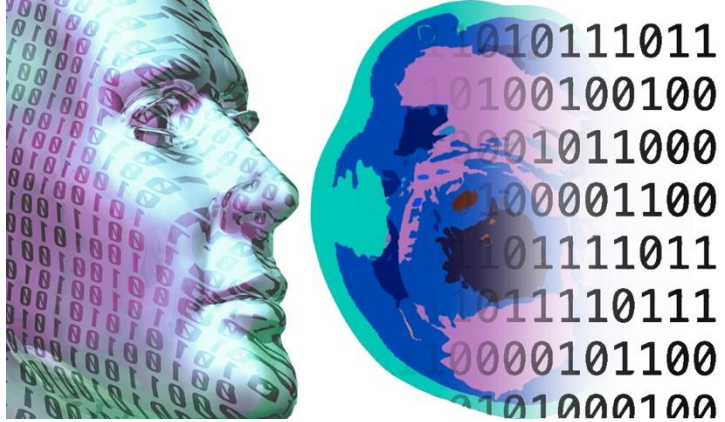


Learning and characterizing cellular dynamics with Neural and Graph ODE Networks (Part 1)

Smita Krishnaswamy

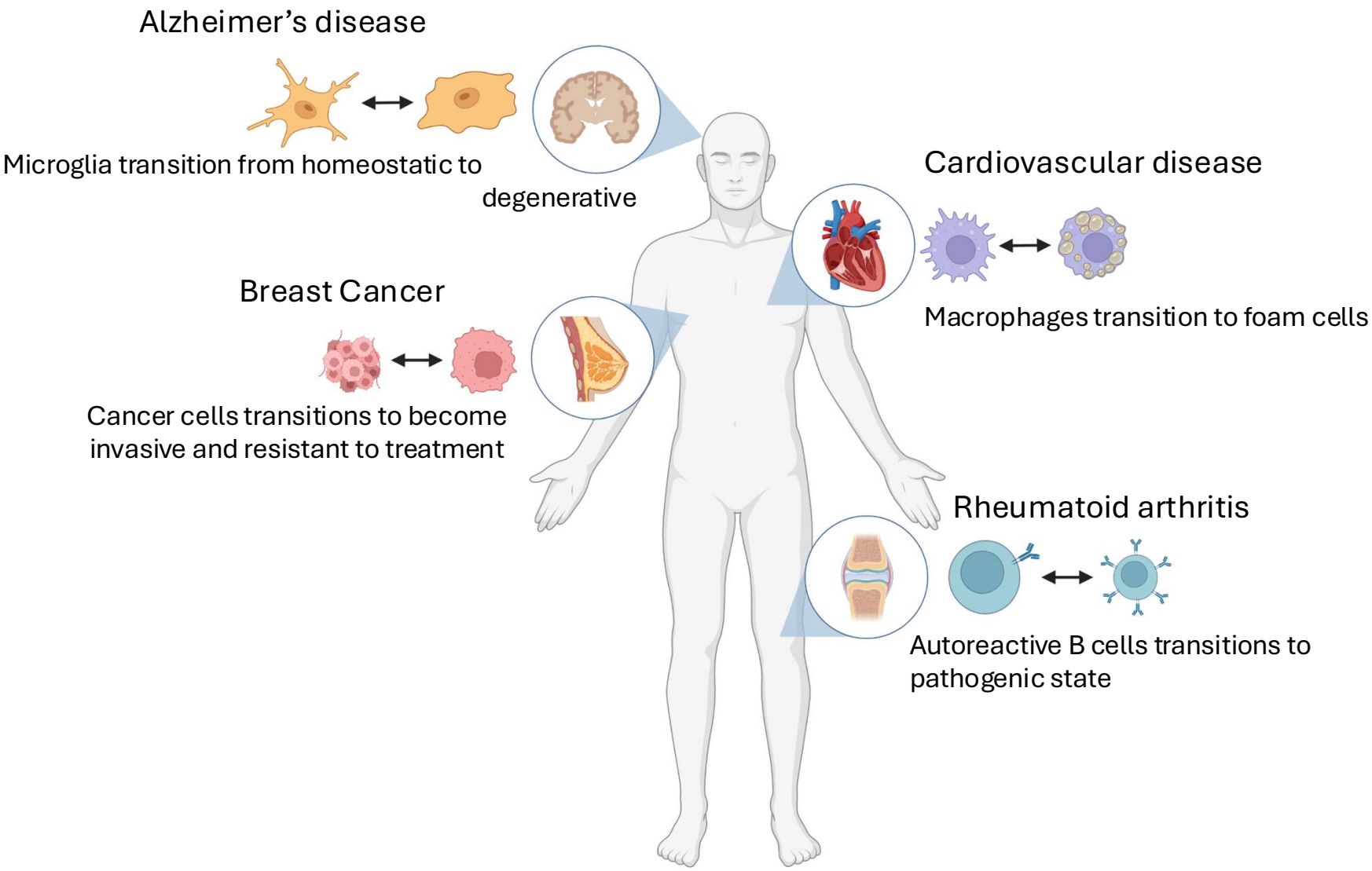
Department of Computer Science, Department of Genetics

Program for Applied Math, Program for Computational Biology & Bioinformatics

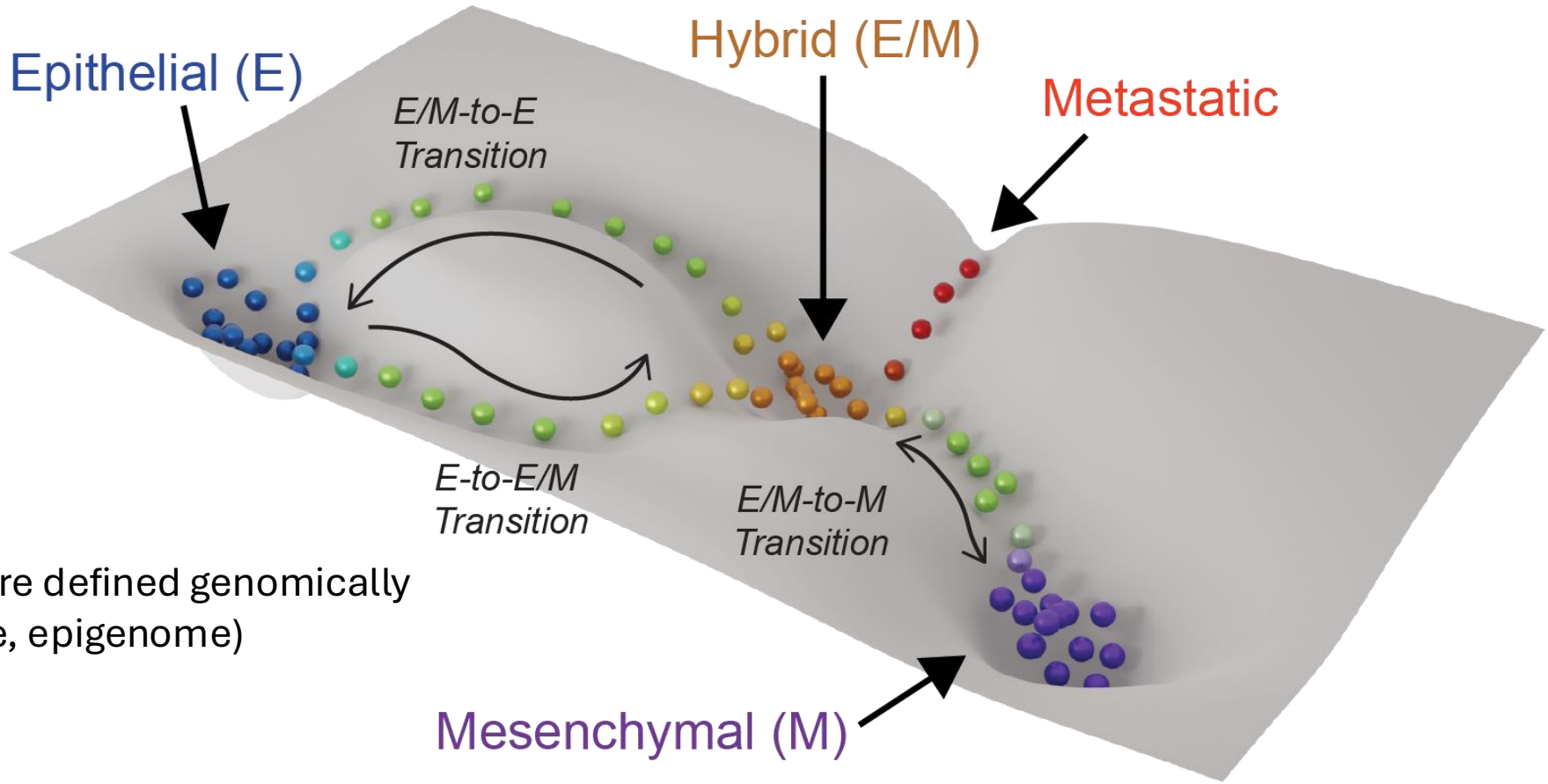
Program for Interdepartmental Neuroscience

Wu Tsai Institute Center for Neurocomputation and Machine Intelligence

Diseases involve dynamic cellular state changes

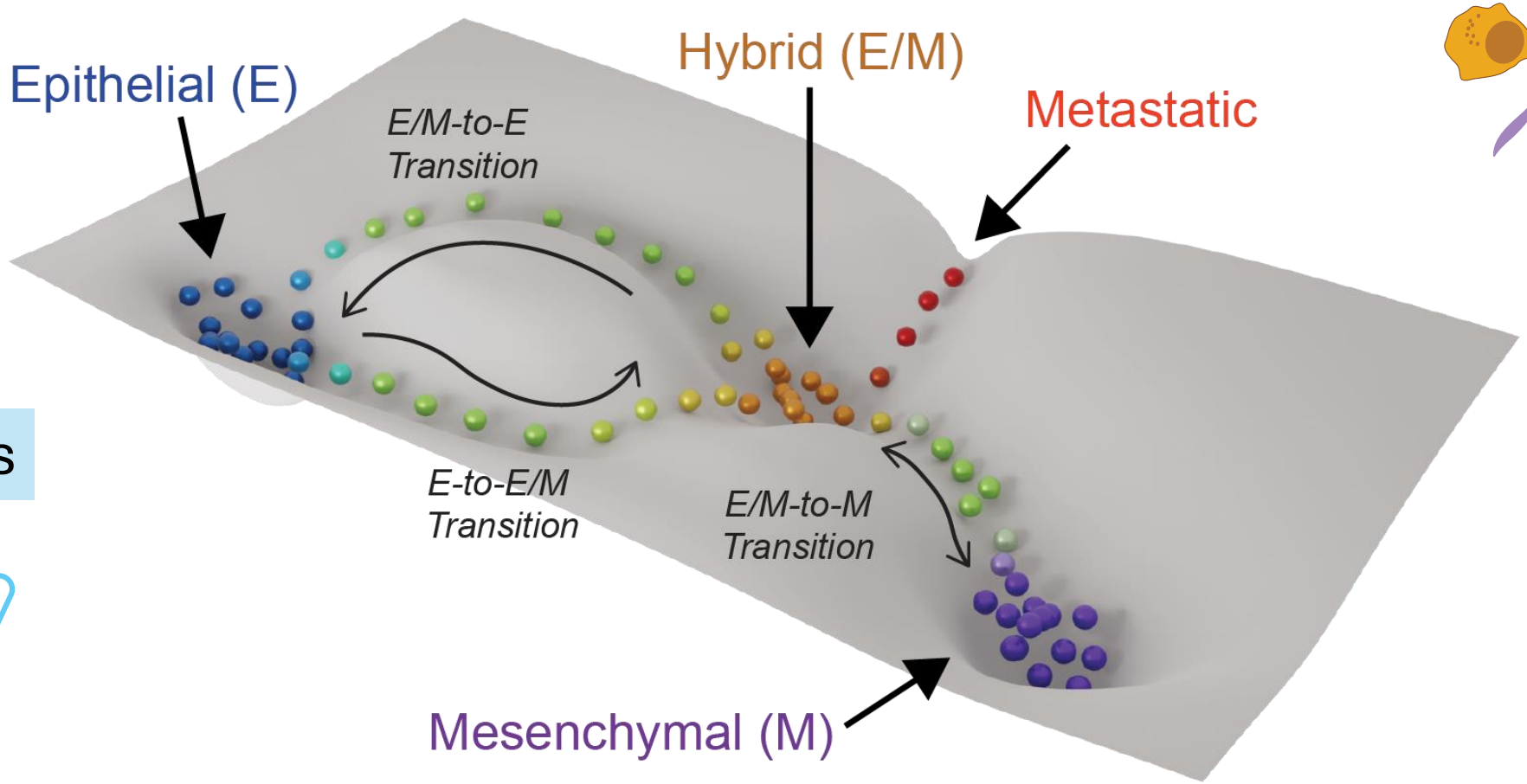


Cell states are based on genomics/epigenomics/signaling

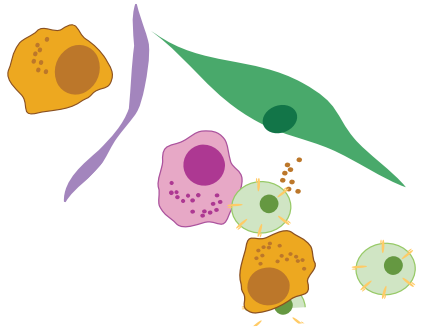


These states are defined genomically (transcriptome, epigenome)

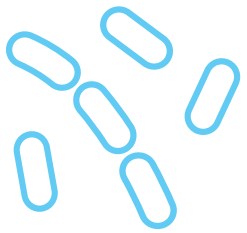
Therapies can cause switches between states



