NICHETIME: SPATIOTEMPORAL GENERATIVE MODELING OF CELLULAR MICROENVIRONMENT EVOLUTION VIA DUAL-SCALE OPTIMAL TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning continuous-time evolution of structured systems from unpaired observations remains a fundamental challenge in generative modeling. Existing trajectory inference and generative approaches often assume conservative dynamics or overlook the hierarchical organization and directional constraints of spatiotemporal biological systems. We introduce **NicheTime**, a continuous-time generative framework for modeling non-conservative spatiotemporal dynamics through transport-guided flow matching. NicheTime addresses three key challenges: unknown correspondence across temporal observations, multi-scale structural evolution, and irreversible dynamics with variable mass flow. NicheTime develops a dual-scale optimal transport coupling that first aligns coarse-grained structures and then refines local correspondences, while allowing flexible mass variation to capture non-conservative transitions. Based on these transport plans, we propose a direction-aware Graph Transformer flow matching architecture that incorporates temporal ordering into message propagation and jointly generates system states and spatial configurations through neural ODE integration. We evaluate NicheTime on three time-resolved spatial transcriptomics datasets covering embryonic development, neural development, and aging. Extensive experiments show that NicheTime improves trajectory reconstruction, state fidelity, and spatial consistency over existing approaches, highlighting the potential of transport-guided flow matching for learning complex structured dynamics.m

1 INTRODUCTION

Learning continuous-time dynamics from unpaired cross-sectional observations requires inferring how distributions evolve without access to individual trajectories. This problem is especially challenging when observations are structured, since correspondence may depend on both global organization and local interactions, while changes in population composition can violate strict mass-preserving assumptions.

Optimal transport (OT) provides a principled way to couple unpaired distributions, and flow matching enables continuous-time generative modeling Lipman et al. (2022); Liu et al. (2022); Albergo & Vanden-Eijnden (2022); Eijkelboom et al. (2024). However, existing approaches typically address these components separately. Single-scale OT can overlook hierarchical correspondence, balanced transport can impose unrealistic marginal constraints under heterogeneous population changes, and standard graph or point-cloud models do not explicitly incorporate directional information into local message passing. These limitations make it difficult to learn structured dynamics from unpaired snapshots with both spatial organization and temporal asymmetry.

Time-resolved spatial transcriptomics provides a representative instance of this setting. Each snapshot records cellular states together with spatial coordinates, but measurements at different timepoints are obtained from different specimens and therefore lack cell-level correspondence Moses & Pachter (2022); Vu et al. (2022); Nobori et al. (2023). Tissue organization is naturally hierarchical, with cells interacting within local niches whose composition and spatial arrangement change over time. In addition, proliferation, cell loss, density heterogeneity, and sampling variability make

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

strict mass-preserving alignment inappropriate, while developmental progression provides an approximate temporal ordering of cellular states.

We introduce **NicheTime**, a continuous-time generative framework that combines dual-scale optimal transport with direction-aware graph flow matching. NicheTime first aligns source and target niches and then estimates cell-level correspondences within matched niche pairs. A unified regularized OT formulation supports balanced, unbalanced, and semi-relaxed coupling, allowing unequal source contributions as local population composition changes. The resulting transport plans define correspondences for conditional flow matching. To model structured dynamics, NicheTime represents each niche as a cell graph and parameterizes the conditional vector field with a Graph Transformer. We further introduce a pseudotime-guided graph layer that softly attenuates messages inconsistent with the inferred temporal ordering, providing directional inductive bias without enforcing a hard graph orientation.

Our contributions are summarized as follows.

- We develop a **dual-scale optimal transport framework** that couples unpaired snapshots at both niche and cell levels, with flexible marginal constraints to accommodate heterogeneous population changes across time.
- We introduce an **explicit graph representation of spatial niches** for flow matching, preserving cell-cell relational structure while jointly modeling molecular states and spatial organization.
- We propose a **pseudotime-guided graph layer** that incorporates approximate temporal ordering through soft directional message passing, attenuating backward information propagation without imposing hard directional constraints.

2 Related Work

2.1 Trajectory Inference and Flow Matching from Unpaired Snapshots

Inferring temporal dynamics from cross-sectional observations has been studied extensively in single-cell and spatial omics. RNA-velocity methods estimate local temporal direction from transcriptional kinetics Aivazidis et al. (2025), while optimal-transport-based approaches infer couplings between cell populations observed at different timepoints Klein et al. (2025); Huguet et al. (2022); Shen et al. (2025). In spatial transcriptomics, the problem is further complicated by the lack of cell-level correspondence across destructive measurements and by changes in spatial organization over time Moses & Pachter (2022); Vu et al. (2022); Nobori et al. (2023). Recent work therefore considers generative models that jointly evolve cellular states and spatial coordinates at the level of local microenvironments Sakalyan et al. (2026).

Flow matching provides a simulation-free framework for learning continuous transformations between distributions Liu et al. (2022); Lipman et al. (2022); Albergo & Vanden-Eijnden (2022), with variational formulations extending the choice of conditional objectives Eijkelboom et al. (2024). When observations are unpaired, the learned flow depends on the coupling used to construct source–target training pairs. Existing applications commonly rely on a single transport plan between individual observations or pooled representations Huguet et al. (2022); Shen et al. (2025); Sakalyan et al. (2026). NicheTime instead constructs correspondence at both niche and cell resolutions before learning the conditional flow.

2.2 Optimal Transport with Structured and Relaxed Correspondence

Optimal transport provides a principled framework for matching empirical distributions under a prescribed cost. Balanced OT constrains the transported marginals to match the source and target distributions exactly, whereas unbalanced and semi-relaxed formulations relax one or both marginal constraints. Such relaxation is useful when observed populations differ in local abundance or when a strictly mass-preserving correspondence is not appropriate.

Most trajectory-inference applications formulate transport at a single resolution Klein et al. (2025); Shen et al. (2025). For spatial microenvironments, coupling pooled niche representations captures

coarse correspondence but does not directly resolve relationships among the cells contained within each niche Sakalyan et al. (2026). NicheTime adopts a dual-scale construction in which niche-level transport identifies corresponding local environments and cell-level transport subsequently estimates fine-grained correspondence within matched niche pairs. The same regularized objective supports balanced, unbalanced, and semi-relaxed marginal constraints, allowing the correspondence assumption to be adjusted without changing the overall coupling framework.

2.3 Graph Modeling of Spatial and Directional Structure

Graph neural networks provide a natural representation for spatially organized cellular data by explicitly encoding relationships between neighboring cells Zhao et al. (2023). Common architectures such as GCN, GAT, and GIN aggregate information along graph edges Yao et al. (2019); Veličković et al. (2017); Chen et al. (2019), while Graph Transformers combine local relational structure with attention-based interactions over longer ranges Shehzad et al. (2026). In contrast, point-cloud architectures treat cells primarily as permutation-invariant sets and do not explicitly represent local cell–cell edges unless such relations are introduced separately Sakalyan et al. (2026).

Standard graph message passing is typically agnostic to temporal ordering. In developmental settings, however, cells may be associated with an approximate ordering such as pseudotime, providing additional information about the direction of state progression. Existing undirected aggregation does not distinguish messages that follow or oppose this ordering. NicheTime explicitly represents each niche as a cell graph and introduces a pseudotime-guided graph layer that uses temporal ordering to modulate edge-wise message propagation. The resulting mechanism imposes a soft directional bias rather than a hard graph orientation, allowing local deviations from the estimated ordering.

3 Methodology

As illustrated in Figure 1, NicheTime learns continuous dynamics between unpaired structured observations by jointly addressing cross-time correspondence and structure-aware generation. We represent each observation as a collection of local cell graphs and infer latent correspondences through optimal transport at niche and cell resolutions. These correspondences define the training distribution for a conditional flow model, whose representation incorporates relational structure and approximate temporal ordering. For reproducibility, the implementation of NicheTime is available in an anonymous repository at `https://anonymous.4open.science/r/NicheTime-review-ICLR2027`.

3.1 Problem Setup

We formulate learning from unpaired observations as continuous-time generative modeling over structured states. A snapshot at time t is represented as a collection of cellular states

$$\mathcal{D}_t = \{(x_i, p_i, c_i, pt_i)\}_{i=1}^{N_t}, \tag{1}$$

where $x_i \in \mathbb{R}^{d_x}$ denotes the expression embedding, $p_i \in \mathbb{R}^2$ represents spatial coordinates, c_i denotes the cell category, and pt_i is an optional temporal ordering prior. Given two unpaired observations $\mathcal{D}_0$ and $\mathcal{D}_1$, our goal is to learn a conditional vector field that transforms source states into target states while preserving relational and spatial organization.

