost mapping between individual cells (Schiebinger et al 2019, Zeira et al 2022, Klein et al 2025, Liu & Halmos et al 2025), building cell-to-cell trajectories

Existing Methods Infer Cell-Cell Coupling Independent of Cell Type

Cell-to-cell Coupling (Waddington OT, DeST-OT moscot)



However, these methods **do not**:

- Learn cell-types (assume cell-type inference is *distinct*)
- Find a **differentiation map** *jointly* with cell-type

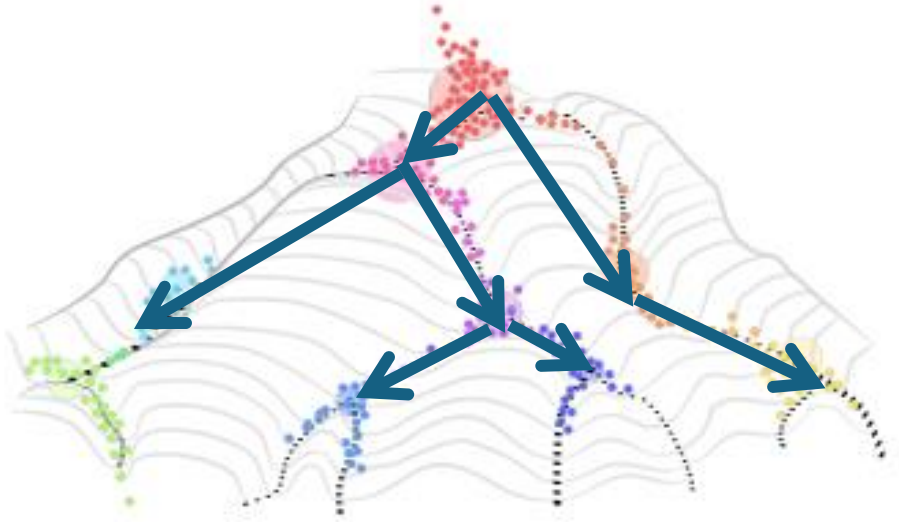
Finding Latent Trajectories over Latent Cell-State

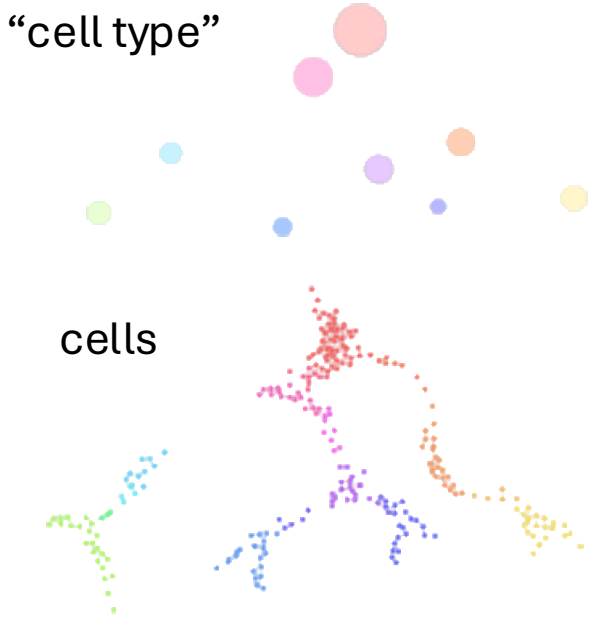
Our work addresses:

2. What are the cell-states or types which "index" the temporal process of development?

3. What is the *differentiation map* between these cell types?

Cell-type and Trajectory Inference



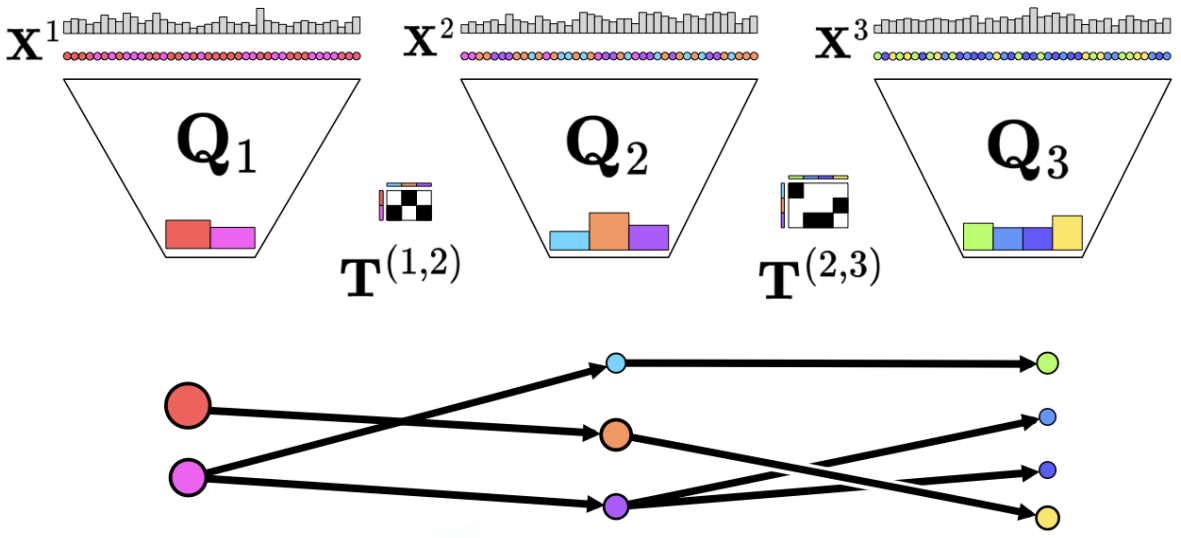


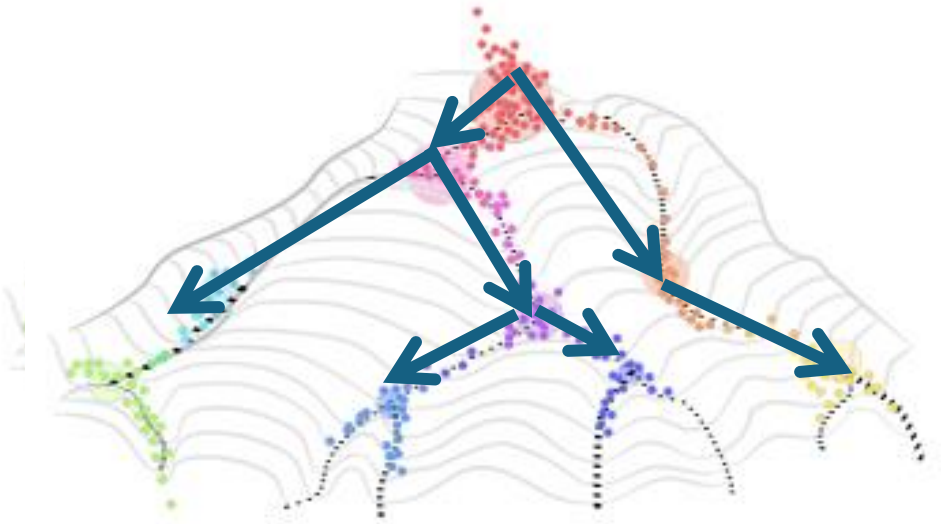
Hidden Markov Optimal Transport (HM-OT)

(1) Discovers latent cell types and aligns individual cells to them.

(2) Maps between the cell types while minimizing an optimal transport cost.

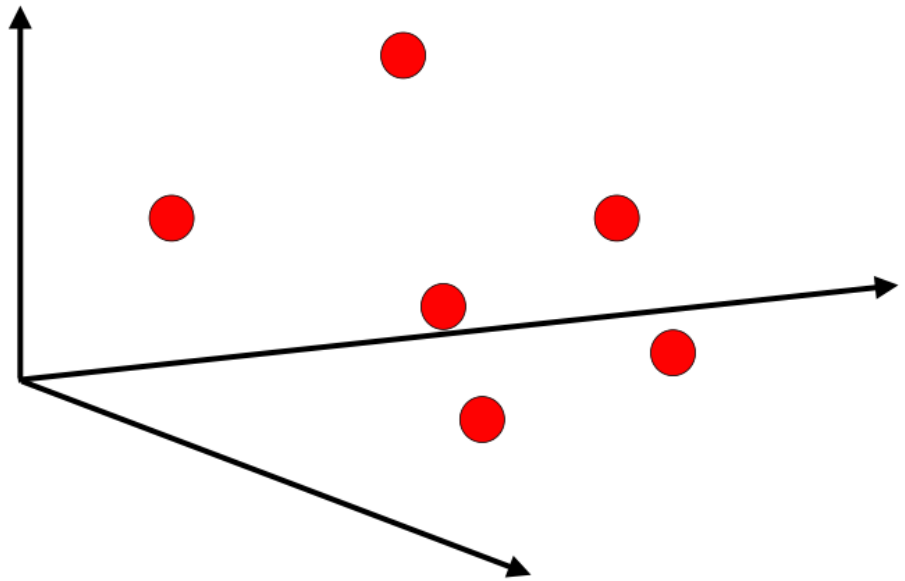
(3) Uses *low-rank optimal transport* (Forrow et al '19, Scetbon et al '20, Lin et al '21, Halmos et al '24) to do (1) and (2) simultaneously across multiple timepoints.





Optimal Transport (source: shamelessly lifted from Marco Cuturi!)