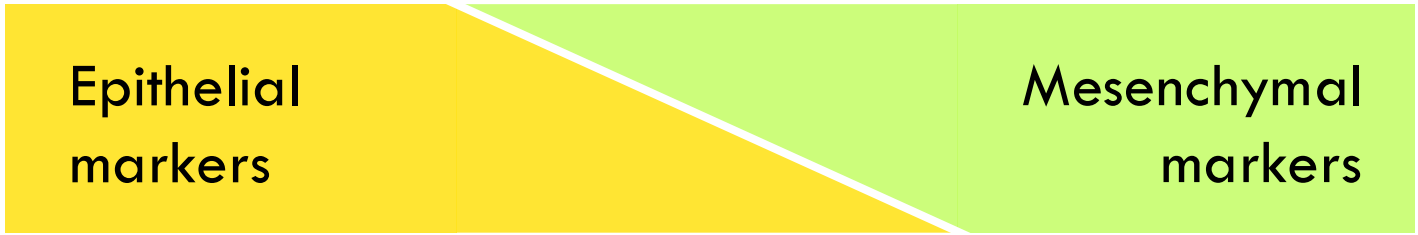
Microenvironment



Therapies



Cancer metastasis often involves such transitions



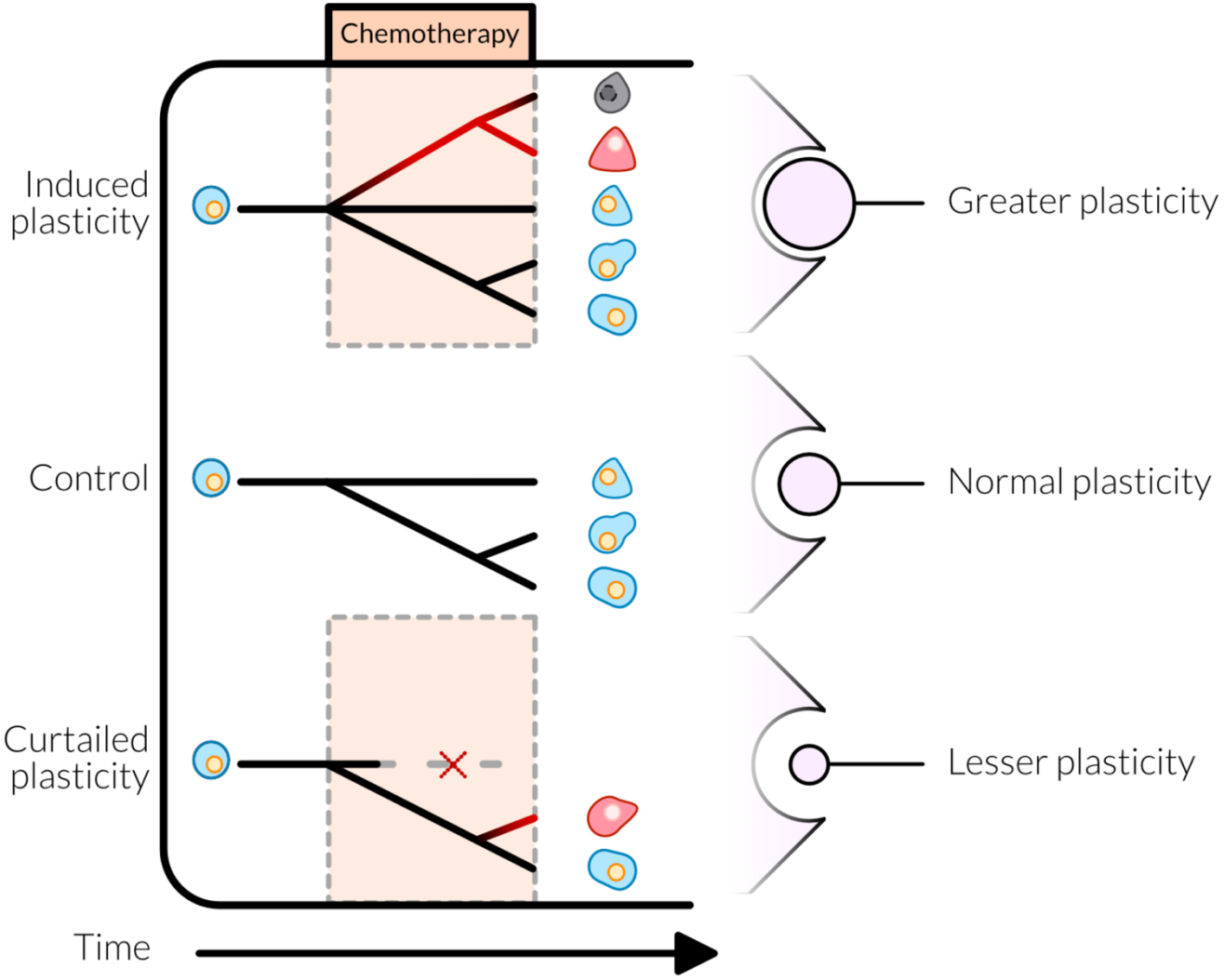
e.g. E-cadherin, Laminin-1, Syndecan

e.g. Vimentin, Snail, Zeb N-Cadherin

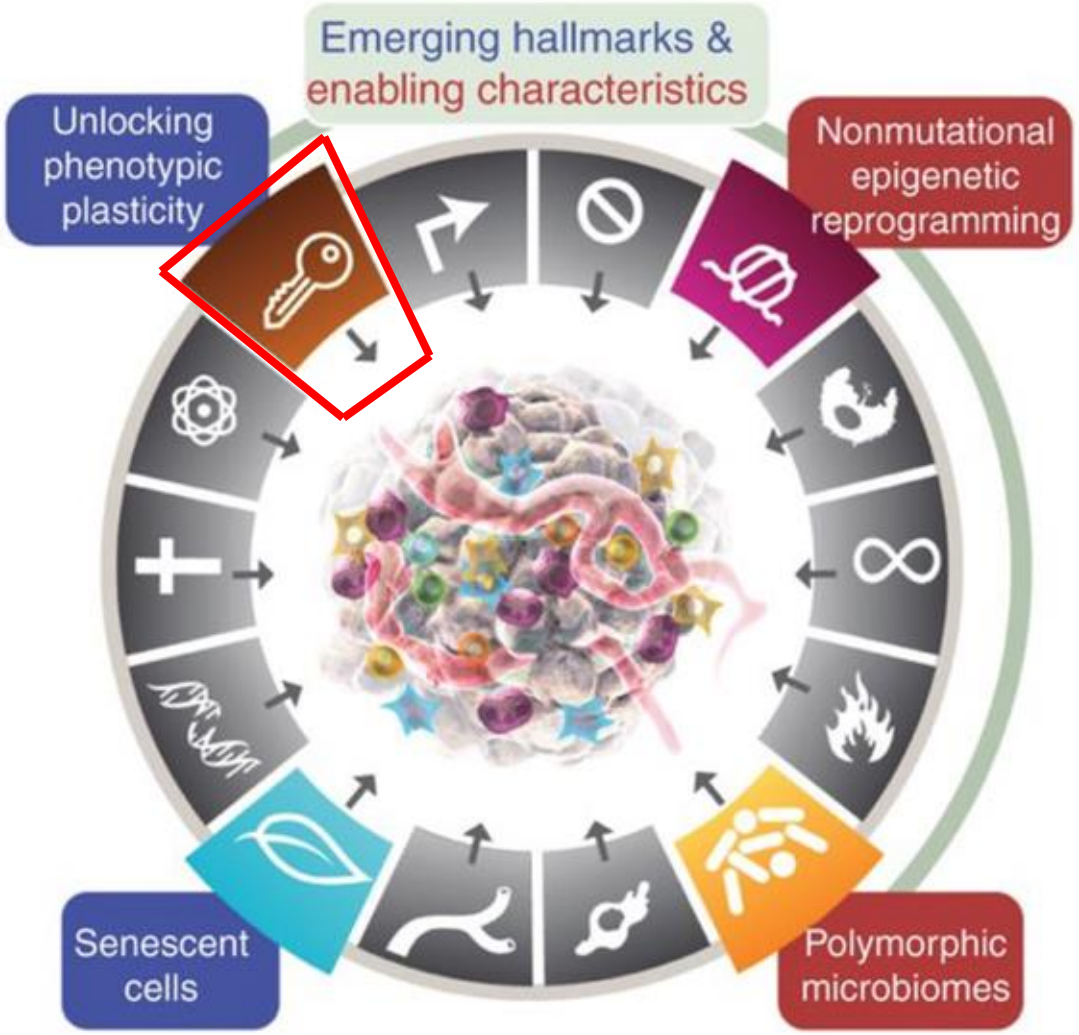
These are characterized by transcriptomic and epigenomic and signaling changes-–not necessarily genetic.

Cancer Plasticity

Plasticity refers to the ability to switch between states



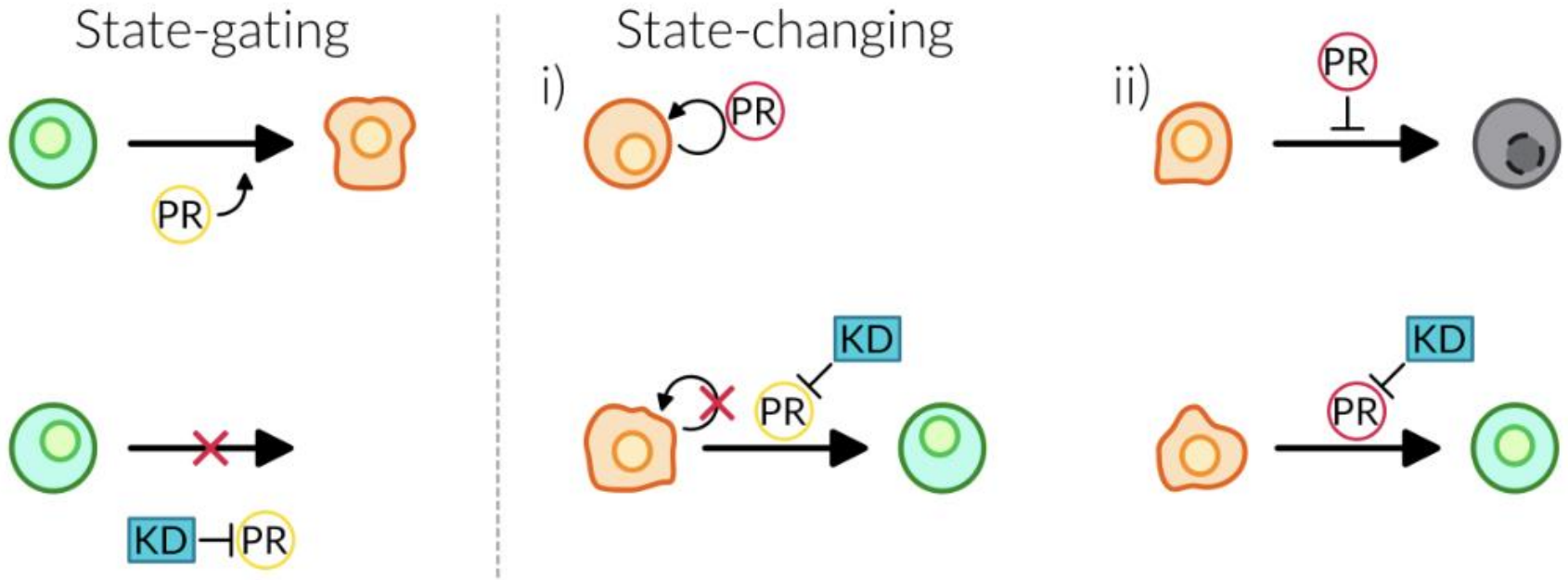
Cellular plasticity now recognized as a hallmark of cancer



- Hallmarks of cancer are established mechanisms by which tumors grow and persist
- Often targets of therapeutic intervention
- Cancer plasticity is a newly recognized hallmark of cancer
- No therapies approved that target cancer plasticity

Hanahan et al. Cancer Discovery, 2022

New Opportunities for Therapy



Connection with mutations

- Mutations can result altered plasticity landscapes
- Many mutations can create the same change (convergence in phenotype)
- Number of positions on a genome vs intrinsic dimensionality of transcriptional space
  - 8000 mutations in glioma but only a handful of states

Good news: The state space is mappable from single cell data

MINI REVIEW | AUGUST 05 2022

Mapping Phenotypic Plasticity upon the Cancer Cell State Landscape Using Manifold Learning

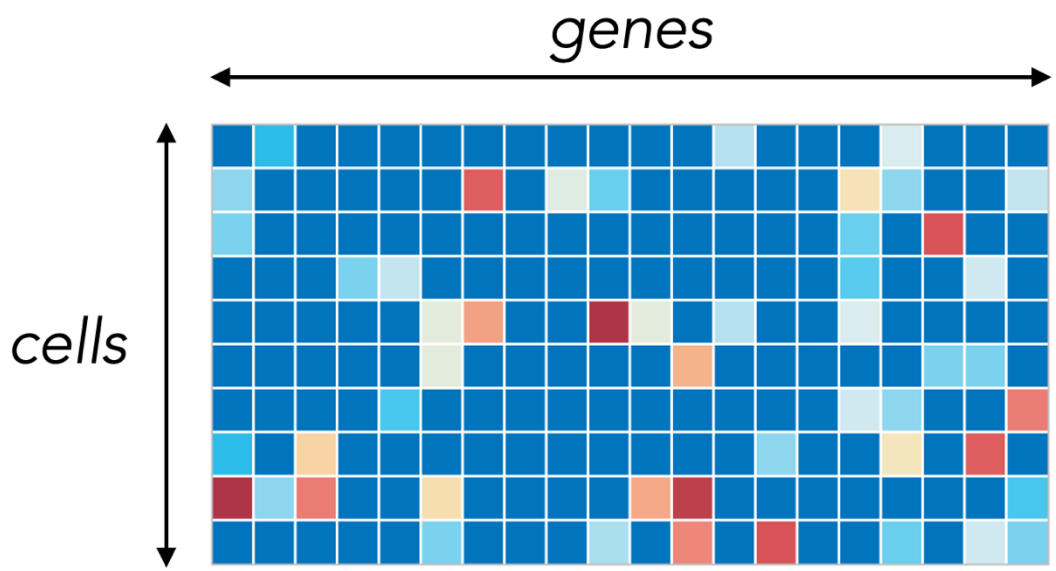
Daniel B. Burkhardt; Beatriz P. San Juan; John G. Lock; Smita Krishnaswamy; Christine L. Chaffer

+ Author & Article Information

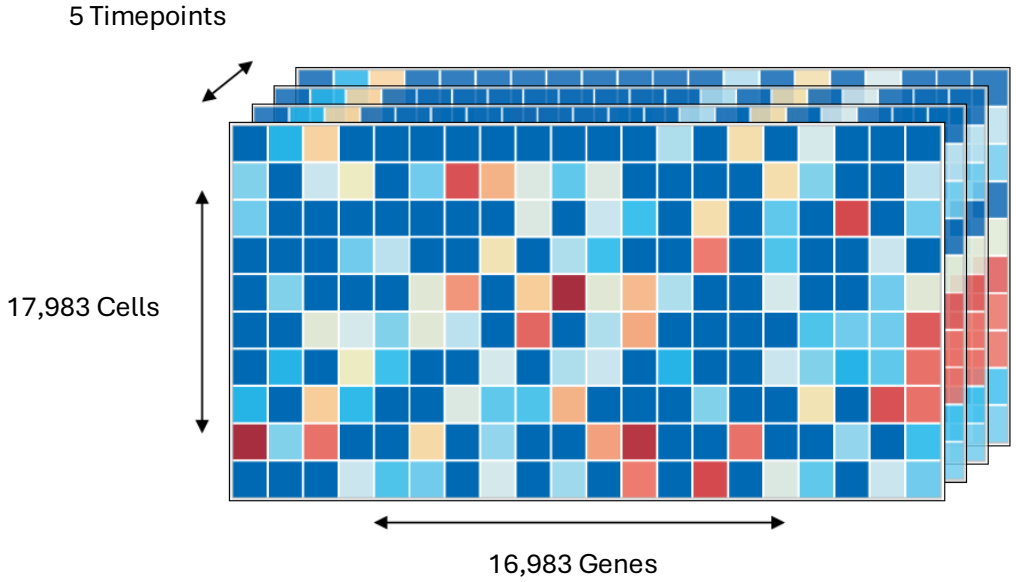
Cancer Discov (2022) 12 (8): 1847–1859.

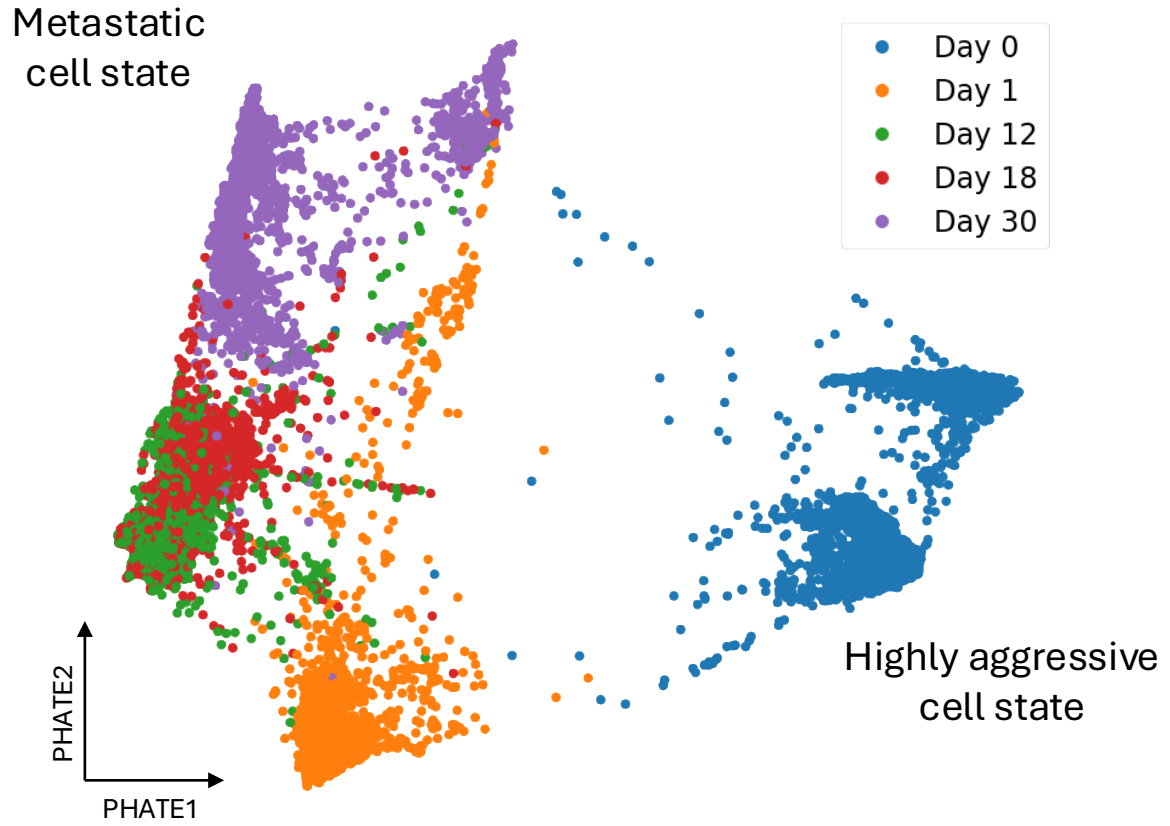
https://doi.org/10.1158/2159-8290.CD-21-0282 **Article history**

The data



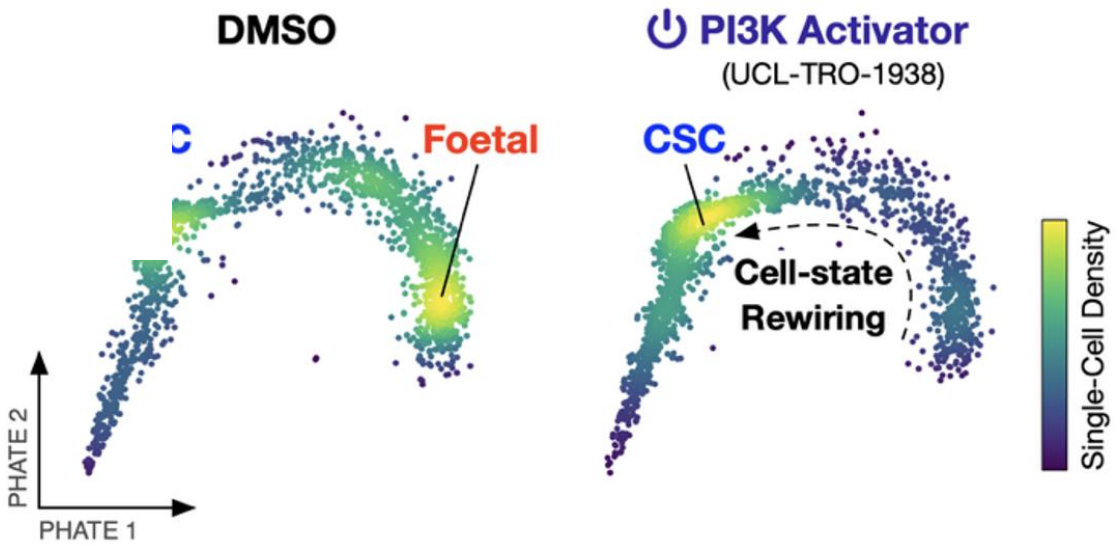
Short time courses are sometimes available



