Learning Continuous-Time Dynamics of Tissue Remodeling from Spatial Transcriptomic Snapshots with Biology-Aware Flow Matching

Pinar Demetci *
Eric and Wendy Schmidt Center
Broad Institute of MIT and Harvard
Cambridge, MA 02142
demetci@broadinstitute.org

Gabriella M. González†
Brown University
185 Meeting Street, Box G-L215
Providence, RI 02912
gabriella_gonzalez@brown.edu

Abstract

Understanding the molecular mechanisms of tissue regeneration requires reconstructing continuous spatiotemporal dynamics from sparse, irregularly sampled data. We introduce a computational framework that learns these dynamics using a multi-marginal flow matching approach. Our model explicitly accounts for cell proliferation and signaling, and constrains inferred trajectories to the data manifold to ensure biological plausibility. To capture the influence of the microenvironment, we incorporate cell-cell signaling representations. For efficient training, we use semi-balanced Fused Gromov-Wassertein couplings to pick multi-marginal samples for training and define conditional reference paths based on cubic Hermite splines. We then decompose the learned velocity field into spatial and gene expression components, disentangling the factors behind tissue migration from molecular state transitions. Applied to an axolotl brain regeneration and ulcerative colitis datasets, our method successfully reconstructs tissue composition at unseen time points, particularly under irregular sampling. It also recovers cell type-specific signaling influences such as Wnt and TGF-β signaling in neural maturation and interleukin signaling in inflammation. This work provides a scalable and interpretable framework for studying dynamic tissue remodeling from limited spatial transcriptomic observations.

1 Introduction

The development of virtual tissue models capable of predicting cellular responses to interventions represents a critical frontier in AI-driven drug discovery. Recent advances in spatial transcriptomic sequencing have created unprecedented opportunities to capture the molecular and spatial architecture of tissues across time, with temporal datasets now spanning diverse biological processes, such as ulcerative colitis [1], brain aging [2], and axolotl brain regeneration [3]. However, the resource-intensive and destructive nature of these technologies imposes fundamental limitations: it is impossible to experimentally track individual cells over time, and temporal data consist only of cross-sectional snapshots from sparse, irregularly sampled time points. This creates a critical computational challenge to reconstruct continuous-time cellular dynamics from discrete observations while incorporating biological inductive biases essential for tissue-level modeling, such as cell growth and death dynamics, local cell-cell signaling, and microenvironmental factors. Such computational frameworks are particularly crucial for diseases where cell-cell communication drives pathogenesis, including inflammatory

*Corresponding author. Long-term contact information: pinardemetci@gmail.com

†Work partially conducted during research internship. Preferred e-mail address for communication is gabim.gonzalez4@gmail.com

Proceedings of the Latinx in AI Workshop @ NeurIPS-25.

bowel diseases like ulcerative colitis, where disrupted epithelial-immune cell signaling perpetuates chronic inflammation and impairs tissue repair [1]. By accurately modeling these dynamic cellular interactions, virtual tissue models could enable prediction of therapeutic interventions and accelerate the development of targeted treatments that restore healthy tissue function.

Existing computational models that learn tissue trajectories from spatial transcriptomic snapshots can generally be categorized into two groups: (1) methods that simply connect cells from adjacent time points without learning a generative model of the tissue dynamics, such as SpaTrack [4] and Moscot [5], and (2) methods that learn a continuous-time generative model, such as stVCR [6], STORIES [7], and GENOT [8]. The first group cannot simulate intermediate or future states, learns simplistic transitions between largely separated time points, and lacks a notion of dynamics beyond one-step transitions. While generative models address this, they come with limitations: stVCR requires computationally expensive ODE integration in each training step and assumes unrealistic rigid slice alignment, which limits generalization to unseen time points. STORIES avoids ODE integration by learning a neural network–parameterized potential landscape, optimized to minimize debiased Fused Gromov-Wasserstein (FGW) loss between predicted and observed cell distributions. However, training remains slow due to the quadratic cost of FGW computation at each step. GENOT leverages the computationally efficient conditional flow matching framework to learn entropic optimal transport couplings between cell populations, enabling stochastic trajectory modeling with support for arbitrary cost functions, including Fused Gromov-Wasserstein distances. While it has not yet been applied to reconstructing tissue dynamics from time-series spatial transcriptomic data, its framework is extensible to that setting. However, GENOT is currently limited to modeling pairwise distributions and does not support trajectories over multiple, irregularly sampled time points. While there are computational methods proposed to infer single-cell trajectories from multiple, irregularly sampled time points, such as MMFM and MMSFM [9, 10], these lack explicit mechanisms to model spatial transcriptomic data, where slices from different time point snapshots are typically not aligned in the same coordinate system. Lastly, none of these models explicitly incorporate local cell-cell signaling or spatial microenvironment information, which are known factors to that influence cell fate and tissue patterning.

To address this gap, we propose a new computational framework that learn continuous-time dynamics from spatial transcriptomic snapshots that are sparse and irregularly sampled. Our model integrates spatial contextual information on cells like intercellular signaling, and supports scalable inference across multiple time points. In preliminary experiments applied to reconstruct tissue dynamics in mouse intestinal model of ulcerative colitis and axolotl brain regeneration, we show that our approach enables both improved accuracy in trajectory reconstruction and can generate hypothesis on signaling factors that drive these trajectories. This positions our method as a flexible and data-efficient tool for studying a variety of complex tissue remodeling processes.

2 Method

Given spatial transcriptomic snapshots $\mathcal{D}_t = (\mathbf{X}_t, \mathbf{S}_t)$ from time points $t \in \{t_0, t_1, \ldots, t_T\}$, where $\mathbf{X}_t \in \mathbb{R}^{n_t \times g}$ denotes gene expression (i.e. *"transcriptomic"*) measurements for n_t cells and g genes, and $\mathbf{S}_t \in \mathbb{R}^{n_t \times d}$ specifies the corresponding spatial coordinates in d-dimensional space, our proposed method, CHRONOTILE , aims to learn the continuous-time dynamics of cells' transcriptomic states and spatial organization, and to identify their molecular and microenvironmental contributors. This involves several challenges:

- Spatial coordinates, $\mathbf{S}_t$, are typically not aligned across time points in a shared coordinate system due to variations in tissue slice orientation;
- Gene expression data, $\mathbf{X}_t$, are noisy and high-dimensional, with thousands of genes measured per cell;
- Temporal sampling is typically sparse and irregular;
- Tissue architecture evolves due to temporal changes in cells' transcriptomic states, their spatial migration, proliferation and death. These changes are often influenced by microenvironmental factors, such as cell-cell signaling;

Considering these challenges, CHRONOTILE uses a multi-marginal flow matching framework to learn a neural network–parameterized velocity field that models the continuous-time evolution of

tissue architecture in terms of both cellular transcriptomic states and spatial organization. To do so, CHRONOTILE performs the following steps:

1. **Learning lower-dimensional representations of transcriptomic states:** CHRONOTILE first learns lower-dimensional representations $\mathbf{z} \in \mathbb{R}^p$ of transcriptomic profiles using manifold-regularized variational autoencoders (VAEs), which promote smooth flows over the data manifold (Section **A.1**).
2. **Computing cell proliferation–aware couplings:** For each pair of consecutive time points (t_i, t_{i+1}), it computes a probabilistic coupling between cells in snapshots $\mathcal{D}_{t_i}$ and $\mathcal{D}_{t_{i+1}}$ (Section **2.1.1**). This coupling accounts for the VAE-embedded transcriptomic states, intra-snapshot spatial geometry, and estimated cell proliferation rates. These proliferation rates are inferred from expression levels of genes associated with cell growth and death (Section **2.2**).
3. **Pre-aligning spatial transcriptomic slices:** Using the computed couplings, it estimates a global similarity transform to align spatial coordinates across each consecutive snapshot pair, thus establishing a shared coordinate system. The resulting aligned spatial coordinates (Section **A.4**), $\mathbf{s}'$, are used in subsequent steps.
   Steps 1–3 prepare the spatial transcriptomic data for modeling continuous-time dynamics via flow matching.
4. **Incorporating cell-cell signaling:** To account for microenvironmental influences, CHRONOTILE scores pathway-specific signaling input received by each cell from its spatial neighbors (Section **2.3**). The top 20 most variable signaling pathways are selected and transformed into a feature vector $\mathbf{c}_i^{\text{signal}}$, which is concatenated with each cell's latent state and spatial coordinates, and used as inputs in the next step.
5. **Learning continuous-time dynamics via multi-marginal flow matching:** Finally, given these inputs, CHRONOTILE learns a neural velocity field via multi-marginal flow matching (Section **2.1**), capturing the joint evolution of cells' transcriptomic states and spatial positions over continuous time. To generate training trajectories for flow matching, CHRONOTILE fits monotonic cubic Hermite splines between cells linked across multiple time points by the couplings computed in Step 2. By analyzing the partial derivatives of the learned velocity field with respect to specific inputs (e.g., gene expression or signaling features), we identify molecular and microenvironmental drivers of tissue reorganization in a cell-type–specific manner.

We host the source code on `https://github.com/pinardemetci/CHRONOTILE`.

2.1 Learning continuous-time spatiotranscriptomic dynamics with biology-aware multi-marginal flow matching

We represent each cell at time t with a joint spatiotranscriptomic state vector $\mathbf{y} = [\mathbf{z}, \mathbf{s}'] \in \mathbb{R}^{(p+d)}$, where $\mathbf{z} \in \mathbb{R}^p$ is a lower-dimensional embedding of the transcriptomic profile generated via the manifold-regularized variational auto-encoder (described in Section **A.1**) that encourages flows to progress along the data manifold, and $\mathbf{s}' \in \mathbb{R}^d$ denotes the spatial coordinates, "corrected" after we align snapshots from different time points into a shared coordinate system (described in Section **A.4**).

We model the evolution of each cell's state via a neural ODE that parameterizes the velocity field underlying cell state changes with a neural network v_θ:

$$\frac{d\mathbf{y}}{dt} = v_\theta(\mathbf{y}, \phi(t), \mathbf{c}), \tag{1}$$

where $\phi(t) \in \mathbb{R}^4$ denotes 4-dimensional sinusoidal time embeddings [11] that provide rich temporal context to the network and $\mathbf{c} \in \mathbb{R}^{20}$ provides additional signaling information received by cells from their local spatial neighborhood in the form of signaling pathway scores on the top 20 most variable pathways (Section **2.3**). The velocity field $v_\theta : \mathbb{R}^{(p+d)+4+20} \times \mathbb{R} \to \mathbb{R}^{p+d}$ is a neural ordinary differential equation (ODE) that predicts the instantaneous change in the joint spatio-transcriptomic state, $\mathbf{y}$, at time t and is constructed as a feed-forward network with four hidden layers and smooth activations (GeLU default).

Traditionally, such neural ODEs have been trained by repeatedly solving forwards ODE integrations at every gradient step, which is computationally expensive and can become numerically unstable

when the dynamics are stiff. The conditional flow matching (CFM) framework [12] has introduced a simpler and more efficient training scheme that side-steps simulations, where the velocity field is regressed onto an analytic reference field $u_t(\mathbf{z}_t \mid \mathbf{z}_0, \mathbf{z}_1)$ conditioned on sampled pairs $(\mathbf{z}_0, \mathbf{z}_1)$ representing endpoints of predefined reference paths between two snapshots:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\, \mathbb{E}_{(\mathbf{z}_0,\mathbf{z}_1)\sim\pi}\, \mathbb{E}_{\mathbf{z}_t\sim p_t(\,\cdot\,\mid \mathbf{z}_0,\mathbf{z}_1)} \big\| v_\theta(\mathbf{z}_t, t) - u_t\big(\mathbf{z}_t \mid \mathbf{z}_0, \mathbf{z}_1\big)\big\|_2^2 \tag{2}$$

Here, π is a joint distribution over consecutive time points that can represent either independent sampling from marginals [12] or structured couplings [13]. The CFM objective in Equation 2 is limited to modeling flows between two time points. To handle multiple time points, we adopt the multi-marginal flow matching (MMFM) loss [9]:

$$\mathcal{L}_{\text{MMFM}}(\theta) = \mathbb{E}_{\boldsymbol{\xi}\sim q(\boldsymbol{\xi})}\, \mathbb{E}_{t\sim\mathcal{U}[0,1],\, \mathbf{z}\sim p_t(\mathbf{z}\mid\boldsymbol{\xi})} \left[\|v_\theta(\mathbf{z}, t) - u_t(\mathbf{z} \mid \boldsymbol{\xi})\|_2^2\right] \tag{3}$$

where $\boldsymbol{\xi} = (\mathbf{z}_0, \mathbf{z}_1, \ldots, \mathbf{z}_K)$ is a multi-temporal trajectory sampled across all $K+1$ time points. Unlike [9], we construct the coupling distribution $q(\boldsymbol{\xi})$ using semi-balanced entropic Fused Gromov-Wasserstein (FGW) optimal transport [14, 15, 16] (Section **2.1.1**). The choice of FGW allows for finding probabilistic couplings between cells from consecutive time points, while accounting for similarities in both their transcriptomic cell states and their spatial context. Additionally, the semi-balanced formulation allows us to incorporate estimates of net cell proliferation, which lets cells with higher estimated proliferation rates to be matched with more cells in the next consecutive time point. To obtain the analytic reference field $u_t(\cdot \mid \boldsymbol{\xi})$, we draw anchor points from the training data, $\boldsymbol{\xi} = (\mathbf{z}_0, \ldots, \mathbf{z}_K)$, sampled based on these couplings, and interpolate between them using smooth splines (Section **2.1.2**), which accommodate irregularly spaced time points and provides closed-form derivatives for quick evaluation and back-propagation during training. We specifically choose monotonic cubic Hermite splines for analytical properties described in Section **2.1.2**.