Structured niche representation. To preserve local relational structure, we first partition each observation into spatially localized cell groups. For a spatial center μ_u, the corresponding niche contains the cells within a radius r:

$$\mathcal{V}_u = \{i \mid \|p_i - \mu_u\|_2 \le r\}, \tag{2}$$

Based on the cells within each niche, we construct a graph

$$G_u = (\mathcal{V}_u, \mathcal{E}_u), \tag{3}$$

For each cell i, we summarize the composition of its spatial neighborhood as

$$\rho_i = \frac{1}{k} \sum_{j \in \mathrm{kNN}(i)} \mathrm{onehot}(c_j). \tag{4}$$

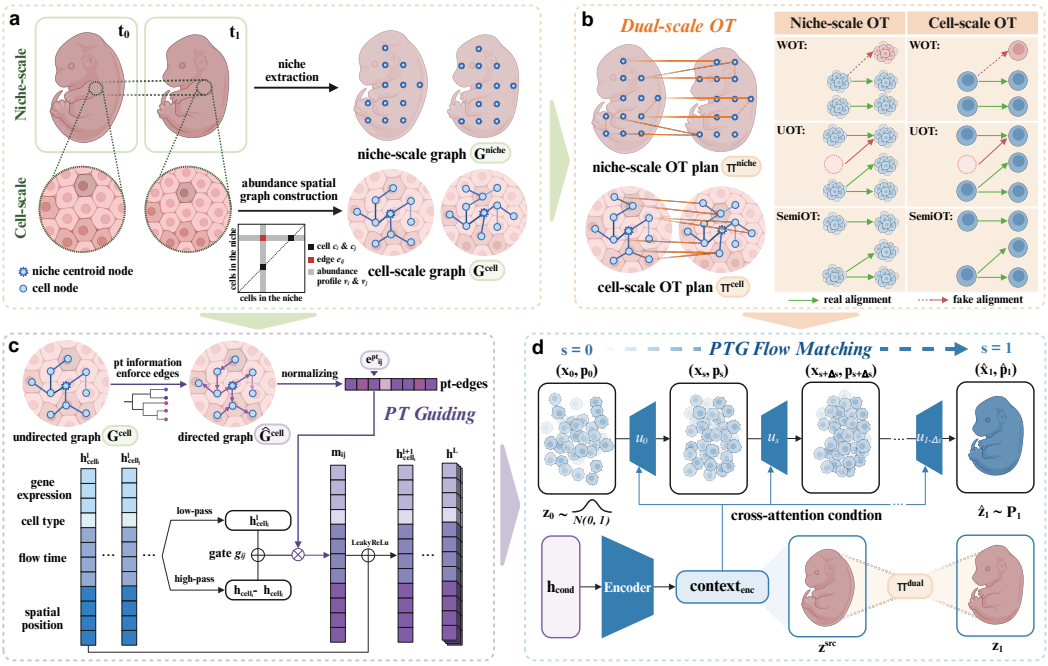


Figure 1: **Overview of NicheTime.** NicheTime learns continuous-time evolution from unpaired structured observations through three components. **(a)** Each observation is represented as a hierarchy of spatial graphs, where local niches preserve cell-scale relational structure. **(b)** A dual-scale optimal transport module estimates cross-time correspondence at both niche and cell resolutions under flexible marginal constraints. **(c)** A direction-aware graph flow matching model uses the inferred correspondence to learn a continuous transformation of cellular states and spatial coordinates. **(d)** Conditioned on the PTG-encoded source and guided by the OT plans, a neural ODE transports noise to the target, jointly evolving cell expressions and spatial coordinates.

The relation between neighboring cells i and j is weighted by

$$e_{ij} = \cos(\rho_i, \rho_j)\mathbb{I}\left[\cos(\rho_i, \rho_j) > \tau\right]. \tag{5}$$

The resulting graph explicitly encodes local cell-cell relations while retaining node-scale feature and spatial information. The same structured representation is used for both correspondence estimation and conditional flow learning.

3.2 Dual-Scale Optimal Transport Coupling

Flow learning requires correspondences between source and target states, which are unavailable for unpaired observations. A global cell-scale coupling captures fine-grained correspondence but does not explicitly account for the organization of cells into local structures and becomes expensive for large observations. Conversely, coupling only aggregated niches loses cell-scale resolution. We therefore estimate correspondence at two complementary scales. A niche-scale plan first identifies related local structures, after which cell-scale plans refine correspondence within matched niche pairs. (see Appendix A for details)

We use the following regularized transport operator throughout both scales

$$\mathcal{T}_\epsilon(C, a, b) = \arg\min_{\Pi \geq 0} \langle C, \Pi \rangle + \epsilon H(\Pi) + \mathcal{R}_s(\Pi \mathbf{1}, a) + \mathcal{R}_t(\Pi^\top \mathbf{1}, b), \tag{6}$$

where C is the transport cost, a and b are source and target marginals, and $H(\Pi)$ is the entropic regularizer. Different choices of marginal penalties recover balanced, unbalanced, and semi-relaxed

transport

$$(\mathcal{R}_s, \mathcal{R}_t) = \begin{cases} (\iota_a, \iota_b), & \text{balanced}, \\ (\lambda_s D_{\mathrm{KL}}, \lambda_t D_{\mathrm{KL}}), & \text{unbalanced}, \\ (\lambda_s D_{\mathrm{KL}}, \iota_b), & \text{semi-relaxed}, \end{cases} \tag{7}$$

where ι_a and ι_b denote hard marginal constraints. The semi-relaxed formulation preserves the target marginal while allowing unequal source contributions, providing a flexible correspondence model when local compositions differ across observations.

Niche-scale coupling. Let

$$\mathcal{N}_0 = \{\mathcal{V}_u^0\}_{u=1}^{M_0}, \qquad \mathcal{N}_1 = \{\mathcal{V}_v^1\}_{v=1}^{M_1} \tag{8}$$

denote source and target niche collections. Each niche is summarized by its mean feature and spatial coordinate

$$\bar{x}_u^t = \frac{1}{|\mathcal{V}_u^t|} \sum_{i \in \mathcal{V}_u^t} x_i, \qquad \bar{p}_u^t = \frac{1}{|\mathcal{V}_u^t|} \sum_{i \in \mathcal{V}_u^t} p_i. \tag{9}$$

We define the niche-scale transport cost as

$$C_{uv}^{\mathrm{niche}} = \alpha_{\mathrm{niche}} \|\bar{x}_u^0 - \bar{x}_v^1\|_2^2 + (1 - \alpha_{\mathrm{niche}}) \|\bar{p}_u^0 - \bar{p}_v^1\|_2^2, \tag{10}$$

which yields

$$\Pi^{\mathrm{niche}} = \mathcal{T}_\epsilon(C^{\mathrm{niche}}, a, b). \tag{11}$$

Cell-scale coupling. For each niche pair (u, v) sampled according to Π^{niche}, we refine correspondence between their constituent cells. The cell-scale cost is

$$C_{ij}^{\mathrm{cell}} = \alpha_{\mathrm{cell}} \|x_i^0 - x_j^1\|_2^2 + (1 - \alpha_{\mathrm{cell}}) \|p_i^0 - p_j^1\|_2^2, \tag{12}$$

and the corresponding plan is

$$\Pi_{uv}^{\mathrm{cell}} = \mathcal{T}_\epsilon(C_{uv}^{\mathrm{cell}}, a_u, b_v). \tag{13}$$

Sampling (i, j) from Π_{uv}^{cell} provides a source graph condition G_u^0 together with a corresponding target state

$$z_j^1 = [x_j^1 \| p_j^1]. \tag{14}$$

These transport-induced samples define the training distribution used by the conditional flow model.

3.3 Pseudotime-Guided Graph Flow

Given the source graph G_u^0, NicheTime learns a structured representation that conditions the continuous-time generative process. Conventional symmetric message passing encodes relational structure but does not distinguish whether neighboring information follows or opposes the available temporal ordering. We therefore introduce a pseudotime-guided graph layer that applies a soft directional modulation to local message propagation.

For each source node, we construct the initial representation

$$h_i^{(0)} = \left[E_x(x_i \| \mathrm{onehot}(c_i)) \| \Phi(s) \| E_p(p_i)\right], \tag{15}$$

where $\Phi(s)$ denotes the flow-time embedding. Pseudotime is normalized by its within-observation rank

$$\widetilde{pt}_i = \frac{\mathrm{Rank}(pt_i)}{|\mathcal{D}_t|}. \tag{16}$$

For a message propagated from node j to node i, we modulate the structural edge weight as

$$\widetilde{e}_{ij} = e_{ij} \exp\left[-\beta \mathrm{ReLU}(\widetilde{pt}_j - \widetilde{pt}_i)\right]. \tag{17}$$

This operation attenuates messages that oppose the inferred temporal ordering without imposing a hard directional constraint.

The directional weights are combined with an adaptive low- and high-frequency message operator

$$m_{ij} = g_{ij} W_{\mathrm{lo}}(\widetilde{e}_{ij} h_j) + (1 - g_{ij}) W_{\mathrm{hi}} \left[\widetilde{e}_{ij}(h_i - h_j)\right], \tag{18}$$

where

$$g_{ij} = \sigma\left(w_g^\top [h_i \| \Phi(s) \| \widetilde{pt}_i \| \widetilde{pt}_i - \widetilde{pt}_j]\right). \tag{19}$$

Node representations are updated according to

$$h_i^{(l+1)} = \mathrm{LeakyReLU}\left(h_i^{(l)} + \sum_{j \in \mathcal{N}(i)} m_{ij}\right). \tag{20}$$

The PTG layer captures local structure and directional information. We further apply global self-attention over the cell tokens within each niche to model dependencies beyond the local graph neighborhood. The resulting representation

$$H_u^0 = \mathrm{Enc}_\theta(G_u^0, s) \tag{21}$$

serves as the structured condition for continuous-time generation.