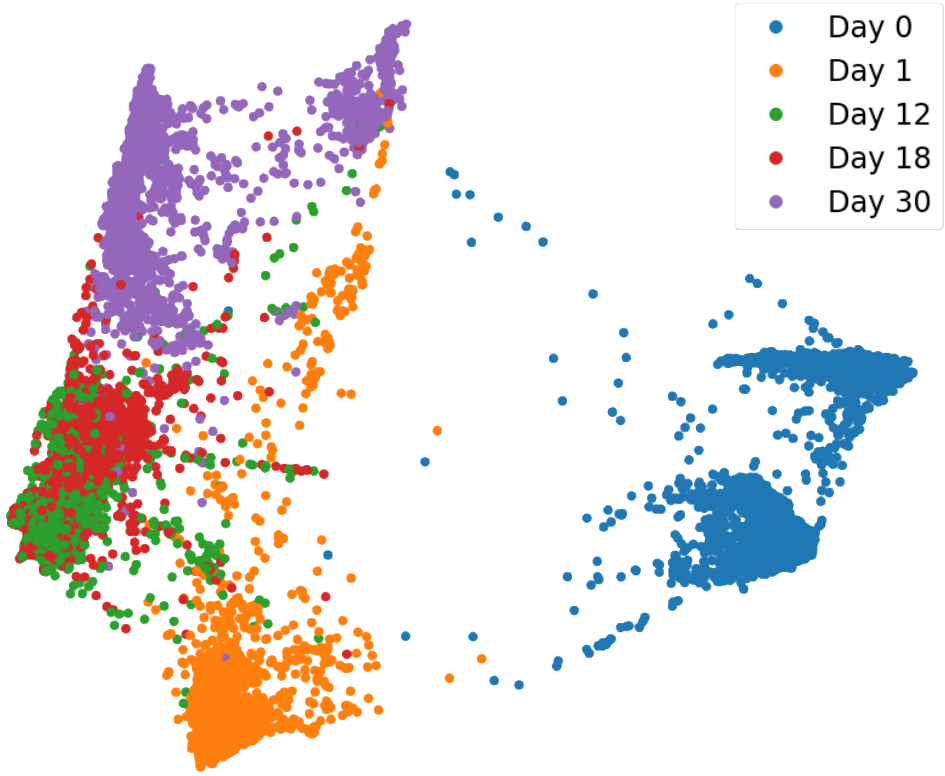


Christine Chaffer Lab

Examples



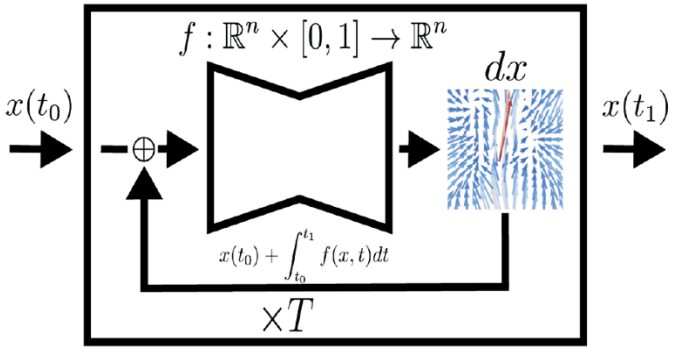
Colorectal cancer



Breast Cancers

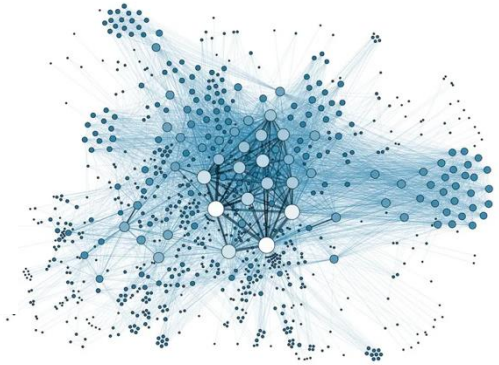
Learning and modulating dynamics

Learn cell-state dynamics



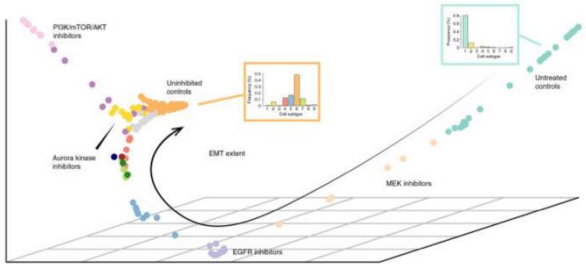
TrajectoryNet
MIOflow

Build networks,
Find targets



Granger
RITINI

Phenoscapes

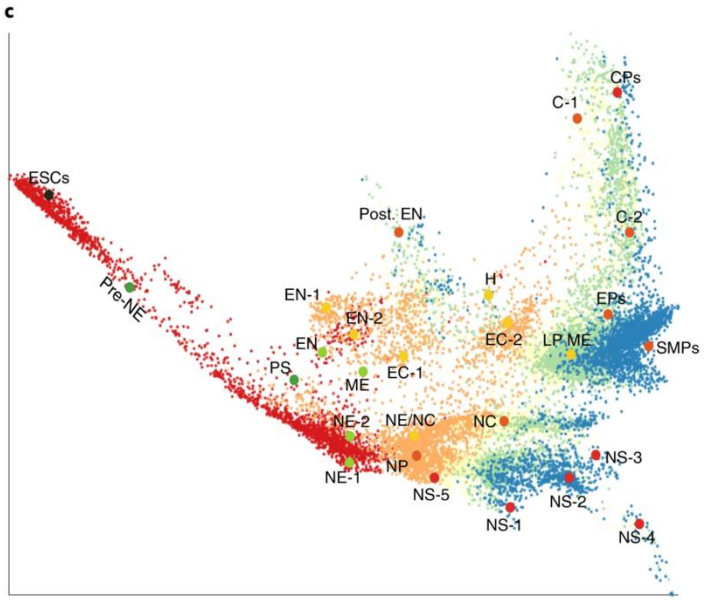


Diffusion EMD
Trellis

Putting data in a geometry preserving space …

PHATE Visualization

(Moon K et al. Nature Biotechnology 2019)



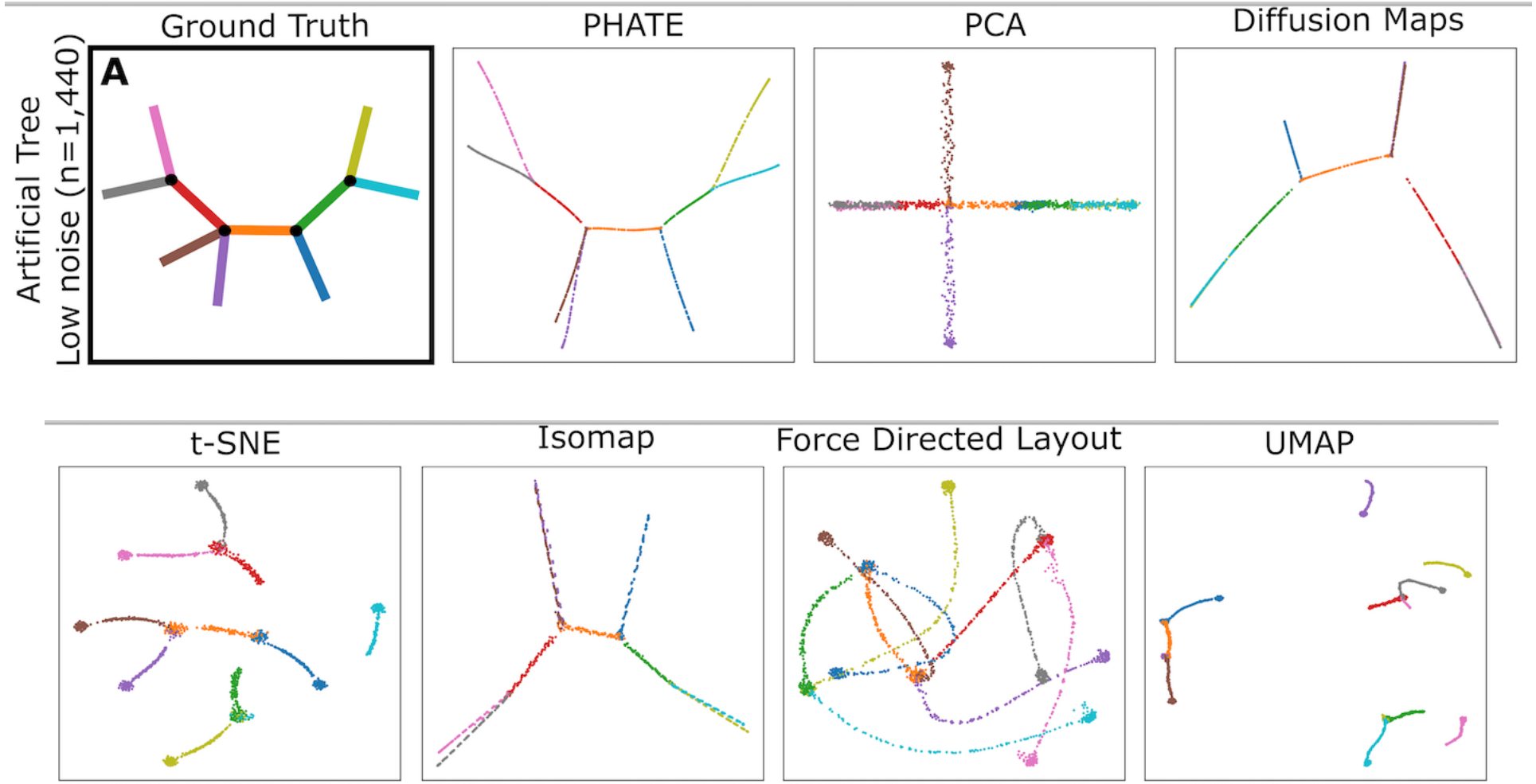
Kevin Moon
David van Dijk
Zheng Wang
Scott Gigante
Dan Burkhardt
William Chen
Natalia Ivanova
Guy Wolf
Smita Krishnaswamy

Main idea: a metric-preserving dimensionality reduction algorithm that naturally emphasizes trajectory structure

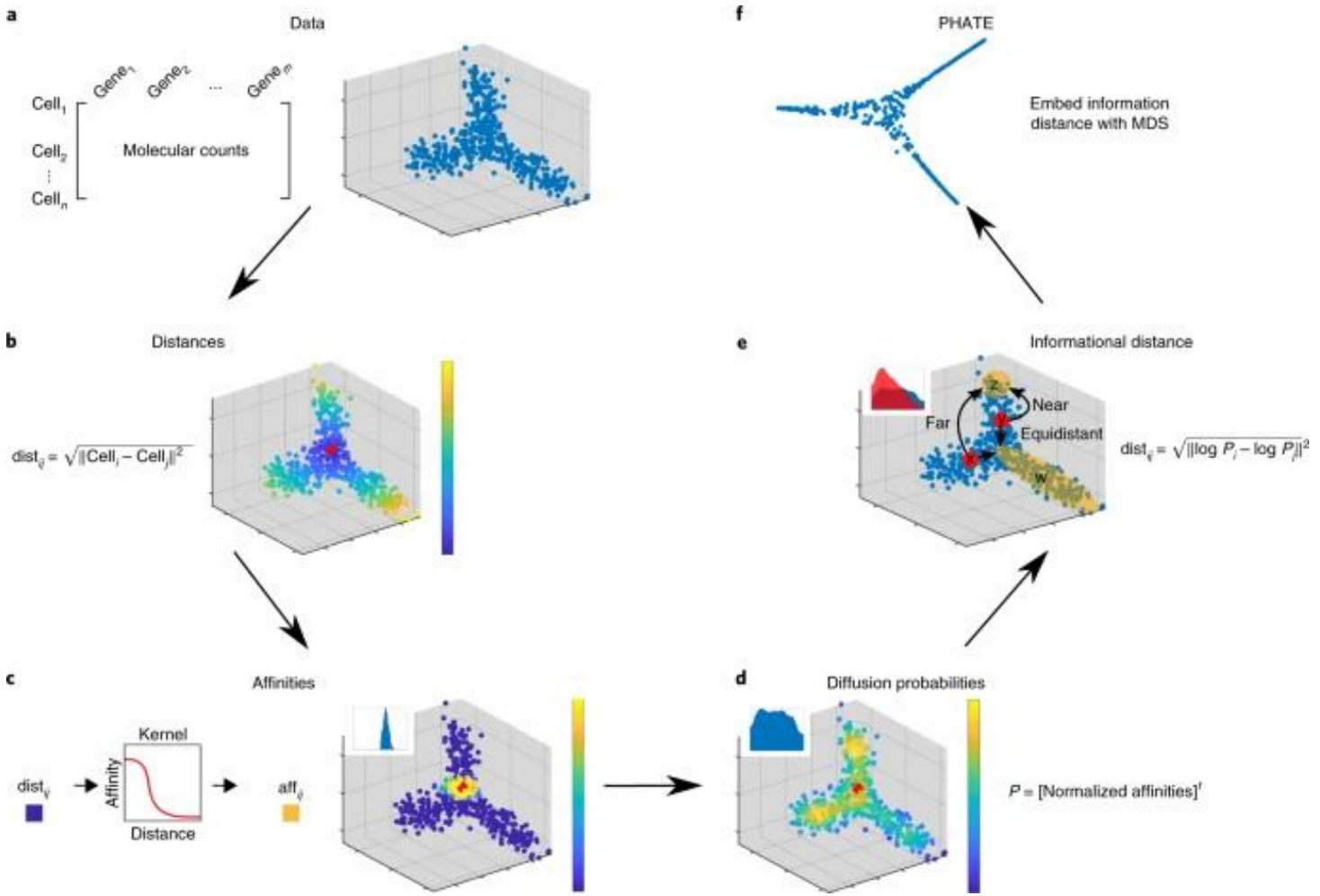
Goal: Create a visualization that preserves data geometry in low dimensions

- Denoise the data – don't let noise cloud the measurements
- Capture geometry– an embedding where distances have meaning (like diffusion maps)
- Compress into visualizable dimensions –allow flexible compressibility into 2/3 dimensions
- Allow for scalability to large modern datasets

Preserving geometry (toy dataset)

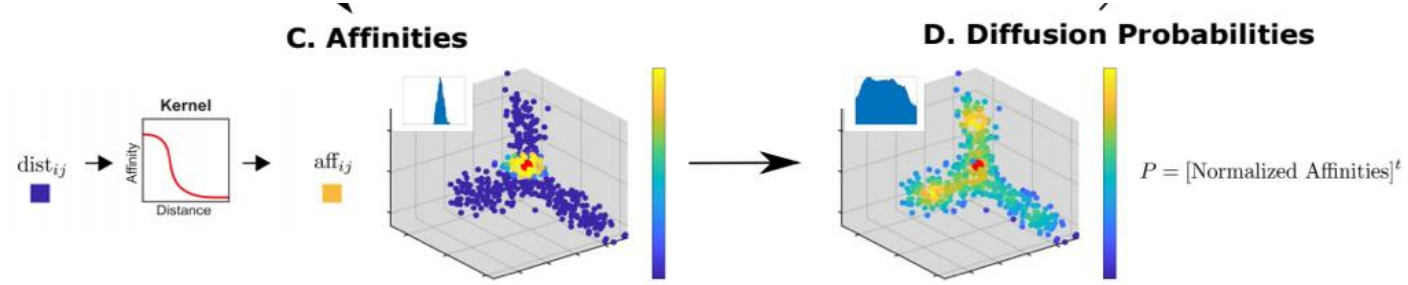


The PHATE Algorithm



Global Contextualization

- Step 0. Denoise the connectivity by diffusion
- Step 1. Each cell is represented as a t-step diffusion probability to ALL other cells



$C_1 = [\ p_{11}\ p_{12}\ \ldots\ldots p_{1n}]$
$C_2 = [\ p_{21}\ p_{22}\ \ldots\ldots p_{2n}]$
...
$C_n = [\ p_{n1}\ p_{n2}\ \ldots\ldots p_{nn}]$

- Step 3. Compute potential distance using diffusion

Probabilities

Step 4. Preserve in low dimensions using Metric MDS

Potential Distance and PHATE

Definition *The t-step* ***potential distance is*** $\wp^t(x,y) = ||\log(P^t(x,.)) - log(P^t(y,.))||_2$

Proposition [Moon et al] *Given a diffusion process defined by a fixed-bandwidth Gaussian kernel, the potential distance can be written as follows.*

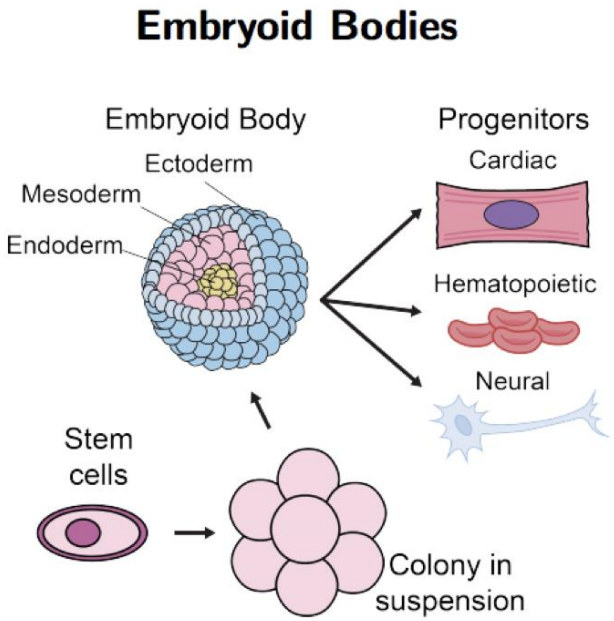
$$\wp^t(x,y) = \sum_{z \in X} log^2 \left(\frac{1+<\Phi^{\frac{t}{2}}(x), \Phi^{\frac{t}{2}}(z)>}{1+<\Phi^{\frac{t}{2}}(y), \Phi^{\frac{t}{2}}(z)>}\right)^{1/2}$$

Key Features of the Potential Distance

$$\wp^t(x,y) = \sum_{z \in X} log^2 \left(\frac{1+<\Phi^{\frac{t}{2}}(x), \Phi^{\frac{t}{2}}(z)>}{1+<\Phi^{\frac{t}{2}}(y), \Phi^{\frac{t}{2}}(z)>}\right)^{1/2}$$