2.1.1 Picking training samples for reference paths

For a given pair of consecutive time points, e.g. t_1 and t_2, we compute the semi-balanced FGW coupling between cells as:

$$\mathbf{P}^*(t_1, t_2) = \arg\min_{\mathbf{P}\geq 0} \alpha \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \|\mathbf{z}_i - \mathbf{z}_j\|^2 P_{ij} + (1-\alpha) \sum_{i,k=1}^{n_1}\sum_{j,l=1}^{n_2} \|D^s_{ik} - D^s_{jl}\|^2 P_{ij} P_{kl} + \epsilon\, \text{KL}(\mathbf{P}||\mathbf{a}\otimes\mathbf{b}) + \lambda \text{KL}(\mathbf{P}\mathbf{1}||\mathbf{a}) \tag{4}$$

where $D^s_{ik} = \|s_i - s_k\|^2$ gives the squared Euclidean distance between the spatial coordinates of cells i and k from time point t_1. Essentially, this objective considers both the transcriptomic similarity (via embeddings $\mathbf{z}$) and spatial structural alignment (via intra-time-point spatial distances, $\mathbf{D}^s$), balanced by the hyperparameter $\alpha \in [0, 1]$. Crucially, since $\mathbf{D}^s$ holds relative pairwise distances within a time point, it considers the tissue's geometric structure without requiring the slices to be aligned across time points. Additionally, since the embeddings $\mathbf{z}$ are computed with a manifold-aware VAE, the term $\|\mathbf{z}_i - \mathbf{z}_j\|^2$ approximates transcriptomic manifold distances. The vectors $\mathbf{a}$ and $\mathbf{b}$ are the prescribed mass distributions over cells from time points t_1 and t_2, respectively, where $\mathbf{a}$ initialized based on gene-expression-derived proliferation scores (Section 2.2) and assume $\mathbf{b}$ is uniform. The total outgoing coupling probability, a_i, for cell i is high when it has a high profileration rate estimate. The semi-balanced regularization term, $\lambda\text{KL}(\mathbf{P}\mathbf{1}||\mathbf{a})$, allows deviation from imperfect growth priors and the entropic regularization term, $\epsilon\,\text{KL}(\mathbf{P}||\mathbf{a}\otimes\mathbf{b})$, enables efficient computation and allows coupling probabilities to spread across multiple cells.

Based on these couplings, we pick training samples to define reference paths. Following the "rolling window" strategy in [10], we start by sampling a pair of aligned data points from the initial time points and then extend it to a triplet. To do this, we first sample a source index i from the outgoing mass distribution $a_i = \sum_j \mathbf{P}^*_{ij}(t_1, t_2)$ and then sample its destination, j, at the next time point from the row-normalized conditional

$$(\mathbf{y}_i^{t_1}, \mathbf{y}_j^{t_2}) \sim \pi_{12}, \qquad \pi_{12}(j \mid i) = \frac{P^*_{ij}(t_1, t_2)}{\sum_{j'} P^*_{ij'}}.$$

Given this middle cell $y_j^{t_2}$, we sample its successor from the next coupling by

$$\mathbf{y}_k^{t_3} \sim \pi_{23}(\,\cdot \mid j), \qquad \pi_{23}(k \mid j) = \frac{P^*_{jk}(t_2, t_3)}{\sum_{k'} P^*_{jk'}}.$$

Giving us sampled triplets $(\mathbf{y}_i^{t_1}, \mathbf{y}_j^{t_2}, \mathbf{y}_k^{t_3})$ for training.

2.1.2 Defining reference paths

These sampled triplets $(\mathbf{y}_i^{t_1}, \mathbf{y}_j^{t_2}, \mathbf{y}_k^{t_3})$ serve as control points (or "anchor points") for defining smooth and continuous *local reference paths*. By fitting splines over rolling windows of three consecutive time points, we capture intermediate dynamics more accurately than pairwise interpolation and obtain trajectories that remain well-behaved even when sampling intervals are uneven. This multi-point construction enforces temporal smoothness while accommodating irregular sampling, resulting in more stable and biologically plausible reference paths [10].

For each triplet, we concatenate two monotonic cubic Hermite segments, $\mu_{12}(t)$ for the time interval $[t_1, t_2]$ and $\mu_{23}(t)$ for the time interval $[t_2, t_3]$, that meet at the middle point $\mathbf{y}_j^{t_2}$.

For the segment $\mu_{12}(t)$, with the rescaled time $\tau_{12} = (t - t_1)/(t_2 - t_1) \in [0, 1]$, we define:

$$\mu_{12}(t) = h_{00}(\tau_{12})\,\mathbf{y}_i^{t_1} + h_{10}(\tau_{12})\,m_{t_1} + h_{01}(\tau_{12})\,\mathbf{y}_j^{t_2} + h_{11}(\tau_{12})\,m_{t_2},$$

and similarly, for the segment $\mu_{23}(t)$, using $\tau_{23} = (t - t_2)/(t_3 - t_2)$, we define:

$$\mu_{23}(t) = h_{00}(\tau_{23})\,\mathbf{y}_j^{t_2} + h_{10}(\tau_{23})\,m_{t_2} + h_{01}(\tau_{23})\,\mathbf{y}_k^{t_3} + h_{11}(\tau_{23})\,m_{t_3}, \qquad m_{t_3} = s_+.$$

where the Hermite basis polynomials are $h_{00} = 2\tau^3 - 3\tau^2 + 1,\ h_{10} = \tau^3 - 2\tau^2 + \tau,\ h_{01} = -2\tau^3 + 3\tau^2,\ h_{11} = \tau^3 - \tau^2$.

The slopes m_{t_1}, m_{t_2}, and m_{t_3} indicate the starting, middle, and end point slopes, respectively. We define the starting and end point slopes as "secant slopes", i.e. straight-line rate of change between the two data points they interpolate: $\mathbf{y}_i^{t_1}$ and $\mathbf{y}_j^{t_2}$ for m_{t_1} and $\mathbf{y}_j^{t_2}$ and $\mathbf{y}_k^{t_3}$ for m_{t_3}:

$$m_{t_1} = \frac{\mathbf{y}_j^{t_2} - \mathbf{y}_i^{t_1}}{t_2 - t_1}, \qquad m_{t_3} = \frac{\mathbf{y}_k^{t_3} - \mathbf{y}_j^{t_2 1}}{t_3 - t_2}$$

The middle point slope, m_{t_2}, is chosen according to the Fritsch-Carlson monotonicity rule so the cubic never overshoots its two end-points and every interpolated state remains inside the convex hull of its neighboring observations. To achieve this, if the incoming and outgoing secant slopes m_{t_1} and m_{t_3} disagree in sign, we force $m_{t_2} = 0$; otherwise we clip the slope to the harmonic mean:

$$m_{t_2} = \begin{cases} 0, & m_{t_1} - m_{t_3} \le 0, \\ \dfrac{2m_{t_1} - m_{t_3}}{m_{t_1} + m_{t_3}}, & \text{otherwise}, \end{cases}$$

With these two segments, $\mu_{12}(t)$ and $\mu_{23}(t)$, we set

$$\mu(t) = \begin{cases} \mu_{12}(t), & t \in [t_1, t_2], \\ \mu_{23}(t), & t \in [t_2, t_3], \end{cases}$$