3.4 Direction-Aware Conditional Flow Matching

For a target state $z^1 = [x^1 \| p^1]$ sampled from the transport-induced correspondence, we draw an independent base sample $\xi \sim \mathcal{N}(0, I)$ and construct the linear probability path

$$z_s = (1-s)\xi + s z^1, \qquad s \in [0,1]. \tag{22}$$

The source observation does not constitute the base endpoint of this path. Instead, its graph representation H_u^0 conditions the generation of the matched target state.

The conditional flow model takes the intermediate state z_s, flow time s, and source representation H_u^0 as input. Following the GLVFM formulation (Appendix B), it predicts the target feature and spatial states

$$\hat{x}_\theta(z_s, s, H_u^0), \qquad \hat{p}_\theta(z_s, s, H_u^0). \tag{23}$$

Because feature and spatial variables have different reconstruction characteristics, we use modality-specific objectives. The training loss is

$$\begin{aligned}\mathcal{L}_{\mathrm{GLVFM}}(\theta) = \mathbb{E}\Big[&\lambda_x \|\hat{x}_\theta(z_s, s, H_u^0) - x^1\|_2^2 \\ &+ \lambda_p \|\hat{p}_\theta(z_s, s, H_u^0) - p^1\|_1\Big],\end{aligned} \tag{24}$$

where λ_x and λ_p control the relative contributions of feature and spatial reconstruction. The learned conditional flow defines a time-dependent vector field $v_\theta(z_s, s \mid H_u^0)$, which is integrated during inference from the Gaussian base distribution

$$\frac{dz_s}{ds} = v_\theta(z_s, s \mid H_u^0), \qquad z_{s=0} \sim \mathcal{N}(0, I). \tag{25}$$

The terminal state $z_{s=1}$ provides the generated cellular features and spatial coordinates conditioned on the source structured observation. (see Appendix F for algorithms and Appendix G for implementation details)

4 Experiments

4.1 Experimental Setup

We evaluate NicheTime on three time-resolved spatial transcriptomics datasets: Mouse Embryogenesis (MED) Klein et al. (2025); Chen et al. (2022), Axolotl Brain Development (ABD) Wei et al. (2022), and Mouse Brain Aging (MBA) Sun et al. (2025). These datasets cover diverse temporal scales, spatial organizations, and dataset sizes, providing a comprehensive evaluation of structured dynamics modeling. Unless otherwise specified, NicheTime uses the SemiOT-SemiOT configuration for both niche- and cell-level transport. (see Appendix D for further analysis)

Table 1: Comparison with NicheFlow Sakalyan et al. (2026), a representative generative framework for spatial microenvironment trajectory modeling. Lower is better for MSE and MAE; higher is better for 1NN-F1 (Appendix C) . Bold: best; underline: second-best. (Results reported under the SemiOT-SemiOT configuration).

Method	ABD			MBA			MED		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
NicheFlow	1.5828 ±0.0009	0.9447 ±0.0003	0.6278 ±0.0008	1.5414 ±0.0006	0.9565 ±0.0002	0.2847 ±0.0003	**1.3775** ±0.0014	**0.8921** ±0.0004	0.6634 ±0.0021
$\text{NicheTime}_{\text{NONE}}$	1.9339 ±0.0033	1.0525 ±0.0010	0.6128 ±0.0023	1.7073 ±0.0044	1.0084 ±0.0013	**0.3014** ±0.0016	1.6511 ±0.0059	0.9879 ±0.0019	0.6395 ±0.0036
$\text{NicheTime}_{\text{PTG}}$	**1.4189** ±0.0023	**0.8970** ±0.0008	**0.6659** ±0.0012	**1.2637** ±0.0012	**0.8535** ±0.0004	0.2584 ±0.0003	1.5636 ±0.0034	0.9595 ±0.0012	**0.6642** ±0.0050

Table 2: Quantitative comparison of dual-scale optimal transport formulations across niche and cell scales. Lower is better for MSE and MAE; higher is better for 1NN-F1. Bold: best; underline: second-best, within each niche-scale group. The SemiOT-SemiOT configuration is adopted for all subsequent main experiments; see Section 4.3.

Niche-scale	Cell-scale	ABD			MBA			MED		
		MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
WOT	WOT	**1.6488** ±0.0011	**0.9635** ±0.0004	**0.5785** ±0.0011	1.6185 ±0.0005	0.9795 ±0.0001	0.2836 ±0.0004	**1.5036** ±0.0043	**0.9341** ±0.0013	0.6392 ±0.0018
	UOT	1.6789 ±0.0020	0.9735 ±0.0007	0.5535 ±0.0020	**1.5642** ±0.0008	0.9634 ±0.0002	0.2869 ±0.0004	1.5333 ±0.0027	0.9429 ±0.0009	**0.6479** ±0.0027
	SemiOT	1.7105 ±0.0020	0.9823 ±0.0005	0.5238 ±0.0012	1.5644 ±0.0005	**0.9630** ±0.0002	**0.2886** ±0.0004	1.5484 ±0.0024	0.9484 ±0.0008	0.6382 ±0.0017
UOT	WOT	1.7721 ±0.0058	1.0062 ±0.0016	0.4074 ±0.0012	1.3487 ±0.0005	**0.8867** ±0.0002	0.2422 ±0.0007	1.5719 ±0.0046	0.9636 ±0.0012	0.6482 ±0.0053
	UOT	**1.5677** ±0.0036	**0.9454** ±0.0011	0.6409 ±0.0021	**1.3450** ±0.0007	0.8871 ±0.0002	**0.2639** ±0.0011	**1.5559** ±0.0052	**0.9580** ±0.0015	**0.6748** ±0.0056
	SemiOT	1.5725 ±0.0039	0.9465 ±0.0012	**0.6432** ±0.0017	1.4661 ±0.0009	0.9262 ±0.0003	0.2196 ±0.0004	1.5796 ±0.0058	0.9656 ±0.0014	0.6496 ±0.0047
SemiOT	WOT	1.4209 ±0.0033	0.8982 ±0.0010	**0.6684** ±0.0010	1.2818 ±0.0012	0.8616 ±0.0004	0.2385 ±0.0005	1.5509 ±0.0052	**0.9549** ±0.0018	**0.6854** ±0.0040
	UOT	1.4543 ±0.0022	0.9102 ±0.0007	0.6650 ±0.0018	1.3507 ±0.0008	0.8828 ±0.0003	0.2294 ±0.0003	**1.5504** ±0.0039	0.9559 ±0.0015	0.4499 ±0.0071
	SemiOT	**1.4189** ±0.0023	**0.8970** ±0.0008	0.6659 ±0.0012	**1.2637** ±0.0012	**0.8535** ±0.0004	**0.2584** ±0.0003	1.5636 ±0.0034	0.9595 ±0.0012	0.6642 ±0.0050

4.2 Comparison with Existing Methods

We compare NicheTime with NicheFlow Sakalyan et al. (2026), the closest existing generative framework that jointly models cellular states and spatial coordinates of local microenvironments.

Table 1 evaluates the proposed framework against NicheFlow Sakalyan et al. (2026), the closest existing generative approach for modeling cellular microenvironment dynamics. We compare NicheFlow with $\text{NicheTime}_{\text{NONE}}$ to assess the contribution of replacing single-scale transport with dual-scale correspondence at the niche and cell levels, while $\text{NicheTime}_{\text{PTG}}$ further incorporates explicit graph structure and pseudotime-guided message passing. The results indicate that dual-scale correspondence improves reconstruction quality in several settings, while the additional PTG module further enhances performance on multiple datasets and metrics, highlighting the complementary roles of multi-resolution correspondence and direction-aware relational modeling.

4.3 Ablation Study of Dual-Scale Transport

Table 2 evaluates different marginal constraints within the proposed dual-scale transport framework, complementing Table 1, which compares single- and dual-scale correspondence. By combining balanced, unbalanced, and semi-relaxed transport at the niche and cell levels, we assess how correspondence assumptions affect generation quality across datasets and metrics.

Within the SemiOT niche-scale group, SemiOT-SemiOT achieves the best performance on five of nine metrics, while SemiOT-WOT is stronger on several others, particularly on MED. We therefore adopt SemiOT-SemiOT as the default configuration because it combines strong overall performance with relaxed marginal constraints at both scales, which better reflect heterogeneous population changes during development. We additionally report SemiOT-WOT results in Appendix E to assess sensitivity to a stricter cell-level constraint.

4.4 Ablation Study of Graph Representation and Directional Modeling

Table 3 evaluates the contribution of explicit graph representation by comparing NicheTime with graph-free and standard graph neural architectures. The graph-free variant treats each niche as an unordered collection of cells, while GCN, GAT, and GIN introduce conventional symmetric message passing. The proposed PTG layer further incorporates temporal ordering into graph propagation.