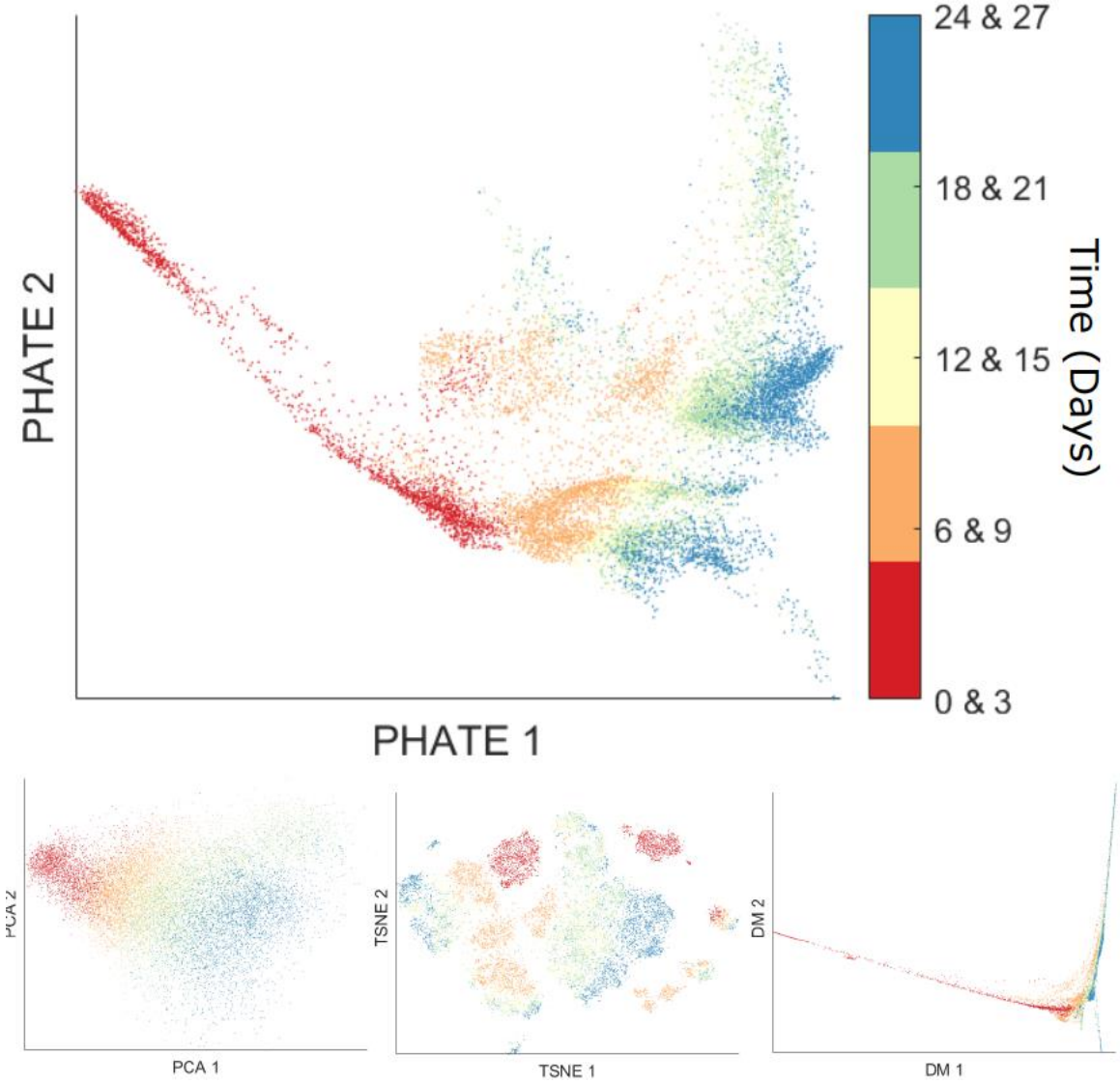
- It is a triplet distance between (x,y) that sums over diffusion inner products to all other points z
- It damps diffusion probabilities using a log factor which increases the importance of far points z in contextualizing (x,y)
- It accounts for the covariance between random walks arising from x vs those arising from y

Stem Cell Development



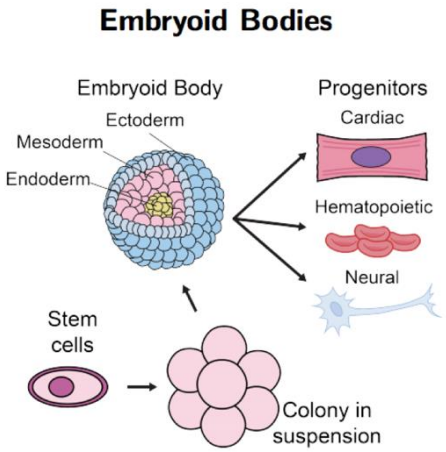
27 day time course

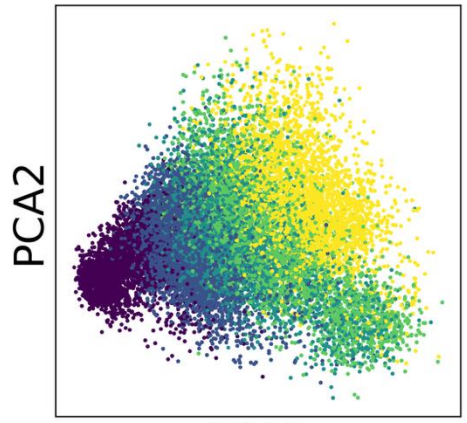
scRNA-seq

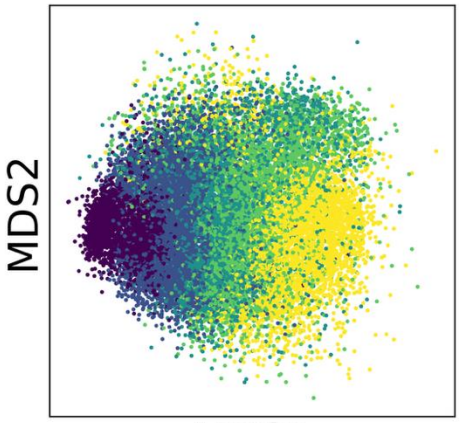


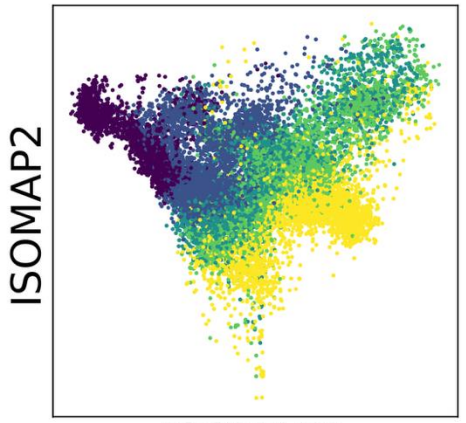
PCA tSNE Diffusion Maps

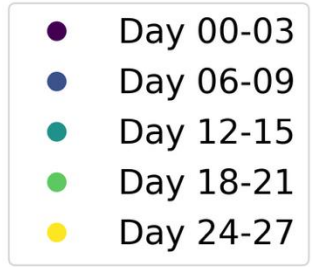
Embryoid Body Comparison

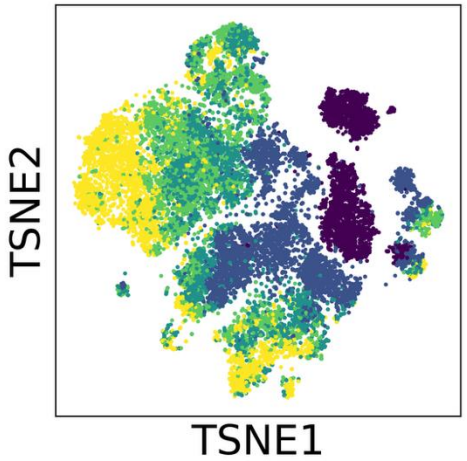


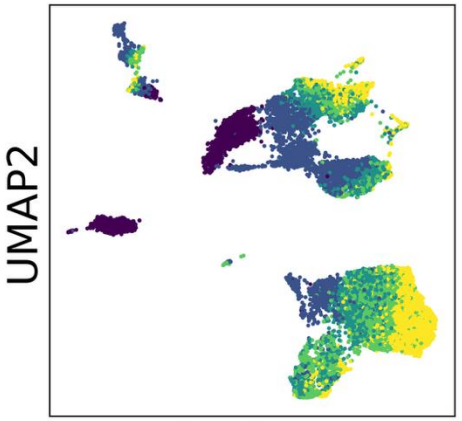


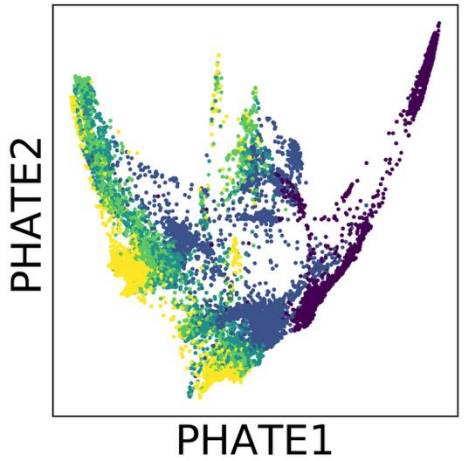


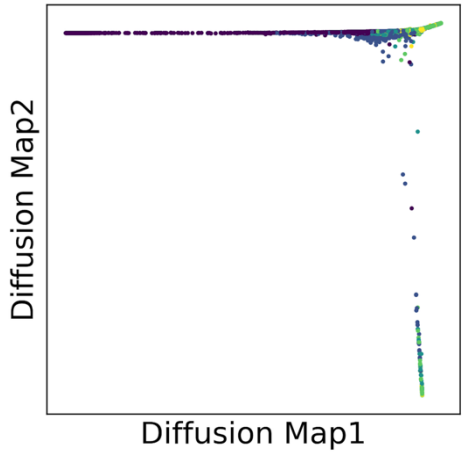




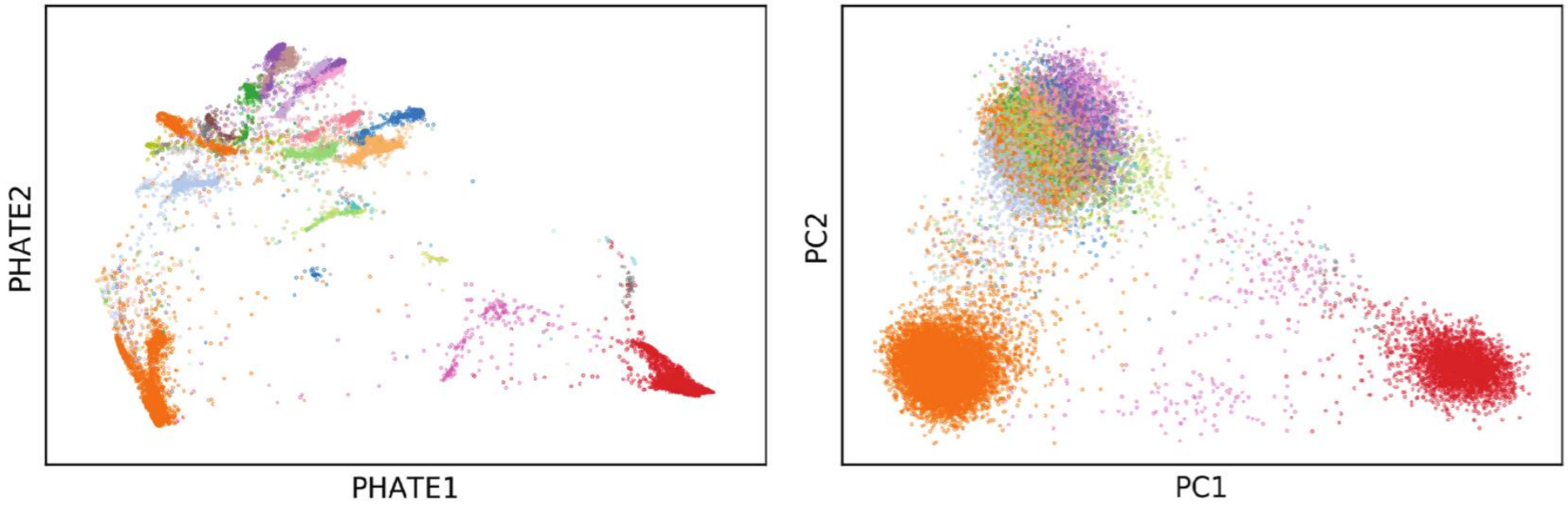








PHATE preserves cluster and trajectory structure, denoises to reveal more local structure



Article | Published: 03 December 2019

Visualizing structure and transitions in high-dimensional biological data

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhar
Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Nata
& Smita Krishnaswamy

Nature Biotechnology **37**, 1482–1492 (2019) | Cite this article

Nat Comput Sci. Author manuscript; available in PMC: 2023 Sep 8.

Published in final edited form as: Nat Comput Sci. 2023 Mar 27;3(3):240–253. doi: 10.1038/s43588-023-00419-0

Multi-view manifold learning of human brain-state trajectories

Erica L Busch [1], Jessie Huang [2], Andrew Benz [3], Tom Wallenstein [2], Guillaume Lajoie [4,5], Guy Wolf [4,5], Smita Krishnaswamy [2,6,7,8,9,✉], Nicholas B Turk-Browne [1,8,9]

Article | Published: 28 February 2022

Multiscale PHATE identifies multimodal signatures of COVID-19

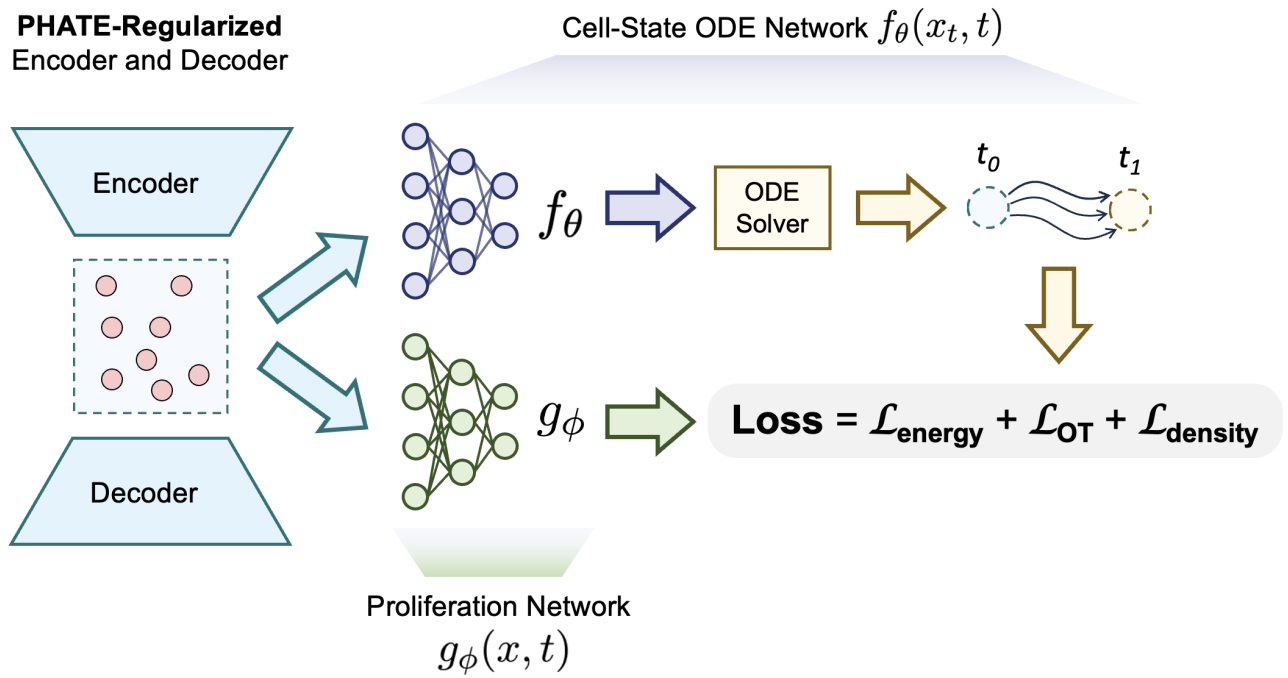
Manik Kuchroo, Jessie Huang, Patrick Wong, Jean-Christophe Grenier, Dennis Shung, Alexander Tong, Carolina Lucas, Jon Klein, Daniel B. Burkhardt, Scott Gigante, Abhinav Godavarthi, Bastian Rieck, Benjamin Israelow, Michael Simonov, Tianyang Mao, Ji Eun Oh, Julio Silva, Takehiro Takahashi, Camila D. Odio, Arnau Casanovas-Massana, John Fournier, Yale IMPACT Team, Shelli Farhadian, Charles S. Dela Cruz, … Smita Krishnaswamy + Show authors