Taking the time-derivative of the composite path gives the *reference vector field*, conditioned on the control-point context $\mathcal{C} = \big(\mathbf{y}_i^{t_1}, \mathbf{y}_j^{t_2}, \mathbf{y}_k^{t_3}\big)$:

$$u\big(\mathbf{y}, t \,\big|\, \mathcal{C}\big) = \frac{d\mu_{\mathcal{C}}(t)}{dt} = \begin{cases} \dfrac{1}{t_2 - t_1}\ \dfrac{d\mu_{12}}{d\tau_{12}}\big(\tau_{12}\big), & t \in [t_1, t_2], \\ \dfrac{1}{t_3 - t_2}\ \dfrac{d\mu_{23}}{d\tau_{23}}\big(\tau_{23}\big), & t \in [t_2, t_3], \end{cases}$$

with $\tau_{12} = (t - t_1)/(t_2 - t_1)$ and $\tau_{23} = (t - t_2)/(t_3 - t_2)$.

The multi-marginal conditional flow matching training loss is then:

$$\mathcal{L} = \ \mathbb{E}_{w\sim\mathcal{W}}\ \mathbb{E}_{\mathcal{C}_w}\ \mathbb{E}_{t\sim\mathrm{Unif}(t_w, t_{w+2})} \Big\| v_\theta\big(\mathbf{y}, t, \mathbf{c}\big) - u\big(\mathbf{y}, t \,\big|\, \mathcal{C}_w\big) \Big\|^2.$$

where $\mathcal{W} = \{\,1, \ldots, K-2\}$ indexes the $K-2$ overlapping time windows (t_w, t_{w+1}, t_{w+2}).

Because we define local Hermite paths over an overlapping rolling window of three time points and each local path shares the middle slope, the reference paths we define and their first derivative with

respect to time (i.e. the conditional reference vector field) are continuous, so cells do not "teleport" or suffer an instantaneous jump in velocity. Unlike natural cubic splines used in [9], samples from interpolating time points are guaranteed to remain in the convex hull of the observed data from consecutive time points and the second derivative (i.e. acceleration) is allowed to jump at t_2, which is more suitable for biological systems with transcriptional bursts. Because each Hermite segment depends on at most three control points, evaluating $\mu(t)$ and its derivative $\dot{\mu}(t)$ is $\mathcal{O}(1)$ and adds negligible overhead.

2.2 Computing initial cell proliferation rate estimates

We initialize cell growth rate estimates based on the net proliferation rates for each cell, predicted from the observed expression levels of genes that are annotated in the MSigDB [**?**] Hallmark gene sets on apoptosis, G2M checkpoint and E2F targets. For a cell k from time point t_i, we use scanpy's `scanpy.tl.score_genes` function to generate raw "apoptosis" and "proliferation" scores, $S_k^{\text{prolif.}}$ and $S_k^{\text{apop.}}$, that we then turn into a single "net proliferation rate" estimate by first computing an unnormalized score:

$$g'_k \;=\; \zeta_k^{\text{prolif.}} - \zeta_k^{\text{apop.}},$$

where $\zeta_k^{\text{prolif.}}$ and $\zeta_k^{\text{apop.}}$ are the z-scores of the raw scores, computed to bring the two onto a comparable scale:

$$\zeta_k^{\text{prolif}} \;=\; \frac{S_k^{\text{prolif}} - \mu_{\text{prolif}}}{\sigma_{\text{prolif}}}, \qquad \zeta_k^{\text{apop}} \;=\; \frac{S_k^{\text{apop}} - \mu_{\text{apop}}}{\sigma_{\text{apop}}},$$

where $\mu_\bullet$ and $\sigma_\bullet$ are the mean and (population) standard deviation of the corresponding signature across all cells. When the apoptosis scores dominate the proliferation scores, the g'_k term can take negative values. Since we use net growth scores to represent the level of density change, we normalize g'_k to obtain non-negative scores:

$$g_k \;=\; \begin{cases} g'_k, & g'_k \geq 0, \\ \dfrac{1}{-\,g'_k}, & g'_k < 0, \end{cases} \qquad g_k \in \left[\, 1/(-g'_{\min}),\; g'_{\max} \right].$$

When computing the semi-balanced OT coupling between cells from consecutive time points t_i and t_{i+1}, we first initialize the source marginal $\mathbf{a}$ defined over cells from time point t_i with these growth term estimates. Then, we multiply $\mathbf{a}$ by a constant factor so that the ratio of its total mass to that of the (unit-mass) target marginal $\mathbf{b}$ defined over t_{i+1} matches the ratio of the number of cells observed in snapshots t_i and t_{i+1}:

$$\frac{\sum_k a_k}{\sum_l b_l} \;=\; \frac{n_{t_i}}{n_{t_{i+1}}} \quad\Longrightarrow\quad \sum_k a_k = \frac{n_{t_i}}{n_{t_{i+1}}},$$

ensuring that the transported mass reflects the different numbers of cells present in the two consecutive spatial transcriptomic snapshots.

2.3 Incorporating cell-cell signaling

For every cell i, we summarize the quantity and composition of signaling input received from its local neighbors. Since absolute spatial distances are not always available or standardized across datasets, we construct a k-nearest neighbor (k-NN) graph with a default $k = 50$, connecting each cell to its 50 nearest neighbors based on aligned spatial coordinates.

To estimate pathway-specific communication along edges, we adapt the ligand–receptor framework of *CellChat* [17], which couples a curated ligand–receptor database ($\sim$ 2,000 pairs grouped into pathways $\mathcal{P}$) with a mass–action model. For every directed edge $(i \to j)$ and pathway P, we compute

$$w_{ij}^{(P)} \;=\; \frac{1}{|P|} \sum_{(L,R)\in P} \frac{\langle x_{i,L}\rangle \, \langle x_{j,R}\rangle}{\langle x_{i,L}\rangle + \langle x_{j,R}\rangle},$$

where $\langle x_{k,L}\rangle$ denotes the average CPM-normalized expression of ligand L in sender cell k, and similarly for receptor R in the receiver. Note that CPM stands for "Counts Per Million", a standard

normalization method that scales gene expression so each cell's total counts sum to one million in order to correct for differences in sequencing depth across cells. A permutation test over shuffled cell labels converts these raw scores into probabilities of communication; we retain the expected value as the edge weight. For each pathway, the total incoming signal to cell i is aggregated as

$$\iota_{i,P} \;=\; \sum_{(j,i)\in E} w_{ji}^{(P)}, \qquad P \in \mathcal{P}.$$

We retain the $P^\star = 20$ pathways with the highest variance across cells, apply a $\log(1+\cdot)$ transform followed by z-scoring, and obtain the signaling feature vector:

$$\mathbf{c}_i^{\text{signal}} \;=\; \text{zscore}(\log(\iota_{i,P^\star}+1)) \;\in\; \mathbb{R}^{P^\star}.$$

The resulting $\mathbf{c}_i^{\text{signal}}$ encodes the relative strength of the most variable signaling pathways impinging on cell i, and is concatenated with the latent state and spatial coordinates when evaluating the velocity field v_θ.