The results show that explicitly modeling cell-level relations improves structured state generation compared with the graph-free variant. Moreover, PTG achieves competitive performance compared

Table 3: Quantitative comparison of graph architectures within the proposed framework. We evaluate the proposed PTG against a baseline without graph convolutions (NONE) and standard symmetric message-passing networks (GCN, GAT, GIN). Bold: best; underline: second-best.

GNN layer	ABD			MBA			MED		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
NONE	1.9339 ±0.0033	1.0525 ±0.0010	0.6128 ±0.0023	1.7073 ±0.0044	1.0084 ±0.0013	0.3014 ±0.0016	1.6511 ±0.0059	0.9879 ±0.0019	0.6395 ±0.0036
GCN	1.8464 ±0.0026	1.0302 ±0.0007	0.6181 ±0.0024	1.7330 ±0.0031	1.0137 ±0.0010	0.3035 ±0.0015	1.5935 ±0.0075	0.9690 ±0.0024	**0.6887** ±0.0058
GAT	1.9225 ±0.0023	1.0503 ±0.0008	0.6359 ±0.0025	1.6960 ±0.0036	1.0017 ±0.0012	**0.3093** ±0.0020	1.7057 ±0.0083	1.0037 ±0.0026	0.6518 ±0.0040
GIN	1.8768 ±0.0038	1.0383 ±0.0011	0.6468 ±0.0015	1.6975 ±0.0031	1.0030 ±0.0010	0.3027 ±0.0014	1.6267 ±0.0059	0.9795 ±0.0020	0.6523 ±0.0039
PTG	**1.4189** ±0.0023	**0.8970** ±0.0008	**0.6659** ±0.0012	**1.2637** ±0.0012	**0.8535** ±0.0004	0.2584 ±0.0003	**1.5636** ±0.0034	**0.9595** ±0.0012	0.6642 ±0.0050

Table 4: Quantitative ablation study of the PTG layer. We compare the full PTG architecture against its two structural variants. Bold: best; underline: second-best.

GNN layer	ABD			MBA			MED		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
$\mathrm{PTG}_{\mathrm{woH}}$	1.4459 ±0.0028	0.9072 ±0.0010	0.6652 ±0.0023	1.2948 ±0.0008	0.8648 ±0.0002	0.2239 ±0.0002	1.5794 ±0.0031	0.9641 ±0.0012	**0.6957** ±0.0045
$\mathrm{PTG}_{\mathrm{woPT}}$	1.8746 ±0.0053	1.0375 ±0.0016	0.6263 ±0.0017	1.6896 ±0.0033	1.0011 ±0.0012	**0.3070** ±0.0018	1.6449 ±0.0061	0.9856 ±0.0020	0.6395 ±0.0026
PTG	**1.4189** ±0.0023	**0.8970** ±0.0008	**0.6659** ±0.0012	**1.2637** ±0.0012	**0.8535** ±0.0004	0.2584 ±0.0003	**1.5636** ±0.0034	**0.9595** ±0.0012	0.6642 ±0.0050

with standard graph architectures, indicating that temporal direction provides additional information beyond spatial connectivity alone.

Table 4 further examines the internal components of PTG. Removing pseudotime guidance ($\mathrm{PTG}_{\mathrm{woPT}}$) substantially reduces performance, demonstrating the importance of directional information propagation. Removing the high-pass component ($\mathrm{PTG}_{\mathrm{woH}}$) also affects reconstruction quality, suggesting that adaptive feature aggregation helps preserve local structural variation.

4.5 Analysis of Generated Structured States

In addition to aggregate metrics, we evaluate whether NicheTime preserves meaningful molecular and spatial structures in its generated states. We focus this analysis on the ABD dataset, which contains heterogeneous spatial organization, multiple developmental stages, and well-annotated cell types. To facilitate a comparative presentation of our predictions of cellular developmental states, we computed pseudotime values separately for the ground-truth (GT) data and the predicted data, and internally normalized each dataset to remove batch effects.

Figure 2 further evaluates whether the generated states preserve structured variation in the feature space. We use a heatmap to show the correlations among different dimensions, with darker colors indicating higher correlations. The high-correlation blocks are concentrated along the diagonal, demonstrating that our prediction model closely reconstructs the GT characteristics in terms of cell-type clustering and expression lineages. The strong correspondence between these principal dimensions indicates that NicheTime captures coordinated feature evolution beyond merely matching global feature distributions.

Qualitative evaluations in Figures 3 show that, across the ABD trajectory, the predicted tissues preserve large-scale spatial organization. We use point-cloud visualizations in space to show that our model's predictions correspond well to the GT cell types. Furthermore, in terms of spatial location distributions, although minor discrepancies appear near volatile peripheral boundaries, the major internal compartments and developmental gradients remain intact. This confirms that our model can effectively predict the true states and spatial distributions of developing cells.

Together, these analyses provide additional evidence that the proposed transport-guided flow model preserves both feature-level and spatial-structural properties during continuous-time generation.

5 Conclusion

We introduced NicheTime, a continuous-time generative framework for learning dynamics from unpaired structured observations. NicheTime addresses three key challenges in this setting by explicitly modeling relational structure, estimating cross-time correspondence at multiple resolutions through dual-scale optimal transport, and incorporating temporal ordering into flow-based generation with a pseudotime-guided graph layer. Experiments on three time-resolved spatial transcriptomics datasets demonstrate that NicheTime improves state reconstruction, identity preservation, and spatial consis-

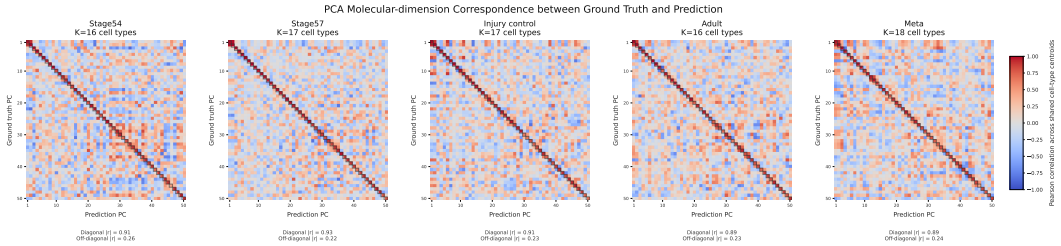


Figure 2: Correspondence of molecular dimensions between real and predicted ABD tissues. Each entry shows the Pearson correlation between PCA dimensions across shared centroids for each cell type. A stronger diagonal correlation indicates that molecular variation is preserved for specific dimensions.

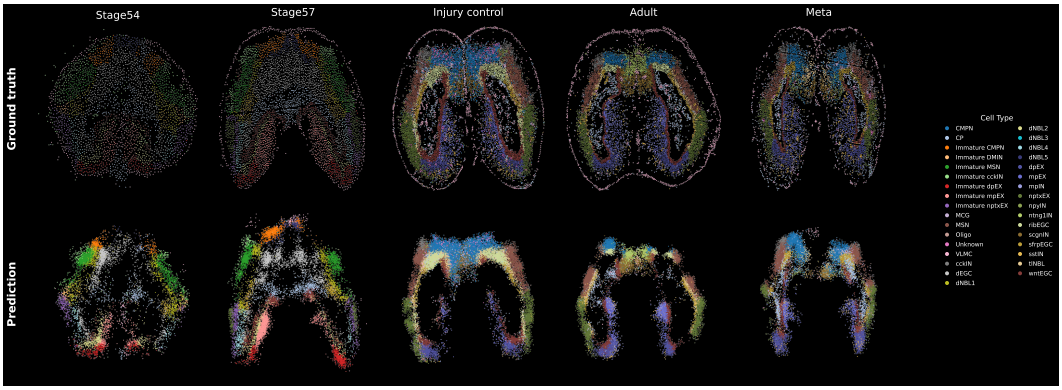


Figure 3: Spatial organization of cell types in real and predicted ABD tissues.

tency over existing approaches, while ablation studies validate the contribution of each component. These results highlight the effectiveness of combining transport-based correspondence learning with direction-aware structured flow matching for modeling complex continuous-time evolution from unpaired observations. We discuss the limitations of the current framework and potential directions for future work in Appendix H.

THE USE OF LARGE LANGUAGE MODELS (LLMS)

In this work, we used generative AI tools for the following purposes. We used AI-based code completion while writing our implementation, limited to line-level autocompletion of boilerplate, utility, logging, and plotting code. We also used generative AI tools to help write scripts that compute summary statistics over our experimental outputs in this paper. Finally, we used generative AI tools to edit and polish the writing of this paper for improved readability, including grammar correction, rephrasing for clarity, and improving the fluency of author-written text.

We did not use generative AI tools for any task requiring mandatory disclosure. Specifically, all research ideas, hypotheses, theoretical models, mathematical claims and proofs, experimental design and methodology, and interpretation of results in NicheTime were conceived, carried out, and written by the authors without the use of generative AI. The design of our method and its algorithmic implementation were likewise carried out by the authors; no AI tool was used to decide or generate the core method logic. No synthetic data were generated with AI tools, and no AI assistance was used for translation, dataset cleaning or reformatting, or qualitative and thematic data analysis.