Nature Biotechnology **40**, 681–691 (2022) | Cite this article

Learning dynamics, forecasting future cell states …

Manifold Interpolating flows

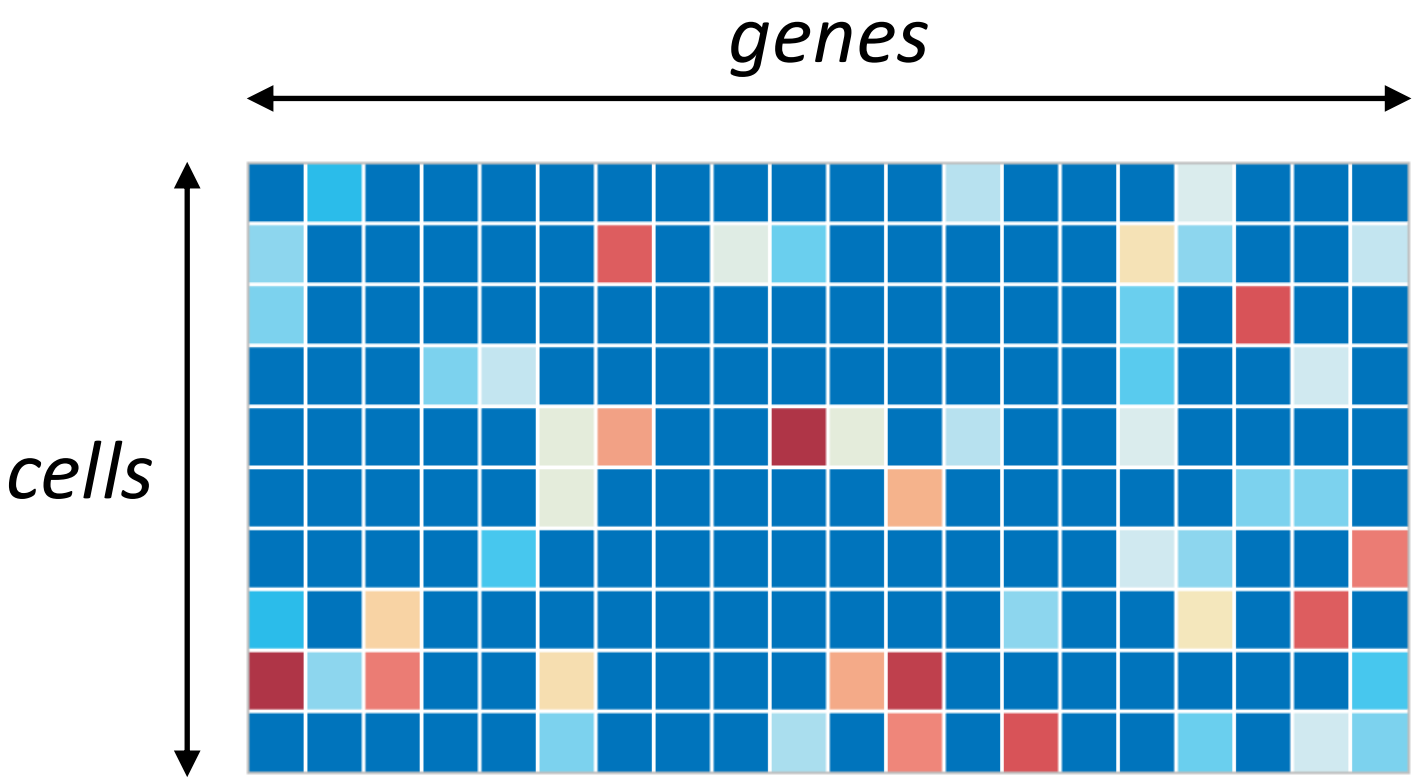
(Huguet, Magruder, et al. NeurIPS 2022, Sun et al. in revision at *Nature Genetics*)

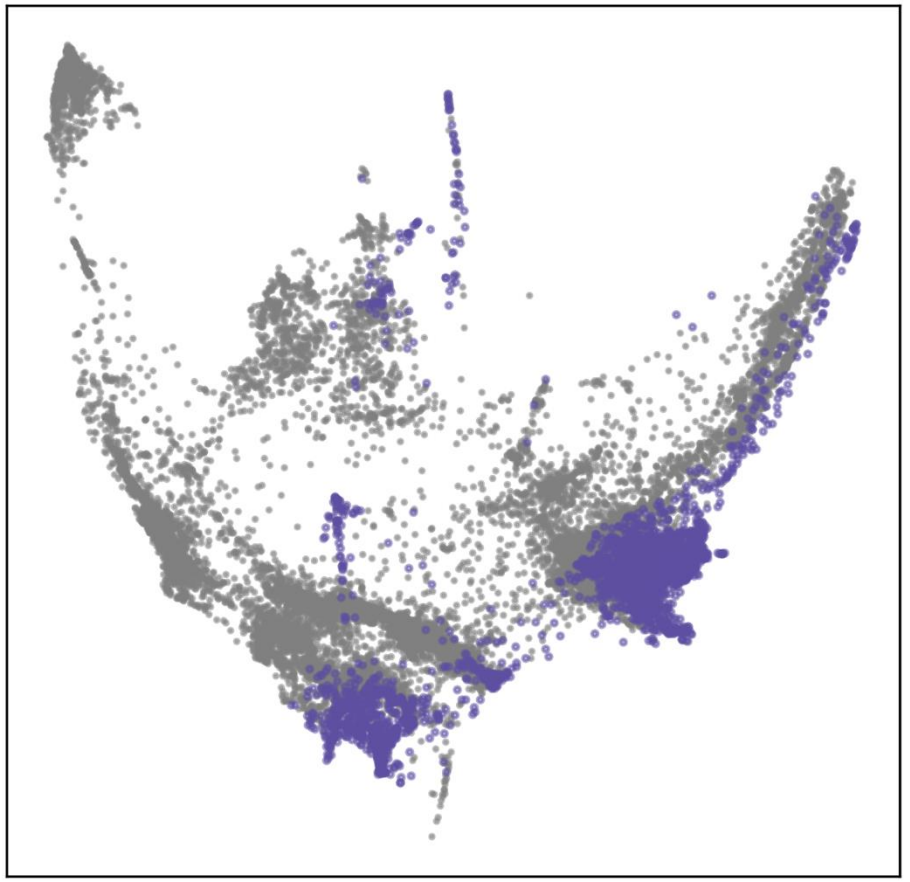


Xingzhi Sun
Guillaume Huguet
Alex Tong
Manik Kuchroo
Sumner Magruder
Guy Wolf

Main idea: Flow data in a manifold embedding penalized by optimal transport

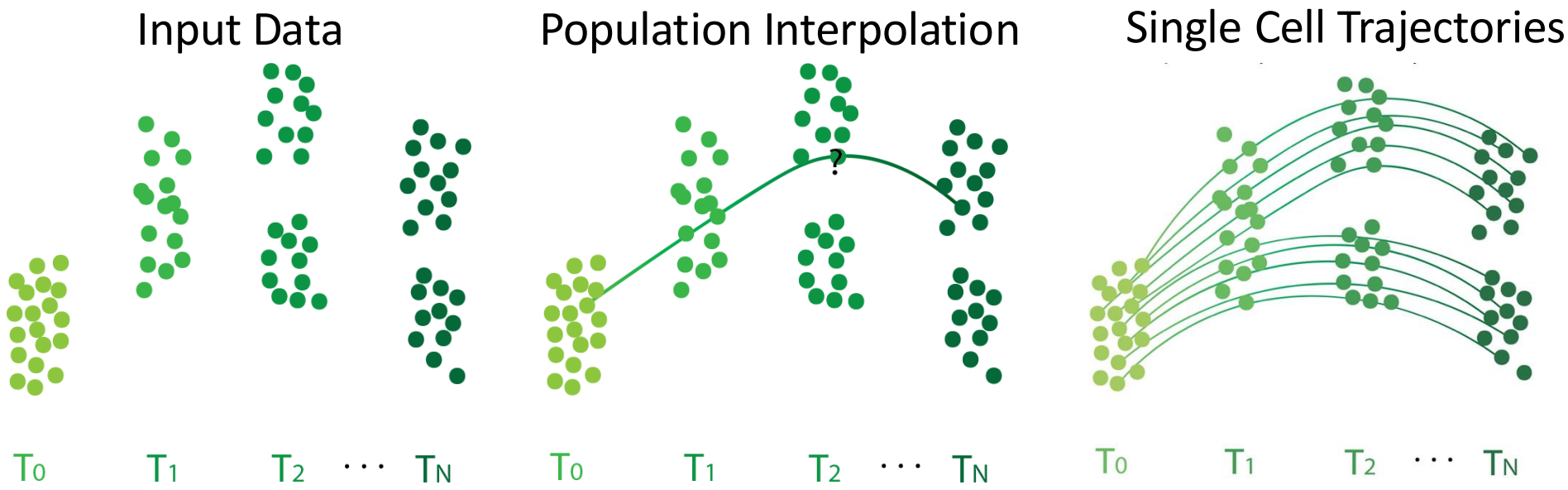
Inferring Continuous Flow in Static Snapshots





Natalia Ivanova Lab

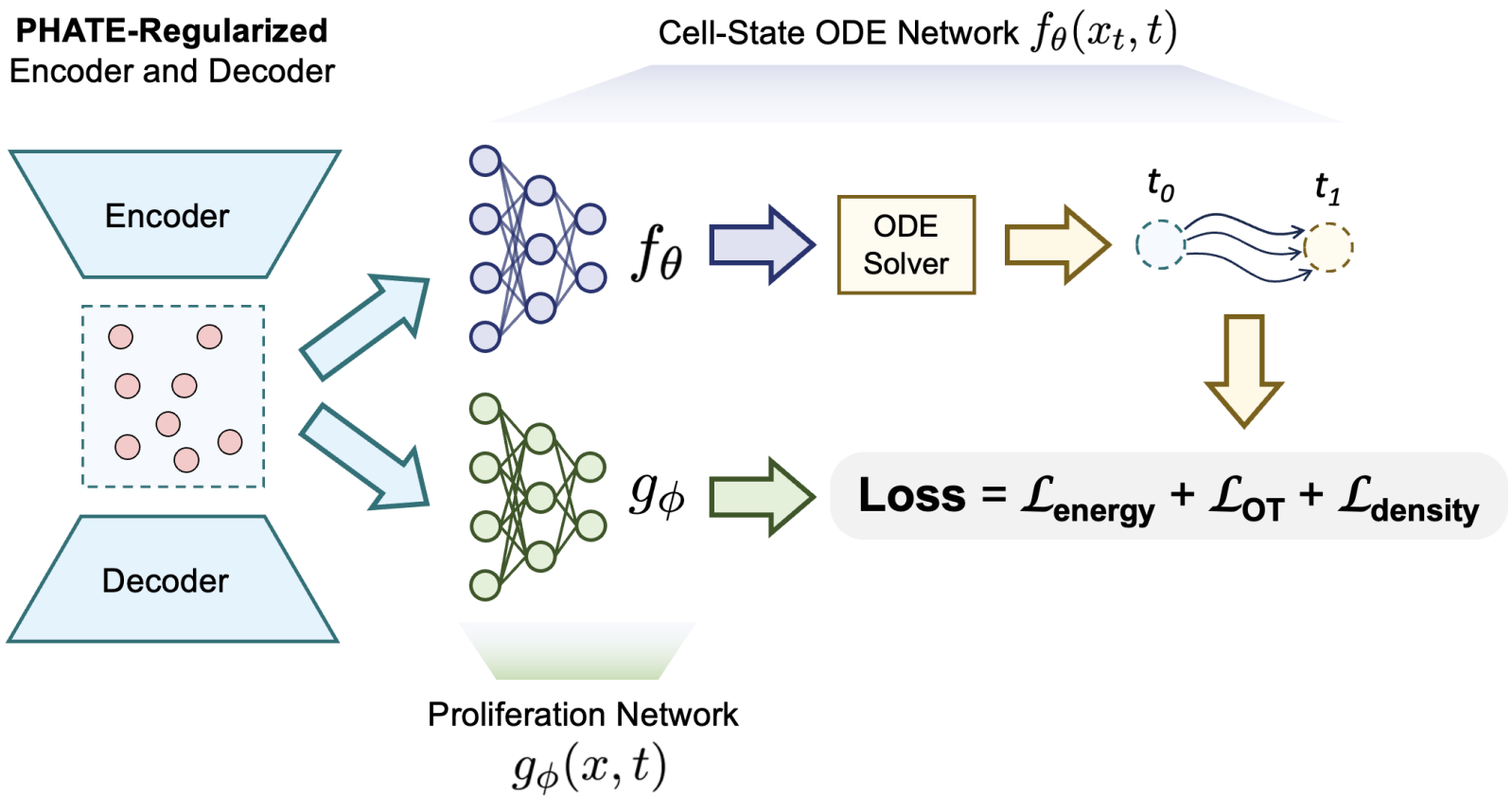
Modeling Dynamics from static snapshot data



Neural ODEs perform dynamic optimal transport to “move” the cellular population efficiently from one timepoint to another.

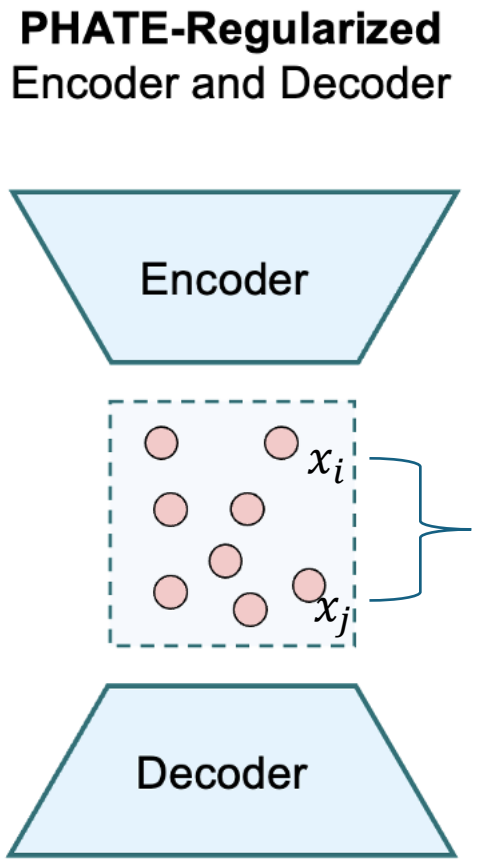
Single cell trajectories are left in the “wake” of this transport.

MIOflow Architecture



[Huguet et al. NeurIPS 2022, Sun et al. In preparation]

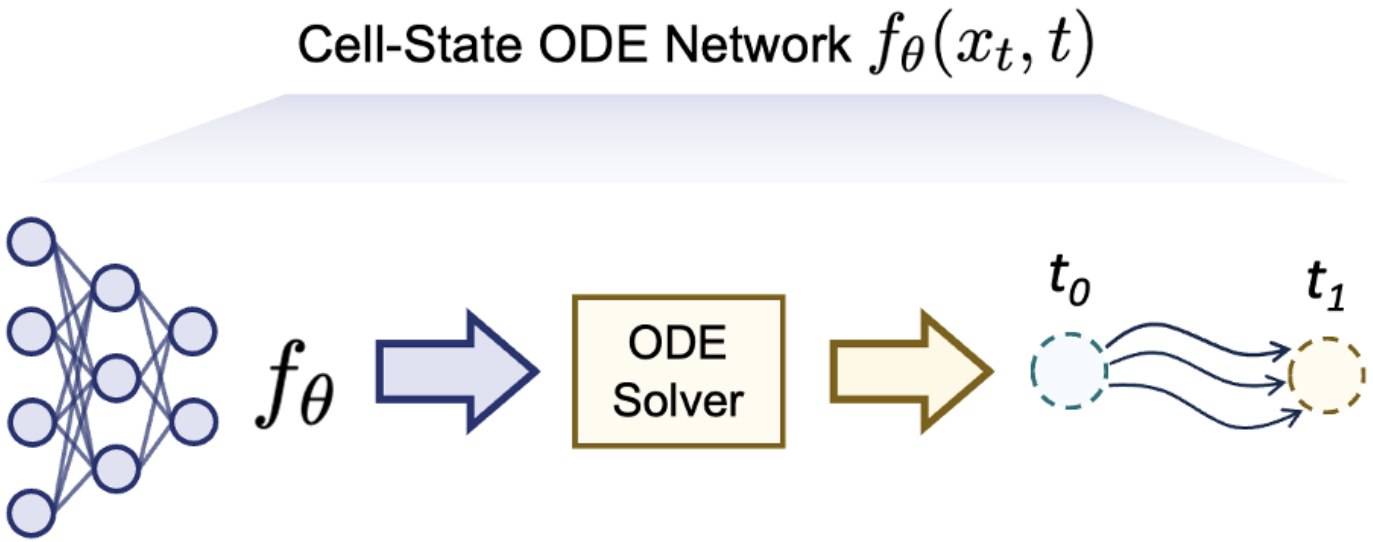
PHATE latent space for MIOflow



Regularized to preserve potential distances in latent space

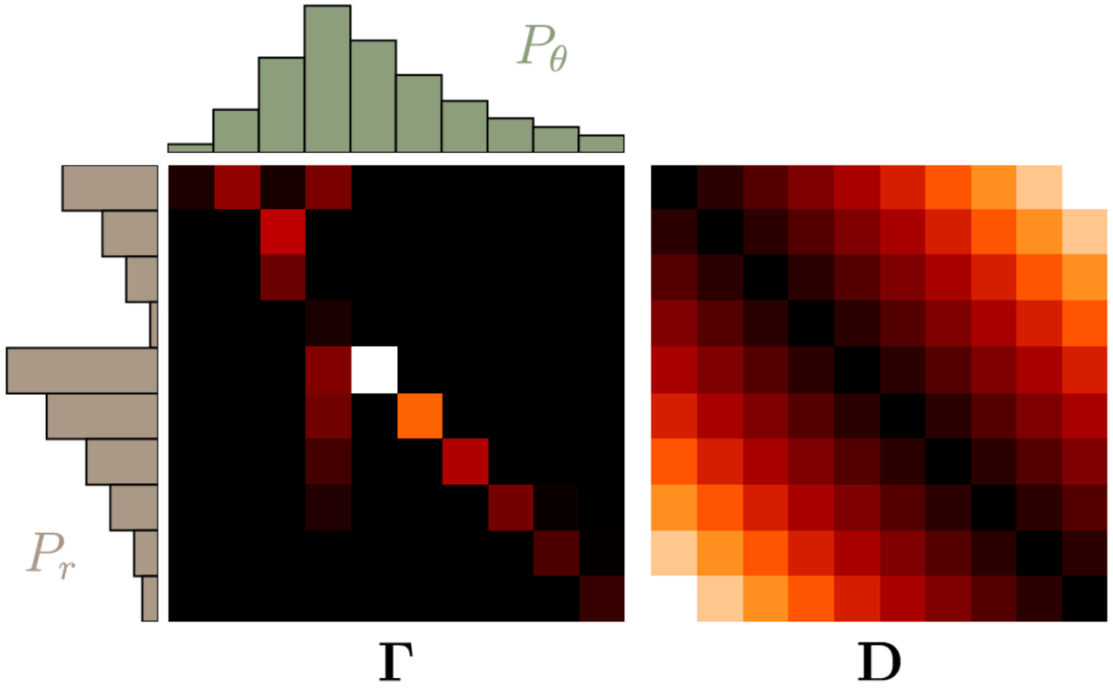
$$L(\phi) = \frac{2}{N}\sum_i \sum_j (||\phi(x_i) - \phi(x_j)||_2 - \wp^t(x_i, x_j))^2$$

Flow (derivative network)



Self-supervised dynamic OT: Penalized by matching marginals at collected timepoints using static OT (or KL-divergence), as well as path smoothness!