2.4 Investigating molecular drivers

For a cell i observed at time t we construct the cell state input $\mathbf{y}_i = \left[\, \mathbf{z}_i, \mathbf{s}_i' \,\right]$, where $\mathbf{z}_i$ is the latent embeddings of the transcriptomic state and $\mathbf{s}_i'$ is the spatial coordinates and compute $\mathbf{c}_i$, which encodes a cell's incoming signaling descriptors. We then compute its velocity $\mathbf{v}_i \;=\; v_\theta\big(\mathbf{y}_i, \phi(t), \mathbf{c}_i\big) \;\in \mathbb{R}^{g+d}$. In order the separate the cell state transitions from spatial migration, we decompose the velocity into two components: $\mathbf{v}_i^z \in \mathbb{R}^g$ and $\mathbf{v}_i^s \in \mathbb{R}^d$ by separating the transcriptomically relevant dimensions and the spatially relevant dimensions (i.e. last 2 dimensions in our applications). To rank each of the input components by their contribution to either the cell state transitions $\mathbf{v}_i^z$ or spatial migration $\mathbf{v}_i^s$, we compute the partial derivative of the velocity field with respect to the input (i.e. Jacobian columns):

$$\mathbf{J}_{i,r} \;=\; \frac{\partial\, v_\theta(\mathbf{y}_i, t, \mathbf{c}_i)}{\partial x_r}\bigg|_{\mathbf{x}=\mathbf{x}_i} \in \mathbb{R}^p.$$

which shows how much the instantaneous velocity of cell i would change with infinitesimal increase in the value of the input x_r, Because gene counts influence the field only through the embedding $\mathbf{z} = f_{enc}(\mathbf{x})$, the instantaneous effect of gene g is obtained by composing derivatives:

$$\underbrace{\mathbf{J}_{i,g}^{\text{gene}}}_{\in\mathbb{R}^{p+d}} = \mathbf{J}_i^{(v)}\; \frac{\partial f_{enc}(\mathbf{x})}{\partial x_g}\bigg|_{\mathbf{x}=\mathbf{x}_i},$$

where $\mathbf{J}_i^{(v)} = \partial v_\theta / \partial \mathbf{z}$ is the block of $\mathbf{J}_{i,r}$ corresponding to $\mathbf{z}$. We then score feature r (or gene g) by the directional derivative and compute the dot product with the cell's velocity, which helps us determine whether the input further pushes the cell in the same direction of transition and gives us a notion of a driver score to rank input factors:

$$q_{i,r} \;=\; \mathbf{v}_i^\top\, \mathbf{J}_{i,r}, \qquad q_{i,g}^{\text{gene}} \;=\; \mathbf{v}_i^\top\, \mathbf{J}_{i,g}^{\text{gene}}. \tag{1}$$

Then, a population-level importance score is obtained by averaging over a cell set $\mathcal{S}$:

$$\bar{q}_r \;=\; \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} q_{i,r}.$$

For a group of related features G (e.g. all components of a pathway vector) we summarise with the L_2 norm $S_{i,G} = \big\|\{s_{i,r} : r \in G\}\big\|_2$.

3 Results

We apply our method to reconstruct the spatio-temporal dynamics of two tissues. The first is the axolotl brain [3], which provides 2-dimensional spatial transcriptomic snapshots across the course of tissue repair. In this study, researchers induced tissue regeneration by surgically resecting a

portion of the telencephalon area of the brain and then collected 2D spatial transcriptomic slices from different animals at seven stages: 2, 5, 10, 15, 20, 30, and 60 days post injury (dpi). The second dataset is a mouse model of ulcerative colitis [1], which also contains 2D spatial transcriptomic snapshots from mouse colon during dextran-sodium-sulfate (DSS)-induced colitis. It captures spatial transcriptomic measurements during inflammation and recovery at four time points: baseline (day 0), early inflammation (day 3), peak disease (day 9), and post-recovery (day 21). Both datasets contain replicates from multiple individual animals per time point. While these could be used in incorporating sample-sample variability in future work, we pick one snapshot per time point in the current work to start with a simpler scenario.

We first train our model in a leave-one-out setting and evaluate its ability to reconstruct the held-out time points as Figure 1 visualizes. We compare reconstruction performance against two baselines: STORIES [7] and GENOT [8], where GENOT. In Table 1, we demonstrate more accurate reconstruction, especially in later time points with larger temporal gaps between snapshots. We then investigate the signaling pathways contributing to cell state transitions in specific cell types at different time points in the axolotl brain regeneration dataset. Figure 2 visualizes the ranking of signaling pathways, which shows time-varying cell-type-specific signaling logic during regeneration [3]. In neuroblasts, canonical developmental pathways such as Wnt, Notch, and FGF dominate neurogenesis in early stages post-injury, consistent with their known roles in promoting proliferation and lineage commitment during neurogenesis [18, 19]. By Day 7, a shift toward TGF-β and VEGF signaling suggests progression into differentiation and integration phases, aligning with literature on tissue remodeling and neurovascular coupling [20, 21]. On the other hand, microglia show a transition from pro-inflammatory signaling (e.g., interleukins, TNF) at Day 3 toward reparative and regulatory signals (e.g., TGF-β, VEGF) by Day 7, in line with established temporal dynamics of immune cell polarization in central nervous system repair [22]. Together, these findings support the biological fidelity of our approach and highlight its utility in revealing coordinated spatio-temporal programs of regeneration.

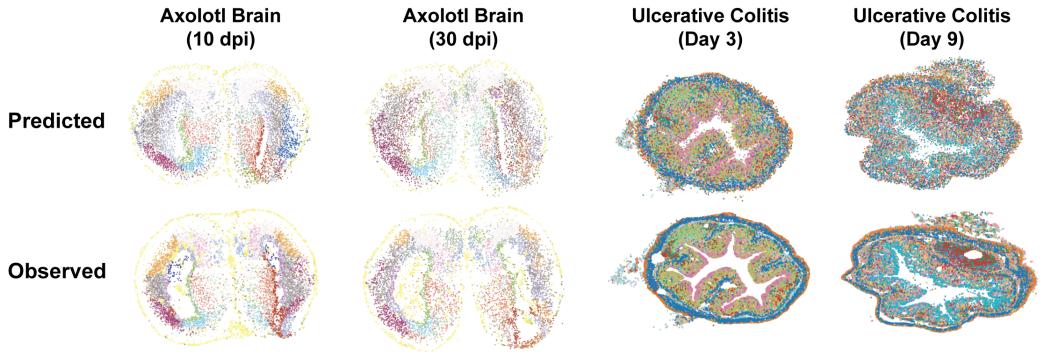

Figure 1: Visualization of reconstructed slices

Table 1: Benchmarking reconstruction error (measured by Wasserstein distance W_1).