A puzzle:

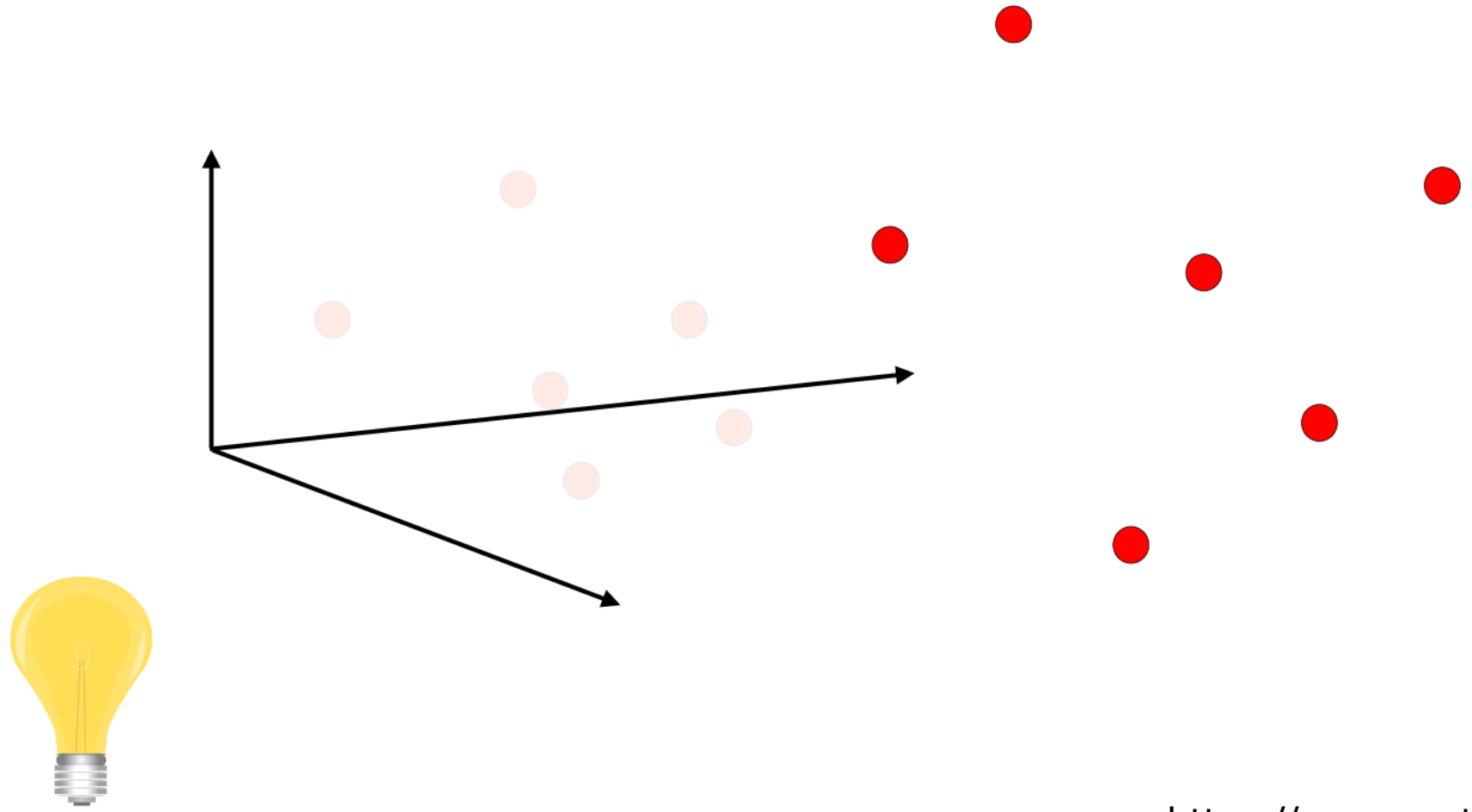


https://marcocuturi.net/ot.html

Optimal Transport

https://marcocuturi.net/ot.html

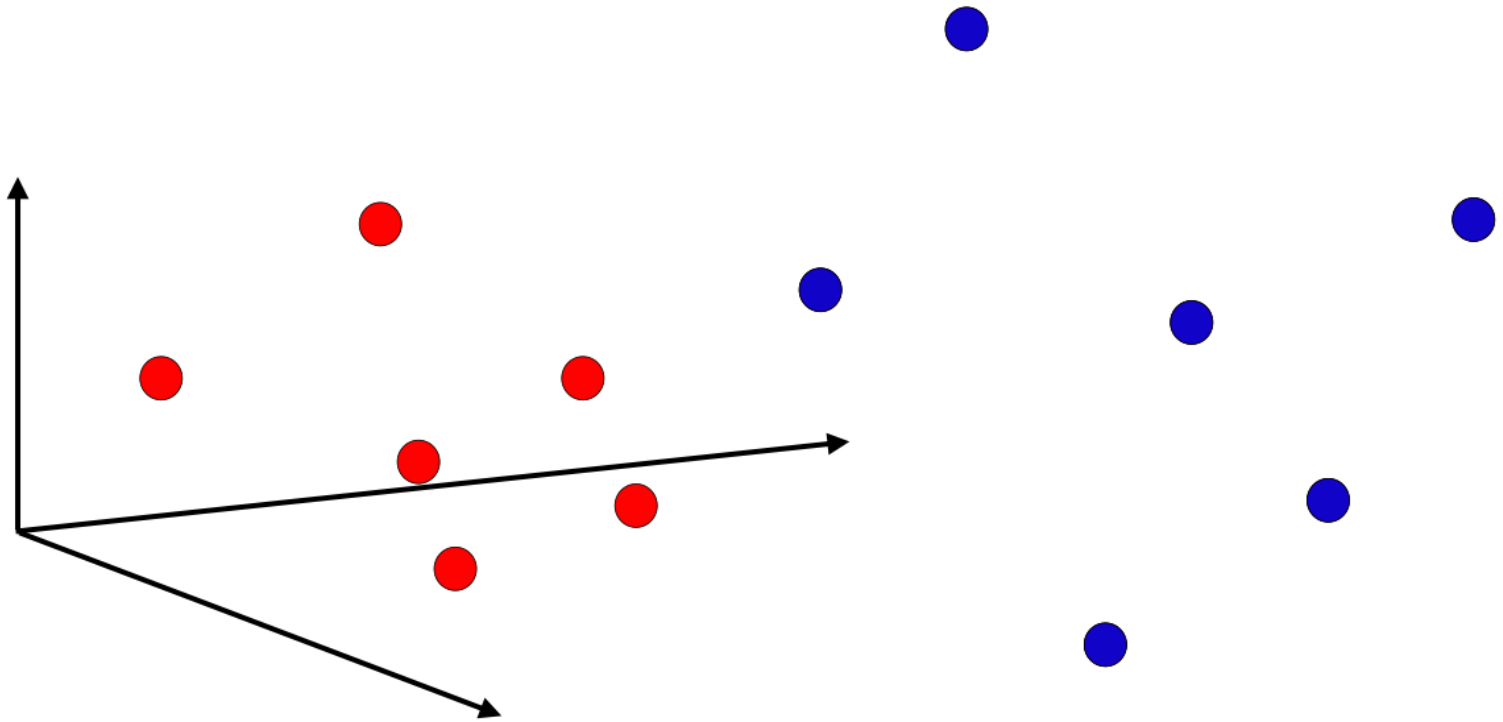
Optimal Transport



https://marcocuturi.net/ot.html

Optimal Transport

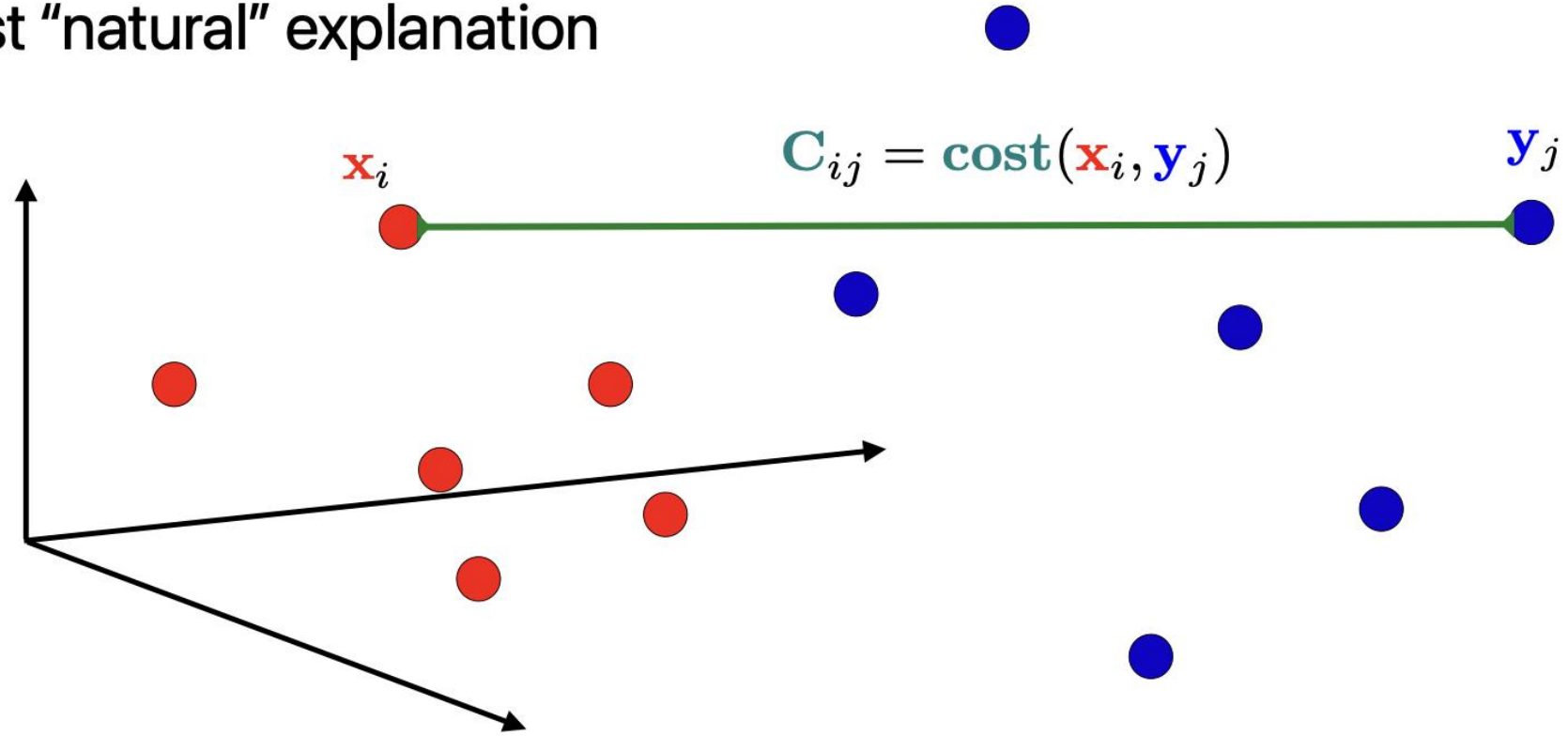
The puzzle is: who went where?