All AI-assisted content was reviewed and verified by the authors. AI-assisted code, including the summary-statistics and plotting scripts, was inspected and tested for correctness by the authors, and we verified all reported statistics and figures against the raw experimental outputs. All AI-assisted text edits were checked to ensure that no technical claim, result, or citation was altered or introduced

by the tools. The authors take full responsibility for the final content of this work, including all text, claims, and artifacts.

REFERENCES

Alexander Aivazidis, Fani Memi, Vitalii Kleshchevnikov, Sezgin Er, Brian Clarke, Oliver Stegle, and Omer Ali Bayraktar. Cell2fate infers rna velocity modules to improve cell fate prediction. *Nature methods*, 22(4):698–707, 2025.

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M Ibrahim, Andrew J Hill, Fan Zhang, Stefan Mundlos, Lena Christiansen, Frank J Steemers, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, 566(7745):496–502, 2019.

Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

Zhengdao Chen, Soledad Villar, Lei Chen, and Joan Bruna. On the equivalence between graph isomorphism testing and function approximation with gnns. *Advances in neural information processing systems*, 32, 2019.

Floor Eijkelboom, Grigory Bartosh, Christian A Naesseth, Max Welling, and Jan-Willem van de Meent. Variational flow matching for graph generation. *Advances in Neural Information Processing Systems*, 37:11735–11764, 2024.

Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. *Advances in neural information processing systems*, 35:29705–29718, 2022.

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.

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

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5):534–546, 2022.

Tatsuya Nobori, Alexander Monell, Travis A Lee, Jingtian Zhou, Joseph Nery, and Joseph R Ecker. Time-resolved single-cell and spatial gene regulatory atlas of plants under pathogen attack. *BioRxiv*, pp. 2023–04, 2023.

Halmos Peter, Liu Xinhao, Gold Julian, Chen Feng, Ding Li, and Benjamin J Raphael. Dest-ot: Alignment of spatiotemporal transcriptomics data. *Cell Systems*, 16(2), 2025.

Kristiyan Sakalyan, Alessandro Palma, Filippo Guerranti, Fabian Theis, and Stephan Günnemann. Modeling microenvironment trajectories on spatial transcriptomics with nicheflow. *Advances in Neural Information Processing Systems*, 38:103328–103371, 2026.

Ahsan Shehzad, Feng Xia, Shagufta Abid, Ciyuan Peng, Shuo Yu, Dongyu Zhang, and Karin Verspoor. Graph transformers: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2026.

Xunan Shen, Lulu Zuo, Zhongfei Ye, Zhongyang Yuan, Ke Huang, Zeyu Li, Qichao Yu, Xuanxuan Zou, Xiaoyu Wei, Ping Xu, et al. Inferring cell trajectories of spatial transcriptomics via optimal transport analysis. *Cell systems*, 16(2), 2025.

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, 2025.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Tam Vu, Alexander Vallmitjana, Joshua Gu, Kieu La, Qi Xu, Jesus Flores, Jan Zimak, Jessica Shiu, Linzi Hosohama, Jie Wu, et al. Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis. *Nature communications*, 13(1):169, 2022.

Xiaoyu Wei, Sulei Fu, Hanbo Li, Yang Liu, Shuai Wang, Weimin Feng, Yunzhi Yang, Xiawei Liu, Yan-Yun Zeng, Mengnan Cheng, et al. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022.

Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 7370–7377, 2019.

Shen Zhao, De-Pin Chen, Tong Fu, Jing-Cheng Yang, Ding Ma, Xiu-Zhi Zhu, Xiang-Xue Wang, Yi-Ping Jiao, Xi Jin, Yi Xiao, et al. Single-cell morphological and topological atlas reveals the ecosystem diversity of human breast cancer. *Nature Communications*, 14(1):6796, 2023.

A Optimal Transport Details

To detail the optimal transport computations, we first define the notations used throughout this appendix. Let $\Pi \equiv \Pi_{uv}^{\text{cell}}$ denote the cell-scale transport plan of a matched niche pair (u, v), with $N_1 = |\mathcal{V}_u^0|$ and $N_2 = |\mathcal{V}_v^1|$ representing the source and target niche sizes, respectively. The inter-slice cost is denoted by $C = C^{\text{cell}}$, as defined in Eq. (9). A central component of our unbalanced (UOT) and semi-relaxed (SemiOT) formulations is the generalized Kullback-Leibler (KL) divergence, introduced in the marginal penalties of Eq. (8), which relaxes strict mass conservation. The divergence is defined as

$$D_{\text{KL}}(p \,\|\, q) = \sum_i \Big(p_i \log \tfrac{p_i}{q_i} - p_i + q_i \Big) \geq 0, \tag{26}$$

which vanishes if and only if $p = q$.

Leveraging this formulation, all three OT variants introduced in Eq. (8) can be efficiently computed using a single, unified entropic scaling solver, differing exclusively in the strictness of their marginal constraints. By defining the Gibbs kernel $K = \exp(-C/\epsilon) \in \mathbb{R}_+^{N_1 \times N_2}$ and expressing the transport plan as $\Pi = \operatorname{diag}(u)\, K\, \operatorname{diag}(v)$ with scaling vectors $u \in \mathbb{R}_+^{N_1}$ and $v \in \mathbb{R}_+^{N_2}$, the solver employs alternating updates:

$$u \leftarrow \Big(\frac{a}{Kv} \Big)^{\zeta_s}, \qquad v \leftarrow \Big(\frac{b}{K^\top u} \Big)^{\zeta_t}, \tag{27}$$

where the exponent ζ for a marginal constraint penalized by a KL divergence with weight λ is given by $\zeta = \lambda/(\lambda + \epsilon)$. A strict (hard) marginal constraint is recovered in the limit $\lambda \to \infty$, yielding $\zeta = 1$. Consequently, the three coupling modes are seamlessly instantiated through specific exponent configurations:

$$\begin{aligned} \text{WOT: } & \zeta_s = \zeta_t = 1, \\ \text{UOT: } & \zeta_s = \tfrac{\lambda_s}{\lambda_s + \epsilon}, \;\; \zeta_t = \tfrac{\lambda_t}{\lambda_t + \epsilon}, \\ \text{SemiOT: } & \zeta_s = \tfrac{\lambda_s}{\lambda_s + \epsilon}, \;\; \zeta_t = 1. \end{aligned} \tag{28}$$

The relative behavior of balanced, unbalanced, and semi-relaxed transport under non-conservative mass flow has been analyzed by Peter et al. (2025). We note that the time variable governing this transport is a *generative* one, namely the flow-matching time $s \in [0, 1]$ introduced in Section 3.3, and should not be conflated with biological time. NicheTime learns a continuous transport between two *observed* snapshot distributions rather than modeling a continuous biological clock; predicting biologically unobserved timepoints, such as a held-out intermediate stage, is beyond the present scope and left for future work.

This unified framework provides distinct physical interpretations for each mode. In SemiOT, the target update ($\zeta_t = 1$) acts as an exact Sinkhorn projection, enforcing $\Pi^\top \mathbf{1}_{N_1} = b$ at every iteration to ensure all target cells are accounted for. Conversely, the source update ($\zeta_s < 1$) performs a KL-relaxed step, permitting the total transported source mass to deviate from the prior a; this crucial relaxation absorbs net local cellular growth or death prior to matching. In practice, unless the density weighting described in Section 3.2 is enabled, the cell-scale source prior defaults to a uniform distribution $a = \mathbf{1}_{N_1}/N_1$, and the fixed target prior to $b = \mathbf{1}_{N_2}/N_2$. Finally, this identical iterative procedure is directly applied to the macroscopic cost C^{niche} to obtain the niche-scale plan Π^{niche}.

B Gaussian-Laplacian Variational Flow Matching

We derive the GLVFM objective of Eq. (18), following the variational view of flow matching Eijkelboom et al. (2024) and its mixed-family instantiation for spatial point clouds Sakalyan et al. (2026). We write the joint cell state as $z = (x, p)$ with expression $x \in \mathbb{R}^{d_x}$ and coordinates $p \in \mathbb{R}^2$, and $u_s(\cdot)$ for the flow-matching velocity field at flow time s, which the network v_θ of Appendix F approximates.

Variational flow matching. The flow carries the prior $z_0 \sim \mathcal{N}(0, I)$ to the target $z_1 \sim \mathbb{P}_1$ along $z_s = sz_1 + (1-s)z_0$. Its marginal field is the posterior expectation of the conditional field, $u_s(z) = \mathbb{E}_{p_s(z_1|z)}[u_s(z \mid z_1)]$, with $p_s(z_1 \mid z)$ the endpoint posterior. VFM recasts learning as variational inference, approximating $p_s(z_1 \mid z)$ by $q_s^\theta(z_1 \mid z)$ and minimizing

$$\mathcal{L}_{\text{VFM}}(\theta) = -\,\mathbb{E}_{s\sim\mathcal{U}(0,1),\, z_1\sim\mathbb{P}_1,\, z\sim p_s(z|z_1)}\big[\log q_s^\theta(z_1 \mid z)\big]. \tag{29}$$

Mean parameterization. For the straight-line interpolant, the conditional field $u_s(z \mid z_1) = (z_1 - z)/(1-s)$ is linear in z_1, so

$$u_s(z) = \big(\mathbb{E}_{p_s(z_1|z)}[z_1] - z\big)/(1-s) \tag{30}$$

depends on the posterior only through its mean. It therefore suffices to learn the posterior mean, which we identify with the decoder endpoint $\hat{z}_\theta(z_s, s) = (\hat{x}_\theta, \hat{p}_\theta)$; and since only per-dimension means enter u_s, the posterior may be fully factorized, each factor free to take its own family provided its mean is matched.