	Axolotl brain (10 dpi)	**Axolotl brain (30 dpi)**	**Ulcerative colitis (Day 3)**	**Ulcerative colitis (Day 9)**
CHRONOTILE (ours)	**0.021**	**0.034**	0.044	**0.076**
STORIES	0.039	0.057	**0.041**	0.101
GENOT	0.092	0.142	0.160	0.155

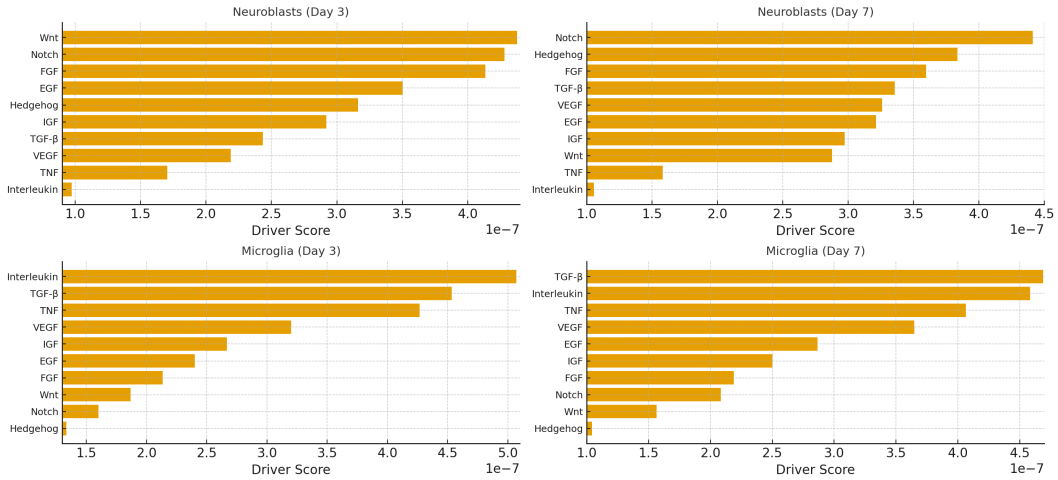

Figure 2: Ranking of signaling pathways most influential in transcriptomic cell state transitions in axolotl brain regeneration dataset.

4 Discussion

Our results demonstrate that biology-aware multi-marginal flow matching can reconstruct continuous spatio-transcriptomic trajectories from sparse, irregular snapshots while explicitly accounting for cell–cell signaling and net population change. By coupling semi-balanced FGW plans with Hermite-spline reference paths, the learned vector field avoids slice-alignment assumptions and prevents “teleportation". Our preliminary results show this framework enables us to investigate the signaling pathways responsible for cell state transitions. This work-in-progress reflects a promising direction but requires further investigation of molecular drivers across various tissue backgrounds. Additionally, the current framework does not consider signaling effects from 3-dimensional neighborhood, cell niche effects beyond signaling or individual variability from multiple samples within the same time point, which is left for future work.

Acknowledgments and Disclosure of Funding

P.D.’s contributions were partly supported by a fellowship from the Eric and Wendy Schmidt Center at the Broad Institute. G.G.’s contributions, largely involving running baselines, were carried out during an independent research internship between August - October 2025.

References

[1] Paolo Cadinu, Kisha N Sivanathan, Aditya Misra, Rosalind J Xu, Davide Mangani, Evan Yang, Joseph M Rone, Katherine Tooley, Yoon-Chul Kye, Lloyd Bod, Ludwig Geistlinger, Tyrone Lee, Randall T Mertens, Noriaki Ono, Gang Wang, Liliana Sanmarco, Francisco J Quintana, Ana C Anderson, Vijay K Kuchroo, Jeffrey R Moffitt, and Roni Nowarski. Charting the cellular biogeography in colitis reveals fibroblast trajectories and coordinated spatial remodeling. *Cell*, 187(8):2010–2028.e30, April 2024.

[2] Eric D. Sun, Olivia Y. Zhou, Max Hauptschein, Nimrod Rappoport, Lucy Xu, Paloma Navarro Negredo, Ling Liu, Thomas A. Rando, James Zou, and Anne Brunet. Spatial transcriptomic clocks reveal cell proximity effects in brain ageing. *Nature*, 638(8049):160–171, Feb 2025.

[3] Xiaoyu Wei, Sulei Fu, Hanbo Li, Yang Liu, Shuai Wang, Weimin Feng, Yunzhi Yang, Xiawei Liu, Yan-Yun Zeng, Mengnan Cheng, Yiwei Lai, Xiaojie Qiu, Liang Wu, Nannan Zhang, Yujia Jiang, Jiangshan Xu, Xiaoshan Su, Cheng Peng, Lei Han, Wilson Pak-Kin Lou, Chuanyu Liu, Yue Yuan, Kailong Ma, Tao Yang, Xiangyu Pan, Shang Gao, Ao Chen, Miguel A. Esteban, Huanming Yang, Jian Wang, Guangyi Fan, Longqi Liu, Liang Chen, Xun Xu, Ji-Feng Fei, and Ying Gu. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022.

[4] Xunan Shen, Lulu Zuo, Zhongfei Ye, Zhongyang Yuan, Ke Huang, Zeyu Li, Qichao Yu, Xuanxuan Zou, Xiaoyu Wei, Ping Xu, Yaqi Deng, Xin Jin, Xun Xu, Liang Wu, Hongmei Zhu, and Pengfei Qin. Inferring cell trajectories of spatial transcriptomics via optimal transport analysis. *Cell Systems*, 16(2), Feb 2025.

[5] Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Lama Saber, Changying Jing, et al. Mapping cells through time and space with moscot. *Nature*, 638(8052):1065–1075, 2025.

[6] Qiangwei Peng, Peijie Zhou, and Tiejun Li. stvcr: Spatiotemporal dynamics of single cells. *bioRxiv*, 2025.

[7] Geert-Jan Huizing, Jules Samaran, Daniele Capocefalo, Anna Audit, Gabriel Peyré, and Laura Cantini. Stories: learning cell fate landscapes from spatial transcriptomics. *bioRxiv*, 2025.

[8] Dominik Klein, Théo Uscidda, Fabian Theis, and Marco Cuturi. Genot: Entropic (gromov) wasserstein flow matching with applications to single-cell genomics. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 103897–103944. Curran Associates, Inc., 2024.

[9] Martin Rohbeck, Charlotte Bunne, Edward De Brouwer, Jan-Christian Huetter, Anne Biton, Kelvin Y. Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. In *The Thirteenth International Conference on Learning Representations*, 2025.

[10] Justin Lee, Behnaz Moradijamei, and Heman Shakeri. Multi-marginal stochastic flow matching for high-dimensional snapshot data at irregular time points. *arXiv preprint arXiv:2508.04351*, 2025.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, volume 30, 2017.

[12] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[13] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023.

[14] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.

[15] Alexis Thual, Quang Huy Tran, Tatiana Zemskova, Nicolas Courty, Rémi Flamary, Stanislas Dehaene, and Bertrand Thirion. Aligning individual brains with fused unbalanced gromov wasserstein. *Advances in neural information processing systems*, 35:21792–21804, 2022.