Optimal Transport Plan



Minimize

$$c = \sum_{(x,y)} \Gamma_{x,y} .* D_{x,y}$$ Transport cost

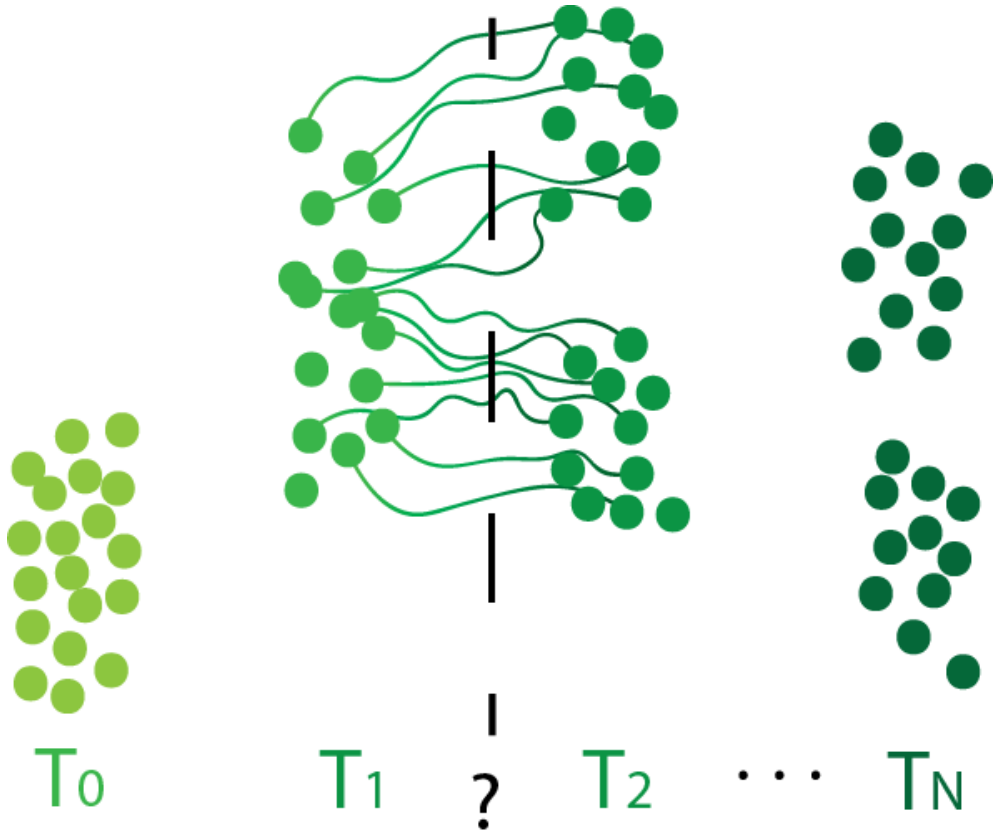
Constraints

$$\sum_x \Gamma(x,y) = P_r(y)$$ Row marginal

$$\sum_y \Gamma(x,y) = P_\theta(x)$$ Column marginal

Minimal cost transport plan defines static optimal transport

Dynamic optimal transport: efficient paths



Circuitous paths are not biological realistic

$$L_e := \lambda_e \sum_{i=1}^{T-2} \int_i^{i+1} \|f(x_t, t)\|_2^2 dt, \text{ where } x_{t_1} \in \mathsf{X}_{t_1}$$

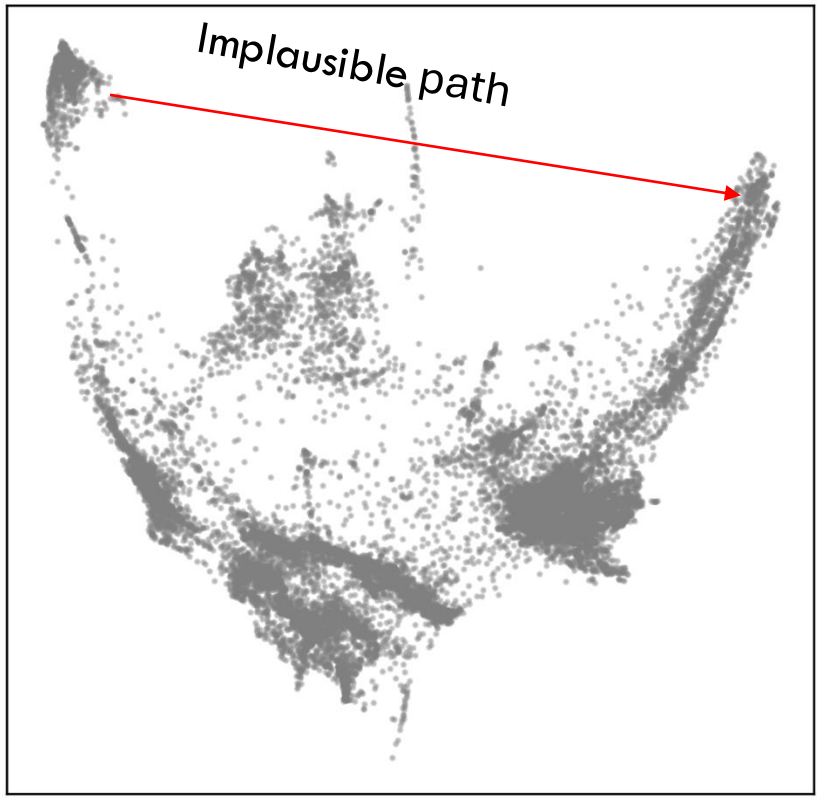
Regularized Flow gives DOT

Thrm [Huguet et al.] We consider a time-varying vector field f(x,t) defining the trajectories $f(X_t, t)dt$ *with density* ρ_t*, and a dissimilarity between distributions that satisfies* $D(\mu, \nu) = 0$ *iff* $\mu = \upsilon$*. Given these Assumptions,* $\exists\ \lambda > 0$ *such that*

$$W_2(\mu, \nu)^2 = \inf_{X_t} E\left[\int_0^1 \|f(X_t, t)\|_2^2 dt\right] + \lambda D(\rho_1, \nu)\ \ s.t.\ \ X_0 \sim \mu.$$

Moreover W_2 *has ground geodesic manifold distance.*

Density loss to enhance transport on manifolds



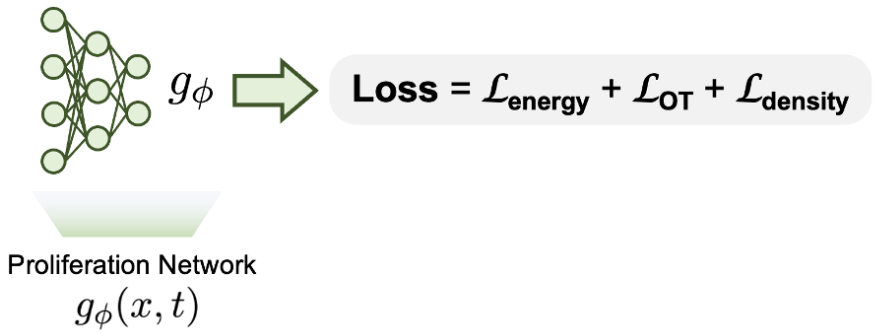
Cells have to transition through allowable parts of the state space

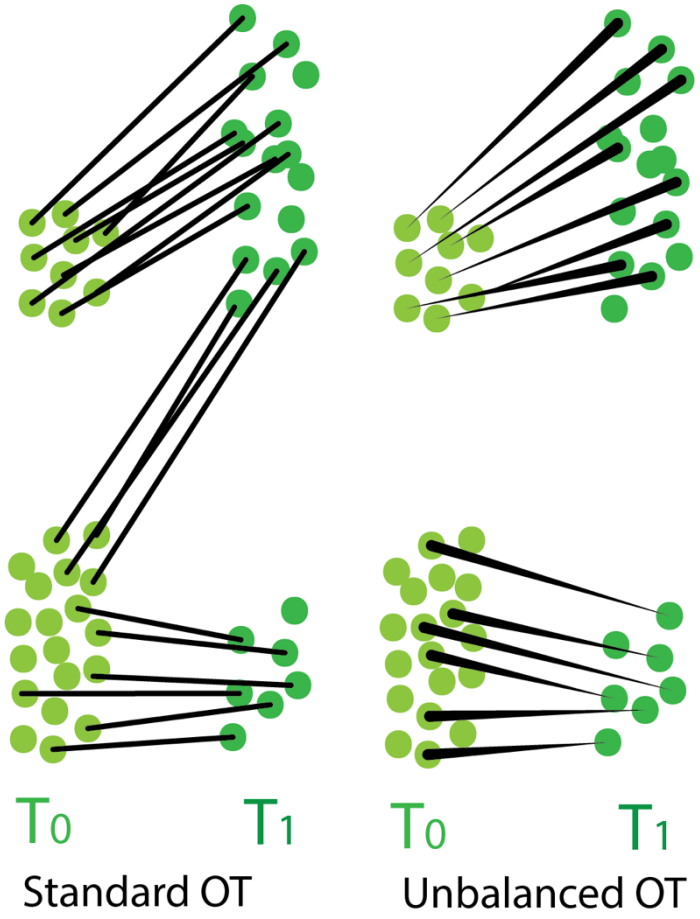
We penalize paths that go through empty space

$$l_{\text{density}}(x,t) = \frac{1}{k}\sum_{j=1}^{k}\max(0, d_{(j)}(x(t)) - h),$$

Cell Death and Growth

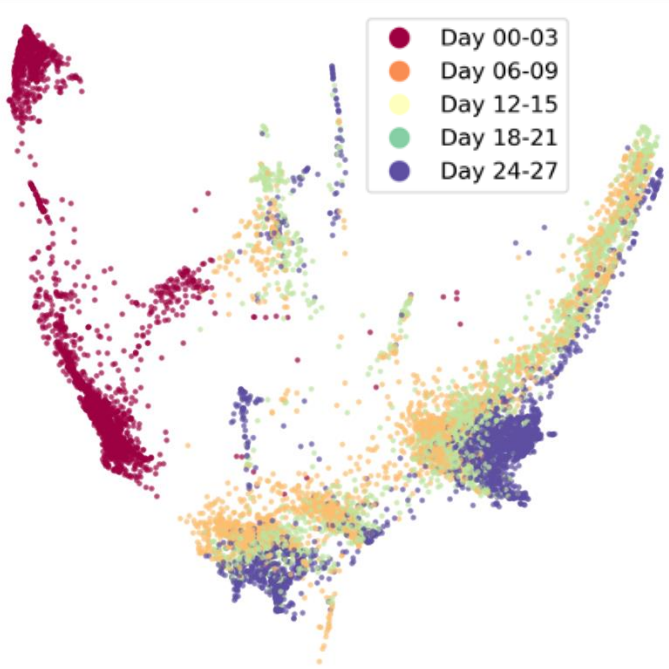
- Allowing cells to “die” instead of moving them to implausible locations
- Learn a growth rate for each cell using another network