Gaussian-Laplacian factorization. We use this freedom to fit each modality with a matching distribution, a unit-variance Gaussian on every expression dimension and a unit-scale Laplace on every coordinate:

$$q_s^\theta(z_1 \mid z_s) = \prod_{d=1}^{d_x} \mathcal{N}\big(x_1^d;\, \hat{x}_\theta^d,\, 1\big) \prod_{k=1}^{2} \text{Laplace}\big(p_1^k;\, \hat{p}_\theta^k,\, 1\big). \tag{31}$$

Up to constants, their negative log-likelihoods are $\frac{1}{2}\big(x_1^d - \hat{x}_\theta^d\big)^2$ and $\big|p_1^k - \hat{p}_\theta^k\big|$; summing over dimensions yields an ℓ_2 error on expression and an ℓ_1 error on coordinates, and reinstating the modality weights λ_x, λ_p recovers Eq. (18). The choice is biologically motivated: the sharply peaked, heavy-tailed Laplace places coordinates precisely and keeps the boundaries between tissue niches crisp while tolerating the occasional migrating cell, whereas the Gaussian matches the smooth, continuous variation of the PCA expression embedding.

C Evaluation Metrics

We employ five complementary metrics that probe distinct biological properties of the generated microenvironment. All are computed on the held-out target snapshot.

MSE and MAE. Given N cells with predicted expression vectors $\hat{\mathbf{y}}_i$ and ground-truth vectors $\mathbf{y}_i$,

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_2^2, \qquad \text{MAE} = \frac{1}{N}\sum_{i=1}^{N} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\|_1. \tag{32}$$

MSE penalizes large deviations quadratically and is sensitive to errors in highly expressed marker genes; MAE is more robust to such outliers. Lower is better. Here $\mathbf{y}_i$ is the *observed* target cell

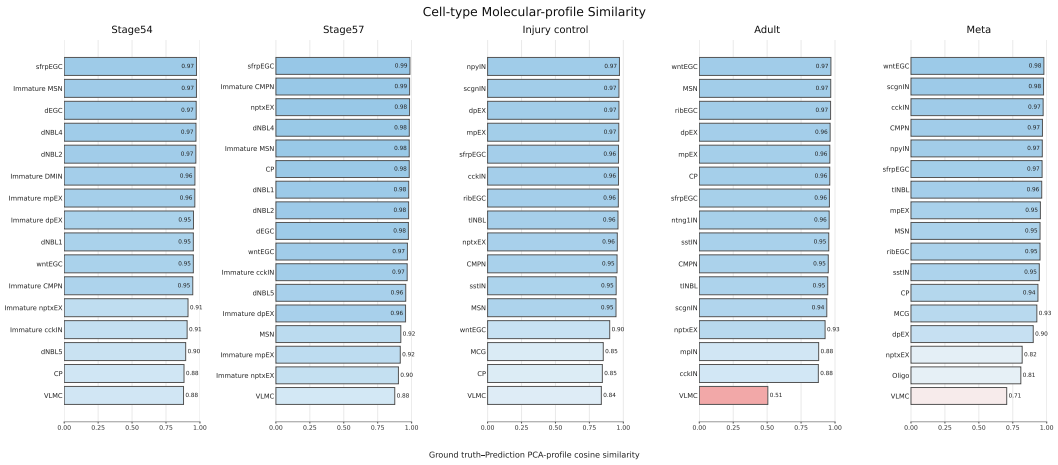


Figure 4: Molecular fidelity across cell types between real and predicted ABD tissues. Bars report the cosine similarity of PCA expression profiles for shared cell types within each target stage; higher values indicate better performance.

matched to the prediction $\hat{\mathbf{y}}_i$ under the shared OT protocol used for all methods; the pairing-free spatial (PSD/SPD, below) and cell-type distributional (Section 4.5) analyses complement these paired errors.

1NN-F1. This metric quantifies whether each generated cell lands in the correct cell-type neighborhood of the real tissue. Every generated cell $\hat{i}$ carries a ground-truth type $c_{\hat{i}}$ inherited from its source correspondence under the OT plan (Eq. (10)). We embed generated and real cells in the same expression space, assign each generated cell the label of its nearest *real* cell, $\hat{c}_{\hat{i}} = c_{\mathrm{NN}(\hat{i})}$ with $\mathrm{NN}(\hat{i}) = \arg\min_{j \in \text{real}} \|\hat{x}_{\hat{i}} - x_j\|_2$, and report the macro-averaged F1 between $\{c_{\hat{i}}\}$ and $\{\hat{c}_{\hat{i}}\}$ over all cell types. A cell thus scores correctly only when it is generated into the local expression neighborhood populated by cells of its own type, so 1NN-F1 rewards both type-resolved fidelity (expression placed in the right region) and type balance (every type recovered, via the macro average), while penalizing collapse onto a few dominant types. Higher is better.

PSD and SPD. Let $\{\hat{\mathbf{p}}_i\}_{i=1}^N$ be predicted coordinates and $\{\mathbf{p}_j\}_{j=1}^M$ ground-truth coordinates:

$$\mathrm{PSD} = \frac{1}{N} \sum_{i=1}^{N} \min_{1 \le j \le M} \|\hat{\mathbf{p}}_i - \mathbf{p}_j\|_2, \tag{33}$$

$$\mathrm{SPD} = \frac{1}{M} \sum_{j=1}^{M} \min_{1 \le i \le N} \|\mathbf{p}_j - \hat{\mathbf{p}}_i\|_2. \tag{34}$$

PSD measures spatial precision: whether each predicted cell lies close to the true tissue geometry. SPD measures spatial recall: whether the true tissue is fully covered by the predicted point cloud. Lower is better for both.

D Additional Spatial Analysis

For fair comparison, all methods score generated cells against held-out target cells under a shared OT matching protocol. To complement these pairing-dependent metrics, we employ pairing-free measures, specifically the Chamfer-style spatial distances (PSD/SPD) and the cell-type distributional analyses (Section 4.5).

Figure 4 evaluates feature consistency between generated and observed states across cell populations. The predicted states maintain high similarity for most populations, while residual discrepancies highlight challenging cell populations where reconstruction remains more difficult.

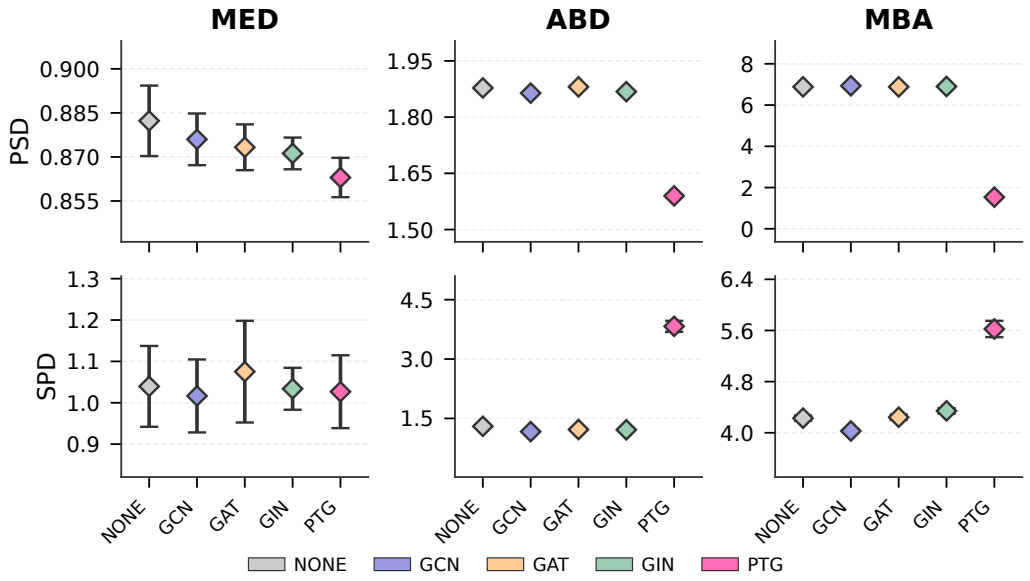


Figure 5: Spatial metrics for different GNN architectures. Lower is better for both metrics.