[16] Cédric Vincent-Cuaz, Rémi Flamary, Marco Corneli, Titouan Vayer, and Nicolas Courty. Semi-relaxed gromov-wasserstein divergence with applications on graphs. *arXiv preprint arXiv:2110.02753*, 2021.

[17] Suoqin Jin, Christian F Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V Plikus, and Qing Nie. Inference and analysis of cell-cell communication using cellchat. *Nature communications*, 12(1):1088, 2021.

[18] Sebastian B Arredondo, Daniela Valenzuela-Bezanilla, Muriel D Mardones, and Lorena Varela-Nallar. Role of wnt signaling in adult hippocampal neurogenesis in health and disease. *Front. Cell Dev. Biol.*, 8:860, September 2020.

[19] Itaru Imayoshi, Masayuki Sakamoto, Masahiro Yamaguchi, Kensaku Mori, and Ryoichiro Kageyama. Essential roles of notch signaling in maintenance of neural stem cells in developing and adult brains. *Journal of Neuroscience*, 30(9):3489–3498, 2010.

[20] Emily A Meyers and John A Kessler. TGF-β family signaling in neural and neuronal differentiation, development, and function. *Cold Spring Harb. Perspect. Biol.*, 9(8), August 2017.

[21] Kunlin Jin, Yonghua Zhu, Yunjuan Sun, Xiao Ou Mao, Lin Xie, and David A. Greenberg. Vascular endothelial growth factor (vegf) stimulates neurogenesis in vitro and in vivo. *Proceedings of the National Academy of Sciences*, 99(18):11946–11950, 2002.

[22] Roberta Ramos Cavalcanti, Fernanda Martins Almeida, Ana Maria Blanco Martinez, and Camila Marques Freria. Neuroinflammation: targeting microglia for neuroprotection and repair after spinal cord injury. *Frontiers in Immunology*, Volume 16 - 2025, 2025.

[23] Xinhao Liu, Ron Zeira, and Benjamin J Raphael. Paste2: partial alignment of multi-slice spatially resolved transcriptomics data. *bioRxiv*, 2023.

A Appendix

A.1 Learning manifold-preserving representations

The vector field v_θ could be trained using any latent representation of the transcriptomic profiles, concatenated with spatial coordinates. In order to encourage flows to move on the data manifold, in our applications, we train a "manifold-aware" variational autoencoder (VAE). For this, we first estimate manifold distances with a neighborhood graph and use these to regularize the embedding space learned in VAE training. In large datasets, since computing manifold distances can be inefficient both in terms of computational time and memory, we give the users an option to use shortest graph distances using Dijkstra's algorithm instead.

A.2 Estimating Manifold Distances.

We estimate manifold distances by building a k-nearest neighbor (k-NN) graph of cells based on the top 50 principal components (PCs) of log-normalized counts, using cosine distance and $k = 30$ by default. On this graph, we define a symmetric affinity matrix

$$K_{ij} = \begin{cases} \exp\bigl(-\|x_i - x_j\|_2^2/\varepsilon\bigr), & (i,j) \text{ is an edge,} \\ 0, & \text{otherwise,} \end{cases}$$

where ε is the median edge length. While K gives us a notion of local similarity between cells, it is not a true manifold metric. It only considers one-hop distances and raw affinities vary with sampling density (e.g. dense regions of the graph will have inflated values). To correct for density bias, we construct the Markov diffusion operator, $\mathbf{P}$:

$$P = D^{-1}\widetilde{K} \quad \text{with}$$
$$\widetilde{K} \;=\; Q^{-1}KQ^{-1}, \qquad Q_{ii} = \sum\nolimits_j K_{ij}, \qquad D_{ii} = \sum\nolimits_j \widetilde{K}_{ij}.$$

Here, the left-right scaling by the local kernel mass Q^{-1} removes first-order density bias, and the subsequent row-stochastic normalization by D^{-1} turns P into a probability-transition matrix, where each cell diffuses the same total mass per step, regardless of the number of neighbors. However, $\mathbf{P}$ still only reflects one-step transitions, capturing purely local geometry. To compare cells that are connected through multi-hop paths, we let the random walk evolve over a data-driven t^* number of steps. The t-step distribution $P_{i:}^t$ describes where a particle that started at cell i is found after t steps. While small t values only probe micro-neighborhoods, large values risk mixing transition across entire branches. We let t be a user-determined hyperparameter in the default setting. However, we also give an option to automatically determine a data-driven diffusion time t^* by locating the knee of the spectral-entropy curve, which yields the largest scale that preserves manifold structure without over-smoothing. For this, let $(\lambda_\ell)_{\ell \geq 0}$ be the eigenvalues of P with $\lambda_0 = 1 > \lambda_1 \geq \lambda_2 \geq \cdots$. For each time t, we compute the spectral entropy $H(t) \;=\; -\sum_{\ell \geq 1} \lambda_\ell^{2t} \, \log \lambda_\ell^{2t}$, and select t^* at the knee point of the $H(t)$ curve, which gives the largest scale before over-smoothing. Using this single, data-driven scale t^* we define pairwise manifold distances between cells as:

$$d_{ij} = \left\| P_{i:}^{t^*} - P_{j:}^{t^*} \right\|_1^{1/2}, \tag{3}$$

which serves as our ground-truth manifold metric. In rare datasets, the spectral-entropy curve may be nearly flat with no clear knee. In such a case, we check the second eigenvalue λ_1 of P to estimate if the random walk mixes too slowly because the walk is locally trapped in well-separated clusters ($\lambda_1 \geq 0.98$) or if it mixes fast due to the high homogeneity in the data ($\lambda_1 < 0.98$). In the latter scenario, we default to local scale $t^* = 1$ and in the former, we set $t^* = \lceil 1/(1-\lambda_1) \rceil$, which is expected to the random walk run long enough to move beyond each cell's immediate neighbourhood and smooth out local noise, yet stop before it mixes probability across distinct branches.

A.3 VAE Training.

Given the pre-computed manifold distances, we train a $\beta-$VAE with an isotropic Gaussian prior $p(z) = \mathcal{N}(0,1)$ in two stages. For the initial 10% of epochs ("warm-up stage"), only the reconstruction loss (Gaussian likelihood on log-counts) and an annealed KL term are optimized, allowing the

decoder to stabilize:

$$\mathcal{L}_{\text{warm}} \;=\; \mathcal{L}_{\text{recon}} \;+\; \beta_{\text{KL}}\,\mathcal{L}_{\text{KL}}, \quad \text{with} \quad \beta_{\text{KL}}\colon 0 \to 1 \;\text{ linearly annealed}$$