https://marcocuturi.net/ot.html

Optimal Transport

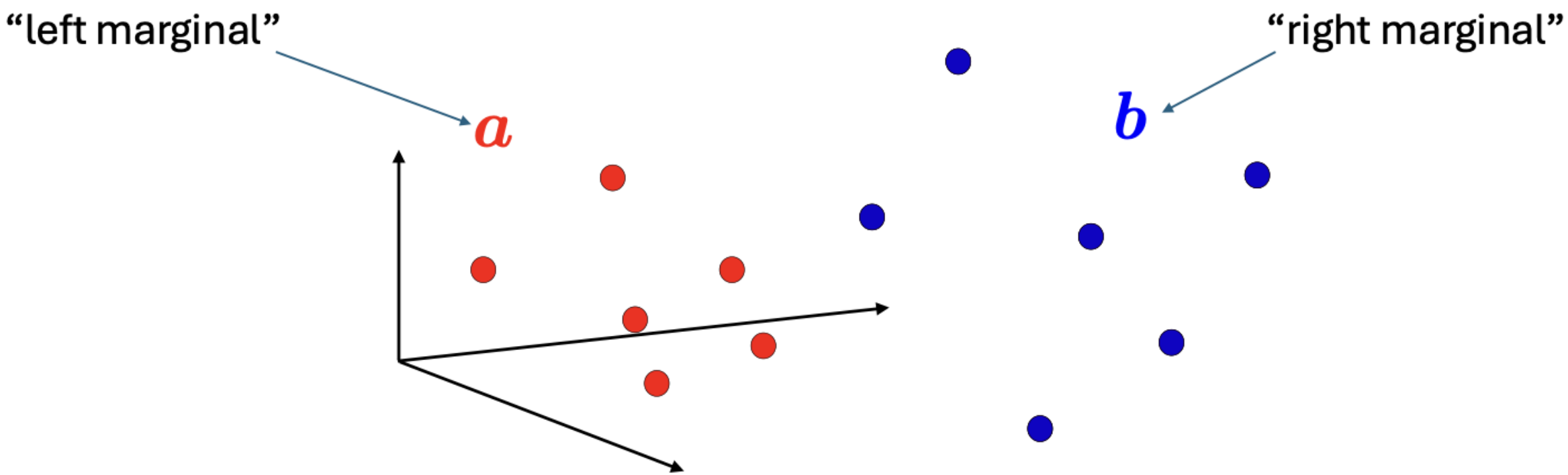
Goal: Find most “natural” explanation



https://marcocuturi.net/ot.html

Optimal Transport Feasible Set

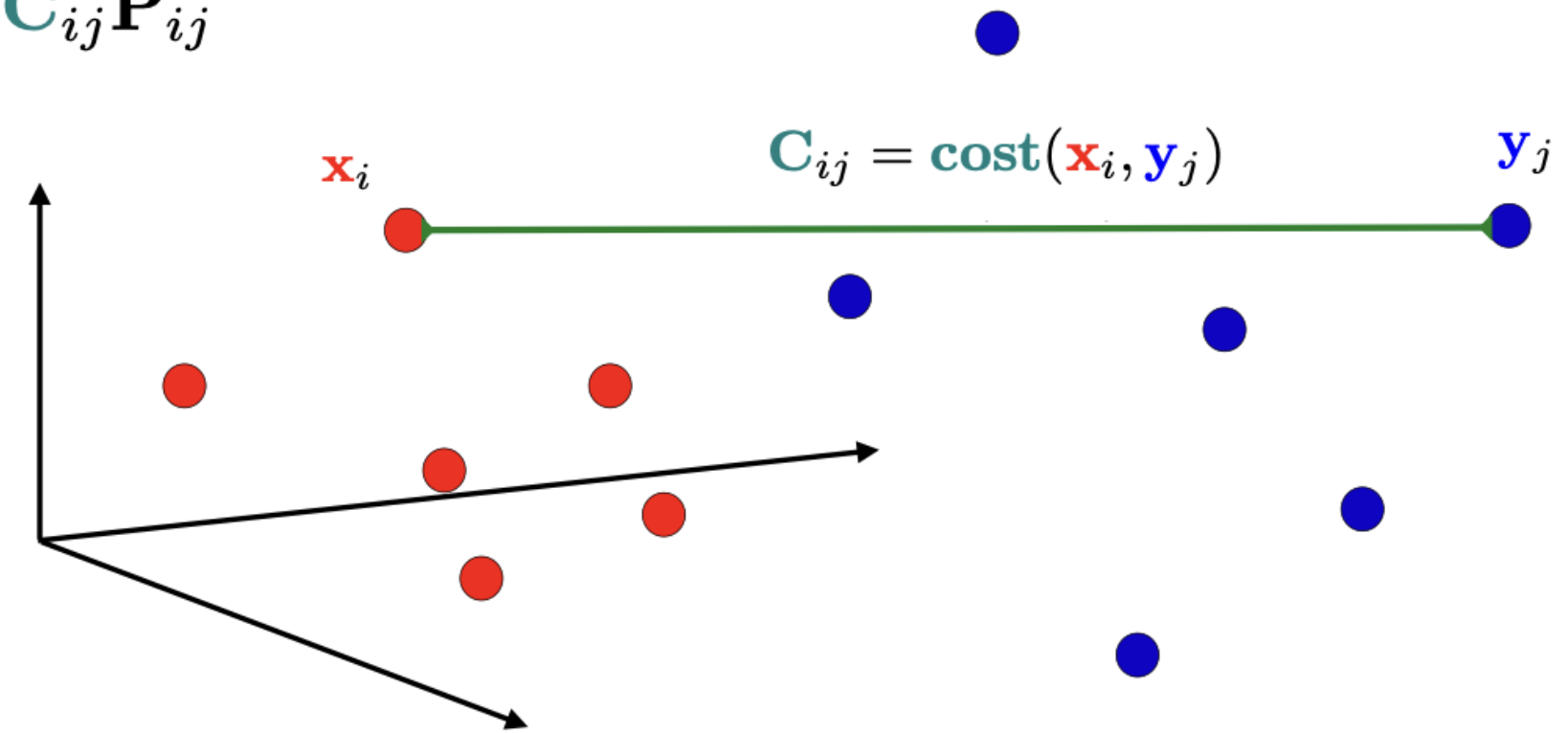
$$\Pi(\boldsymbol{a}, \boldsymbol{b}) = \{\mathbf{P} \in \mathbb{R}_{+}^{n \times m} : \mathbf{P}\mathbf{1}_m = \boldsymbol{a},\ \mathbf{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$



Wasserstein Problem

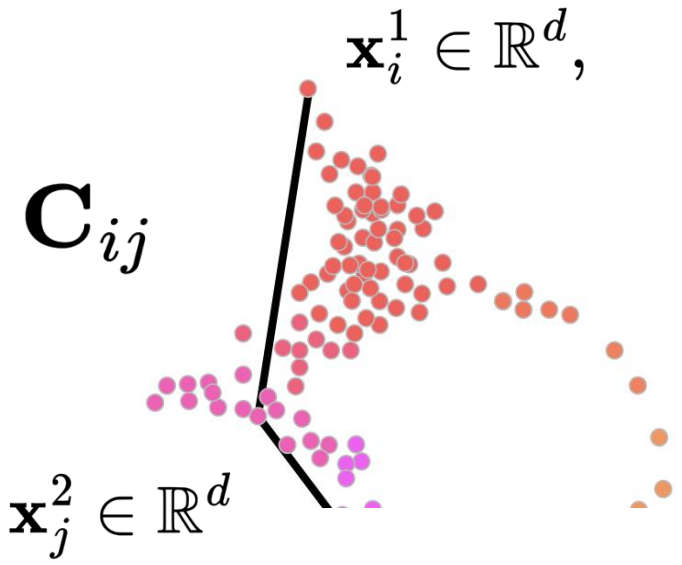
$$\mathbf{P}^\star = \underset{\mathbf{P}\in\Pi(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \mathbf{C}, \mathbf{P} \rangle$$

$$= \underset{\mathbf{P}\in\Pi(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \sum_{i,j} \mathbf{C}_{ij}\mathbf{P}_{ij}$$



Why use OT for temporal alignment?

The most "natural" alignment minimizes the transcriptional distance between these cells, represented with cost C_{ij}



$$\mathbf{C}_{ij} := \mathbf{C}_{ij}^{(1,2)} = \|\mathbf{x}_i^1 - \mathbf{x}_j^2\|_2$$

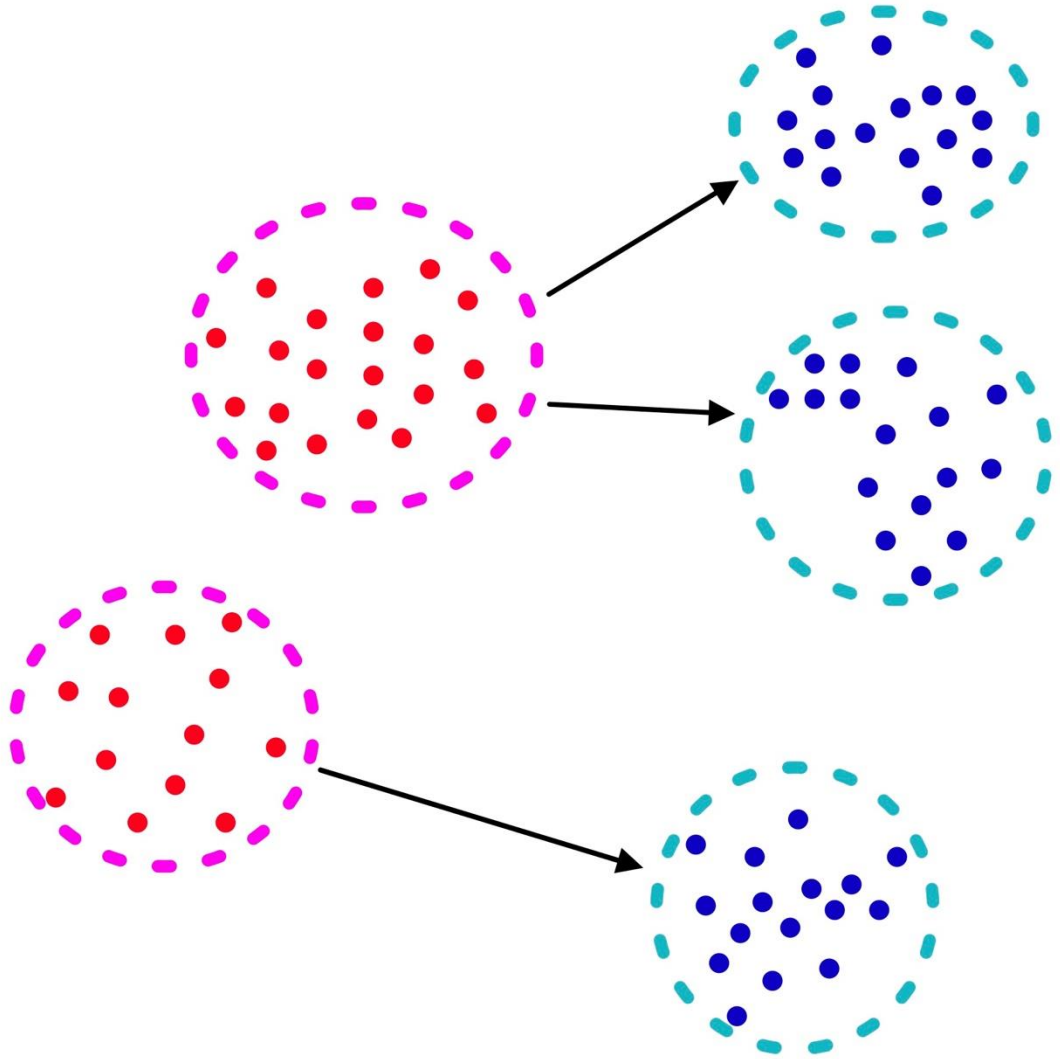
*Low-Rank** Wasserstein Problem

Often the data has some cluster structure, e.g. cell types, and the most interesting biological question is to understand the mapping at that resolution.