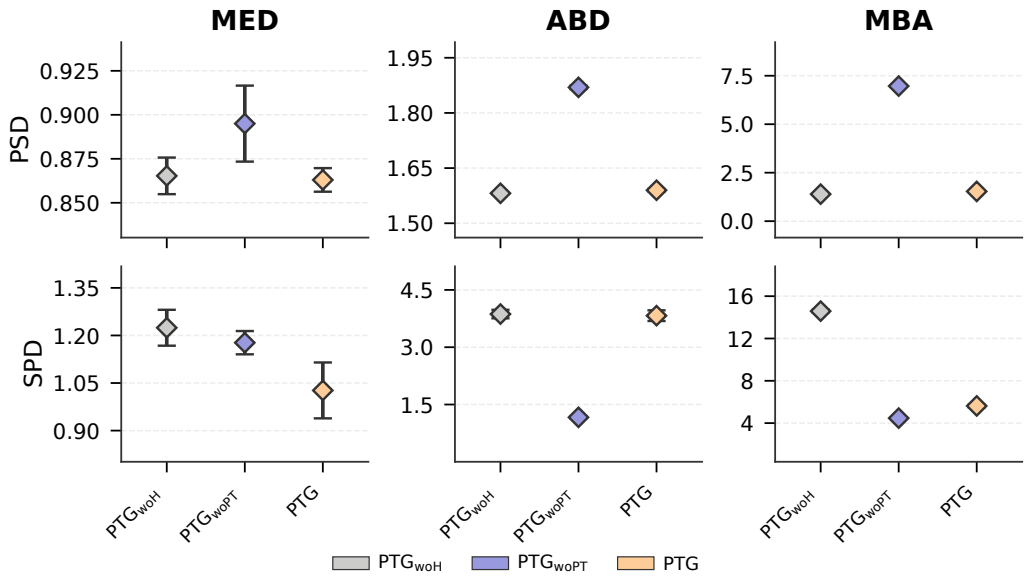


Figure 6: Spatial metrics for the PTG ablation study. Lower is better for both metrics.

Quantitative metrics in Figures 5 and 6 reveal a trade-off between spatial precision (PSD) and coverage (SPD). PTG balances performance on the short MED trajectory but induces spatial mode collapse on longer trajectories like ABD and MBA. Incorporating pseudotime guidance improves both expression accuracy and spatial precision, reducing PSD by up to 77.7% on MBA. This improvement occurs because the directional prior concentrates cells along the dominant developmental axis. This concentration is the coverage side of an accuracy–coverage trade-off: it sacrifices uniform coverage and leaves peripheral regions underrepresented.

Our ablation study confirms this trade-off stems directly from the directionality constraint. Removing pseudotime guidance ($\mathrm{PTG}_{\mathrm{woPT}}$) restores coverage and recovers SPD to 4.471 on MBA, but

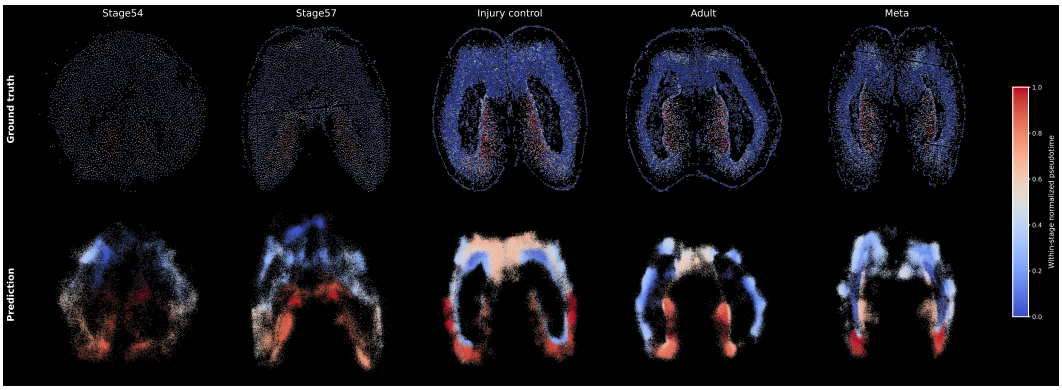


Figure 7: Spatial organization of distribution of developmental pseudotime in real and predicted ABD tissues. Pseudotime is normalized within each stage to $[0, 1]$ to enable the comparison of developmental stages across tissues.

at the expense of a $\sim 33\%$ drop in expression accuracy. Conversely, the high-pass filter mitigates this side effect. Removing the filter ($\mathrm{PTG}_{\mathrm{woH}}$) severely degrades SPD on MBA, confirming that boundary enhancement signals actively counteract spatial over-concentration. Still, achieving high peripheral fidelity under a strong directional prior remains a persistent challenge, which coverage-regularized objectives may help address.

E Supplementary Results under SemiOT-WOT Configuration

As discussed in Section 4.3 of the main text, while the SemiOT-SemiOT coupling configuration remains biologically faithful to non-conservative mass flow (cell growth and death), the SemiOT-WOT configuration forces strict matching at the cell level. This can marginally improve metrics related to pure predictive reconstruction tasks. Here we report the baseline comparison and the structural graph ablation study under the SemiOT-WOT setting.

Table 5: Comparison with the prior state-of-the-art NicheFlow Sakalyan et al. (2026) under the SemiOT-WOT configuration. Lower is better for MSE and MAE; higher is better for 1NN-F1. Bold: best; italic: second-best.

Method	MED			ABD			MBA		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
NicheFlow	**1.3775** ±0.0014	**0.8921** ±0.0004	**0.6634** ±0.0021	*1.5828* ±0.0009	*0.9447* ±0.0003	*0.6278* ±0.0008	*1.5414* ±0.0006	*0.9565* ±0.0002	*0.2847* ±0.0003
$\mathrm{NicheTime}_{\mathrm{NONE}}$	1.6739 ±0.0059	0.9952 ±0.0019	0.6374 ±0.0029	1.9610 ±0.0051	1.0599 ±0.0014	0.6210 ±0.0024	1.7191 ±0.0040	1.0109 ±0.0012	**0.3111** ±0.0015
$\mathrm{NicheTime}_{\mathrm{PTG}}$	*1.4938* ±0.0043	*0.9362* ±0.0017	*0.6584* ±0.0038	**1.4347** ±0.0021	**0.9032** ±0.0006	**0.6544** ±0.0017	**1.2543** ±0.0011	**0.8513** ±0.0004	0.2563 ±0.0010

Table 6: Quantitative comparison of spatial graph architectures under the SemiOT-WOT configuration. Bold: best; italic: second-best.

GNN layer	MED			ABD			MBA		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
NONE	1.6739 ±0.0059	0.9952 ±0.0019	0.6374 ±0.0029	1.9610 ±0.0051	1.0599 ±0.0014	0.6210 ±0.0024	1.7191 ±0.0040	1.0109 ±0.0012	**0.3111** ±0.0015
GCN	1.6355 ±0.0073	0.9842 ±0.0023	**0.6618** ±0.0041	1.9185 ±0.0034	1.0494 ±0.0012	0.6310 ±0.0022	*1.6766* ±0.0033	*0.9966* ±0.0009	0.2998 ±0.0012
GAT	*1.6139* ±0.0062	*0.9762* ±0.0021	0.6463 ±0.0026	*1.8867* ±0.0026	*1.0411* ±0.0009	*0.6401* ±0.0012	1.7822 ±0.0036	1.0294 ±0.0011	0.2971 ±0.0021
GIN	1.6564 ±0.0041	0.9896 ±0.0012	0.6336 ±0.0040	1.9553 ±0.0025	1.0602 ±0.0006	0.6156 ±0.0016	1.7039 ±0.0028	1.0057 ±0.0010	*0.3002* ±0.0010
PTG	**1.4938** ±0.0043	**0.9362** ±0.0017	*0.6584* ±0.0038	**1.4347** ±0.0021	**0.9032** ±0.0006	**0.6544** ±0.0017	**1.2543** ±0.0011	**0.8513** ±0.0004	0.2563 ±0.0010

Table 7: Quantitative ablation study of the PTG layer under the SemiOT-WOT configuration. Bold: best; italic: second-best.

GNN layer	MED			ABD			MBA		
	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1	MSE	MAE	1NN-F1
$\mathrm{PTG}_{\mathrm{woH}}$	*1.5794* ±0.0031	*0.9641* ±0.0012	**0.6957** ±0.0045	*1.4459* ±0.0028	*0.9072* ±0.0010	*0.6652* ±0.0023	*1.2948* ±0.0008	*0.8648* ±0.0002	0.2239 ±0.0002
$\mathrm{PTG}_{\mathrm{woPT}}$	1.6449 ±0.0061	0.9856 ±0.0020	0.6395 ±0.0026	1.8746 ±0.0053	1.0375 ±0.0016	0.6263 ±0.0017	1.6896 ±0.0033	1.0011 ±0.0012	**0.3070** ±0.0018
PTG	**1.4938** ±0.0043	**0.9362** ±0.0017	*0.6584* ±0.0038	**1.4347** ±0.0021	**0.8970** ±0.0008	**0.6659** ±0.0012	**1.2543** ±0.0011	**0.8513** ±0.0004	*0.2563* ±0.0010

Table 5 shows that our advantage over NicheFlow does not depend on the cell-scale coupling. Under SemiOT-WOT, NicheTime$_{\text{PTG}}$ still strictly outperforms NicheFlow on ABD and retains the best regression accuracy on MBA, reducing MSE by 18.6% and MAE by 11.0%, while NicheFlow keeps its advantage only on the short MED trajectory. Crucially, the two couplings differ by at most 4.5% in MSE for NicheTime$_{\text{PTG}}$ on any dataset, and SemiOT-SemiOT is in fact marginally better on ABD. Strict mass conservation therefore buys little predictive accuracy, and the biologically principled configuration adopted in the main text incurs a negligible cost. The graph-free variant again leads on MBA 1NN-F1, reproducing the coverage-precision trade-off of Section 4.4.