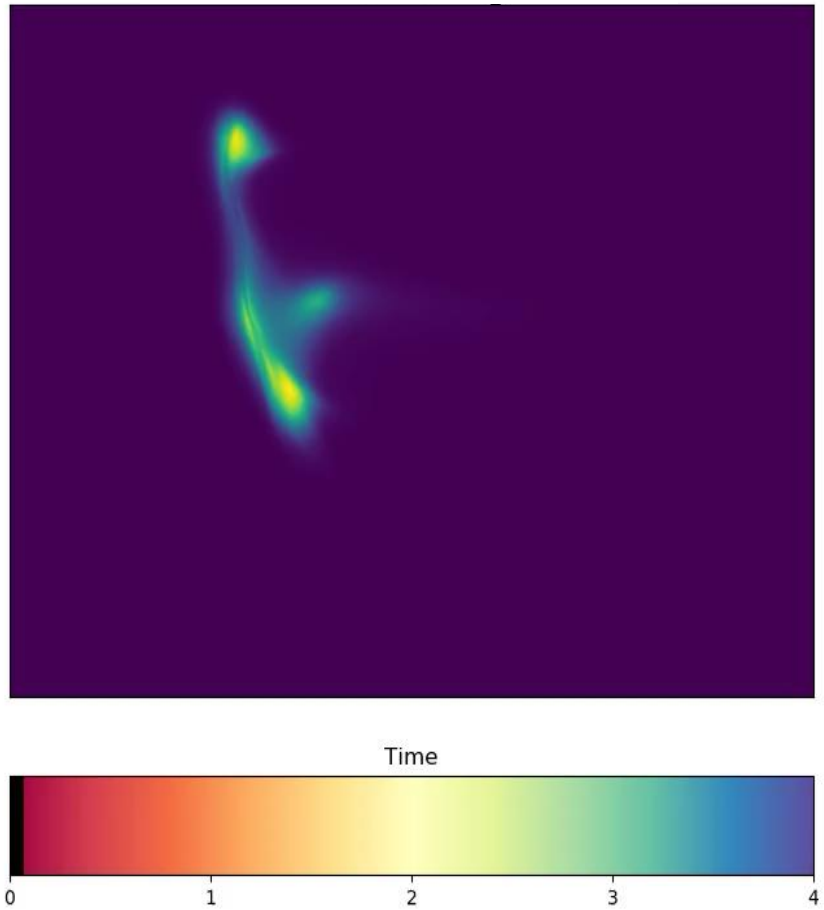


Example: Embryoid body differentiation Trajectories

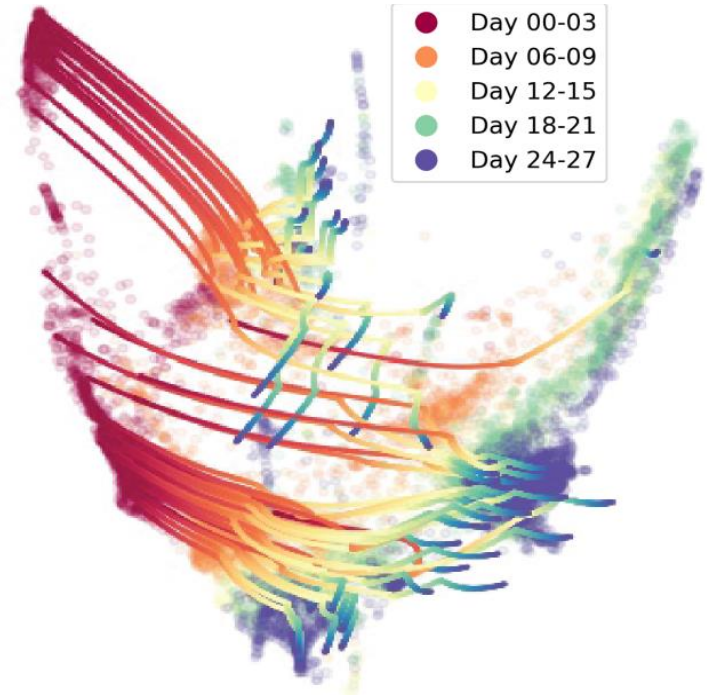
Raw Data



Dynamic OT

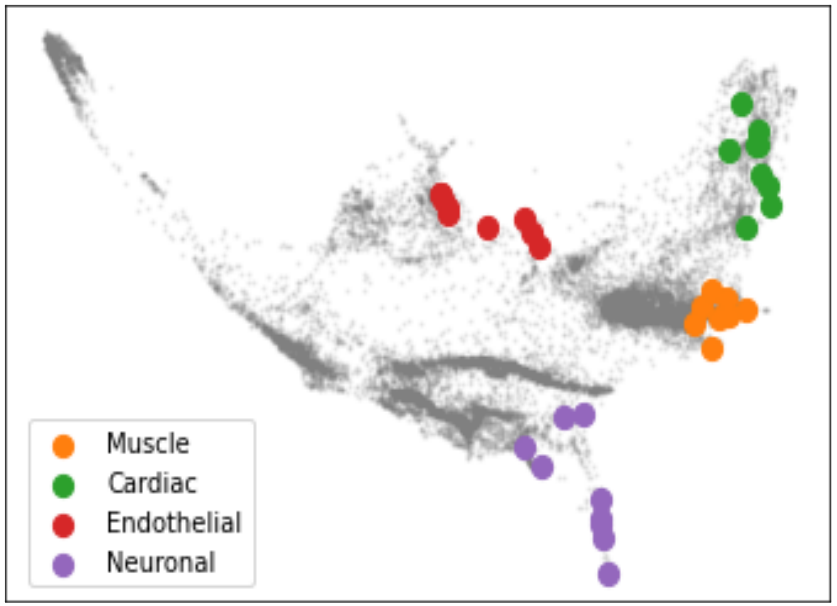


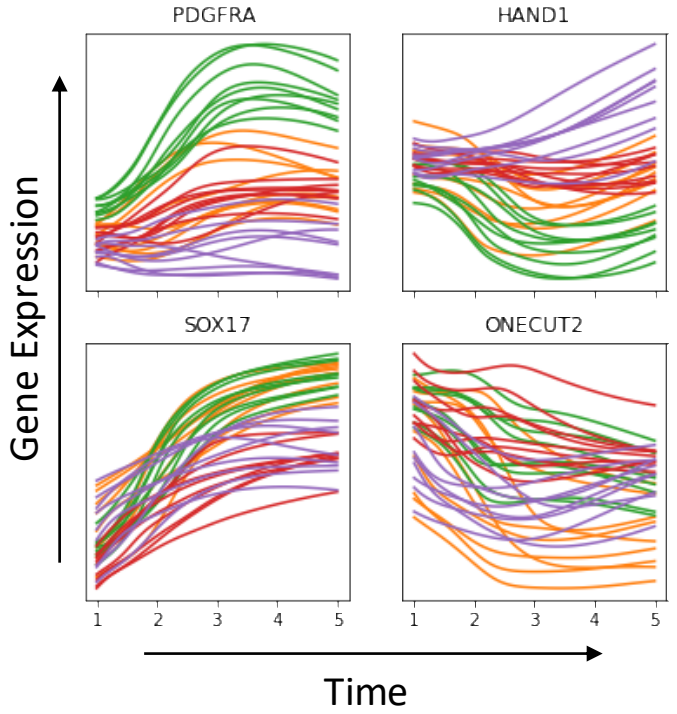
Output



Ivanova Lab

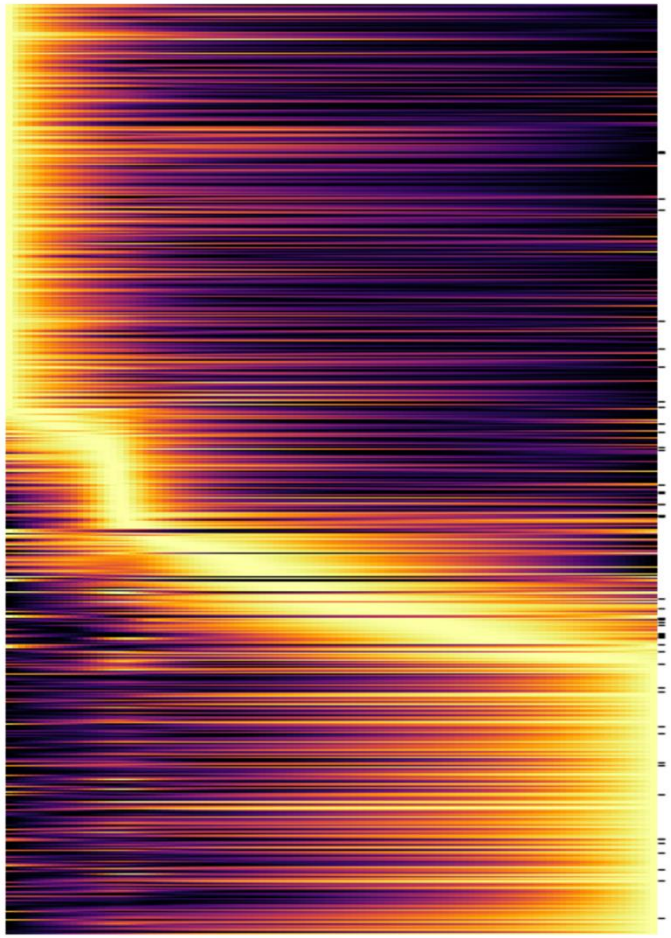
Single gene dynamics come along with cell dynamics



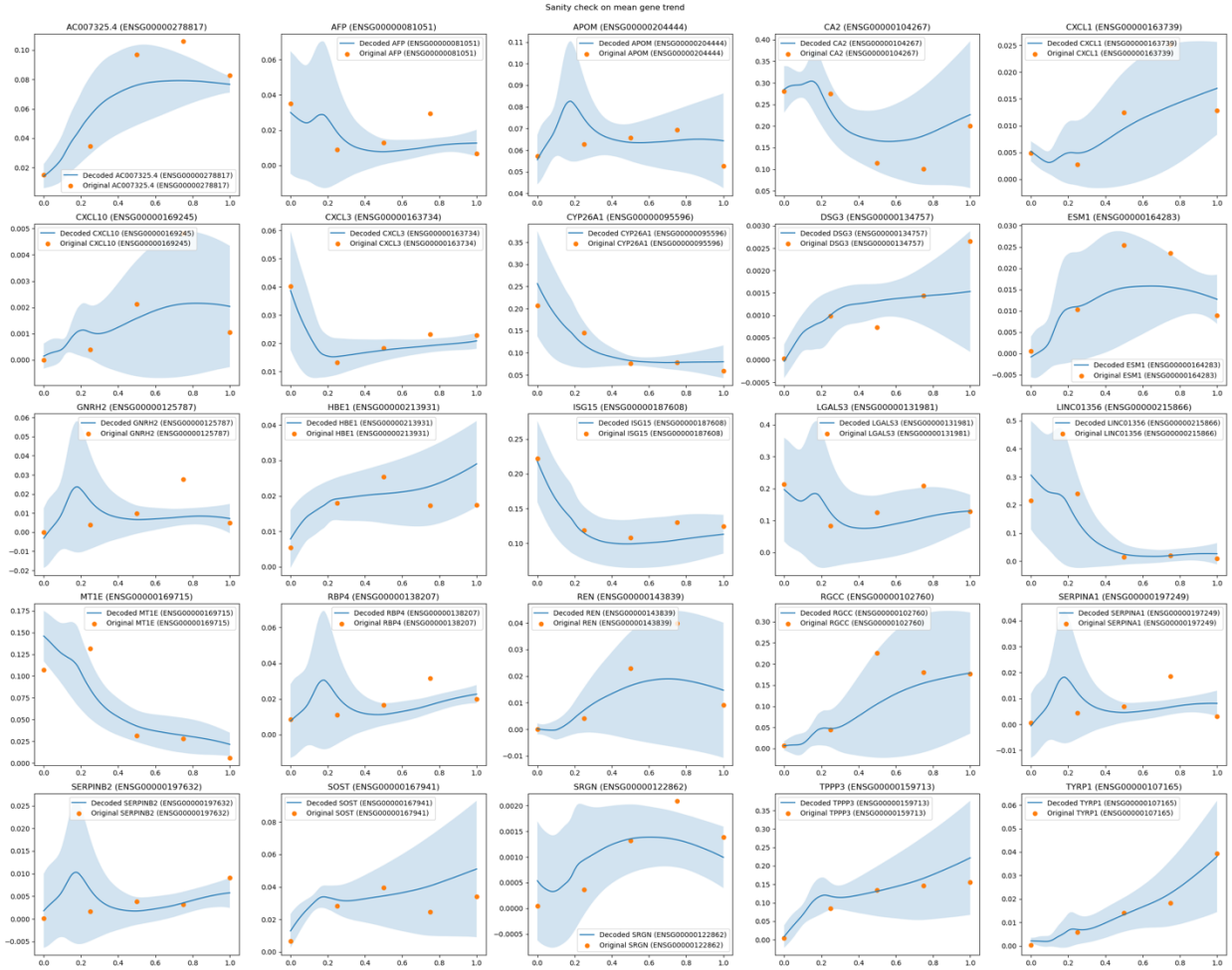


Tong et al. ICML, 2020

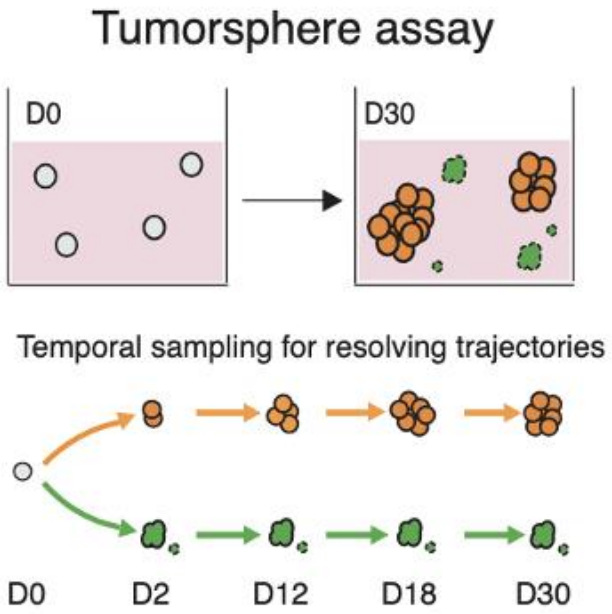
Decoded Gene Trajectories

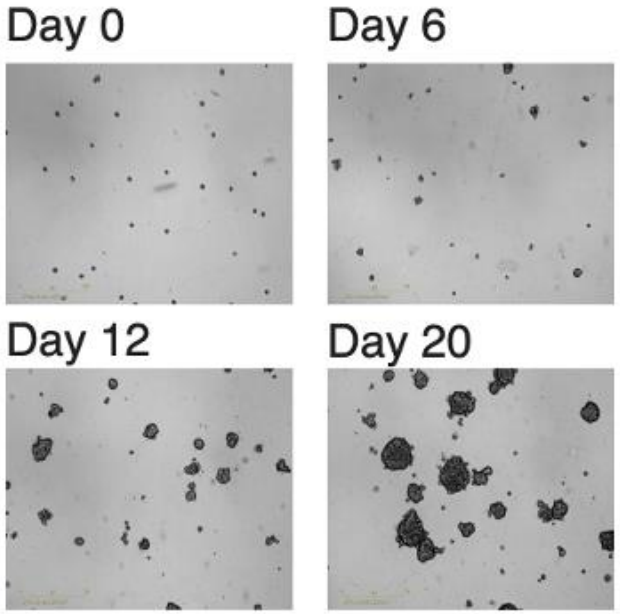


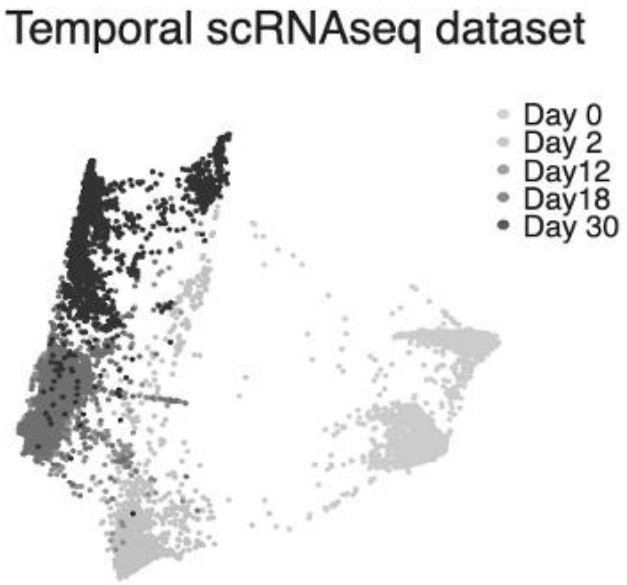
Genes ordered by peak time



Time-course mammosphere experiment

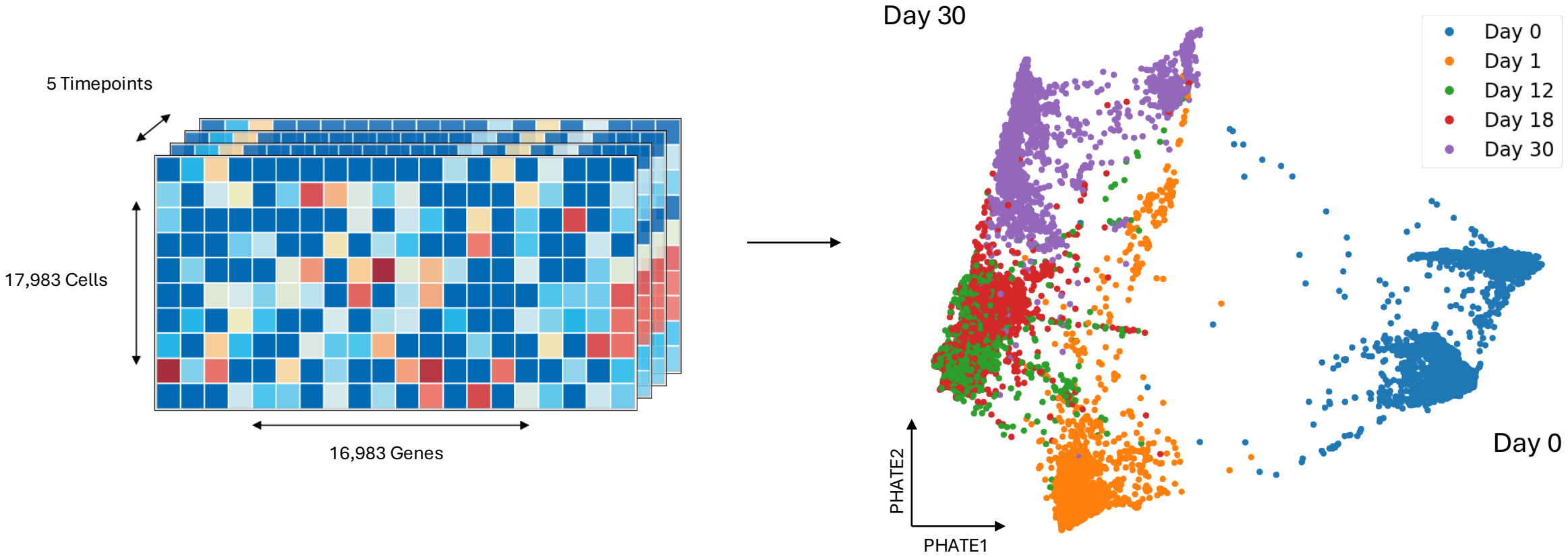




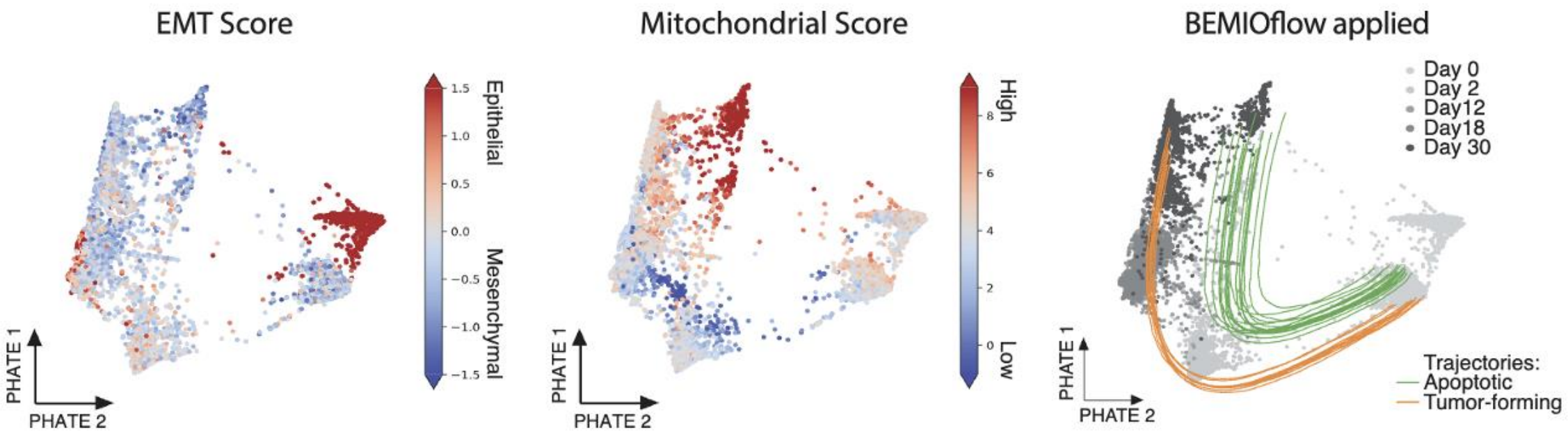


Christine Chaffer Lab

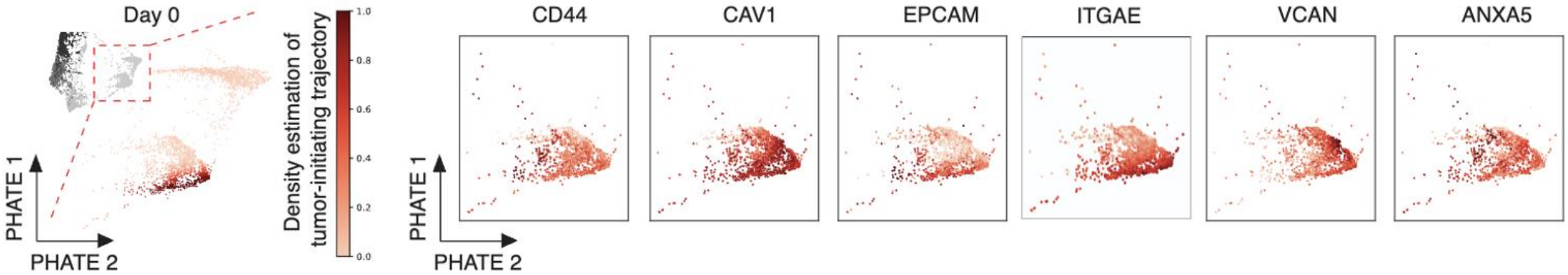
Time course triple negative breast cancer data



Dynamic optimal transport of cells

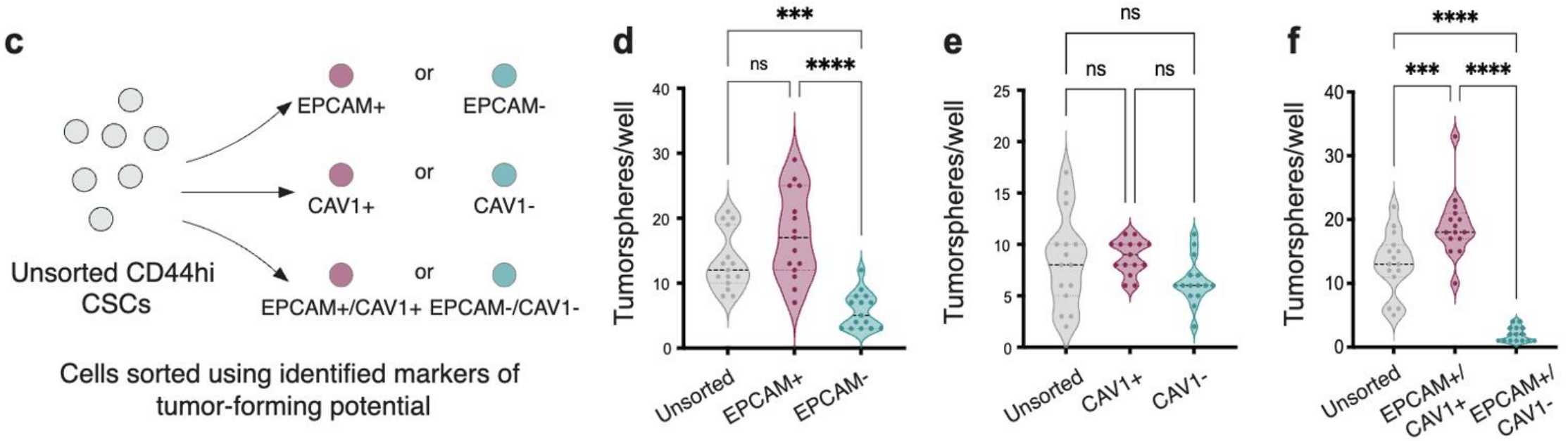


Cell-of-Origin Analysis (Cancer stem cells)