Consider a modification of the puzzle:

(1) What are the "best" cell types

(2) Which cell types transitioned to which?



* Latent-coupling (LC) formulation
(Lin et al 2021, Halmos et al 2024)

Low-Rank Wasserstein Problem

$$\mathbf{P}^\star = \arg\min_{\mathbf{P}\in\Pi_{\mathbf{g}_1,\mathbf{g}_2}(\mathbf{a},\mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle$$

LC (latent coupling) factorization of P imposes *rank constraint* and decomposes P into 3 factors while keeping it a feasible coupling

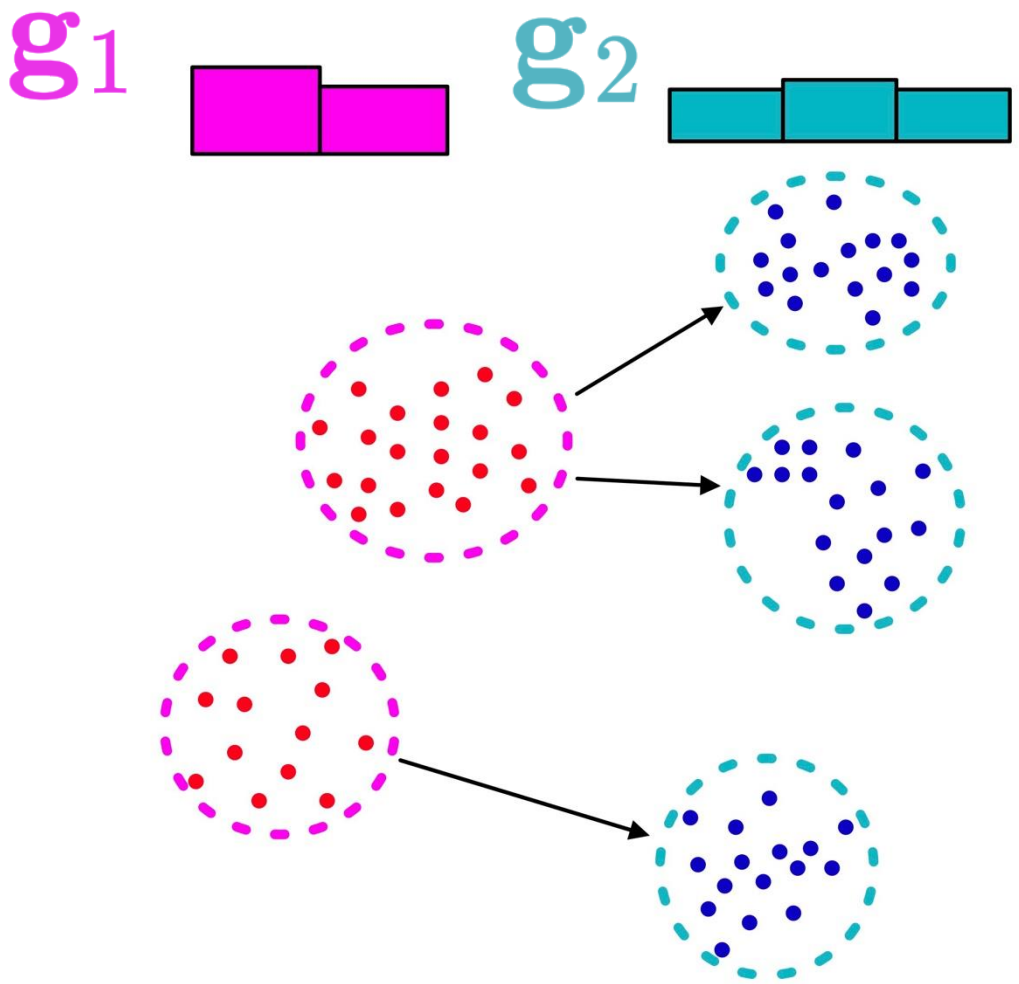
$$\mathbf{P} \in \Pi_{\mathbf{g}_1,\mathbf{g}_2}(\mathbf{a}, \mathbf{b})$$

$$\implies \mathbf{P} = \mathbf{Q}_1 \mathrm{diag}(1/\mathbf{g}_1) \mathbf{T} \mathrm{diag}(1/\mathbf{g}_2) \mathbf{Q}_2^{\mathrm{T}}$$

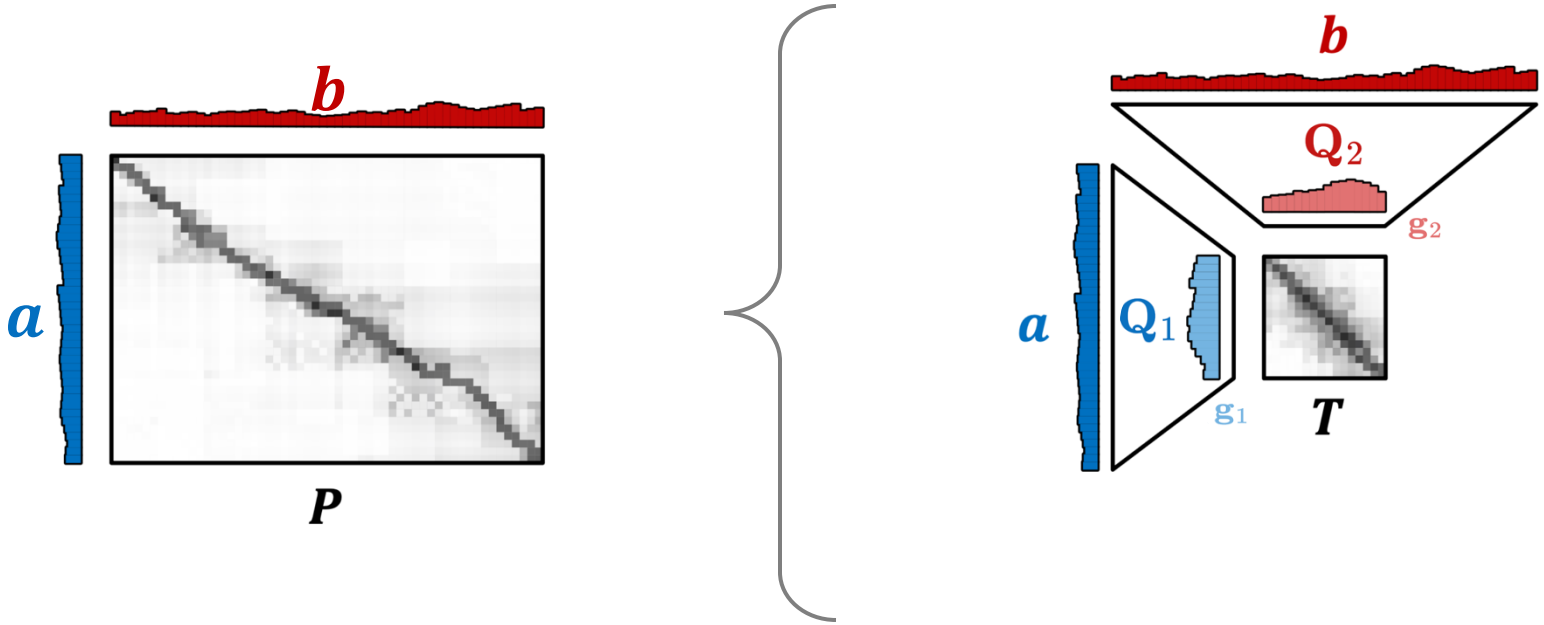
$\mathbf{Q}_1 \in \Pi(\mathbf{a}, \mathbf{g}_1)$ — Couples point distribution at time 1 to cell-type distribution at time 1

$\mathbf{T} \in \Pi(\mathbf{g}_1, \mathbf{g}_2)$ — Couples cell-type distribution at time 1 to cell-type distribution at time 2

$\mathbf{Q}_2^{\mathrm{T}} \in \Pi(\mathbf{g}_2, \mathbf{b})$ — Couples cell-type distribution at time 2 to point distribution at time 2



Low-Rank Optimal Transport: A Special Parametrization for a Coupling



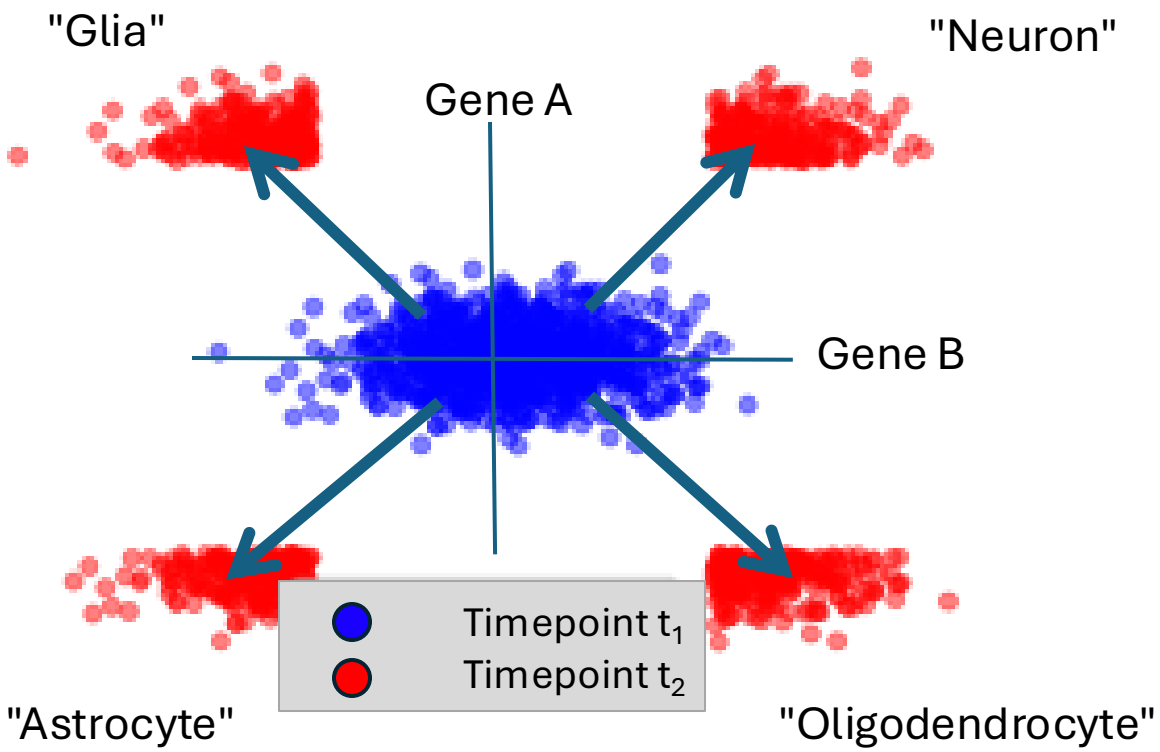
Factor relaxation with latent coupling (FRLC)

Low rank approximation of optimal transport

Halmos*, Liu*, Gold*, R. NeurIPS (2024)

Why use Low-Rank OT for temporal alignment?