Tables 6 and 7 reproduce the architectural conclusions of the main text with near-identical margins. PTG attains the best MSE and MAE on all three datasets, reducing MSE over the graph-free baseline by 10.8%, 26.8% and 27.0% on MED, ABD and MBA, whereas the strongest symmetric competitor gains at most 3.8% and GAT is even worse than using no graph on MBA. The bottleneck is thus not graph capacity but the absence of directionality, since symmetric aggregation over a developmentally heterogeneous niche mixes mature and progenitor representations irrespective of how correspondences are formed. Consistently, removing pseudotime damping remains the dominant failure mode, inflating MSE by 10.1%, 30.7% and 34.7% on MED, ABD and MBA, a penalty that grows with trajectory length as directional errors compound. Removing the high-pass gate costs far less in expression error but governs neighborhood structure in the direction predicted by its role as a boundary detector: it raises 1NN-F1 to 0.6957 on the sharply partitioned MED trajectory while dropping it to 0.2239 on the diffuse MBA trajectory. The contribution of the directed spatial backbone is therefore orthogonal to the choice of cell-scale coupling.

F Algorithms

Both procedures share one backbone: PTG over the source niche's cell graph G^{cell}, encoded *once* into context$_{\text{enc}}$, then a Transformer decoder that cross-attends to it. Throughout, $\mathbf{z} = (\mathbf{x}, \mathbf{p})$ is the joint cell state (expression $\|$ coordinates) and $s \in [0, 1]$ the flow time.

Algorithm 1 NicheTime training

Require: snapshots $\mathcal{D}_{t_0}, \mathcal{D}_{t_1}$; OT variant; network v_θ; rate η; weights λ_x, λ_p

repeat
  sample matched niche pair (u, v) from Π^{niche}; draw cell pairs $(\mathbf{z}^{\text{src}}, \mathbf{z}_1) \sim \Pi^{\text{cell}}_{uv}$ ▷ dual-scale OT, Sec. 3.2
  context$_{\text{enc}} \leftarrow \text{Encoder}\big(\text{PTG}(G^{\text{cell}}) \oplus h_{\text{cond}}\big),\ h_{\text{cond}} = (\mathbf{z}^{\text{src}}, t_0)$
  sample $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),\ s \sim \mathcal{U}(0, 1);\quad \mathbf{z}_s \leftarrow s\,\mathbf{z}_1 + (1 - s)\,\mathbf{z}_0$ ▷ interpolant
  $(\hat{\mathbf{x}}_\theta, \hat{\mathbf{p}}_\theta) \leftarrow \text{Decoder}\big((\mathbf{z}_s, t_1, s),\ \text{cross-attend context}_{\text{enc}}\big)$
  optionally refine $(\hat{\mathbf{x}}_\theta, \hat{\mathbf{p}}_\theta)$ via Post-GNN on dynamic k-NN from $\hat{\mathbf{p}}_\theta$
  $\mathcal{L} \leftarrow \lambda_x \|\hat{\mathbf{x}}_\theta - \mathbf{x}_1\|_2^2 + \lambda_p \|\hat{\mathbf{p}}_\theta - \mathbf{p}_1\|_1$ ▷ GLVFM; $\times$ mass if UOT
  $\theta \leftarrow \theta - \eta\, \nabla_\theta \mathcal{L}$
until converged
return trained network v_θ

G Datasets and Implementation Details

G.1 Datasets

We benchmark NicheTime on three time-resolved spatial transcriptomics datasets (Table 8) deliberately chosen to span a comprehensive range of biological scenarios, measurement modalities, and dataset scales.

G.2 Preprocessing

Each snapshot is normalized and log-transformed; expression is reduced to a 50-dimensional embedding by PCA on highly variable genes. Spatial coordinates are rescaled to a common unit range. Cell-type labels are adopted directly from the source annotations. Pseudotime is estimated by an upstream trajectory-inference tool and rank-normalized as in Eq. (12) of the main text.

Algorithm 2 NicheTime inference

Require: trained v_θ; source niche cell graph G^{cell} with $\mathbf{z}^{\text{src}}$; ODE steps T; rebuild interval κ

context$_{\text{enc}} \leftarrow$ Encoder$\big(\text{PTG}(G^{\text{cell}}) \oplus h_{\text{cond}}\big),\ h_{\text{cond}} = (\mathbf{z}^{\text{src}}, t_0)$ ▷ computed once
$\mathbf{z} \leftarrow \mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ▷ init state (noise, $s = 0$)
for $n = 0$ to $T - 1$ **do**
  $s \leftarrow n/T$
  if Post-GNN used **and** $n \bmod \kappa = 0$ **then** rebuild dynamic k-NN graph from coords of $\mathbf{z}$
  end if
  $(\hat{\mathbf{x}}_\theta, \hat{\mathbf{p}}_\theta) \leftarrow$ Decoder$\big((\mathbf{z}, t_1, s),$ cross-attend context$_{\text{enc}}\big)$, then Post-GNN if used
  $\mathbf{v} \leftarrow (\hat{\mathbf{z}}_\theta - \mathbf{z})/(1 - s)$ ▷ field from endpoint $\hat{\mathbf{z}}_\theta = (\hat{\mathbf{x}}_\theta, \hat{\mathbf{p}}_\theta)$
  $\mathbf{z} \leftarrow \mathbf{z} + \frac{1}{T}\,\mathbf{v}$ ▷ Euler step
end for
return $\mathbf{z}$ as prediction at t_1

Table 8: Summary of benchmark datasets.

Dataset	Abbr.	Time Points	Cell Types	Cells	Technology
Mouse Embryogenesis	MED	3	24	54K	Stereo-seq
Axolotl Brain Development	ABD	6	33	36K	Stereo-seq
Mouse Brain Aging	MBA	20	18	1.5M	MERFISH

G.3 Implementation Details

Pseudotime computation. All pseudotime values used across our benchmark datasets are obtained via **Monocle 3** Cao et al. (2019), a trajectory-inference framework that learns a principal-graph-based trajectory in low-dimensional cell-state space and assigns each cell a continuous developmental time coordinate. The raw pseudotime scores produced by Monocle 3 are subsequently rank-normalized as described in Eq. (12) of the main text before being provided to the model.

On the use of pseudotime. Pseudotime enters the model only as a *directional prior* on message passing (Eq. (13)): it orients edges and damps backward propagation in pseudotime, but it never supplies the target expression. The learning signal comes entirely from OT correspondences to the *observed, held-out* target snapshot, while pseudotime is computed on the source snapshot alone. The model is thus told which way time flows rather than what the future state is, so no information about the prediction target leaks into the guidance. This mirrors standard practice in RNA velocity and trajectory inference, where a directionality estimate conditions the generative dynamics without determining them.

H Limitations

NicheTime has three main limitations, each suggesting a direction for future work.

Trade-off between accuracy and coverage. By concentrating generated cells along the dominant developmental axis, the soft pseudotime prior improves expression fidelity but reduces peripheral coverage and cell-type neighborhood diversity on longer trajectories such as ABD and MBA (Appendix D). The high-pass gate only partially offsets this; coverage-regularized objectives are left to future work.

Training and evaluation between observed endpoints. NicheTime learns a transport between two observed snapshots rather than a continuous biological clock, so the generative flow time $s \in [0, 1]$ is distinct from biological time (Appendix A). Extrapolation to unobserved stages, such as a held-out intermediate timepoint, is left to future work.

Dependence on upstream pseudotime. The directional prior relies on a pseudotime ordering from an upstream tool (Monocle 3; Appendix G), and being soft (finite β) it tolerates local ordering error without leaking target information. A global scalar pseudotime cannot capture branch-specific order in bifurcating lineages; robustness to noise and alternative estimators remains to be evaluated.

Table 9: NicheTime hyperparameters used in the main experiments.

Group	Hyperparameter	Value
Data	expression PCs	50
	spatial coordinate dim.	2
Niche graph	radius cutoff r	0.15
	spatial neighbors k	10
	abundance threshold τ	0.1
Backbone	token embedding width	128 (120 rounded)
	token split (x, t, p)	$40/40/40$
	feedforward width	256
	attention heads	4
	encoder / decoder layers	2 / 2
	pre-GNN layers	1
	post-refinement GNN	enabled
Flow matching	ODE solver	Euler
	integration steps	20
	feature loss weights λ_x	1.0
	position loss weights λ_p	1.0
Optimization	optimizer	AdamW
	learning rate	10^{-3}
	weight decay	10^{-5}
	GNN learning-rate scale	0.1
	GNN warmup steps	2000