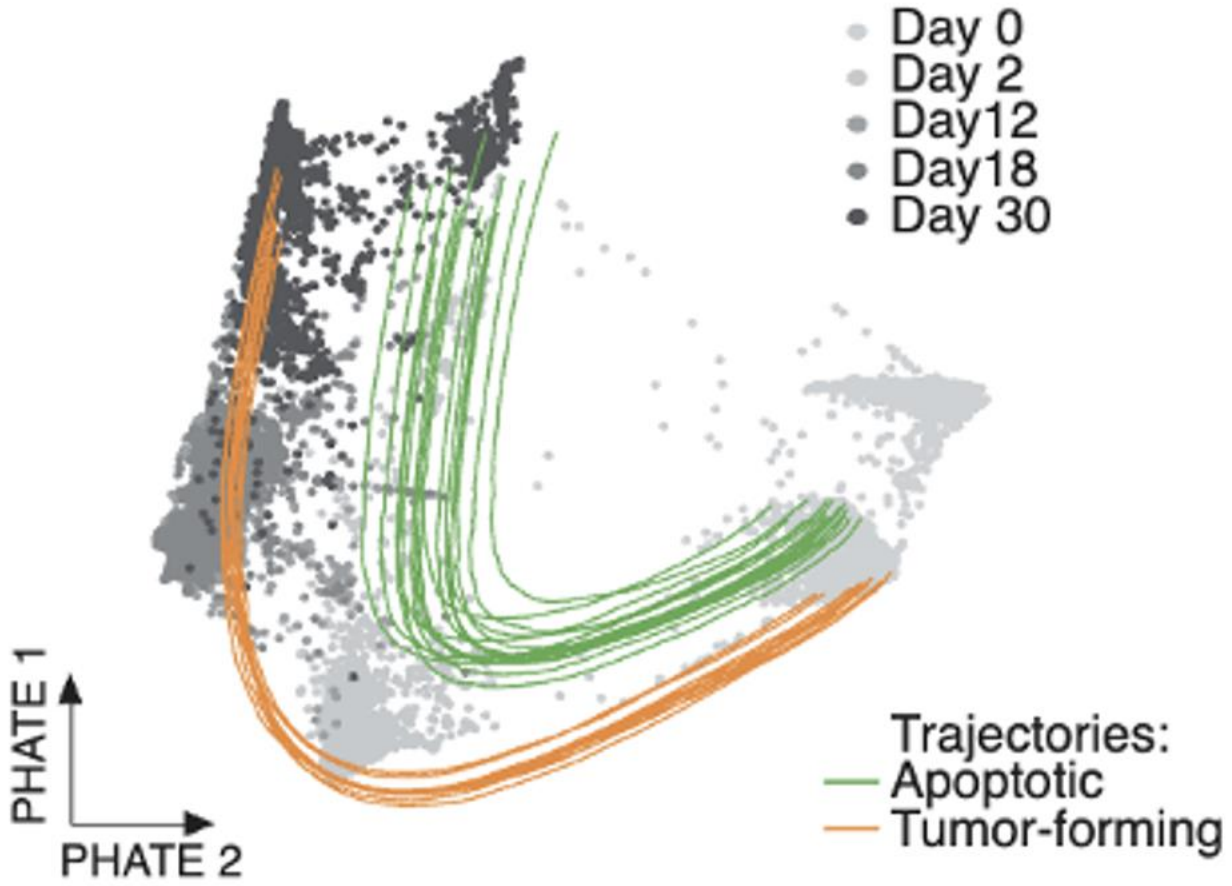


Zooming into Day 0 one can see that the trajectories tend to start in very specific cell states

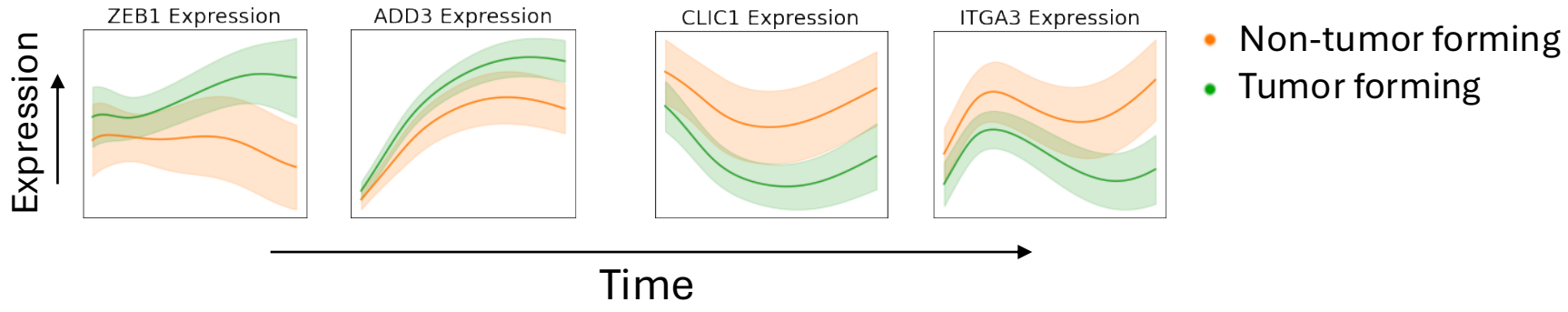
Validation by sorting

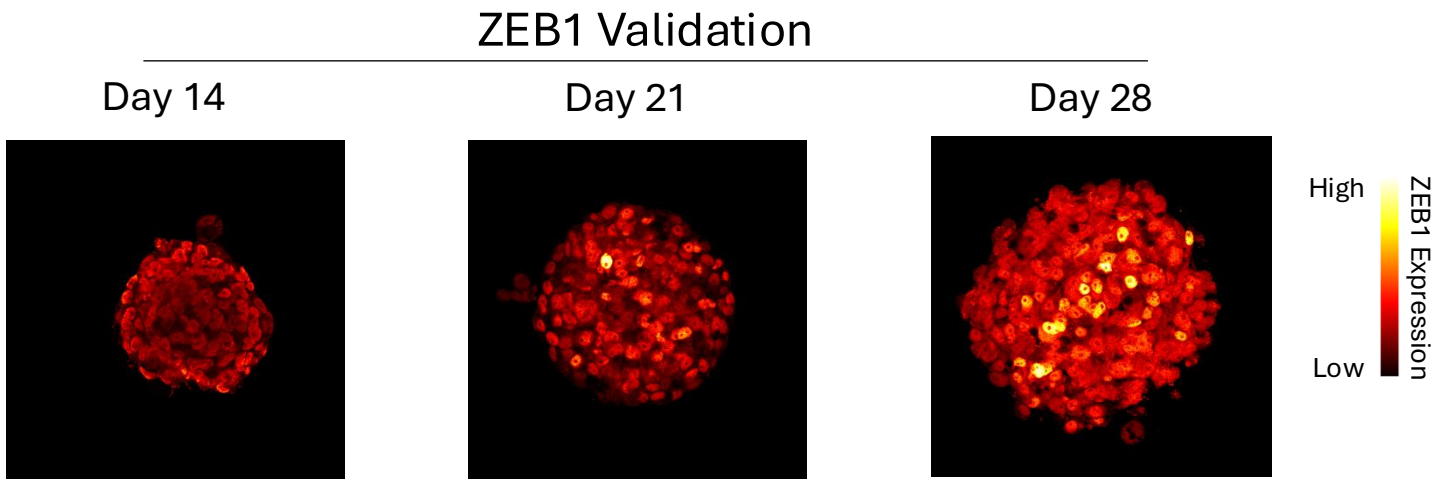


Single-cell trajectories over time

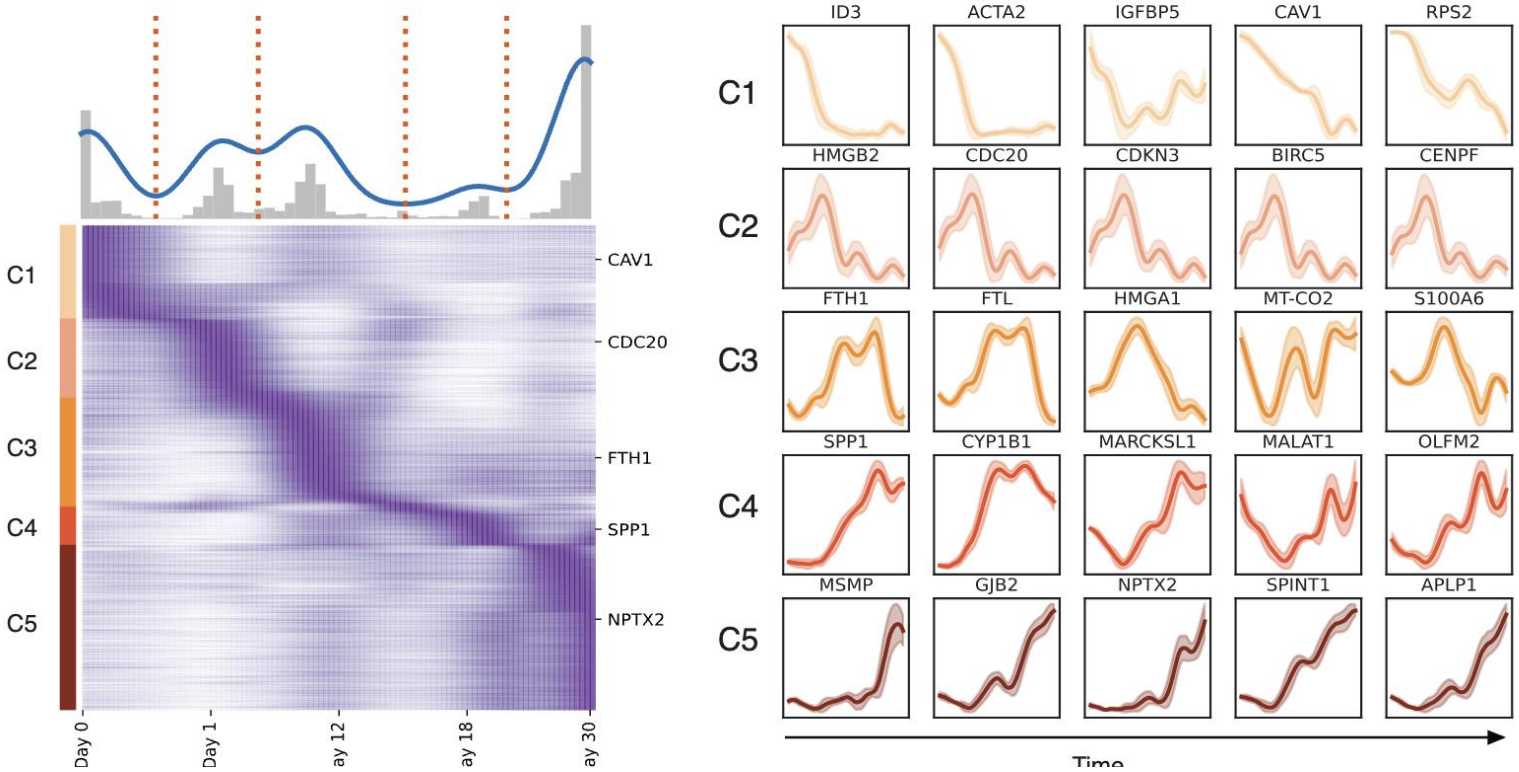


Gene dynamics comparisons





Timed-gene expression modules



Manifold Interpolating Optimal-Transport Flows for Trajectory Inference

Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, Smita Krishnaswamy

Advances in Neural Information Processing Systems 35 (NeurIPS 2022) Main Conference Track

TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics

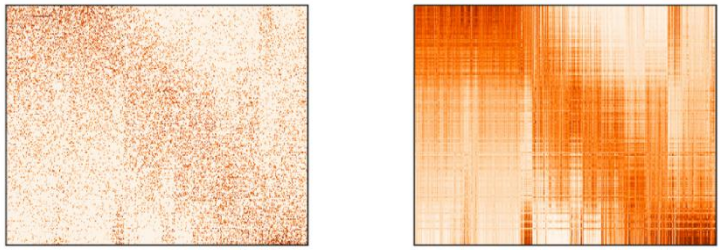
Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, Smita Krishnaswamy *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119:9526-9536, 2020.

1 Revealing dynamic temporal trajectories and underlying
2 regulatory networks with *Cflows*

3 Xingzhi Sun[1*], Shabarni Gupta[2,3*] Alexander Tong[1,4,5*], Manik Kuchroo[6,7*], Dhananjay
4 Bhaskar[1*], Chen Liu[1], Aarthi Venkat[8], Beatriz P. San Juan[2,3], Laura Rangel[2,3], Vanina
5 Rodriguez[2,3], John G. Lock[9], Christine L. Chaffer[2,3§], Smita Krishnaswamy[1,5,10§]

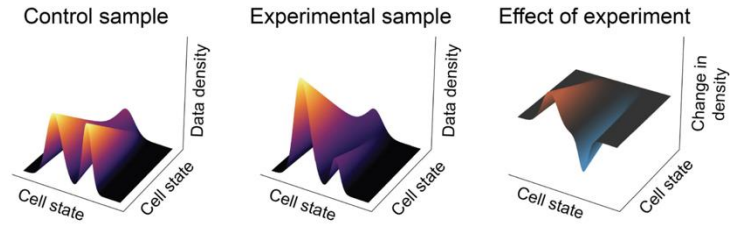
Denoising – MAGIC

van Dijk et al. Cell

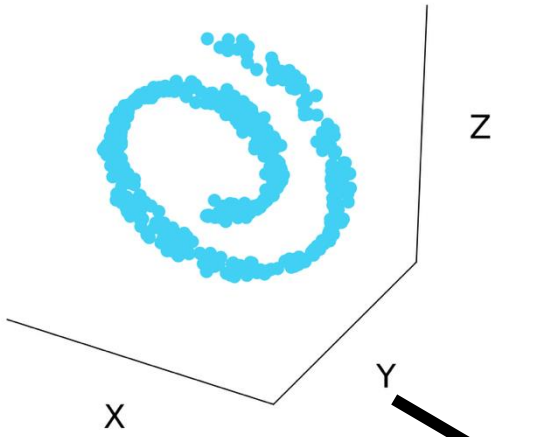


Density estimation--MELD

Burkhardt et al. Nature Biotech



Manifold Learning



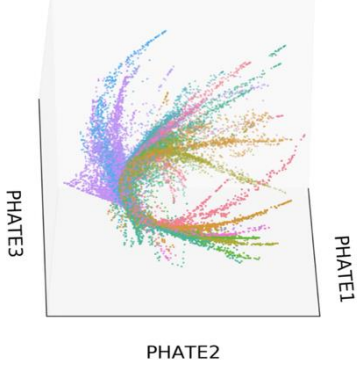
Dimension Reduction – PHATE/ Multiscale PHATE/T-PHATE/ Geo

Moon et al. Nature Biotech

Kuchroo et al. Nature Biotech

Busch et al. Nature Comp Science

Huguet et al. NeurIPS 2023



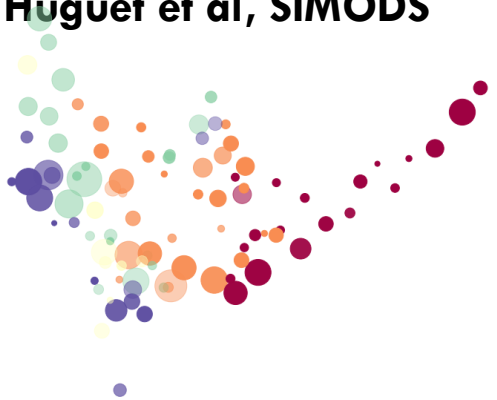
Manifold topology—Diffusion Condensation

Brugnone et al. IEEE Big Data

Moyle et al. Nature

Kuchroo et a. Nature Biotech

Huguet et al, SIMODS



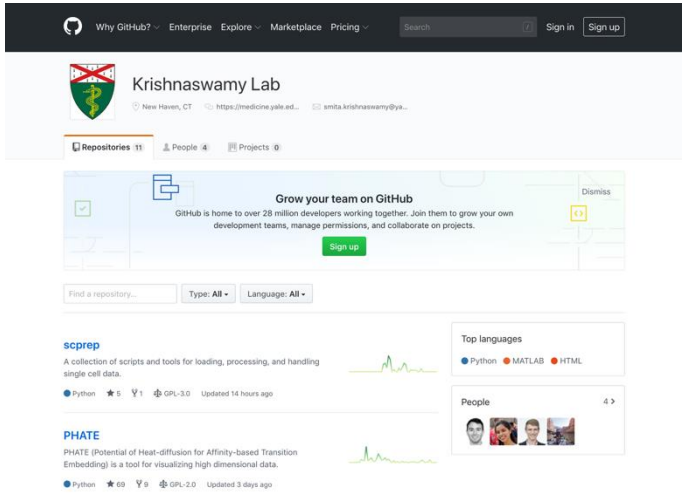
Krishnaswamy Lab Resources

Website

https://www.krishnaswamylab.org

GitHub

github.com/KrishnaswamyLab



Twitter

@krishnaswamylab