Sometimes the clustering / cell types at one timepoint are *not enough* to build a map of cell differentiation!

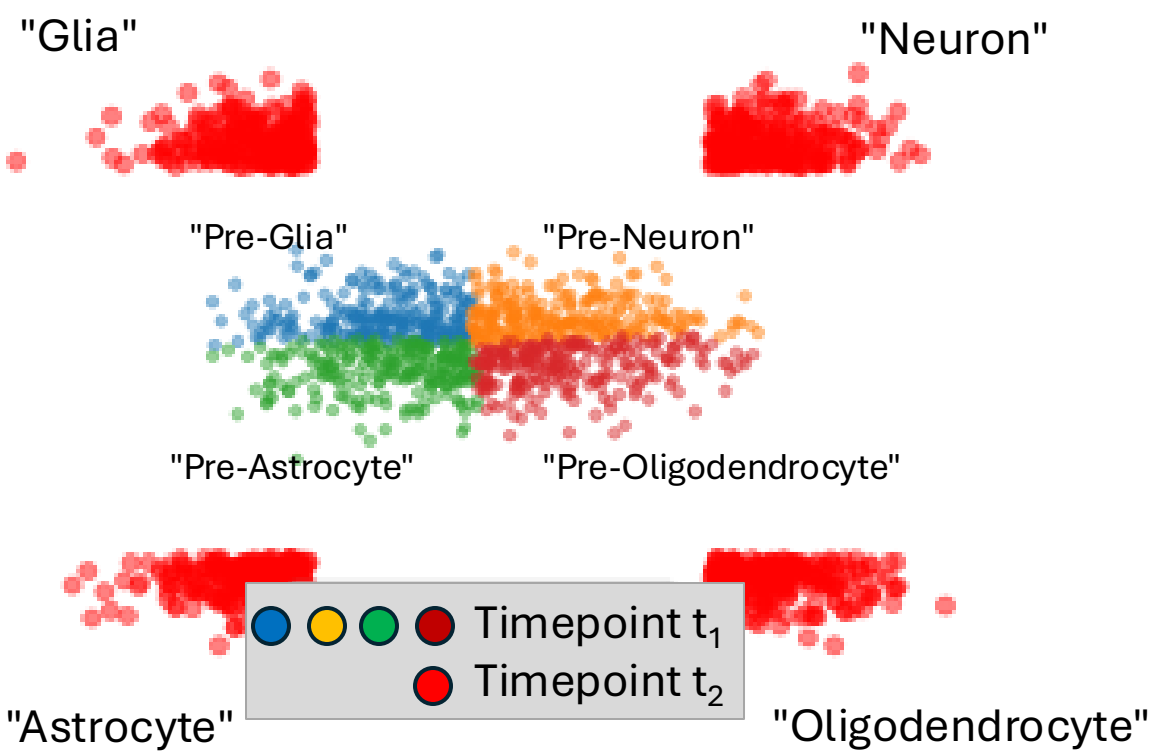


Why use Low-Rank OT for temporal alignment?

Need to leverage temporal information to get both differentiation map and cell-state correct!

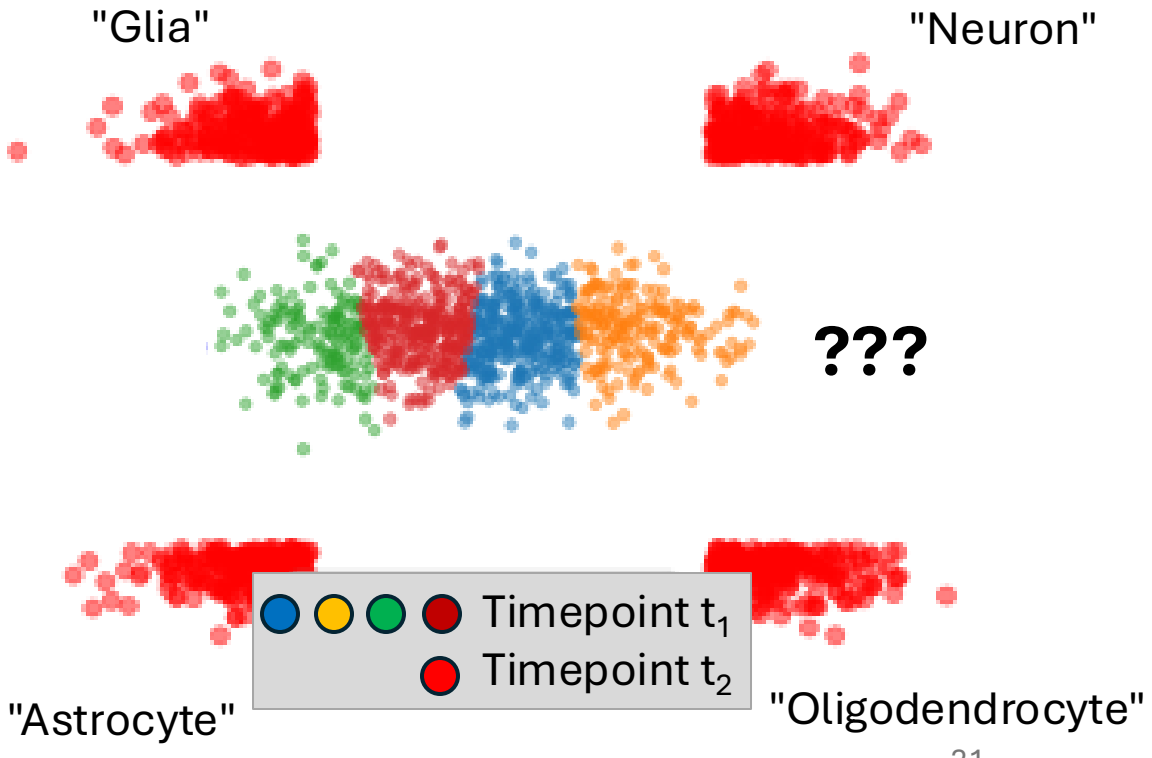
HM-OT / Low-Rank OT:

Clusters by ancestry using multiple timepoints.



Single-timepoint clustering (e.g. k-means)

Only uses information from one timepoint.

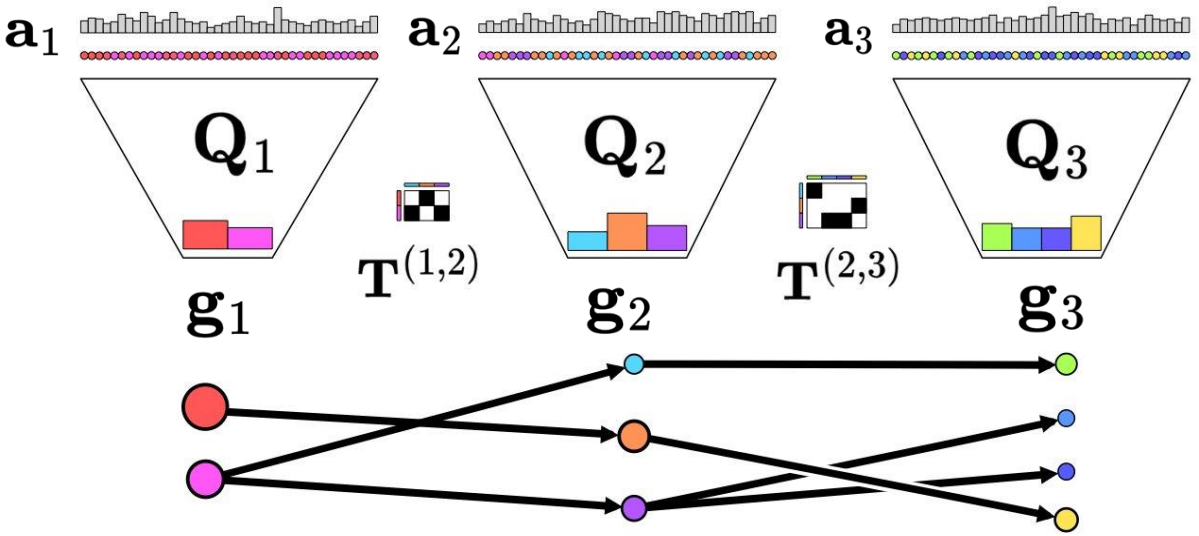


Hidden-Markov Optimal Transport

Problem: Given empirical distributions $(a_t)_{t=1,..,N}$ find the latent factors $(Q_t)_{t=1,..,N}$ and differentiation maps $(T^{(t,t+1)})_{t=1,..,N-1}$ that minimize the Wasserstein distance traveled by the clusters through time.

$$\min_{\mathbf{Q},\mathbf{T}\,:\,(\mathbf{Q}_t,\mathbf{Q}_{t+1},\mathbf{T}^{(t,t+1)})\in \mathsf{LC}_{\mathbf{a}_t,\mathbf{a}_{t+1}}(r_t,r_{t+1})} \sum_{t=1}^{N-1} \langle \mathbf{C}^{(t,t+1)}, \mathbf{P}^{(t,t+1)} \rangle_F$$

$$\mathbf{P}^{(t,t+1)} := \mathbf{Q}_t \mathrm{diag}(1/\mathbf{g}_t) \mathbf{T}^{(t,t+1)} \mathrm{diag}(1/\mathbf{g}_{t+1}) \mathbf{Q}_{t+1}^T$$



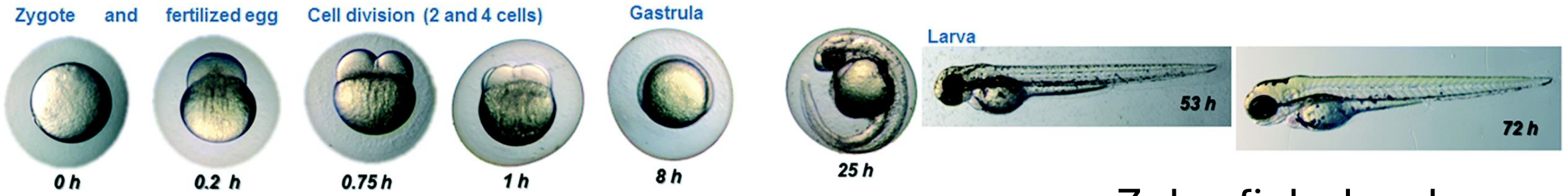
HM-OT: Algorithm

- Computes MAP estimates of Q_t , $T^{(t,t+1)}$, g_t using an algorithm analogous to forward-backward for Hidden Markov Models (HMM)

- Highly flexible in terms of input information! One can either run it unsupervised and learn all variables or fix/initialize any subset of the following and learn the rest:
  - Cell-type proportions ($\mathbf{g}_t$) [i.e. if you know there are "rare" cell-types]
  - Cell to cell-type mappings ($\mathbf{Q}_t$) [i.e. if you know cell-types]

- More algorithmic details are in the paper!