For a mini-batch size B, the training loss would be:

$$\mathcal{L}_{\text{recon}} = \frac{1}{B}\sum_{i=1}^{B}\|x_i - \hat{x}_i\|_2^2,$$

$$\mathcal{L}_{\text{KL}} = \text{KL}(q_\phi(z|x)||p(x)) = \frac{1}{2B}\sum_{i=1}^{B}\sum_{d=1}^{D}\Big(\mu_{id}^2 + \sigma_{id}^2 - 1 - \log\sigma_{id}^2\Big),$$

where x_i is the log-normalized expression of cell i, $\hat{x}_i$ is the decoder output, and $q_\phi(z|x_i) = \mathcal{N}(\mu_i, \text{diag}(\sigma_i^2))$ is the encoder distribution. Following this, we introduce a triplet regularizer term $\mathcal{L}_{\text{tri}}$ that aligns Euclidean latent distances with the manifold metric:

$$\mathcal{L}_{\text{tri}} = \frac{1}{B}\sum_{i=1}^{B}\big[\|z_a - z_p\|_2 - \|z_a - z_n\|_2 + m\big]_+$$

Here, each anchor a is a batch sample, the positive sample p is one of its nearest neighbors, while the negative sample n is a distant cell with $d_{an} > \tau$, where the threshold τ reflects the 90^{th} percentile of all d_{ij}. We choose $m = 0.5$ as a margin large enough to prevent the optimizer from collapsing all points onto one another (if $m = 0$) but also small compared with the typical latent spread ($\sqrt{D}$, which would be ~ 3 for $D = 10$). The full objective used after the warm-up phase is:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + \beta\,\mathcal{L}_{\text{KL}} + \lambda_{\text{tri}}(t)\,\mathcal{L}_{\text{tri}} \quad \text{with} \quad \lambda_{\text{tri}}\colon 0 \to 5 \;\text{ linearly annealed}$$

Since the triplet term needs only one positive and one negative sample per anchor, memory and compute stay $O(B)$ per batch, while the latent Euclidean metric converges to the diffusion-based geodesic over multiple batches and epochs throughout the training.

A.4 Aligning spatial transcriptomic snapshots in a shared coordinate system

Given the semi-balanced FGW coupling $\mathbf{P}^*(t_i, t_{i+1}) \in \mathbb{R}_{\geq 0}^{n_i \times n_{i+1}}$ (Eq. 4), we convert this soft correspondence into an explicit rigid transformation that places two 2-D *spatial transcriptomic snapshots* (henceforth, "slices") in a shared Cartesian frame. The procedure extends the weighted Procrustes/Kabsch solution of PASTE2 [23] to the unbalanced setting in which $\mathbf{P}^*\mathbf{1} \neq \mathbf{a}$ and $\mathbf{1}^\top\mathbf{P}^* \neq \mathbf{b}$.

First, coupling weights are normalised as

$$w_{ij} = \frac{P^*_{ij}}{\sum_{i',j'} P^*_{i'j'}}, \qquad \omega_i = \sum_{j=1}^{n_{i+1}} w_{ij}, \qquad \upsilon_j = \sum_{i=1}^{n_i} w_{ij},$$

with $\sum_{i,j} w_{ij} = 1$; entries below a threshold $\tau = 10^{-4}$ are discarded.

Let $s_k^{t_i} \in \mathbb{R}^2$ denote the spatial coordinates of cell k in snapshot t_i and $s_\ell^{t_{i+1}} \in \mathbb{R}^2$ those of cell ℓ in snapshot t_{i+1}. Weighted centroids are

$$\mu_{t_i} = \frac{\sum_k \omega_k\, s_k^{t_i}}{\sum_k \omega_k}, \qquad \mu_{t_{i+1}} = \frac{\sum_\ell \upsilon_\ell\, s_\ell^{t_{i+1}}}{\sum_\ell \upsilon_\ell},$$

and centred coordinates are $\tilde{s}_k^{t_i} = s_k^{t_i} - \mu_{t_i}$ and $\tilde{s}_\ell^{t_{i+1}} = s_\ell^{t_{i+1}} - \mu_{t_{i+1}}$.

The weighted orthogonal Procrustes problem builds the cross-covariance

$$\mathbf{C} = \sum_{k=1}^{n_i}\sum_{\ell=1}^{n_{i+1}} w_{k\ell}\,\tilde{s}_k^{t_i}\left(\tilde{s}_\ell^{t_{i+1}}\right)^\top \in \mathbb{R}^{2\times 2},$$

whose singular-value decomposition $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ yields the optimal rotation $\mathbf{R} = \mathbf{U}\mathbf{V}^\top \in \text{SO}(2)$ that minimises $\sum_{k,\ell} w_{k\ell}\|\tilde{s}_k^{t_i} - \mathbf{R}\tilde{s}_\ell^{t_{i+1}}\|^2$.

Applying the rigid transform gives aligned coordinates

$$s'^{\,t_{i+1}}_{\ell} = \mathbf{R}\, s^{t_{i+1}}_{\ell} + \big(\mu_{t_i} - \mathbf{R}\mu_{t_{i+1}}\big), \qquad \ell = 1, \ldots, n_{i+1},$$

where the translation vector is $\mathbf{t} = \mu_{t_i} - \mathbf{R}\mu_{t_{i+1}}$.

For a temporal series $\{t_1, \ldots, t_K\}$, successive pairs (t_k, t_{k+1}) are aligned to obtain $(\mathbf{R}_k, \mathbf{t}_k)$; composing these yields

$$\big(\mathbf{R}^{(m)}, \mathbf{t}^{(m)}\big) = \Big(\mathbf{R}_1 \cdots \mathbf{R}_{m-1},\ \mathbf{t}_1 + \sum_{p=2}^{m-1} \mathbf{R}_1 \cdots \mathbf{R}_{p-1}\mathbf{t}_p\Big), \qquad m > 1,$$

so that every snapshot ultimately resides in a single common 2-D coordinate system.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We demonstrate experimental results based on our exploration so far that support the claims made in the introduction and abstract.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We acknowledge the preliminary nature of our results as this is a work-in-progress. We also discuss methodological limitations in the Discussion section, such as unaccounted-for biological factors and technical confounders in our model.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: Our work does not involve any theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We detail dataset preprocessing steps as well as all the details needed to independently implement our model in the Methods and Appendix sections. We provide also a link to the GitHub repository, where we host our source code and include information on and links to the pre-processed datasets we analyze in this study.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide the link to our public code repository in the Methods section, along with the links to the datasets.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide sufficient detail on experimental details in the paper to make sense of the results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Our preliminary results did not include repeat experiments but future work will involve statistical analysis.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the details of the compute environment our model was developed and tested in in the README file of our public repository. As we do not present a large-scale model (e.g. a foundation model), most modern computing setups would be expected to successfully run our code.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We reviewed the NeurIPS Code of Ethics and confirm that we conform to it.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification:We do not believe our work has a direct societal impact to discuss. Our main area of impact is to potentially enable novel biological insights.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We do not believe our model has a high risk of misuse or requires any safeguards to be put in place.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets and computational tools used in this work are properly cited. We rely exclusively on publicly available data and open-source academic software.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: This work introduces a new computational framework, with its source code and standard documentation made publicly available. The documentation will be continually improved as we work on further developing the methodology.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA] .

    Justification: The paper does not involve any human subjects or crowd-sourced work.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve research with human subjects, crowdsourcing or any other work that requires an IRB approval.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: Core method development in this research does not involve any LLMs; however, LLMs have been used for improving writing and grammar.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.