Zebrafish development is a well-studied model of organismal development and cell differentiation



Zebrafish development

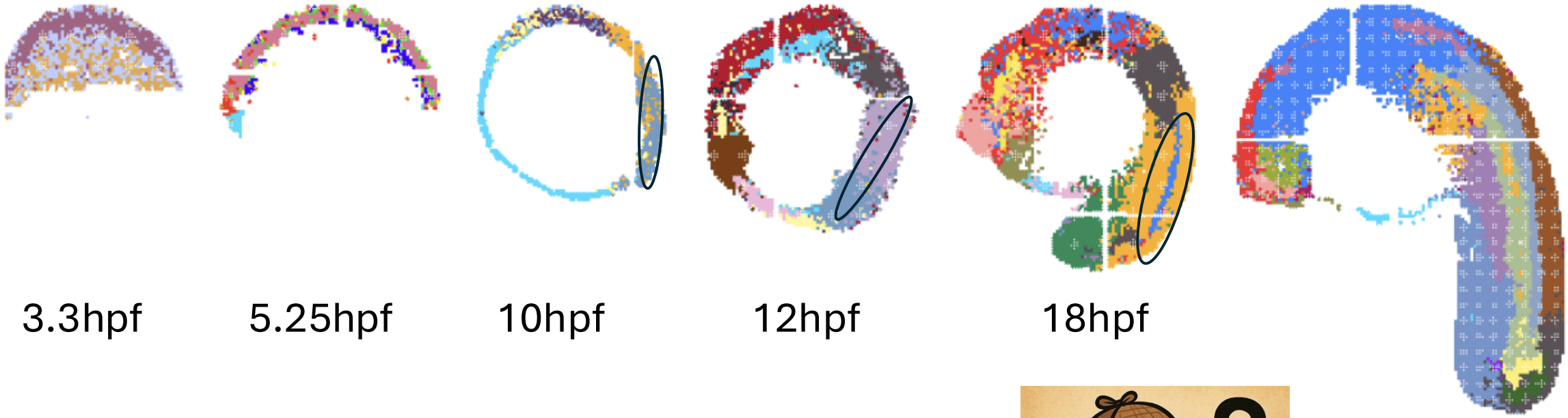
[García-Cambero,et al. *Env.Sci.* 2019]

Spatiotemporal transcriptomics

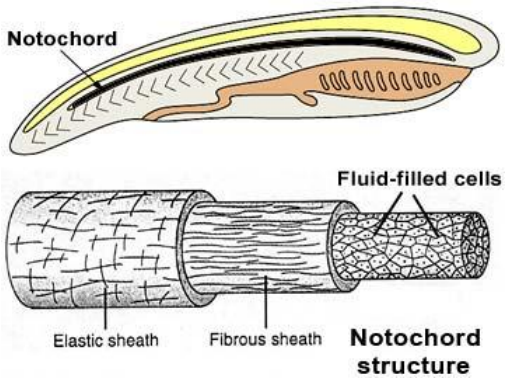
Stereo-Seq [Liu et al. *Developmental Cell* (2022)]

Spatiotemporal transcriptomics of zebrafish embryogenesis

Liu et al. *Developmental Cell* (2022)

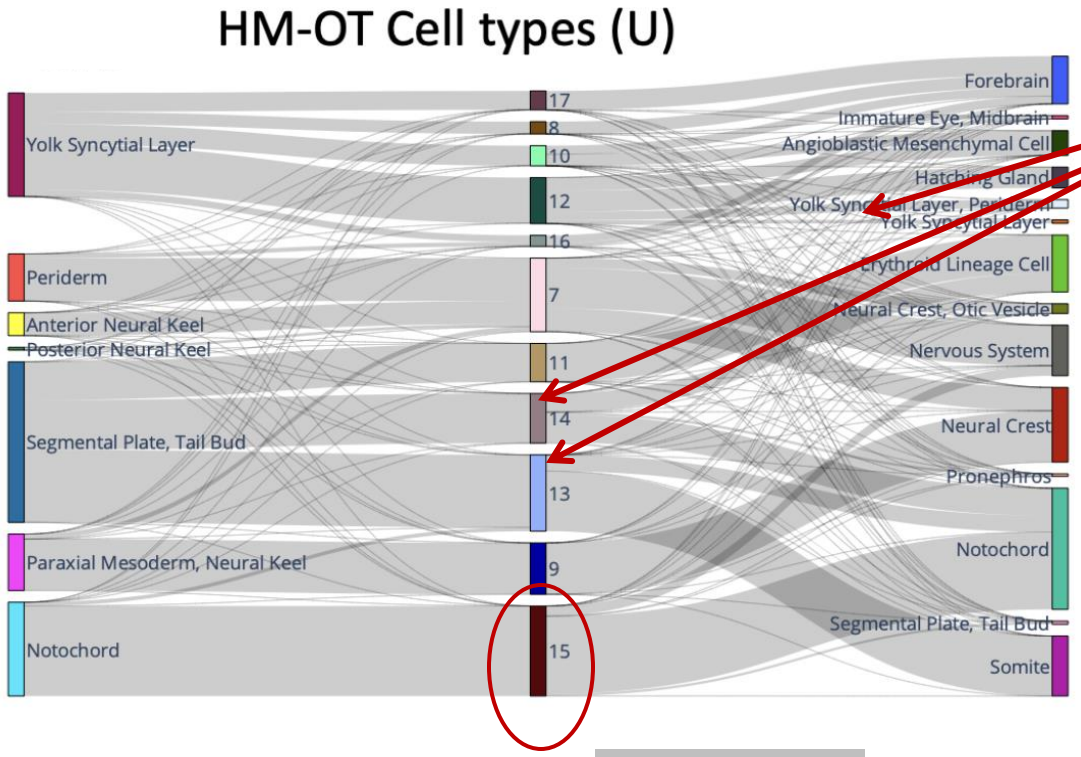


3.3hpf 5.25hpf 10hpf 12hpf 18hpf 24hpf



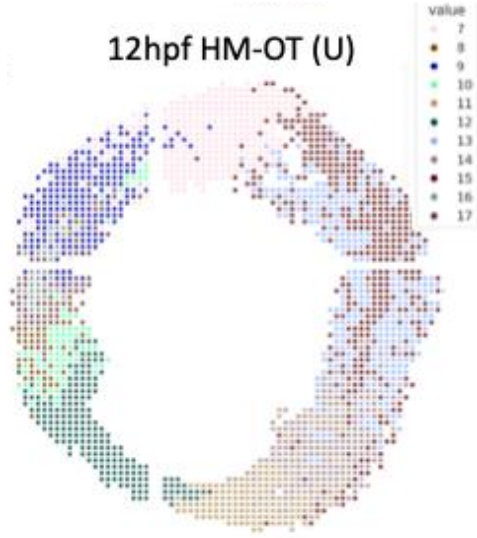
Notochord cell-type "disappears" in the published annotation!

HM-OT: Clustering and Differentiation Map of Zebrafish



Corrects transitions which were incorrect with annotated cell-types

12hpf



Notochord recovered!

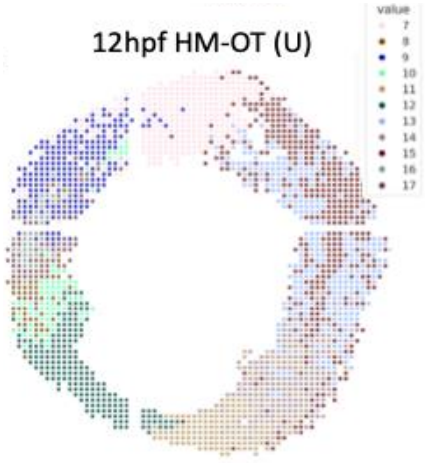
Stereo-Seq data and cell type annotations from Liu et al. *Developmental Cell* (2022)

HM-OT: Differentiation Map of Zebrafish

HM-OT inferred cell-types substantially improve the NPMI (pointwise mutual information) to ground truth trajectories relative to annotated types:

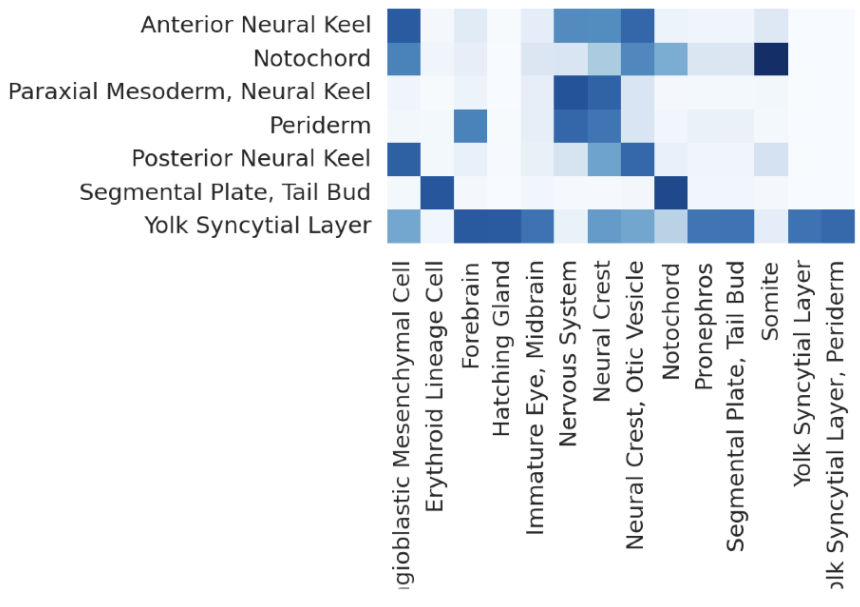
12hpf

12hpf HM-OT (U)

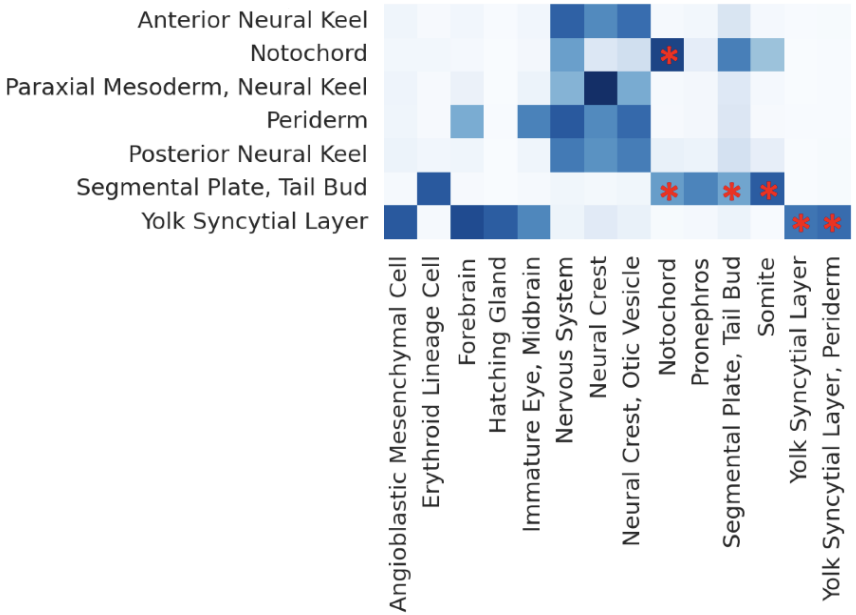


NPMI of Cell-types between 10hpf and 18hpf

HM-OT (S)

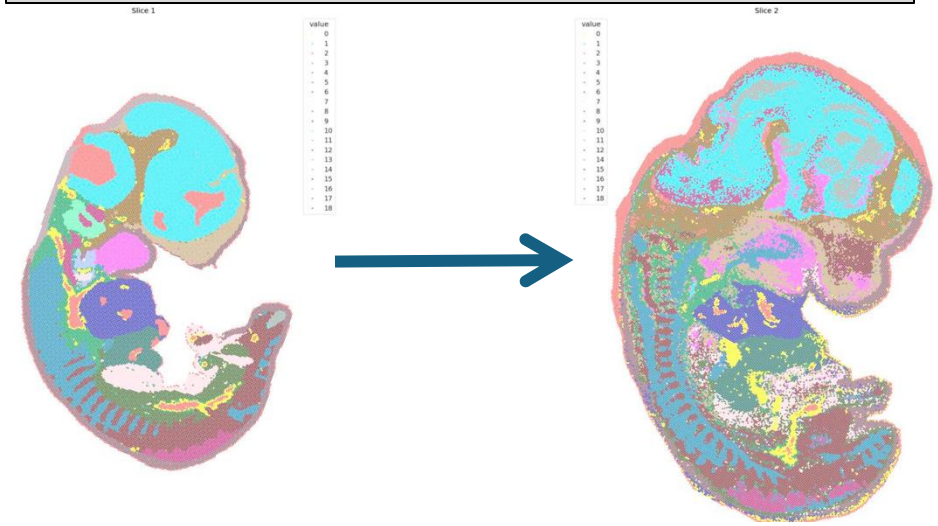


HM-OT (U)

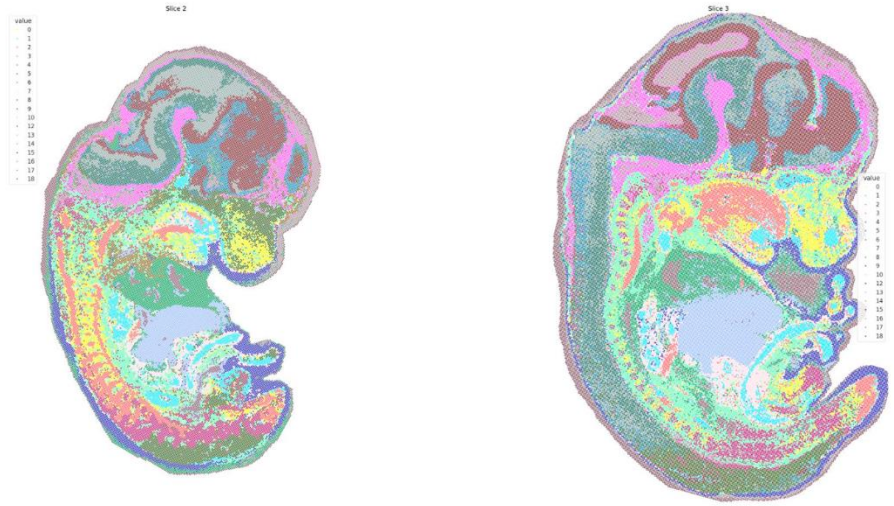


HM-OT is a Flexible Toolbox for (Co) Clustering

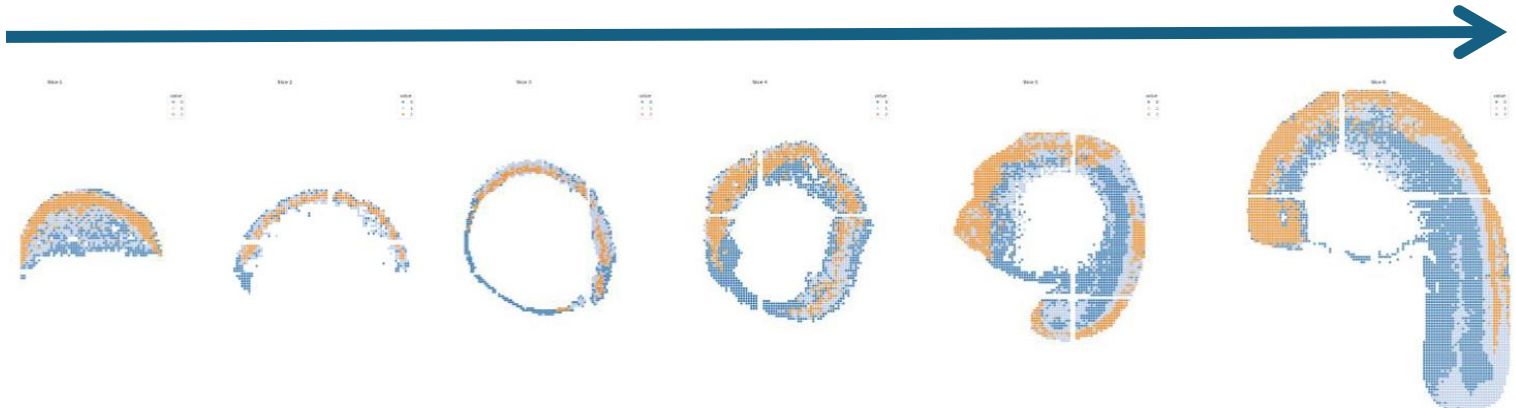
Transfer known clusters forwards or backwards in time to other data

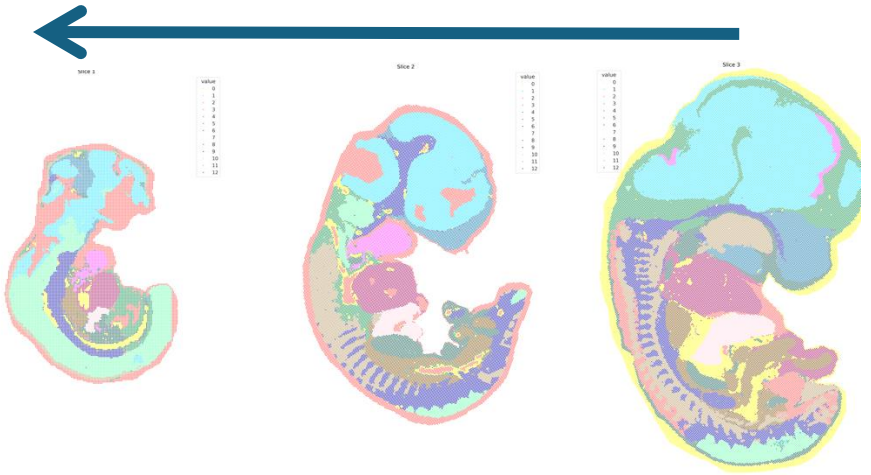


Learn cell state/type from scratch to minimize HM-OT objective



Project or co-cluster cell-types forward and backward in time through differentiation map

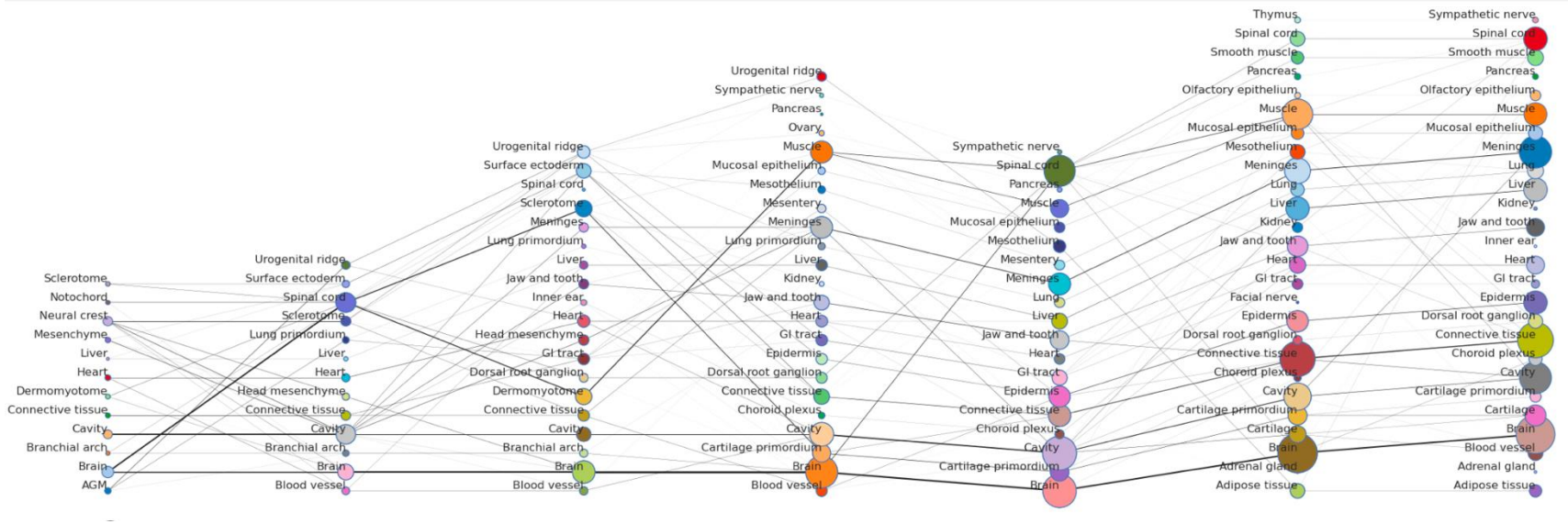




Large-Scale Inference of Differentiation Maps

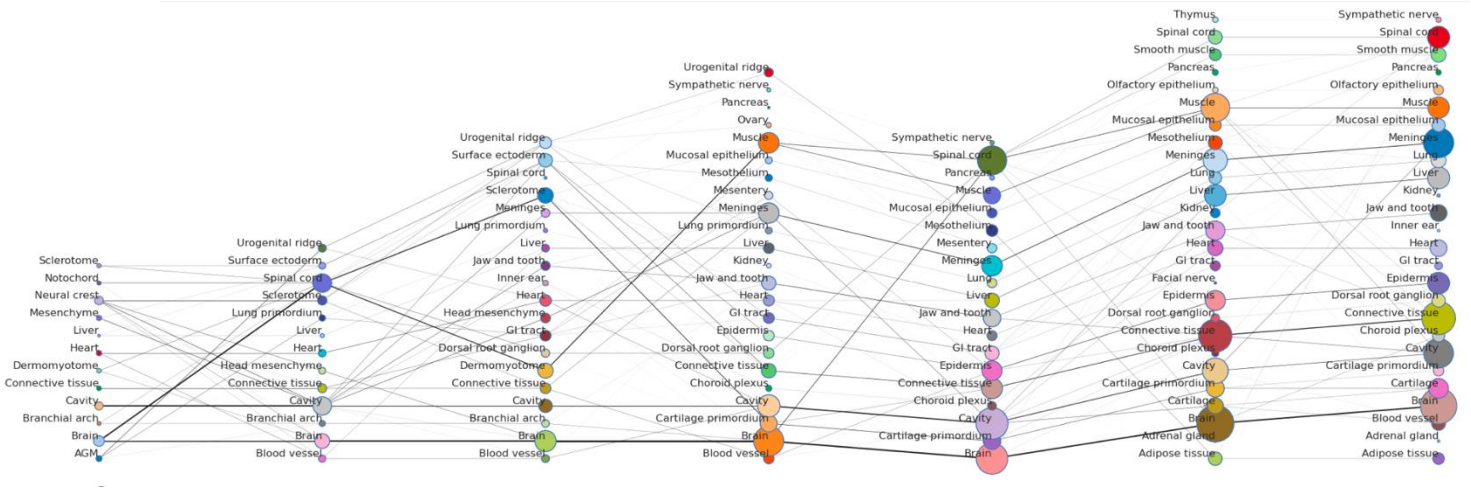
Lightning fast and space-efficient; can scale maps to millions of points!

Spatial (Stereo-Seq) Mouse Development (Chen et al '22)

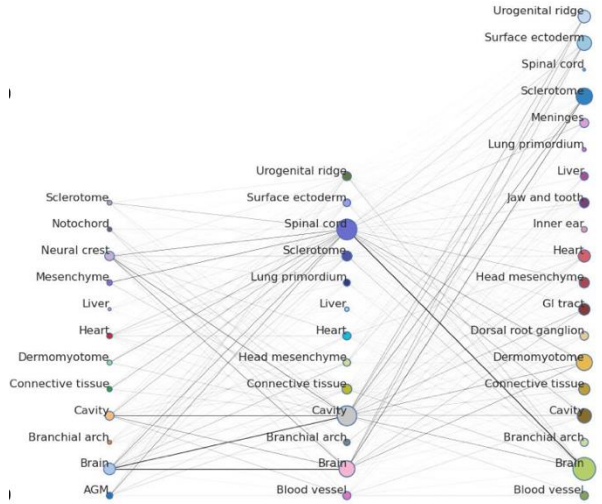


Temporal (Single-Cell) Mouse Embryogenesis (Qiu et al '24)

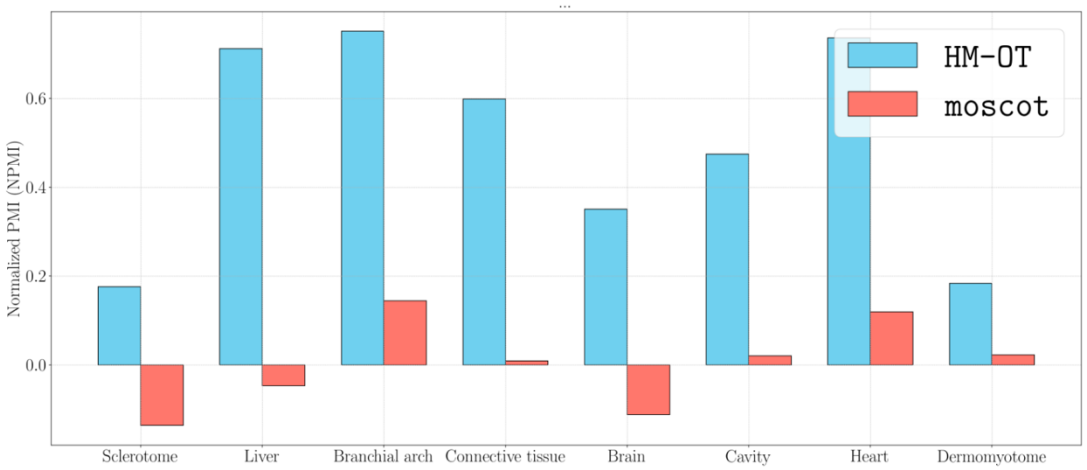
Large-Scale Inference of Differentiation Maps



HM-OT



moscot spatiotemporal



Other true transitions identified:

- urogenital ridge transitioning to ovary (NPMI = **0.560**, E11.5-12.5)
- lung primordium transitioning to lung (NPMI=**0.927**, E12.5-13.5)
- dermomyotome transitioning to muscle (NPMI=**0.968**, E11.5-12.5)
- sclerotome transitioning to cartilage primordium (NPMI=**0.866**, E11.5-12.5)
- surface ectoderm transitioning to the epidermis (NPMI=**0.705**, E11.5-12.5)

Spatial (Stereo-Seq) Mouse Development (Chen et al '22)

Summary

HM-OT: a scalable algorithm to infer differentiation maps, discover temporal co-clusters, and track cell-types through time and space.

- HM-OT introduces a novel factorization of optimal transport to model cell-type differentiation
- Optimizes this factorization across the full time-series of temporal transcriptomics data

Thank you!

https://github.com/raphael-group/HM-OT/

Acknowledgments

Raphael Group

Prof. Ben Raphael
Dr. Brian Arnold
Dr. Metin Balaban
Dr. Uthsav Chitra
Dr. Julian Gold
Dr. Cong Ma
Dr. Uyen Mai
Dr. Hirak Sarkar
Dr. Palash Sashittal
Dr. Yihang Shen
Dr. Alexander Strzalkowski
Dr. Hongyu Zheng
Viola Chen
Gillian Chu
Peter Halmos
William Howard-Snyder
Gary Hu
Akhil Jakatdar
Xinhao Liu
Sereno Lopez-Darwin
Henri Schmidt
Ahmed Shuaibi
Richard Zhang
Clover Zheng

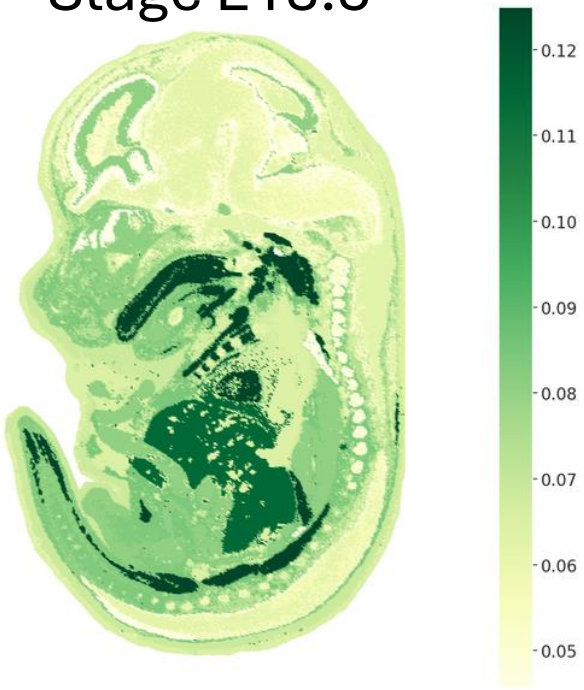
PRINCETON UNIVERSITY

NIH NATIONAL CANCER INSTITUTE Informatics Technology for Cancer Research

SCHMIDT FUTURES

Scaling DeST-OT with low rank optimal transport

Stage E13.5

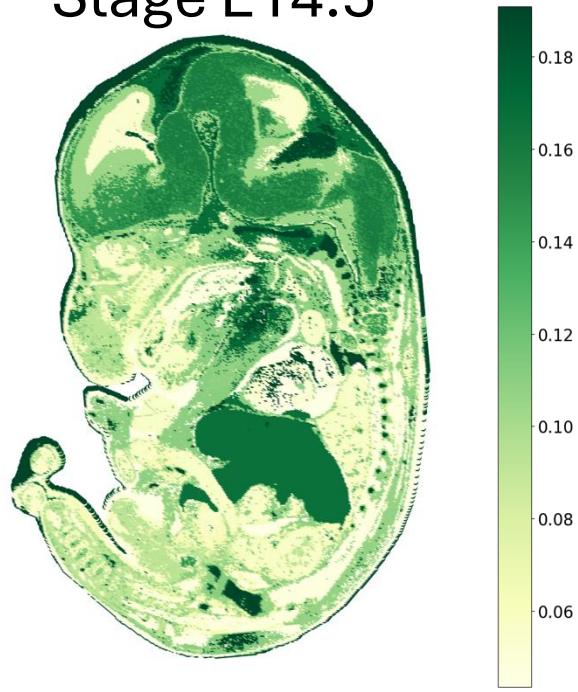


77K spots

Stereo-seq of mouse embryo

[Chen et al. *Cell*, 2022]

Stage E14.5



102K spots